DEVELOPMENT OF INTELLIGENT CHARACTER RECOGNITION MODEL INSPIRED BY COGNITIVE EXPERIMENTS

Chetan Sudarshan Ralekar

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2021

DEVELOPMENT OF INTELLIGENT CHARACTER RECOGNITION MODEL INSPIRED BY COGNITIVE EXPERIMENTS

by

Chetan Sudarshan Ralekar

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

NOVEMBER 2021

Dedicated to my family...

Certificate

This is to certify that the thesis titled **"Development of Intelligent Character Recognition Model Inspired by Cognitive Experiments"** being submitted by **Chetan Sudarshan Ralekar** to the **Department of Electrical Engineering**, Indian Institute of Technology Delhi, for the award of **Doctor of Philosophy** is a record of bonafide research work carried out by her under my guidance and supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

	Chaudhury
Professor Tapan Kumar Gandhi	Professor Santanu Chaudhury
Department of Electrical Engineering	Department of Electrical Engineering
Indian Institute of Technology Delhi	Indian Institute of Technology Delhi
New Delhi - 110016, India.	New Delhi - 110016, India.

Acknowledgements

I want to express my sincere gratitude to my advisors **Prof. Tapan Kumar Gandhi** and **Prof. Santanu Chaudhury**. The thesis would not have its present state without their guidance and encouragement. Both of my supervisors have always motivated me to strive for excellence and put in my hundred percent to the job at hand. My sincere thanks to Tapan sir and Santanu sir for believing in my sincere efforts and kept my morale high when things were not turning up as planned. It is rightly said if you get the right 'Guru' to guide you, then the sky is the only limit. I feel fortunate to have such eminent researchers as my mentors.

I would like to thank the members of my thesis committee: **Prof. Brejesh Lall, Prof. Anup Singh**, and **Prof. Sumeet Agarwal**, for their insightful suggestions, which encouraged me to widen my research from various perspectives. I am grateful to **Prof. Pawan Sinha, MIT USA** for his thoughtful suggestions and comments on my work. I want to thank all the participants who have actively participated in the studies. Thanks to the Ethics committee of IIT Delhi for approving the protocol of the experiment.

I am thankful to all my lab-mates Jyoti, Amit sir, Ashirbad, Rijul, Shipra ma'am, Kritika, Sidharth, Sangeeta ma'am, Chandrabhushan, Anupam, Amita, Shefali and post-doctoral fellows Dr. Taranjit, Dr. Rohit from Neurocomputing lab for keeping the environment in the lab lively. Special thanks to my friend and collaborator, Shubham Choudhary, who has helped me explore the world of Keras, tensor-flow, and PyTorch. Thanks to Shruti ma'am, Shilpi ma'am, Manaswi, Jobin sir, and seniors from multimedia lab Dr. Meera, Dr. Anupama Ray, Dr. Ritu Garg for supporting me. I am thankful to my fellow collaborators Punyajoy, Nishad, Jatin Jain, Rupesh, for assisting me in work. I enjoyed working on several projects with summer and GI-

PEDIA interns Afshan, Varsha Suresh, Ambuj, Saumya Das, Riddhi, Ankur and course project with Sukruth and Swatilekha. Thanks for letting me be your student mentor.

I extend my gratitude to Tapan sir and Pawan sir for introducing me to Project Prakash. I am thankful to the entire Project Prakash family for their support. I want to thank Dr. Paula Rubio-Fernandes (University of Oslo, Norway) and Dr. Shlomit Ami-Ben (Post-doc, MIT) for collaborating with me on various research projects, and Prof. Ashwini Vaidya and Prof. Samar from HSS IIT Delhi, for insightful discussions on various technical and non-technical topics.

My Ph.D. journey will be incomplete if I do not mention my association with SPIC MACAY club and Music club, IIT Delhi. Both of these clubs have helped me grow and showcase my talent. Thanks to all the members of these clubs for making me a part of the team representing IIT Delhi in various competitions and platforms. I am thankful to Padma Shri Prof. Kiran Seth, founder of SPIC MACAY, Dr. Pragya Jain and Dr. Rachana Prasad for believing me and letting me anchor various concerts of eminent artistes. These extracurricular activities have helped to keep my mind fresh which has resulted in improved concentration during work. I have been very fortunate in having Anshul, Vinay, Deepak, Usham, and Soumava as my best buddies at IIT, who stood by me through the good and bad. The journey could not have been easy without the constant support from you guys. Thanks to my friends from SM group especially, Arunima, Aparna, Dr. Kritika, Sanchayan and Laxman and on-capmus friends Rajib, Surpriyo for their support.

Last but not least, I acknowledge the most important contribution of my parents and sister, who formed part of my vision and taught me good values that matter in life. Their unfailing love and support have always been my strength. Their patience and sacrifice will remain my inspiration throughout my life. I would like to express my heartiest gratitude towards my best friend Mayur, who patiently supported me through all my efforts and encouraged me to pursue my Ph.D. Finally, I am thankful to God and all the well-wishers who bestowed their blessings on me.

Chetan Ralekar

Abstract

Character recognition has been one of the most contemporary and challenging areas of research for decades. The past research on developing character recognition models using image processing and computer vision-based techniques shows that 'character recognition' is assumed to be a pure machine vision optimization problem. However, humans are still better at recognizing many characters especially distorted, ornamental or calligraphic characters as compared to the highly sophisticated recognition models. Although understanding the mechanism of character recognition by humans may give us some cues leading to better recognition abilities, the appropriate methodological approach of using these cues has not been much explored for developing character recognition models. Therefore, in this thesis, we propose to perform some cognitive experiments with humans to obtain some cues which will be used to develop intelligent character recognition models.

We start the investigation by understanding the reading behavior followed by distorted and ornamental character recognition through the analysis of the eye movement data. The experimental results show that humans direct their gaze to specific regions contributing to the character identity and recognize the characters using those key character regions. To highlight the character regions influencing the model's decision, we have proposed a 'modified Grad-CAM visualization technique', which is one of the contributions in the domain of explainable AI. Qualitative comparison between visualization maps and eye-fixation maps reveals that the deep learning model considered similar regions in character, which humans have fixated on, in the case of correctly classified characters. On the other hand, when the focused regions are different for humans and deep nets, the characters are typically misclassified by the latter. Both

visualization maps and fixation maps are used to develop a visual explanation guided attention model (VEGAM). Furthermore, this thesis presents distinct attention models such as the deeply supervised and collaborative attention models that signify the importance of foveal and parafoveal vision in recognition. The experimental design, data analysis, and results presented in this thesis demonstrate the effectiveness of the proposed approaches.

Owing to a sophisticated vision system, vast visual experience, rich cognitive and perceptual processes, humans have a distinct edge over machines in the recognition task. The behavioral experiments using Bubble and proposed RISE (Random image structure evolution) stimuli emphasize the presence of some higher-level cognitive processes aiding successful recognition. In this era of AI, researchers are striving hard to build systems that can match human intelligence. One way to bridge the gap between machine and human intelligence is to design the artificial systems using cues derived from their natural counterpart, such as through cognitive experiments on human participants. This thesis, henceforth, motivates the need for substantial efforts in exploring various aspects of the cognitive processes to develop the next generation collaborative models for mimicking the human intelligence in AI-driven Engineering systems.

सार

करैक्टर रिकग्निशन (अक्षर पहचान) दशकों से सबसे समकालीन और चुनौतीपूर्ण अनुसंधान के क्षेत्रों में से एक रहा है। इमेज प्रोसेसिंग (छवि प्रसंस्करण) और कंप्यूटर दृष्टि-आधारित तकनीकों का उपयोग करके करैक्टर रिकग्निशन मॉडल विकसित करने के पिछले कई शोधों से पता चलता है कि 'करैक्टर रिकग्निशन' को सिर्फ मशीन दृष्टि अनुकूलन समस्या माना जा रहा है। हालांकि, अत्यधिक परिष्कृत मान्यता मॉडल की तुलना में मनुष्य अभी भी कई कैरक्टर्स को; विशेष रूप से, विकृत, सजावटी या सुलेख अक्षरों को पहचानने में बेहतर हैं। यद्यपि मनुष्यों द्वारा करैक्टर रिकग्निशन के तंत्र को समझने से हमें कुछ संकेत मिल सकते हैं, जिससे बेहतर पहचान (रिकग्निशन) क्षमता हो सकती है। किन्तु करैक्टर रिकग्निशन मॉडल विकसित करने के लिए इन संकेतों का उपयोग करने के उपयुक्त पद्धतिगत दृष्टिकोण का अधिक पता नहीं लगाया गया है। इसलिए, इस थीसिस में, हम कुछ संकेत प्राप्त करने के लिए मनुष्यों के साथ कुछ संज्ञानात्मक प्रयोग करने का प्रस्ताव करते हैं, जिनका उपयोग बुद्धिमान करैक्टर रिकग्निशन मॉडल को विकसित करने के लिए किया जाएगा।

हम आंखों की गति के विश्लेषण के माध्यम से पढ़ने के व्यवहार को समझकर विकृत और सजावटी करैक्टर रिकग्निशन की जांच शुरू करते हैं। प्रायोगिक परिणामों से पता चलता है कि मनुष्य करैक्टर रिकग्निशन में योगदान देने वाले विशिष्ट क्षेत्रों (रीजन) की ओर अपनी आँखे टिकाता हैं और उन प्रमुख क्षेत्रों का उपयोग करके करैक्टर (अक्षर) को पहचानते हैं। मॉडल के निर्णय को प्रभावित करने वाले करैक्टर क्षेत्रों को उजागर करने के लिए, हमने एक 'संशोधित ग्रैड-कॉम' विज़ुअलाइज़ेशन तकनीक' का प्रस्ताव दिया है, जो एक्सप्लेनेबल एआई के क्षेत्र के योगदानों में से एक है। विज़ुअलाइज़ेशन मैप्स और आई-फिक्सेशन मैप्स के बीच गुणात्मक तुलना से पता चलता है कि डीप लर्निंग मॉडल ने करैक्टर में समान क्षेत्रों को माना, जिसे मनुष्यों ने ठीक से वर्गीकृत वर्णों के मामले में तय किया है। दूसरी ओर, जब मनुष्यों और डीप लर्निंग मॉडल के केंद्रित क्षेत्र भिन्न होते थे, तो करैक्टर को आम तौर पर मॉडल द्वारा गलत वर्गीकृत किया गया । विज़ुअल एक्सप्लनेशन गाइडेड अटेंशन मॉडल (वीईजीएएम) विकसित करने के लिए विज़ुअलाइज़ेशन मैप्स और फिक्सेशन मैप्स दोनों का उपयोग किया जाता है। इसके अलावा, यह थीसिस अलग-अलग अटेंशन मॉडल प्रस्तुत करता है , जैसे कि डीपली सुपरवाइज़ और कोलॅबोरेटिव मॉडल्स , जो मान्यता में फोवियल और पैराफॉवेल दृष्टि के महत्व को दर्शाते हैं। इस थीसिस में प्रस्तुत प्रयोगात्मक डिजाइन, डेटा विश्लेषण और परिणाम प्रस्तावित दृष्टिकोणों की प्रभावशीलता को प्रदर्शित करते हैं।

एक परिष्कृत दृष्टि प्रणाली, विशाल दृश्य अनुभव, समृद्ध संज्ञानात्मक और अवधारणात्मक प्रक्रियाओं के कारण, मनुष्यों की 'पहचान कार्य' में मशीनों पर एक अलग बढ़त है। बबल और प्रस्तावित RISE (रैंडम इमेज स्ट्रक्चर इवोल्यूशन) स्टिम्युलि का उपयोग करने वाले प्रयोग कुछ उच्च-स्तरीय संज्ञानात्मक प्रक्रियाओं की उपस्थिति पर जोर देते हैं; जो सफल अक्षर पहचान कार्य की सहायता करते हैं। एआई के इस युग में, शोधकर्ता ऐसे सिस्टम बनाने के लिए कड़ी मेहनत कर रहे हैं जो मानव बुद्धि से मेल खा सकें। मशीन और मानव बुद्धि के बीच की खाई को पाटने का एक तरीका यह है की हम मानव प्रतिभागियों पर संज्ञानात्मक प्रयोगों के माध्यम से प्राप्त संकेतों का उपयोग कृत्रिम प्रणालियों को डिजाइन करने में करे। यह थीसिस, अब से एआई-संचालित इंजीनियरिंग सिस्टम में मानव बुद्धि की नकल करने के लिए अगली पीढ़ी के सहयोगी मॉडल विकसित करने के लिए संज्ञानात्मक प्रक्रियाओं के विभिन्न पहलुओं की खोज में पर्याप्त प्रयासों की आवश्यकता को प्रेरित करता है।

Contents

List of Figures

List of Tables