

DISSERTATION
FROM COGNITION 2 APPLICATIONS
Bridging the Gap Between Formal Education & AI

Thesis approved by
the Department of Computer Science
of the Technische Universität Kaiserslautern
for the award of the Doctoral Degree

DOCTOR OF ENGINEERING
(DR.-ING.)

to

Junaid Younas

Date of the Defense: 2022-12-21
Dean: Prof. Dr. Jens Schmitt
Reviewer: Prof. Dr. Paul Lukowicz
Reviewer: Prof. Dr. Faisal Shafait (NUST, Islamabad, Pakistan)

Junaid Younas: *From Cognition 2 Applications* – Bridging the Gap Between Formal Education & AI

SUPERVISOR:
Prof. Dr. Paul Lukowicz

CONTACT INFORMATION:
junaid.younas@dfki.de

There is only one place where your existence matter without any ifs
and buts, where LIFE begins and LOVE never ends, its FAMILY.

Dedicated to the loving memories of My Father Muhammad Younas

1965 – 2021

&

My beloved son, Muhammad Hassan, whom I still have to hold in
my arms.

ABSTRACT

This thesis focuses on novel methods to establish the utility of wearable devices along with Machine Learning (ML) and pattern recognition methods for formal education and address the open research questions posed by existing methods. Firstly, state-of-the-art (SotA) methods are proposed to analyse the cognitive activities in the learning process, i.e., reading, writing, and their correlation. Furthermore, this thesis presents real-time applications in wearable space as an experimental tool in Physics education and an air-writing system.

There are two critical components in analysing the reading behaviour, i.e., **WHERE** a person looks at (gaze analysis) and **WHAT** a person looks at (content analysis). This thesis proposes novel methods to classify the reading content to address the WHAT AT component. The proposed methods are based on a hybrid approach, which fuses the traditional Computer Vision (CV) methods with Deep Neural Networks (DNNs). When evaluated on publicly available datasets, these methods yield SotA results to define the structure of the document images. Moreover, extensive efforts were made to refine and correct the ICDAR2017 Page Object Detection (ICDAR2017-POD) dataset and a completely new Figure and Formula Detection (FFD) dataset.

Traditionally, handwriting research focuses on character and number recognition without looking into the type of writing, i.e., text, math, and drawing. This thesis reports multiple contributions for online handwriting classification. First, it presents a public dataset for online handwriting classification **OnTabWriter**, collected using iPen and an iPad. In addition, a new feature set is introduced for online handwriting classification to establish the benchmark on the proposed dataset to classify handwriting as plain text, mathematical expression, and plot/graph. An ablation study is made to evaluate the performance of the proposed feature set in comparison to existing feature sets. Lastly, this thesis evaluates the importance of context for online handwriting classification.

Analysing reading and writing activities individually is insufficient to provide insights to identify the student's expertise unless their correlations are analysed. This thesis presents a study where reading data from wearable eye-trackers and writing data from sensor pen are analysed together in correlation to correlate the expertise of the users in Physics education with their actual knowledge. Initial results show a strong correlation between an individual's expertise and understanding of the subject.

Augmented Reality (AR) & Virtual Reality (VR) applications can play a vital role in making classroom environments more interactive and engaging both for teachers and learners. To validate the hypothesis, different applications are developed and evaluated. First, smart glasses are used as an experimental tool in Physics education to help the learners perform experiments by providing assistance and feedback on Head Mounted Display (HMD) in understanding acoustics concepts. Second, a real-time application of air-writing with the finger on an imaginary canvas using a single Inertial Measurement Unit (IMU) as the Finger Air Writing System (FAirWrite) system is also presented. FAirWrite system is further equipped with Deep Learning (DL) methods to classify the air-written characters.

ACKNOWLEDGMENTS

First, I would like to thank and pay my gratitude to the Lord of the universe Almighty Allah, who enabled me to complete and write this dissertation, and His countless blessings during this formidable journey. Peace, Prayers, and Blessings upon the only perfect man ever lived, my lord, the greatest leader of all times Muhammad (S.A.W) and his family.

*The ultimate Creator
of Universe*

I want to thank Prof. Dr Paul Lukowicz, my advisor, for the opportunity to conduct my research in the Embedded Intelligence group at German Research Center for Artificial Intelligence (DFKI) Kaiserslautern. This dissertation would not have been possible without his ongoing support, supervision, feedback, freedom to work independently, being available all the time for his input, and critical analysis to keep me following the right direction. I will be indebted to Paul for the rest of my life for his role in multiple ways during this phase of my life. I wish I could thank him enough. I would like to thank Prof. Dr Faisal Shafait for his role and help from the start of this work to accepting the role as external supervisor towards the end of this thesis. Finally, I thank Prof. Dr Leo van Waveren for agreeing to chair my Ph.D. commission and providing valuable external feedback.

*Supervisors &
Commission*

I want to thank Dr Sheraz Ahmed for going out of the way to help me out during this journey. To me, you are always a mentor. I learned a lot from you during our technical discussions, guidance, and how to stay focused to achieve the goals, both at professional and personal levels. I have developed a special relationship with you in the last few weeks of my stay in Germany. I would also like to thank Dr Imran Malik for his role during my Ph.D. I thank all my friends here in DFKI and Germany for helping me out and being there to spend quality time over the last decade, making me feel at home, miles away from my family. I have gathered countless memories during our joint dinner sessions, travels, and coffee chats. I would also like to thank all my students whom I had the honor to supervise or work on their projects and thesis. These experiences have been a great learning opportunity for myself.

Friends & Students

A special thanks to my mother, Jamila Younas, for all the sacrifices she made for my comfort, right from the moment I was born, from the day you slept on the wet side of the bed to put me sleep on the drier side, to the day you sold your necklace to ensure my comfort leaving for Germany for the first time, to this day. A mere thank-you is not enough for all your efforts. All the prayers for My late father

Parents

Muhammad Younas, whose vision started with holding my finger to teach me how to walk on a journey to follow this path to make it to this day to write these lines; one of the most important and biggest achievement of my life so far. May YOU rest in ETERNAL PEACE.

DFKI & colleagues

I would also like to thank the [DFKI](#), my colleagues, and my office mates. Jane has been a real help throughout this journey for her role in managing all the administrative stuff on my behalf, her help in reviewing my papers for the English language and helping me out with my deficiencies in German. I am thankful to all my colleagues at the Embedded Intelligence group for their support and help in creating a pleasant and inspiring work environment. You have always been welcoming and accommodating.

Family

Last but not least, my better half Naila Junaid, I can't thank enough for her sacrifices, support, and being there every moment, although we were physically thousands of miles apart. She also took on my responsibilities in raising our kids, when my father was ill and on countless other occasions. I am grateful to my kids, Fatima and Hussnain, for understanding my position and spending most of their childhood in the absence of their father. I am thankful to my siblings for their support, prayers, and extended help during my Ph.D. journey.

I am thankful to everyone I forgot to mention for his help, support, and prayers.

— Thank you, Junaid

GRANTS

This work has been supported partially by the Higher Education Commission ([HEC](#)), Pakistan, under their scholarship program for masters leading to Ph.D.

PUBLICATIONS AS PART OF THIS THESIS

Parts of the research and material (including figures, tables and algorithms) in this thesis have already been published in:

J. Kuhn, P. Lukowicz, M. Hirth, A. Poxrucker, J. Weppner, and J. Younas. "gPhysics—Using Smart Glasses for Head-Centered, Context-Aware Learning in Physics Experiments." In: *IEEE Transactions on Learning Technologies* 9.4 (2016), pp. 304–317.

J. Younas, M. Z. Afzal, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "D-StaR: A Generic Method for Stamp Segmentation from Document Images." In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 01. 2017, pp. 248–253.

J. Younas, S. Fritsch, G. Pirkl, S. Ahmed, M. I. Malik, F. Shafait, and P. Lukowicz. "What Am I Writing: Classification of On-Line Handwritten Sequences." In: *Intelligent Environments (Workshops)*. Vol. 23. Ambient Intelligence and Smart Environments. IOS Press, 2018, pp. 417–426.

S. Bian, V. F. Rey, J. Younas, and P. Lukowicz. "Wrist-Worn Capacitive Sensor for Activity and Physical Collaboration Recognition." In: *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 2019, pp. 261–266.

J. Younas, S. T. R. Rizvi, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "FFD: Figure and Formula Detection from Document Images." In: *2019 Digital Image Computing: Techniques and Applications (DICTA)*. 2019, pp. 1–7.

J. Younas, S. A. Siddiqui, M. Munir, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "Fi-Fo Detector: Figure and Formula Detection Using Deformable Networks." In: *Applied Sciences* 10.18 (2020).

J. Younas, H. Margarito, S. Bian, and P. Lukowicz. "Finger Air Writing - Movement Reconstruction with Low-Cost IMU Sensor." In: *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. MobiQuitous '20. Darmstadt, Germany: Association for Computing Machinery, 2020, pp. 69–75.

J. Younas, M. I. Malik, S. Ahmed, F. Shafait, and P. Lukowicz. "Sense the pen: Classification of online handwritten sequences (text, mathematical expression, plot/graph)." In: *Expert Systems with Applications* 172 (2021), p. 114588.

J. Younas, H. Margarito, and P. Lukowicz. "FAirWrite - Movement Reconstruction and Recognition Using a Low-cost IMU." In: *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 2022, pp. 298–303.

J. Younas and P. Lukowicz. "Cognitive Ability Classification using On-body Sensors." In: *[UbiComp/ISWC '22 Adjunct] Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Cambridge, United Kingdom: Association for Computing Machinery, 2022.

CONTENTS

Publications as Part of this Thesis	ix
List of Figures	xiv
List of Tables	xvi
Acronyms	xvii
I Introduction & Background	
1 Introduction	3
1.1 Motivation	3
1.2 Research Questions	5
1.3 Contributions	6
1.4 Overview	8
2 State of the Art	11
2.1 Page Object Detection	11
2.1.1 Traditional Approaches	12
2.1.2 Deep Learning Approaches	13
2.2 On-line Handwriting Classification	15
2.2.1 Traditional Methods	16
2.2.2 Deep Learning Methods	18
2.3 Applications of Wearable's in Formal Education	19
II Learning Activity Classification & Performance Evaluation	
3 Content Classification: Off-line Modality	25
3.1 Motivation	28
3.2 Related Work	29
3.3 Methodology	33
3.3.1 Fi-fo image representation	33
3.3.2 Figure and Formula Detection (FFD)	35
3.3.3 Figure and Formula (Fi-fo) Detector	37
3.4 Datasets	39
3.4.1 ICDAR2017 Page Object Detection (ICDAR2017-POD)	40
3.4.2 Figure and Formula Detection (FFD) Dataset	42
3.4.3 PublayNet	43
3.4.4 Evaluation Protocol	44
3.5 Results and Discussions	44
3.5.1 Results of FFD approach	44
3.5.2 Results of Fi-fo Detector approach	47
3.5.3 Ablation Study	48
3.5.4 Discussions	50
4 Content Classification: online modality	53
4.1 Motivation	56
4.2 Related work	58
4.3 Methodology	60

4.3.1	Data Collection and Pre-processing	60
4.3.2	Feature Extractor	61
4.3.3	Classifiers	64
4.4	Dataset & Evaluation protocol	66
4.4.1	Overview	66
4.4.2	Dataset	67
4.4.3	Feedback	67
4.4.4	Evaluation Protocol	68
4.4.5	Optimization Parameters	69
4.5	Results and Discussion	69
4.5.1	Results	69
4.5.2	Discussion	73
5	Cognitive Abilities: A performance analysis	79
5.1	Motivation	82
5.2	Related Work	83
5.3	System Overview	85
5.3.1	Meta information	86
5.3.2	Digital pen	86
5.3.3	Eye-tracker	87
5.4	Data Collection and organization	88
5.4.1	Data Collection	88
5.4.2	Data Organization	88
5.4.3	Feature Extraction	89
5.5	Initial data Exploration	90
III Applications of Wearable Sensors in Classrooms		
6	FAirWrite: Finger Air-writing System	97
6.1	Motivation	100
6.2	Related Work	101
6.3	FAirWrite System	103
6.3.1	Finger-worn Sensor Design	104
6.3.2	Trajectory Reconstruction	105
6.3.3	Classifier	107
6.3.4	Finger Air Writing System (FAirWrite) User interface	108
6.4	Experiment Set-up	108
6.4.1	Data Collection	108
6.4.2	Evaluation Protocol	110
6.5	Results and Discussions	111
6.5.1	User Quality Appreciation	111
6.5.2	Model-based evaluation	111
6.5.3	Discussion	113
IV Supplementary Work		
7	dStaR	119
7.1	Motivation	121
7.2	Related Work	122

7.3	dStaR: The Presented System	124
7.3.1	Domain Adaptation and Transfer Learning	125
7.3.2	VGG-Net	125
7.3.3	Fully Convolutional Networks (FCNs)	125
7.4	Evaluation	127
7.4.1	Dataset	127
7.4.2	Evaluation Protocol	127
7.4.3	Results and Discussion	128
8	Applications of Wearables: gPhysics & WristSense	133
8.1	gPhysics	136
8.1.1	Motivation	136
8.1.2	Experimentation & Study Design	137
8.1.3	Results	138
8.2	WristSense	140
8.2.1	Motivation	140
8.2.2	Related Work	141
8.2.3	Capability Exploration	142
8.2.4	Collaborative Work Monitoring	145
v	Conclusion	
9	Summary	151
9.1	Achievements and Discussion	151
9.2	Scope and Outlook	155
	Bibliography	159
	Index	187
	Academic Curriculum Vitae: Junaid Younas	189

LIST OF FIGURES

Figure 1.1	Scope and contributions of this thesis	5
Figure 3.1	Fi-Fo Detector Outline	34
Figure 3.2	Fi-Fo Image representation	34
Figure 3.3	FFD pipeline with its components	35
Figure 3.4	Examples from ICDAR2017-POD dataset	40
Figure 3.5	Problems with ICDAR2017-POD dataset	41
Figure 3.6	Examples from FFD dataset	43
Figure 3.7	FFD driven results using Faster-RCNN	46
Figure 3.8	FFD driven results using mask-RCNN	46
Figure 3.9	Fi-Fo detector driven visual results on corrected dataset	49
Figure 3.10	Annotation problems highlighted by Fi-Fo detector	51
Figure 4.1	System overview	60
Figure 4.2	Data collection set-up	61
Figure 4.3	Bagging classifier	65
Figure 4.4	Content distribution in a dataset	68
Figure 4.5	Person dependent results	71
Figure 4.6	Person independent results	72
Figure 4.7	Example of perfection	72
Figure 4.8	Visual Results for graph class	75
Figure 4.9	Visual results for Math class	75
Figure 4.10	Classification output of complex page	75
Figure 4.11	Ablation study on proposed feature set	76
Figure 4.12	Individualised feature evaluation	78
Figure 5.1	On-body sensors to monitor cognitive activities	86
Figure 5.2	Pen sensors data	87
Figure 5.3	Data collection using eye trackers	88
Figure 5.4	Expertise estimation based on cognitive abilities	92
Figure 6.1	FAirWrite System in real life application	97
Figure 6.2	Examples of Air-writing	101
Figure 6.3	FAirWrite system architecture	104
Figure 6.4	Finger-worn sensor design	105
Figure 6.5	Sensor motion with respect to air motions	105
Figure 6.6	Adaptable canvas definition	106
Figure 6.7	OS-CNN model architecture	108
Figure 6.8	FAirWrite user interface	109
Figure 6.9	Examples of trajectory reconstruction using FAir-Write	110
Figure 6.10	Confusion matrix of OS-CNN classifier	112
Figure 6.11	Obvious confusions in model-based evaluation	113

Figure 6.12	Failed examples of trajectory reconstruction . . .	114
Figure 7.1	Examples of stamp variants	121
Figure 7.2	dStaR architecture	124
Figure 7.3	dStaR: an overview	126
Figure 7.4	Overlapping stamps: visual results	129
Figure 7.5	Failed examples of dStaR	130
Figure 8.1	gPhysics in application with Google glass . . .	137
Figure 8.2	Wondering & Curiosity using smart glasses . .	139
Figure 8.3	Cognitive load while using smart glasses . . .	139
Figure 8.4	Experimentation time: A comparison	140
Figure 8.5	Touch a chair with prototypes at a chair and on wrist	143
Figure 8.6	Approaching a door with prototype attached at the doorknob and on wrist	144
Figure 8.7	Walking by detection with prototypes attached on wrists	144
Figure 8.8	Collaborative work site	146
Figure 8.9	process of the collaborative work	147

LIST OF TABLES

Table 3.1	Overview of ICDAR2017-POD with its corrected version	42
Table 3.2	Overview of FFD dataset	43
Table 3.3	Overview of FFD driven results	45
Table 3.4	Comparative results of Fi-Fo detector	47
Table 3.5	Fi-Fo detector results on ICDAR2017-POD (corrected) dataset	48
Table 3.6	Ablation study on Fi-Fo image representation	49
Table 4.1	Overview of the proposed feature set	64
Table 4.2	Personal preferences:Notebooks versus Tablets	68
Table 4.3	Overview: Person dependent results	70
Table 4.4	Overview: Person independent results	71
Table 4.5	Detailed Person dependent results	72
Table 4.6	Detailed person independent results	74
Table 4.7	Results with context information	74
Table 4.8	Relevance based feature selection	77
Table 6.1	User quality appreciation results	111
Table 6.2	Model-based Evaluation	112
Table 7.1	dStaR results on overlapping stamps	129
Table 7.2	dStaR results on coloured stamps	130
Table 7.3	dStaR results on black stamps	130
Table 7.4	DStaR results on random test set	131

ACRONYMS

DFKI	German Research Center for Artificial Intelligence
HEC	Higher Education Commission
AR	Augmented Reality
VR	Virtual Reality
MR	Mixed Reality
AI	Artificial Intelligence
PoI	Point of Interest
DNN	Deep Neural Network
Fi-fo	Figure and Formula
dStaR	Deep Stamp Recognition
HMD	Head Mounted Display
FCN	Fully Convolutional Network
ML	Machine Learning
DL	Deep Learning
FFD	Figure and Formula Detection
ICDAR ₂₀₁₇ -POD	ICDAR ₂₀₁₇ Page Object Detection
FAirWrite	Finger Air Writing System
POD	Page Object Detection
CV	Computer Vision
IMU	Inertial Measurement Unit
SotA	state-of-the-art
OCR	Optical Character Recognition
CCA	Connected Component Analysis
PDF	Portable Document Format
HMM	Hidden Markov Model
CRF	Conditional Random Field
CNN	Convolutional Neural Network
SSD	Single Shot Detector
YOLO	You only Look Once
WAP	Watch, Attend, and Parse
SVM	State Vector Machine
KNN	K Nearest Neighbour

RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
BLSTM	Bidirectional Long-Short Term Memory
QR	Quick Response
FPN	Feature Pyramid Network
RCNN	Region Based Convolutional Neural Networks
RPN	Region Proposal Networks
FFT	Fast Fourier Transform
MSE	Mean Square Error
NLPR	National Laboratory of Pattern Recognition
PAL	Pattern Analysis and Learning
IoU	Intersection over Union
mAP	mean Average Precision
AP	Average Precision
ResNet	Residual Network
RoI	Region of Interest
NMS	non-maximum suppression
GPU	Graphical Processing Unit
DCN	Deformable Convolutional Network
RFCN	Region-based Fully Convolutional Networks
TP	true positives
FP	false positive
FN	false negatives
RGB	Red, Green, Blue
STEM	Science, Technology, Engineering, and Mathematics
GDTW	Gaussian Dynamic Time Warping
DTW	Dynamic Time Warping
RBF	Radial Basis Function
MRF	Markov Random Field
RF	Random Forest
DT	Decision Tree
ET	Extra Tree
GBM	Gradient Boosting Machine
GRU	Gated Recurrent Unit
USA	United State of America
SNE	Distributed Stochastic Neighbour Embedding

HCI	Human Computer Interaction
GUI	Graphical User Interface
MEMS	Micro Electro Mechanical System
SI	International System of Unit
API	Application Programming Interface
OS	Omni-scale
DoF	Degrees of Freedom
ASCII	American Standard Code for Information Interchange
PCA	Principal Component Analysis
ISODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
FC	Fully Connected
HBC	Human Body Capacitance
IG	Intervention Group
CG	Control Group
ANCOVA	Analysis of Covariance

Part I

INTRODUCTION & BACKGROUND

INTRODUCTION

As you start to walk on the way, the way appears....

M. Rumi

1.1 MOTIVATION

Applications of wearable sensors vary from activity monitoring to activity recognition and classification in various fields to assist the process for improved performance and enhanced skills. These advances in wearable technology open up a huge opportunity to reform formal education with the vision toward smart and personalized classrooms to formulate the instructions and interactions tailored to student's individual needs and strengths. This thesis explores the potential of Artificial Intelligence (AI)-based methods for the classification of cognitive abilities to bridge the gaps in formal education. Furthermore, Augmented Reality (AR) and Virtual Reality (VR) scenarios applications are considered to aid the learning process with the goal of better engagement and enhanced learning experience to envision smart classrooms.

AI plays a vital role in reforming formal education

Learning is a continuous process of stimulating cognitive abilities by involving the learners in different activities, i.e., reading, writing, and correlation between the two (observation). Reading activity is one of the fundamental academic skills to instigate the analytical process, which helps develop the mind. A recent research study [99] established the importance of reading behaviour analysis as university-level reading completely differs from school-level reading because of requirements of deeper analysis, critical thinking, and problem-solving skills. Reading activity depends on two important factors (i). Reading content (WHAT you read) and (ii). Gaze (WHERE you look). Traditionally, the reading behaviour is analysed by focusing only on the WHERE AT component using gaze tracking methods [112, 117, 136] but completely neglecting the WHAT AT component. This thesis focuses on the WHAT AT factor to signify the importance of content during reading activity and its relevance in reading behaviour analysis. Combining content analysis with gaze tracking will help to highlight the Point of Interests (PoIs) and their relevancy to analyse the student's progress for the given task. Automatic information extraction from document page images requires the detection and un-

Content analysis is fundamental for reading activity evaluation

derstanding of page objects such as tables, figures, formulas, etc. To address the problem, state-of-the-art (SotA) and novel Deep Neural Network (DNN) methods are leveraged to define the structure of reading material, processed as document images (printed or digital), by highlighting figures and formulas. Reading material is processed as document images, whether in printed or digital form, and the content is broadly categorised as text, mathematical expressions, and figures. This thesis proposes novel and generic methods to define the structure of document images (printed or digital) by highlighting figures and formulas.

Looking into the type of writing provides useful insights about writing activity

Writing activity helps to keep the record/notes of an individual's learning for future use. It stimulates the cognitive process to express and convey the understanding of the task at hand. Traditionally, Writing activity analysis was limited to handwriting recognition [16, 146, 249], forgery detection [78, 153, 155], and verification [58, 122]. Only a little has been done in the field of online handwriting classification to look into the type of writing, i.e., writing text, producing mathematical formulas, and/or drawing plots/graphs. Online handwriting classification task is far from trivial, as it involves inter-person and intra-person variations both in temporal and spatial domains. It is very important to evaluate the type of writing by looking into what is being written, to process and investigate the handwriting activity to evaluate learning progress in formal education. To fill the gap, this thesis proposes a novel and comprehensive feature set for online handwriting classification with an ablation study on multiple Machine Learning (ML) and Deep Learning (DL) classifiers to establish the efficacy of the proposed feature set. Furthermore, a completely new dataset is made publicly available to enable the handwriting research community to research and advances in the online handwriting classification domain. Another important factor in online handwriting classification is context information; using context information with ML classifiers can significantly improve the outcome of the classification process, an important contribution of this thesis for online handwriting classification. Online handwriting classification is an important step in evaluating handwriting activity from a formal education perspective for evaluation and feedback estimation for every learner. It can also serve as a preliminary step for handwriting recognition systems to assist them in better performance by filtering the input data for the type of writing they are aimed for, i.e., text recognition and formula recognition.

Wearable sensors are a way forward to monitor classroom activities

It has been established in multiple research studies that when learners are provided with feedback on their performance during learning activities results in substantial gains [25, 109, 280]. Incorporating modern technological developments in classrooms to innovate the interactions and instructions to strengthen the feedback mech-

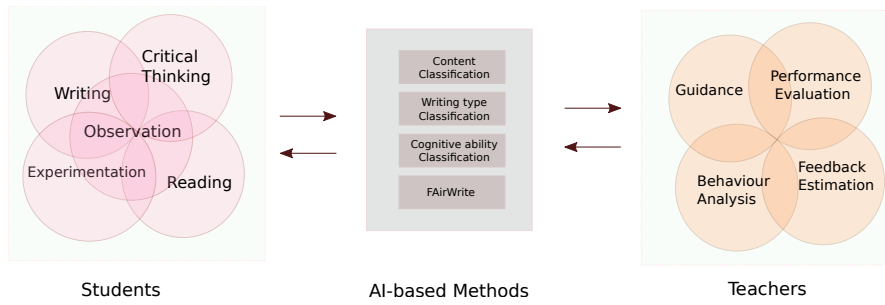


Figure 1.1: Scope and research contributions of this thesis

anisms based on individual strengths and weaknesses is further emphasized. Recently different methods have been presented to estimate the student's self-confidence estimation to provide feedback based on reading activity, handwriting activity, multiple choice questions, and stress estimation using gaze tracking in literature [74, 111, 113, 170]. There is an open research area to estimate the feedback, expertise and confidence estimation, and performance assessment in the classrooms based on the correlation of cognitive abilities, i.e., reading and writing skills. A study is presented in this thesis to estimate the feedback based on expertise and performance for the individual tasks and the whole activity by assessing the reading behaviour, writing behaviour, and correlation between the two. The presented study also helps to highlight the strengths and weaknesses of the learners for given tasks.

The world is focusing on building meta-verse infrastructure with the goal to connect human beings with new immersive and imaginary worlds. These advancements open up the opportunity to develop applications to cater to the educational needs of the students and prepare the teachers to connect the physical world with technological leaps in AR and VR [63, 164, 171]. Multiple applications are presented to establish the utility of AR and VR in classrooms, online learning, in laboratories to assist the students in performing the learning activities [23, 42, 189, 252], learning while interacting, and learning by experience. This thesis uses the "gPhysics" application which uses Google Glass to enable the students to perform experiments to understand the acoustic principles by providing real-time assistance and feedback on Head Mounted Display (HMD) to evaluate the impact of smart glasses on learning outcomes. Another application of Mixed Reality (MR) in combination with AI in the classroom to air-write using finger gestures and then classifying these writing snippets onto digits and characters using Deep Learning (DL) with the SoTA results.

*Mixed reality is
future of formal
education*

1.2 RESEARCH QUESTIONS

This dissertation focuses on the basic question of **How to incorporate the Artificial Intelligence (AI) in combination with wearable sen-**

AI can bridge the gaps in teacher-student relationship

sors in formal education with the goal to strengthen the teachers-learners relationship by bridging the gaps between AI and formal education. One possible way forward in this direction is by monitoring the routine cognitive activities for deeper insights (learning analytics) and assisting them in performing these activities for better engagement (applications & tools). These advancements will contribute to an enhanced learning experience, improved learning outcomes, and provide insights about the progress of learning activities. Figure 1.1 presents a brief overview of this dissertation's scope and research contributions in a broader spectrum. The main research question can be further elaborated by decomposing it into the following two sub-questions, each concerning the desired goals:

1. **Question:** How Artificial Intelligence (AI) be used to analyse and evaluate the cognitive activities in the learning process?

AI-based methods can be used for cognitive ability classification

Goal: Development and implementation of AI-based methods to enable the assessment of the cognitive activities, i.e., reading, writing, and their correlation in the classroom. Proposing novel hybrid methods to combine state-of-the-art (SotA) DNNs methods with traditional Computer Vision (CV) techniques to define the structure of document images. Classification of page objects structurally defines the documents, which enables to analyse the reading behaviour by combing WHAT AT and WHERE AT components of the reading activity. Propose methods for on-line handwriting classification using machine-based knowledge to provide insights about the writing activity. Finally, explore the possibilities to correlate these cognitive activities for performance evaluation and feedback estimation.

2. **Question:** How to integrate bearable computing in formal education to aid the learning process?

Wearable sensors present great potential as an interactive tool in classrooms

goal: Exploring venues to use wearable sensors as interactive tools to assist the learners in performing cognitive activities for better engagement and improve the overall learning experience. Studying the utility of smart glasses and HMDs as an experimental tool in Physics education. Enabling the users to perform the writing gestures in the air to incorporate AR and VR scenarios in classrooms without affecting the natural cognitive process.

1.3 CONTRIBUTIONS

The following are the major contributions of this thesis.

1. Traditional reading behaviour analysis methods focus on gaze data only without inspecting the content and its relevancy in

the reading text. This thesis presents methods to bridge this gap by introducing Figure and Formula (**Fi-fo**) Detector to detect the page objects from document images. In contrast to existing methods, the proposed method is based on a hybrid approach that fuses the traditional Computer Vision (**CV**) based image representation with state-of-the-art (**SotA**) Deep Neural Networks (**DNNs**) to extract the figures and formulas from documents images. **Fi-fo** Detector helps to define the structure of documents by detecting figures and formulas with an f1-score of 0.954 and 0.922, to address the open research question in the reading behaviour analysis. Defining the structure of the document enables the existing methods to incorporate content analysis techniques to highlight the Point of Interests (**PoIs**) and their contextual relevancy.

Figure and Formula (Fi-fo) Detector defines the structure of document images by classifying page objects

2. It is important to have quality datasets to drive active research in content classification and analysis. However, during the evaluation of **Fi-fo** detector, several inconsistencies were found in the original **ICDAR2017-POD** dataset. Therefore, in this thesis, explicit efforts were made to refine and correct the annotations of **ICDAR2017-POD** dataset. Moreover, a new dataset named Figure and Formula Detection (**FFD**) is also curated and made publicly available for Page Object Detection (**POD**) to define the structure of document images.
3. Writing is one of the most important and common activities in the classroom (and much still needs to be done for handwriting classification). Online handwriting classification is a task far from trivial, as it involves interpersonal and intra-personal variations; online handwriting classification remains an open research area. To bridge this gap, this work presents a newly collected dataset **onTabWriter** using iPad ¹ and Apple pencil ², consisting of 12,139 natural handwriting sequences from 30 different participants without any constraints. The collected dataset is made available for the research community to contribute to online handwriting classification, i.e., text, mathematical expressions, graphs/drawings.
4. Distinct features are vital in the performance of Machine Learning (**ML**) classifiers; more refined features result in better classifier performance. The research community has been adopting handwriting recognition feature sets for handwriting classification problems. This thesis presents a new feature set for online handwriting classification for the very first time to the best

Quality datasets are key to quality methods

onTabWriter dataset helps in online handwriting classification

Classification results rely on the quality of features

¹ Ipad Pro

² Apple Pencil

of the author's knowledge. Furthermore, this thesis covers a deeper analysis to signify the relevance of every individual feature for online handwriting classification problems.

A comprehensive ablation study establishes the efficacy of proposed feature set

5. To validate the effectiveness of the proposed feature set, this thesis presents an extensive ablation study and comparative analysis with existing feature sets. We also investigate the significance of context for the problem at hand. Ablation study and comprehensive comparative analysis on **onTabWriter** reveals that the new proposed feature set can capture a rich representation of handwritten sequences, which in turn results in superior performance on the task of online handwriting classification.

On-body sensor set-up helps in monitoring cognitive activities in classroom

6. This thesis presents proof of work to correlate the cognitive activities in the learning process to provide insights about performance and feedback estimation. The presented methods also help in evaluating the expertise of the learner for a given task by assessing the reading behaviour, writing behaviour, and their correlations. The derived results from the evaluation will help to identify the weaknesses and strengths in a particular area, resulting in improvement of the overall learning process.

Smart glasses improve engagement in interactive learning activities

7. This thesis uses an Mixed Reality (MR) application "gPhysics" to explore the potential of smart glasses as an experimental tool in Physics education. gPhysics application helps the learners perform experiments by providing visual assistance and feedback on HMD. It is observed that using smart glasses as an experimental tool increases curiosity and results in better engagement for the whole experimentation time.

Fingers are ideal replacement of pens to write in AR and VR

8. This thesis presents a real-time application of air-writing in augmented/virtual reality using a single low-cost Inertial Measurement Unit (IMU) as the Finger Air Writing System (FAirWrite) system. The proposed system intuitively captures the finger motions in the air and reconstructs its trajectory as air-writing in real-time. Despite system development, systematic evaluation of the proposed system for reconstructing such noisy trajectories and classifying them into digits and characters exploiting the potential of DNNs lays another major contribution of this research work.

1.4 OVERVIEW

This thesis is structured as follows:

Starting with the motivation [Section 1.1](#), followed by research scope in [Section 1.2](#) and contributions in [Section 1.3](#), and structure of this thesis are defined in [Chapter 1](#). [Chapter 2](#) is divided into three major parts. In [Section 2.1](#), an overview of the recent technological developments and *SotA* methods related to Page Object Detection (*POD*) are covered. A literature overview related to online handwriting classification is covered in [Section 2.2](#). In [Section 2.3](#), a detailed analysis of wearable applications with regard to *MR* in formal education is covered. The rest of this thesis is split into two major parts, one for each of the main goals elaborated in [Section 1.2](#): Cognitive Ability Classification ([Part II](#)) and Applications of wearable sensors in formal education ([Part III](#)).

Introductory part of the dissertation

In [Part II](#), [Chapter 3](#) presents methods for content classification using *DL* in combination with traditional *CV* techniques, comprehensive evaluation of presented methods, and *POD* datasets. [Chapter 4](#) presents a comprehensive analysis of online handwriting classification, which includes data collection, data processing, feature extraction and a new feature set for online handwriting classification, evaluation of features and dataset using Machine Learning (*ML*) and Deep Learning (*DL*) classifiers, and importance of context in online handwriting classification. We also cover the importance and relevance of context for online handwriting classification. *POD* and handwriting classification can help to analyse the reading and writing activity individually, and it is very important to analyse and evaluate both activities in correlation to dive into performance evaluation and feedback estimation based on the expertise of the learners, which is covered in [Chapter 5](#).

AI-based methods & their applications for cognitive ability classification

[Part III](#) of this dissertation emphasizes on applications of wearable sensors and gadgets as an interactive tool to perform learning activities. [Chapter 6](#) presents an air-writing tool *Finger Air Writing System (FAirWrite)* presented with multi-dimensional applications. *FAirWrite* system captures hand movement gestures as air-writing snippets in *AR VR* using single *IMU* without requiring any reference surface. *FAirWrite* system records the air-writing gestures and reconstructs the trajectory in real time for visual feedback on writing progress. This thesis also presents a large collection of data from independent users to assist the research community in air-writing classification. The evaluation results of *FAirWrite* system, including quantitative results and a qualitative error analysis using *DL* methods, are covered.

Applications of Wearable sensors in formal education

Associated Research is made part of this thesis in [Part IV](#) *Deep Stamp Recognition (dStar)* system is presented in [Chapter 7](#), a novel and generic approach to detect the stamps from administrative documents using *FCN*. [Chapter 8](#), is divided into two parts, where research outcomes are a part of collaborative work. First part [Section 8.1](#)

Associated research

presents a study on the influence of smart glasses on learning outcomes while performing experimentation in a classroom environment. In the later part [Section 8.2](#), an innovative approach is presented to sense the potential variations caused by disturbance or movements using near-field electric field information.

The thesis is concluded in [Part v](#) with a summary, limitations of proposed methods, and hints toward future work.

STATE OF THE ART

This chapter is dedicated to the foundations this work is built upon. As this dissertation focuses on methods and applications to bridge the gaps between formal education and Artificial Intelligence (AI), this chapter is split into three major sections: [Section 2.1](#) covers the details of related work and state-of-the-art (SotA) methods for content analysis, i.e., detection of key objects from document page images and classification of online handwriting in [Section 2.2](#), which serve as basic components for bridging. [Section 2.3](#) is further dedicated for applications of wearables related to formal education keeping in view the perspective of Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR).

2.1 PAGE OBJECT DETECTION

Page objects on document images have very low inter-class variance and can easily be confused with each other. Hence, [POD](#) is a vital step to define the structure of documents, which is inherent to structured reading order, whether handwritten or printed. Once document structure is defined, it enables the eye-tracking research community to relate the relevancy and importance of reading text with [PoIs](#) using the gaze information for deeper analysis to evaluate reading behaviour, an essential activity in formal education. Different page objects are possibly present on a document image, i.e., text, mathematical expressions (formulas), figures including plots and graphs, lists, tables, etc. This work focuses on detecting figures and formulas from document images, and the rest of the page objects are considered text. [POD](#) from document images is a well-recognised problem and has received noticeable attention of the document analysis community with the surge of [DL](#)-based object detection methods. This section summarizes relevant and [SotA](#) work for [POD](#).

Significance of [POD](#) for content analysis

Page Objects like figures, formulas, and tables are an integral part of the document, as they preserve significant information in a confined space. [POD](#) related work is mainly divided into two types in this section depending on methodologies adapted to process the document images. (i) Traditional approaches rely on commonly used [CV](#) methods and a hand-crafted set of features to highlight the page objects on a document image. (ii). [DL](#) based methods which utilize the potential of [DNNs](#) to extract the page objects from document images relying on feature extraction using Convolutional Neural Networks ([CNNs](#)).

[POD](#) methodologies are broadly categorised as traditional & [DNNs](#) approaches

2.1.1 Traditional Approaches

CV techniques

This section provides a history and brief overview of the techniques based on traditional **CV** methods such as binarization, Connected Component Analysis (**CCA**), distance transform, localization, and keyword extraction methods to highlight the page objects present on document images. Several different methods have been presented to perform formula detection [50, 115, 186], figure detection [41, 92, 100], and table detection [76, 123, 215].

Table detection using traditional methods

Several different methods have been presented in the past to locate tables from document images by employing traditional techniques. Some prior methods [67, 76, 250] based on line definition and white spaces were proposed to define the structure of tables in document images. Some other techniques rely on open source Optical Character Recognition (**OCR**) based methods to detect tables from heterogeneous documents [215, 239]. Another method to detect tables from non-raster Portable Document Format (**PDF**) files by applying *pdftotext* Linux utility to extract feature vectors followed by Hidden Markov Model (**HMM**) is presented by Silva et.al. [222]. The performance of all these methods depends either on the structure of a document, the location of tables on a document, or the format of the tables themselves, a major limitation of traditional methods.

Figure detection using traditional approaches

Figures are a significant part of document images. A famous Chinese quote; "one picture is worth ten thousand words", highlights the importance and information preserved in figures. Figures present in document images are very diverse; hence a challenging task to detect them using traditional methods. There are a few methods to detect them from document images. Hirayama et al. [100] presented a block segmentation method to detect the figure areas from document images. The proposed method segments the whole page into blocks based on border lines, and then blocks are unified using column information, followed by text and figure area detection using projection profile methods. Ha et al. [92] presented a similar idea which also uses projection profiles and **CCA** to segment images from document page images using the XY-cut algorithm. These methods might look very simple today, but they opened up a new research paradigm for the document image analysis community to detect figures from document images.

formula detection using CV techniques

Formulas are a precise way to express relations as mathematical expressions, and detecting them is a complex task. Formula detection is a key vehicle in transcribing document images into electronic form. Kacem et al. [118] present a **CCA** and bounding box based method to segment the formulas from document images. The formula segmentation process is carried out in two separate steps based on global

and local segmentation. Local segmentation processes the formulas from plain text and global segmentation segments standalone formulas from text based on heuristics-defined primary labeling. Fateman et al. [68] presented a parsing-based technique to extract the mathematical content from document images. The parsed information is stored as latex expression for further use and processing. Iwattsuki et al. [115] present a method to identify math zones by applying Conditional Random Fields (CRFs) using a manually annotated corpus from PDF documents. The proposed method uses layout (font types) and linguistic features (context n-grams).

A major drawback of custom feature engineering and heuristics-based approaches is that their outcome depends heavily on the quality of features, which also rely on domain knowledge. Dependence on custom features also restricts the performance of traditional approaches to single-class object detection. Diversification in a single class may also result in degraded performance of traditional approaches, like in the case of figures which may include images, plots, graphs, drawings, etc. Therefore, custom feature-based methods are not suitable for multi-object classification and with diverse datasets.

Limitations of CV techniques

2.1.2 Deep Learning Approaches

As traditional approaches learn from the preprocessed features as input to the system based on defined heuristics, limiting their performance to a specific scope. On the other hand, DNNs learn the patterns and correlations of the representations from raw input data to extract the features. These representation-based features broaden the canvas of DNNs to various advanced problems to efficiently deliver accurate and reliable results without the intervention of human expertise. With the popularity of CNNs to classify multi-class objects with SotA results, POD problem from document images also got the attention. This section provides an overview of SotA techniques for tables, figures, formulas, and multi-page object detection from document images.

DL for POD

Schreiber et al.[211] present a DL based approach utilizing Faster-Region Based Convolutional Neural Networks (RCNN) method [197] to detect tables from document images. They also employ the transfer learning method to apply a pre-trained model from natural scene images to document images domain. The proposed method detects tables and also attempts to recognise the table structure. Gilani et al. [79] also present a similar approach to detect tables from document images by adding a preprocessing step to the input image. They transform Red, Green, Blue (RGB) input image using different distance transform methods, i.e., Euclidean distance transform [31], linear distance transform [66], and max distance transform [194], to each input image channel. A major shortcoming of the proposed work

Table detection using CNNs

is that it does not provide any information on the advantages/disadvantages of using the image transformation process. A recent [SotA](#) end-to-end method using CDeC-Net is proposed by Agarwal et al. [1] detect tables from document images.

*Formula detection
using DNNs*

There are different ways to process formulas/mathematical expressions, i.e., formula detection from document images, formula recognition from offline handwritten expressions, and online formula detection. Here, we emphasize only on [SotA](#) methods for formula detection from document images. Ohyama et al. [173] presented a [CNN](#)-based image conversion technique for extracting mathematical expressions from page images. The proposed approach deploys a similar architecture to U-Net [198] to convert the input page image to a mathematical expressions only image as output. A Single Shot Detector ([SSD](#))-based [145] method to locate formulas from page images using a sliding window method is presented as ScanSSD by Mali et al. [151]. Phong et al. [187] applied You only Look Once ([YOLO](#)) network to detect formulas from page images followed by an end-to-end method for recognition of detected formulas using Watch, Attend, and Parse ([WAP](#)) network.

*Graphical object
detection using
CNNs*

Saha et al.[204] present an end-to-end trainable neural network to detect graphical objects from document images. They employ two [SotA](#) object detection networks: Faster-[RCNN](#) [197] and Mask-[RCNN](#) [97] into page images domain. The evaluation process is carried out to detect tables, figures, and formulas on [ICDAR2017-POD](#) [72], [ICDAR-2013](#) [81], and [UNLV](#) [216] datasets. The proposed method did not include the current and even recent [SotA](#) in the evaluation process.

*POD using
Detectron*

Xu et al. [263] utilize the Mask-[RCNN](#) architecture for [POD](#) from document page images. The proposed methods use Feature Pyramid Network ([FPN](#)) [143] with Residual Network ([ResNet](#))₁₀₁ [98] as backbone trained for document images. The proposed method is exhaustively evaluated on six different datasets, including [ICDAR2017-POD](#) [72], along with a synthetically generated dataset to detect tables, figures, and formulas from document page images. The proposed method also skips the recent [SotA](#) methods for [POD](#) in the evaluation process.

*POD using
spatial-related
relation & vision*

A very recent work by Bi et al. [20] presents a novel method, namely spatial-related relation and vision for [POD](#). The proposed network is a combination of three sub-networks: vision feature extraction network, relation feature aggregation network, and result in refinement network. These sub-networks enable the proposed methods to use a combination of vision and spatial-relation features (context information). The proposed method covers a comprehensive ablation study to establish the utility and effectiveness of each sub-network

and extensive evaluation on three different datasets; ICDAR2017-POD, Article regions [227], and PubLayNet [284].

The recent *SotA* method to detect page objects from document images uses a combination of traditional *CV* methods, *CNN*, and clustering techniques presented by [140]. The proposed method detects the tables, figures, and formulas from document images with *SotA* results on ICDAR2017-POD dataset. The proposed method is a combination of multiple sub-networks and relies on heuristics-driven processing heads, which can limit its performance to generic scenarios.

Recent SotA method for POD

In our work, we focus on the methods that create a balance between commercially existing methods by addressing their limitation to improve their scope of applications in real-life scenarios. Such as, commercially available *OCRs* can detect text and tables from document images, but methods to detect remaining page objects, i.e., figures and formulas, are still missing. So this work presents the methods to detect and segment the figures and formulas from document images to improve the performance and reliability of existing systems.

2.2 ON-LINE HANDWRITING CLASSIFICATION

Handwriting analysis is a vast field of research that includes the applications areas in forensic sciences, postal services, education, the banking sector, and e-commerce. There are two methods for handwriting data acquisition, i.e., offline and online. In the offline data acquisition process, only spatial information in the form of an image are conserved. In online data acquisition, temporal information with spatial information are also considered, and data is stored and available for processing as time-series sequences. There are several ways to process the handwriting data, such as handwriting recognition, a key step to transform document images into digital documents, signature verification to verify the authenticity of the documents, forgery detection, writer identification, writing mode detection, and handwriting classification.

Applications of handwriting classification

Cognitive ability classification in classrooms is one of the key contributions of this thesis, and handwriting is a fundamental activity in the learning process. This section familiarizes the readers with the background and *SotA* methods for online handwriting classification. This section is further divided into two subsections. Section 2.2.1 provides details about the background knowledge and conventional methods for handwriting classification. Section 2.2.2 covers the details of recent *SotA* methods and approaches applying *DNNs* for online handwriting classification problem.

Handwriting classification for cognitive ability classification

2.2.1 Traditional Methods

Handwriting analysis has been successfully used as a biometric tool

The known history of handwriting analysis tracks back to the second quarter of 20th century before the start of the second world war when handwriting samples were used as a forensic tool for person identification. Nottingham [167] and Milwaukee [12] police adopted the Lee and Abbey system of handwriting classification by collecting a database of handwriting samples to keep the record of photographed prisoners in order to reform the criminal investigation process. This opened up new venues for the research community to systematically investigate handwriting behavior by looking into the features and characteristics of every individual's handwriting. Smith et al. [226] presented the first known feature set for handwriting classification in 1954. The proposed feature set considered six factors as features of an individual's handwriting. The proposed feature set enabled the person classification based on its handwriting samples by examining the speed, size, slant, spacing, pressure, and form of the writing. The first four features were categorised as developed characteristics and the last two as unconscious behaviour that may vary for individuals.

Handwriting classification system to identify the law violators

In 1959, Livingston et al. [12] presented a new handwriting and pen-printing classification system for criminal investigation to identify law violators. The proposed system investigates specific letter factors writing style of $\{e, r, \dots, k, S\}$, and general writing factors $\{\text{slant, capital connections, the}$. The proposed system was successfully put into search and identification experience and approved by the Milwaukee police department into operation.

Digital systems and online handwriting classification

With the advent of computer-based pens and handwriting systems in the last decade of 20th century, online handwriting processing got a boost. Its analysis broadened to person identification, writing style classification, and handwriting classification. Schomaker et al. [208] presented a method to classify the writers and writing styles for online handwriting recognition. They used Kohonen neural network method for the segmentation of words to obtain a stroke alphabet followed by the process of writer classification using a probabilistic stroke transition network. They also present a comprehensive database for online data exchange and recognizer benchmark as UNIPEN dataset [91], a vastly used dataset for online handwriting recognition problems.

Synthetic parameters for handwriting classification

A new set of synthetic parameters based on the fractal behaviour of handwriting is presented by Bouletreau [26]. They present a set of four parameters: the fractal dimension of writing computed from the slope of the central zone, the secondary dimension based on the last zone of writing, the legibility rating of the image based on an iterative

division of the evolution graph, and the implication index computed as a delta of first two features. Their synthetic parameters result in improved performance in classifying handwriting into families than conventional ones.

Willems et al. [254] presented a feature set for mode detection in natural online pen inputs. The presented feature set consists of six global features length, area, compactness, eccentricity, circular variance, and closure, and two structural features, curvature and perpendicularity, to classify pen trajectories into handwriting, lines, arrows, and geometric shapes. The proposed feature set is evaluated using K Nearest Neighbour (KNN) by achieving classification accuracy of 98.7 on unseen test data.

Mode detection for online pen inputs

Jain et al. [116] present a hierarchical approach to extract homogeneous regions from online handwritten documents. Firstly, documents are segmented into regions of text and non-text strokes, followed by text region classification into plain text and unruled tables, and non-text regions classified as drawings and ruled tables. Rossignol et al. [200] present a preliminary system for distinct online handwriting into text and different drawing classes. Bishop et al. [24] present a system to separate the text from graphics from online handwritten strokes using stroke information, gaps between two strokes, and temporal characteristics of stroke to train a probabilistic classifier.

Homogeneous regions classification from handwritten documents

Willema et al. [253] presented a mode detection system for online pen input in the crisis management domain. The presented method employs a Bayesian belief network to combine the classification results with context information to improve the overall performance of the system. The presented system distinguishes the different pen inputs as deictic gestures, handwritten text, and iconic objects with an error rate of 4.0%.

Handwriting classification in crisis management

Kumara et al. [135] proposed a novel system for the classification of writer-independent online handwriting classification based on a large margin approach. The proposed method starts with describing a scheme for interpolation of time-series by the sum of polynomials using Reproducing Kernel Hilbert Space. The derived interpolations are processed by large margin formulation to achieve the [SoTA](#) results on multiple datasets.

Using large margins for handwriting classification

Garlapati et al. [75] presented a classification system for offline handwritten and printed text to improve the performance of Optical Character Recognition (OCR) System. The proposed system mainly consists of three sub-processes: text localization (binarization and morphological operations), feature extraction (structural and visual intensity features), and classification (State Vector Machine (SVM) clas-

Handwriting classification to assist OCRs

sifier). The proposed system achieves the [SotA](#) results for printed and handwritten text classification on IAM dataset [157].

2.2.2 Deep Learning Methods

*DL for on-line
handwriting
classification*

Deep Learning (DL) methods paved the new way to process handwriting analysis in multiple dimensions, from writing hand detection, handwriting recognition, signature verification, writing mode detection, [OCRs](#) to handwriting classification. This section will cover details of [DNNs](#)-based networks to provide an overview of the progress and recent advancements in the online handwriting analysis domain.

*n-grams & HMMs
for handwriting
classification*

Toselli et al. [235] presented an automatic handwriting recognition and classification method based on [HMMs](#) and n-grams. The proposed system is based on two-stage finite-state models. The first phase [HMMs](#) and n-gram language models are employed to recognise the handwritten sequences, followed by n-gram text classification models to classify the recognised text in the second phase.

*IAMonDo dataset
for online
handwriting
classification*

Indermuhle et al. [108] presented the IAMonDo database online handwritten documents for mode detection problem. The IAMonDO dataset opened up a new research dimension in online handwriting processing. They also present a Bidirectional Long-Short Term Memory ([BLSTM](#)) neural network-based approach to detect the writing mode in online handwritten documents [107]. The proposed approach is one of the earliest methods to utilise neural networks for mode detection problem in online handwritten documents.

*Using CRFs for
text/non-text
classification*

Delaye et al. [54] present a [CRFs](#) based method to classify text/non-text classification from online handwritten documents. Using [CRF](#) enables the authors to incorporate contextual information, i.e., spatial and temporal relationships between the strokes. They also present an improved version of their system in another work [55]. The proposed method is evaluated on the IAMonDo dataset to classify text and non-text strokes with state-of-the-art ([SotA](#)) results.

*Graph Attention
Networks for
contextual stroke
classification*

Ye et al. [267] present a graph attention networks base method to classify the contextual strokes in online handwritten documents. In graph networks, strokes are treated as nodes, and temporal and spatial interactions between strokes represent edges and the whole document as a graph. Graph convolutions are combined with attention mechanisms to dynamically aggregate neighborhood features to enable the whole network to learn context-aware features. They evaluate the proposed method on the IAMonDo dataset to report the superior efficacy of their methods.

Polotskyi et al. [190] propose a neural network-based method to classify online handwritten strokes as text and non-text without using context information. The proposed system adopts the features from the online handwriting recognition domain to the online handwriting classification domain. The proposed approach is tested on a publicly available IAMonDO dataset and a Samsung Mobile Handwriting Document (MHWD, MHWD_M) dataset with impressive results. The authors further claim that online handwriting classification results are better when no context information is used in comparison to when using context information.

DNNs for on-line handwritten stroke classification

Grygoriev et al. [88] present a hierarchical approach using 1D Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) for online handwritten stroke classification. The proposed architecture used 1D CNN on a lower hierarchical level and RNN on an upper hierarchical level. The authors claim the SotA performance of the proposed model not only in terms of accuracy but also for computation costs and memory consumption.

Hierarchical approach for online stroke classification

Digitisation is a key factor in storing information in modern-day life. Digitising the handwriting activity in the classroom is a nearly impossible task with existing systems because of diverse writing activities in the classrooms, i.e., writing text, maths, drawing figures, and plotting graphs with intra-personal variations. In our work, we focus on addressing the problem of online handwriting classification to address the shortcomings of existing systems. We proposed SotA methods to classify the online handwriting sequences into text, formulas, and plot/graphs. Classification of handwritten sequences results in defining the structure of handwritten documents, leading to a systematic evaluation of handwriting activity.

2.3 APPLICATIONS OF WEARABLE'S IN FORMAL EDUCATION

Technological advancements and the proliferation of affordable wearable devices such as bracelets, rings, glasses, watches, and embedded in clothing are introducing humankind to new immersive and imaginary worlds. These advancements have made Augmented Reality (AR) and Virtual Reality (VR) more viable and desirable in many domains to perform routine activities as wearable devices become ever more integrated into everyday life. This urges the need to take advantage of advancements to educate the children and prepare the teachers to capitalise on these opportunities. This section covers the background knowledge and recent applications of Augmented Reality (AR) and Virtual Reality (VR) used in classrooms, focusing on classroom activities in formal education. In the first part, we focus on background knowledge and commonly used devices for AR and VR

Wearable's have huge potential in formal education

scenarios to familiarise the reader with the topic. In the later part, an overview of the applications is presented to make use of [AR](#) and [VR](#) in formal education.

Outlook

The term Virtual Reality ([VR](#)) and Augmented Reality ([AR](#)) got the technological craze in the 1990s [63] but got the attention and adopted by the masses in 2010s in multiple domains with the introduction of Head Mounted Displays ([HMDs](#)) and wearable gadgets. These applications are in vast domains, including gaming, healthcare, real estate, marketing and advertisement, fitness and training, manufacturing, education, etc. It is essential to adapt [AR](#) and [VR](#) applications to deploy them in educational programs so they best meet the requirements of learners and the scope of educational needs. Adopting these technologies in education facilitates learning, knowledge acquisition, lower cognitive load, and increased attention. Moreover, implementing these technologies in classrooms results in improved engagement compared to traditional learning methods. It enables the learners to interact in an immersive world to understand abstract information and complex representation learning.

AR & VR in education

[AR](#) and [VR](#) technologies enable the seamless connection of the digital and physical domains that combine real and virtual information to convey abstract information using unique visual and interactive experiences. Recent studies and surveys [23, 43, 171] showed the encouraging trend to deploy these technologies in classrooms and their acceptability among the learners. Combining [AR](#) scenes with traditional learning activities enhances the learner's problem-solving skills, motivation, involvement, and engagement [172]. These technologies also empower the instructors to adapt their instructions keeping in view of the individual learners' needs by analysing their preferences to help them improve their performance [171]. A teacher observed, "[AR](#) textbook provides a multi-sensory approach to learning that links text, image, sound, and movement and is a highly motivational communication format", which further adds, "no question that [AR](#) will prove to be a highly effective medium both for entertainment and education", after an [AR](#)-based storytelling workshop conducted for school students [161].

Studying Kirchhoff's law in AR

Kapp et al. [119] present an [AR](#) experiment study for high school students to understand Kirchhoff's law in electric circuits. Students first built the circuits and then performed real-time measurements using smart glasses. The presented system enables students to understand the conceptual evaluation of the current and voltage relationship without requiring them to perform repetitive measurements while working with real setup and data.

Nguyen [171] presents a study to investigate the perception of AR and VR applications from a student perspective, an essential question considered by instructors and course designers. They designed 16 weeks of AR/VR technologies course to examine the learning behaviour of students based on five activities: learning the basics, self-learning, working with projects, scaffolding to support students, and students' evaluation. Study shows an encouraging trend among students to adopt the necessary tools and create applications with various topics of their interest.

AR & VR applications improve creativity among students

Plunkett et al. [189] present a method to incorporate AR in the laboratory to perform chemistry experiments using a simple smartphone. AR enables the projection of virtual information onto a real-world scenario. The use AR notecards to demonstrate Organic Chemistry reactions and mechanisms using the HP Reveal application. The physical AR notecards contain a Quick Response (QR) code, reactants, chemical substrate, and reaction direction pointing to an unrevealed product. Scanning the AR notecard using HP Reveal shows a simulation displaying the product's chemical structure along with electron movements. This application enables chemistry students to understand the chemical reaction process in classrooms and assists them in performing experiments in the laboratory.

Understanding chemical reactions using AR

Yu et al. [279] present a comprehensive study on evaluating the impact of AR in different experimental conditions. They explore the use of AR learning tool to facilitate the students to understand magnetic field concepts. The study also assesses the anxiety level of students, learning motivation, and learning performance to investigate the impact of AR while performing the learning activity. Results show that the learners perform better in AR experiment setup compared to traditional material, understanding abstract and complicated concepts, improved learning gains, and with higher concentration. Students in AR setup also show higher motivation, positive attitude, and lower cognitive load.

Students show higher engagement while learning through AR

Our work on applications of wearable devices in formal education mainly focuses on two main directions. Firstly, we present a study on the use of HMDs to assist students while demonstrating practical skills during science experiments, to help them perform the activity more efficiently with a more profound and better understanding of the topic. In another application, we present a wearable system to interact within the classrooms to enable them to perform the writing activity in the air without needing a reference surface of visual feedback. These applications encourage the research community for deeper analysis and investigation to introduce the applications and systems to enable the learners to interact with their learning environments in MR.

Part II

LEARNING ACTIVITY CLASSIFICATION &
PERFORMANCE EVALUATION

Content classification is a fundamental step in document image processing for downstream tasks such as intelligent document structure definition, document editing, and content understanding. Documents are classified into two types on an abstract level: handwritten and printed documents, and there are different mediums for document creation and acquisition, i.e., handwritten, text files, web documents, Portable Document Format (PDF), images, etc. To cater to all the diverse methods to create and save documents using a single system requires converting them to document images. Specialized algorithms are needed to process and interpret information present in document images. This chapter of the thesis focuses on methods to enable the POD from document images to automate the process of document segmentation and information extraction, an essential and vital step in content analysis. There are several other advantages of document structure definition besides content analysis, such as digitization, document editing, data accessibility, and information retrieval. One of the significant applications of content classification methods in formal education is to evolve reading behaviour analysis techniques with deeper insights. Content classification enables the existing gaze-tracking methods to correlate the gaze information with the reading text and their relevancy for further evaluation. Major contributions of this chapter are highlighted as follows:

Content classification is fundamental in document structure understanding

- Hybrid approaches that fuse traditional Computer Vision (CV) methods with Deep Learning (DL) for refined representation learning to detect heterogeneous objects in document images, figures and formulas in particular.
- Detecting the problems in the ground truth of ICDAR2017 Page Object Detection (ICDAR2017-POD) [73] dataset and refinement of said dataset to eliminate disproportions and confusions.
- Curation of a new Figure and Formula Detection (FFD) dataset for POD from document images to benchmark the proposed methods and to assist the development of generic systems in POD domain.
- Ablation study of the proposed method on a large publicly available dataset ICDAR2017-POD to justify the efficacy of the proposed approach.

*Structure of the
chapter*

The Rest of the chapter is structured as follows. [Section 3.1](#) states the problem statement, challenges, and motivation for [POD](#) problem from document images. [Section 3.2](#) summarizes the recent developments and state-of-the-art ([SotA](#)) systems in [POD](#) domain. [Section 3.3](#) covers the details of systems developed during this research work: [Fi-fo](#) Detector and [FFD](#), along with a detailed analysis of the methodology to detect figures and formulas from document images. [Section 3.4](#) presents an overview of the datasets and evaluation protocols followed in furnishing the performance of the proposed methods. [Section 3.5](#) covers the details of the obtained results, along with a comprehensive discussion to feature the highlights and weaknesses of the proposed methodologies.

The author of this thesis has published the content, figures, and tables included in this chapter in the following publications. The author of this dissertation has written all the text taken from the mentioned publications and the text in this chapter itself. The publication list included in this chapter refers as follows:

- Younas J. et al. (2020), Fi-fo Detector: Figure and Formula Detection using Deformable Networks. In: Applied Sciences 10.18 (2020)[276]
- Younas J. et al. (2019), FFD: Figure and Formula Detection from Document Images. In: 2019 Digital Image Computing: Techniques and Applications (DICTA), (2019)[275]

3.1 MOTIVATION

Why information extraction is needed from documents

Digitization of document images is a growing need for commercial and non-commercial entities, i.e., banks, industries, educational institutes, libraries, etc. Aside from record-keeping, it significantly improves the availability of data just at a click and/or a tap from anywhere in the world, at any time. These digitized documents can be processed in an automated fashion, given that the information contained in those documents can be extracted reliably. Reliable extraction of information from documents has been a major focus of the document analysis community for decades [3, 124, 219].

Significance of figures and formulas in documents

Figures are an integral part of a range of different types of documents as they portray the maximum amount of information in the least amount of space/time. On the other hand, formulas are the best way to express these relations symbolically, leveraging the power of mathematics at its core. Detection of formulas and figures from document images is a challenging task as document images are composed of multi-level information. The information encapsulated in a document includes title, author details, corresponding text, figures, formulas, and many other related objects.

Challenges

Figure detection from document images is a challenging and crucial task. Figure detection is a prefatory step in document image processing systems, enabling these systems to discriminate between textual and non-textual regions present in a document. Figures that are usually present in document images include layout design, block diagrams, natural images, and plots/graphs. Decorative graphics, i.e., long lines and "rules," are not considered figures in this work. Similarly, formulas are presented as a 2-dimensional arrangement, with distinctive structural features compared to the plain text, which is 1-dimensional. They can portray complex inter-relationship between different entities in a concise form. The advantages of the ability to recognize formulas are twofold: (i) Formula detection eases the dissemination and retrieval of mathematical knowledge from document images and (ii) enhances the performance of text recognition systems like Optical Character Recognition (OCR) as the conventional text processing pipeline should not be executed on those regions producing counter-productive transcriptions [32, 225]. Figure and formula detection from document images is a challenging task as figures and formulas are usually spread widely across the document images at varying locations. Likewise, figure and formula appearance rely massively on the document format, style layout, orientation, aspect ratio, and other factors. Therefore, it is not easy to detect figures and formulas directly from document images, which could be a potential reason existing commercial and open-source tools lack support for this functionality. Moreover, as table detection is already available in

commercial OCRs [225], e.g., Tesseract, Abby, therefore table detection is not considered for evaluation in this work.

Significant efforts have been made in the past to segment out the different page-objects in document images. Most of the early approaches heavily relied on heuristics-which are task-specific- and thus fail to generalize to novel scenarios [50, 92]. Deep Learning (DL) based models have been leveraged for this segmentation in the recent past [79, 140, 211, 268]. All of these methods involve a significant amount of pre or post-processing based on hand-designed heuristics. A recent attempt has been made by Siddiqui et al. [219] to incorporate deformable CNNs for the analysis of document images. However, the potential of DL methods in combination with traditional computer vision approaches hasn't been well explored in this context.

Limitations of existing POD approaches

This chapter of the thesis presents generic, data-driven, and end-to-end methods for the detection of figures and formulas in document images. The proposed methods leverage the potential of a combination of CV techniques to further boost the capabilities of the DNNs. We particularly leverage a novel combination of traditional approaches, which includes inverse distance transform, Connected Component Analysis (CCA), and the gray-scale version of the raw input image to strengthen the capabilities of the deep models further, as color features are not particularly useful in telling these page-objects apart. The transformed image representations are termed as Fi-fo image representations. The first method, Figure and Formula Detection (FFD), employs faster-Region Based Convolutional Neural Networks (RCNN) [197] and mask-RCNN [97] as deep models in our approach to detect figures and formulas from document images. In the second method, Figure and Formula (Fi-fo) Detector successfully uses the Feature Pyramid Network (FPN) with deformable convolutions to identify the figures and formulas occurring at different scales, orientations, and aspect ratios. The proposed methods are evaluated on the publicly available ICDAR2017-POD competition dataset and a newly curated dataset FFD.

Highlights

3.2 RELATED WORK

Document image processing is an interesting topic among the Computer Vision (CV) research community. Significant progress has been made in this domain, which includes heuristic-based, Convolutional Neural Network (CNN) based, statistics-based-such as CRFs & Graph trees, and/or combination of these methods [50, 92, 115, 121, 186]. Heuristics include color-based features, shape-based features, geometric features, and key point descriptors. Deep Learning (DL) based approaches use CNNs [121], Region Proposal Networks (RPN) [79], and deformable CNN [219]. Tasks performed on document images

There are different ways for POD from document images

include (but are not limited to) textual and non-textual region discrimination, graphics, and page object detection, which includes text, formulas, and figures.

*X-Y cut algorithm
for image
segmentation*

Ha et al. [92] presented a fundamental method for image segmentation based on the recursive X-Y cut algorithm. Their method used a projection profile based on the spatial configuration of Connected Component Analysis (CCA), which extracts columns from document page images. It might appear a simple method today, but it has opened up a new research direction in page object segmentation and detection decades back.

*Using OCR for
image segmentation*

Chiu et al. [41] presented an OCR based picture detection method from document images. OCR is applied to detect text regions, followed by segmentation methods to mask them out, and finally, non-textual regions are clustered together. Segmentation is further improved using caption information in post-processing. This approach depends not only on the performance of the OCR but also on the subsequent steps, which are to be executed precisely to achieve better results.

*Saliency-based CNN
for table and chart
detection*

Kavasidis et al. [121] presented a saliency-based CNN for table and chart detection from digitized images. They applied saliency detection on input images to preserve contextual information. FCN is used as a base detector followed by fully-connected CRFs for localizing tables and charts. They evaluated the presented method on the extended version of the ICDAR-2013 dataset. This approach is not only multi-step but an extended version of the used dataset is not publicly available to draw comparisons.

*OCR in combination
with CNNs for
document image
segmentation*

Yi et al. [268] presented a page object detection method using region proposal CNNs, followed by a custom algorithm to refine proposed regions along with a CNN classifier for object category classification. It first pre-processes the input image by applying a component-based region proposal algorithm customized for document images, which extracts the rough region proposals at the initial stage and prunes them later. The refined region proposals are fed to the CNN model for classification. The results of CNN models are finally post-processed by a dynamic algorithm to optimize the detected region proposals. They evaluated their system on a private dataset and considered four-page objects for classification, i.e., text lines, figures, formulas, and tables.

*CRFs-based method
for POD from PDF
documents*

Iwatsuki et al. [115] presented a CRF based method to extract formulas and mathematical zones from PDF documents. Their method uses layout features like font, style, and linguistic features such as n-gram context to build their CRF model. Phong et al. [186] developed a new method for detecting mathematical expressions. They used OCR to analyse layout, text lines, and expressions. Features are extracted

from expressions using the Fast Fourier Transform (FFT) and Mean Square Error (MSE). State Vector Machine (SVM) classifiers were applied on extracted features to classify mathematical expressions and formulas.

Gao et al. [73] presented a combination of a Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) models to detect formulas from PDF documents. A combination of CNN and RNN models enables this method to preserve both character and visual features for formula detection. They applied bottom-up and top-down strategies to generate formula region candidates, followed by feature extraction networks (CNNs & RNNs) and post-processing for refined formula region. However, this work can only be applied to PDF documents, which is a considerable limitation when discussing document images.

Using a combination of CNN & RNN for formula detection

Deep Learning (DL), in the recent past, has become the center of attention for research in the document analysis community. Gilani et al. [79] and Schreiber et al. [211] leveraged Faster-RCNN for table detection. Siddiqui et al. [219] additionally equipped the Faster-RCNN model with deformable property to gain significant improvements over prior state-of-the-art (SotA). National Laboratory of Pattern Recognition (NLPR)-Pattern Analysis and Learning (PAL) [72] are the winner of the ICDAR2017 Page Object Detection (ICDAR2017-POD) competition. They presented a multi-stage approach for the classification of figures, formulas, and tables from document images using the connected components of the input image, SVM classifiers, CNN based CRFs, Faster-RCNN, and normal CNNs. Finally, final results are achieved by integrating the intermediate results of these stages.

Using DNNs for document image segmentation

A recent SotA POD system is proposed by Li et al. [140]. They proposed a hybrid model, which is a combination of deep structured prediction and supervised clustering for page object detection. First, they extract columns and then line regions of document images. They used Conditional Random Field (CRF) formulated Convolutional Neural Network (CNN) with unary and pairwise potentials to classify and cluster primitive region proposals from line regions. After classification, the same class clusters are merged to get page objects. Their presented approach comprises partially trainable networks with heuristics-driven pre-processing and post-processing heads, which might limit its application in a generic scenario.

Recent SotA system for POD

Saha et al. [204] presented the most recent method for Page Object Detection (POD) in document images. Their approach is based on mask-RCNN for figure, formula, and table detection in document images. Although the presented approach has no strings (pre and/or post-processing) attached to it, the authors did not compare their ap-

Graphical object detection from document images

proach against the *SotA* methodology [140] in evaluation. Moreover, they used different evaluation metrics rather than following the standards introduced in the *ICDAR2017-POD* competition for Page Object Detection (*POD*). Thus their approach is not considered for comparison in this work.

*Figure detection
from PDF documents*

Recently, systems have been presented for parsing, classifying, and localizing figures from *PDF* documents. Siegel et al. [220] presented a method to parse figures from *PDF* documents, parsed figures are then classified using graph-based *CNNs*. They also present a figure classification dataset, namely "FigureSeer". Clark et al. [44] present another method, "PDFFigures 2.0", to parse and classify figures from *PDF* documents along with a new dataset. Siegel et al. [221] present "DeepFigures", a deep neural method for detecting figures from *PDF* documents. These methods deal with *PDF* documents but not document images as in our case, so they are out of scope for comparison in the presented work.

*Methods chosen for
comparison*

In this chapter, the author of this work considers the methodology proposed by Li et al. [140] for drawing comparisons being the state-of-the-art (*SotA*) system at the time of publication, and with *NLPR-PAL* [72], i.e., the winner of the *ICDAR-2017* competition on Page Object Detection (*POD*) from document images.

*POD using
meta-data from PDF
& DNNs*

After the publication of the research work included in this chapter, recent developments in *POD* domain are also made part of this thesis. Li et al. [138] presented a benchmark suite to train and evaluate cross-domain *POD* models. Each dataset included in the benchmark suite is composed of document images with bounding box annotations for page objects, raw *PDF* files to preserve meta-data for additional information along with page images, and *PDF* rendering layers, comprising of text, vector and raster layers, to preserve structural abstraction of the *PDF* pages. The proposed *POD* model is built on the top of Feature Pyramid Network (*FPN*) object detection network with three additional modules: feature pyramid alignment, region alignment, and rendering layer alignment modules to combine knowledge of natural image domain with document image domain.

*Using ensemble of
DNNs for POD*

An ensemble-based methodology to use different Deep Neural Network (*DNN*) for Page Object Detection (*POD*) task is presented by Vo et al. [245]. The proposed methodology fuses the detection results of two Deep Learning (*DL*) models (*Faster-RCNN* and *RPN*) to take advantage of the strengths of both networks and result in better performance. The authors benchmark the proposed methodology on *ICDAR2017-POD* dataset using Intersection over Union (*IoU*) and mean Average Precision (*mAP*) as evaluation metrics. The authors of the paper did not follow the standard evaluation metrics introduced in

ICDAR₂₀₁₇-POD competition, and therefore evaluations are not directly comparable to prior methodologies.

Li et al. [139] present SelfDoc, a self-supervised, task-agnostic representation learning framework for document images. The proposed framework uses semantic components, such as text block, heading, and figure, as building blocks making full use of linguistic, structural layout, and visual information. The proposed framework implements a cross-modality encoder to enable the cross-modal learning of textual and visual representations. SelfDoc framework learns generic representations from unlabelled data and then, later, fine-tunes for downstream tasks needing significantly fewer data, such as document entity recognition, document classification, and document clustering.

*Self-supervised
learning for
document imaging*

3.3 METHODOLOGY

Multiple Page Object Detection (POD) problems are first introduced in ICDAR₂₀₁₇-POD competition with a release of a public dataset of document images annotating the page objects such as tables, figures, and formulas. DNNs-based object detectors are mainly composed of two main networks to perform the object classification. At the initial stage, CNNs is used as a base network to learn and extract the features during the training phase, followed by a classification network to segment the desired objects from the input image utilizing the features learned in the last step. We also adopt the state-of-the-art (SotA) object detecting methods to classify the page objects from document images taking advantage of transfer learning for domain adoption from the natural image classification networks to document image classification networks. In the first approach, Figure and Formula Detection (FFD) leverage the mask-RCNN and Faster-RCNN as deep models to detect figures and formulas from document images. In the second approach, we introduce Figure and Formula (Fi-fo) Detector, an end-to-end data-driven deep model powered by deformable Feature Pyramid Network (FPN) to extract figures and formulas from document images. We use Fi-fo image representation (instead of the raw image) as input to both networks. The Fi-fo image representation uses distance transform, Connected Component Analysis (CCA), and the original gray-scale image. We stack these three representations together before feeding them to the network; for more details, refer to Section 3.3.1. Detailed methodologies of proposed approaches are explained in the following subsections.

*Overview of
proposed
methodologies*

3.3.1 Fi-fo image representation

Deep Neural Networks (DNNs) dominate the counterparts when it comes to natural scene image processing, whether it is classification,

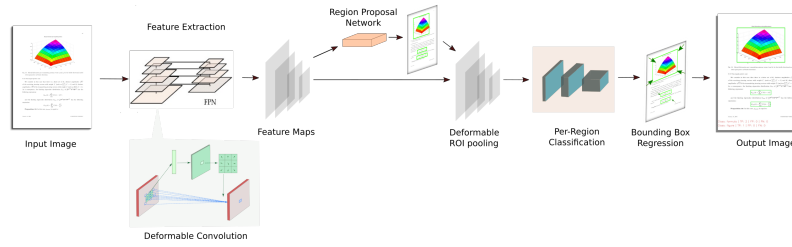


Figure 3.1: Proposed Fi-Fo Detector Outline based on Deformable FPN

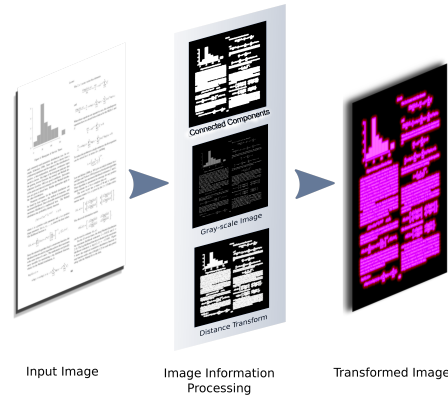


Figure 3.2: Fi-Fo Image representation

segmentation, or object detection. However, document images are very different from natural scene images. Page objects appear at varying positions and differ among the documents depending on the document's format. Therefore, preserving the contextual information is very important, aiding the classifier in learning the desired representation more efficiently. Fi-Fo image representation transforms the document image to appear as close as a natural scene image. It preserves the original image information in the form of a gray-scale image. Image transformation has already been used for Page Object Detection (POD) and segmentation in the past. Ha et al. [92] used vertical projection profiles to extract column regions and draw bounding boxes around connected components. Bukhari et al. [34] used CCA for document image segmentation, while Gilani et al. [79] used distance-based profiles, which were fed to the final classifier to extract the table structure from document images.

Image transformation helps in document image segmentation

Fi-fo Image representation uses color, distance, and CCA transforms

We use color transform, Connected Component Analysis (CCA), and distance transform to generate Fi-fo image representations. The gray-scale image retains the original information of the input image in a single channel. CCA is applied horizontally to identify regions in the image. Distance transform conserves the precise distance between page objects and blank regions. Additionally, we also reverse the distance transform, i.e., the maximum value occurs at the textual regions and diffuses smoothly as a function of the distance to the textual regions. We stack these representations together to feed them

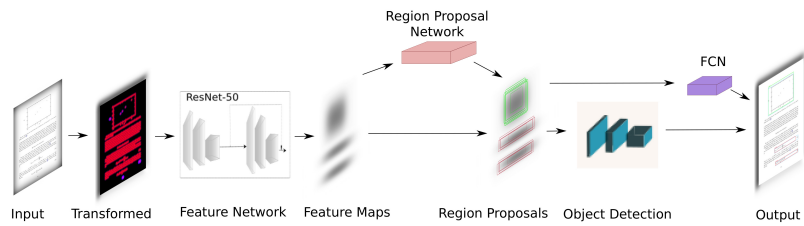


Figure 3.3: FFD pipeline with all its components.

to the network. Figure 3.2 shows the Fi-fo image representation with intermediate information in every channel.

3.3.2 Figure and Formula Detection (FFD)

We adhere to faster-RCNN [197] and mask-RCNN [97] for FFD architecture built upon the pre-trained Residual Network (ResNet)-50 [96] on the ImageNet¹ dataset. Using a pre-trained network enables the proposed approaches of domain adaptation from natural scene images to document images by taking advantage of transfer learning. Transfer learning is a remedy to the need for extensive resources in the training phase for DNNs data-driven nature. Transfer learning is an important aspect of DNNs, as it avoids over-fitting along with better resource utilization because of a useful initialization point.

CNNs used to implement FFD

3.3.2.1 Faster-RCNN

Faster-RCNN [197] has been successfully used for table detection from document images in the recent past [79, 211, 245]. Faster-RCNN is a combination of three networks: a feature extraction backbone, a Region Proposal Networks (RPN) to generate bounding boxes for potential candidates present in an input image, and a region classification network with bounding box regression to classify the Region of Interest (RoI). We refer readers to [197] for further details of faster-RCNN. In FFD, the transformed input image is fed to the feature extraction backbone, which not only generates the feature map but also preserves the shape and structure of the original image. Using pre-trained weights from a state-of-the-art (SotA) image classification network with final layers sheared off is a common practice to overcome the large dataset requirements, as training on these large-scale datasets transforms the initial layers of the network into a generic feature extractor. Pre-trained ResNet-50 [96] up to the final convolutional layer of 4th-stage is used as the feature extractor in FFD.

Faster-RCNN is one of the earliest models to solve complex CV problems

Region Proposal Networks (RPN) predicts bounding boxes of all possible candidate regions commonly termed as anchors and their possibility of being foreground or background based on overlap. It

¹ ImageNet

Region Proposal Networks (RPN) highlights the potential candidates regions

also refines the anchors. Input to RPN is a feature map, the output of the feature extraction backbone. RPN is a small convolutional network, which transforms $x \times x$ spatial input into a lower-dimensional features. These features are used for bounding box regression and classification. In FFD, we used four different anchor scales along with three aspect ratios, resulting in a total of 12 anchors. Multiple anchors help the network overcome variability in terms of size present in real-world objects.

RPN are followed by a detection or classification network, usually known as RCNN. RCNN takes the input from both the feature network and RPN to generate the final class label and bounding box offsets for every input region. By doing so, the detection network crops the features from the feature network using bounding boxes fed from RPN to classify the object present inside the bounding box.

3.3.2.2 Mask-RCNN

Mask-RCNN performs both detection and segmentation task

Mask-RCNN [97] shares the same network of faster-RCNN as explained in Section 3.3.2.1 with an additional module by implementing a Fully Convolutional Network (FCN) in parallel to the last-stage classification network to generate pixel-level binary masks for every Region of Interest (RoI). Using FCN enables the mask-RCNN to encode input objects in a spatial layout by mask representation. It also implements RoI alignment to preserve explicit per-pixel spatial correspondence of input RoI features. With an additional FCN module, mask-RCNN segments the detected object by generating binary masks parallel to bounding boxes and classification scores. Mask-RCNN is used for the very first time to detect objects from document images to the best of the author's knowledge at the time of publication. We refer interested readers to [97] for details of mask-RCNN.

3.3.2.3 Model Configuration

Parameters are vital for optimization of DNNs

Input images are rescaled to the size of $1,000 \times 1,200$ before feeding them to the network. A single image per batch is used. We used the Detectron implementation [80] of both faster-RCNN and mask-RCNN, including pre-trained weights of ResNet-50. Extracted features till the final 4th-stage convolutional layer of pre-trained ResNet-50 is used as the backbone in both models. 4 different anchor scales of $[32 \times 32, 64 \times 64, 128 \times 128, 256 \times 256]$ with 3 aspect ratios of $[1:2, 1:1, 2:1]$ are used in this implementation. All models are trained for 100 epochs with a learning rate of 0.001 with the learning rate scheduling. A non-maximum suppression (NMS) threshold of 0.3 in combination with a class score is used on region proposals for bounding box regression. The confidence threshold to retain the prediction is set to 0.6. All models were trained on a single 1080Ti Graphical Processing Unit (GPU).

3.3.3 Figure and Formula (Fi-fo) Detector

3.3.3.1 Residual Network (ResNet)-101

We used pre-trained ResNet-101 [96] as the backbone of Fi-fo Detector. ResNet-101, as the name implies, consists of 101 convolution layers stacked together in 33 residual blocks, where each block consists of three convolutional layers. As our focus is on the implementation of deformable CNNs for document images, regular ResNet-101 is transformed into its deformable variant. To achieve deformable functionality in ResNet-101, regular higher-level convolution layers namely res(5a,5b,5c)_branch2b are replaced with their deformable counterparts. We initialized deformable layers with zero offsets to benefit from transfer learning, making deformable convolutional layers equivalent to their non-deformable counterparts.

ResNet-101 forms the backbone of Fi-fo Detector

3.3.3.2 Deformable Convolutional Network (DCN)

The proposed method is based on Deformable Convolutional Network (DCN) [48, 287]. Convolutional Neural Networks (CNNs) learn the relevant feature representation depending on the task at hand [269]. These features are extracted in every layer using filters. Filters in lower convolutional layers usually capture textures and preliminary objects, which include gradients, textures, materials, and colors. In contrast, filters in higher convolutional layers describe more abstract objects and their parts [17]. In traditional CNNs, the convolutional layer samples the input feature maps at fixed locations, which the subsequent layers carry forward, resulting in a fixed and known geometric transformation. Using the fixed grid for the detection of objects occurring at different scales and different transformations is not ideal. DCNs address these constraints of traditional CNNs by introducing two additional modules to existing Deep Neural Networks (DNNs), namely (i) the deformable convolution and (ii) deformable RoI-pooling. Regular convolutional layers are augmented with a 2D-offset convolutional layer to form the deformable convolution layer. Regular convolution operates on a uniform grid as its receptive field, whereas deformable convolution leverages the additional offset layers to augment the uniform grid conditioned on the input. The adaptive receptive field allows filters in convolution layers to adapt to different scales and transformations. Since objects like figures and formulas appear at vastly different scales, the deformable property significantly helps cope with these intense input variations. The mathematical formulation of deformable convolution is explained in [48] as:

DCN use an adaptive receptive field to adapt to different scales and transformation

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \times x(p_0 + p_n + \Delta(p_n)) \quad (3.1)$$

Offsets are used for sampling at irregular locations in DCN

where \mathcal{R} defines the offsets from the point under consideration (p_0) in a regular-grid pattern, x represents the input, y represents the output feature map while w represents the filter weights. Considering a 3×3 convolutional layer, the set $\mathcal{R} = (-1, -1), (-1, 0), (-1, 1), (0, -1), (0, 0), (0, 1), (1, -1), (1, 0), (1, 1)$ defines this regular-grid comprising of the 9 positions within the receptive field of the filter. In deformable convolution, sampling is done on irregular locations determined by the offset. The offset is defined as $\Delta(p_n)$, which augments the predefined offsets to deform the receptive field of the filter arbitrarily. Both the features as well as the offsets are learned by back-propagation of gradients. Since these offsets are fractional, they are implemented via bilinear interpolation. For simplicity, let us consider $p = p_0 + p_n + \Delta(p_n)$. Hence, the operation can be represented as:

$$x(p) = \sum_q G(q, p) \times x(q) \quad \text{where} \quad G(q, p) = g(q_x, p_x) \times g(q_y, p_y) \quad (3.2)$$

where q in Eq. (3.2) enumerates all the possible spatial locations on the feature map x , G is the bilinear interpolation kernel and g is defined to be $g(a, b) = \max(0, 1 - |a - b|)$. Region proposals are an integral part of object detection methods, achieved using RoI-pooling, which converts an arbitrary-sized input region into a fixed-size feature representation. Regular RoI-pooling divides the RoI into $k \times k$ spatial bins. Similar to the deformable convolution, deformable RoI-pooling introduces additional offsets to spatial bins. This can be mathematically written as Eq. (3.3).

$$y(i, j) = \sum_{p \in \text{bin}(i, j)} \frac{x(p_0 + p + \Delta p_{ij})}{n_{ij}} \quad (3.3)$$

where $\text{bin}(i, j)$ defines a bin over spatial locations for feature aggregation ($\lfloor i \frac{w}{k} \rfloor \leq p_x < \lceil (i+1) \frac{w}{k} \rceil$, $\lfloor j \frac{h}{k} \rfloor \leq p_y < \lceil (j+1) \frac{h}{k} \rceil$), and n_{ij} represents the number of items in $\text{bin}(i, j)$. We refer readers to [48, 287] for a comprehensive introduction to the deformable convolutional layers.

3.3.3.3 Network architecture

Three different variants of DCN are used to evaluate their efficacy

The proposed Fi-fo detector is based on deformable Feature Pyramid Network (FPN) [143], which integrates features from multiple scales within a single forward pass, transforming it into a faster variant of multi-scale detection. This makes it capable of better-handling objects of small sizes. As a comparison, we also include results from deformable Faster-RCNN [197] and deformable Region-based Fully Convolutional Networks (RFCN) [49], which were the most dominant architectures before FPN. All these models are augmented with deformable convolutions along with the replacement of conventional RoI-pooling layer with deformable RoI-pooling.

The deformable convolutions explicitly generate offsets for every location in feature maps, making the process a memory-intensive operation. Therefore, all the models used for our experiments are built upon the ResNet-101, converted to a deformable network by replacing 3 higher level traditional layers into deformable counterparts to aid multi-scale feature extraction. This adoption enables us to leverage a deformable ResNet-101 as the base model for all the models used in our experiments.

The performance of DNNs rely heavily on the amount of data available for training, making them data-hungry [230]. Since the initial layers of the network are generic feature extractors, the initial layers trained on a large corpus of images are adapted as the feature extractor in our case, which are fine-tuned for the document analysis task during training. This is commonly referred to as transfer learning in the literature, where the learned knowledge is transferred from one problem to another [219].

Transfer learning helps to achieve better results and domain adaptation

3.3.3.4 Model Configuration

We used deformable ResNet-101 as the backbone of our deformable detection models, along with model weights trained on the ImageNet dataset as described previously. After deformable pooling, we keep the rest of the object detection pipeline intact- including per-region classification and bounding-box regression. Using pre-trained weights enables the proposed approach for domain adaptation from natural scene images to document images. We trained three different variants of deformable models, which include deformable Faster-RCNN, deformable FCN, and deformable FPN. We used three different anchor ratios for all our models and were set to [0.5, 1, 2]. We used five different anchor scales for RFCN and Faster-RCNN set to [2, 4, 8, 16, 32]. FPNs have built-in features for multi-scale detection because of their top-down architecture, so only a single anchor scale of [8] is used. We trained our models for 50 epochs with a learning rate of 0.000125 (with a learning rate schedule). We used aspect-aware image resizing with a max image size of $1,280 \times 800$. All models were trained on a single NVIDIA V-100 GPU.

Network configuration for different variants of DCN

3.4 DATASETS

This section covers the details of the datasets along with evaluation protocols followed during this research. The following section familiarises the reader of this thesis with the details of publicly available POD datasets for document image processing. Moreover, this section also covers the details of evaluation metrics followed as standard for POD problem.

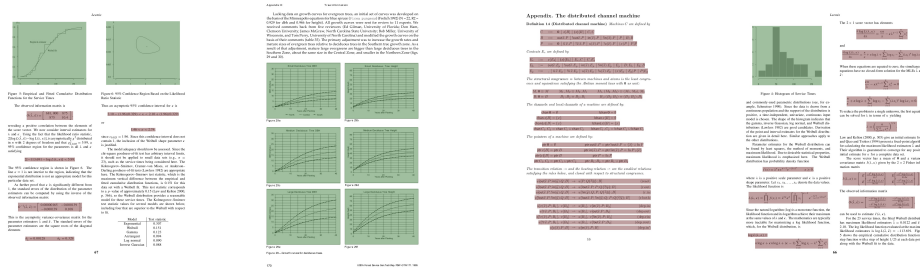


Figure 3.4: Annotated examples from ICDAR2017 Page Object Detection (ICDAR2017-POD) dataset (formulas are labeled as the brown color and figures as green color).

3.4.1 ICDAR2017 Page Object Detection (ICDAR2017-POD)

ICDAR2017-POD is benchmark dataset for POD problem from document images

We used the publicly available ICDAR2017 Page Object Detection (ICDAR2017-POD) competition dataset [72] to benchmark the performance of our model. ICDAR2017-POD was released recently for a competition focused on the figure, formula, and table detection from document images. The dataset is comprised of page document images from 1,500 scientific papers available at CiteSeer². This dataset comprises 2,417 document images in the English language, segregated into 1,600 train and 817 test document images. The dataset exhibits high variability in terms of format and page layout. Page layout styles include single-column, double-column, and multi-column pages. Various formulas, figures, tables, and other page objects are spread across the document images.

The page objects include textual content, page title, captions, headings, etc., but only figures, tables, and formulas were annotated for the task. Every document image is accompanied by a corresponding *.xml* file in PASCAL-VOC format annotated ground-truth. Page objects are annotated by rectangular coordinates to generate bounding boxes. Figure 3.4 shows some images from the ICDAR2017-POD dataset along with the corresponding ground-truth information.

3.4.1.1 Faulty Annotations

Annotation issues of ICDAR2017-POD dataset

Initial experiments led us to the discovery of the problems in annotations of the ICDAR2017-POD dataset. Some of these problems are highlighted in Figure 3.5, where green, blue, and brown color annotate figures, formulas, and tables, respectively. There were missing annotations, which include formulas, figures, and tables, as shown in Figure 3.5a. There were occasions where page objects were mislabelled, i.e., formulas annotated as tables, text annotated as formulas, and so on. As shown in Figure 3.5c, text lines are labeled as formulas,

² CiteSeer

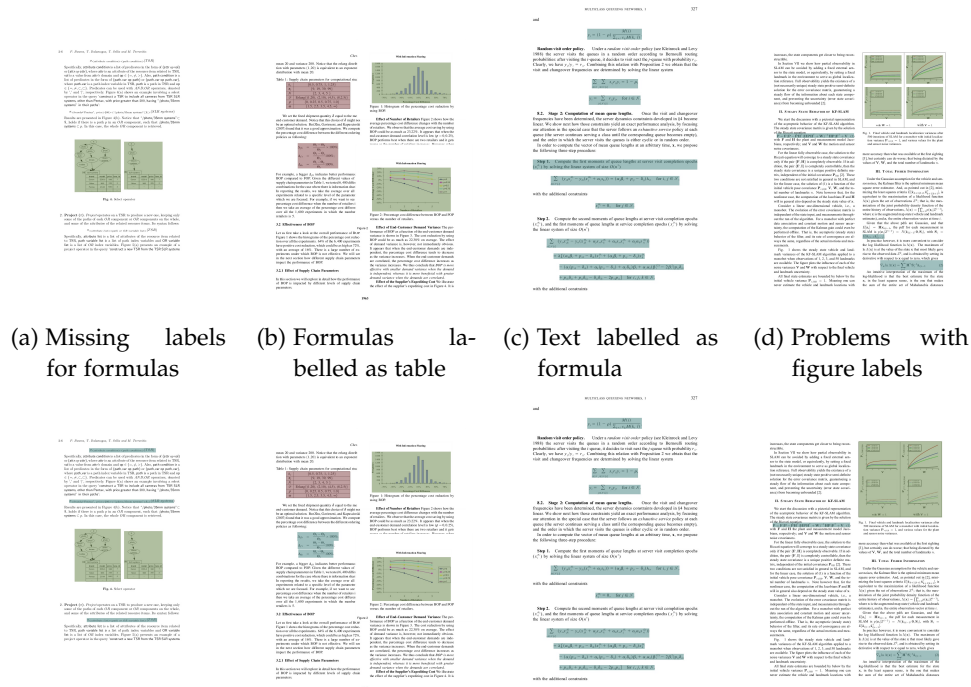


Figure 3.5: Problems with ICDAR2017-POD dataset (1st row) in comparison to ICDAR2017-POD (corrected) dataset (2nd row) (green, blue, and brown color represent figures, formulas, and tables, respectively, thanks to Fi-fo detector)

confusing the system to distinguish text from formulas. Similarly, in Figure 3.5b, a block of formulas is labeled as a table, again creating vagueness for the system. There were inconsistencies in the figure’s annotation, as shown in Figure 3.5d. Common problems with figure annotation include over-segmentation, under-segmentation, and inconsistent labeling for multi-panel figures, i.e., in some instances, objects enclosed inside a solid boundary are annotated as a single figure considering the outer bounding box. In contrast, the outer boundary is neglected for other similar instances, and enclosed objects are annotated as individual figures. These irregularities in original annotations penalized the data-driven systems, which resulted in overall performance degradation. So, it is very important to cleanse the dataset to avoid problems for the research community in the future.

3.4.1.2 ICDAR2017 Page Object Detection (ICDAR2017-POD) (corrected)

While evaluating the ICDAR2017 Page Object Detection (ICDAR2017-POD), *Annotations are corrected, updated, and made publicly available* Fi-fo Detector highlighted the problem in the ground truth of the dataset such as missing, confused, and irregular labels. As Fi-fo Detector is a data-driven approach, it relies not only on the quantity of data but also on the quality of data is very important. Confusions in the

Table 3.1: Overview of [ICDAR2017-POD](#) (corrected) with class-wise comparison and modifications in [ICDAR2017-POD](#) dataset.

Class	ICDAR2017-POD	ICDAR2017-POD (corrected)	# of files modified
Figure	2939	2912	135
Formula	5427	5463	156
Table	1016	1053	30

given data penalize the system resulting in degrading performance, as in the original [ICDAR2017-POD](#) dataset. Therefore, we manually inspected the entire dataset to update missing annotations and fine-tune the confusing ones. While updating the [ICDAR2017-POD](#) dataset, we did not add or remove any image from the dataset. Rather, we only updated the annotations to minimize the ambiguities and inconsistencies present in the dataset following the existing labeling conventions, as shown in [Figure 3.5](#). In the [ICDAR2017-POD](#) dataset, some decorative graphics were annotated as figures, and the rest were ignored. So, the only thing we completely changed is the removal of decorative graphics as figures from the original dataset. [Figure 3.10c](#) is a perfect example to demonstrate where the header-rule line was annotated as a figure, but the footer-rule line was ignored. All these problems contribute to the performance degradation of any data-driven system.

Bringing consistency and clarity in the annotations will help in achieving generalized systems for [POD](#) problem. 273 annotation files were updated in total. The updation included the removal of false labels, updating missing labels, and bringing uniform labeling conventions. An overview of the [ICDAR2017-POD](#) (corrected) dataset along with the [ICDAR2017-POD](#) dataset is presented in [Table 3.1](#). The updated annotations have been publicly released as the [ICDAR2017-POD](#) (corrected) dataset to aid future research in this direction³.

3.4.2 *Figure and Formula Detection (FFD) Dataset*

Multiple datasets help in the cross-evaluation of data-driven systems

[ICDAR2017-POD](#) competition dataset [72] was the largest publicly available dataset for Page Object Detection ([POD](#)) to the best of the authors' knowledge during the course of this research work. There is a real need for a publicly available dataset for cross-evaluation and to achieve generalization for data-driven systems, in particular. Therefore, we collected and manually annotated a dataset named [FFD](#), particularly targeted toward formulas and figure detection. The dataset consists of 680 document images from 100 scientific papers in the English language available at *arXiv*⁴. The collected document images were taken from journals and/or conferences of different disciplines to cover a variety of page formats, layouts, and styles. Page objects present in every document image also show diversity and variability.

³ [ICDAR2017-POD](#) (corrected)

⁴ [Arxiv](#)

Table 3.2: Dataset content details including numbers of objects present in training and test set

Split	document images	figures	formulas
Train	480	681	1,212
Test	200	308	708
Total	680	989	1,929

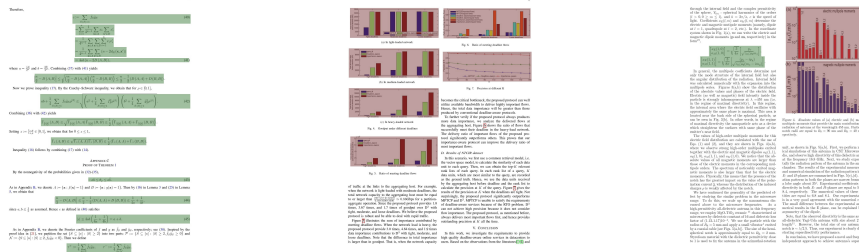


Figure 3.6: Examples of annotated document images from the FFD dataset; green colour annotates formulas, brown colour represents figures

We manually annotated only two classes, i.e., figures and formulas. Example document images from FFD dataset are visualized in Figure 3.6. The 1,929 formulas and 993 figures are in the FFD dataset. Every document image carries a corresponding *.xml* with annotated ground-truth information in PASCAL-VOC format. Out of 680 document images, 70% are used in the training set, and the remaining 30% are placed in the test set. FFD dataset will be made publicly available for the research community to aid research in this direction⁵.

3.4.3 PublayNet

Zhong et al. [284] publish the largest dataset for document layout analysis called PubLayNet. The dataset consists of over 360K automatically annotated page images. The dataset annotates five page objects, i.e., text, title, list, figure, and table. One of the major annotations missing in the PubLayNet dataset is "formula" objects. The availability of a large dataset for document layout analysis is a key performance factor in data-driven methods to automate the process of Page Object Detection (POD). PubLayNet dataset is published during or after the publication of this work. Therefore, it is not included in the evaluation.

PubLayNet is the largest publicly available POD dataset

⁵ JacFFD Dataset

3.4.4 Evaluation Protocol

Following standard evaluation protocols helps in fair comparison

We follow the evaluation protocol defined for all our presented methods in the [ICDAR2017-POD](#) competition. We compute true positives (TP), false positive (FP), and false negatives (FN) during the testing phase. These results are computed by evaluating the test set on Intersection over Union (IoU) threshold of 0.6, & 0.8 for calculation of the given metrics, which means at least 60% of the object is correctly detected in prior and 80% in later threshold settings. Results are reported using the metrics of precision, recall, f1-score, Average Precision (AP), and mean Average Precision (mAP).

Precision

The precision metric evaluates how accurate are the system's predictions. It is calculated as follows:

$$\text{Precision} = \frac{\text{correct detection}}{\text{total detection}} \quad (3.4)$$

Recall

The recall metric is the measure of how well a system performs in finding all positive examples (TPs) given in the test set. It is given by,

$$\text{Recall} = \frac{\text{correct detection}}{\text{total ground-truth annotations}} \quad (3.5)$$

mean Average Precision (mAP)

mean Average Precision (mAP) is computed as an average of maximum precision at different recall levels. The mathematical formulation of mean average precision is given as:

$$\text{mAP} = \frac{1}{|Q|} \sum_{r=1}^Q \text{AP}_r \quad (3.6)$$

All of the results reported in this thesis chapter are generated using the official evaluation code provided by the [ICDAR2017-POD](#) competition organizers.

3.5 RESULTS AND DISCUSSIONS

This section covers the comprehensive evaluation results of the proposed approaches in this chapter. Moreover, this section also discusses the merits and weaknesses of the proposed methods, along with a detailed comparison of existing [SotA](#) methods.

3.5.1 Results of [FFD](#) approach

We evaluate the presented approach using [faster-RCNN](#) and [mask-RCNN](#) on [FFD](#) (our collected dataset) and the publicly available [ICDAR2017-POD](#)

Table 3.3: Comparison of FFD approach with existing state-of-the-art (SotA) methods using ICDAR2017-POD and FFD dataset

Method	Class	IoU = 0.6				IoU = 0.8			
		Precision	Recall	f1-score	AP	Precision	Recall	f1-score	AP
NLPR-PAL [72]	Formula	0.901	0.929	0.915	0.839	0.888	0.916	0.902	0.816
ICDAR2017-POD	Figure	0.920	0.933	0.927	0.849	0.892	0.904	0.898	0.805
Li et al. [140]	Formula	0.930	0.953	0.942	0.878	0.921	0.944	0.932	0.863
ICDAR2017-POD	Figure	0.948	0.940	0.944	0.896	0.921	0.913	0.917	0.850
Faster-RCNN	Formula	0.894	0.889	0.897	0.873	0.760	0.570	0.650	0.671
ICDAR2017-POD	Figure	0.894	0.900	0.897	0.862	0.811	0.801	0.806	0.787
Mask-RCNN	Formula	0.894	0.921	0.907	0.897	0.788	0.835	0.811	0.776
ICDAR2017-POD	Figure	0.894	0.918	0.905	0.886	0.805	0.828	0.816	0.794
Faster-RCNN	Formula	0.916	0.89	0.903	0.875	0.596	0.577	0.591	0.448
FFD dataset	Figure	0.890	0.899	0.894	0.851	0.770	0.781	0.776	0.750
Mask-RCNN	Formula	0.898	0.913	0.905	0.892	0.711	0.723	0.717	0.621
FFD dataset	Figure	0.908	0.905	0.906	0.894	0.809	0.814	0.811	0.791

dataset for Page Object Detection (POD). We report results on the standard IoU threshold of 0.6 and 0.8 defined for the ICDAR2017-POD competition. Mask-RCNN delivers the best results for figure detection on IoU threshold of 0.6, f1-score of 0.906 with a precision of 0.908 and recall of 0.905.

Faster-RCNN detected figures on the IoU threshold of 0.6 with the precision and recall of 0.89 and 0.899, which translated into f1-score of 0.894 while evaluating on the FFD dataset. On the IoU threshold of 0.6, formulas are detected with the precision of 0.916, recall of 0.89 and f1-score of 0.903. When IoU threshold is increased to 0.8, numbers for faster-RCNN on formulas detection dropped to the f1-score of 0.591 with a precision and recall of 0.596 and 0.577, respectively. Similarly, results for figure detection also drop to 0.77, 0.781, and 0.776 in terms of precision, recall, and f1-score, respectively. Average Precision (AP) for formula detection is 0.875 and 0.851 for figure detection.

Faster-RCNN results

Mask-RCNN produced better results in comparison to faster-RCNN for both figure and formula detection, as shown in Table 3.3. Figures are detected with a precision of 0.908, and recall is translated into numbers as 0.905 and f1-score measures to 0.906 on the IoU threshold of 0.6. Formulas are detected with a precision of 0.898, and recall, and f1-score are 0.913 and 0.905, respectively. Precision for figure detection on IoU 0.8 is calculated as 0.809 with a recall of 0.814 and f1-score of 0.811. Numbers for formula detection realized to 0.711, 0.723, and 0.717 as precision, recall, and f1-score, respectively. AP for figure detection comes as 0.894 and that for formulas is 0.892.

Mask-RCNN results

On the ICDAR2017-POD dataset, FFD approach performed equally well as the results show in Table 3.3. On the IoU threshold of 0.6, both faster-RCNN and mask-RCNN were competitive in terms of performance. When IoU threshold is increased to 0.8, a significant drop

FFD results on ICDAR2017-POD dataset

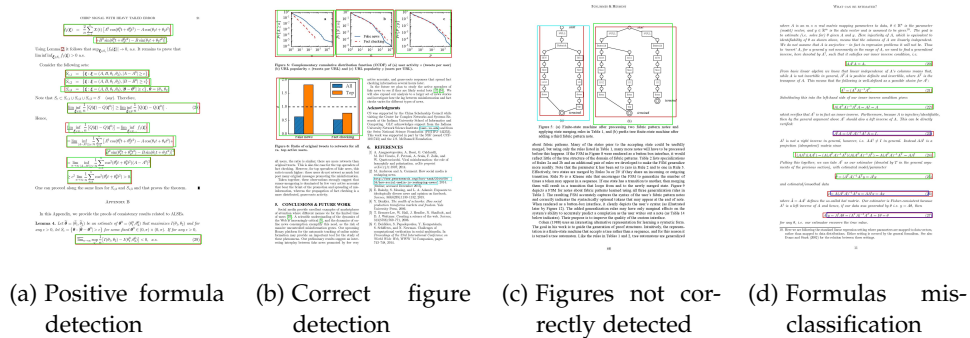


Figure 3.7: Analysis of results generated by FFD detector using Faster-Region Based Convolutional Neural Networks (RCNN) on FFD dataset, red colour annotates ground truth and false negatives (FN), green colour highlights true positives (TP), cyan annotates false positive (FP) for figures and blue colour represents false positive (FP) for formulas.

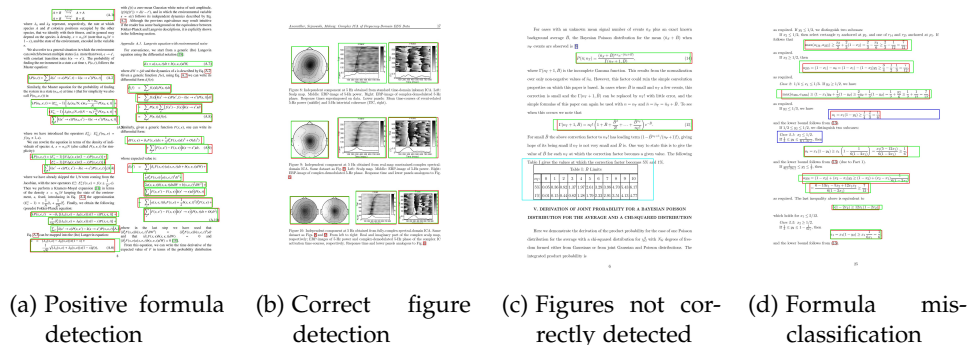


Figure 3.8: Analysis of results generated by FFD detector using mask-RCNN on FFD dataset, red colour annotates ground truth and false negatives (FN), green colour highlights true positives (TP), cyan annotates false positive (FP) for figures and blue colour represents false positive (FP) for formulas.

Table 3.4: Comparison of *Fi-fo* Detector with existing state-of-the-art (*SotA*) methods using *ICDAR2017-POD* annotations both for training and testing.

ICDAR2017-POD									
Method	Class	IoU = 0.6				IoU = 0.8			
		Precision	Recall	f1-score	AP	Precision	Recall	f1-score	AP
NLPR-PAL [72]	Formula	0.901	0.929	0.915	0.839	0.888	0.916	0.902	0.816
	Figure	0.920	0.933	0.927	0.849	0.892	0.904	0.898	0.805
Li et al. [140]	Formula	0.93	0.953	0.942	0.878	0.921	0.944	0.932	0.863
	Figure	0.948	0.940	0.944	0.896	0.921	0.913	0.917	0.85
Deformable Faster-RCNN	Formula	0.882	0.738	0.803	0.660	0.638	0.534	0.582	0.337
	Figure	0.929	0.872	0.899	0.660	0.855	0.802	0.828	0.720
Deformable RFCN	Formula	0.914	0.918	0.916	0.915	0.832	0.836	0.834	0.826
	Figure	0.904	0.920	0.912	0.903	0.86	0.875	0.867	0.864
<i>Fi-fo</i> detector	Formula	0.909	0.927	0.918	0.911	0.856	0.878	0.867	0.854
	Figure	0.918	0.883	0.90	0.894	0.871	0.838	0.854	0.861

in numbers for formulas detection using faster-RCNN is observed in comparison to mask-RCNN. On the *IoU* threshold of 0.6, mask-RCNN recognized figures with a precision of 0.894 against the recall of 0.918, and *f1*-score is computed as 0.905. Formulas are detected with precision, recall, and *f1*-score of 0.894, 0.921, and 0.907, respectively. On the *IoU* threshold of 0.8, figures and formulas are detected with the *f1*-score of 0.816 and 0.811, respectively.

3.5.2 Results of *Fi-fo* Detector approach

Detailed results of the *Fi-fo* detector are furnished in the Table 3.4 and Table 3.5. We evaluate the deformable variants of Faster-RCNN, RFCN, and FPN using both the original and updated annotations of *ICDAR2017-POD* dataset. Feature Pyramid Network (FPN) forms the basis of the *Fi-fo* detector as it outperformed other deformable variants for object detection problems, owing to its multi-scale detection capabilities. On *IoU* threshold of 0.6, formulas were detected with the precision and recall of 0.909 and 0.927, along with an *f1*-score and Average Precision (AP) of 0.918 and 0.911, respectively. Once the *IoU* threshold was increased to 0.8, precision and recall dropped to 0.856 and 0.878 with *f1*-score and AP of 0.867 and 0.854, respectively. Considering figures, the obtained numbers were 0.918, 0.883, 0.90, and 0.894 in terms of precision, recall, *f1*-score, and AP, respectively. For the *IoU* threshold of 0.8, precision, recall, and *f1*-score went down to 0.871, 0.838, and 0.854, respectively. Using *ICDAR2017-POD* dataset, visual results looked convincing, as both formulas and figures were detected properly, but numbers were surprisingly not up to the expectations keeping in mind the potential of FPN implemented with deformable convolution, as shown in Table 3.4.

Results produced by *Fi-fo* detector on *ICDAR2017-POD* (corrected) outperformed the existing *SotA* system by a large margin. On *IoU* thresh-

Fi-fo detector provides *SotA* results for figure & formula detection

Table 3.5: Comparison of **Fi-fo** Detector with existing **SotA** methods using **ICDAR2017-POD** in training and **ICDAR2017-POD** (corrected) for testing.

Trained:ICDAR2017-POD, Tested:ICDAR2017-POD (corrected)									
Method	Class	IoU = 0.6				IoU = 0.8			
		Precision	Recall	f1-score	AP	Precision	Recall	f1-score	AP
Li et al. [140]	Formula	0.935	0.331	0.489	0.312	0.877	0.310	0.459	0.274
	Figure	0.918	0.292	0.443	0.271	0.888	0.283	0.429	0.253
Fi-fo detector	Formula	0.949	0.945	0.947	0.967	0.897	0.893	0.895	0.941
	Figure	0.930	0.932	0.931	0.97	0.899	0.900	0.899	0.952

Comparison of **Fi-fo** detector with existing **SotA** approaches

old of 0.6, **Fi-fo** detector achieved the **f1-score** of 0.947 with the Average Precision (**AP**) of 0.967 in comparison to Li et al. [140] system’s **f1-score** of 0.489 and **AP** of 0.312, for formula detection. Similarly, there is an enormous difference in the results of figure detection produced by **Fi-fo** detector and Li et al. [140] on **ICDAR2017-POD** (corrected). **Fi-fo** detector detected the figures with **f1-score** of 0.931, whereas Li et al. methods results in **f1-score** of 0.271, as shown in Table 3.5. On **IoU** threshold of 0.8, **AP** for Li et al. was computed to 0.253 and 0.274 in comparison to **Fi-fo** detector’s Average Precision (**AP**) score of 0.941 and 0.952 for figure and formula detection, respectively.

3.5.3 Ablation Study

Fi-fo image representation helps **DNNs** learn better

We present an ablation study, which covers the potential of **Fi-fo** image representation compared to raw image representation, along with an analysis of deformable networks concerning non-deformable counterparts. Starting with Red, Green, Blue (**RGB**) images, we noticed the problems in annotations. Fixing the annotations resulted in minor improvements but was still not impressive, as highlighted in Table 3.6, which took us to image information processing, which resulted in significant improvement in performance. Moreover, we present a comparative analysis to establish the utility and effectiveness of the **ICDAR2017-POD** (corrected) dataset with the existing state-of-the-art (**SotA**) systems.

Fi-fo image representation in combination with deformable neural networks

To further validate the performance and potential of **Fi-fo** detector, the proposed ablation study also covers a comparative analysis of the potential of deformable and non-deformable neural networks. We observed a clear performance boost from raw image representation to **Fi-fo** image representation and non-deformable neural networks to deformable neural networks; results are furnished in Table 3.6. The **SotA** performance is achieved using the combination of **Fi-fo** image representation with deformable neural networks. Using an **IoU** threshold of 0.6 as per **ICDAR2017-POD** competition standards, we achieved **SotA** for formula detection with a precision of 0.957 and recall of 0.952 which translates to **f1-score** of 0.954. Results with an **IoU** of 0.8 are 0.913,



Figure 3.9: Analysis of results generated by **Fi-fo** Detector using **ICDAR₂₀₁₇-POD** (corrected) dataset, green colour highlights true positives (**TP**), blue colour signifies false positive (**FP**) for formulas, magenta color flags false positive (**FP**) for figures, and red color annotates false negatives (**FN**) for both classes.

Table 3.6: An ablation study on the performance of **Fi-fo** detector using raw image representation, **Fi-fo** image representation, using non deformable Feature Pyramid Network (**FPN**), and deformable **FPN**, both for **ICDAR₂₀₁₇-POD** and **ICDAR₂₀₁₇-POD** (corrected).

ICDAR ₂₀₁₇ -POD										
Method	Image Representation	Class	IoU = 0.6				IoU = 0.8			
			Precision	Recall	f1-score	AP	Precision	Recall	f1-score	AP
Fi-fo detector Deformable	Raw	Formula	0.867	0.918	0.892	0.893	0.780	0.826	0.802	0.780
		Figure	0.860	0.869	0.864	0.847	0.818	0.827	0.822	0.799
Fi-fo detector Non Deformable	Fi-fo	Formula	0.867	0.874	0.871	0.917	0.712	0.694	0.703	0.837
		Figure	0.856	0.821	0.838	0.929	0.801	0.739	0.769	0.889
Fi-fo detector Deformable	Fi-fo	Formula	0.909	0.927	0.918	0.911	0.856	0.878	0.867	0.854
		Figure	0.918	0.883	0.90	0.894	0.871	0.838	0.854	0.861
ICDAR ₂₀₁₇ -POD (corrected)										
Fi-fo detector Deformable	Raw	Formula	0.949	0.945	0.947	0.973	0.897	0.893	0.895	0.967
		Figure	0.930	0.932	0.931	0.971	0.897	0.90	0.899	0.959
Fi-fo detector Non Deformable	Fi-fo	Formula	0.910	0.927	0.918	0.953	0.860	0.877	0.868	0.928
		Figure	0.879	0.822	0.850	0.948	0.847	0.792	0.819	0.958
Fi-fo detector Deformable	Fi-fo	Formula	0.957	0.952	0.954	0.949	0.913	0.908	0.910	0.898
		Figure	0.931	0.913	0.922	0.905	0.901	0.885	0.893	0.870

0.908, and 0.91 in terms of precision, recall, and f1-score for formula detection, respectively. In figure detection, at **IoU** threshold of 0.6, **Fi-fo** detector achieved a precision of 0.931, recall of 0.913 along with an f1-score of 0.905. Setting the **IoU** threshold to 0.8 translated into a precision of 0.901, recall of 0.885, and f1-score of 0.893. In terms of Average Precision (**AP**), **Fi-fo** detector outperformed other methods by a significant margin. At **IoU**=0.6, Average Precision (**AP**) of figure and formula detection was found to be 0.949 and 0.905, while at **IoU**=0.8, the **AP** of 0.898 and 0.870 for formulas and figures was achieved, respectively.

3.5.4 Discussions

The proposed approaches **FFD** and **Fi-fo** detector in this chapter present promising results and encourage the need for end-to-end and data-driven approaches for **POD** problem. Both approaches establish the connotation of object detection deep models for document images. In the case of the **FFD** approach, a comprehensive and detailed evaluation is presented using different object detection networks across multiple datasets. Results furnished in [Section 3.5.1](#) demonstrate the convergence strength of **FFD**, as it achieves competitive results on the **FFD** dataset in comparison to the **ICDAR2017-POD** (about three times larger) dataset, keeping in mind **DNNs** are data-driven methods. Moreover, **FFD** approach can be adapted to any real-world scenario for figure and formulas detection with minimal efforts, such as re-training the proposed with a small number of examples enabling the system adoption to multiple scenarios, validated by results furnished in [Section 3.5.1](#). The results of **FFD** approach are further improved by **Fi-fo** detector approach, as discussed in [Section 3.5.2](#) and [Section 3.5.3](#), which is still state-of-the-art (**SotA**) approach for figure and formula detection from document images upto the best of authors knowledge. In this section, we will discuss the merits and demerits of **Fi-fo** detector further.

End-to-end and data-driven approaches perform better for POD problem

Strengths of Fi-fo detector

As exhibited in [Figure 3.10](#), **Fi-fo** detector works fine, correctly detecting the page objects, i.e., figures and formulas in particular, but the same is not reflected in terms of numbers. Upon investigating the results, we discovered the irregularities and inconsistencies in the original annotations available for the **ICDAR2017-POD** dataset. There were clear examples of missing annotations for page objects, as the case in [Figure 3.10b](#)- where annotations for formulas were missing. Confusions between figure and table annotations are briefly shown in [Figure 3.10d](#). [Figure 3.10c](#) establishes the case of inconsistent labelling for figure annotations. There were examples of over-segmented ground truth where captions or text lines were annotated along with a figure or table, as shown in [Figure 3.5c](#). These inconsistencies have been discussed in detail in [Section 3.4.1.2](#).

Data-driven approaches have a clear edge over heuristic-defined methods

Problems in the original annotations of the **ICDAR2017-POD** dataset led us to update the annotations, which included removing discrepancies, confusion, and adding missing labels. After addressing the problems found in **ICDAR2017-POD** dataset, a clean dataset is presented as **ICDAR2017-POD** (corrected). To establish a fair comparison with existing **SotA** methods, it is necessary to present their results on **ICDAR2017-POD** (corrected). As Li et al. [140] method is not an end-to-end system, it combines trainable and heuristic-based parts. Upon request, the authors [140] excused themselves from providing us with their system because of its complexities but provided the result files on **ICDAR2017-POD**

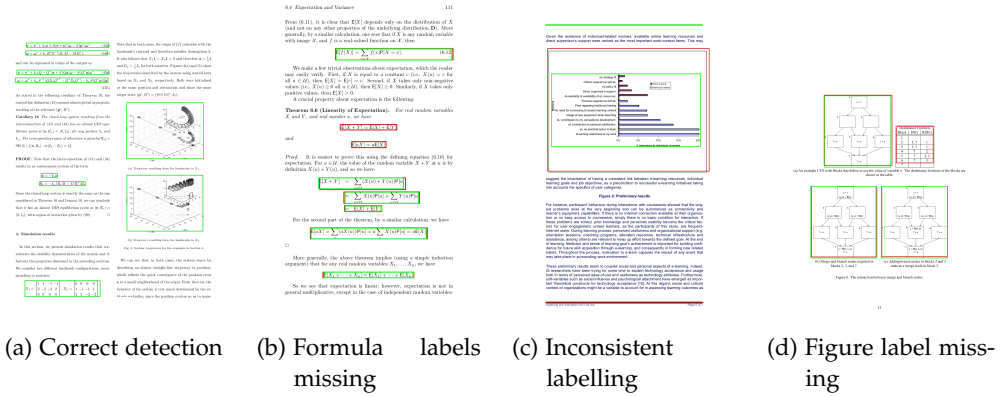


Figure 3.10: **Fi-fo** detector evaluation results on **ICDAR2017-POD** dataset with original annotations (green color represents annotated objects; while red color highlights the detection by **Fi-fo** detector but missing in original annotations).

dataset for comparison. It limits the scope of comparison on the corrected dataset, as both the system and results on a corrected dataset from the Li et al. [140] were not accessible. So, in the given circumstances, we opted for the best possible way to establish a fair comparison with the existing **SotA** system. We trained **Fi-fo** detector using **ICDAR2017-POD** dataset. At the same time, evaluation is performed on **ICDAR2017-POD** (corrected) for both **Fi-fo** detector, and Li et al. [140], and the results are furnished in Table 3.5. One of the potential reasons for a significant decline in the performance of the Li et al. method on updated annotations might be its inability as an end-to-end system. Since the dataset had inconsistencies, **SotA** methods had to leverage hand-defined heuristics, which catered for these inconsistencies. Upon removal of these inconsistencies from the dataset, the heuristics themselves had to be adapted and updated, which is a major shortcoming of heuristics-based systems. Given results establish the weakness of the existing **SotA** system both on the **ICDAR2017-POD** as well as the **ICDAR2017-POD** (corrected) dataset. Using **ICDAR2017-POD**, their system failed to detect and highlight the missing or wrong labels. Similarly, it failed to capitalize on the **ICDAR2017-POD** (corrected). It is worth emphasizing again that all the reported results on the original and updated annotations in this chapter are generated using the official evaluation code released by organizers of the **ICDAR2017-POD** competition [73].

Despite reporting significantly high metric scores, Li et al. [140] method could not find any inconsistencies in the **ICDAR2017-POD** dataset. Their approach relies heavily on heuristics and pre/post-processing, specifically tuned to the inconsistencies in the dataset. This is one of the primary reasons why purely data-driven techniques were not found to be very effective for this dataset compared to hand-defined

Fi-fo detector has clear edge over existing SotA method

heuristics. This provides a clear edge to *Fi-fo* detector over Li et al. [140] method, the existing state-of-the-art (*SotA*) method for Page Object Detection (*POD*), in terms of generalization. *Fi-fo* Detector demonstrated to be a generic network by pointing out discrepancies in the *ICDAR2017-POD* dataset, highlighted in *Figure 3.10*. Secondly, *Fi-fo* detector does not rely on any pre/post-processing, rather simple image transforms providing a clear edge not only in terms of efficiency and computation costs but also delivering the *SotA* results, as shown in *Figure 3.9a* & *Figure 3.9b*. We also report results from the ablation study where we removed components from the *Fi-fo* detector to identify the contribution of the individual components to the system.

Deformable networks results in better performance for POD from document images

The obtained results highlight the superiority of the deformable model family for Page Object Detection (*POD*) task, where the models either outperformed or achieved performance on par with heuristic-based methods. Moreover, it is also demonstrated that the *Fi-fo* detector shows progressive performance moving from *ICDAR2017-POD* to *ICDAR2017-POD* (corrected) at every subsequent step. In contrast, existing *SotA* failed to do so, as shown in the *Table 3.6*. Since *Fi-fo* detector is a data-driven approach, significant improvements in performance could be achieved by increasing the amount of training data. Results can further be boosted by post-processing particularly for the detected figure regions using Computer Vision (*CV*) approaches, considering the case of *Figure 3.9c* into account.

Fi-fo detector still remains SotA for figure & formula detection

In this chapter we proposed multiple approaches to detect figures and formulas from document images using a novel combination of *Fi-fo* image representation and *DNNs*. The *SotA*s results made part in this chapter establish the utility of *Fi-fo* image representations to complement the performance of Deep Learning (*DL*) models. We did fine-tune the annotations of *ICDAR2017-POD* and make them publicly available as *ICDAR2017-POD* (corrected) along with a complete newly curated *FFD* dataset to demonstrate the generic behaviour of the proposed approaches. Moreover, few approaches [95, 245] have been presented after the publication of *Fi-fo* detector, but *Fi-fo* detector still remains the state-of-the-art (*SotA*) performer in detecting page objects from document images, figures and formulas in particular at the time of writing this thesis.

During the 21st century, Machine Learning (ML) and Deep Neural Networks (DNNs) showed a phenomenal stride in various fields of life. These methods are widely adopted to use technological breakthroughs to innovate common practices in formal education. Using Artificial Intelligence (AI) methods in formal education can lead to better data analysis and foresight to assist and improve learning and teaching experiences. Writing is one of the fundamental classroom activities, and it is very important to investigate the writing behaviour of the learners to evaluate the learning process. Students in general, and Science, Technology, Engineering, and Mathematics (STEM) education in particular, are required to understand and express their knowledge skills to interpret complex relationships using different writing types such as text, formulas, and drawing plots/graphs. Existing systems such as Optical Character Recognitions (OCRs) and other handwriting recognition tools commonly used to process handwriting information work well with a single type of writing, such as text or math, but they fail when input is a combination of different writing types. This dissertation chapter focuses on the methods to classify writing into text, mathematical expressions, and plot/graph classes. The proposed methods can serve as a preliminary step for handwriting recognizers. Furthermore, the proposed methods can be used to analyse the writing behaviour of students to look into the individual's progress and to provide customized feedback based on the individual's strengths and weaknesses.

*Importance of
classifying of
writing into
multiple types*

The major contribution of this chapter to the classification of online written sequences is highlighted as follows:

- Curation of a novel online handwriting dataset "onTabWriter" captured using a digital/sensor pen (Apple pencil) and digital/sensor screen (iPad). The captured data are continuous streams of multi-dimensional points analysed and processed to classify handwritten sequences into plain text, mathematical expressions, and plots/graphs.
- Presentation of a new feature set for online handwritten sequence classification. The proposed feature set consists of 49 features incorporating the writing style, shape, size, speed, and variations factor. To the best of the authors' knowledge, the presented feature set has not previously been used for handwriting classification.

- Benchmarking the proposed feature set using sense the pen dataset on various Machine Learning (ML) and Deep Learning (DL) classifiers along with the comprehensive comparisons with existing state-of-the-art (SotA) methods for online handwritten sequence classification.
- An ablation study is performed to examine the impact and performance of the proposed feature set compared to the existing feature sets on multiple evaluation metrics for online handwritten sequence classification.
- Evaluation of using context information in combination with ML methods and its impact on online handwriting classification.

Chapter outline

The rest of the chapter is structured as follows. [Section 4.1](#) introduces the readers of this to the problem of online handwriting classification, challenges and recent advances in the domain. [Section 4.2](#) presents an overview of recent work and state-of-the-art (SotA) methods for the problem of online handwriting classification. In [Section 4.3](#), precise information about the presented feature set and details about ML and DL classifiers are covered. [Section 4.4](#) covers the methodology of collecting the database and presents salient features of data, followed by evaluation protocol and parameters to tune the ML and DL classifiers for optimal performance. Results are furnished and discussed in [Section 4.5](#), which also covers the strengths and weaknesses of the presented feature set along with a comprehensive ablation study on the impact of features included in the proposed feature set. A detailed comparison of the proposed feature set with the existing feature sets is also part of [Section 4.5](#).

The author of this thesis has published the content, figures, and tables included in this chapter in the following publications. The author of this dissertation has written all the text taken from the mentioned publications and the text in this chapter itself. More details about the publications included in this chapter are as follows:

- Younas J. et al., What am I Writing: Classification of online handwritten sequences. In: Intelligent Environments, 2018 [271]
- Younas J. et al., Sense the pen: Classification of online handwritten sequences (text, mathematical expression, plot/graph). Expert Systems with Applications, 172:114588, 2021 [272]

4.1 MOTIVATION

Sensor-based systems result in enhanced interactions while performing routine tasks

We live in a digital age where different sensors or sensor-based devices surround us. Display and sensor technologies, when combined, provide new ways for users to interact with their surroundings to perform routine activities. For handwriting, in particular, the increasing influence of digital devices (e.g., digital ink and mobile devices like cellphones, iPad, tablets, etc.) has attracted the research community's attention [61, 83, 104, 127, 240, 266]. Use of these devices paved the way for writing-behaviour analysis like handwriting classification (classifying the handwritten samples/sequences into text, graphics, or formulas, etc.) [15, 53, 107, 271], handwriting recognition (recognizing what is written - Optical Character Recognition (OCR)) [146, 147, 156], and writer identification (identifying the writer of handwritten text/signature) [122, 148, 205, 213].

Handwriting analysis is broadly categorized into two types

Handwriting is broadly categorized as (a) offline and (b) online. Offline handwriting is usually produced on paper with a normal ink pen. In offline handwriting, only spatial information is stored, and temporal information is not available. So, offline handwriting processing systems are provided with handwriting information in the form of images. Writing on digital displays or writing with digital pens on special/ordinary paper is termed online handwriting. In online handwriting, pen movements are recorded as a continuous stream of points; therefore, temporal information about handwriting is also available with spatial information. Online handwriting is processed in the form of time-series sequences. Note that every person has a specific writing style, which may vary when they write different modalities, like plain text, mathematical expression, plots/graphs, etc. This makes handwriting classification, whether online or offline, quite challenging and interesting.

Historical perspective of handwriting classification

Handwriting classification is important from a historical perspective as well. During the first half of the 20th century, handwriting classification systems were developed as a biometric tool. For example, Milwaukee police [12] and Nottingham police [167] adopted handwriting classification systems to assist in the criminal investigation process and to keep records of citizens. Furthermore, German police used handwriting as a biometric feature during the second world war [167]. Till the 90's, mostly handwritten templates were used for writing classification and writer identification based on a predefined feature set [12, 226]. Later, traditional handwriting devices are coupled with sensors to broaden the research scope, particularly for online handwriting, and commercial systems for online handwriting analysis (classification and recognition) were reported [53, 206, 271].

Today, handwriting classification systems find applications in education, banking, postal services, and forensic science. For example, online handwriting classification systems serve as a basis to analyse the performance of students while attempting solutions to different tasks (writing mathematical expressions, plain text, or plotting graphs) [61, 104, 127, 240, 266]. Similarly, in the banking sector and forensic science, automatic handwriting classification systems can facilitate segmentation/extraction of different modalities (like handwritten text or mathematical expressions) from documents which could further facilitate experts in performing document verification [58, 152, 155, 217]. Moreover, handwriting classification can serve as an important step to improve the performance of handwriting recognition systems by classifying data first and then passing it to handwriting recognition systems.

Applications of handwriting classification are widespread

This work focuses on classifying the online handwritten sequences to automatically look into the type of writing before any subsequent processing to address the limitations of existing systems. First, we collect a novel database for online handwriting classification and recognition recorded. The collected data is processed into sequences and annotated with labels as plain text, mathematical expressions, and plots/graphs. Each sequence is transformed into the 49-dimensional vector using a novel feature set proposed in this work for online handwritten sequence classification problems in particular and can also be adapted for other online handwriting processing methods. We also demonstrate its significance in developing a novel approach to classify handwritten sequences in plain text, mathematical expressions, and plots/graphs. These feature vectors train the ML and DL classifiers to classify the input sequences into text, math, and plot/graph class. Classification of handwriting type enables the teachers with deeper insights into students' progress during the exposition of complex tasks and concepts, which further helps them provide personalised feedback based on individual weaknesses and strengths.

Looking into the type of handwriting

We note that a few datasets, originally collected for online handwriting recognition, have been used for handwriting classification [108, 146]. These datasets, however, are either better suited to recognition, e.g., *IM-OnDB* - collected as handwritten notes of English text on a whiteboard, or for mode detection, i.e., identifying handwritten strokes at every point in document creation (e.g., *IM-OnDo*). The dataset presented in this chapter contains handwritten sequences that could be readily used for online classification of handwritten modalities into text, mathematical expressions, and/or plots/graphs.

Significance of collected dataset

4.2 RELATED WORK

Features play a vital role in handwriting classification

This section presents a historical perspective and a detailed overview of the recent developments in online handwriting classification, including features, datasets, and methodologies with their strengths and weaknesses. Smith et al. [226] presented a feature set to classify handwriting for the first time to the best of the author's knowledge in 1954. The presented feature set is a combination of developed and unconscious behavior: speed, size, slant, and spacing are the four features to incorporate the developed factors and handwriting, whereas pressure and form (defined as idiosyncrasies of handwriting) are the two features from unconscious behaviour. In 1959, Livingston et al. [12] presented a handwriting and pen-printing classification system to identify law violators. They highlighted 12 factors of printed style lettering done with pencils, pens or other writing instruments, which can be used to classify an individual's handwriting.

Synthetic parameters for handwriting classification

Bouletreau et al. [26] presented a new family of synthetic parameters to classify handwriting into different families. The proposed synthetic parameters are based on the fractal analysis of writing behaviours. The proposed fractal behaviours of an individual's handwriting are appearance, implication, rapidity, juxtaposition, and direction. The proposed approach serves as a preliminary step in handwriting recognition.

A flexible framework for online handwriting segmentation

Delaye et al. [53] presented a flexible framework to segment online handwritten documents, i.e., text lines, non-text objects, and mathematical symbols. Their proposed approach is based on single-linkage agglomerate clustering built upon a feature set for pairwise distance definition. They also present a combination of features to improve online handwritten document segmentation.

Writer classification using Kohonen network

Schomaker et al. [208] present a ML-based approach to identify and classify the writers based on their writing style. The proposed approach records the pen-tip displacement data, which then, based on velocity, is segmented into strokes. A 1-d feature vector represents each stroke. These feature vectors are used to train the Kohonen network to classify writers. They used discriminant analysis and clustering techniques to classify writing styles into different families.

Handwriting classification based on discrimination

Bahlmann et al. [15] present a new Gaussian Dynamic Time Warping (GDTW) kernel by combining State Vector Machines (SVMs) and Dynamic Time Warping (DTW) for the classification of online handwriting. The proposed approach creates class boundaries based on discrimination rather than relying on modeling assumptions. They evaluated their approach on the UNIPEN handwriting dataset [91]. Ahmad et al. [2] presented the development of a hybrid model for

online handwriting classification. Their system can classify digits, lower-case, and upper-case letters using State Vector Machines (SVMs) with Radial Basis Function (RBF) kernel. They reported results on UNIPEN and IRONOFF [244] handwriting datasets.

Delaye et al. [52] presented an automatic handwritten document segmentation method to extract graphical objects from online handwritten documents. Their method is based on hierarchical Conditional Random Fields (CRFs). The proposed methodology is evaluated using the IAM-onDo dataset [108]. Delaye et al. [56] presented a text/non-text classification system based on CRFs for online handwritten documents. They first calculated the CRFs for text and non-text strokes, then integrated context information to improve the classification results.

*Text/non-text
classification using
CRFs*

Phan et al. [185] presented a Deep Learning (DL) based classifier for classifying online handwritten documents into text and non-text parts. They used Recurrent Neural Networks (RNNs) and Long-Short Term Memory (LSTM) networks to evaluate their system on the Japanese ink documents database Kondate [160] and IAM-OnDo database. Inatani et al. [106] present a comparison of Markov Random Fields (MRFs) and Conditional Random Fields (CRFs) to separate text versus non-text strokes from online handwriting Japanese documents. The proposed approach also evaluates the impact of context information on the Kondate Japanese ink dataset.

*Using RNNs & LSTM
for handwritten
document
classification*

Weber et al. [206] presented a system to classify ink traces into either text or graphics for mode detection. They also presented a set of features for online handwriting classification and recognition tasks. They benchmark their presented feature set using Machine Learning (ML) classifiers on the IAM-OnDo database. Indermuele et al. [107] presented a Bidirectional Long-Short Term Memory (BLSTM) based neural network approach for text and non-text stroke detection. Individual strokes are transformed into feature vectors to train and test the presented model. They also reported results on the IAM-OnDo database.

*Classification of ink
traces for mode
detection*

In this section, we will also introduce the readers to the publicly available datasets for online handwriting analysis, including handwriting recognition, handwriting mode detection, and handwriting classification. IAM-onDo dataset [108] is the widely used and most popular publicly available dataset for mode detection. With the introduction of the IAM-onDo dataset, text and non-text classification tasks got the attention of the research community, as this database contains contents with formal and informal text, diagrams, tables, drawings, and figures. IAM-onDo dataset content is collected from 189 writers in the form of 941 online handwritten documents. Kon-

*Datasets for online
handwriting
analysis*

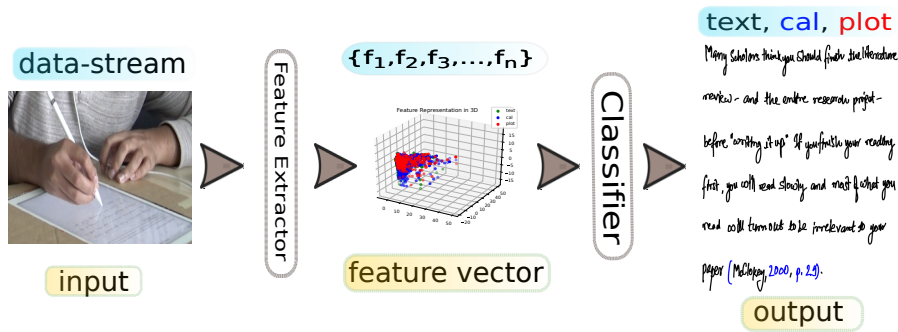


Figure 4.1: System overview

date is another publicly available dataset of online handwritten patterns of mixed objects such as text, figures, tables, maps, and diagrams in the Japanese language. The dataset contains the handwriting traces with ground-truth tags from 100 writers.

Multiple datasets for online handwriting analysis using sensor pen

Ott et al. [175] presented a set of new datasets for online handwriting processing, such as handwriting recognition (character recognition), mathematical expression recognition (digit recognition), and sequence classification (word recognition). These datasets preserve the natural writing behaviour of the writers as the data are collected using a sensor pen with normal/traditional paper to write on. The proposed datasets can be used for various tasks in the online handwriting domain.

4.3 METHODOLOGY

Overview of proposed methodology

This section presents a detailed overview of data collection, which later contributed to the compilation of the dataset. Feature extractors transform input sequences into feature vectors, which are then used to train the machine-learning classifiers. An overview of the methodology followed in the current work is shown in Figure 4.1.

4.3.1 Data Collection and Pre-processing

Data is collected by using Apple pencil & iPad

The data collection process starts with an *iOS*-based application for iPads, which provides the functions for creating new documents, storing data, managing, and exporting existing documents to other devices for further processing. Document templates are used to create new documents, which predefine the document's structure. Each new document is assigned a unique ID, thus allowing multiple copies of a single template. Tasks are distributed along the pages depending on the nature of the task (text, mathematical expressions, or plots), allowing users to navigate back and forth. Apple pencil is used to perform these tasks, as shown in Figure 6.4. Writing data is recorded at a rate of 240 points per second. These points contain information about the

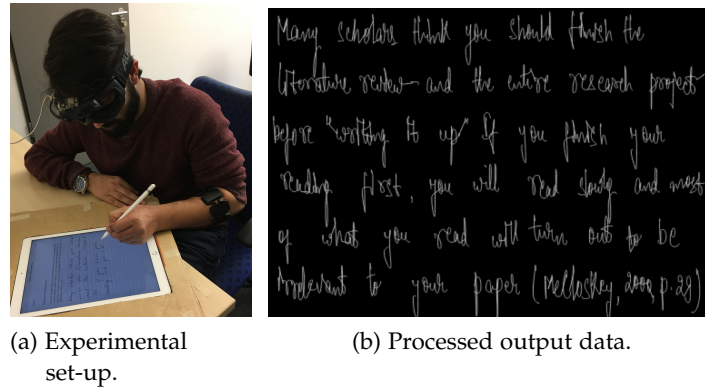


Figure 4.2: Data collection set-up

pencil's location on the touch-screen, the force of the touch, altitude & azimuth angle, and time¹. The writing of a user is rendered in view by linear interpolation between successive points. There is an option to use the eraser, which enables the user to undo writing mistakes.

We also discuss steps to refine the collected data into a proper dataset for further use by the research community without reinventing the wheel. When a person writes, data are stored as a continuous stream of points in a .csv file. In addition to handwritten sequences, data contain document identifiers and page identifiers. So, preprocessing was done to cleanse the data, segment it into strokes and sequences, and remove the structural information of the document and pages. Every segmented sequence represents a single word or expression. Stroke is the term used in handwriting processing to refer to the data written and collected during a single pen-down and pen-up action. Successive strokes are gathered to form a sequence, which refers to a piece of meaningful information such as a word, mathematical expression, or plot/graph. The sequences may vary in length and compose of single or multiple strokes. After preprocessing, the data look the same as was written on the iPad, as shown in Figure 4.2b.

Pre-processing transforms raw sensor data into meaningful information

4.3.2 Feature Extractor

Different features have been presented in the literature [16, 147, 148, 176] for handwriting recognition and classification. We present a new feature set, which includes some of the existing features along with a variety of new features to interpret vast handwriting behaviours and writing types into meaningful information for Artificial Intelligence (AI) classifiers. A detailed comparison of existing and newly proposed features is presented in Table 4.1. The proposed feature set

Features aid ML methods to perform better

¹ See <https://developer.apple.com/documentation/uikit/uiview> documentation of the UIView class

is used to transform online handwritten sequences into feature vectors. Sequences are recorded at the rate of 240Hz, where at every point, p_i is recorded with information of time-stamped (x, y) coordinates, angles, and force, defined as $p_i = (x_i, y_i, f_i, \dots, t_i)$.

*Transforming data
into features*

A stroke starts with the pen-down movement of the Apple pencil writing on the iPad and ends with the next pen-up movement. Thus, a stroke is defined by a sequence of points, $s_i = [p_1, p_2, \dots, p_n]$, for a time interval, $t_i = [t_1, t_2, \dots, t_n]$, when the pencil-tip is in contact with iPad, whereas a sequence refers to a meaningful expression. A sequence can be composed of single or multiple strokes, $seq_i = [s_1, s_2, s_3, \dots, s_n]$. Every sequence is considered an independent entity and is transformed into feature vectors, $v_i = [f_{1_i}, f_{2_i}, f_{3_i}, \dots, f_{n_i}] \in \mathbb{R}^f$, which were later used to train ML and DL classifiers. Our presented feature set consists of the following 49 features contributing to achieving state-of-the-art (SotA) performance.

- The sequence length, L_s (1), is the total length of strokes present in a sequence.

$$L_s = \sum_{i=1}^n \text{len}(s_i) \quad (4.1)$$

- Time of a sequence, T_s (2), total time in seconds taken to complete a sequence.

$$T_s = [t_n - t_1]_s \quad (4.2)$$

- Sequence displacement, Δs (3), the shortest possible distance in pixels of pencil movement for a given segmented sequence.

$$\Delta s = \sqrt{\Delta x^2 + \Delta y^2} \quad (4.3)$$

- Sequence distance, D (4), the sum of displacement of consecutive points present in a sequence.

- Sequence height and width, (5), (6), sequence height is defined by the difference of the maximum and minimum value of y -values present in the sequence, $\text{height} = \max[y_i] - \min[y_i]$, while sequence width is the difference of x -values, $\text{width} = \max[x_i] - \min[x_i]$.

- Sequence slope or gradient, m (7), slope or gradient is a measure of steepness and direction of the line.

$$m = \frac{\Delta y}{\Delta x} \quad (4.4)$$

- Speed (8), the rate at which a given sequence is produced.
- Velocity (9), rate of change of the displacement for a given sequence.

- Strokes count (10), number of times the pen made contact with the screen to complete a sequence.
- Average stroke distance (11), average stroke distance is calculated by averaging the total distance with the number of strokes in sequence.
- Maximum and minimum distance of stroke (12), (13), a maximum and minimum distance of strokes, which are present in a given sequence.
- Maximum and minimum stroke length (14), (15), maximum and minimum length of strokes that are present in a given sequence.
- Maximum and minimum stroke time (16), (17), the maximum time taken to produce a stroke in sequence as well as the minimum time for a stroke.
- Mean stroke length (18), average length of strokes present in a sequence.
- Mean stroke time (19), average time taken to produce strokes of a sequence.
- Mean slope (20), average slope of strokes present in a sequence.
- Vicinity aspect (21), the aspect of the trajectory of a given sequence.

$$\frac{\Delta y - \Delta x}{\Delta y + \Delta x} \quad (4.5)$$

- Vicinity curliness (22), the length of given sequence divided by $\max(\Delta x, \Delta y)$.
- Linearity (23), we define the linearity of a sequence by the average squared distance of strokes present in the sequence to the straight line.
- Maximum and a minimum of force (24), (25), maximum and minimum of the pen force used to produce a given sequence.
- Range of force (26), the difference between maximum and minimum force values for a given sequence.
- Mean force (27), average force applied for a given sequence.
- Variance and standard deviation of force (28), (29).
- Variance and standard deviation (30), (31) of the rate of change during segmented sequence Δt .
- x,y-skew (32), (33), skewness is a measure of the amount and direction of departure from horizontal symmetry for a given sequence.

Table 4.1: Overview of the proposed feature set.

Liwicki et al. [146]	Additional features (Phase-I)	Additional features (Phase-II)
Speed	Sequence length	Sequence distance
Vicinity aspect	Sequence time	Sequence height & width
Vicinity curliness	Sequence displacement	Stroke count
Slope	Velocity	Average stroke distance
Linearity	Force range	Max. & Min. stroke distance
	Mean force	Max. & Min. stroke length
	Force variance	Max. & Min. stroke time
	Variance & Std. Δt	Mean stroke length
	x,y skew	Mean stroke time
	x,y kurtosis	Mean slope
	Variance & Std. $\Delta x, \Delta y$	Max. & Min. force
	Variance & Std. of direction angles	Force std.
	Variance & Std. of slope	Variance & Std. x-values
		Variance & Std. y-values

- x,y-kurtosis (34), (35), kurtosis is a measure of height and sharpness of the central peak for a given sequence.
- Variance of $\Delta x, \Delta y$ (36), (37), the rate of change of pixels in both horizontal and vertical direction.
- Variance and standard deviation (38), (39) of x-values.
- Variance and standard deviation (40), (41) of y-values.
- Standard deviation of $\Delta x, \Delta y$ (42), (43).
- Variance of direction angles of a given sequence (44), (45), measure of variance sin and cosine angles between consecutive pixel for a given sequence.
- Standard deviation of direction angles (46), (47).
- Variance and standard deviation of gradient of a given sequence (48), (49).

4.3.3 Classifiers

Multiple bagging and boosting classifiers are used

This section covers a detailed analysis of **ML** and **DL** classifiers used for evaluation. We used different classifiers based on bagging and boosting algorithms. Bagging algorithms are simple ensemble techniques that merge various classification models using voting strategies like average voting, majority voting, etc. Observations are chosen differently for individual models using the bootstrap process, which helps achieve better generalization. Bagging classifiers are elaborated generically in [Figure 4.3](#). On the other hand, boosting algorithms are ensemble techniques that build models using a sequential learning process where observations are chosen based on classification error.

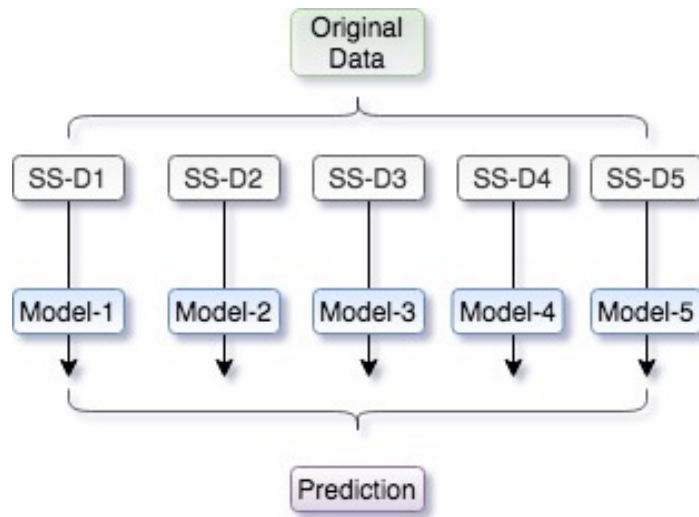


Figure 4.3: Bagging classifier

This results in better performance at every subsequent step. The following subsections explain various classification models used in this ablation study.

4.3.3.1 Random Forest (RF)

Random Forest (RF) is considered a very effective Machine Learning (ML) algorithm for predictions. RF is a meta estimator [30] that follows the bagging technique. It uses Decision Trees (DTs) as a basic building block. Multiple DTs are combined to form a forest named RF. Each DT in the forest uses random sub-samples of training data and is built independently. Distribution is the same for all the trees present in the forest. For classification results, RF uses the majority vote method to produce more diverse and robust results.

Uses Decision Trees (DTs) & random sub-samples to learn

4.3.3.2 Bagging Classifier

A bagging classifier [29] is an ensemble algorithm that fits base classifiers, each on a random subset of data. It aggregates averages of individual predictions using a popularity vote method to estimate the final result. A bagging classifier is a way to reduce the variance of base estimators by introducing randomization, resulting in a significant performance boost.

Majority voting to aggregate individual predictions

4.3.3.3 Extra Tree (ET)

An Extra Tree (ET) classifier [77] is a meta-estimator based on the bagging technique. It builds an ensemble of unpruned decision or regression trees. The main difference of ET classifier with other ensemble methods is that rather than using a random subset of data, it uses complete data to build individual trees. Secondly, it splits the

Bootstrap aggregation for regression & classification

nodes by choosing cut points entirely at random. The final prediction is achieved by aggregating the individual outputs using majority voting in classification problems and averaging them in regression problems.

4.3.3.4 Gradient Boosting Machine (GBM)

Sequential forward learning model

Gradient Boosting Machines (GBMs) [71] is an ensemble machine-learning algorithm using the boosting technique. GBM is a sequential learning model that builds an additive model forward stage-wise. Every model in the subsequent step learns from the model's errors in the previous step. Error is minimized by defining loss functions, i.e., Mean Square Error (MSE). Predictions are updated by using gradient descent and applying a learning rate. Final predictions are made where the error is minimum and predicted values are close to actual ones.

4.3.3.5 Recurrent Neural Network (RNN)

RNNs are ideal for including temporal dynamics in sequence data

Recurrent Neural Networks (RNNs) are known for sequence data handling because of their ability to handle temporal information using self-connected hidden layers. The hidden layer implements input, forget, and output gates to regulate the dependencies. LSTMs [102] and Gated Recurrent Units (GRUs) [89] are commonly used RNNs for handwriting recognition and classification. LSTM units implement memory gates to store the memory at different stages, enabling them to carry the early-stage features to later stages, allowing longer-distance dependencies. Gated Recurrent Units (GRUs) also keep the temporal information without implementing memory gates, making them adaptive to different time scales to store the dependencies. Memory gates solve the problem of vanishing gradients in back-propagation, resulting in improved performance. We implement the GRUs based neural network for this work.

4.4 DATASET & EVALUATION PROTOCOL

4.4.1 Overview

Details of data collection process

20 participants (14 males, 6 females) took part in the study. 18 participants were right-handed, and 17 participants (12 males, 5 females) had first-time writing experience on digital devices, iPads in this case. These participants were students from different disciplines and geographical regions, i.e., Germany, Pakistan, India, Cuba, Venezuela, and the United State of America (USA). The data collection was constraint-free as there were no time restrictions, and all participants were free to write text, mathematical expressions, and/or plot graphs the way they wanted. Constraint-free writing enabled the students to write

with their natural writing behaviour, which brings diversity to data collection, as there are intra-personal writing variations as well as inter-personal writing variations.

During experiments, participants solved different exercises based on the instruction material provided to them. Exercises include text reproduction, creative writing, copying mathematical expressions, solving fundamental calculus problems, and drawing easy graphs. Exercises were kept simple and elaborated so that every participant could understand them. The difficulty level increased as participants progressed with the solutions.

Dataset can be used to differentiate creative writing & copying

4.4.2 Dataset

Considering the classroom environment, note creation is a common activity, and the best way to monitor and track the progress is by online handwriting analysis. These notes mostly consist of plain text, whether structured, i.e., list, caption, and part of a table, or unstructured, i.e., normal text, mathematical expressions that include numerical representation, formulas, axis-markers, and graphs/plots. First of all, strokes are extracted from raw sequences. Every stroke is visualized individually, and then sequences are created out of these strokes manually to minimize the influence of segmentation errors. To generate the ground truth, every sequence is annotated as text, mathematical expression, and plot/graph. Our presented dataset consists of 12,139 labeled sequences, as content breakdown is presented in [Figure 4.4a](#). 65% of sequences belong to text class, 25% of the sequences from mathematical expression class, and remaining 10% from plot/graph class. Class-wise spread of both train and test set using t-Distributed Stochastic Neighbour Embedding (SNE) visualizations are shown in [Figure 4.4b](#) and [Figure 4.4c](#). The dataset, namely onTabWriter, containing stroke information and labeled sequence information, is publicly available for the research community to explore the online handwriting classification venue further.²

Contents of collected dataset

4.4.3 Feedback

After completing writing tasks, participants were requested to fill out a feedback form. Based on feedback, 75% of participants found attempting solutions to calculus problems the most difficult. In comparison, 60% of participants felt more stressed solving mathematical expressions than producing text and drawing graphs. 70% of participants felt more comfortable while copying text and solutions while remaining like creative tasks. We also asked participants about their preferences and provided regular writing notebooks or digital

Experience of using digital devices for handwriting

² [OnTabWriter Dataset](#)

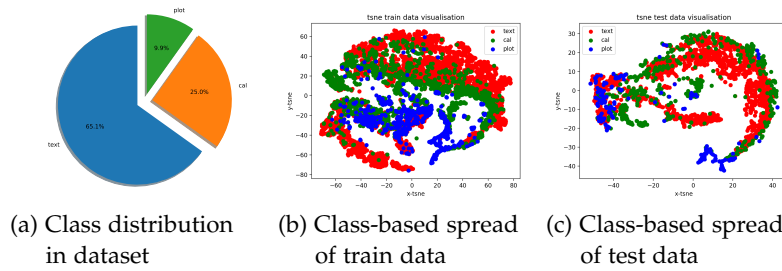


Figure 4.4: Contents present in a dataset with their representation and distribution in train and test set

Table 4.2: Personal preference to write on traditional notebooks versus tablets

Task	Traditional Notebook (%)	Tablets(%)
All	30	70
Text writing	35	65
Mathematical expressions	40	60
Graphs	50	50

devices. As depicted, for every task, most writers preferred writing on tablets, and statistics are presented in Table 4.2. Besides, almost every first-time user reported gradual improvement in writing ease and comfort with more writing practice and showed interest in adopting digital devices for writing in the future.

4.4.4 Evaluation Protocol

Data split for training and testing

In this section, we will discuss the data split used to train and test Machine Learning (ML) and Deep Learning (DL) classifiers, along with the introduction of evaluation metrics to report results. The dataset is analysed in person-dependent settings where train and test data are split in a way that both contain writing data from all participants. We also report results in a person-independent set-up where data is split into train and test sets so that a participant’s data can be used either in the training phase or testing it but cannot in both. Person independent set-up helps to establish the generalization of our presented approach by reporting results on totally unseen writing data. 4:1 split transforms the dataset into train and test sets. 80% of data is used in the training and optimization phase and the rest 20% is used to test the model and report results.

Evaluation metrics

We used the sci-kit library [183] to train and test our Machine Learning (ML) classifier and Pytorch library [182] to implement and evaluate Deep Neural Networks (DNNs). Accuracy defines the cor-

rectness of a model and the most commonly used metric to report the results. When input data is biased or polarised, precision and recall become more relevant metrics. Therefore, we report results as the most relevant metrics in the online handwriting research community, i.e., accuracy, precision, recall, and f1-score. Precision is defined as the ability of a system to distinguish a positive sample from a negative one. In contrast, recall is the competence of a system to classify positive samples, mathematically defined as follows:

$$\text{Precision} = \frac{\text{correctDetections}}{\text{totalDetections}} \quad (4.6)$$

$$\text{Recall} = \frac{\text{correctDetections}}{\text{totalSamples}} \quad (4.7)$$

As we have close numbers for precision and recall for different classifiers, the f1-score represents the results, the harmonic mean of precision and recall. f1-score is mathematically defined as:

$$\text{f1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.8)$$

4.4.5 Optimization Parameters

Random Forest (RF), Extra Tree (ET), bagging classifier, and Gradient Boosting Machine (GBM) classifiers are used in this study to establish the significance of the proposed feature set. Decision Trees (DTs) are used as a base estimator. All parameters are found after empirical evaluation. The number of trees is set to 199. We used the maximum available features to train and test all of our models. Criterion used for RF is 'entropy', for ET 'gini', and for GBM 'MSE'. 'balanced' weight mode is used to address the partiality and bias in the data. We tried different variants of LSTM and GRU networks. Both networks are three layers deep and [64, 128, 256] hidden units for each layer. A learning rate of 0.001 is used with a batch size of 100. All networks are trained for 30 epochs.

Empirical evaluation of network parameters for optimization

4.5 RESULTS AND DISCUSSION

4.5.1 Results

This section presents a comprehensive comparison of the efficacy of feature sets proposed for online handwriting recognition and classification for person-dependent and person-independent setups. Moreover, an ablation study on the impact of individual and combination of features present in the proposed feature set is also part of this section along with the evaluation of using the context information for online handwriting classification. In person-dependent data split and using a bagging classifier, a successful classification rate of 90.0%

Person-dependent results in a nutshell

Table 4.3: Comparison of person dependent performance of ML & DL classifiers on different feature sets

Feature set	Classifiers	Overall result %			
		Accuracy	Precision	Recall	f1-score
Liwiki et al. [146]	Bagging classifier	86.4	86.2	86.4	86.3
	Extra Tree (ET)	85.5	85.2	85.5	
	Gradient Boosting Machine (GBM)	85.2	85.1	85.2	85.3
	Random Forest (RF)	85.5	85.5	85.5	85.5
Phase-I results	Bagging	79.4	79.1	79.4	79.2
	Extra Tree (ET)	80.2	79.9	80.2	80.5
	Gradient Boosting Machine (GBM)	79.8	79.6	79.8	79.7
	Random Forest (RF)	77.9	78.5	77.9	78.2
Final Results(Phase-II)	Bagging classifier	90.0	89.8	90.0	89.9
	Extra Tree (ET)	90.3	90.2	90.3	90.2
	Gradient Boosting Machine (GBM)	90.5	90.4	90.5	90.4
	Random Forest (RF)t	90.1	89.9	90.1	90.0
	Gated Recurrent Unit (GRU)	89.7	89.1	89.7	89.1

is achieved with precision 89.8 and recall of 90.0, as reported in Table 4.3. The bagging classifier produced the highest score for the text class by correctly classifying 95.9% sequences with the precision and recall of 92.0 and 95.9. Mathematical expressions are classified with a precision of 86.2, recall is 79.4, and classification accuracy is 79.4. Numbers reported for plot/graph using bagging classifier are 78.4, 84.6, and 78.4 percent in terms of accuracy, precision, and recall, respectively, as shown in Table 4.5.

*Detailed
classification results
of ML classifiers*

When it comes to RF classifier, overall accuracy, precision, and recall, each calculates to 90.1, 89.9, and 90.1. Class-wise accuracy is 96.3, 78.9, and 79.9 percent, as shown in Table 4.5. The precision score is 92.4, 86.7, and 82.5 with a recall of 96.0, 78.9, and 79.9 for text, mathematical expressions, and plot/graph class, respectively. Extra Tree (ET) classifier shows marginally lesser performance than GBM classifier with the accuracy, precision, and recall score of almost 90.3. Text classification score is the best for ET classifier with the accuracy of 97.2%, while results for mathematical expressions class are the lowest with the accuracy of 77.7%, and graph/plot class is classified with the accuracy of 79.5%. Precision & recall scores for text, mathematical expressions, and graph class are 91.1 & 97.2, 90.4 & 77.7, and 86.9 & 79.5, respectively.

*GBM outperforms its
counterparts in
person-dependent
data split*

GBM produces the overall best results with an accuracy score of 90.5% and outperforms all its counterparts. The precision and recall score for GBM is 90.4 and 90.5. Mathematical expressions and plot/graph class are predicted with the classification accuracy of 81.3 and 79.9 percent, the precision of 86.1 and 84.5, and recall of 81.3 and 79.9. The classification rate of 95.8 is achieved by GBM for text classification with the precision & recall of 93.0 & 95.8.

*Results in
comparison to
existing feature sets*

In a person-dependent set-up, all classifiers perform convincingly well, producing overall classification results with overall accuracy and f1-score in the range of 90 ± 0.5 . Gradient Boosting Machine (GBM)

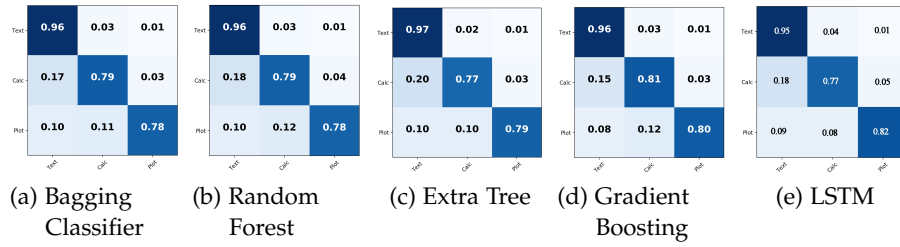


Figure 4.5: Normalized confusion matrices for person dependent results

Table 4.4: Comparison of person-independent performance on different feature sets

Feature set	Classifiers	Overall result %			
		Accuracy	Precision	Recall	f1-score
Liwiki et al. [146]	Bagging classifier	82.0	82.0	82.0	82.0
	Extra Tree (ET)	80.8	80.7	80.8	80.7
	Gradient Boosting Machine (GBM)	82.5	82.3	82.5	82.4
	Random Forest (RF)	81.2	81.4	81.2	81.3
Phase-I results	Bagging	78.1	77.7	78.1	77.9
	Extra Tree (ET)	78.2	78.0	78.2	78.1
	Gradient Boosting Machine (GBM)	77.0	76.4	77.0	76.7
	Random Forest (RF)	78.3	77.8	78.3	78.0
Final Results(Phase-II)	Bagging classifier	81.0	83.1	81.0	82.0
	Extra Tree (ET)	88.9	88.7	88.9	88.8
	Gradient Boosting Machine (GBM)	87.3	87.4	87.3	87.3
	Random Forest (RF)	88.6	88.5	88.6	88.5
	Gated Recurrent Unit (GRU)	87.4	87.7	87.4	87.5

classifier produced overall the state-of-the-art (SotA) results, as reported in Table 4.3. Similarly, Gradient Boosting Machine (GBM) outperforms its counterparts with the highest success rate in classifying individual classes i.e., text, and mathematical expressions with the f1-score of 94.1, and 83.6 as shown in Table 4.5. All of our presented models produce results with the f1-score of 90 ± 0.5 compared to the best results with f1-score of 86.3 achieved by Liwicki et al. [146] feature set using extra tree classifier. Our presented approach also outperforms their method in per-class computed results, as shown in Table 4.5.

In person-independent data split, Extra Tree (ET) classifier produced the best results with an overall accuracy of 88.9% and f1-score of 88.8. The precision score of the ET classifier is 88.7 with a recall of 88.9. The bagging classifier surprisingly doesn't perform well in a person-independent set-up with overall classification accuracy and f1-score of 81.0 and 82.0. The overall precision and recall score for the bagging classifier is 83.1 and 81.0. Accuracy, precision, recall, and f1-score of Random Forest (RF) is 88.6%, 88.5, 88.6, and 88.5, respectively. With GBM, the same is not true as in a person-dependent set-up, as shown in Table 4.4. 87.3% of the sequences were predicted correctly with the precision of 87.4, recall of 87.3 and f1-score of 87.3. Liwicki et al.'s [146] best results are 82.5 and 82.4 for accuracy and f1-score using Gradient Boosting Machine (GBM) classifier.

ET classifier produced SotA results in person independent data split

Table 4.5: Class-wise detailed results of different feature sets on various classifiers in person dependent set up on newly proposed dataset.

Feature set	Classifiers	Text %				Mathematical expressions%				Graph%			
		Accuracy	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score
Liwiki et al. [146]	Bagging classifier	93.7	89.9	93.7	91.8	71.9	81.5	71.9	76.4	74.8	73.0	74.8	73.9
	Extra Tree (ET)	94.4	88.1	94.4	91.1	66.5	82.3	66.5	73.6	74.8	73.3	74.8	74.0
	Gradient Boosting Machine (GBM)	92.0	89.8	91.8	90.8	73.2	78.1	73.2	75.6	72.8	71.3	72.8	72.0
	Random Forest (RF)	92.6	90.1	92.6	91.3	70.9	79.8	70.9	75.1	76.0	68.7	76.0	72.2
Phase-I results	Bagging	88.7	85.6	88.7	87.1	68.2	68.8	68.5	68.2	67.1	74.4	67.1	70.6
	Extra Tree (ET)	90.9	84.6	90.9	87.6	67.6	71.5	67.6	69.5	66.1	77.3	66.1	69.4
	Gradient Boosting Machine (GBM)	88.6	86.1	88.6	87.3	70.0	69.7	70.0	69.8	67.3	74.4	67.3	70.7
	Random Forest (RF)	83.8	88.0	83.8	85.8	70.0	65.9	70.0	67.9	71.3	67.6	71.3	69.4
Final Results(Phase-II)	Bagging classifier	95.9	92.0	95.9	93.9	79.4	86.2	79.4	75.1	78.4	84.6	78.4	72.2
	Extra Tree (ET)	97.2	91.1	97.2	94.1	77.7	90.4	77.7	82.7	79.5	86.9	79.5	81.4
	Gradient Boosting Machine (GBM)	95.8	93.0	95.8	94.1	81.3	86.1	81.3	83.6	79.9	84.5	79.9	83.0
	Random Forest (RF)	96.0	92.4	96.0	94.1	78.9	86.7	78.9	82.6	79.9	82.5	79.9	81.2
	Gated Recurrent Unit (GRU)	95.2	91.7	95.2	93.4	77.0	95.2	77.0	85.1	82.2	93.4	82.2	86.9

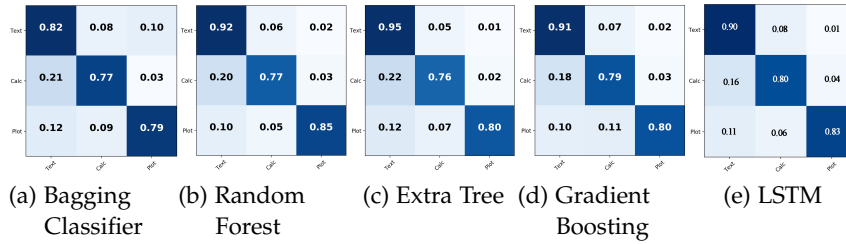


Figure 4.6: Normalized confusion matrices for person independent results

Person independent results using ML classifiers

Text class is 94.4 times correctly classified by ET classifiers with the precision of 90.8 and recall of 94.4. RF classifier predicts text class with numbers 92.9, 92.0, and 92.9 in terms of accuracy, precision, and recall, respectively. There is a significant decrease in the number of text class predictions for the bagging classifier, as reported in Table 4.6.

Text classification accuracy is 82.3%, with precision and recall scores of 90.3 and 82.3, respectively. GBM produced results with the accuracy of 91.2%, and precision & recall is 92.3 & 91.2 for text class. Mathematical expression class is correctly classified at the rate of 76.6, 77.5, 76.4, and 79.0 with a precision of 73.8, 80.7, 81.4, and 75.9 for bagging classifier, RF, ET, and GBM, respectively. The recall rate of mathematical expressions is 76.6, 77.5, 76.4, and 79.0, respectively. Plot/graph class

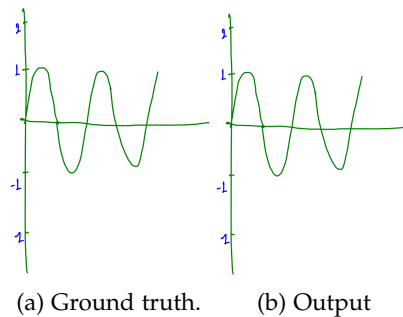


Figure 4.7: A perfect classification result (green color annotates plot, while blue color is labeled as mathematical expressions).

is classified by bagging classifier with the accuracy of 79.4, precision & recall score is 51.6 & 79.4, results for RF classifier are 84.9, 82.7, and 84.9, respectively. Accuracy of GBM and ET classifier to predict plot class is 80.0% and 79.0% with the precision & recall of 81.0 & 80.0 and 88.3 & 79.0.

We also evaluated the presented feature set using Deep Learning (DL) methods. LSTM and GRU models were trained and tested for person-dependent and person-independent setups. The best results are achieved by using 256 hidden units with GRU network. In person-dependent set-up, overall accuracy of 89.7% with the precision, recall, and f1-score of 89.1, 89.7, and 89.1, respectively. Results are 87.4, 87.7, 87.4, and 87.5 for accuracy, precision, recall, and f1-score, respectively, for person-independent set-up. In a person-dependent set-up, GRU network achieved the best results among all the graph/plot class classifiers with an f1-score of 86.9. Similarly, in person-independent set-up GRU delivered the best results for the mathematical expression class with the f1-score of 85.0. Results furnished in Table 4.5 & Table 4.6 demonstrate the competitive results for both ML compared to DL method, which establishes the relevance and importance of the presented feature set for online handwriting classification.

Results of Deep Learning (DL) approaches for on-line handwriting classification

4.5.2 Discussion

We trained the same classifiers on Liwicki et al. [146] proposed feature sets using the proposed dataset for a detailed and fair comparison. Our proposed feature set achieved the best results in both person-dependent and person-independent setups and outperformed its counterparts in all metric scores. The computational load is also minimal, which makes it ideal for real-time use in low-cost systems. Moreover, results achieved by our proposed feature set are also superior for every class, i.e., text, mathematical expressions, and plotting graphs classification by a margin.

Fair comparison is essential for the efficacy of proposed approaches

When there is a clear pattern and structure in the writing of a participant, ideal results, regardless of the sequence class, are produced, as shown in Figure 4.7b, with ground truth file in Figure 4.7a. Figure 4.5 shows that every classifier performed distinctly on classifying text because writing text exhibits a clear pattern and distinguished writing behavior. Few text sequences produced by writing single or few strokes are confused either with mathematical expressions or plot/graph, classification results compared to ground-truth file, as shown in Figure 4.9a & Figure 4.9b.

Clear writing patterns help the network to learn better

Considering segmented sequences composed of individual strokes, it is very hard to tell whether they belong to text or calculation class without looking into context, i.e., considering neighboring strokes and

Table 4.6: Class-wise detailed results of different feature sets in person independent set up on new proposed dataset.

Feature set	Classifiers	Text %				Mathematical expressions%				Graph%			
		Accuracy	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score	Accuracy	Precision	Recall	f1-score
Liwiki et al. [146]	Bagging classifier	88.4	87.2	87.8	88.4	70.1	68.8	70.1	69.4	65.3	76.1	65.3	70.3
	Extra Tree (ET)	88.2	85.7	88.2	86.9	66.6	65.9	66.6	66.2	62.7	81.1	62.7	70.7
	Gradient Boosting Machine (GBM)	89.3	87.5	89.3	88.4	71.1	69.6	71.1	70.3	61.7	76.0	61.7	68.1
	Random Forest (RF)	86.5	87.7	86.5	87.1	70.4	67.3	70.4	68.8	69.6	70.3	69.6	69.9
Phase-I results	Bagging	91.9	80.0	91.9	85.5	57.8	73.6	57.8	64.7	69.5	77.1	69.5	73.1
	Extra Tree (ET)	94.7	78.6	94.7	85.9	55.5	75.8	55.5	64.1	65.5	79.7	65.5	71.9
	Gradient Boosting Machine (GBM)	90.0	80.4	90.0	84.9	57.6	71.4	57.6	63.8	69.1	72.7	69.1	70.9
	Random Forest (RF)	89.8	82.7	89.8	86.1	60.7	72.6	60.7	66.1	72.1	71.8	72.1	71.9
Final Results(Phase-II)	Bagging classifier	82.3	90.3	82.3	86.1	76.6	73.8	76.6	75.2	79.4	51.6	79.4	62.5
	Extra Tree (ET)	94.4	90.8	94.4	92.6	76.4	81.4	76.4	78.8	79.0	88.3	79.0	83.4
	Gradient Boosting Machine (GBM)	91.2	92.3	91.2	91.7	79.0	75.9	79.0	77.4	80.0	81.0	80.0	80.5
	Random Forest (RF)	92.9	92.0	92.9	92.5	77.5	80.7	77.5	79.1	84.9	82.7	84.9	83.8
	Gated Recurrent Unit (GRU)	90.5	92.8	90.5	91.6	80.1	90.5	80.1	85.0	83.1	91.6	83.1	87.1

Table 4.7: Evaluation of using context information in on-line handwriting classification

Classifier	Overall Accuracy(%)	Text (%)	Calculation (%)	Graph (%)
Classifier	80	87	70	73
Classifier+Context information	92	98	86	78

Using contextual information helps to improve the classification process

sequences or by a specific symbol associated with a particular class. Most of the sequences in the calculation are very short and composed of few characters, very similar to text. However, when combined with mathematical symbols, there is a significant difference between text and calculation. This can be achieved by retaining information about neighbouring individuals, which means if preceding and following sequences belong to the same class, the probability of a specific sequence belonging to that class is much higher. Therefore, this work also investigates the importance and effectiveness of context information for online handwriting classification and results show a significant boost for text and calculation class in particular. Overall results are improved by 12%. Results for text class are improved to 98%, and for calculation class increment of 16% is noticed, as shown in Table 4.7.

Proposed feature set delivers SotA results

The proposed feature set makes every classifier capable of classifying free-style and constraint-free online handwritten sequences into text, mathematical expressions, and/or plots/graphs. Furthermore, it can classify minority class sequences among the majority class sequences, i.e., a single mathematical expression correctly classified within a text block, as shown in Figure 4.9b. As discussed earlier, every sequence is considered an individual sequence despite its position or context in the text. Results establish the significance of the presented feature set for online handwritten sequences. Every classifier in an ablation study performs equally well to produce the state-of-the-art (SotA) results in person-dependent and person-independent data split.

Math class often gets confused with Text class

Producing mathematical formulas, their derivations, and solutions

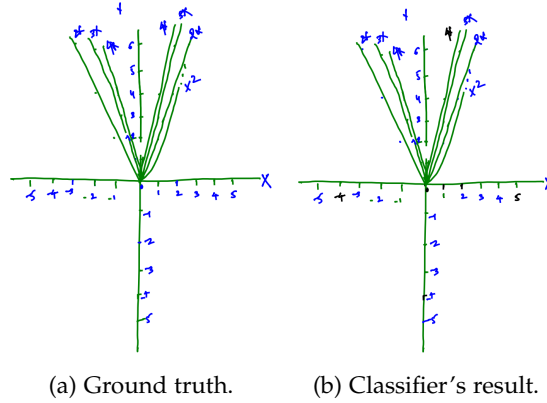


Figure 4.8: An example of plot classification with annotated ground-truth (complex scenario) (green color annotates plot and blue color is for mathematical expressions, while black color highlights text).

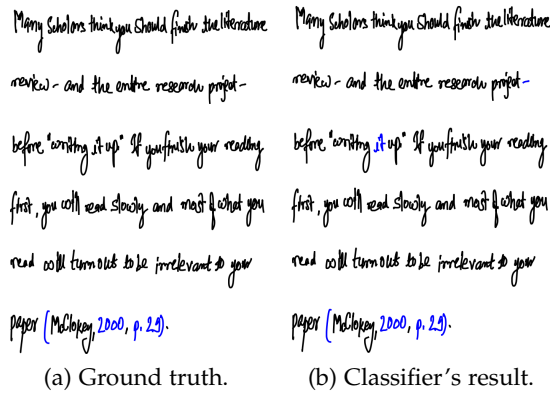


Figure 4.9: An example of text classification along with ground-truth (black color annotates text, while blue color is labeled as mathematical expressions).

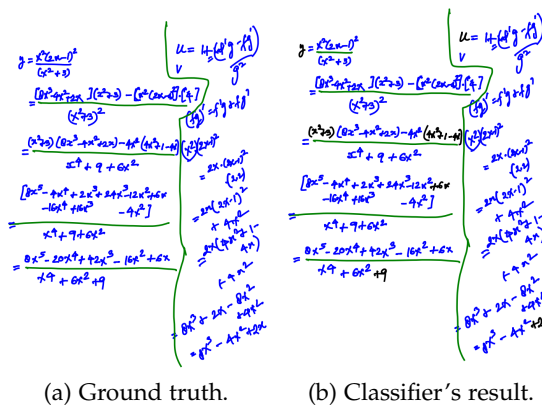


Figure 4.10: Classification of mathematical expressions class in complex scenario (blue color annotates mathematical expressions and green color annotates plot, while the black color is labeled as text).

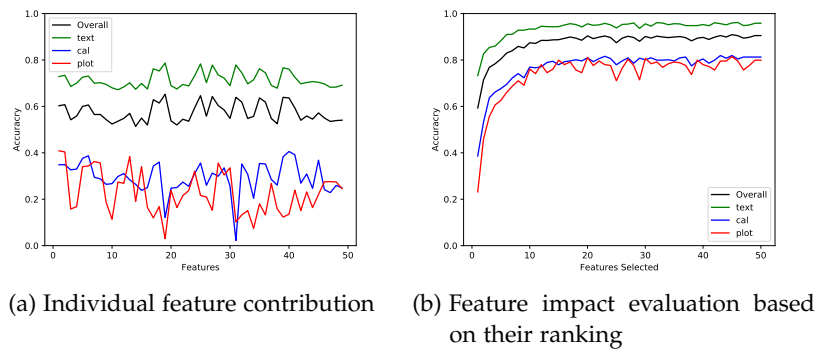


Figure 4.11: Ablation study of our presented feature set.

have a very close resemblance to text. Therefore, the mathematical expressions class gets confused with the text class, as shown in [Figure 4.10a](#) & [Figure 4.10b](#). Mathematical expressions and text writing are very different from plot classes; therefore, there is lesser confusion between these classes, as results also demonstrate. The plot/graph sequence visually shows a clear pattern, but markers and ticks are single-stroke sequences and are treated as independent sequences. Some confusions between text and/or mathematical expressions are shown in [Figure 4.8a](#) & [Figure 4.8b](#).

Ablation study for comparison with existing feature sets

The feature set presented by Liwicki et al. [146] was mainly for handwriting recognition but has been adopted for handwriting classification. Meanwhile, the paraphernalia for online handwriting recording has advanced much over the last decade, i.e., the introduction of sensor pens, and smart pens to write on paper and digital screens. These devices record pen pressure, force, time, and angles along with (x,y) coordinates, which can help develop better handwriting classification systems. The ablation study results on the presented feature set, a combination of existing and new features, institute a better understanding of the problem, as shown in [Table 4.3](#) & [Table 4.4](#). Furthermore, the relevancy of features in the presented feature set is also demonstrated by the Machine Learning (ML) classifier's results compared to the Deep Learning (DL) classifier. As handwriting classification is a very different task than handwriting recognition, a new feature set focused on online handwriting classification will help the research community further push the boundaries in this direction.

Exploring the significance of individual features

We also discuss the significance of features present in our feature set. First, individual features are evaluated for their impact as a whole and for every class, as shown in [Figure 4.11a](#). By evaluating individual features, we get an insight into the performance and importance of individual features along with in-class comparison. Moreover, we also rank the features based on their importance and significance in

Table 4.8: Selected features based on their relevance.

Liwicki et al. [146]	Newly proposed features	
Speed	Sequence displacement	Sequence distance
Vicinity aspect	Velocity	Sequence height
Vicinity curliness	Mean force	Max. & Min. stroke distance
Slope	Variance & std. Δt	Max. & Min. stroke length
	x,y skew	Max. stroke time
	x,y kurtosis	Mean slope
	Variance $\Delta x, \Delta y$	Max. force
	Variance of direction angles	Force std.
	Variance & std. y-values	Variance x-values

the given set, starting from the maximum and dropping the least important feature at every subsequent step, as shown in Figure 4.11b. The most important features in the proposed feature set are maximum stroke distance, the variance of Δy , vicinity curliness, a variance of y-values, etc. The least important features in the rankings are minimum force, sequence time, the average change in directional angles, etc.

Once we have insights about the impact of individual features that can be further utilized to evaluate and find out important features for individual classes, as shown in Figure 4.12. Simple peaks in the individual feature impact graph provide information on overall important features along with important features for text, calculation, and plot class. Ten features found to be most contributing to text classification, which include stroke, sequence, and time-related features and results are visualized in Figure 4.12a. Similarly, 14 features influenced most in the classification of mathematical expressions and results are shown in Figure 4.12b. For plot/graph class classification, 12 features are marked as high impact and results are shown in Figure 4.12c. Common features in all the classes are stroke and sequence distance, length, time, and standard deviation and variance along a horizontal and vertical axis. We combine these individual features to form a superset of important features, and results are presented in Figure 4.12d. We achieved nearly the same performance as the full feature set by using selected features based on their rankings. Details of selected features are provided in Table 4.8, which contains a few features from Liwicki et al. [146] and newly proposed features.

*Important features
for every individual
data class*

We also evaluate existing feature sets on our proposed dataset to establish the utility, superiority, and fair comparison with our presented feature set. Although the feature set presented by Liwiki et al. [146, 148] focused on handwriting recognition, even then, our proposed feature set yields not only overall superior results but also surpasses their results for text detection. It also outperforms results pro-

*Proposed feature set
outperforms for
every individual
metric*

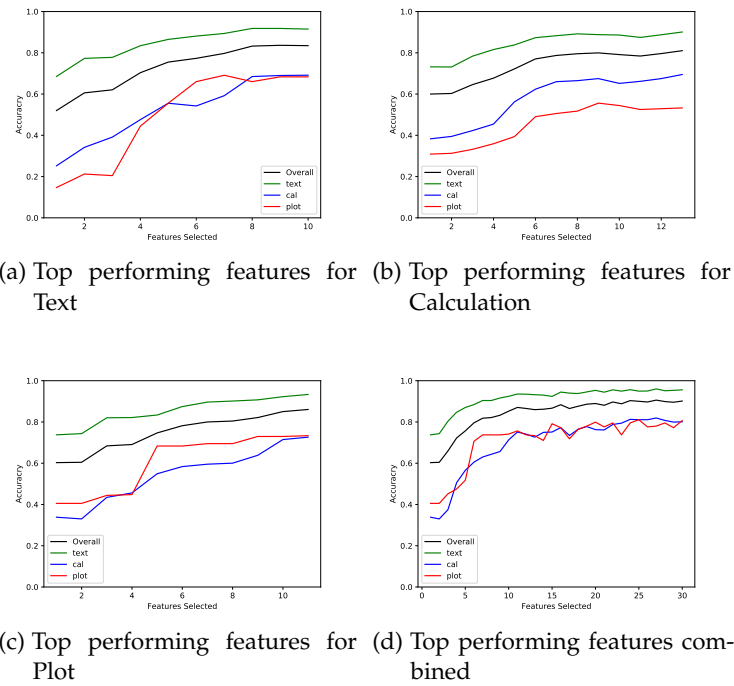


Figure 4.12: Evaluation of top performing features from the proposed feature set.

duced using existing feature sets to classify mathematical expressions and plot/graph class by a margin. We evaluate the existing feature set on the best-performing classifier in this study, yet every classifier produced better results on the newly proposed feature set.

COGNITIVE ABILITIES: A PERFORMANCE ANALYSIS

The foundation of this thesis lies in the two basic concepts for technological inclusion in formal education, i.e., analysis of cognitive abilities and applications to perform the activities to assist the learning process with an improved learning experience. [Chapter 3](#) and [Chapter 4](#) present the methods to evaluate and investigate cognitive activities, i.e., reading and writing, of the learning process on their own. Although reading and writing are individual activities for cognitive ability analysis, there is a dire need for methods to analyse and investigate these activities in correlation to each other. This chapter of the dissertation presents some inceptive research along with preliminary findings to evaluate the behaviour of understanding the problem and concepts while performing cognitive activities during the tasks assigned to them based on prior skills and knowledge about the topic. The process starts with the data collection from learners during cognitive activities using different on-body sensors. On-body sensors track the progress while processing the physical content and working on problem-solving. Combining the information collected from multiple sources and processing them using data science and Artificial Intelligence (AI) tools to analyse and study the individual's approach to attempt the solutions while learning and producing representations and interacting with them. These tools help to improve the overall learning process by providing insights about the individual's cognitive and affective requirements, such as behaviour analysis, feedback estimation, and performance evaluation, a major step in the direction of need-based learning and teaching systems.

Cognitive ability analysis using on-body sensor information

The major contributions of this chapter are as follows:

- Collection of data using multiple on-body sensors while working on introductory Physics course problems, varying difficulty level of the tasks to investigate the behaviour while attempting solutions based on the understanding of the topic and expertise in the domain.
- Presentation of a feature set for evaluating the cognitive process while attempting solutions to the tasks during the cognitive activities and attempting the solution in a ubiquitous environment.
- Insights about the learners' performance based on their cognitive abilities while problem-solving by looking deeper into the

cognitive activities such as reading and writing and the correlation between the two. These findings provide information about the student's understanding of the problem, confidence level while attempting solutions, cognitive load while problem-solving and expertise.

*Structure of this
chapter*

The composition of this chapter is as follows: [Section 5.1](#) introduces its readers to the problem, challenges, motivation, and solution to evaluate the cognitive activities and their correlations for insights about the progress of learners during classroom activities. Recent developments and advances to evaluate the reading behaviour, writing behaviour, performance evaluation and feedback estimation based on these cognitive activities are made part of [Section 5.2](#). [Section 5.3](#) covers the methodology followed in this work with brief details of individual components. [Section 5.4](#) presents the data collection and organization process along with the feature set details for cognitive ability classification. Results are furnished in [Section 5.5](#) with discussions about the methodology followed, its strengths, weaknesses, and prospects.

The author of this thesis has published the content, figures, and tables included in this chapter in the following publications. The author of this dissertation has written all the text taken from the mentioned publications and the text in this chapter itself. The publication list included in this chapter refers as follows:

- Younas J. et al. (2022), Cognitive Ability Classification using On-body Sensors In: Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11–15, 2022, Cambridge, United Kingdom[277]

5.1 MOTIVATION

learning helps to develop cognitive abilities

One of the main objectives of formal education is to instigate the curiosity to learn new things and stimulate the cognitive abilities of the students to develop problem-solving skills. *Cognitive abilities* are defined as the capacity of an individual: "to learn, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience"[188]. Humankind has been learning since its existence through different means, and the key sources to acquire knowledge are observation, reading, and writing activities. The methods to acquire knowledge and learn are also evolving since then. Similarly, methods to monitor and evaluate the learning progress are required to develop and updated as per needs and requirements to facilitate the enhanced learning experience and outcomes.

Technological interventions can be used as a bridge in student-teacher relationship

The learning habits of every individual vary based on his cognitive abilities, interests, and preferences. Even within a person, learning behaviours differ for cognitive activities such as reading, writing, and observation. These cognitive activities help students to develop how to effectively read, write, think, analyse, remember, solve, understand, and enable themselves to make these skills function together to develop intellect and achievement. On the other side of the learning process, teachers/instructors are equally important and play a vital role in steering the cognitive potential in the learning contexts by deciding what to teach and how to teach it. Teachers also play an important role in students' future academic careers and lives by influencing expectations, grading, and contributing toward self-conception and a sense of achievement. It is very hard for a teacher to track and monitor the progress of every student to cater to his/her individual needs and preferences using the same static and traditional teaching approaches. To achieve the goal, technological interventions can act as a bridge in a student-teacher relationship by assisting in performing and tracking the activities and evaluating the behaviours and performance in a better and improved way.

On-body sensors have huge potential to assist the learning process

Smart gadgets and technological aids have significantly assisted cognitive functioning in learning, becoming increasingly important in the educational domain. Major developments have been observed in the recent past to integrate technological developments and smart gadgets in formal education to aid learners and teachers to strengthen the student-teacher interaction bonding [110, 130, 132]. History of reading activity monitoring by tracking eye movements to analyse the reading behaviour tracks back to 19th century [246]. With the advent of eye-tracking tools, the focus of the gaze-tracking community shifted to evaluate the cognitive aspects. The concept of intelligent and interactive documents has been presented recently [114], where the reading documents on digital display are changed dynami-

cally based on the reader's interest. Dynamic handling of documents results in better engagement and focus during the reading activity. Another work classifies the comprehension levels using eye-tracking methods to evaluate their reading behaviour and problem-solving skills [110]. Similarly, other gaze-tracking methods are used to classify the visual attention to understand concepts in Science, Technology, Engineering, and Mathematics (STEM) education [130], attention analysis and performance evaluation for the same activities using different modes of instructions [202], and visual attention to monitoring performance while Physics experimentation [18, 125, 131]. Another study [101] uses video analysis to evaluate the cognitive load for multiple representation learning. Most gaze tracking methods focus on reading activity only after looking deeper into writing activity, a major limitation of these methods. Using implicit sensor information for cognitive ability classification to foster representational competence for teachers and learners is still an open research area to the best of the author's knowledge.

The work presented in this chapter focuses on cognitive ability classification analysing cognitive activities in correlation using an on-body sensor platform. The primary motivation behind our work is using a combination of on-body sensors to record the cognitive activities in the classroom, such as reading, writing, and problem-solving. Participants (Physics students) of different comprehension levels were presented with information material to familiarize themselves with typical introductory Physics concepts and then attempt a solution to the related tasks. Reading and writing data are processed separately and in correlations for insights about the behaviour and progress of learners while attempting the solutions to the tasks, such as short questions, numericals, drawings, and explanations. Initial results provide useful insights into the problem-solving skills and behaviour of participants, such as confidence-score, cognitive load, and expertise. The insights extracted implicitly using sensor data are similar to the prior classification of individuals as experts and novices based on their understanding of the subject. Teachers can use the derived insights to indicate the weaknesses and strengths of an individual student to help them give feedback to the learners based on their cognitive and affective requirements.

Cognitive ability classification using on-body sensor information

5.2 RELATED WORK

Using on-body sensors is a common approach in human activity recognition. These activities cover a wide range of applications in health, fitness, education, automation, education, and various other fields of life. Mobility, low-power consumption, readily available at a low price, and embedded inside daily-use gadgets make these sen-

There are multiple ways of using wearable sensors in learning activities

sors ideal for data collection and activity recognition tasks, particularly for cognitive ability classification in the classroom. Eye-tracking tools are a way to go for the research community to capture the information about the gaze of students while performing the activities such as reading, scanning information, problem-solving, and performing physical tasks [110, 130, 132, 202]. Similarly, some methods focused on visual attention while reading activity to detect stress level and attention estimation [18, 114, 125]. Recently, on-body sensors and smart gadgets have been employed to explore the potential of performing cognitive activities in Augmented Reality (AR) and Virtual Reality (VR) scenarios [119, 132, 234, 260]. There are methods for handwriting recognition and classification [235, 267, 283], which can also be used to track the writing progress but a combination of cognitive activities, i.e., reading and writing, are not explored at best.

*Gaze-tracking helps
in the note-taking
process by
highlighting key
points*

Nguyen et al. [170] present a gaze-based note-taking system while attending online lecture videos without diverting attention from the content. The proposed system combines offline video analysis with online gaze monitoring functionality. Offline video analysis identifies and annotates the important contents in the videos, whereas online gaze monitoring part highlights the key content for note-taking. Moreover, the system automatically controls the video speed or pauses it by analysing the user's attention while taking notes. The result supports the system's utility for online education, enabling users to take notes with lower cognitive load and better concentration levels.

*Gaze analysis to
reorient student's
attention*

D'Mello et al. [47] present an intelligent tutoring system using gaze information. The proposed system aims to promote engagement and dynamically detect the students' interest levels. They used eye-tracking information to monitor the period of disengagement and gaze-reactivity methods to reorient the student's attention. Whenever the tutor found any disengagement in the gaze tutor, he used one of the predefined statements to regain the attention of students. The results show that using the gaze-reactivity technique improves engagement during the lessons, learning gains, and minimal impact on student's motivation.

*Analysing
representation
learning skills to
assess expertise*

Mozaffari et al. [38] presented a study on representational competence in Physics using gaze-tracking. Students were presented with the metainformation using vectorial representation and data representation in tables and diagrams. Students were instructed to solve the Physics problems using metainformation. Students were also categorized as experts, intermediates and novices. The study assessed the effectiveness of different representation learning to analyse the problem-solving skills based on students' expertise.

*Evaluate
comprehension levels
by analysing reading
behaviour*

Ishimaru et al. [111] present a study using eye-tracking glasses to

foster the concept of an intelligent book. They present the meta information and related task to observe the behaviours of the students while reading activity and then attempt a solution to the given tasks. The findings of the study confirm that the reading behaviour of the students is directly related to their comprehension levels, as students with higher comprehension do not look for hints for solutions during problem-solving. They also categorized the learner into three classes, i.e. novice, intermediate, and expert, based on reading time and attempting solution time in combination with scores of the task.

Dinehart et al. [59] presented important research on handwriting analysis in early childhood education. They explore the role of handwriting in the early educational development of young children and improvements in teaching practices to improve 'readiness' handwriting skills.

Importance of writing activity in childhood learning

Another method uses document summarization techniques to assist reading activity [242]. The authors proposed an intelligent reading assistant system that can help to improve the cognitive capability of students while reading digital documents. Klein et al. [125] used gaze-tracking methods to explore visual attention behaviour while taking the test of understanding kinematics graphs in school students. The study further reveals that deeper analysis of gaze data highlights the discrimination between correct and incorrect answers solutions. Similar findings in a predict-observe-explain setting are reported in [131]; a low confidence rating requires more time to complete the task than confident students.

Intelligent systems to assist cognitive abilities

5.3 SYSTEM OVERVIEW

Research studies have established that aiding human activities with technological interventions improves performance [64]. Schneider et al. [207] presented a review and use of sensors and their application in learning. Our presented system is centered around the students aided with on-body sensors, i.e., eye tracker, sensor pen, and meta information, to track their performance while performing cognitive activities. On-body sensor information enables the teachers to look deeper into the behaviours of the students and then assist the teachers with useful insights to interact and address individual requirements effectively to foster the concept of adaptive and need-based learning. The complete system with its components is shown in [Figure 5.1](#). Details of individual components are explained in the following subsections:

On-body sensor information helps to look into problem-solving skills

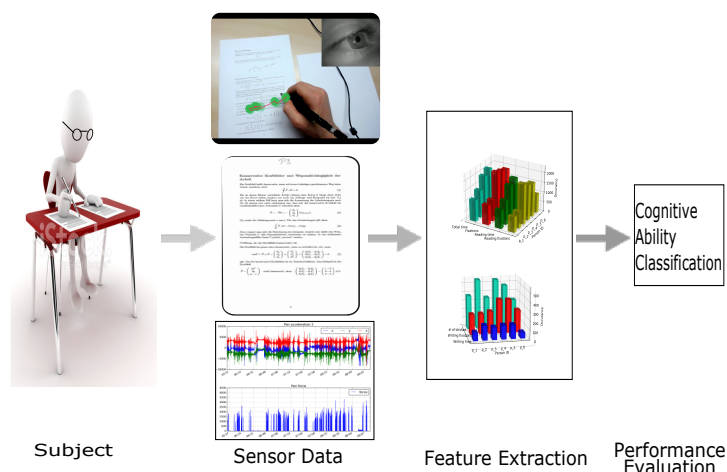


Figure 5.1: System overview: Using on-body sensors during cognitive activities for performance evaluation.

5.3.1 Meta information

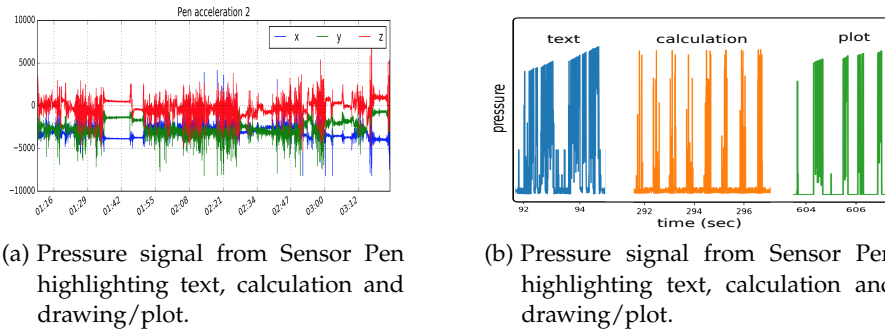
Refers to printed document used in collecting data

Multiple representations help people in effective learning of complex ideas [6]. Meta information consisted of instructional material in the document presented to participants. Instructional material defines the structure of documents and includes reading material and a set of exercises to be worked out by participants for evaluation. Reading material introduces the participants to the topic with background knowledge such as descriptive text, mathematical equations, and figures/graphs. The aim of presenting meta-information is to assist participants to understand the problem by providing relevant and useful information and evaluating their cognitive abilities while attempting solutions.

5.3.2 Digital pen

Digital pens are an ideal tool for collecting natural handwriting data

There is a strong relationship between cognitive load and handwriting production [13, 14]. Yu et al. [278] present a study on the analysis of writing features to evaluate cognitive load. Sensor pens and graphics tablets are standard means to capture handwriting progress. Stabilo digi-pen is a commercially available tool for the research community to capture handwriting progress using multiple embedded sensors without disturbing cognitive process. Stabilo digi-pen is a sensor pen equipped with an Inertial Measurement Unit (IMU) to record the writing progress by capturing the pen's orientation, acceleration, gyroscope, and compass information. The raw acceleration signal recorded is shown in figure 5.2a. IMU capture data tracks the pen movement, a vital step to reconstruct the handwriting. It is also mounted with an internal pressure sensor to record pressure data while handwriting. Figure 5.2b shows the significance of pres-



(a) Pressure signal from Sensor Pen highlighting text, calculation and drawing/plot.

(b) Pressure signal from Sensor Pen highlighting text, calculation and drawing/plot.

Figure 5.2: Pen sensors data

sure signal while producing normal text, equation or drawing figure. Pressure signals can also be used to track pen-up and pen-down movement. IMU signals, in combination with pressure signals, can be used to reconstruct handwriting. Its major advantage over other digital handwriting apparatuses is that it can be used to write on normal papers, allowing users to produce natural handwriting snippets. Pen data is updated at every 5ms. Motion and pressure data of the pen is transferred using a micro-USB port at the rear end of the pen in real-time.

5.3.3 Eye-tracker

Reading behaviour is a very important aspect of cognitive ability classification to look into details of reading behaviours of experts and novices [141, 196]. Capturing eye movements to understand behaviour while performing cognitive activities is promising, as the eyes play a vital role during these activities. In formal education, gaze analysis can tell a lot about the student's approach and behaviour while performing tasks such as reading, writing, and problem-solving [38, 111, 125]. Eye-tracker are commonly used devices to capture and track the eye-movement. Two types of eye-tracking tools are commercially available and readily used by the research community for gaze analysis: Remote eye-trackers and mobile eye-trackers. Mobile eye trackers are wearable devices used as headsets or glasses, whereas eye trackers fixed on/inside a remote display or a screen are called remote/ stationary eye trackers. Nowadays, modern Virtual Reality (VR) and Augmented Reality (AR) tools also come with eye-tracking functionality without requiring any specific or additional hardware and are termed software-based eye-tracking tools. In this study, Pupil [120] monocular mobile eye-tracker is used, as shown in figure 5.3a. It is a head-mounted, lightweight and plug and play USB device. The pupil eye-tracker embeds a pupil camera with a real-world scene camera. The pupil camera is an infrared device that captures eye movement, and the scene camera records the point of in-

Eye-tracking tools help to monitor gaze movement in correlation to real-world data

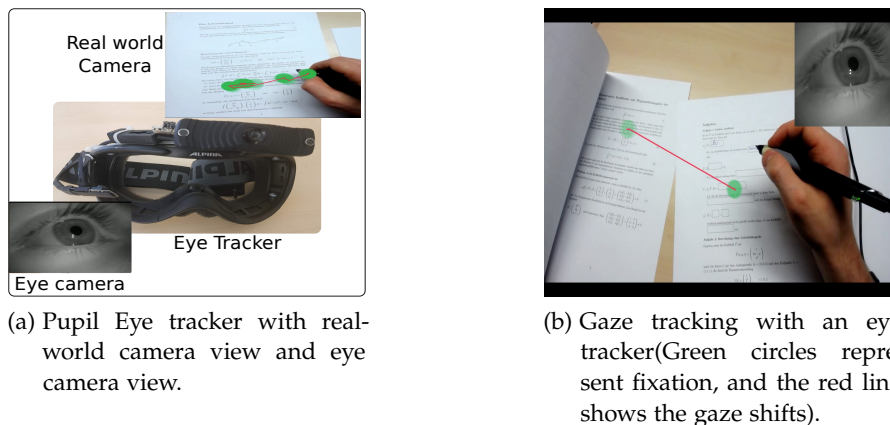


Figure 5.3: Data collection using eye trackers

terest of participants. Pupil eye tracker can detect pupils, track pupil movement, real-time gaze mapping, and perform many other functions. Gaze mapping of the pupil eye tracker used in our study is shown in [Figure 5.3b](#).

5.4 DATA COLLECTION AND ORGANIZATION

5.4.1 Data Collection

Dynamics of data collection

Six participants (4 males and 2 females) volunteered for the study. Out of six participants, 2 were students and 4 were researchers from the Physics department. Before attempting solutions, based on their understanding and self-assessment of the topic, three participants were categorized as novices and rest three as experts.

Data is collected while solving complex representation tasks

Participants were presented with instructional material containing reading material and related tasks for problem-solving. The reading material provided information related to gravitational field theory to familiarize the participants with the topic. Exercises were designed to test factual knowledge and knowledge transfer skills. Exercises were kept simple with the increased difficulty at every subsequent step to challenge the advanced problem-solving skills in the field. Solutions to the exercises required hand-written explanatory answers, mathematical calculations, and complex representations such as drawing figures and plots.

5.4.2 Data Organization

Processing raw sensor data into useful information is a critical step

Once data is collected, the first step toward data organization is synchronizing the data stream from multiple on-body sensors. Hand-writing progress is captured by using [IMU](#) and pressure sensor in

Stabilo digi-pen. Eye movements are recorded with the aid of a head-mounted Pupil eye-tracker. Once data are synchronized, data is pre-processed and segmented into multiple parts for deeper analysis and insights. For handwriting data term "sketched segment" is used to represent a sequence, as producing data includes hand-written text, mathematical calculations, and drawings of figures and/or plots. Handwriting data is processed and segmented using pressure sensor information.

5.4.3 Feature Extraction

A common approach followed to analyse the on-body raw sensor data is processing it into meaningful representations. In this work, we present a novel set of 12 features to analyse the behaviours for cognitive ability classification. Our proposed workflow consists of multiple on-body sensors to capture the progress while performing cognitive activities in the learning process. Pen's sensor data are used to extract the features providing insights about handwriting progress and can be used to analyse the writing activity. Eye-trackers data are used to extract the eye-gaze information while performing both the Reading and writing activity. Some other features are extracted with the combination of both sensors' data. These features enable and assist the teachers by providing insights about the progress of individual activities and the whole cognitive process involved during problem-solving. Fixations and saccades are two commonly used metrics to interpret and process eye-tracking data. A fixation is defined as the period for which gaze is engaged to a specific object or point of interest. The normal period for fixation is between 60 to 1000ms. Following are the details to describe individual features included in the proposed feature set:

Features help to understand the key points of data

- Total fixations, the total number of fixations that occurred for a participant during the whole time of the study, whether he/she is reading, writing or thinking.
- Reading fixations, number of fixations that occurred while reading or thinking process during the whole time of the study.
- Writing fixations, a number of fixations happened only during the writing activity while attempting solutions to the problems included in the study.
- Average consecutive read fixations, the average fixation count while reading or thinking before switching to writing.
- Average consecutive write fixations, the average number of consecutive fixations when the pen tip is in contact with the paper to produce writing before switching back to reading activity or thought process.

- Total time, time taken in seconds by a participant to complete the given tasks.
- Reading time, time in seconds consumed by a participant for reading, thinking and understanding the problem.
- Writing time, time in seconds consumed by a participant for writing activity.
- Sketched segments, count of written segments produced as text, mathematical calculation, plots and diagrams.
- Shift between read and write, number of times a participant was engaged in cognitive process or consulting instruction material while producing solutions.
- Average time difference between strokes, the average time a participant takes between two consecutive pen usages.
- Writing pressure is a key factor in estimating stress while performing writing activities. Stabilo diig-pen measures the writing pressure by force exerted on the pen-tip while writing.

The feature set presented in this study uses eye tracker information, pen sensor data and a combination of both to represent raw sensor information into a substantial feature vector.

5.5 INITIAL DATA EXPLORATION

Tchalenko et al. [233] present a study on an "eye-hand" strategy to copy and produce drawings for experts and novices. Alamargot et al. [8] presented an "Eye and Pen" method for looking into the dynamics of the writing process based on visual attention. We employ the proposed feature set to evaluate its effectiveness and relevance in cognitive ability classification. Figure 5.4 highlights observable differences in behaviours of experts and novices for cognitive activities.

CONFIDENCE SCORE Self-confidence is an essential factor in learning. Feedback based on self-confidence during cognitive activities helps the learners improve performance [158]. Figure 5.4c shows that experts need lesser time to think while attempting solutions. They also provide precise and compact solutions to problems, indicating higher self-confidence. Similarly, Figure 5.4c provides valuable insights about initial cognition, as novices spent not only more time completing the task but also understanding the task, as compared to experts.

Higher confidence level helps to show better performance & results

COGNITIVE LOAD Hochberg et al. [101] conducted a study to demonstrate that lower cognitive load results in higher learning achievements. Figure 5.4a shows that the novices need more time between pen usage. It means they experienced a higher cognitive load for attempting solutions to exercises. It is also observed that the novices consulted more often to information material while attempting solutions to exercise as shown in Figure 5.4b. Also indicates they needed help recalling the necessary information. During the reading activity, we found out that experts go through the abstract-level details of the topics without going into much detail. When they consult the reading material while attempting solutions, they directly focus on the Point of Interests (PoIs), as shown in Figure 5.4c. On the other hand, novices tried to read and understand every detail present in the material, and while attempting solutions, they had to search through the reading material to find the related information, so they needed more time to process the information with higher cognitive loads.

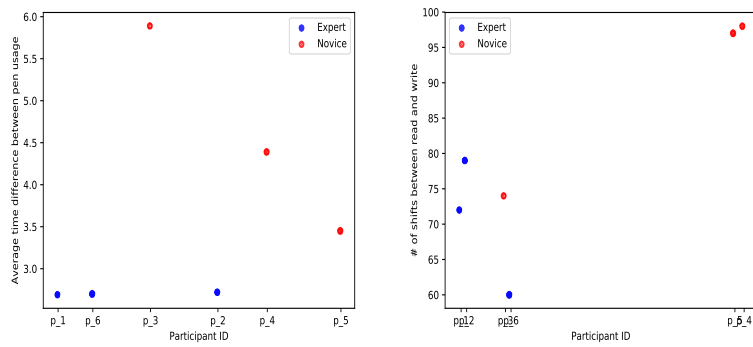
Novices experience higher cognitive load for complex representation learning

EXPERTISE defines one's understanding and knowledge in a particular field. Brueckner et al. [33] explored the influence of expertise on representation learning across multiple domains in a recent study. In this work, we also explore the difference in behaviours of novices and experts based on their expertise. Figure 5.4 shows notable differences in the behaviours of experts and novices in attempting solutions, recalling concepts, required lesser time to understand and produce solutions for complex representation learning. Figure 5.4d shows the visible difference in results of experts and novices, even when the normalised sum of all features in the feature set is used. When the problem is difficult to understand, the participant requires more time to process the information, which results in more reading fixations. We also analyse the time to process the reading material and the number of fixations. As mentioned in [196], there is a clear difference between the reading time of novices and experts. Expert participants require much less time than novices; this study validated the same, as shown in Figure 5.4c.

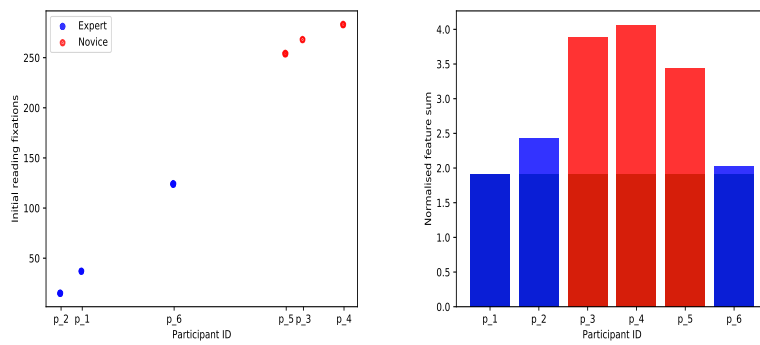
Experts show higher expertise during problem-solving & understanding

We present a feature set to explore the difference between the behaviour of experts and novices when they are exposed to factual and knowledge transfer-based exercises. The proposed features give teachers meaningful insights about differences in behaviours for analysis, understanding, and approach to attempting tasks for experts and novices. Experts exhibit different behaviour right from the beginning, whether analysing the problem, understanding the problem and/or formulating the solutions, taking less time and producing quick solutions skipping intermediate steps and producing abstract solutions. It is also observed that novices require more effort, time and cognitive load from analysing the task to translating their understanding on paper. Novices go through details and process every in-

Proposed features help to differentiate learners based on their problem-solving skills



(a) sketched segments vs. time difference in pen usage. (b) Gaze shifts between reading & writing



(c) total time spent vs. time taken for (d) Normalised sum of all features cognition

Figure 5.4: Cognitive ability comparison for experts and novices.

termediate step on paper. As results presented in this study, whether it is reading time, writing time, reading fixations, writing fixations or time difference between consecutive pen usages, novices' behavior differs from that of experts. These insights are helpful for both learners and teachers to analyse the performance based on an individual's strengths and weaknesses and address them accordingly, which helps improve the overall learning process.

On-body sensors enable the teachers to look deeper into the behaviours of the students, assist them in interacting and addressing the individual's requirements, and foster the concept of need-based learning. Initial data exploration reveals that implicit sensor information can be used as an aid for the teachers to provide needs-based individual feedback. The encouraging results and findings demand further research to develop mental models using on-body sensors for adaptive teaching and learning analytic systems. It is also encouraged to utilize the Artificial Intelligence (AI)-based methods for assistance and deeper insights about the learning progress and activities to assist the teachers in better formulating their instructions and interactions addressing the needs and preferences of each student. These technological interventions and assistance can help to improve the overall learning process by delivering personalized-focused interactions and education.

On-body sensors present a great opportunity for cognitive ability classification

Part III

APPLICATIONS OF WEARABLE SENSORS IN
CLASSROOMS

FAIRWRITE: FINGER AIR-WRITING SYSTEM

This dissertation is built around two main research areas, i.e., methods to analyse cognitive activities in the classroom and applications of smart and wearable systems in performing cognitive tasks, to incorporate technological developments in formal education. This part of the thesis focuses on applications of Mixed Reality (MR) in a combination of Artificial Intelligence (AI) to enhance the learning environment.

Handwriting is a common and vital activity in the classrooms, air-writing enabling systems deemed a default choice to induct technological developments and smart gadgets in classrooms. Air-writing systems can be a handy tool for daily classroom activities. For example, instructors can interact with content to elaborate the complex concepts on digital display without interrupting the momentum of instructions/lectures by simple air-writing motions. Similarly, students can create notes of important points that can be transcribed and stored on their smart devices using air-writing without affecting the cognitive process. The second section of this chapter presents Finger Air Writing System (FAirWrite), a novel air-writing system with applications for VR and AR scenarios such as education, construction sites, offices etc. The proposed system uses a single Inertial Measurement Unit (IMU) to capture the finger air-writing motions without requiring a reference surface and then reconstruct the captured motions as writing trajectories in real time. The proposed system leverages the potential Deep Neural Networks (DNNs) to recognise and classify the written numerics (0-9) and characters (A-Z). The main contributions of the presented work are as follows:

Air-writing systems have applications in various fields

- Finger Air Writing System (FAirWrite) system development to intuitively record the casual motions in the air with a finger. The captured motions track the air-writing trajectory using a

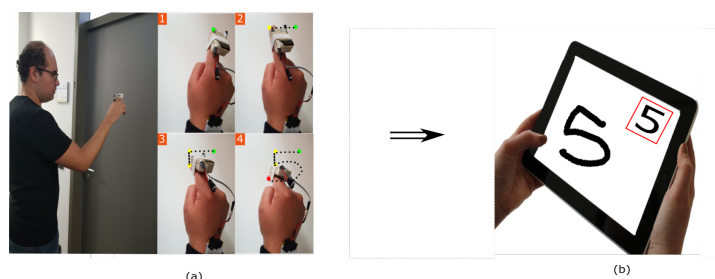


Figure 6.1: FAirWrite system at use. (a). Writing on an imaginary canvas imitating intermediate steps (b). Trajectory reconstruction and recognition in real-time.

single low-cost Inertial Measurement Unit (IMU) worn on the index finger without requiring a reference surface.

- Development of a simple and real-time handwriting trajectory reconstruction algorithm to extract the air-writing motions and project them onto a 2-D digital canvas in a human-readable format.
- Collection of air-writing dataset from 100 participants that can be used to recognize and segment the air-writing motions from various other activities. Moreover, the collected dataset consists of numerals (0-9), lowercase letters (a-z), and upper-case letters (A-Z) that can be used for various air-writing recognition and classification tasks.
- A Graphical User Interface (GUI) to interact with the system and control the functionalities of the system.
- Along with the qualitative evaluation, a systematic evaluation based on Deep Neural Networks (DNNs) to recognise and classify the air-writing mappings onto numerics (0-9) and letters (A-Z), delivering state-of-the-art (SotA) results.

*Structure of this
chapter*

Rest of the [Chapter 6](#) is structured as follows: [Section 6.1](#) introduces the readers to the problem, the need for an air-writing system, challenges in the domain, and motivation for developing FAirWrite system. [Section 6.2](#) covers the literature review and recent developments in trajectory reconstruction of air-writing using vision- and sensor-based systems. [Section 6.3](#) explains the working of the FAirWrite system and methods to reconstruct the trajectory of air-writing in real-time along with the introduction of Deep Learning (DL) classifiers used in this work for classification of air-writing trajectories. [Section 6.4](#) of this chapter describes the data collection process, features of the collected dataset, and evaluation protocol to train and test the proposed classifiers. Results are furnished in [Section 6.5](#) in addition to discussing the strengths and weaknesses of the proposed system and classification models.

The author of this dissertation has published the content, figures, and tables included in this chapter in the following publications. The author of this dissertation has made major or partial contributions to the work published in the following publications. Content, figures, and tables might be reproduced in this chapter. More details about the publications included in this chapter are as follows:

- Younas J. et al., Finger air writing - movement reconstruction with low-cost IMU sensor. In: 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MobiQuitous '20, page 69–75, New York, NY, USA, 2020. [273]
- Younas J. et al., Fairwrite - movement reconstruction and recognition using a low-cost IMU. In: 2022 IEEE International Conference on Pervasive Computing and Communications Workshops (PerComWorkshops), 2022 [274]

6.1 MOTIVATION

IMUs are ideal tool to capture activities

The notion of seamlessly embedding digital interactions in everyday activities is a central concept of ubiquitous computing. Activity and context recognition mapping physical actions, particularly motions and postures, onto digital information is a key component of such seamless interactions. These digital interactions enable end users to use their daily wearable gadgets, i.e., smartphones, smartwatches, smart glasses, and ear-pods, to interact with the surrounding environments with improved user experiences. Inertial Measurement Units (IMUs) are an integral part of such smart and wearable gadgets, with the applications spread widely. Their applications include the area of sports training and bio-mechanics [90, 166], health-care [285], education [236, 248, 271], position estimation [position1], activity recognition [232, 259], gesture recognition [218, 281], Virtual Reality (VR) [257], and Augmented Reality (AR) [169] to almost every field of life.

There are multiple ways to capture air-writing motions

Air-based gesture recognition and trajectory reconstruction is not a new problem, and substantial work has been done in the last two decades, particularly, which included vision-based systems [37, 165, 174], sensor-based systems [9, 51, 134, 179] and depth-based vision systems [7, 37]. Vision-based systems were first used by Oka et al. [174] to track fingertip movement and recognise the geometric shapes trajectories by using an infra-red camera and color sensors. Similarly, Mukherjee et al. [168] used webcam videos to track fingertip detection and recognise air-writing. Vision-based systems limit the mobility, range, and canvas, limiting the concept of air-writing, such as restricted mobility and writing within the camera's range. On the other hand, sensor-based systems commonly use IMUs and leap motion sensors to track and reconstruct the air-writing trajectories. This work focuses on a specific type of motion tracking and recognition: finger-ring-based "air-writing" that mimics a "sticky note" interaction paradigm. Thus, people can casually write a short note by moving their finger "in the air" when passing by a location. Such a note is then virtually attached to the location and is shown as text to an appropriate person when they pass the same location.

Finger air-writing poses multiple challenges

Finger-based air-writing systems are encouraged in recent research work [237], establishing that finger-drawn gestures appear similar to pen-drawn gestures in multiple aspects. The problem with the most sensor-based system either limits mobility, requiring a real surface to write on or poses challenges for general users to adapt them to their natural handwriting patterns, i.e., glove-based handwriting system [9] or holding a mobile phone in their hands for writing [179]. Air-writing systems pose multiple challenges regarding virtual boundary definition, catering to spatio-temporal variability and segmentation

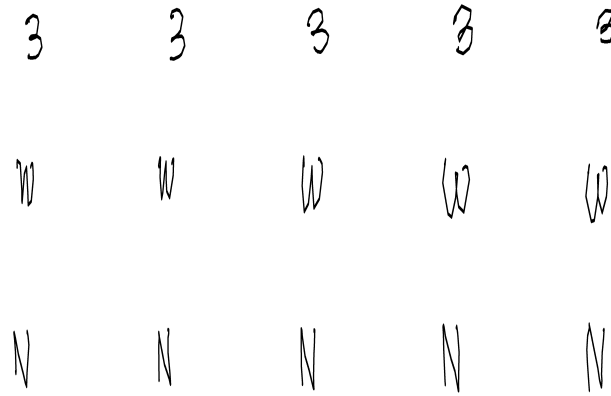


Figure 6.2: Examples of Air-writing highlighting the variations for same user in repeated cycles

ambiguities and making them easier to adapt to and write in the air. Trajectory reconstruction is a very complicated task as every user has their specific writing style, speed, and orientation, resulting in different trajectories for the same text by the same user as shown in [Figure 6.2](#). Sensor internal errors and inaccuracies accumulate over time, which commonly result in drift for trajectory reconstruction, making the task far from trivial. We address these crucial components of the idea: the ability to extract 2-D trajectories from casual finger motion using a single low-cost Inertial Measurement Unit (IMU) without requiring any reference surface and catering to the sensor's internal errors. The extracted trajectories are reconstructed on the remote display to provide the real-time experience/feedback to interact with the system making it very usable and easier to adopt, as in traditional handwriting methods. Finger Air Writing System (FAirWrite) system focuses on the challenges such as mobility, virtual boundary-free screen, system to be easily worn on a finger for air-writing, and real-time applications, along with addressal of several other limitations posed by vision-based and sensor-based systems

6.2 RELATED WORK

Wearable gadgets have been widely used to evaluate hand movements for activity recognition, gesture recognition, and air-writing acquisition. This includes various air-writing systems not requiring any real surface using wearable devices, i.e., smart-watch [264], Myo-[armband \[51\]](#), [IMU-based systems \[9, 179\]](#). This section covers a brief and comprehensive overview of recent developments and systems for online handwriting acquisition and recognition.

Digital pens are the commonly used and commercially available

Sensor pens are commonly used for digital writing

tools to record handwriting progress and make them digitally available for processing. These digital pens use Micro Electro Mechanical Systems (MEMSs) to record and store the pen movements during the writing process on normal or special paper. The MEMS based digital pens [210, 247] mainly use IMUs (accelerometer, gyroscope, and magnetometer), and some use pressure sensors to record the pen up and pen down activities as well. A few are commercially available for research purposes exclusively, i.e., *stabilo-digipen*¹. Although digital pens are very easy to adopt and used for handwriting cannot be used for air-writing.

Vision-based systems have their limitations to capture air-writing

Vision-based systems are theoretically the most appropriate apparatus to perform and record air-writing trajectories, as they do not put any constraint on the user to hinder the natural writing process. Studies established the utility of a vision-based system to track the air-writing by capturing the finger movements using 2-D cameras [37, 165, 168] and depth cameras [7]. Vision-based systems address the limitation of digital pens but put the mobility and range constraint on the end-user when it is out of the camera's range or dealing with an obscure environment.

IMUs are ideal tool to capture air-writing gestures

The answer to the range limitation of the camera lies in the use of IMU-based wearable devices [9, 60, 178] to enable the end-user to air-write without requiring additional equipment, i.e., additional sensor set-ups, displays, screens, etc. Smartwatches [264] is the most commonly used wearable gadget for arm, hand, and finger gesture recognition using a built-in accelerometer and gyroscope. Myo-armband-based solution [51] for handwriting recognition in the air has been presented recently with its scope limited to digits recognition only. A MEMS-based handwriting system for flat surfaces using IMUs is introduced in [60]. A single trajectory reconstruction method for lowercase English alphabets is introduced in [179], further improved with a handwriting recognition module in [178]. Some discussed systems are limited to numerics recognition, few are for the lowercase English alphabet only, and others must be intact with a surface to produce handwriting. This opens up an opportunity for a wearable sensor-based handwriting system, which can be used in real-time and multiple scenarios, i.e., real life, VR and AR.

Using ML to capture & recognise writing gestures

A proof-of-concept is presented in [9] to enable its users to air-write on an imaginary blackboard. The proposed two-staged approach spots and recognizes text from continuous character gestures using IMUs. State Vector Machines (SVMs) are used in the spotting stage to identify the data segments containing the writing activity. They used Hidden Markov Model (HMM) in conjunction with a statistical language model for the recognition stage. The presented method does

¹ *Stabilo Digi-Pen*

not involve trajectory reconstruction but directly uses the Machine Learning (ML) algorithms for spotting and recognition.

Dash et al. [51] present a real-time Myo-armband-based air-writing system to write digits on a virtual screen. They also present a "dif2viz" method to map the air-writing trajectory on a 2-D plane using orientation angle information from built-in IMU in Myo-armband. They used a combination of Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) to recognise the air-written digits. The presented work's scope and applications are limited to being used for digits only.

Myo-armbands can be used to air-write digits

Pan et al. [179] recently presented a smartphone-based air-writing and trajectory reconstruction system. It reconstructs the single-stroke lowercase English letters using a built-in IMU sensor in a smartphone. The proposed system is further improved in [178] along with the additional Machine Learning (ML) module for handwriting recognition using Dynamic Time Warping (DTW) and Convolutional Neural Networks (CNNs). Their approach is limited to single-stroke characters. It requires a precise and complex calibration process every time the system is put in use, a common practice for low-cost IMU sensors to compensate for system-induced errors. The proposed trajectory reconstruction algorithm is complex, and trajectory reconstruction is in multiple stages, and limitation of its working for the single-stroke letter only limits its scope as a real-time system.

Using smartphones as air-writing tool

In summary, every system has its own merits and demerits, limiting their suitability to air-writing because of their size [178], placement [9], complexity [51], and functionality limited to numerics recognition, lowercase English alphabets only, requiring a surface to produce handwriting. This opens up an opportunity for a wearable sensor-based handwriting system, which can be used in real-time and multiple scenarios, i.e., real life, VR and AR. FAirWrite system bridges the gaps. It overcomes the limitation of existing air-writing systems by equipping end users with a system apparent to use and adapt to without affecting the natural writing behaviour.

There is still room to improve air-writing tools

6.3 FAIRWRITE SYSTEM

This section covers the details of a finger-worn air-writing system, built using a single IMU as Finger Air Writing System (FAirWrite), a complete system imitating real-world application is shown in Figure 6.1. The proposed system architecture is elaborated by decomposing it into three parts (i) The hardware part, to develop a finger-worn sensor using a low-cost IMU to record the air-writing motions of a user during the process of producing the writing snippets. (ii) A trajectory reconstruction algorithm to project air-writing on a 2-D plane

Overview of FAirWrite system

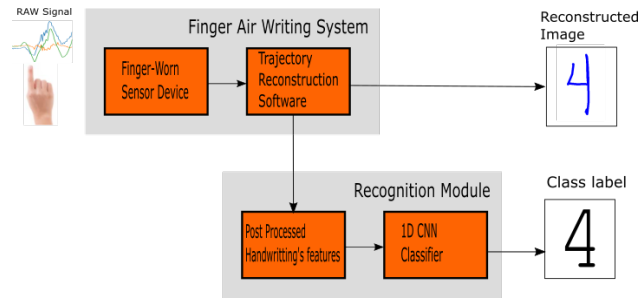


Figure 6.3: FAirWrite system architecture

in real-time, which includes calibration, error compensation, and visualization to provide real-time feedback. (iii) A classifier/recogniser based on Deep Learning (DL) approach to classifying the air-written trajectories as words and digits and then recognizing the individual sequences.

Complete FAirWrite system architecture is shown in Figure 6.3 with details of individual components in the following subsections.

6.3.1 Finger-worn Sensor Design

The finger-worn sensor is an embedded device comprising of a low-cost MEMS IMU ², a control unit ³, and a Bluetooth transmitter powered by 3.7V power supply. The IMU sensor is mounted on an Adafruit board ⁴ that interacts with the sensor's registers to convert the sensor output data to International System of Units (SIs). The IMU sensor is a 9-Degrees of Freedom (DoF) device equipped with a tri-axial 14-bit accelerometer and a tri-axial 16-bit gyroscope. This way, the output signal obtained from the set of Application Programming Interface (API) provides an acceleration vector $[a_x, a_y, a_z]$ in m/s^2 , the angular velocity vector $[w_\phi, w_\theta, w_\psi]$ from the gyroscope in degree/s, and magnetic field strength vector $[m_x, m_y, m_z]$ in μT . For the control unit, the finger-worn sensor uses an Arduino Pro Mini 3.3V ⁵ with an AT-Mega328P ⁶ micro-controller. Arduino unit gets the data vectors from Adafruit, encodes them in American Standard Code for Information Interchange (ASCII) format, and transmits the data via Bluetooth sensor at the sampling rate of 30Hz. For trajectory reconstruction on remote systems keeping the mobility of the user in mind, Bluetooth connectivity ensures uninterrupted data transmission while air-writing along with data transmission to remote display in real-time, keeping in view the real-life applications. The composition of the finger-worn sensor is elaborated in Figure 6.4.

Construction of
Finger-worn sensor

-
- 2 Bosch IMU Sensor
 - 3 Arduino Pro
 - 4 Adafruit board
 - 5 Arduino Pro
 - 6 ATMEGA328P

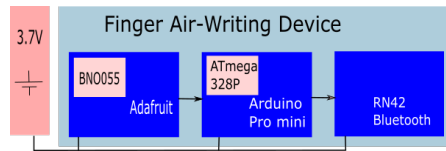


Figure 6.4: Finger-Worn Sensor design

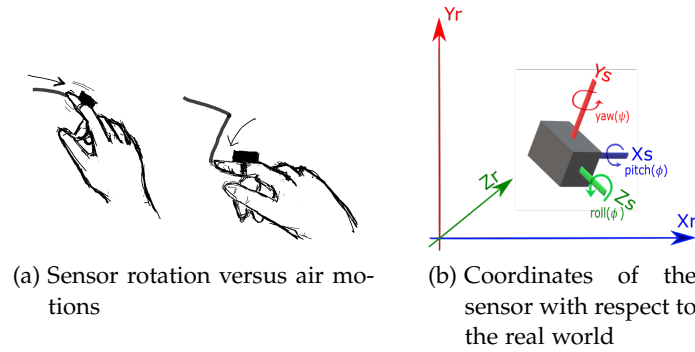


Figure 6.5: Sensor motion with respect to air motions for trajectory reconstruction.

6.3.2 Trajectory Reconstruction

Each data sample consists of a time-series $D_i = \{M_i^1, M_i^2, \dots, M_i^{\tau_i}\}$ consisting of τ_i time steps, each element is $M_i^t = \begin{bmatrix} a_i^t & g_i^t & m_i^t \end{bmatrix}$ where the a_i^t , g_i^t , and m_i^t are the three-axis vector of the accelerometer, gyroscope and magnetometer, respectively. In order to reconstruct the air-writing trajectory in real-time on a remote display for visual feedback, the FAirWrite system uses the angular velocity vector information, i.e., roll, pitch, and yaw. The trajectory reconstruction process is simple and divided into two steps.

Trajectory reconstruction is done in two main parts

- **Attitude Estimation:** To map the data from the sensor's frame of reference to a real world's frame of reference using sensor fusion algorithm [Figure 6.5b](#).
- **Visualization on a Projection Screen:** Transformation of data into real-world coordinate sequences for projection on display to facilitate real-time feedback on the progress of air-writing.

Attitude Estimation

When the user holds the sensor on the finger, the coordinate system of the device is different from the coordinate system in the real world shown in [Figure 6.5a](#). This makes the change of rotation measured by the sensor directly incompatible with the real world's frame of reference. To be able to make a trajectory reconstruction from the displacement of the sensor, it is necessary to obtain the absolute orien-

Conversion from sensors coordinates to real-world coordinates

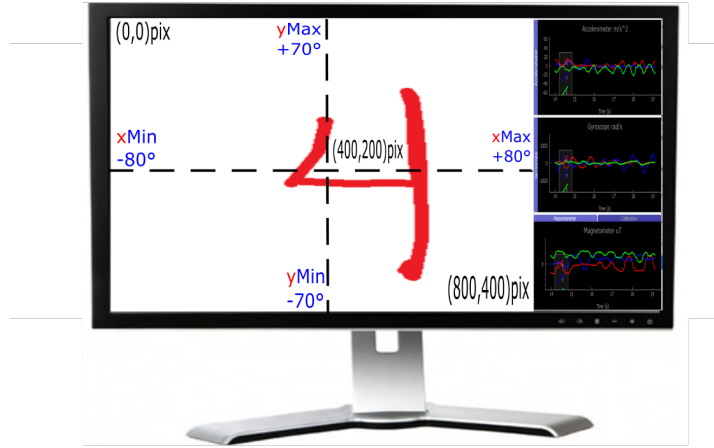


Figure 6.6: The canvas size in pixels and x & y coordinates limits in degrees to plot in screen

tation of the sensor in coordinates of the real world. For this purpose, an accurate, computationally efficient, and custom sensor-fusion algorithm based on Madgwick filter [150] is proposed to transform sensor orientation to real-world coordinates along with removal of noise interference and sensor drift. This filter applies quaternion representation to the orientation angles obtained from angular rate, acceleration, and earth’s magnetic field.

The resulting quaternion $q_{(est)}$ is the estimated orientation in the real-world frame and is represented as:

$$q_{est} = q_w + q_x + q_y + q_z \tag{6.1}$$

where q_x, q_y, q_z is the vector part of q_{est} , while q_w is the scalar part of the vector.

The quaternion can represent any arbitrary orientation in 3-D space, but we are only interested in the 2-D space represented by Euler angles. The Conversion between quaternion to Euler angle is obtained by:

$$\begin{bmatrix} \phi \\ \theta \\ \psi \end{bmatrix} = \begin{bmatrix} \arctan \frac{2(q_w q_x + q_y q_z)}{1 - 2(q_x^2 + q_y^2)} \\ \arcsin(2(q_w q_y - q_z q_x)) \\ \arctan \frac{2(q_w q_z - q_x q_y)}{1 - 2(q_y^2 + q_z^2)} \end{bmatrix} \tag{6.2}$$

where ϕ , θ , and ψ are the roll, pitch, and yaw, respectively. Figure 6.5b shows the relationship of IMU coordinates X_s, Y_s , and Z_s to the real-world coordinates X_r, Y_r , and Z_r .

Visualization to a 2D canvas

In 2nd phase of the trajectory reconstruction algorithm, visualization

*Translating
real-world
coordinates for
display on screen*

of air-writing to a 2-D screen is done based on the rate of change of the finger-worn sensor's orientation is proposed. This rate of change is translated to a series of X and Y coordinates and updates the screen at a speed imperceptible to the human eye, which makes it ideal for real-time visualization. We use the Euler angles obtained in (6.2) to plot the trajectory on screen. Pitch(θ) is used to estimate the X coordinates on screen, and yaw(ψ) for Y coordinates. Whenever the user starts writing, the center of the screen is set as an initial point. After initialization, we calculate gain to relate the finger movement in the air to a 2-D display regardless of size and resolution. The gain is a fixed value for both X and Y.

$$\text{GainX} = \frac{\text{CanvasLENGTH}}{x_{\max} - x_{\min}} \quad (6.3)$$

$$\text{GainY} = \frac{\text{CanvasHEIGHT}}{y_{\max} - y_{\min}} \quad (6.4)$$

where $x_{\max} = 80$, $x_{\min} = -80$, $y_{\max} = 70$, $y_{\min} = -70$ degrees.

In the second step, the difference in the current angle and the previous angle is calculated and multiplied by a gain to determine how many pixels the cursor is moved with reference to the previous data sample to calculate the current position of the cursor on the screen:

$$\text{angle}_x(t) = (\theta(t) - \text{angle}_x(t-1)) * \text{GainX} \quad (6.5)$$

$$\text{angle}_y(t) = (\psi(t) - \text{angle}_y(t-1)) * \text{GainY} \quad (6.6)$$

where $\text{angle}_x(t)$ and $\text{angle}_y(t)$ estimate the difference in angles from the previous value with a gain defined by the min and max values on the screen.

Lastly, the obtention of X and Y coordinates to plot the trajectory reconstruction on the screen.

$$\text{screenPos}_x(t) = \text{angle}_x(t) + \text{screenPos}_x(t-1) \quad (6.7)$$

$$\text{screenPos}_y(t) = \text{angle}_y(t) + \text{screenPos}_y(t-1) \quad (6.8)$$

6.3.3 Classifier

FAirWrite system implements the Omni-scale (OS)-CNN [231] Deep Learning (DL) model to classify the air-written digits and characters. OS-CNN rethinks the time series classification by learning the classifier and kernel size simultaneously by implementing a special design of kernel size configuration to enable the system to cover all possible scales of the receptive fields using a limited number of kernel sizes.

In our proposed approach, we implemented the OS-CNN comprising of three convolutional layers, where the kernel size is prime numbers from 1 to N in the first two layers, to cover all possible receptive fields of any odd numbers in $(0, 2N)$. 3rd convolution layer has the kernel

OS-CNN are SotA for time series classification

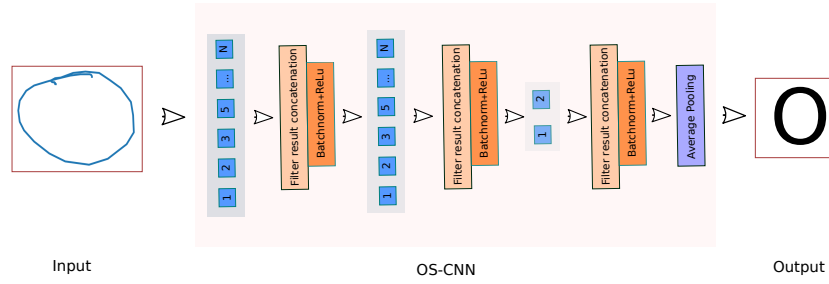


Figure 6.7: Omni-scale (OS)-Convolutional Neural Network (CNN) model architecture.

size of one and two to cover receptive fields of any integer in $(0, 2N)$, followed by a global average pooling layer and a fully connected layer. The OS-CNN model architecture is shown in Figure 6.7. The input sequences are re-sampled with the median of the dataset to address the problem of varying length sequences. A batch size of 200 learning rate of $1e^{-4}$ is used with the gradient descent loss function and the model was trained for 50 epochs.

6.3.4 Finger Air Writing System (FAirWrite) User interface

It is very important to interact with the system in real-time

FAirWrite user interface is a Graphical User Interface (GUI) application developed using PyQT framework [193] to enable the end user to interact with the finger-worn system in real time. Its functionalities include receiving the sensor data, constructing the air-writing trajectories, providing visual feedback, and other control functionalities, as shown in Figure 6.8. FAirWrite user interface utilizes simple image processing algorithms to enable the user to write words and sentences, as shown in Pan 1 of the Figure 6.8. Users can toggle between the writing modes. Pan 2 represents the canvas to display the current air-writing character. Pan 3 shows the raw sensor information, with Pan 4 displaying the sensor's calibration status. Lastly, pan 5 is the control panel of the system's functionalities.

6.4 EXPERIMENT SET-UP

6.4.1 Data Collection

A wide-spread of participants contributed to data collection

A dataset from 100 participants (61 male, 39 female) is collected using FAirWrite system with complete freedom to keep the writing process as natural as possible for every individual. The participants came from different geographical regions, i.e., Europe, America, Asia, and the African continent; most participants belonged to India, Germany, Mexico and other regions. Including participants from different origins brought diversity and variation to the dataset regarding writing habits, e.g., speed, size, and style. Most of the participants were uni-

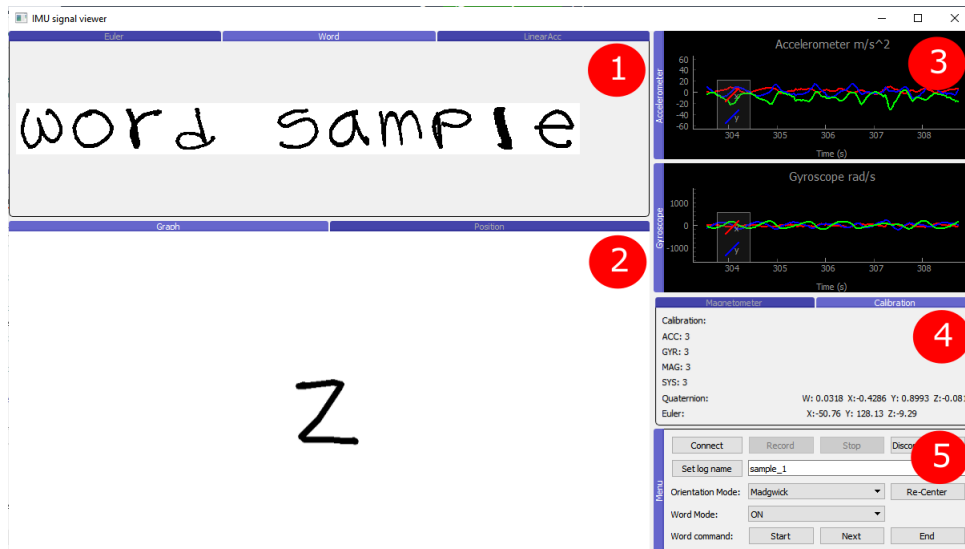


Figure 6.8: FAirWrite user interface to interact with the system.

versity students depict higher education levels. The age of the participating individuals ranged from 20 to 35 years, and most were right-handed. Left-handed participants also preferred wearing the sensor on their right hand to write with and reported no difficulties using it. Initially, it took time for most participants to adapt to the air-writing. However, after a few writing strokes, most participants got comfortable writing with it and reported their ease of using the system.

FAirWrite system provides us 9-Degrees of Freedom (DoF) data from an accelerometer, gyroscope, and magnetometer in m/s^2 , degrees/s, and μT , respectively. The used system provides the continuous data stream at the rate of 30Hz and is transmitted to a remote display via Bluetooth. Before actual data collection, participants were given time to use the system to adapt to the system's working and understand its functionality. Participants were advised to use the FAirWrite system on the index finger. Every participant was asked to write numerals (0-9), lowercase alphabets (a-z), and upper-case alphabets (A-Z). Trajectory reconstruction of collected data symbols is shown in Figure 6.9 The FAirWrite is calibrated for every new data recording by placing it on a flat surface for a few seconds to minimize internal errors.

We also requested participants to contribute to data collection for multiple repeated cycles. Repetitions allowed the system to incorporate the intra-person along with inter-person handwriting variability as shown in Figure 6.2, different writing trajectories for the same characters in different repetitions, writing speed, writing stroke order, and writing style to bring the diversity in the collected dataset. In total, the dataset included about 15000 air-writing sequences. Every sample is visualized manually to remove human errors in the collection process.

How to understand the collected data

Air-writing dataset includes both intra- & inter-person writing variations





































Label	Trajectory	Label	Trajectory	Label	Trajectory	Label	Trajectory
0		1		2		3	
4		5		6		7	
8		9		A		B	
C		D		E		F	
G		H		I		J	
K		L		M		N	
O		P		Q		R	
S		T		U		V	
W		X		Y		Z	

Figure 6.9: Examples of trajectory reconstruction of characters and digits present in FAirWrite dataset.

6.4.2 Evaluation Protocol

Two different evaluations are used to test the performance of FAirWrite system

FAirWrite system is aimed to be a real-time air-writing trajectory reconstruction and recognition system; we follow more than one evaluation protocol, i.e., user quality appreciation (human-based) and model-based evaluation. In user quality appreciation, participants were shown random samples from data collection. They were requested to recognise them and label trajectory reconstruction as good, normal, and bad based on the character’s appearance.

Data split and evaluation metrics

In model-based evaluation, we used Machine Learning (ML) and Deep Learning (DL) algorithms to classify and recognise the air-written characters. At the time of publication of the paper [274], data from 30 participants were available, and the rest of the participant’s data was collected afterward. The collected dataset is divided into train and test sets for model-based evaluation. The upper-case and lowercase alphabet labels are merged to avoid misclassification due to similarities in their shapes. The dataset is split using person independent mechanism so that the participant’s data is either used in the train set or the test set but not in both to establish the system’s generic behaviour and robustness. Train-set consists of the data from 21 participants and the rest of the 9 participant’s data in test-set. We evaluate the system on K Nearest Neighbours (KNNs) in combination with Dynamic Time Warping (DTW), OS-CNNs, and Bidirectional Long-Short Term Memory (BLSTM) models. Overall results are reported regarding classification accuracy, and individual results are shown using the confusion matrix metric.

Table 6.1: User Quality appreciation results

Symbol	Recognized (%)	Good (%)	Normal (%)	Bad (%)
Numerics (0-9)	95.91	79.55	14.81	5.64
Lower-case letters(a-z)	89.93	81.22	13.27	5.51
Upper-case letters(A-Z)	93.04	78.49	14.66	6.85

6.5 RESULTS AND DISCUSSIONS

The main intent of any writing system, whether handwriting or air-writing, is the human beings, so we evaluated the performance of FAirWrite system on human feedback along with model-based evaluation using intelligent models.

6.5.1 User Quality Appreciation

In human-based evaluation, at the end of the activity, the participants were shown random images of air-written digits and characters for recognition. After recognition, we also asked them to label the quality of trajectory reconstruction as good, normal, and poor. The complete results are furnished in Table 6.1. Most of the participants were able to recognize the numerals with the highest overall accuracy of 95.91%, out of which 79.55% as good, 14.81% as normal, and only 5.64% of all the recognised sequences are labeled as bad. Upper-case letters are recognized with the accuracy of 93.04%, where 78.49% of characters are labeled as good, 14.66% as normal and 6.85% are marked with poor labels. Lowercase letters are recognized with the accuracy of 81.22, 13.27, and 5.51 percent as good, normal, and poor, respectively, with an overall recognition rate of 89.93.

*Human-based
evaluation of
trajectory
reconstruction*

6.5.2 Model-based evaluation

In model-based evaluation, we evaluated the FAirWrite system using multiple Machine Learning (ML) and Deep Learning (DL) models, and results are furnished in Table 6.2. The best results are achieved by using the combination of two Machine Learning (ML) techniques, the combination of KNN with DTW results in overall recognition accuracy of 94%, where digits and letters are recognized at the rate of 85.4 and 95.5, respectively. In the case of Deep Learning (DL) models OSCNN achieved the overall accuracy of 88% with class-wise accuracy of 81.8% and 91.3% for numerals and alphabet recognition. We also evaluate the FAirWrite performance on BLSTM models with the overall accuracy of 67%, the digit recognition accuracy of 63.5% and letter recognition accuracy of 68%, which to our surprise, is on the lower side.

*Performance of ML
& DL classifiers for
classification*

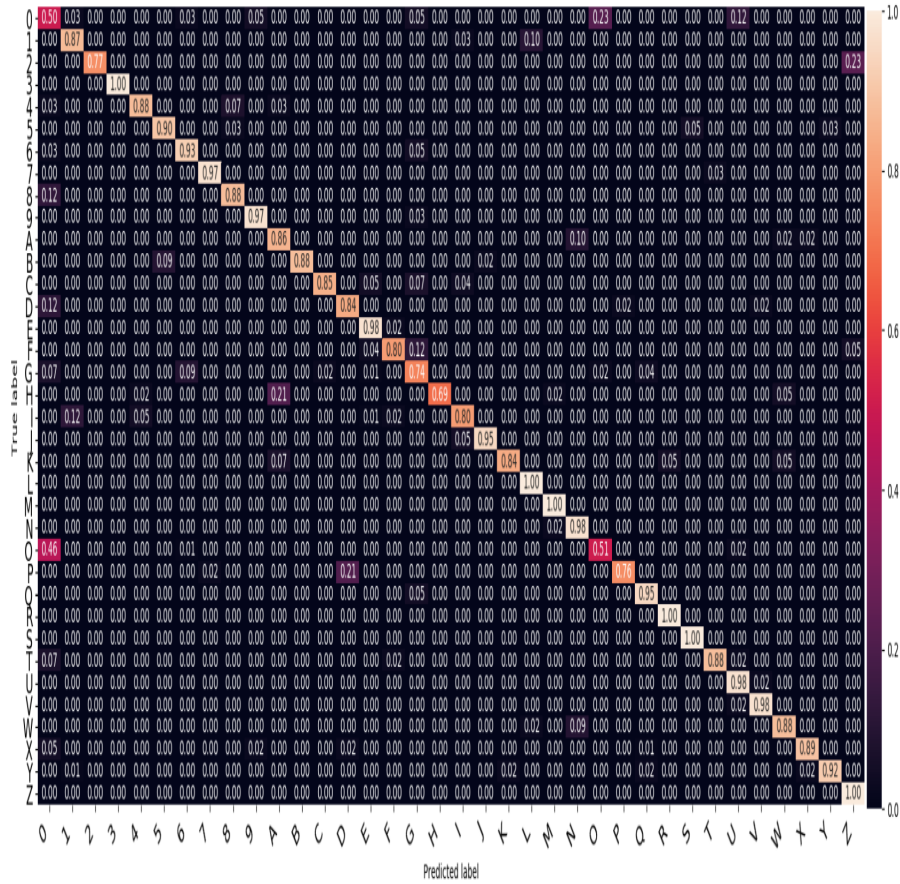


Figure 6.10: Confusion matrix of individual classification results of OS-CNN classifier.

Table 6.2: Model-based Evaluation Results

Classifier	Accuracy(%)		
	Overall	Numerics (0-9)	Letters(A-Z)
1D-CNN	88	81.8	91.3
BLSTM	67	63.5	68
DTW+KNN	94	85.4	95.5











Ground Truth	Reconstruction of Correct Prediction	Incorrectly Predicted Label	Reconstruction of Incorrect Prediction
0		0	
H		A	
G		6	
P		D	
2		Z	

Figure 6.11: Examples of trajectory reconstruction of obvious confusions in model-based evaluation. Black trajectory reconstruction represents the ground truth, and trajectory reconstruction in red represents the incorrectly classified samples.

6.5.3 Discussion

Based on user quality appreciation, we infer that numbers and letters, both uppercase and lowercase, that are neither too short nor too long to write are more likely to produce better results after trajectory reconstruction. For numerics, 3, 4, and 7 are the most recognized against 8, 9, and 5 as unrecognised. In upper-case letters, L, V, and C are mostly recognized, whereas I, E, and H recognition rates are not as expected. In lowercase letters, m, w, and v are recognized with the highest rate and i, j, and t are recognized at the lowest rate. A few examples of trajectory reconstruction of each class, which users find hard to recognize, are shown in [Figure 6.12](#).

In model-based evaluation, the recognition results for most characters are encouraging, as shown in [Figure 6.10](#). The digits (0-9) are correctly classified with the accuracy of 86.7 percent, lowercase letters with 87.9%, and upper-case letters with 88.6%. We merged the labels of uppercase and lowercase letters to minimize the confusion as most are of the same shape, which results in improved letter recognition at the accuracy of 91.3%. In digits, 3, 7, and 9, and in letters, L, M, and R are recognised with higher rates. The letter 'O' and 'o', '6' and 'G', 'p' and 'D', and 'Z' and '2' are mostly confused, as they look the same in appearance, as shown in [Figure 6.11](#). This problem can be addressed using the context information of the previous and next characters. Moreover, a combination of DTW with KNN delivered the best results, but the costs of the system (about 6.3sec/sample) are on the higher side. Therefore it is not recommended for real-time evaluation. On

Trajectory reconstruction performance for individual characters & numbers

Detailed results of model-based evaluation

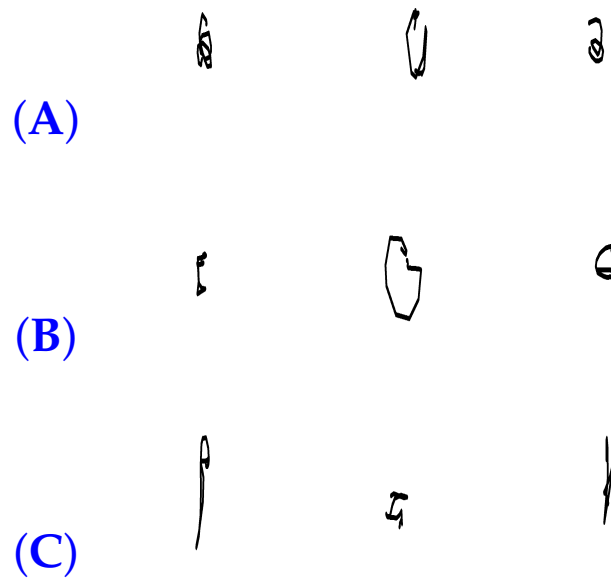


Figure 6.12: Failed examples for digits, lowercase letters, and uppercase trajectory reconstruction. (A) shows results of '8', '9', and '3'. (B) shows letters 'E', 'G', and 'S'. (C) show results of 'f', 'z', and 'k'.

the other hand, *OS-CNN* processes 5 samples per second, making it ideal for real-time use despite slightly lesser results.

IMU sensors are known for internal errors, i.e., noise and drift, accumulating over time. To evaluate the problem, we asked participants to vary the writing speed in repeated cycles to evaluate its impact on trajectory reconstruction. If the person writes too fast or otherwise, the results are not as good as expected, whereas better trajectory reconstruction is achieved at normal and slow writing speed. A comparative analysis of the results for the same character produced by the same user at varying speed is presented in [Figure 6.2](#). Moreover, we also present a simple mechanism to air-write words and sentences using the *FAirWrite* system by taking advantage of simple image processing algorithms, a major lacking functionality in existing systems.

*Exploring
functioning
limitations of IMUs*

*Comparison with
existing methods*

To draw a fair comparison with the existing system, *Airscript* system [51] presented by Dash et al. and *Smartphone IMU* system [179] are selected. *Airscript* uses *Myo*-armband, and its functionality is limited to writing and recognizing the numerals only (0-9). It requires full arm movement to air-write, as the sensor is worn on the forearm and requires a complicated calibration process every time before the system is used. The *smartphone IMU* base system for air-writing was presented in [179], which was further improved in [178]. This system is also limited to a single trajectory lowercase letter only, and carrying smartphones in hands to write with affects the natural writing pro-

cess. On the contrary, **FAirWrite** is worn on a finger, making it ideal for air-writing on an imaginary canvas because of its size and placement, as explained in [237]. Moreover, it offers complete functionality to write the numerals (0-9) and letters (A-Z), along with a classification module to classify the air-written characters into digits and alphabets. The simplified approach for trajectory reconstruction in combination with cost-effective Deep Learning (DL) models **OS-CNN** for the classification of air-written characters gives it a clear edge and makes it ideal for real-time use.

We introduce a novel and generic method called **FAirWrite** system to create the documents by air-writing with a finger on a virtual screen and with no limits of spatio-temporal boundaries. The proposed system is based on a single low-cost **IMU** with the applications to write in Virtual Reality (**VR**) and Augmented Reality (**AR**) scenarios. We also present a simple and real-time method to reconstruct the air-written trajectories on a 2-D display. An enhanced **FAirWrite** user interface enables the users to interact with the system and for real-time feedback. We evaluated the performance of **FAirWrite** system by user quality appreciation and exploring deep-learning methods to report the state-of-the-art (**SotA**) results. Both recognition methods achieved an overall accuracy of about 95%. The collected dataset from 100 participants is made publicly available for the benefit of the research community.

*Strengths &
limitations of
FAirWrite system*

Part IV

SUPPLEMENTARY WORK

Documents play a key role in the administrative functioning of an organisation. Documents are also one of the major forms of communication and exchange of information between organisations and within the organisations. Contracts, wills, invoices, certificates, and other documents used by organisations are often signed and stamped to ensure their authenticity. The attachment of stamps to the documents signifies their relevance and authenticity. Stamp detection and recognition is important to the automated processing and understanding of administrative documents. Stamp segmentation demands semantic analysis of the document images as stamps on documents may contain complex backgrounds and surround by unwanted data. This chapter of the dissertation focuses on stamp segmentation from scanned document images, isolating them from background data and other information present in document images.

Stamps highlight the significance of documents

The major contributions of this work are as follows:

- A stamp segmentation approach Deep Stamp Recognition ([dStaR](#)) is presented in this chapter. [dStaR](#) is the first end-to-end and trainable approach for stamp segmentation from various document images. The proposed approach uses Fully Convolutional Network ([FCN](#)) for semantic analysis of documents to extract stamps, the first methodology to adopt Deep Neural Networks ([DNNs](#)) for stamp detection problem.
- The proposed approach is evaluated on a publicly available stamp dataset. Evaluation results show that the presented approach outperforms the state-of-the-art ([SotA](#)) methods for stamp segmentation and achieves pixel-based precision and recall of 87% and 84%, respectively.

The rest of the chapter is structured as follows: [Section 7.1](#) presents the problem statement, motivation to solve the problem, and proposed solution to address the mentioned problem. [Section 7.2](#) summarizes not only the previous work done for stamp detection and verification domain but also provides an overview of the methodologies proposed after the publication of this work. [Section 7.3](#) elaborates the [dStaR](#) approach for stamp segmentation and detection with the details for the networks adopted for the proposed approach. [Section 7.4](#) presents the evaluation methodology, details of the publicly available

Structure of this chapter

stamp segmentation dataset, evaluation criterion, results of the proposed approach, and a detailed comparison with existing methods. The author of this dissertation has partially published the contents, figures, and tables in the following publications. The author originally wrote all the text, figures, and tables in the mentioned publications and this dissertation. More details about the publications included in this chapter are as follows:

- Younas J. et al., D-star: A generic method for stamp segmentation from document images. In: 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), pages 248–253, 2017 [[270](#)]



Figure 7.1: Textual, graphical, official and fun purpose stamps

7.1 MOTIVATION

Stamps are considered a mark of authenticity and originality of documents. Stamps mark the documents with creation, distribution, and storage information. Large organizations receive and send thousands of documents (including legal, financial, and security documents, bank receipts, checks, and utility bills) daily. These documents have single or multiple stamps on them. Every stamp on a document highlights the significance and purpose of the document. Stamps largely vary in shape, size, and color from organization to organization as well as within departments of an organization. In general, stamps appear in textual, graphical, regular (official), and irregular (fun purposes) shapes, as shown in [Figure 7.1](#).

Stamps are of gross importance in documents

Stamp segmentation is an important part of the automated classification and verification of documents. Stamps, however, may overlap text, logos, and/or other information in documents. Furthermore, the orientation of stamps varies from document to document, and different scanning environments also add their overhead. Hence, proper and correct stamp segmentation from scanned documents is a challenging problem.

Stamp segmentation is a challenging task

In the past, various approaches have been presented for stamp detection [5, 162]. Most of these approaches used different sets of stamp features, including color [162, 241], shape [19], and local keypoint descriptors [5, 57, 70] to separate stamps from logos, text, and other information present in scanned document images. These approaches may serve specific organizations which use a predefined set of stamps based on color, shape, or features. However, none of these approaches can be applied to develop a generic system for stamp detection and segmentation applicable to a vast majority of stamps with different colors, shapes and textures, especially for overlapping stamp segmentation. There are a few approaches published after the publication of mentioned work in this chapter [62, 195], we will discuss them in [Section 7.2](#) and the possible comparison in [Section 7.4.3](#).

Limitations of existing stamp detection methods

Deep Learning (DL) has been successfully applied for object classification and detection [129, 223, 286]. While DL has seen success

FCNs are tailor-made for semantic analysis of document images

with a breakthrough paper by Krizhevsky et al. [129], the history of successful Deep Learning (DL)-based methods for handwriting recognition [82] is rather old. For pixel-level image labeling, DL methods have been applied for binarisation and layout analysis [39, 181, 214]. However, Fully Convolutional Networks (FCNs) remains unexplored in this context. A closely related work [282] uses FCNs for detecting text in natural scenes. Stamps are not location-specific or content-specific objects in the document images. Therefore, stamp segmentation requires the semantic analysis of document images to segment them from overlapping objects or surrounding document objects such as text, figures, tables, and logos. FCN is the SotA approach for semantic analysis in natural-scene images, and their potential for semantic analysis of documents is still unexplored to the best of the author's knowledge(at the time of publication).

dStaR uses FCNs for stamp segmentation

Based on FCNs, we present a generic approach to segment stamps from scanned document images. The presented approach, named dStaR, can detect unseen stamps of any shape, color, size, and orientation. Moreover, dStaR can detect overlapping stamps. This is the first method to use deep learning for stamp detection. We used a Fully Convolutional Network (FCN) to segment stamp masks from scanned document images. Contour refinement is applied to the predicted masks for pixel-based evaluation and reforming the original stamps. The proposed method is evaluated on a publicly available stamp detection and verification dataset [162] where it yields pixel-based precision and recall of 87% and 84%, respectively.

7.2 RELATED WORK

Heuristic-base approaches

Different methods have been proposed to detect the stamps from document images in the past. Most of these methods used heuristics-based approaches to detect and classify stamps. Some approaches used color-based features to segment stamps, while other approaches used geometric features with key-point descriptors to extract the stamps from scanned document images.

Color profiles for stamp detection

One of the earliest methods to detect seal imprints and signatures from checks of Japanese banks was proposed by Ueda et al. [241], assuming signatures, seal imprints and backgrounds to be different from each other. It uses the color information RGB 3D to detect stamps and signatures from images. The proposed technique fails when any of the three clusters is not monochromatic.

XY-cut algorithm for stamp detection

Micenkova et al. [162] present an automatic segmentation and verification system to detect and verify the stamps from scanned document images. This approach is based on color segmentation of documents in $Y C_b C_r$ color space. Candidate solutions are extracted using

XY-cut algorithm [128]. Candidate solutions are further processed using geometrical and color-based features to extract the stamp region. This approach does not address black stamp detection. It also fails to detect the stamp when the stamp background matches the color of the stamp. The extended model with stamp verification is presented in Micenkova et al. [163].

Ahmed et al. [5] presented a part-based feature extraction method. It uses a two-step approach to classify stamps from non-stamp regions using geometrical features. First, it computes the key points and then descriptors from these key points. Their presented approach outperforms other methods in detecting black stamps, while the results reported for colored images in [5] are on the lower side.

Part-based feature extraction method

An outliers-based approach has been presented to detect the stamps and logos from scanned document images by Dey et al. [57]. This approach assumes that the documents only consist of text, stamps and logos. Considering stamps and logos as outliers, it divides the document into foreground and background using Principal Component Analysis (PCA) and color information, treating both as separate images. These images are then individually processed further for pixel-level evaluation to segment out the stamps and logo regions from document images.

Considering stamps as outliers for detection

Another shape-specific segmentation approach is presented in Forczmański et al. [70]. This approach uses color space transformation to look for potential color stamps, followed by different object detection algorithms to compute the shape descriptors. Isolated regions are extracted from scanned documents and classified using computed shape descriptors. This approach potentially addresses the detection of well-defined shapes (official stamps) regardless of stamp color.

Shape descriptors for stamp detection

Bhalgat et al. [19] proposed a shape-specific stamp segmenting approach using exemplar features. This approach uses unsupervised learning methods to extract the dictionary items for stamp shapes. Feature vectors are extracted using a single Convolution layer with 4×4 quadrant max-pooling. A dictionary ranking item scheme is used for the recognition of stamps. This approach produces excellent results for only oval-shaped stamps.

Exemplar features & supervised learning approach for stamp detection

Note that Dey et al. [57] also presented an outliers-based approach to detect stamps and logos from scanned document images. It is a highly fragile and prone to error approach as it is explicitly based on a large set of experimentally computed parameters generated from the whole dataset. These computed parameters are then applied to the same dataset for evaluation. Hence, the comparison does not stand valid with this approach.

Limitation of outliers-based approach

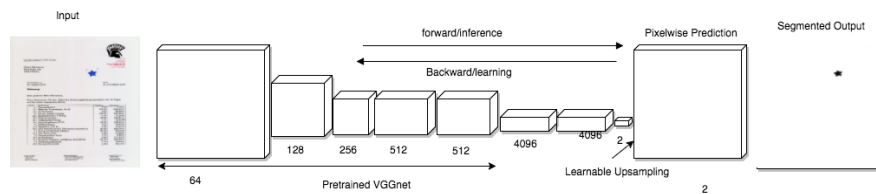


Figure 7.2: dStaR architecture with input image and pixel-level segmentation result of FCN

Related work published after presented work

The discussed methods use heuristic approaches and have their constraints in particular scenarios. Deep Learning (DL) approach has not been used to detect the stamps from the images until time up to the best of author's knowledge. Rajab et al. [195] presented a method to segment the stamps from document images using local k-means and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) algorithms. The proposed methodology uses clustering techniques to isolate stamp clusters from background clusters and then refine those clusters to extract the binary mask for the stamp area. The evaluation protocol and dataset used to evaluate the proposed approach rule out the fair comparison with our proposed methodology.

Combining supervised & unsupervised learning for stamp detection

Duy et al. [62] present an approach for stamp segmentation and verification. The proposed approach is a combination of unsupervised learning methods and State Vector Machines (SVMs). An unsupervised learning machine method detects all the objects on the document image, followed by an SVM model to classify stamp and no stamp regions. At the last step of the proposed methodology, second SVM classifiers are used to verify the authenticity of the classified stamps at the prior step as forged or authentic stamps.

7.3 DSTAR: THE PRESENTED SYSTEM

Overview of proposed approach

This section details the presented approach Deep Stamp Recognition (dStaR) for stamp segmentation in document images. Figure 7.2 shows architecture of dStaR, which uses a FCN to generate semantic segmentation of input images. The generated segments are pixel-level maps of stamp location in the original scanned document images. The FCN's generated stamp maps are then post-processed using connected component analysis to detect the exact stamp pixels from the input image. Usually, DL based approaches require a lot of training data. However, the publicly available "Stamp Detection and Verification" dataset we used contains only 400 scanned document images. Therefore, to resolve this problem, we used the concept of domain adaptation and transfer learning to train the Fully Convolutional Network (FCN).

7.3.1 Domain Adaptation and Transfer Learning

In this paper, we adapted the domain of general-purpose object detection and segmentation from natural scene images to the segmentation of stamps in document images. Both of these domains are different. Furthermore, to compensate for the problem of the non-availability of a large dataset, we used the concept of transfer learning. Transfer learning is defined as a transfer of knowledge from a learned task to a new task [177]. In Convolutional Neural Networks (CNNs), transfer learning refers to using learned features from a pre-trained network for a new task. Pre-trained network is particularly useful when we need more data to train a new network. In dStaR, we used a pre-trained VGG-Net16 [223] for transfer learning. The VGG-Net16 was trained on PASCAL-VOC-2011 dataset [65]. VGG-Net16 was preferred as the backbone because it produced better segmentation results despite inference time on the higher side [19].

Domain adaptation enhances the performance of DNNs

7.3.2 VGG-Net

VGG-Net is runner up for ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. It is an advance CNN architecture, which takes the fixed size input of 224×224 RGB images. The filter size used in VGG-Net convolution layers is 3×3 , the smallest possible receptive field size to capture the features. The convolution stride is fixed to 1 pixel. After convolution, spatial pooling is carried out by 5 max-pooling layers with a 2×2 pixel window, with a stride of 2. Input images are processed through a stack of these convolution layers followed by three Fully Connected (FC) layers. For detailed information, we refer our readers to [223].

VGG-Net is used as backbone of dStaR

To adapt the pre-trained VGG-Net to the problem of stamp segmentation and to improve the performance of FCN, we removed the FC layers and used the output of 5th max-pool layer for fully convolution processing. Figure 7.2 provides an overview of the architecture of pre-trained VGG-Net used in dStaR. By removing FC layers from VGG-Net, we can process input images of arbitrary size.

7.3.3 Fully Convolutional Networks (FCNs)

Fully Convolutional Networks (FCNs) can process arbitrary-sized images due to the absence of Fully Connected (FC) layers in the network at the end, which requires fixed size input. FCNs are mainly used for semantic segmentation of images in which pixel-level output is generated by combining context from higher layers and information from lower layers while retaining spatial information. Deconvolution layers are used for decoding the embeddings generated by the encoder. No learning is needed for deconvolution layers as these are initialized as

FCNs are widely used for semantic analysis

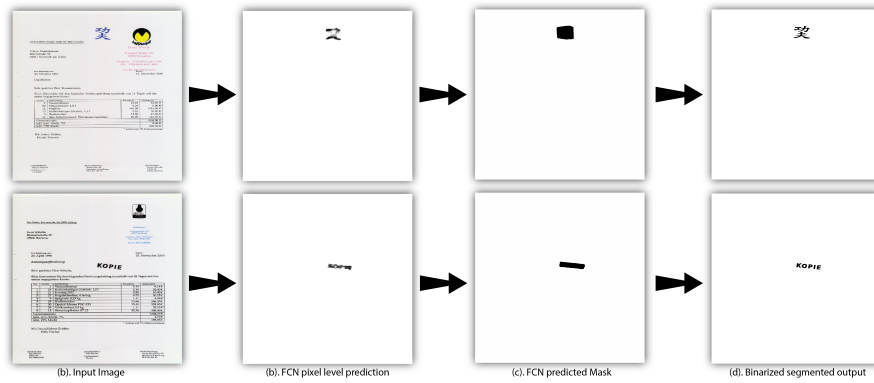


Figure 7.3: dStaR overview, (a) Two different input images, one containing graphical and colored stamp, while the other input image with a black and textual stamp on it, (b) shows the results for both scanned document images for pixel-level predictions, (c) shows the processed predicted masks from FCN, and (d) shows the detected and segmented binary stamps.

bilinear up-sampling layers. FCNs are given priority over conventional CNNs because of following advantages [149]:

- Highly computational efficient networks. It takes about 100 milliseconds to process 1000×1000 image in the training phase.
- FCN generates pixel-level masks for every corresponding class because of their ability to perform segmentation at higher levels.
- FCN can be built on any state-of-the-art (SotA) CNN, e.g., Alexnet, ResNet, Image-Net, Google-Inception models, rendering vast scope of adaptability. Using pre-trained networks significantly boost and fine-tunes the performance of FCNs, helping in very fast convergence even on very small datasets.

Benefits of FCNs over traditional CNNs

As FCN processing is on pixel-level, it needs per-pixel annotations for training. Our focus in the scanned document images is stamp regions. So, we used the annotations containing only the pixel-level stamp masks, resulting in a binary classification task for FCN.

dStaR network architecture

We used convolution layers from pre-trained VGG-Net. Three fully convolution layers on top of VGG-Net were added to perform FCN functionality. The kernel size for the first, second and third FC layers are 7×7 , 1×1 , and 1×1 , respectively. The output of size 512 generated by the fifth max-pool layer of VGG-Net served as input of the first fully convolution layer. The output of the first fully convolution layer of size 4096 is the input of the second fully convolution layer, which generates the same output as the input. The output of the second fully convolution layer is then processed in the last fully convolution layer for pixel-level prediction of every class.

As we used the FCN8-s, it means 3 levels upscaling or deconvolution are done to produce per pixel classification at stride 8. FCN8-s uses predictions from max-pooling layer 5,4 and 3 respectively, of stride at 8 to generate the pixel-level predictions as shown in Figure 7.2.

We tried images of different sizes as input of FCN for optimal performance because the number of scanned document images is very less in the context of Deep Learning (DL). We used RGB images of size 1000×1000 as input to our FCN. FCN was trained with pixel-level binary masks, marking stamps only as Point of Interests (PoIs). The training was done for 10 epochs. We used the batch size of 2 in the training phase. To fine-tune the network parameters and optimize the performance, the learning rate of .0001 was used.

Network parameters for optimized performance

FCN generates the pixel-level predictions for stamp regions. Figure 7.3 shows the complete work-flow of the dStaR with intermediate results on each step. The pixel-level predictions are post-processed using Connected Component Analysis (CCA) to generate FCNs masks for stamp segmentation. These predicted masks are used to extract stamp pixels from the input image. The stamp-segmented images are then converted to binary images for evaluation purposes.

Refinement of network output

7.4 EVALUATION

7.4.1 Dataset

For evaluation of dStaR, we used a publicly available stamp detection and verification dataset¹ [19]. This dataset contains 400 document images scanned at 200, 300, and 600 dpi resolution. The scanned document images contain printed text, stamps (textual and non-textual), logos, and signatures. The dataset contains stamps of varying sizes, shapes, and colors. 341 (out of 400) scanned document images contain single or multiple stamps; the remaining 59 images have no stamps. Out of 341 scanned document images, 80 contains black stamps, and 241 contains colored stamps. In 55 scanned document images, stamps are overlapped with text, logos, and/ or signatures. For every scanned image, there are two ground-truth images available; one containing the pixel-level information and the other containing the bounding box information for each stamp. So, this dataset can be used for region classification and pixel-level evaluation.

Stamp detection dataset

7.4.2 Evaluation Protocol

For the Evaluation of dStaR, the dataset has been split into train and test sets with different configurations. We used document images

Test & train split for evaluation

¹ Stamp segmentation and verification dataset

scanned at 200 dpi resolution. The train set contains 90% of scanned document images, and the remaining 10% are used to evaluate the proposed approach. As we evaluate [dStaR](#) for three different categories of scanned document images, we customized the test sets with only colored, black, and overlapping stamps, respectively. Furthermore, we evaluated our presented approach with a system-generated random test set and class-balanced data (equally distributed samples from every category) for overall performance evaluation.

We evaluated the [dStaR](#) for pixel-level detection of stamps, and therefore, the most relevant evaluation metrics are precision and recall [184]. Precision is the intuitive ability of a classifier to distinguish a negative sample from a positive one. It is computed as:

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad (7.1)$$

The recall is the ability of a classifier to classify all the positive samples. It is computed as:

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad (7.2)$$

Evaluation metrics

In equations 7.1 & 7.2, tp, fp, and fn denote the true positives, false positives, and false negatives, respectively. True positives refer to the predicted number of pixels belonging to stamps. False positives refer to the number of pixels predicted as stamps but do not belong to stamps. False negatives specify the number of stamp pixels the system fails to predict.

7.4.3 *Results and Discussion*

Results are compared considering different aspects

We present a detailed comparison of our presented approach, considering different aspects (segmentation of colored, monochrome, and overlapping stamps) with the [SotA](#) approaches. Our presented approach is independent of the shape, size, color, and orientation of stamps with regard to text or logos present in scanned document images. Results used for the comparisons are computed when only the mentioned stamp category was present in the test set, with the rest of the scanned document images in the training set.

Results for segmentation of overlapping stamps

Segmentation of overlapping information is a very difficult task in information segmentation and classification [4, 154]. [Table 7.1](#) shows the results when [dStaR](#) is tested on overlapping stamps. These stamps overlap with the background text and/or logos at different positions. [dStaR](#) outperforms the [SotA](#) in segmenting overlapping stamps by a large margin. It correctly segments the overlapping stamps with pixel-level precision of 74% and recall of 77%. [Figure 7.4](#) shows some overlapped stamps and their binary segmented results by [dStaR](#). [Mícenkova et al. \[162\]](#) reported the pixel-level recall and precision of 69%

Table 7.1: Performance Evaluation of dStaR on overlapping stamps

Approach	Precision(%)	Recall(%)
dStaR	74	77
Micenkova et al. [162]	68	69
Ahmed et al. [5]	Not Reported	

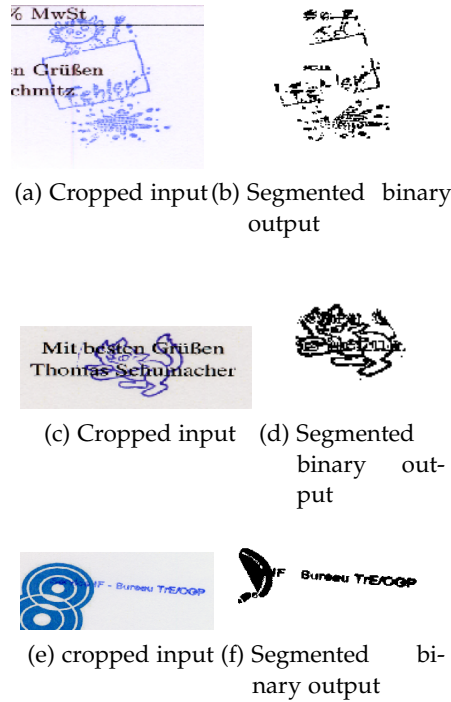


Figure 7.4: Overlapping stamps detected by dStaR successfully (Overlapping images with predicted binary outputs)

and 68%, respectively, for overlapping stamps. Ahmed et al. [5] did not present results for overlapping stamps in the dataset but mentioned that their approach fails to detect the severely overlapping stamps.

Furthermore, we also evaluate dStaR from another dimension, i.e., colored stamp detection and segmentation. Table 7.2 provides results of colored stamp detection. For colored stamps, the presented approach reports the pixel-level precision and recall of 92.7% and 84.3%, respectively. Micenkova et al. [162] reported pixel-level precision and recall of 82.7% and 82.8%, respectively. Their approach fails to detect the stamps when their color matches the background, whereas dStaR can also be used successfully in such scenarios. Ahmed et al. [5] approach is, although independent of color and shape of stamps, as it uses part-based key points and feature descriptors for stamp detec-

Evaluating dStaR for coloured stamps

Table 7.2: dStaR in comparison with the SotA approaches for coloured stamps

Approach	Precision(%)	Recall(%)
dStaR	92.7	84.3
Ahmed et al. [5]	62	57
Micenkova et al. [162]	82.7	82.8



Figure 7.5: Stamps dStaR failed to segment out.

tion, logos are also misclassified as stamps [5]. Therefore, their results are on the lower side with pixel-level precision and recall of 62% and 57%, respectively.

Evaluating dStaR for monochrome stamps

Table 7.3 elaborates the precision and recall comparison of dStaR with the existing SotA approaches for black stamps in scanned document images. Micenkova et al. [162] approach assumes the stamps as colored objects only by processing the stamps document images for YC_b C_r color clusters. These color clusters are used for the segmentation and detection of stamps. When it comes to black stamps, this approach does not stand valid (applicable). Ahmed et al. [5] approach report the pixel-level precision and recall of 83% and 73%, respectively, in comparison to the dStaR 93.75% and 50.2%, for black stamps.

Overall results on randomly generated test-set

Table 7.4 reports the overall stamp segmentation results of dStaR. The SotA approaches do not report their overall results. dStaR, however, achieve pixel-level precision and recall of 87% and 84%, respectively. Note that the test and training set division for evaluations (presented

Table 7.3: dStaR evaluation for black stamps w.r.t the SotA approaches

Approach	Precision(%)	Recall(%)
dStaR	93.75	50.2
Ahmed et al.	83	73
Micenkova et al.	Not Applicable	

Table 7.4: Overall performance of dStaR on randomly generated test set

Approach	Precision(%)	Recall(%)
dStaR	87	84
Micenkova et al. [162]	Not Reported	
Ahmed et al. [5]	Not Reported	

in [Table 7.4](#)) have been made the system randomly and autonomously without any human intervention.

This dissertation chapter is divided into two parts, each as an outcome of collaborative work. The first part focuses on applying wearable glasses to study and analyse their impact on learning outcomes while performing Physics experiments. The second part focuses on another application of wearable sensors for indoor activity sensing using near-field electric field principles. Consumer-friendly hardware of smart gadgets results in growing reliance on and frequent use of them in everyday life activities. Smart gadgets are gaining great attention to develop applications in the field of medicine, sports, heritage, gaming, entertainment and many other research areas using Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) in a combination of AI. This increased influence of smart gadgets and their applications urges the need for their adoption in formal education to benefit learners and instructors and enhance the overall learning experience. Smart glasses are a great platform to use as an assistive tool to experiment and interactively learn complex concepts. The first section of this chapter presents a study evaluating the impact of Google Glass on learning outcomes for performing Physics experiments. Contributions of the author of this thesis in this regard are summarised as follows:

- Evaluating the impact of smart glasses on learning outcomes during Physics experimentation using gPhysics application. The study evaluates wondering, curiosity, learning achievement, cognitive load, and experimentation time among the students during the task to study the relationship between the frequency of the sound generated by hitting a glass of water and the amount of water in the glass.

Smart wearable systems are referred to as a set-up of number of sensors attached to the human body to collect data and receive commands from the user. Different Smart wearable systems employ multiple sensors to measure the proportion of the human body for human activity detection and recognition. Human activity recognition plays a significant role in the advancements of human-interaction applications in healthcare, personal fitness, and smart devices. Given smart watches' wide and increasing popularity, the wrist is a compelling location for placing sensors. On the other hand, only specific information, such as hand/arm motions and selected physiological signals,

Wearable devices present a vast application paradigm to assist the learning process

Exploring wearable for activity sensing

are readily available at the wrist. In this dissertation chapter, we explore a novel wrist-worn sensing approach that allows information not typically associated with the wrist or the arm to be acquired by exploring the ubiquitous near-field electric phenomena. Major contributions of this work are as follows:

- A use case to demonstrate the collaborative work by two people is recorded by deploying our prototypes both at surrounding objects and on wrists, presenting the feasibility of collaborative work monitoring by sensing the variation of the near-field electric field.

The rest of the chapter is divided into two major parts; [Section 8.1](#) covers the details of a study to evaluate the impact and effectiveness of using Google Glass as an experimentation tool in Physics education. [Section 8.1.1](#) presents the case for using smart glasses, Google glass in our case Experimentation and study design is part of [Section 8.1.2](#). Finally, results are presented in [Section 8.1.3](#). [Section 8.2](#) covers the details of the WristSense application. [Section 8.2.1](#) states the problem, motivation along with possible solutions, and proposed solution to use wrist-based sensors for human activity recognition. [Section 8.2.2](#) familiarizes the readers of this dissertation with the recent developments for wrist-based sensing devices and their applications. Background knowledge and details about the proposed sensing prototype are provided in ???. [Section 8.2.3](#) explores the possible applications of the proposed prototype, such as touch sensing, proximity sensing, and activity sensing.

For the First part, the author of this dissertation has contributed to the experiment conducted for the study and evaluation of results related to the gPhysics application using Google Glass only. For the second part, the author of this dissertation has contributed to brainstorming, performing the experiments, and content-writing, with the major contributions of the first author, Sizhen Bian, in the publication. Content such as text, figures, and tables included in this chapter are taken from the mentioned publication. More details about the publications included in this chapter are as follows:

- Kuhn J. et al., gPhysics—Using smart glasses for head-centered, context-aware learning in physics experiments. In: IEEE Transactions on Learning Technologies, 9(4):304–317, 2016 [132]
- Bian S et al., Wrist-worn capacitive sensor for activity and physical collaboration recognition. In: IEEE International Conference on Pervasive Computing and Communications Workshops(PerCom Workshops), pages 261–266, 2019 [22]

8.1 GPHYSICS

8.1.1 Motivation

Smart glasses can be used to present data in multiple representations

Smart glasses such as Google Glass [228] are a new class of wearable computers combining classical HMDs with multiple sensors for head-centered, contact-free, sensor-based, and context-aware interactions. Google Glass combines an HMD, a headphone, a multi-touch track-pad, head motion sensing, eye-blink detection, a microphone, a first-person camera, storage, and communication capabilities. The use of smart glasses enables its users to seamlessly blend their interactions in physical and digital worlds, presenting an ideal ground to prosper in the direction of MR applications. Smart glasses allow the analysis of physical data that can be presented to the user with different kinds of representation, i.e., explanatory text, tables, graphs, pictures, and equations. These advantages of smart glasses present a compelling case for their applications in smart experimentation and problem-solving in science education.

Applications and educational properties of wearable glasses

The educational properties of wearable technologies are widespread [27] in literature across different fields. Google Glass has been used in medical training role-play activities using simulations to train in AR, first-person viewpoint, and recording for analysis and observation [261]. Other applications of Google Glass in the medical field are broadcasting a procedure on HMD [126], video recording students during standardized patient encounters [238], evaluating their interpersonal communication skills and non-verbal behaviours [238], and tele-monitoring trainees for ongoing procedures such as shoulder surgery [144] and cardiac ultrasonography [203]. An application of Google Glass to augment information about paintings in an art gallery is present in [137]. Smart Glass implementations of smart glasses have focused more on traditional and general applications in an educational setting, such as recording actions and presenting and sharing information. However, there is little progress or work concerning the use of wearable technologies in education to examine implementation in a regular context (at the time of publication, leaps of progress have been observed since then).

Google glasses can help students through interactive learning

Wearable technologies, in general, and smart glasses, in particular, present a huge opportunity for interactive and engaging educational tools helping students learn complex concepts in STEM education. When aided with multiple representations, students can perform better in learning outcomes and problem-solving in Physics education [6]. This sets an ideal stage to explore their potential for the topic of acoustics, as multiple representations are important for learning content with oscillograms, frequency spectra, etc. "gPhysics" presents a MR application of Google glass as an experimental tool for students as-

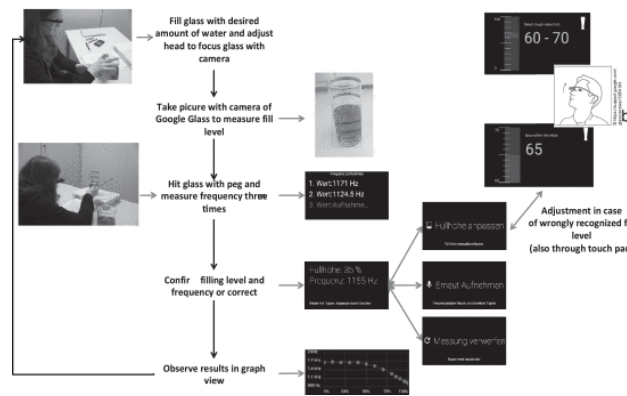


Figure 8.1: An overview of performing experimentation steps using gPhysics application on Google glass

sisting them in performing the physics experiments to perform acoustic related experiments [133]. This work uses the gPhysics application to study the effectiveness of smart glasses in smart experimentation in learning outcomes. Outcomes of the study reveal that smart glasses help to foster wondering, curiosity, and learning achievement and reduce cognitive load and experimentation time. Further details are provided in the following subsections.

8.1.2 Experimentation & Study Design

In this work, students use the gPhysics application to study the relationship between the fill level in the glass and the frequency of the generated tone and the relationship between the water fill level in the glass and the frequency of the induced tone by damping the vibration of the glass while hitting the glass with a wooden peg. Figure 8.1 shows the complete experimentation process imitating intermediate steps using the gPhysics Google Glass application. The first activity of the design experimentation is to adjust the initial fill level of the water glass using a double blink of the right eye or tap on the touchpad. Then participants filled the glass to the desired mark and confirmed the water level shown on the Google glass display with the possibility to adjust it. After confirmation, participants were directed to the measuring menu and asked to measure the frequency value by hitting the water-filled glass with a wooden peg. The tone frequency is recorded automatically, and after confirmation, a graph is displayed to show the relationship between the current frequency and water-fill level. The same procedure is repeated for given filled levels to complete the experiment. The whole experiment is repeated after fixing the hair tie to the top of the water glass.

Experimental design and process

Participants were 8th-grade students from a German high school divided into three groups based on smart gadgets they used to per-

Study design

form experimentation. Students who performed the experiments with Google Glass were placed in Intervention Group (IG), android tablet group as Control Group (CG)-1, and participants who used the SpectrumView application on iOS tablets were placed in CG-2. IG and CG-1 group were assisted by respective gadgets to perform activities and tasks, e.g., measuring frequency, fill-level, and plotting graphs, whereas CG-2 performed these activities manually. In total, 46 students participated in the study; out of them, 19 students were placed in IG, 13 in CG-1, and 14 in CG-2. Before starting the study, students grades in Physics, Mathematics, and German were gathered, and pre-test measures were to evaluate their familiarity with the use of smartphones and tablets and curiosity concerning smartphones, tablets, and Google Glass. A training session was also conducted for students to familiarize them with using smart devices and applications. After experimentation, post-test measures were administered on participants state of curiosity about the learning experience, use of the device (based on which group they belong to), wondering, cognitive load, and learning achievement for comparison with the pre-test measures. Analysis of Covariance (ANCOVA) technique is used to analyse the results of the study where wondering, cognitive load, curiosity, experimentation time, and learning achievement in Physics were considered as dependent variables. Treatment group (IG, CG-1 or 2) and gender were considered as independent variables; covariates of the study were curiosity, trait, grades from current school results in Mathematics, Physics, and German along with the familiarity of students with mobile devices and the pre-test data on learning achievements. η_p^2 is used as measure of effect size.

8.1.3 Results

Google glass fosters the wondering and curiosity

The three groups differed significantly with large effect sizes concerning the variables wondering: ($F(2,46) = 6.02; p < .01; \eta_p^2 = 0.23$); curiosity state(total): ($F(2,46) = 10.17; p < .001; \eta_p^2 = 0.36$); curiosity state(formal): ($F(2,46) = 6.94; p < .05; \eta_p^2 = 0.27$), curiosity state (informal): ($F(2,46) = 8.11; p < .01; \eta_p^2 = 0.31$). The effects in total as shown in Figure 8.2 are related to a higher value of each dependent variable in the IG reveals that using gPhysics application fosters both the wondering, curiosity in total and curiosity state with regard to informal learning. The curiosity state in formal learning remains consistent among the gPhysics users but differs for CG-2 (SpectrumView tablet) group.

Students using Google Glass for experimentation experience higher cognitive load

In terms of cognitive load, it is observed that IG using gPhysics on Google Glass experienced higher cognitive load in comparison to CG-1 & CG-2, as shown in Figure 8.3. The total cognitive load variable: ($F(2,46) = 4.12; p < .05; \eta_p^2 = 0.17$); cognitive load specific to device:

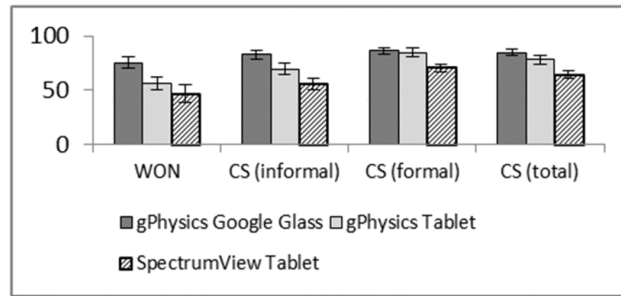


Figure 8.2: Wondering & curiosity state in terms of mean values & standard error

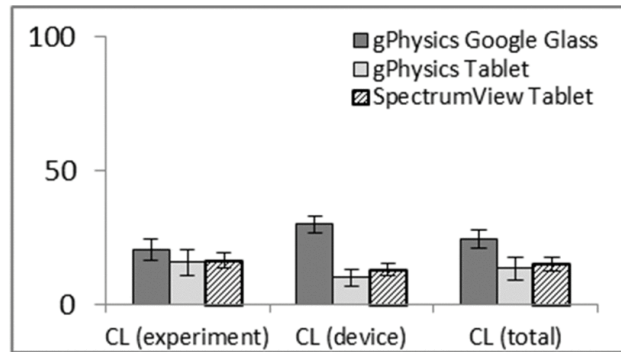


Figure 8.3: Cognitive load in terms of mean values & standard error

($F(2,46) = 14.46; p < .001; \eta_p^2 = 0.47$). It is also observed that grades in Mathematics also had a significant influence on the dependent variable; ($F(1,46) = 7.45; p < .05; \eta_p^2 = 0.17$). Therefore, besides the correlation of the grade in Mathematics trait with learning achievement in this topic, it shows the difference between the groups regarding their grades in Mathematics explained 17% of the variance of learning achievement in acoustics.

The same trend has been observed for differences in experimentation time and cognitive load variables. The three groups differed significantly with large effect sizes concerning the experimentation time for all three repetitions. Figure 8.4 shows that the IG was faster than CG-2 group but slower than CG-1 group. For both the CG-1 and IG group, the execution time significantly decreases between the first and second experiments. This is a clear indication of learning associated with the device. However, between the second and third experiment, the execution time decrease for IG and increases for CG-1. This indicates a much stronger learning effect for IG.

Using Google Glass for experimentation indicates positive learning effects

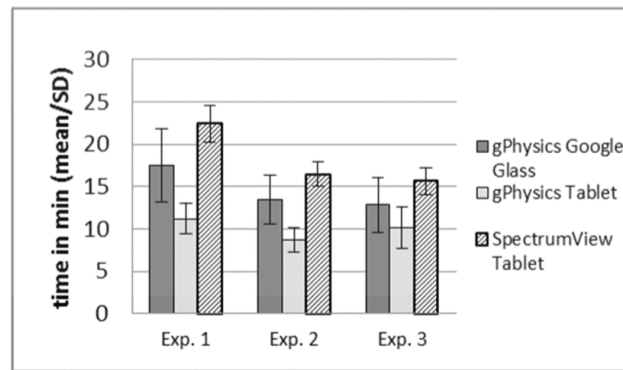


Figure 8.4: Experimentation time (mean values & standard error) of three groups for three experimental procedures

8.2 WRISTSENSE

8.2.1 Motivation

Electric field information can be used for activity sensing

An ambient electric field is ubiquitous as every subject carries a certain amount of charge, even an insulator [229]. For example, when a human body moves on the ground, the tribological interaction [36] between body and ground will generate an electrostatic charge on the human body, thus setting up a static electric field between the human body and ground [69]. Appliances at home also radiate electric fields [191]. Those fields could be distorted by surrounding disturbances or self-movement. For example, a refrigerator, as a radiator of an electric field, its field can be distorted by an intruder like a human body. Walking can cause the variation of potential on the body, namely the variation of an electric field from the body to the ground, which could also be described by variation of Human Body Capacitance (HBC).

Limitations of vision-based systems in ambient environment

Vision-based systems use cameras that restrict the mobility and freedom of users [201, 262]. To address the limitation of vision-based systems, on-body sensors such as motion sensors [45, 85, 180] and capacitive sensors [40, 192] are widely used for proximity sensing in ambient environments. Capacitive sensing techniques analyse the variations in capacitance (near-field) caused by human body intervention for activity sensing in ubiquitous environments. Moreover, this work can use capacitive sensing remotely, on-body as wearable, and/or hybrid.

Physical activity sensing using finger-worn sensor

The core idea behind the presented work is that electric field and capacitance changes related to the body can be sensed at any chosen body location, including the wrist, thus significantly enhancing the information that can be extracted from a wrist-worn device. Since such information includes not just the activities of the user her/him-

self but also changes in the environment, a key application that we explore is performing physical activity-related tasks in collaboration, which is difficult to capture using other sensors [87]. Furthermore, we demonstrate how the proposed prototype can detect motions of various body parts beyond the wrist, such as touch and proximity between users and objects.

8.2.2 Related Work

Different types of sensors are ubiquitous for human activity recognition for several reasons, such as low power consumption, mobility, readily availability, low price, and ease of wearing/attach to the body. There is a growing acceptance of capacitance-based sensing to monitor and recognise human activity among the research community. Thus by sensing the variation of the human body-related near-field electric field, a wide range of applications can be covered [105, 142, 224]. Zimmerman et al. [288] presented a use case of on-body capacitive sensing for user interfaces for the very first time. Afterwards application of capacitive sensing are explored for activity monitoring [40, 93], proximity sensing [86, 103], and touch sensing [192, 199].

Different sensors have applications for human activity recognition

Harland et al. [93] proposed remote, off-body sensing of the electrical activity of the heart at distances up to 1m from the body to high-resolution electrocardiograms. Pouryazdan et al. [192] used electric potential sensors to sense hair touch and restless leg movement. Grosse-Puppendahl et al. [86] developed a Platypus system to localize and identify people by remotely and passively sensing changes in their body electric potential. Cheng et al. [40] presented an on-body capacitive sensing approach to monitor human activities and physiology-related information at multiple body locations.

Human activity sensing

Wimmer et al. [255] present a "CapToolKit" for realizing capacitive sensing applications for Human Computer Interaction (HCI). The proposed system is very easy to install, cheap to build, and easy to adopt for multiple applications. In another work, Wimmer et al. [256] introduced two pieces of activity-sensing furniture, CapTable and CapShelf, using networked capacitive sensors. The proposed hardware can extract activity patterns based on hand and body motion information.

Toolkit & activity sensing furniture

Valtonen et al. [243] proposed a system that uses capacitive sensing information for indoor positioning and activity recognition in smart homes. The proposed system is capable of positioning a person at floor level and monitoring its interaction with the surrounding items in smart homes. The proposed system relies on the conductivity of the human body and the capacitive coupling of low-frequency signals between electrodes from the floor and the environment. The authors

Capacitive sensing for indoor positioning & activity recognition

demonstrated the efficacy of the proposed system to monitor a person unobtrusively without compromising an individual's privacy.

*Gym workout
counting &
recognition*

After the publication of this work, Bian et al. [21] presented an approach for full-body gym workout counting and recognition using the prototype presented in this work. The results show that capacitive sensing can recognise human body activity when the sensor is attached to a body part not directly engaged in activity movement. The proposed method achieved the average counting accuracy competitive with motion sensors attached directly to the part of the body involved in activity movement.

8.2.3 Capability Exploration

*Performance
exploration for
real-world use-cases*

To explore the capability of our sensing modality, we deployed our prototypes on various objects involved in the human body's action to monitor collaborative work. Plenty of the related capacitive coupling-based works, developed for an ambient intelligence scenario, focused on single side context, either perceiving information from the actuator of action [45, 46, 85] or from the reactor of an action [10, 11, 28]. For example, Cohn et al. [45] mentioned the ability of capacitive sensing by recording the repetitive motion of the body, and Arshad et al. [10] developed a floor-sensing model for elderly tracking and fall detection. However, integrating our sensor both in the actuator side and reactor side in the environment will provide complete information and thus provide a better understanding of the interaction of the individuals with the environment, as both the source and receiving end action will generate signals. Thus we set up a simple collaborative task in which two participants are involved and interact with the ambient environment. First, we describe the basic sensing ability and the background principle of our prototype in the following sections, which enable the monitoring of a whole group work process.

8.2.3.1 Touch Sensing

*When touch happens,
capacitance changes*

Touch is one of the basic interactions between people and their surroundings. Sensing approaches like infra-red camera [159], pressure sensor [199], acoustic signal [94], etc., are used to detect this action. The basis of touch sensing of our prototype is that when touch happens, the human body will supply a different path for the charge on the object to flow to a lower potential plate (sinking charge [288]) until the potential difference disappears. Once the charge flow-caused voltage variation is observed, a touch event can be detected.

*WristSense can
detect touch action*

Figure 8.5 shows the potential variation of two related prototypes, one is attached to a chair with an internal metal structure and paint on the surface, and another is attached to a person's wrist. The po-

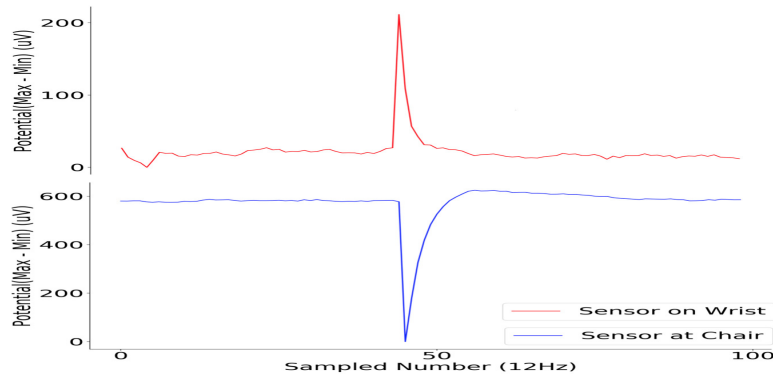


Figure 8.5: Touch a chair with prototypes at a chair and on wrist

tential variation happens at the same time but in the opposite direction. Before the touch, the potential of the touch point at the chair side is higher than the potential at the human body's hand side. The charge will stop flowing when the touch position at both sides shares the same potential. Then, the potential will be balanced to its former level at each electrode position. Taking the hand away from the chair will not cause charge flow anymore because there is no potential difference on both sides. That is to say, the prototype can detect touch action, but the 'remove' action is beyond the sensor's capability.

8.2.3.2 Proximity Sensing

Proximity detection is a primary sensing approach in Ambient Intelligence scenarios. Unlike camera [201], light [35], capacitance-based proximity has the advantage of low power consumption and effortless system establishment. The basic background of this sensing approach resides in the proximity caused by dielectric or distance variation in a capacitor. Capacitive proximity sensing allows not only detection when an object is approaching but also distance estimation, as the scale of capacitance variation is strictly related to the proximity distance.

Capacitance-based proximity sensing

Figure 8.6 shows the process when a participant walks to a door from a 1.5 meter distance, touches the doorknob, and returns to his original spot. An accelerometer is attached to the right calf of the participant. The potential variation of the prototype at the doorknob shows the proximity of a human body, which implies that the distance could also be estimated by its variation scale. The arrows showed when the participant touched the doorknob, causing charge flow. The potential variation direction implies that the human body was sinking charge from the doorknob. Figure 8.7 shows a potential variation on the wrist when P2 walked by P1 two times with the nearest distance of 1m and 0.5m, where P1 just stood still.

Capacitance information can be used for distance estimation

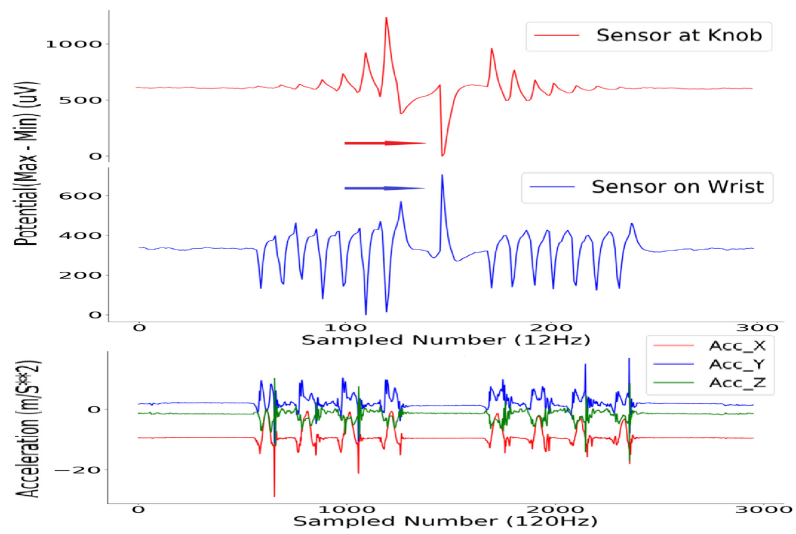


Figure 8.6: Approaching a door with prototype attached at the doorknob and on wrist

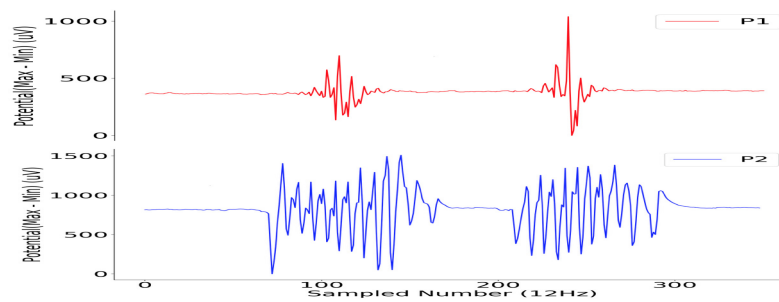


Figure 8.7: Walking by detection with prototypes attached on wrists

8.2.4 Collaborative Work Monitoring

Based on the sensing capability described above, monitoring collaborative work with a capacitance-based sensing approach is thus feasible. Event and motion from a human body can be obtained by distributing our prototypes at the working site and on moving bodies. Traditionally, multiple-person activity monitoring was fulfilled using a vision-based system [84], which can collect very detailed information about group content, but requires a high computational demand and may raise privacy concerns. We set up a simple collaborative task where two people must move a shelf from one spot to another and assemble two shelves. Figure 8.8 illustrates the working place, where prototypes are attached to two shelves, one doorknob, a toolbox, and on the wrists of two involved participants. Those objects will assist in having a better understanding of participants' actions. One common feature of those objects is their interior metal structure and paint on the surface. Each Participant was wearing an accelerometer on the calf.

*Capacitive sensing
for collaborative
work*

Figure 8.9 depicts the process of this collaborative work. Arrows label the event actions, and straight lines label the motion actions. Here are the fundamental steps: 1; for the beginning, P1 and P2 lift their legs 10 times (P11, P12, P21, P22); 2, P1 walks to P2 (P13), and they shake their hands (P1a, P2a); 3, P1 walks to shelf one (P14), walking by the door (Da), touches shelf one (P1b, S1a) and tries to lift it; 4, P1 finds it too large and not convenient for one person to carry, and calls P2 to help. P2 walks to P1 (P23), also walking by the door (Db), touches shelf one (P1c, P2b, S1b); 5, P1 and P2 lift shelf one, go to shelf two (P15, P24, S11, S21), walking by the door again (Dc, Dd); 6, They drop down shelf one, P1 manages to keep shelf one and two together (P1d, S1c, S2a); 7, P2 goes to the toolbox (P25) and takes it (P2c, Ta), then goes back to P1 (P26, T3); 8, P2 hands over the toolbox to P1 (P1e, P2d, Tb, S1d, S2b), and walks away (P27); 9, P1 uses some wire from toolbox to tie shelf one and two together, leaving the toolbox on the ground, and moves the shelves to another nearby spot (P16, S11, S22); 10, P1 walks to the toolbox (P17), takes it (P1f, Tc), then returns it back to its original place (P18, T4, Td); 11, P1 returns back to his original place (P19), walking by P2, and they lift leg for several times to finish the whole task (P110, P28). During the whole procedure, there are several signals to be declared. First, because the Toolbox is near the original spot of P2, P2's lift leg action can be perceived by the Toolbox (T1, T2, T5). Second, when P1 and P2 are coupled strictly by shelf one, their entire capacitance to ground is approximately doubled, so the Walking caused capacitance variation ratio with the total capacitance decreased (P15, P24, S11, Dc, Dd). Third, some gradually changing signals implies the approaching or leaving of a participant, like the potential variation before Ta, Tc, P2a, after P2a and during S21.

*Use-case for
collaborative work
monitoring*

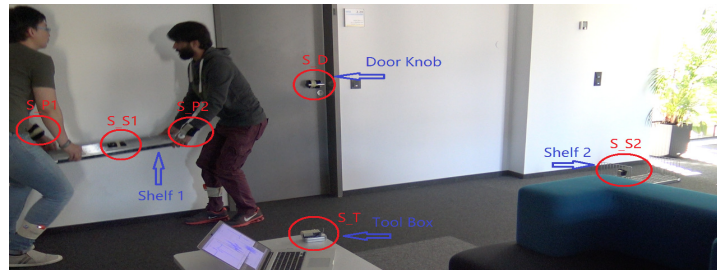


Figure 8.8: Collaborative work site

*Capacitive sensing
applications are
widespread*

The above-described potential variations from a simple collaborative work show a feasible human activity monitoring access, with exact time synchronization of all the prototypes, actions from involved participants, like touch, motion and approximate position, could be detected. This could be used in a wide range of ambient intelligence scenarios, like ambient assistive living, factory works, etc.

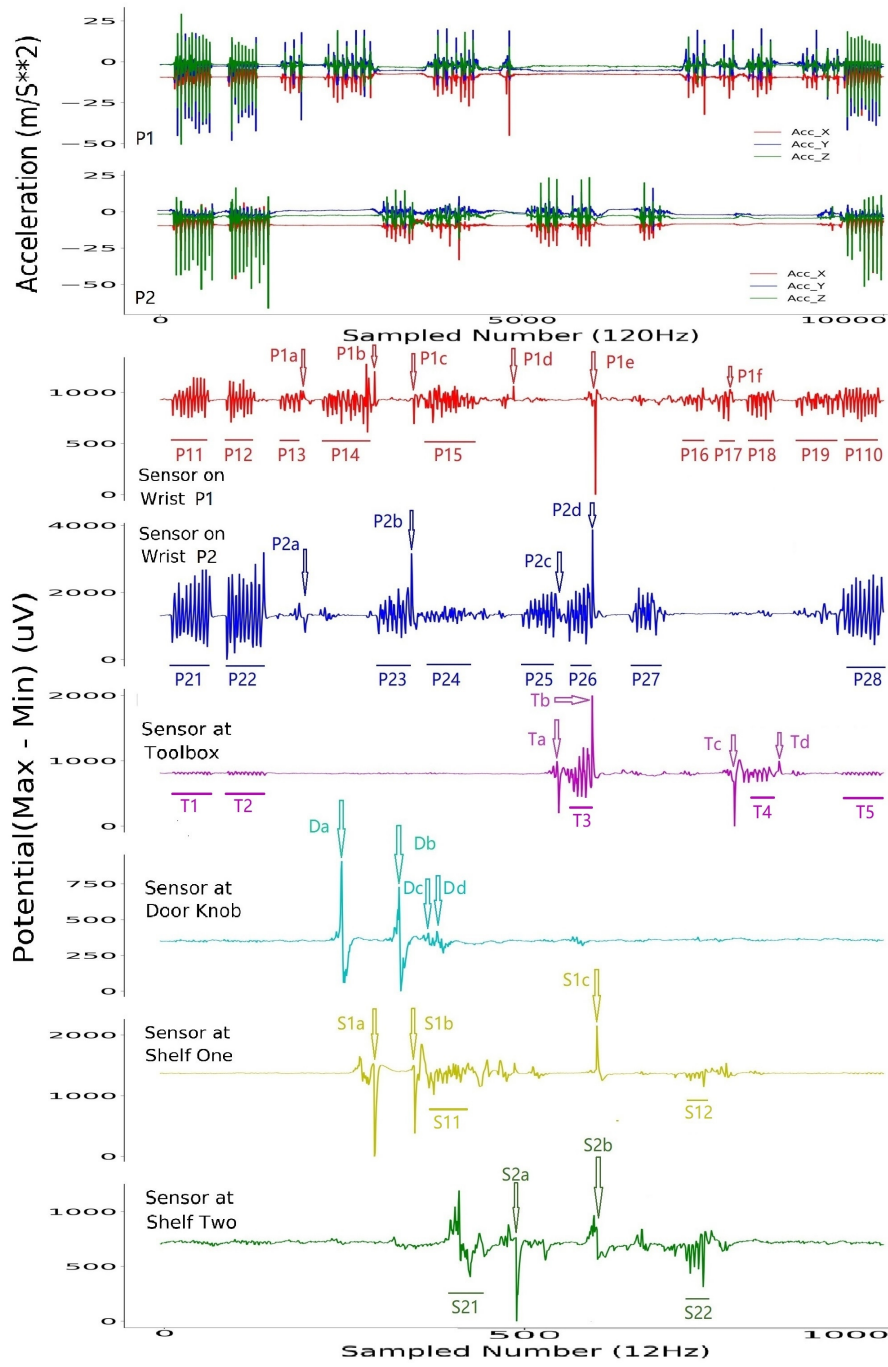


Figure 8.9: process of the collaborative work

Part V

CONCLUSION

SUMMARY

Every beginning has an end and every end has a new beginning....

Santosh Kalwar

The journey of this dissertation started with working on conducting experiments using the "gPhysics" application for Google Glass in collaboration with the Physics department at the University of Kaiserslautern. I contributed in conducting experiments, gathering the data, and processing it to drive the results with other group members. Meanwhile, Google Inc. decided to shelve the Google Glass project, but at the same time, [DNNs](#) started evolving at a tremendous pace. The impressive performance of [DNNs](#) delivering [SotA](#) results for the problems varying from simple image classification to object detection, segmentation, and activity recognition tasks tabled opportunities for research communities to explore venues and their applications in various fields of life. The growing paradigm of [DL](#)-based methods opened up a new door to looking for methods and domains to incorporate the [AI](#) in formal education to improve the learning environment; the key research question of this dissertation defined in [Chapter 1](#).

Conception of the research topic

Formal education is a vast field and is evolving day by day, incorporating technical advances, this dissertation touches on multiple aspects of cognitive ability classification and applications of wearable sensors in the context of formal education in limited scope yet in the best possible ways. [Section 9.1](#) discusses the strengths of proposed methodologies and pens down the achievements of this dissertation while answering the research questions raised in [Section 1.2](#). The beauty of the research is that it is never-ending, but the scope and research questions keep evolving with time to cater to the demands of the ever-growing world; a full stop or end of research kills the purpose and idea itself. As there is always room for improvement, similar is the case of not only this dissertation but should be for every dissertation. Limitations of the work presented in this thesis along with hints laying a roadmap of what can be done in the future are discussed in [Section 9.2](#).

Outline

9.1 ACHIEVEMENTS AND DISCUSSION

Summary of contributions

This dissertation explores the use of AI-based methods to assist the process of formal education. Traditionally, students and instructors have direct relations and connections with each other. Instructors convey the information to learners in the form of instructions mostly verbally, relying on literature such as books and, lately, digital content and videos as well. Learners absorb these instructions and further aid themselves with reading and writing activities to understand the elaborated concepts and develop knowledge skills. Besides, instructors perform another very important role in the learners' lives, monitoring the learning progress, evaluating the knowledge-developing skills, and an advisory role that might be pivotal for shaping their future lives. Monitoring, analysing, and assisting them in performing these two-way interactions are the major areas constituting the contributions of this dissertation. Part II of this dissertation discusses the approaches for content classification, handwriting classification, and correlation between the reading and writing behaviour for feedback and performance evaluation. In Part III, details of applications of wearable and smart gadgets are presented to assist the learners in performing classroom activities for enhanced experience are the major research areas touched on in this dissertation. As discussed, both AI and formal education are such vast fields that the laid contributions are just the tip of the iceberg.

Defining the structure of the document can help largely to analyse the reading activity

Content is where formal education starts with, instructors rely on it for teaching, and students need to keep learning. Educational content comes in multiple forms, such as books, essays, blogs, videos, presentations, etc. There are multiple ways to analyse the content in formal education, and the most common of them is using gaze information to track the progress during the reading activity [109]. Gaze information provides information about the eye-movement but does not consider the relevancy of reading content during the reading activity. This thesis explores another direction to analyse the content using DNNs. The area/form of content is limited to printed documents in digital format from publicly available POD datasets. We consider two main classes, i.e., figures and formulas for classification and the rest of the content on document images are considered as text. We present a novel combination of Fi-fo image representation and DNN to detect figures and formulas from document images. Fi-fo image representation utilizes traditional CV techniques to transform input image to complement the performance of DL models. The presented approach is generic, as it can detect page objects, i.e., figures and formulas, despite varying page formats and layouts. During the evaluation of Fi-fo detector, it is revealed that there are many confusions and inconsistencies in the ICDAR2017-POD dataset ground-truth. DNNs does not rely only on the quantity of the data; the data quality is also

Fi-fo detector achieves the SotA performance for POD

key in performance. We fine-tuned the dataset to remove inconsistencies and confusion and presented it as the ICDAR-2017 POD (corrected) dataset. During the course of this work, no other [POD](#) datasets were publicly available, so we also propose a new publicly available [FFD](#) dataset for figure and formula detection for cross-validation of proposed methods. The [SotA](#) results furnished in [Section 3.5](#) establish the efficacy of proposed approaches for content classification from document images. Our proposed approaches [FFD](#) or [Fi-fo](#) detector, when combined with existing methodologies such as gaze tracking, will result in deeper and more meaningful insights about the reading activity and help in a better understanding of reading behavior.

Writing is another daily practice activity in the classrooms with manifold importance; it is a basic form of communication in formal education, it capitulates the learning process, helps learners to display critical thinking skills, and provides them a foundation to express themselves. Moreover, written assignments and exams are the long-established way to gauge the outcomes and performance of the learning process. The second major contribution of this dissertation is in the online handwriting classification domain, to classify the writing into text, mathematical expression, and/or plot/graphical expressions. In [Chapter 4](#), we propose a novel feature set to classify online handwritten sequences into text, mathematical expressions, and plots/graphs. The presented feature set is evaluated using various [ML](#) and [DL](#) classifiers yielding the [SotA](#) results on unseen data. Data itself is very critical for representation learning of both [ML](#) and [DL](#) classifiers and a scarce resource during the conception of this work. We present a new public dataset for constraint-free online handwriting classification using a sensor pen and touch display collected in classroom set-up. Additionally, data collection involves reproducing simple representation learning with a lower cognitive load and producing complex representation with a higher cognitive load. This dataset is challenging as the contributors were allowed to write in freestyle - where sequences lack clear patterns- causing difficulty for classifiers. Classification of handwriting into different writing types enables the tracking and monitoring of the progress of writing activity, which can be further used to analyse the behaviour of learners during producing complex representations for performance evaluation and feedback estimation to foster the idea of personalised and need-based learning.

"Reading and writing are intricately intertwined. One is the inverse of the other: Reading is the inhale; writing is the exhale. They depend on each other, and when we find time to practice both, the students are winners," states Mary K. Tedrew in her book, "Write, Think, Learn". In formal education, reading and writing activities happen simultaneously, and it can be inferred that both have strong correlations, as one activity complements the other along with overall

Classifying handwriting into its types helps to monitor the writing activity

onTabWriter dataset will help in evaluating complex representation learning

performance evaluation using correlation of reading & writing activity

growth in learners' creative skills of learners[209, 212]. Performance evaluation and knowledge skills are predominantly assessed by designing the tasks, which combine reading and writing activities and involve the cognitive process. Chapter 5 presents a study's initial findings among Physics students to correlate their expertise based on their reading and writing behaviours during the complex representation learning. We used multiple on-body sensors to enable the teachers for deeper insights into the behaviours of the students, assist them in interacting and addressing the individual's requirements, and foster the concept of need-based learning. We also present a feature set to explore the difference between the behaviour of experts and novices when they are exposed to factual and knowledge transfer-based exercises. The proposed features lend a hand to the teachers with meaningful insights about differences in behaviours for analysis, understanding, and approach to attempt the tasks for experts in contrast to novices. Experts exhibit different behaviour from the beginning, whether analysing the problem, understanding the problem and/or formulating the solutions, taking less time, producing quick answers, skipping intermediate steps and producing abstract solutions. Initial data exploration reveals that implicit sensor information can be used as an aid for the teachers to provide needs-based individual feedback.

Implicit sensor information helps to look into knowledge skills

Applications of wearable systems in combination with AI for formal education

Applications of smart and wearable systems are the other major area that accounts for the contribution of this dissertation. Due to the advantages smart wearables offer to make them a default choice for multiple applications in healthcare, biomedicine, workplaces, education, and other fields of life [265]. Two areas are touched on in this dissertation to explore the application of wearable systems in formal education; the influence of smart glasses in cognition and a finger-worn prototype for air-writing without requiring any additional tool.

Smart glasses can positively influence the learning outcomes

An application of Google Glass, "gPhysics", is introduced to explore the efficacy of smart glasses as an experimental tool in Physics education [251]. The author of this dissertation contributes in collaboration with the authors of "gPhysics" to evaluate the influence of smart glasses while performing experimentation in terms of engagement, curiosity, execution time, and cognitive load. gPhysics application enables the students to perform automated measurements and plots them to understand acoustic principles visually. Overall results of the study reveal the positive impact and encourage the use of smart glasses in education. IMUs are now integral to almost smart and wearable devices such as smartphones, smart watches, earables, and wristbands. Chapter 6 presents an IMU-based application for AR and VR scenarios by writing with a finger in the air without requiring any reference surface. The proposed methodology is named FAirWrite, which fosters document creation by air-writing on a virtual screen without any constraints of spatio-temporal boundaries. FAirWrite also presents

Realtime air-writing application

a sensor-fusion algorithm to reconstruct the air-written trajectories on a 2-D display in real time. An enhanced GUI enables users to interact with the system and for real-time feedback. FAirWrite system is evaluated for user quality appreciation and exploring deep-learning methods to report SotA results. Both recognition methods achieved an overall accuracy of 95%. The collected dataset is made publicly available for the benefit of the research community. FAirWrite system can improve classroom interaction between teachers and students, particularly during lectures on digital displays. It also enables them to capture and record random thoughts and important points by writing in the air with their fingers only.

Part iv covers the details of projects I did during this course but are not directly contributing to the main research question of this dissertation. Chapter 7 presents a novel and generic approach dStaR for stamp detection using DNNs for the very first time. dStaR can detect unseen stamps of any shape, size, and color. The major advantage dStaR owes over previously presented approaches is that it can detect and segment the overlapping stamps from other information in scanned document images. It also outperforms the SotA approaches by successfully differentiating between stamps and logos, despite of huge similarity between the two. Chapter 8 presents an ultra-low power, capacitance-based prototype capable of sensing human touch, proximity and body activities. We demonstrated its capability with a simple collaborative task, in which the actions of participants were recorded by our prototypes worn on wrists, and assisted by attaching prototypes to other involved objects, that the participants interact with.

Research contributions not directly relating to main research questions

9.2 SCOPE AND OUTLOOK

The foundation of this dissertation lies in the conjuncture of two important aspects of formal education, i.e., applications to assist in performing cognitive activities in the learning environment and methodologies to analyse these activities for deeper insights, performance evaluation, and feedback estimation. The research conducted in the scope of this dissertation has such a broad spectrum that every research question can be interpreted and addressed in multiple ways, as there is always room for improvement in an ever-changing and evolving world. AI-based methods can be incorporated into formal education addressing various aspects and directions of research. Like every research work, the work presented in this dissertation has its limitation beyond the scope of contributions presented in Section 1.3. Different methodologies presented in this dissertation can be deployed for corresponding real-world applications in their defined scope. How-

Limitations

ever, a larger framework is needed to incorporate all these entities to function together to achieve the goals in a larger perspective.

This thesis touches on multiple aspects of incorporating technological interventions in formal education to function together to improve the overall learning experience. At the same time, every chapter of the thesis is self-contained in itself and has been and could be further worked on as an independent topic to fill their own dissertation. It is possible to dig deeper into the topics of every chapter for further elaboration of questions in hand; we will discuss the limitation and prospects of every problem addressed in individual topics along with what is next for the dissertation topic itself.

*What can be done for
POD & possible
applications*

Chapter 3 present the methods of POD for content analysis using DNNs. The proposed methodologies classify page objects such as figures and formulas from document images, and the rest of the content is considered as text. DNNs have shown headlong progress in the last decade, opening new ways to address the problem. A recent and new Publay dataset [284] could be an important avenue for pushing the SotA in this direction. Using attention-based region proposal networks in the current setup can generate improved and better region proposals. Lately, introduced weakly-supervised and self-supervised learning techniques can lead to achieving the goal of generic systems for POD problem using ample amounts of online data with minimal effort. Considering the scope of this thesis, proposed approaches can be put in real-world applications in combination with gaze-tracking methods to relate the PoIs with the content for relevancy to look deeper into the reading behaviours.

*prospects of online
handwriting
classification*

Chapter 4 registers considerable contribution for online handwriting classification. A future direction in this regard could be comparing the presented features in this work with the features learned by DNNs, which will be very interesting and highly encouraged. We also recommend exploring ways to embed context information within the classifier, which could significantly improve the classification results. In Section 4.4, the proposed dataset can be extended further with multi-labeled sequences, i.e., current classification labels along with a copy or creative writing labels to differentiate between copying text, creative writing and attempting solutions. Our presented dataset can also be used to analyse the writing behaviour and classroom performance of students. Online handwriting classification systems could broaden to incorporate gaze-tracking while writing activities for cognitive measurements such as stress level monitoring, expertise, comfort, etc. These measures can help instructors to analyse the process of the writing activity to address personal strengths and weaknesses, a substantial aspect of adaptive and need-based learning models.

Chapter 5 presents a use-case of using an on-body sensor set-up to evaluate cognitive abilities during representation learning. The encouraging results demand further research to develop mental models using on-body sensors for adaptive teaching and learning systems. A future direction in this regard is to explore AI-based methods to visualise cognitive ability classification systems using sensor information. We recommend incorporating the handwriting classifiers presented in Chapter 4 and content classifiers Chapter 3 for deeper insights into respective cognitive activities. These amplification will help achieve a comprehensive framework to comprehensively record, monitor, analyse, and evaluate learners' cognitive activities and abilities. In this area, AI-based system can help instructors. It is also encouraged to involve domain experts in the process of assimilating the demands and requirements of real-world systems. It is also encouraged to explore the venues to employ and test the efficacy of proposed systems in classrooms.

On-body sensors paradigm & AI for learning analytic framework

Part III presents the contributions of applications of wearable systems to assist learners in performing cognitive activities. *gPhysics* application offers a huge stage to investigate the VR, AR, and MR venues for incorporation in formal education. These applications will help achieve an enhanced learning experience, better engagement, improved cognition, and boost motivation during the learning activities. As future work for improving FAirWrite, the functionality of the proposed system should be extended to distinguish between writing and not-writing activity using DNNs. It is also recommended to incorporate the context knowledge to the FAirWrite system for words and sentences for improved performance. There are multiple grounds to explore the use-cases of FAirWrite systems such as education, offices, construction sites etc. In education, it can be used to explore its potential and efficacy for different sign languages and its applications for differently-abled students.

FAirWrite system presents opportunity for exploration for different use-cases

Finally, I want to conclude this dissertation chapter by summarizing my personal views. I'll be concluding an important phase of my life and my journey to higher education in Germany. I will be moving on to a new journey, and a new phase of my life might be with a new role, but it does not mean that all these efforts and learning experiences will remain behind. These experiences gathered over the past decade have positively influenced my life, helping me grow mentally and socially and into a mature person who can see things in broader scope and context. I will try to practice and spread them while passing through different phases of my life ahead.

A personal note on this journey

BIBLIOGRAPHY

- [1] M. Agarwal, A. Mondal, and C. Jawahar. "Cdec-net: Composite deformable cascade network for table detection in document images." In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021, pp. 9491–9498 (cit. on p. 14).
- [2] A. R. Ahmad, M. Khalia, C. Viard-Gaudin, and E. Poisson. "Online handwriting recognition using support vector machine." In: *2004 IEEE Region 10 Conference TENCON 2004*. Vol. A. Nov. 2004, 311–314 Vol. 1 (cit. on p. 58).
- [3] S. Ahmed, F. Shafait, M. Liwicki, and A. Dengel. "A Generic Method for Stamp Segmentation Using Part-Based Features." In: *12th International Conference on Document Analysis and Recognition*. Aug. 2013, pp. 708–712 (cit. on p. 28).
- [4] S. Ahmed, M. I. Malik, M. Liwicki, and A. Dengel. "Signature segmentation from document images." In: *ICFHR*. 2012, pp. 425–429 (cit. on p. 128).
- [5] S. Ahmed, F. Shafait, M. Liwicki, and A. Dengel. "A generic method for stamp segmentation using part-based features." In: *ICDAR*. 2013, pp. 708–712 (cit. on pp. 121, 123, 129–131).
- [6] S. Ainsworth. "DeFT: A conceptual framework for considering learning with multiple representations." In: *Learning and Instruction* 16.3 (2006), pp. 183–198 (cit. on pp. 86, 136).
- [7] M. S. Alam, K.-C. Kwon, M. A. Alam, M. Y. Abbass, S. M. Imtiaz, and N. Kim. "Trajectory-Based Air-Writing Recognition Using Deep Neural Network and Depth Sensor." In: *Sensors* 20.2 (2020) (cit. on pp. 100, 102).
- [8] D. Alamargot, D. Chesnet, C. Dansac, and C. Ros. "Eye and Pen: A new device for studying reading during writing." In: *Behavior Research Methods* 38.2 (May 2006), pp. 287–299 (cit. on p. 90).
- [9] C. Amma, M. Georgi, and T. Schultz. "Airwriting: Hands-Free Mobile Text Input by Spotting and Continuous Recognition of 3d-Space Handwriting with Inertial Sensors." In: *2012 16th International Symposium on Wearable Computers*. 2012, pp. 52–59 (cit. on pp. 100–103).
- [10] A. Arshad, S. Khan, A. Z. Alam, A. F. Ismail, and R. Tasnim. "Capacitive proximity floor sensing system for elderly tracking and fall detection." In: *Smart Instrumentation, Measurement and Application (ICSIMA), 2017 IEEE 4th International Conference on*. IEEE. 2017, pp. 1–5 (cit. on p. 142).

- [11] A. Arshad, S. Khan, A. Z. Alam, K. A. Kadir, R. Tasnim, and A. F. Ismail. "A capacitive proximity sensing scheme for human motion detection." In: *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*. IEEE, 2017, pp. 1–5 (cit. on p. 142).
- [12] O. B. Livingston. "A Handwriting and Pen-Printing Classification System for Identifying Law Violators." In: *Journal of Criminal Law and Criminology* 49 (1959) (cit. on pp. 16, 56, 58).
- [13] A. Baddeley. "Working memory and language: an overview." In: *Journal of Communication Disorders* 36.3 (2003). ASHA 2002, pp. 189–208 (cit. on p. 86).
- [14] A. D. Baddeley, N. Thomson, and M. Buchanan. "Word length and the structure of short-term memory." In: *Journal of Verbal Learning and Verbal Behavior* 14.6 (1975), pp. 575–589 (cit. on p. 86).
- [15] C. Bahlmann, B. Haasdonk, and H. Burkhardt. "Online handwriting recognition with support vector machines - a kernel approach." In: *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*. Aug. 2002, pp. 49–54 (cit. on pp. 56, 58).
- [16] C. Bahlmann. "Directional features in online handwriting recognition." In: *Pattern Recognition* 39.1 (2006), pp. 115–125 (cit. on pp. 4, 61).
- [17] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba. "Network Dissection: Quantifying Interpretability of Deep Visual Representations." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017, pp. 3319–3327 (cit. on p. 37).
- [18] S. Becker, S. Mukhametov, P. Pawels, and J. Kuhn. "Using mobile eye tracking to capture joint visual attention in collaborative experimentation." In: Oct. 2021 (cit. on pp. 83, 84).
- [19] Y. Bhalgat, M. Kulkarni, S. Karande, and S. Lodha. "Stamp processing with exemplar features." In: *arXiv preprint arXiv:1609.05001* (2016) (cit. on pp. 121, 123, 125, 127).
- [20] H. Bi, C. Xu, C. Shi, G. Liu, Y. Li, H. Zhang, and J. Qu. "SRRV: A Novel Document Object Detector Based on Spatial-Related Relation and Vision." In: *IEEE Transactions on Multimedia* (2022), pp. 1–1 (cit. on p. 14).
- [21] S. Bian, V. F. Rey, P. Hevesi, and P. Lukowicz. "Passive Capacitive based Approach for Full Body Gym Workout Recognition and Counting." In: *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2019, pp. 1–10 (cit. on p. 142).

- [22] S. Bian, V. F. Rey, J. Younas, and P. Lukowicz. "Wrist-Worn Capacitive Sensor for Activity and Physical Collaboration Recognition." In: *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 2019, pp. 261–266 (cit. on p. 135).
- [23] M. Billinghurst and A. Duenser. "Augmented Reality in the Classroom." In: *Computer* 45.7 (2012), pp. 56–63 (cit. on pp. 5, 20).
- [24] C. M. Bishop, M. Svensen, and G. E. Hinton. "Distinguishing Text from Graphics in On-Line Handwritten Ink." In: *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition. IWFHR '04*. USA: IEEE Computer Society, 2004, pp. 142–147 (cit. on p. 17).
- [25] P. Black and D. Wiliam. "Assessment and Classroom Learning." In: *Assessment in Education: Principles, Policy & Practice* 5.1 (1998), pp. 7–74 (cit. on p. 4).
- [26] V. Bouletreau, N. Vincent, R. Sabourin, and H. Emptoz. "Synthetic parameters for handwriting classification." In: *Proceedings of the Fourth International Conference on Document Analysis and Recognition*. Vol. 1. Aug. 1997, 102–106 vol.1 (cit. on pp. 16, 58).
- [27] M. Bower and D. Sturman. "What are the educational affordances of wearable technologies?" In: *Computers Education* 88 (2015), pp. 343–353 (cit. on p. 136).
- [28] A. Braun, S. Frank, M. Majewski, and X. Wang. "CapSeat: capacitive proximity sensing for automotive activity recognition." In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM. 2015, pp. 225–232 (cit. on p. 142).
- [29] L. Breiman. "Bagging Predictors." In: *Machine Learning* 24.2 (Aug. 1996), pp. 123–140 (cit. on p. 65).
- [30] L. Breiman. "Random Forests." In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32 (cit. on p. 65).
- [31] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman. "Linear time Euclidean distance transform algorithms." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.5 (1995), pp. 529–533 (cit. on p. 13).
- [32] T. Breuel. "The OCRopus open source OCR system." In: vol. 6815. Jan. 2008, p. 68150 (cit. on p. 28).
- [33] S. Brückner, O. Zlatkin-Troitschanskaia, S. Kuechemann, P. Klein, and J. Kuhn. "Changes in Students' Understanding of and Visual Attention on Digitally Represented Graphs Across Two Domains in Higher Education: A Postreplication Study." In: *Frontiers in Psychology* 11 (Aug. 2020) (cit. on p. 91).

- [34] S. Bukhari, M. Al Azawi, F. Shafait, and T. Breuel. "Document Image Segmentation Using Discriminative Learning over Connected Components." In: *9th International Workshop on Document Analysis Systems*. Boston, Massachusetts, USA, 2010, pp. 183–190 (cit. on p. 34).
- [35] L. Bürgi, R. Pfeiffer, M. Mücklich, P. Metzler, M. Kiy, and C. Winnewisser. "Optical proximity and touch sensors based on monolithically integrated polymer photodiodes and polymer LEDs." In: *Organic Electronics* 7.2 (2006), pp. 114–120 (cit. on p. 143).
- [36] G. Castle. "Contact charging between insulators." In: *Journal of Electrostatics* 40 (1997), pp. 13–20 (cit. on p. 140).
- [37] H. J. Chang, G. Garcia-Hernando, D. Tang, and T.-K. Kim. "Spatio-Temporal Hough Forest for efficient detection–localisation–recognition of fingerwriting in egocentric camera." In: *Computer Vision and Image Understanding* 148 (2016). Special issue on Assistive Computer Vision and Robotics - "Assistive Solutions for Mobility, Communication and HMI", pp. 87–96 (cit. on pp. 100, 102).
- [38] S. S. M. Chanijani, P. Klein, M. Al-Naser, S. S. Bukhari, J. Kuhn, and A. Dengel. "A Study on Representational Competence in Physics Using Mobile Eye Tracking Systems." In: *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: ACM, 2016, pp. 1029–1032 (cit. on pp. 84, 87).
- [39] K. Chen, C.-L. Liu, M. Seuret, M. Liwicki, J. Hennebert, and R. Ingold. "Page segmentation for historical document images based on superpixel classification with unsupervised feature learning." In: *DAS*. 2016, pp. 299–304 (cit. on p. 122).
- [40] J. Cheng, O. Amft, and P. Lukowicz. "Active Capacitive Sensing: Exploring a New Wearable Sensing Modality for Activity Recognition." In: *Pervasive Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 319–336 (cit. on pp. 140, 141).
- [41] P. Chiu, F. Chen, and L. Denoue. "Picture Detection in Document Page Images." In: *10th ACM Symposium on Document Engineering*. Manchester, United Kingdom, 2010, pp. 211–214 (cit. on pp. 12, 30).
- [42] J. Choi, B. Yoon, C. Jung, and W. Woo. "ARClassNote: Augmented Reality Based Remote Education Solution with Tag Recognition and Shared Hand-Written Note." In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 2017, pp. 303–309 (cit. on p. 5).

- [43] S. L. Chu, B. Garcia, and N. Rani. *Research on Wearable Technologies for Learning: A Systematic Review*. Jan. 2022 (cit. on p. 20).
- [44] C. Clark and S. Divvala. "PDFFigures 2.0: Mining Figures from Research Papers." In: *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*. JCDL '16. Newark, New Jersey, USA, 2016, pp. 143–152 (cit. on p. 32).
- [45] G. Cohn, S. Gupta, T.-J. Lee, D. Morris, J. R. Smith, M. S. Reynolds, D. S. Tan, and S. N. Patel. "An ultra-low-power human body motion sensor using static electric field sensing." In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM. 2012, pp. 99–102 (cit. on pp. 140, 142).
- [46] G. Cohn, D. Morris, S. Patel, and D. Tan. "Humantenna: using the body as an antenna for real-time whole-body interaction." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2012, pp. 1901–1910 (cit. on p. 142).
- [47] S. D'Mello, A. Olney, C. Williams, and P. Hays. "Gaze tutor: A gaze-reactive intelligent tutoring system." In: *International Journal of Human-Computer Studies* 70.5 (2012), pp. 377–398 (cit. on p. 84).
- [48] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. "Deformable Convolutional Networks." In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 764–773 (cit. on pp. 37, 38).
- [49] J. Dai, Y. Li, K. He, and J. Sun. "R-FCN: Object Detection via Region-based Fully Convolutional Networks." In: *Advances in Neural Information Processing Systems* 29. 2016, pp. 379–387 (cit. on p. 38).
- [50] A. K. Das, S. P. Chowdhury, S. Mandal, and B. Chanda. "Automated Segmentation of Math-Zones from Document Images." In: *12th International Conference on Document Analysis and Recognition*. Vol. 3. 2003, p. 755 (cit. on pp. 12, 29).
- [51] A. Dash, A. Sahu, R. Shringi, J. Gamboa, M. Z. Afzal, M. I. Malik, A. Dengel, and S. Ahmed. "AirScript - Creating Documents in Air." In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 01. 2017, pp. 908–913 (cit. on pp. 100–103, 114).
- [52] A. Delaye and C. Liu. "Graphics Extraction from Heterogeneous Online Documents with Hierarchical Random Fields." In: *2013 12th International Conference on Document Analysis and Recognition*. Aug. 2013, pp. 1007–1011 (cit. on p. 59).
- [53] A. Delaye and K. Lee. "A flexible framework for online document segmentation by pairwise stroke distance learning." In: *Pattern Recognition* 48.4 (2015), pp. 1197–1210 (cit. on pp. 56, 58).

- [54] A. Delaye and C.-L. Liu. "Text/Non-text Classification in Online Handwritten Documents with Conditional Random Fields." In: *Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 514–521 (cit. on p. 18).
- [55] A. Delaye and C.-L. Liu. "Contextual text/non-text stroke classification in online handwritten notes with conditional random fields." In: *Pattern Recognition* 47.3 (2014). Handwriting Recognition and other PR Applications, pp. 959–968 (cit. on p. 18).
- [56] A. Delaye and C.-L. Liu. "Multi-class segmentation of free-form online documents with tree conditional random fields." In: *International Journal on Document Analysis and Recognition (IJ DAR)* 17.4 (Dec. 2014), pp. 313–329 (cit. on p. 59).
- [57] S. Dey, J. Mukherjee, and S. Sural. "Stamp and logo detection from document images by finding outliers." In: *NCVPRIPG*. 2015, pp. 1–4 (cit. on pp. 121, 123).
- [58] M. Diaz, M. A. Ferrer, D. Impedovo, M. I. Malik, G. Pirlo, and R. Plamondon. "A Perspective Analysis of Handwritten Signature Technology." In: *ACM Computing Surveys (CSUR)* 51.6 (2019), p. 117 (cit. on pp. 4, 57).
- [59] L. Dinehart. "Handwriting in early childhood education: Current research and future implications." In: *Journal of Early Childhood Literacy* 15 (Mar. 2014) (cit. on p. 85).
- [60] Z. Dong, U. C. Wejinya, and W. J. Li. "An Optical-Tracking Calibration Method for MEMS-Based Digital Writing Instrument." In: *IEEE Sensors Journal* 10.10 (2010), pp. 1543–1551 (cit. on p. 102).
- [61] J. Dunn and T. Sweeney. "Writing and iPads in the early years: Perspectives from within the classroom." In: *British Journal of Educational Technology* 49.5 (2018), pp. 859–869 (cit. on pp. 56, 57).
- [62] H. L. Duy, H. M. Nghia, B. T. Vinh, and P. D. Hung. "An Efficient Approach to Stamp Verification." In: *Smart Trends in Computing and Communications*. Singapore: Springer Nature Singapore, 2023, pp. 781–789 (cit. on pp. 121, 124).
- [63] N. Elmqaddem. "Augmented Reality and Virtual Reality in Education. Myth or Reality?" In: *International Journal of Emerging Technologies in Learning (iJET)* 14.03 (Feb. 2019), pp. 234–242 (cit. on pp. 5, 20).
- [64] D. C. Engelbart. "Augmenting human intellect: A conceptual framework." In: *Menlo Park, CA* (1962), p. 21 (cit. on p. 85).

- [65] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. "The pascal visual object classes challenge: A retrospective." In: *IJCV* 111.1 (2015), pp. 98–136 (cit. on p. 125).
- [66] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno. "2D Euclidean distance transform algorithms: A comparative survey." In: *ACM Computing Surveys (CSUR)* 40.1 (2008), pp. 1–44 (cit. on p. 13).
- [67] J. Fang, L. Gao, K. Bai, R. Qiu, X. Tao, and Z. Tang. "A Table Detection Method for Multipage PDF Documents via Visual Separators and Tabular Structures." In: *2011 International Conference on Document Analysis and Recognition*. 2011, pp. 779–783 (cit. on p. 12).
- [68] R. J. Fateman, T. Tokuyasu, B. P. Berman, and N. Mitchell. "Optical Character Recognition and Parsing of Typeset Mathematics1." In: *Journal of Visual Communication and Image Representation* 7.1 (1996), pp. 2–15 (cit. on p. 13).
- [69] T. Ficker. "Electrification of human body by walking." In: *Journal of Electrostatics* 64.1 (2006), pp. 10–16 (cit. on p. 140).
- [70] P. Forczmański and A. Markiewicz. "Low-Level image features for stamps detection and classification." In: *CORES 2013*. 2013, pp. 383–392 (cit. on pp. 121, 123).
- [71] J. H. Friedman. "Stochastic gradient boosting." In: *Computational Statistics & Data Analysis* 38.4 (2002). Nonlinear Methods and Data Mining, pp. 367–378 (cit. on p. 66).
- [72] L. Gao, X. Yi, Z. Jiang, L. Hao, and Z. Tang. "ICDAR2017 Competition on Page Object Detection." In: *14th International Conference on Document Analysis and Recognition*. Vol. 01. 2017, pp. 1417–1422 (cit. on pp. 14, 31, 32, 40, 42, 45, 47).
- [73] L. Gao, X. Yi, Y. Liao, Z. Jiang, Z. Yan, and Z. Tang. "A Deep Learning-Based Formula Detection Method for PDF Documents." In: *14th Int. Conf. on Document Analysis and Recognition*. Vol. 01. 2017, pp. 553–558 (cit. on pp. 25, 31, 51).
- [74] P. Garg, J. Santhosh, A. Dengel, and S. Ishimaru. "Stress Detection by Machine Learning and Wearable Sensors." In: *26th International Conference on Intelligent User Interfaces - Companion. IUI '21 Companion*. College Station, TX, USA: Association for Computing Machinery, 2021, pp. 43–45 (cit. on p. 5).
- [75] B. M. Garlapati and S. R. Chalamala. "A System for Handwritten and Printed Text Classification." In: *2017 UKSim-AMSS 19th International Conference on Computer Modelling Simulation (UKSim)*. 2017, pp. 50–54 (cit. on p. 17).

- [76] B. Gatos, D. Danatsas, I. Pratikakis, and S. J. Perantonis. "Automatic table detection in document images." In: *International Conference on Pattern Recognition and Image Analysis*. Springer. 2005, pp. 609–618 (cit. on p. 12).
- [77] P. Geurts, D. Ernst, and L. Wehenkel. "Extremely Randomized Trees." In: *Mach. Learn.* 63.1 (Apr. 2006), pp. 3–42 (cit. on p. 65).
- [78] S. J. Gideon, A. Kandulna, A. A. Kujur, A. Diana, and K. Raimond. "Handwritten signature forgery detection using convolutional neural networks." In: *Procedia computer science* 143 (2018), pp. 978–987 (cit. on p. 4).
- [79] A. Gilani, S. R. Qasim, M. I. Malik, and F. Shafait. "Table Detection Using Deep Learning." In: *14th Int. Conf. on Document Analysis and Recognition*. 2017, pp. 771–776 (cit. on pp. 13, 29, 31, 34, 35).
- [80] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He. *Detectron*. 2018 (cit. on p. 36).
- [81] M. Göbel, T. Hassan, E. Oro, and G. Orsi. "A Methodology for Evaluating Algorithms for Table Understanding in PDF Documents." In: *Proceedings of the 2012 ACM Symposium on Document Engineering*. DocEng '12. Paris, France: Association for Computing Machinery, 2012, pp. 45–48 (cit. on p. 14).
- [82] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber. "A novel connectionist system for unconstrained handwriting recognition." In: *PAMI* (2009), pp. 855–68 (cit. on p. 122).
- [83] E. Griechisch, M. I. Malik, and M. Liwicki. "On-line Signature Verification Based on Kolmogorov-Smirnov Distribution Distance." In: *14th Int. Conf. on Frontiers in Handwriting Recognition*. Sept. 2014, pp. 738–742 (cit. on p. 56).
- [84] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. "Using adaptive tracking to classify and monitor activities in a site." In: *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*. IEEE. 1998, pp. 22–29 (cit. on p. 145).
- [85] T. Grosse-Puppendahl, E. Berlin, and M. Borazio. "Enhancing accelerometer-based activity recognition with capacitive proximity sensing." In: *International Joint Conference on Ambient Intelligence*. Springer. 2012, pp. 17–32 (cit. on pp. 140, 142).
- [86] T. Grosse-Puppendahl, X. Dellangnol, C. Hatzfeld, B. Fu, M. Kupnik, A. Kuijper, M. R. Hastall, J. Scott, and M. Gruteser. "Platypus: Indoor localization and identification through sensing of electric potential changes in human bodies." In: *Proceedings of the 14th Annual International Conference on Mobile Sys-*

- tems, Applications, and Services*. ACM. 2016, pp. 17–30 (cit. on p. 141).
- [87] A. Gruenerbl, G. Bahle, and P. Lukowicz. “Detecting spontaneous collaboration in dynamic group activities from noisy individual activity data.” In: *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE. 2017, pp. 279–284 (cit. on p. 141).
- [88] A. Grygoriev, I. Degtyarenko, I. Deriuga, S. Polotskyi, V. Melnyk, D. Zakharchuk, and O. Radyvonenko. “HCRNN: A Novel Architecture for Fast Online Handwritten Stroke Classification.” In: *Document Analysis and Recognition – ICDAR 2021*. Cham: Springer International Publishing, 2021, pp. 193–208 (cit. on p. 19).
- [89] Ç. Gülçehre, K. Cho, R. Pascanu, and Y. Bengio. “Learned-Norm Pooling for Deep Feedforward and Recurrent Neural Networks.” In: *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I*. Vol. 8724. Lecture Notes in Computer Science. Springer, 2014, pp. 530–546 (cit. on p. 66).
- [90] R. D. Gurchiek, R. S. McGinnis, A. R. Needle, J. M. McBride, and H. van Werkhoven. “The use of a single inertial sensor to estimate 3-dimensional ground reaction force during accelerative running tasks.” In: *Journal of Biomechanics* 61 (2017), pp. 263–268 (cit. on p. 100).
- [91] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet. “UNIPEN project of on-line data exchange and recognizer benchmarks.” In: *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*. Vol. 2. 1994, 29–33 vol.2 (cit. on pp. 16, 58).
- [92] R. M. Haralick and I. T. Phillips. “Recursive X-Y cut using bounding boxes of connected components.” In: *3rd International Conference on Document Analysis and Recognition*. Vol. 2. Aug. 1995, 952–955 vol.2 (cit. on pp. 12, 29, 30, 34).
- [93] C. Harland, T. Clark, and R. Prance. “Electric potential probes new directions in the remote sensing of the human body.” In: *Measurement Science and technology* 13.2 (2001), p. 163 (cit. on p. 141).
- [94] C. Harrison, J. Schwarz, and S. E. Hudson. “TapSense: enhancing finger interaction on touch surfaces.” In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM. 2011, pp. 627–636 (cit. on p. 142).

- [95] K. A. Hashmi, A. Pagani, M. Liwicki, D. Stricker, and M. Z. Afzal. "Cascade Network with Deformable Composite Backbone for Formula Detection in Scanned Document Images." In: *Applied Sciences* 11.16 (2021) (cit. on p. 52).
- [96] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. June 2016, pp. 770–778 (cit. on pp. 35, 37).
- [97] K. He, G. Gkioxari, P. Dollár, and R. Girshick. "Mask r-cnn." In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969 (cit. on pp. 14, 29, 35, 36).
- [98] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778 (cit. on p. 14).
- [99] J. Hermida. "The Importance of Teaching Academic Reading Skills In First-Year University Courses." In: *RCRN: Pedagogy (Topic)* 3 (June 2009) (cit. on p. 3).
- [100] Y. Hirayama. "A block segmentation method for document images with complicated column structures." In: *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR'93)*. IEEE. 1993, pp. 91–94 (cit. on p. 12).
- [101] K. Hochberg, S. Becker, M. Louis, P. Klein, and J. Kuhn. "Using Smartphones as Experimental Tools—a Follow-up: Cognitive Effects by Video Analysis and Reduction of Cognitive Load by Multiple Representations." In: *Journal of Science Education and Technology* 29 (Apr. 2020) (cit. on pp. 83, 91).
- [102] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory." In: 9.8 (Nov. 1997), pp. 1735–1780 (cit. on p. 66).
- [103] A. Hoffmann, A. Poepfel, A. Schierl, and W. Reif. "Environment-aware proximity detection with capacitive sensors for human-robot-interaction." In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2016, pp. 145–150 (cit. on p. 141).
- [104] C. C. W. Hulls. "Using a Tablet PC for Classroom Instruction." In: *Proceedings Frontiers in Education 35th Annual Conference*. Oct. 2005, T2G–T2G (cit. on pp. 56, 57).
- [105] A. I. Ianov, H. Kawamoto, and Y. Sankai. "Development of a capacitive coupling electrode for bioelectrical signal measurements and assistive device use." In: *Complex Medical Engineering (CME), 2012 ICME International Conference on*. IEEE. 2012, pp. 593–598 (cit. on p. 141).

- [106] S. Inatani, T. Van Phan, and M. Nakagawa. "Comparison of MRF and CRF for Text/Non-text Classification in Japanese Ink Documents." In: *2014 14th International Conference on Frontiers in Handwriting Recognition*. 2014, pp. 684–689 (cit. on p. 59).
- [107] E. Indermühle, V. Frinken, and H. Bunke. "Mode Detection in Online Handwritten Documents Using BLSTM Neural Networks." In: *2012 International Conference on Frontiers in Handwriting Recognition*. Sept. 2012, pp. 302–307 (cit. on pp. 18, 56, 59).
- [108] E. Indermühle, M. Liwicki, and H. Bunke. "IAMonDo-database: An Online Handwritten Document Database with Non-uniform Contents." In: *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. DAS '10. Boston, Massachusetts, USA: ACM, 2010, pp. 97–104 (cit. on pp. 18, 57, 59).
- [109] S. Ishimaru. "Meta-Augmented Human: From Physical to Cognitive Towards Affective State Recognition." doctoralthesis. Technische Universität Kaiserslautern, 2020, pp. XIII, 146 (cit. on pp. 4, 152).
- [110] S. Ishimaru, S. S. Bukhari, C. Heisel, N. Großmann, P. Klein, J. Kuhn, and A. Dengel. "Augmented Learning on Anticipating Textbooks with Eye Tracking." In: *Positive Learning in the Age of Information: A Blessing or a Curse?* Wiesbaden: Springer Fachmedien Wiesbaden, 2018, pp. 387–398 (cit. on pp. 82–84).
- [111] S. Ishimaru, S. S. Bukhari, C. Heisel, J. Kuhn, and A. Dengel. "Towards an Intelligent Textbook: Eye Gaze Based Attention Extraction on Materials for Learning and Instruction in Physics." In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. UbiComp '16 Adjunct. ACM. 2016, pp. 1041–1045 (cit. on pp. 5, 84, 87).
- [112] S. Ishimaru, K. Hoshika, K. Kunze, K. Kise, and A. Dengel. "Towards Reading Trackers in the Wild: Detecting Reading Activities by EOG Glasses and Deep Neural Networks." In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. UbiComp '17. Maui, Hawaii: Association for Computing Machinery, 2017, pp. 704–711 (cit. on p. 3).
- [113] S. Ishimaru, T. Maruichi, K. Kise, and A. Dengel. "Gaze-Based Self-Confidence Estimation on Multiple-Choice Questions and Its Feedback." In: *AsianCHI '20*. Honolulu, HI, USA: Association for Computing Machinery, 2020, p. 8 (cit. on p. 5).

- [114] S. Ishimaru, K. Watanabe, N. Großmann, C. Heisel, P. Klein, Y. Arakawa, J. Kuhn, and A. Dengel. “HyperMind Builder: Pervasive User Interface to Create Intelligent Interactive Documents.” In: *Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. UbiComp '18 Adjunct. ACM. 2018, pp. 357–360 (cit. on pp. 82, 84).
- [115] K. Iwatsuki, T. Sagara, T. Hara, and A. Aizawa. “Detecting In-line Mathematical Expressions in Scientific Documents.” In: *2017 ACM Symposium on Document Engineering*. Valletta, Malta, 2017, pp. 141–144 (cit. on pp. 12, 13, 29, 30).
- [116] K. Jain, A. Namboodiri, and J. Subrahmonia. “Structure in on-line documents.” In: *Proceedings of Sixth International Conference on Document Analysis and Recognition*. 2001, pp. 844–848 (cit. on p. 17).
- [117] H. Jarodzka and S. Brand-Gruwel. “Tracking the reading eye: towards a model of real-world reading.” In: *Journal of Computer Assisted Learning* 33.3 (2017), pp. 193–201 (cit. on p. 3).
- [118] A. Kacem, A. Belaïd, and M. Ben Ahmed. “Automatic Extraction of Printed Mathematical Formulas Using Fuzzy Logic and Propagation of Context.” In: *International Journal on Document Analysis and Recognition* 4 (Dec. 2001), pp. 97–108 (cit. on p. 12).
- [119] S. Kapp et al. “Augmenting Kirchhoff’s laws: Using augmented reality and smartglasses to enhance conceptual electrical experiments for high school students.” In: *The Physics Teacher* 57 (Jan. 2019), pp. 52–53 (cit. on pp. 20, 84).
- [120] M. Kassner, W. Patera, and A. Bulling. “Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction.” In: *UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Apr. 2014) (cit. on p. 87).
- [121] I. Kavasidis, C. Pino, S. Palazzo, F. Rundo, D. Giordano, P. Messina, and C. Spampinato. “A Saliency-Based Convolutional Neural Network for Table and Chart Detection in Digitized Documents.” In: (2019), pp. 292–302 (cit. on pp. 29, 30).
- [122] A. Kholmatov and B. Yanikoglu. “Identity authentication using improved online signature verification method.” In: *Pattern recognition letters* 26.15 (2005), pp. 2400–2408 (cit. on pp. 4, 56).
- [123] T. Kieninger and A. Dengel. “Table recognition and labeling using intrinsic layout features.” In: *International conference on advances in pattern recognition*. Springer. 1999, pp. 307–316 (cit. on p. 12).

- [124] T. Kieninger and A. Dengel. "The T-Recs Table Recognition and Analysis System." In: *Document Analysis Systems: Theory and Practice*. 1999, pp. 255–270 (cit. on p. 28).
- [125] P. Klein, A. Lichtenberger, S. Kuechemann, S. Becker, M. Kekule, J. Viiri, C. Baadte, A. Vaterlaus, and J. Kuhn. "Visual attention while solving the test of understanding graphs in kinematics: An eye-tracking analysis." In: *European Journal of Physics* 41 (Dec. 2019) (cit. on pp. 83–85, 87).
- [126] H. M. Knight, P. R. Gajendragadkar, and A. Bokhari. "Wearable technology: using Google Glass as a teaching tool." In: *Case Reports* 2015 (2015), bcr2014208768 (cit. on p. 136).
- [127] K. Koile and D. Singer. "Development of a Tablet-PC-based System to Increase Instructor-Student Classroom Interactions and Student Learning." In: (Jan. 2006) (cit. on pp. 56, 57).
- [128] M. Krishnamoorthy, G. Nagy, S. Seth, and M. Viswanathan. "Syntactic segmentation and labeling of digitized pages from technical journals." In: *PAMI* 15.7 (1993), pp. 737–747 (cit. on p. 123).
- [129] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." In: *NIPS*. 2012, pp. 1097–1105 (cit. on pp. 121, 122).
- [130] S. Kuechemann, P. Klein, S. Becker, N. Kumari, and J. Kuhn. "Classification of Students' Conceptual Understanding in STEM Education using Their Visual Attention Distributions: A Comparison of Three Machine-Learning Approaches." In: Jan. 2020, pp. 36–46 (cit. on pp. 82–84).
- [131] S. Kuechemann, P. Klein, H. Fouckhardt, S. Gröber, and J. Kuhn. "Students' understanding of non-inertial frames of reference." In: *Physical Review Physics Education Research* 16 (Mar. 2020) (cit. on pp. 83, 85).
- [132] J. Kuhn, P. Lukowicz, M. Hirth, A. Poxrucker, J. Weppner, and J. Younas. "gPhysics—Using Smart Glasses for Head-Centered, Context-Aware Learning in Physics Experiments." In: *IEEE Transactions on Learning Technologies* 9.4 (2016), pp. 304–317 (cit. on pp. 82, 84, 135).
- [133] J. Kuhn, P. Lukowicz, M. Hirth, and J. Weppner. "gPhysics - Using Google Glass as Experimental Tool for Wearable-Technology Enhanced Learning in Physics." In: *Workshop Proceedings of the 11th International Conference on Intelligent Environments, Prague, Czech Republic, July 15-17, 2015*. Vol. 19. Ambient Intelligence and Smart Environments. IOS Press, 2015, pp. 212–219 (cit. on p. 137).

- [134] P. Kumar, R. Saini, S. K. Behera, D. P. Dogra, and P. P. Roy. "Real-time recognition of sign language gestures and air-writing using leap motion." In: *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. 2017, pp. 157–160 (cit. on p. 100).
- [135] K. Kumara, R. Agrawal, and C. Bhattacharyya. "A large margin approach for writer independent online handwriting classification." In: *Pattern Recognition Letters* 29.7 (2008), pp. 933–937 (cit. on p. 17).
- [136] K. Kunze, K. Masai, M. Inami, Ö. Sacakli, M. Liwicki, A. Dengel, S. Ishimaru, and K. Kise. "Quantifying Reading Habits: Counting How Many Words You Read." In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 87–96 (cit. on p. 3).
- [137] M. C. Leue, T. Jung, et al. "Google glass augmented reality: Generic learning outcomes for art galleries." In: *Information and communication technologies in tourism 2015*. Springer, 2015, pp. 463–476 (cit. on p. 136).
- [138] K. Li, C. Wington, C. Tensmeyer, H. Zhao, N. Barmpalios, V. I. Morariu, V. Manjunatha, T. Sun, and Y. Fu. "Cross-Domain Document Object Detection: Benchmark Suite and Method." In: *CoRR abs/2003.13197* (2020) (cit. on p. 32).
- [139] P. Li, J. Gu, J. Kuen, V. I. Morariu, H. Zhao, R. Jain, V. Manjunatha, and H. Liu. "SelfDoc: Self-Supervised Document Representation Learning." In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), pp. 5648–5656 (cit. on p. 33).
- [140] X. H. Li, F. Yin, and C.-L. Liu. "Page Object Detection from PDF Document Images by Deep Structured Prediction and Supervised Clustering." In: *24th International Conference on Pattern Recognition* (2018), pp. 3627–3632 (cit. on pp. 15, 29, 31, 32, 45, 47, 48, 50–52).
- [141] C. Lima Sanches, O. Augereau, and K. Kise. "Estimation of reading subjective understanding based on eye gaze analysis." In: *PLOS ONE* 13.10 (Oct. 2018), pp. 1–16 (cit. on p. 87).
- [142] C.-L. Lin, Y.-M. Chang, C.-C. Hung, C.-D. Tu, and C.-Y. Chuang. "Position estimation and smooth tracking with a fuzzy-logic-based adaptive strong tracking Kalman filter for capacitive touch panels." In: *IEEE Transactions on Industrial Electronics* 62.8 (2015), pp. 5097–5108 (cit. on p. 141).

- [143] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie. "Feature Pyramid Networks for Object Detection." In: *IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 936–944 (cit. on pp. 14, 38).
- [144] B. G. Lindeque, B. A. Ponce, M. E. Menendez, L. O. Oladeji, C. T. Fryberger, and P. K. Dantuluri. "Emerging technology in surgical education: combining real-time augmented reality and wearable computing devices." In: *Orthopedics* 37.11 (2014), pp. 751–757 (cit. on p. 136).
- [145] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. "SSD: Single Shot MultiBox Detector." In: *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 21–37 (cit. on p. 14).
- [146] M. Liwicki and H. Bunke. "HMM-Based On-Line Recognition of Handwritten Whiteboard Notes." In: *Tenth International Workshop on Frontiers in Handwriting Recognition*. Université de Rennes 1. La Baule (France): Suvisoft, Oct. 2006 (cit. on pp. 4, 56, 57, 64, 70–74, 76, 77).
- [147] M. Liwicki and H. Bunke. "Feature Selection for HMM and BLSTM Based Handwriting Recognition of Whiteboard Notes." In: *IJPRAI* 23 (Aug. 2009), pp. 907–923 (cit. on pp. 56, 61).
- [148] M. Liwicki, A. Schlapbach, H. Bunke, S. Bengio, J. Mariéthoz, and J. Richiardi. *Writer Identification for Smart Meeting Room Systems*. Tech. rep. Published in Seventh IAPR Workshop on Document Analysis Systems, DAS, 2006. 2005 (cit. on pp. 56, 61, 77).
- [149] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." In: *ICVPR*. 2015, pp. 3431–3440 (cit. on p. 126).
- [150] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan. "Estimation of IMU and MARG orientation using a gradient descent algorithm." In: *2011 IEEE International Conference on Rehabilitation Robotics*. 2011, pp. 1–7 (cit. on p. 106).
- [151] P. S. Mali, P. Kukkadapu, M. Mahdavi, and R. Zanibbi. "ScanSSD: Scanning Single Shot Detector for Mathematical Formulas in PDF Document Images." In: *ArXiv abs/2003.08005* (2020) (cit. on p. 14).
- [152] M. I. Malik and M. Liwicki. "From Terminology to Evaluation: Performance Assessment of Automatic Signature Verification Systems." In: *Int. Conf. on Frontiers in Handwriting Recognition*. Sept. 2012, pp. 613–618 (cit. on p. 57).

- [153] M. I. Malik, M. Liwicki, A. Dengel, S. Uchida, and V. Frinken. "Automatic Signature Stability Analysis and Verification using Local Features." In: *14th International Conference on Frontiers in Handwriting Recognition (ICFHR-2014)*. IEEE. 2014, pp. 621–626 (cit. on p. 4).
- [154] M. I. Malik, S. Ahmed, F. Shafait, A. S. Mian, C. Nansen, A. Dengel, and M. Liwicki. "Hyper-spectral analysis for automatic signature extraction." In: *BCIGS*. 2015 (cit. on p. 128).
- [155] M. I. Malik, M. Liwicki, and A. Dengel. "Part-based automatic system in comparison to human experts for forensic signature verification." In: *2013 12th International Conference on Document Analysis and Recognition*. IEEE. 2013, pp. 872–876 (cit. on pp. 4, 57).
- [156] S. Mandal, S. R. M. Prasanna, and S. Sundaram. "GMM posterior features for improving online handwriting recognition." In: *Expert Systems with Applications* 97 (2018), pp. 421–433 (cit. on p. 56).
- [157] U.-V. Marti and H. Bunke. "The IAM-database: an English sentence database for offline handwriting recognition." In: *International Journal on Document Analysis and Recognition* 5.1 (2002), pp. 39–46 (cit. on p. 18).
- [158] T. Maruichi and K. Kise. "Writing Behavior as an Estimator of Self-Confidence on English Spelling Questions." In: *AsianCHI '20*. Honolulu, HI, USA: Association for Computing Machinery, 2020, p. 4 (cit. on p. 90).
- [159] N. Matsushita and J. Rekimoto. "HoloWall: designing a finger, hand, body, and object sensitive wall." In: *Proceedings of the 10th annual ACM symposium on User interface software and technology*. ACM. 1997, pp. 209–210 (cit. on p. 142).
- [160] T. Matsushita and M. Nakagawa. "A Database of On-Line Handwritten Mixed Objects Named "Kondate"." In: *2014 14th International Conference on Frontiers in Handwriting Recognition*. 2014, pp. 369–374 (cit. on p. 59).
- [161] J. McKenzie and D. Darnell. "The eyeMagic Book. A Report into Augmented Reality Storytelling in the Context of a children's workshop 2003." In: *New Zealand Centre for Children's Literature and Christchurch College of Education* (2004) (cit. on p. 20).
- [162] B. Micenkov and J. van Beusekom. "Stamp detection in color document images." In: *ICDAR*. 2011, pp. 1125–1129 (cit. on pp. 121, 122, 128–131).
- [163] B. Micenková, J. van Beusekom, and F. Shafait. "Stamp verification for automated document authentication." In: *Computational Forensics*. 2015, pp. 117–129 (cit. on p. 123).

- [164] D. Miller and T. Dousay. "Implementing Augmented Reality in the Classroom." In: *Issues and Trends in Educational Technology* 3.2 (Dec. 2015) (cit. on p. 5).
- [165] S. Misra, J. Singha, and R. Laskar. "Vision-based hand gesture recognition of alphabets, numbers, arithmetic operators and ASCII characters in order to develop a virtual text-entry interface system." In: *Neural Computing and Applications* (2017). cited By 5, pp. 1–19 (cit. on pp. 100, 102).
- [166] Y. Moon, R. S. McGinnis, K. Seagers, R. W. Motl, N. Sheth, J. A. Wright Jr., R. Ghaffari, and J. J. Sosnoff. "Monitoring gait in multiple sclerosis with novel wearable motion sensors." In: *PLOS ONE* 12.2 (Feb. 2017), pp. 1–19 (cit. on p. 100).
- [167] J. Moore. "Handwriting Classification." In: *The Police Journal* 18.1 (1945), pp. 39–61 (cit. on pp. 16, 56).
- [168] S. Mukherjee, S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy. "Fingertip detection and tracking for recognition of air-writing in videos." In: *Expert Systems with Applications* 136 (2019), pp. 217–229 (cit. on pp. 100, 102).
- [169] L. P. Nattamai Sekar, A. Santos, and O. Beltramello. "IMU Drift Reduction for Augmented Reality Applications." In: *Augmented and Virtual Reality*. Cham: Springer International Publishing, 2015, pp. 188–196 (cit. on p. 100).
- [170] C. Nguyen and F. Liu. "Gaze-based Notetaking for Learning from Lecture Videos." In: May 2016, pp. 2093–2097 (cit. on pp. 5, 84).
- [171] V. T. Nguyen, K. Jung, and T. Dang. "Creating Virtual Reality and Augmented Reality Development in Classroom: Is it a Hype?" In: *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 2019, pp. 212–2125 (cit. on pp. 5, 20, 21).
- [172] P. O'Shea, R. Mitchell, C. Johnston, and C. Dede. "Lessons Learned about Designing Augmented Realities." In: *IJGCMS* 1 (Jan. 2009), pp. 1–15 (cit. on p. 20).
- [173] W. Ohyama, M. Suzuki, and S. Uchida. "Detecting Mathematical Expressions in Scientific Document Images Using a U-Net Trained on a Diverse Dataset." In: *IEEE Access* 7 (2019), pp. 144030–144042 (cit. on p. 14).
- [174] K. Oka, Y. Sato, and H. Koike. "Real-time fingertip tracking and gesture recognition." In: *IEEE Computer Graphics and Applications* 22.6 (2002), pp. 64–71 (cit. on p. 100).

- [175] F. Ott, D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler. "Joint Classification and Trajectory Regression of Online Handwriting using a Multi-Task Learning Approach." In: *IEEE Winter Conf. on Applications of Computer Vision (WACV)*. Waikoloa, HI, Jan. 2022 (cit. on p. 60).
- [176] S. Otte, D. Krechel, M. Liwicki, and A. Dengel. "Local Feature Based Online Mode Detection with Recurrent Neural Networks." In: *2012 International Conference on Frontiers in Handwriting Recognition*. Sept. 2012, pp. 533–537 (cit. on p. 61).
- [177] S. J. Pan and Q. Yang. "A survey on transfer learning." In: *IEEE Transactions on knowledge and data engineering* 22.10 (2010), pp. 1345–1359 (cit. on p. 125).
- [178] T. Pan, C. Kuo, H. Liu, and M. Hu. "Handwriting Trajectory Reconstruction Using Low-Cost IMU." In: *IEEE Transactions on Emerging Topics in Computational Intelligence* 3.3 (2019), pp. 261–270 (cit. on pp. 102, 103, 114).
- [179] T. Y. Pan, C. H. Kuo, and M. C. Hu. "A noise reduction method for IMU and its application on handwriting trajectory reconstruction." In: *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. July 2016, pp. 1–6 (cit. on pp. 100–103, 114).
- [180] M. Panwar, S. Ram Dyuthi, K. Chandra Prakash, D. Biswas, A. Acharyya, K. Maharatna, A. Gautam, and G. R. Naik. "CNN based approach for activity recognition using a wrist-worn accelerometer." In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2017, pp. 2438–2441 (cit. on p. 140).
- [181] J. Pastor-Pellicer, M. Z. Afzal, M. Liwicki, and M. J. Castro-Bleda. "Complete System for Text Line Extraction Using Convolutional Neural Networks and Watershed Transform." In: *DAS*. 2016, pp. 30–35 (cit. on p. 122).
- [182] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In: *Advances in Neural Information Processing Systems* 32. Curran Associates, Inc., 2019, pp. 8024–8035 (cit. on p. 68).
- [183] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 68).
- [184] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. "Scikit-learn: Machine learning in Python." In: *Journal of Machine Learning Research* (2011), pp. 2825–2830 (cit. on p. 128).

- [185] T. V. Phan and M. Nakagawa. "Text/Non-text Classification in Online Handwritten Documents with Recurrent Neural Networks." In: *2014 14th International Conference on Frontiers in Handwriting Recognition*. Sept. 2014, pp. 23–28 (cit. on p. 59).
- [186] B. H. Phong, T. M. Hoang, and T. Le. "A new method for displayed mathematical expression detection based on FFT and SVM." In: *4th NAFOSTED Conference on Information and Computer Science*. 2017, pp. 90–95 (cit. on pp. 12, 29, 30).
- [187] B. H. Phong, L. T. Dat, N. T. Yen, T. M. Hoang, and T.-L. Le. "A deep learning based system for mathematical expression detection and recognition in document images." In: *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*. 2020, pp. 85–90 (cit. on p. 14).
- [188] R. Plomin. "Genetics and General Cognitive Ability." In: *Nature* 402 (Jan. 1999), pp. C25–9 (cit. on p. 82).
- [189] K. Plunkett. "A Simple and Practical Method for Incorporating Augmented Reality into the Classroom and Laboratory." In: *Journal of Chemical Education* 96 (Sept. 2019) (cit. on pp. 5, 21).
- [190] S. Polotskyi, I. Deriuga, T. Ignatova, V. Melnyk, and H. Azarov. "Improving Online Handwriting Text/Non-text Classification Accuracy Under Condition of Stroke Context Absence." In: *Advances in Computational Intelligence*. Cham: Springer International Publishing, 2019, pp. 210–221 (cit. on p. 19).
- [191] C. J. Portier and M. S. Wolfe. "Assessment of health effects from exposure to power-line frequency electric and magnetic fields." In: *NIH publication* 98 (1998), p. 3981 (cit. on p. 140).
- [192] A. Pouryazdan, R. J. Prance, H. Prance, and D. Roggen. "Wearable electric potential sensing: a new modality sensing hair touch and restless leg movement." In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM. 2016, pp. 846–850 (cit. on pp. 140, 141).
- [193] PyQT. "PyQt Reference Guide." In: (2012) (cit. on p. 108).
- [194] I. Ragnemalm. "The Euclidean distance transform in arbitrary dimensions." In: *Pattern Recognition Letters* 14.11 (1993), pp. 883–888 (cit. on p. 13).
- [195] M. Rajab and L. George. "Stamps extraction using local adaptive k-means and ISODATA algorithms." In: *Indonesian Journal of Electrical Engineering and Computer Science* 21 (Jan. 2021) (cit. on pp. 121, 124).
- [196] K. Rayner, K. H. Chace, T. J. Slattery, and J. Ashby. "Eye Movements as Reflections of Comprehension Processes in Reading." In: *Scientific Studies of Reading* 10.3 (2006), pp. 241–255 (cit. on pp. 87, 91).

- [197] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." In: *Advances in Neural Information Processing Systems*. 2015, pp. 91–99 (cit. on pp. 13, 14, 29, 35, 38).
- [198] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015, pp. 234–241 (cit. on p. 14).
- [199] I. Rosenberg and K. Perlin. "The UnMousePad: an interpolating multi-touch force-sensing input pad." In: *ACM Transactions on Graphics (TOG)* 28.3 (2009), p. 65 (cit. on pp. 141, 142).
- [200] S. Rossignol, D. Willems, A. Neumann, and L. Vuurpijl. "Mode detection and incremental recognition." In: *Ninth International Workshop on Frontiers in Handwriting Recognition*. 2004, pp. 597–602 (cit. on p. 17).
- [201] N. Roussel, H. Evans, and H. Hansen. "Proximity as an interface for video communication." In: *IEEE MultiMedia* 11.3 (2004), pp. 12–16 (cit. on pp. 140, 143).
- [202] V. Ruf, S. Kuechemann, J. Kuhn, and P. Klein. "Comparison of Written and Spoken Instruction to Foster Coordination between Diagram and Equation in Undergraduate Physics Education." In: *Human Behavior and Emerging Technologies 2022* (Mar. 2022), pp. 1–13 (cit. on pp. 83, 84).
- [203] P. M. Russell, M. Mallin, S. T. Youngquist, J. Cotton, N. Aboul-Hosn, and M. Dawson. "First "Glass" Education: Telementored Cardiac Ultrasonography Using Google Glass-A Pilot Study." In: *Academic Emergency Medicine* 21.11 (2014), pp. 1297–1299 (cit. on p. 136).
- [204] R. Saha, A. Mondal, and C. V. Jawahar. "Graphical Object Detection in Document Images." In: *15th Int. Conf. on Document Analysis and Recognition*. 2019 (cit. on pp. 14, 31).
- [205] H. E. S. Said, K. D. Baker, and T. N. Tan. "Personal identification based on handwriting." In: *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*. Vol. 2. Aug. 1998, 1761–1764 vol.2 (cit. on p. 56).
- [206] Y. T. Schelske, A. Dengel, F. Strauß, M. Liwicki, C. Schoelzel, and M. Weber. "MCS for Online Mode Detection: Evaluation on Pen-Enabled Multi-touch Interfaces." In: *2011 International Conference on Document Analysis and Recognition(ICDAR)*. Vol. 00. Sept. 2011, pp. 957–961 (cit. on pp. 56, 59).
- [207] J. Schneider, D. Börner, P. Rosmalen, and M. Specht. "Augmenting the Senses: A Review on Sensor-Based Learning Support." In: *Sensors* 15 (Feb. 2015), pp. 4097–4133 (cit. on p. 85).

- [208] L. Schomaker, G. Abbink, and S. Selen. "Writer and writing-style classification in the recognition of online handwriting." In: *IEE European Workshop on Handwriting Analysis and Recognition: A European Perspective*. July 1994, pp. 1/1–1/4 (cit. on pp. 16, 58).
- [209] R. Schoonen. "Are reading and writing building on the same skills? The relationship between reading and writing in L1 and EFL." In: *Reading and Writing* 32.3 (2019), pp. 511–535 (cit. on p. 153).
- [210] M. Schrapel, M.-L. Stadler, and M. Rohs. "Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–11 (cit. on p. 102).
- [211] S. Schreiber, S. Agne, I. Wolf, A. Dengel, and S. Ahmed. "Deep-DeSRT: Deep Learning for Detection and Structure Recognition of Tables in Document Images." In: *14th International Conference on Document Analysis and Recognition*. 2017, pp. 1162–1167 (cit. on pp. 13, 29, 31, 35).
- [212] R. I. Segundo Marcos, V. López Fernández, M. T. Daza González, and J. Phillips-Silver. "Promoting children's creative thinking through reading and writing in a cooperative learning classroom." In: *Thinking Skills and Creativity* 36 (2020), p. 100663 (cit. on p. 153).
- [213] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Roure-Alcobé. "Gender Classification by Means of Online Uppercase Handwriting: A Text-Dependent Allographic Approach." English. In: *Cognitive Computation* 8.1 (2016), pp. 15–29 (cit. on p. 56).
- [214] M. Seuret, M. Alberti, R. Ingold, and M. Liwicki. "PCA-Initialized Deep Neural Networks Applied To Document Image Analysis." In: *arXiv preprint arXiv:1702.00177* (2017) (cit. on p. 122).
- [215] F. Shafait and R. Smith. "Table detection in heterogeneous documents." In: *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. 2010, pp. 65–72 (cit. on p. 12).
- [216] A. Shahab, F. Shafait, T. Kieninger, and A. Dengel. "An Open Approach towards the Benchmarking of Table Structure Recognition Systems." In: *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. DAS '10. Boston, Massachusetts, USA: Association for Computing Machinery, 2010, pp. 113–120 (cit. on p. 14).

- [217] A. Sharma and S. Sundaram. "A Novel Online Signature Verification System Based on GMM Features in a DTW Framework." In: *IEEE Transactions on Information Forensics and Security* 12.3 (Mar. 2017), pp. 705–718 (cit. on p. 57).
- [218] N. Siddiqui and R. H. M. Chan. "Multimodal hand gesture recognition using single IMU and acoustic measurements at wrist." In: *PLOS ONE* 15.1 (Jan. 2020), pp. 1–12 (cit. on p. 100).
- [219] S. A. Siddiqui, M. I. Malik, S. Agne, A. Dengel, and S. Ahmed. "DeCNT: Deep Deformable CNN for Table Detection." In: *IEEE Access* 6 (2018), pp. 74151–74161 (cit. on pp. 28, 29, 31, 39).
- [220] N. Siegel, Z. Horvitz, R. Levin, S. Divvala, and A. Farhadi. "FigureSeer: Parsing Result-Figures in Research Papers." In: *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 664–680 (cit. on p. 32).
- [221] N. Siegel, N. Lourie, R. Power, and W. Ammar. "Extracting Scientific Figures with Distantly Supervised Neural Networks." In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. JCDL '18. Fort Worth, Texas, USA, 2018, pp. 223–232 (cit. on p. 32).
- [222] A. C. e Silva. "Learning rich hidden markov models in document analysis: Table location." In: *2009 10th International Conference on Document Analysis and Recognition*. IEEE. 2009, pp. 843–847 (cit. on p. 12).
- [223] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition." In: *arXiv preprint arXiv:1409.1556* (2014) (cit. on pp. 121, 125).
- [224] G. Singh, A. Nelson, R. Robucci, C. Patel, and N. Banerjee. "Inviz: Low-power personalized gesture recognition using wearable textile capacitive sensor arrays." In: *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*. IEEE. 2015, pp. 198–206 (cit. on p. 141).
- [225] R. Smith. "An Overview of the Tesseract OCR Engine." In: *9th Int. Conf. on Document Analysis and Recognition*. Vol. 2. 2007, pp. 629–633 (cit. on pp. 28, 29).
- [226] T. L. Smith. "Six Basic Factors in Handwriting Classification." In: *Journal of Criminal Law and Criminology* (1954) (cit. on pp. 16, 56, 58).
- [227] C. Soto and S. Yoo. "Visual Detection with Context for Document Layout Analysis." In: *EMNLP*. 2019 (cit. on p. 15).
- [228] T. Starner. "Project Glass: An Extension of the Self." In: *IEEE Pervasive Computing* 12.2 (Apr. 2013), pp. 14–16 (cit. on p. 136).

- [229] J. Stern, B. Terris, H. Mamin, and D. Rugar. "Deposition and imaging of localized charge on insulator surfaces using a force microscope." In: *Applied physics letters* 53.26 (1988), pp. 2717–2719 (cit. on p. 140).
- [230] C. Sun, A. Shrivastava, S. Singh, and A. Gupta. "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era." In: *ICCV*. 2017, pp. 843–852 (cit. on p. 39).
- [231] W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, and M. Blumenstein. "Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline." In: *ArXiv* abs/2002.10061 (2020) (cit. on p. 107).
- [232] W. Tao, Z.-H. Lai, M. C. Leu, and Z. Yin. "Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks." In: *Procedia Manufacturing* 26 (2018). 46th SME North American Manufacturing Research Conference, NAMRC 46, Texas, USA, pp. 1159–1166 (cit. on p. 100).
- [233] J. Tchalenko. "Segmentation and accuracy in copying and drawing: Experts and beginners." In: *Vision Research* 49.8 (2009), pp. 791–800 (cit. on p. 90).
- [234] M. Thees, K. Altmeyer, S. Kapp, E. Rexigel, F. Beil, P. Klein, S. Malone, R. Brünken, and J. Kuhn. "Augmented Reality for Presenting Real-Time Data During Students' Laboratory Work: Comparing a Head-Mounted Display With a Separate Display." In: *Frontiers in Psychology* 13 (Mar. 2022), p. 804742 (cit. on p. 84).
- [235] A. Toselli, A. Juan, and E. Vidal. "Spontaneous handwriting recognition and classification." In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. Vol. 1. 2004, 433–436 Vol.1 (cit. on pp. 18, 84).
- [236] N. Toyozumi, J. Takahashi, and G. Lopez. "Trajectory reconstruction algorithm based on sensor fusion between IMU and strain gauge for stand-alone digital pen." In: *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. 2016, pp. 1906–1911 (cit. on p. 100).
- [237] H. Tu, X. Ren, and S. Zhai. "A Comparative Evaluation of Finger and Pen Stroke Gestures." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 1287–1296 (cit. on pp. 100, 115).
- [238] J. Tully, C. Dameff, S. Kaib, and M. Moffitt. "Recording medical students' encounters with standardized patients using Google Glass: providing end-of-life clinical education." In: *Academic medicine* 90.3 (2015), pp. 314–316 (cit. on p. 136).

- [239] S. Tupaj, Z. Shi, C. H. Chang, and H. Alam. "Extracting tabular information from text files." In: *EECS Department, Tufts University, Medford, USA 1* (1996) (cit. on p. 12).
- [240] J. S. Twyman and W. L. Heward. "How to improve student learning in every classroom now." In: *International Journal of Educational Research* 87 (2018), pp. 78–90 (cit. on pp. 56, 57).
- [241] K. Ueda. "Extraction of signature and seal imprint from bankchecks by using color information." In: *ICDAR*. Vol. 2. 1995, pp. 665–668 (cit. on pp. 121, 122).
- [242] P. Vadiraja, A. Dengel, and S. Ishimaru. "Text Summary Augmentation for Intelligent Reading Assistant." In: *Proceedings of the 2nd Augmented Humans International Conference*. AHs '21. ACM. 2021, pp. 319–321 (cit. on p. 85).
- [243] M. Valtonen, T. Vuorela, L. Kaila, and J. Vanhala. "Capacitive indoor positioning and contact sensing for activity recognition in smart homes." In: *Journal of Ambient Intelligence and Smart Environments* 4.4 (2012), pp. 305–334 (cit. on p. 141).
- [244] C. Viard-Gaudin, P. Lallican, S. Knerr, and P. Binter. "The IRESTE On/Off (IRONOFF) dual handwriting database." In: *Proceedings of the Fifth International Conference on Document Analysis and Recognition*. *ICDAR '99 (Cat. No. PR00318)*. 1999, pp. 455–458 (cit. on p. 59).
- [245] N. D. Vo, K. Nguyen, T. V. Nguyen, and K. Nguyen. "Ensemble of Deep Object Detectors for Page Object Detection." In: *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*. *IMCOM '18*. Langkawi, Malaysia: Association for Computing Machinery, 2018 (cit. on pp. 32, 35, 52).
- [246] N. Wade and B. Tatler. "Did Javal measure eye movements during reading?" In: *Journal of Eye Movement Research* 2 (May 2009), pp. 1–7 (cit. on p. 82).
- [247] J. Wang and F. Chuang. "An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition." In: *IEEE Transactions on Industrial Electronics* 59.7 (2012), pp. 2998–3007 (cit. on p. 102).
- [248] J.-S. Wang, Y.-L. Hsu, and J.-N. Liu. "An Inertial-Measurement-Unit-Based Pen With a Trajectory Reconstruction Algorithm and Its Applications." In: *IEEE Transactions on Industrial Electronics* 57.10 (2010), pp. 3508–3521 (cit. on p. 100).
- [249] Q.-F. Wang, E. Cambria, C.-L. Liu, and A. Hussain. "Common Sense Knowledge for Handwritten Chinese Text Recognition." In: *Cognitive Computation* 5.2 (June 2013), pp. 234–242 (cit. on p. 4).

- [250] Y. Wangt, I. T. Phillipst, and R. Haralick. "Automatic table ground truth generation and a background-analysis-based table structure extraction method." In: *Proceedings of Sixth International Conference on Document Analysis and Recognition*. IEEE. 2001, pp. 528–532 (cit. on p. 12).
- [251] J. Weppner, P. Lukowicz, M. Hirth, and J. Kuhn. "gPhysics—Using google glass as experimental tool for wearable-technology enhanced learning in physics." In: *Workshop Proceedings of the 11th International Conference on Intelligent Environments* (2015), p. 212 (cit. on p. 154).
- [252] C. Wickens. "Virtual reality and education." In: *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*. 1992, 842–847 vol.1 (cit. on p. 5).
- [253] D. Willems and L. Vuurpijl. "A Bayesian Network Approach to Mode Detection for Interactive Maps." In: *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*. Vol. 2. 2007, pp. 869–873 (cit. on p. 17).
- [254] D. Willems, S. Rossignol, and L. Vuurpijl. "Mode detection in on-line pen drawing and handwriting recognition." In: *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE. 2005, pp. 31–35 (cit. on p. 17).
- [255] R. Wimmer, M. Kranz, S. Boring, and A. Schmidt. "A Capacitive Sensing Toolkit for Pervasive Activity Detection and Recognition." In: *Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'07)*. 2007, pp. 171–180 (cit. on p. 141).
- [256] R. Wimmer, M. Kranz, S. Boring, and A. Schmidt. "CapTable and CapShelf - Unobtrusive Activity Recognition Using Networked Capacitive Sensors." In: *2007 Fourth International Conference on Networked Sensing Systems*. 2007, pp. 85–88 (cit. on p. 141).
- [257] F. Wittmann, O. Lamercy, R. R. Gonzenbach, M. A. van Raai, R. Höver, J. Held, M. L. Starkey, A. Curt, A. Luft, and R. Gassert. "Assessment-driven arm therapy at home using an IMU-based virtual reality system." In: *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2015, pp. 707–712 (cit. on p. 100).
- [258] S. Won, W. Melek, and F. Golnaraghi. "Position and orientation estimation using Kalman filtering and particle filtering with one IMU and one position sensor." In: *2008 34th Annual Conference of IEEE Industrial Electronics*. 2008, pp. 3006–3010.

- [259] S. Won, W. Melek, and F. Golnaraghi. "Position and orientation estimation using Kalman filtering and particle filtering with one IMU and one position sensor." In: *2008 34th Annual Conference of IEEE Industrial Electronics*. 2008, pp. 3006–3010 (cit. on p. 100).
- [260] S. Wörner, J. Kuhn, and K. Scheiter. "The Best of Two Worlds: A Systematic Review on Combining Real and Virtual Experiments in Science Education." In: *Review of Educational Research* (Apr. 2022), p. 003465432210794 (cit. on p. 84).
- [261] T. S. Wu, C. Dameff, and J. Tully. "Integrating Google Glass into simulation-based training: experiences and future directions." In: *Journal of Biomedical Graphics and Computing* 4 (2014), p. 49 (cit. on p. 136).
- [262] L. Xia and J. Aggarwal. "Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 2834–2841 (cit. on p. 140).
- [263] C. Xu, C. Shi, H. Bi, C. Liu, Y. Yuan, H. Guo, and Y. Chen. "A Page Object Detection Method Based on Mask R-CNN." In: *IEEE Access* 9 (2021), pp. 143448–143457 (cit. on p. 14).
- [264] C. Xu, P. H. Pathak, and P. Mohapatra. "Finger-Writing with Smartwatch: A Case for Finger and Hand Gesture Recognition Using Smartwatch." In: *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. HotMobile '15. Santa Fe, New Mexico, USA: Association for Computing Machinery, 2015, pp. 9–14 (cit. on pp. 101, 102).
- [265] Y. Xue. "A review on intelligent wearables: Uses and risks." In: *Human Behavior and Emerging Technologies* 1.4 (2019), pp. 287–294 (cit. on p. 154).
- [266] J. Yang, H. Pan, W. Zhou, and R. Huang. "Evaluation of smart classroom from the perspective of infusing technology into pedagogy." In: *Smart Learning Environments* (Sept. 2018) (cit. on pp. 56, 57).
- [267] J.-Y. Ye, Y.-M. Zhang, Q. Yang, and C.-L. Liu. "Contextual Stroke Classification in Online Handwritten Documents with Graph Attention Networks." In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 993–998 (cit. on pp. 18, 84).
- [268] X. Yi, L. Gao, Y. Liao, X. Zhang, R. Liu, and Z. Jiang. "CNN Based Page Object Detection in Document Images." In: *14th Int. Conf. on Document Analysis and Recognition*. Vol. 01. 2017, pp. 230–235 (cit. on pp. 29, 30).

- [269] J. Yosinski, J. Clune, A. M. Nguyen, T. J. Fuchs, and H. Lipson. "Understanding Neural Networks Through Deep Visualization." In: *CoRR abs/1506.06579* (2015) (cit. on p. 37).
- [270] J. Younas, M. Z. Afzal, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "D-StaR: A Generic Method for Stamp Segmentation from Document Images." In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 01. 2017, pp. 248–253 (cit. on p. 120).
- [271] J. Younas, S. Fritsch, G. Pirkl, S. Ahmed, M. I. Malik, F. Shafait, and P. Lukowicz. "What Am I Writing: Classification of On-Line Handwritten Sequences." In: *Intelligent Environments (Workshops)*. Vol. 23. Ambient Intelligence and Smart Environments. IOS Press, 2018, pp. 417–426 (cit. on pp. 55, 56, 100).
- [272] J. Younas, M. I. Malik, S. Ahmed, F. Shafait, and P. Lukowicz. "Sense the pen: Classification of online handwritten sequences (text, mathematical expression, plot/graph)." In: *Expert Systems with Applications* 172 (2021), p. 114588 (cit. on p. 55).
- [273] J. Younas, H. Margarito, S. Bian, and P. Lukowicz. "Finger Air Writing - Movement Reconstruction with Low-Cost IMU Sensor." In: *MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. MobiQuitous '20. Darmstadt, Germany: Association for Computing Machinery, 2020, pp. 69–75 (cit. on p. 99).
- [274] J. Younas, H. Margarito, and P. Lukowicz. "FAirWrite - Movement Reconstruction and Recognition Using a Low-cost IMU." In: *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. 2022, pp. 298–303 (cit. on pp. 99, 110).
- [275] J. Younas, S. T. R. Rizvi, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "FFD: Figure and Formula Detection from Document Images." In: *2019 Digital Image Computing: Techniques and Applications (DICTA)*. 2019, pp. 1–7 (cit. on p. 27).
- [276] J. Younas, S. A. Siddiqui, M. Munir, M. I. Malik, F. Shafait, P. Lukowicz, and S. Ahmed. "Fi-Fo Detector: Figure and Formula Detection Using Deformable Networks." In: *Applied Sciences* 10.18 (2020) (cit. on p. 27).
- [277] J. Younas and P. Lukowicz. "Cognitive Ability Classification using On-body Sensors." In: *[UbiComp/ISWC '22 Adjunct] Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Cambridge, United Kingdom: Association for Computing Machinery, 2022 (cit. on p. 81).

- [278] K. Yu, J. Epps, and F. Chen. "Cognitive load evaluation of handwriting using stroke-level features." In: Jan. 2011, pp. 423–426 (cit. on p. 86).
- [279] S. Yu, Q. Liu, J. Ma, L. Huixiao, and S. Ba. "Applying Augmented reality to enhance physics laboratory experience: does learning anxiety matter?" In: *Interactive Learning Environments* (Apr. 2022), pp. 1–16 (cit. on p. 21).
- [280] E. V. Yukhina. "Cognitive Abilities & Learning Styles in Design Processes and Judgements of Architecture Students." Doctor of Philosophy. PhD thesis. 2007-04-30 (cit. on p. 4).
- [281] D. Zhang, X. Wu, and C. Wang. "Fine-Grained and Real-Time Gesture Recognition by Using IMU Sensors." In: *2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS)*. 2017, pp. 747–754 (cit. on p. 100).
- [282] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai. "Multi-oriented text detection with fully convolutional networks." In: *ICVPR*. 2016, pp. 4159–4167 (cit. on p. 122).
- [283] W. Zhao, L. Gao, Z. Yan, S. Peng, L. Du, and Z. Zhang. "Handwritten Mathematical Expression Recognition with Bidirectionally Trained Transformer." In: *Document Analysis and Recognition – ICDAR 2021*. Cham: Springer International Publishing, 2021, pp. 570–584 (cit. on p. 84).
- [284] X. Zhong, J. Tang, and A. Jimeno-Yepes. "PubLayNet: Largest Dataset Ever for Document Layout Analysis." In: *2019 International Conference on Document Analysis and Recognition (ICDAR)* (2019), pp. 1015–1022 (cit. on pp. 15, 43, 156).
- [285] L. Zhou, E. Fischer, C. Tunca, C. M. Brahms, C. Ersoy, U. Granacher, and B. Arnrich. "How We Found Our IMU: Guidelines to IMU Selection and a Comparison of Seven IMUs for Pervasive Healthcare Applications." In: *Sensors* 20.15 (2020) (cit. on p. 100).
- [286] G. Zhu and D. Doermann. "Automatic Document Logo Detection." In: *ICDAR*. 2007, pp. 864–868 (cit. on p. 121).
- [287] X. Zhu, H. Hu, S. Lin, and J. Dai. "Deformable ConvNets v2: More Deformable, Better Results." In: *CoRR* abs/1811.11168 (2018) (cit. on pp. 37, 38).
- [288] T. G. Zimmerman, J. R. Smith, J. A. Paradiso, D. Allport, and N. Gershenfeld. "Applying electric field sensing to human-computer interfaces." In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co. 1995, pp. 280–287 (cit. on pp. 141, 142).

INDEX

- l, 87
- AI, 3, 5, 6, 9, 11, 53, 61, 79, 93, 97, 133, 151, 152, 154, 155, 157
- ANCOVA, 138
- AP, 44, 45, 47–49
- API, 104
- AR, v, 3, 5, 6, 8, 9, 11, 19–21, 84, 87, 97, 100, 102, 103, 115, 133, 136, 154, 157
- ASCII, 104
- BLSTM, 18, 59, 110–112
- CCA, 12, 29, 30, 33, 34, 127
- CG, 138, 139
- CNN, 11, 13–15, 19, 29–33, 35, 37, 107, 108, 110–112, 114, 115, 125, 126
- CRF, 13, 18, 29–31, 59
- CV, v, 6, 7, 9, 11–13, 15, 25, 29, 35, 52, 152
- DCN, 37–39
- DFKI, vii, viii
- DL, v, 4, 5, 9, 11, 13, 18, 25, 29, 31, 32, 52, 54, 57, 59, 62, 64, 68, 70, 73, 76, 98, 104, 107, 110, 111, 115, 121, 122, 124, 127, 151–153
- DNN, 4, 6, 8, 11, 13–15, 18, 19, 29, 31–33, 35, 36, 39, 48, 50, 52, 125, 151, 152, 155–157
- DoF, 104, 109
- dStaR, 9, 119, 122, 124–131, 155
- DT, 65
- DTW, 58, 103, 110–113
- ET, 65, 69–74
- FAirWrite, v, 8, 9, 97, 98, 101, 103–105, 107–111, 114, 115, 154, 155, 157
- FC, 125, 126
- FCN, 9, 30, 36, 39, 119, 121, 122, 124–127
- FFD, v, xi, 7, 25, 26, 29, 33, 35, 36, 42–46, 50, 52, 153
- FFT, 31
- Fi-fo, xi, 7, 26, 29, 33–35, 37, 38, 41, 47–52, 152, 153
- FN, 44, 46, 49
- FP, 44, 46, 49
- FPN, 14, 29, 32, 33, 38, 39, 47, 49
- GBM, 66, 69–74
- GDTW, 58
- GPU, 36, 39
- GRU, 66, 69–74
- GUI, 98, 108, 155
- HBC, 140
- HCI, 141
- HEC, viii
- HMD, v, 5, 6, 8, 21, 136
- HMM, 12, 18, 102
- ICDAR2017-POD, v, 7, 14, 15, 25, 29, 31–33, 40–42, 44, 45, 47–52, 152
- IG, 138, 139
- ILSVRC, 125
- IMU, v, 8, 9, 86–88, 97, 98, 100–104, 106, 114, 115, 154
- IoU, 32, 44, 45, 47–49
- ISODATA, 124
- KNN, 17, 111–113
- LSTM, 59, 66, 69, 73

- mAP, 32, 44
- MEMS, 102, 104
- ML, v, 4, 7, 9, 53, 54, 57–59, 61, 62, 64, 65, 68, 70, 72, 73, 76, 102, 103, 110, 111, 153
- MR, 5, 8, 9, 11, 21, 97, 133, 136, 157
- MSE, 31, 66, 69
- NLPR, 31, 32, 45, 47
- NMS, 36
- OCR, 12, 15, 17, 18, 28–30, 56
- OS, 107, 108, 110–112, 114, 115
- PAL, 31, 32, 45, 47
- PCA, 123
- PDF, 12, 13, 25, 30–32
- POD, 7, 9, 11, 13–15, 25, 26, 29–34, 39, 40, 42, 43, 45, 50, 52, 152, 153, 156
- PoI, 3, 11, 156
- QR, 21
- RBF, 59
- RCNN, 13, 14, 29, 31–33, 35, 36, 38, 39, 44–47
- ResNet, 14, 35–37, 39, 126
- RF, 65, 69–74
- RFCN, 38, 39, 47
- RGB, 13, 48, 122, 125, 127
- RNN, 19, 31, 59, 66
- RoI, 35–38
- RPN, 29, 32, 35, 36
- SNE, 67
- SotA, v, 4–7, 9, 11, 13–15, 17–19, 26, 31–33, 35, 44, 45, 47, 48, 50–52, 54, 62, 71, 74, 98, 107, 115, 119, 122, 126, 128, 130, 151–153, 155, 156
- SSD, 14
- STEM, 53, 83, 136
- SVM, 17, 31, 124
- TP, 44, 46, 49
- USA, 66
- [! (!)3]USB, 87
- VR, v, 3, 5, 6, 8, 9, 11, 19–21, 84, 87, 97, 100, 102, 103, 115, 133, 154, 157
- WAP, 14
- YOLO, 14

ACADEMIC CURRICULUM VITÆ: JUNAID YOUNAS

CONTACT INFORMATION

Email: junaid.younas@dfki.de

EDUCATION

- TU Kaiserslautern *10/2015 - Present*
PhD Student/Researcher, Embedded Intelligence Group, Prof. Lukowicz
Main areas: On-line Content Classification, Document Structure, off-line
Content Classification, Handwriting classification, Performance
Analysis, Air-writing, Deep Learning, Machine Learning
PhD Topic: *From Cognition 2 Applications: Bridging Gaps Between Formal Edu-
cation & AI*
- Higher Education Commission of Pakistan *10/2015 - 05/2019*
Scholar of the *Master Leading to PhD for Faculty Development Program*
- University of Applied Sciences Frankfurt *10/2012 - 03/2015*
Master of Science in Information Technology
Specialization: Communication & Automation Master Thesis:
*Development of Reusable Tests for Automatic Generation of Test Suites
for Value-added Communication Services*
- Higher Education Commission of Pakistan *06/2013 - 03/2015*
Scholar of the *Master Leading to PhD for Faculty Development Program*
- National University of Modern Languages, Islamabad, Pakistan *01/2007 - 12/2010*
Bachelor of Engineering - BE, Telecommunications Engineering
Bachelor Thesis: *Laser-based Communication System*

EXPERIENCE

- TU Kaiserslautern *since 12/2021*
Full-time Researcher, Embedded Intelligence Department, Prof. Lukowicz
- TU Kaiserslautern *01/2017-11/2021*
Part-time Researcher, Embedded Intelligence Department, Prof. Lukowicz
- German Research Center for Artificial Intelligence (DFKI) *since 10/2015*
Guest Researcher, Embedded Intelligence Group, Prof. Lukowicz

SELECTED PUBLICATIONS

- gPhysics—Using Smart Glasses for Head-centered, Context-aware
Learning in Physics Experiments *Journal
Publications*
J. Kuhn, P. Lukowicz, M. Hirth, A. Puxrucker, J. Weppner **J. Younas**
- Sense the Pen: Classification of online Handwritten Sequences
J. Younas, M.I. Malik, S. Ahmed, F. Shafait, P. Lukowicz
- Fi-fo Detector: Figure and Formula Detection using Deformable Net-
works
J. Younas, S.A. Siddiqui, M. Munir, M.I. Malik, F. Shafait, P. Lukowicz, S. Ahmed

*Conference
Publications*

D-star: A Generic Method for Stamp Segmentation from Document Images

J. Younas, M.Z. Afzal, M.I. Malik, F. Shafait, P. Lukowicz, S. Ahmed

What am i writing: Classification of on-line Handwritten Sequences

J. Younas, S. Fritsch, S. Ahmed, M.I. Malik, F. Shafait, P. Lukowicz

Wrist-worn Capacitive Sensor for Activity and Physical Collaboration Recognition

S. Bia, V. Rey, J. Younas, P. Lukowicz

FFD: Figure and Formula Detection from Document Images

J. Younas, S.T. Rizvi, M.I. Malik, F. Shafait, P. Lukowicz, S. Ahmed

Finger Air Writing - Movement Reconstruction with Low-cost IMU Sensor

J. Younas, H. Margarito, P. Lukowicz

FAirWrite - Movement Reconstruction and Recognition using a Low-cost IMU

J. Younas, H. Margarito, P. Lukowicz

Cognitive Ability Classification using On-body Sensors

J. Younas, P. Lukowicz