SEQUENTIAL LABELING FOR RECOGNITION OF IMAGE BASED PATTERNS

ANUPAMA RAY

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

AUGUST 2017

SEQUENTIAL LABELING FOR RECOGNITION OF IMAGE BASED PATTERNS

by

ANUPAMA RAY

DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted

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

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

AUGUST 2017

For you Baba

Certificate

This is to certify that the thesis titled **Sequential Labeling for Recognition of Image Based Patterns** being submitted by **Ms. Anupama Ray** to the **Department of Electrical Engineering**, Indian Institute of Technology Delhi, for the award of **Doctor of Philosophy** is a record of bona-fide 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.

Professor Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Santanu Chaudhury, who has been a huge inspiration in my life. His ideas and his zest for work has been a constant motivation. He has been very generous with his time and always encouraged me for better research. I take this opportunity to thank him for being so understanding and supporting me with all the financial help, without which PhD would be extremely difficult.

I am extremely grateful to the members of my SRC committee: Prof. P. K. Kalra, Dr. Sumantra DuttaRoy and Dr. Brejesh Lall, for their insightful suggestions and constant help which encouraged me to widen my research from various perspectives. I am grateful to the technical staff of Multimedia Lab who allowed me own this lab for the last four years. This lab has been home to me. I thank my seniors Ritu Mam, Meera Mam, and Nisha Mam for their help and advice throughout my PhD. I was lucky to have very friendly labmates like Manoj, Ronak, Deepika and Richa around me. I have been lucky to get the opportunity to team up with a few juniors of mine, some of whom who have been my co-authors. I specially mention Sai Rajeswar, who has seen the long sleepless nights with me in Multimedia Lab without any results. Thank you for being with me, tolerating me and supporting me to get throughout. I have been very fortunate to have friends like Peeyush and Chandni who stood by me through the good and bad. I would like to express my heartiest gratitude towards Karun who stood by me patiently through all my efforts, studied with me, read my papers my thesis, helped me prepare for coding interviews and constantly encouraged me to peruse my PhD. I can never thank you enough!!

This thesis is dedicated to my Baba, it was his dream which I tried to fulfill, I know I

couldn't do my best, but I tried really hard. My mother and sister have been my oxygen tank and no thanks could ever be sufficient. I thank my family for forming my vision, exposing me to the world and allowing me to live the way I wanted. Their infallible love and support has always been my strength. Their patience and sacrifices will remain my inspiration throughout my life.

Anupama Ray

Abstract

Sequential Labeling is a structured output prediction task which requires labeling sequence of inputs by sequence of labels, and the sequences can be of arbitrary length. The label of each input in the input sequence is influenced by the label of the neighboring inputs. This thesis aims to extend the research in sequential labeling for recognition of 2D images. Since text recognition is a sequential labeling problem, we primarily focus on text recognition from images.

In terms of methodology, the first contribution of this work is a novel graph based sequential labeling framework using multiple hypotheses at input and Conditional Random Fields (CRF)/ SVM-HMM as the probabilistic recognizer to capture associations and selecting the best path. We apply this framework for segmentation based OCR and demonstrate results on degraded and noisy documents of printed and handwritten text. The second contribution of this work aims to advance the state-of-art Long Short Term Memory (LSTM), a recurrent neural network by adding dimension to make it structurally suitable to 2D images. We construct a 2D LSTM and stack several layers to create a 2D deep Bidirectional LSTM (BLSTM) network. We use this 2D deep BLSTM architecture within a hypothesis-and-verify architecture along with a learned language model to create a complete script independent, segmentation free OCR architecture for multilingual documents. Our third contribution is an end-to-end trainable Convolutional LSTM (CLSTM) architecture for text in the wild. We propose a framework for iterative text localization using character recognition which shows improvements on state-of-the-art object detection, followed by text recognition using CLSTM. The fourth contribution of this work is a novel end-to-end trainable framework with Convolutional and Fixed point layers, to capture

label dependencies as well as spatial relationships with a faster and guaranteed convergence. We have focused on Indian language scripts since research on Indian scripts are not at par with the Latin scripts. All the frameworks developed are script independent and can be trained on any script.

सार

अनुक्रमिक लेबल एक संरचित आउटपुट भविष्यवाणी कार्य है जिसके लिए लेबल के क्रम से इनपुट के लेबलिंग अनुक्रम की आवश्यकता होती है, और अनुक्रम मनमानी लंबाई का हो सकता है। इनपुट अनुक्रम में प्रत्येक इनपुट का लेबल पड़ोसी इनपुट के लेबल से प्रभावित होता है। इस थीसिस का उद्देश्य 2D छवियों की पहचान के लिए अनुक्रमिक लेबलिंग में अनुसंधान का विस्तार करना है। चूंकि पाठ पहचान एक अनुक्रमिक लेबलिंग समस्या है, इसलिए हम मुख्य रूप से छवियों से पाठ पहचान पर ध्यान केंद्रित करते हैं।

कार्यप्रणाली के संदर्भ में, इस काम का पहला योगदान एक नवीन ग्राफ़ आधारित अनुक्रमिक लेबलिंग ढांचे है जो इनपुट और कंडीशनल रैंडम फ़ील्ड (CRF) / SVM-HMM कई अनुमानों का उपयोग करता है जो एसोसिएशन पर कब्जा करने और सर्वोत्तम मार्ग का चयन करने के लिए संभाव्य पहचानकर्ता है। हम विभाजन के आधार पर ओसीआर के लिए इस ढांचे को लागू करते हैं और मुद्रित और हस्तलिखित पाठ के अपमानित और शोर दस्तावेजों पर परिणाम प्रदर्शित करते हैं।
इस काम का दूसरा योगदान 2D छवियों के लिए संरचनात्मक रूप से उपयुक्त बनाने के लिए आयाम जोड़कर एक लयबद्ध शॉर्ट टर्म मेमोरी (LSTM), एक पुनरावर्ती तंत्रिका नेटवर्क को आगे बढ़ाने का लक्ष्य है। हम एक 2D एलएसटीएम का निर्माण करते हैं और एक 2D गहरी द्विदिश एलएसटीएम (BLSTM) नेटवर्क बनाने के लिए कई स्तरों का ढेर लगाते हैं। हम इस 2D गहरी बीएलएसटीएम आर्किटेक्चर का उपयोग एक परिकल्पना और सत्यापित वास्तुकला के भीतर करते हैं, जिसमें सीखी भाषा मॉडल के साथ-साथ बहुभाषी दस्तावेजों के लिए एक पूर्ण स्क्रिप्ट स्वतंत्र, विभाजन मुक्त ओसीआर आर्किटेक्चर तैयार किया जाता है।

हमारा तीसरा योगदान प्राकृतिक दृश्य चित्रों में पाठ पहचान के लिए एंड-टू-एंड ट्रेनेबल कनवोल्यूशनल एलएसटीएम (CLSTM) वास्तुकला है। वर्णित पाठ स्थानीयकरण के लिए हम ढांचे को प्रस्तावित करते हैं, जो ऑब्जेक्ट का पता लगाने में नवीनतम तकनीकों में सुधार दिखाते हैं, इसके बाद पाठ पहचान के लिये हम सीएलएसटीएम का उपयोग करते है।

इस काम का चौथा योगदान एक परिकल्पनात्मक और निश्चित बिंदु परतों के साथ एक अन्तराल एंड-टू-एंड ट्रेनेबल फ्रेमवर्क है, जो कि लेबल निर्भरता को पकड़ने के साथ-साथ तेज और गारंटीकृत अभिसरण के साथ स्थानिक रिश्तों को भी प्राप्त करता है। हमने भारतीय भाषा लिपियों पर ध्यान केंद्रित किया है क्योंकि भारतीय लिपियों पर शोध लैटिन स्क्रिप्ट के अनुरूप नहीं हैं। विकसित सभी ढांचा स्क्रिप्ट स्वतंत्र हैं और किसी भी स्क्रिप्ट पर प्रशिक्षित किया जा सकता है।

Contents

List of Figures

List of Tables