Part of Word Segmentation and Recognition Techniques for Urdu Handwritten Text

by

Muhammad Kashif Siddhu
(1440211359)

A thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

School of Computer and Communications Engineering
UNIVERSITI MALAYSIA PERLIS

2020

ACKNOWLEDGMENT

All praise be to Allah. I first thank Allah for giving me strength, patience and knowledge to complete this thesis.

Secondly, I would like to thank Dr. Shahrul Nizam Yaakoob for accepting my initial proposal and agreeing to supervise me. Thank you very much Dr. for your guidance in the research and thesis writing. I especially thank you for coming to me for every meeting yourself as your office was far away from my Lab. You were always there whenever I needed your help.

I would also like to thank my field supervisor and mentor Dr. Muhammad Tanvir Parvez for helping to set my research direction even before my PhD has started. Thank you very much sir for sharing and discussing ideas and helping me from the implementation of the algorithms to writing the research papers.

Finally, I want to thank my family for all their support and patience during the period of my PhD studies. This work would never have completed without their prayers and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

POW	Part of Word
UNHD	Urdu Nastal'iq Handwritten Dataset
CC	Connected Component
PC	Primary Component
SC	Secondary Component
GAN	Generative Adversarial Networks
DCGAN	deep Convolutional GAN
LSTM	Long Short Term Memory
SIFT	Scale Invariant Feature Transform
HOG	Histogram of Oriented Gradient
LBP	Local Binary Pattern
DBN	Deep Belief Networks
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks
CDBN	Convolutional Deep Belief Networks
ResNets	Residual Neural Networks
MDLSTM	Multidimensional LSTM
CTC	Connectionist Temporal Classification
UPTI	Urdu Printed Text Image
ReLu	Rectified Linear units
MLP	Multilayer Perceptron
HMMs	Hidden Markov Models
OOV	Out Of Vocabulary
WFST	Weighted Finite State Transducer
PN	Policy Network
SVM	Support Vector Machines
RBF	Radial Basis Function
PHOC	Pyramidal Histogram of Characters
CCA	Canonical Correlation Analysis
PBSC	Pyramid of Bidirectional Character Sequences
DCToW	Discrete Cosine Transform of words
LSDE	Levenshtein Space Deep Embedding
TPP	Temporal Pyramid Pooling

BCE	Binary Cross Entropy loss
PRM	Probabilistic Retrieval Model
DAP	Direct Attribute Prediction
QBS	Query by String
RPN	Region Proposal Network
NMS	non-max suppression
ResNets	Residual Neural Networks
FC	Fully Connected
ROI	region of interest
CSR	Common Subspace Regression
CRNN	Convolutional Recurrent Neural Network
STN	Spatial Transformer Network
GRU	Gated Recurrent Units
BOC	Bag of Characters
AC-GAN	Auxiliary Classifier Generative Adversarial Networks
VGG Net	Visual Geometry Group Net
DR	Detection Rate
SGD	Stochastic gradient descent

LIST OF SYMBOLS

Lx	x coordinate of the lower left edge of the bounding box
Rx	x coordinate of the lower right edge of the bounding box
Wx	Width of bounding box
CCi	Current connected component
CCj	Preceding connected component.
Ly	Lower edge measurement
Uy	Upper edge measurement
D	Discriminator
G	Generator
Z	Random points in the latent space
G(z)	Fake image
LS	Source loss indicating the realness or fakeness of image
LC	Class loss indicating the loss in determining the correct class
S	Stride of the filter
P	Padding in the image
F	The size of the filter
D	Number of filters in a convolutional layer
o2o	One to one detection
N	Number of POWs in the ground truth

Sebahagian daripada Teknik Pengecapan dan Pengiktirafan Kata untuk Teks Urdu Handwritten

ABSTRAK

Tesis ini menjalankan penyelidikan mengenai pengenalan tulisan tangan Urdu. Sistem pengiktirafan tulisan tangan mempunyai beberapa aplikasi mis. pemprosesan borang, automasi pos, pendigitalan dokumen, pemprosesan cek bank, dan lain-lain Pengiktirafan teks dalam dokumen tangan tulisan tangan Urdu adalah di peringkat awal. Salah satu sebab yang kurang mendapat perhatian daripada komuniti penyelidikan adalah kesediaan set data. Dataset pertama dicadangkan hanya dua tahun lalu. Ini adalah karya pertama mengenai pengedaran dokumen handwriting Urdu berdasarkan segmentasi. Tesis ini mencadangkan algoritma segmentasi novel untuk menyegarkan imej baris teks tulisan tangan Urdu ke dalam Bahagian Perkataan (POWs). Oleh kerana dataset yang tersedia adalah kecil dan tidak cukup untuk melatih pengelas berasaskan pembelajaran, teknik pembesaran data digunakan untuk meningkatkan jumlah data. Untuk tujuan ini transformasi affine yang merangkumi putaran dan pembetulan digunakan. Juga, Rangkaian Penguatkuasaan Generik Pembantu Tambahan (ACGANs) yang merupakan variasi rangkaian adversarial Generatif (GAN) digunakan untuk menghasilkan imej yang kelihatan seperti ditulis oleh manusia. Bagi pengiktirafan POW, tiga pengeluar pembelajaran mendalam telah dianalisis iaitu AlexNet, VGG16, dan VGG19. Semua ini adalah Network Neural Convolutional (CNN). Model-model ini telah mencapai keadaan prestasi seni dalam imej semula jadi serta pada imej teks tulisan tangan. Untuk melatih pengelas ini, teknik pembelajaran pemindahan diterapkan. Untuk tujuan ini, model pra-terlatih arkitek ini digunakan. Eksperimen dilakukan pada dataset UNHD tulisan tangan Urdu. Untuk segmentasi POW, eksperimen juga dilakukan pada dataset IFN / ENIT untuk perbandingan dengan algoritma segmentasi yang dicadangkan untuk skrip tulisan tangan tulisan tangan Arab. Hasilnya menunjukkan prestasi yang sangat baik pada kedua dataset ini. Kadar pengesanan 80.22% dicapai pada dataset UNHD dan kadar pengesanan 90.58% dicapai untuk dataset IFN / ENIT. Untuk pengakuan POW, eksperimen dilakukan pada dataset UNHD. Eksperimen dilakukan pada data asal iaitu tanpa tambahan dan juga data tambahan. Hasil menunjukkan peningkatan prestasi yang ketara apabila menggunakan data tambahan. Ketepatan pengiktirafan terbaik 96,48% dicapai pada VGG16 menggunakan data tambahan. Ini adalah 4.48% lebih baik daripada ketepatan pada data asal. Jadi kerja ini menyediakan penyelesaian untuk semua fasa Urdu POW pengenalan saluran mencapai keputusan yang sangat baik. Sebagai karya penumbuk pada teks tulisan tangan Urdu berdasarkan segmentasi, ia boleh menjadi penanda aras untuk penyelidikan masa depan dalam arah ini.

Part of Word Segmentation and Recognition Techniques for Urdu Handwritten Text

ABSTRACT

This thesis conducts research on Urdu handwritten text recognition. The handwriting recognition systems have several applications like forms processing, postal automation, document digitization and bank cheque processing. The text recognition in Urdu handwritten documents is in its infancy. This is the first work on segmentation based Urdu handwritten text recognition. All the three phases required for a segmentation based handwritten text recognition are addressed in this thesis. These phases include segmentation, data augmentation and recognition. This work proposes a novel segmentation algorithm to segment the Urdu handwritten text line images into Parts of Words (POWs). Since the available dataset is small and not have enough data to train a learning-based classifier, a data augmentation technique is designed to increase the amount of data. For this purpose, Auxiliary Classifier Generative Adversarial Networks (ACGANs) which are a variation of Deep Convolutional Generative adversarial networks (DCGANs) are used in combination with affine transformations to generate images that look like written by a human. This is the pioneer work that implements a deep generative model for data augmentation of Urdu POWs. For the POW recognition, three deep learning classifiers have been analyzed namely AlexNet, VGG16 and VGG19. All these are deep Convolutional Neural Networks (CNN). These models have achieved state of the art performance in natural images as well as on handwritten text images. To train these classifiers, the transfer learning technique is applied. For this purpose, pre-trained models of these architectures are used. Experiments are performed on Urdu handwritten dataset named UNHD dataset. For POW segmentation, experiments are also performed on an Arabic handwritten dataset named IFN/ENIT dataset for comparison with other segmentation algorithms proposed in the literature. The results show excellent performance of the proposed segmentation algorithm on all these datasets. A detection rate of 80.22% is achieved on UNHD dataset and a detection rate of 94.73% is achieved for IFN/ENIT dataset. For POW augmentation and recognition, the experiments are performed on UNHD dataset. Experiments are conducted on original data (without augmentation) as well as with augmented data. The results show significant improvement in performance when using the augmented data. The best recognition accuracy of 96.48% is achieved on VGG16 using the augmented data. This is 4.48% better than the accuracy on the original data. Therefore, this work provides solutions for all the phases of segmentation based Urdu handwritten text recognition pipeline achieving excellent results. Being a first work on segmentation based Urdu handwritten text; it can serve as a benchmark for future research in this direction.

CHAPTER 1 : INTRODUCTION

Document analysis is a vast field. It involves tasks like text recognition, layout analysis, writer identification, signature recognition, mathematical equations recognition, word spotting etc. Text recognition is the conversion of the text images (printed or handwritten) into machine editable text. Such a system has many applications like document digitization, postal automation, automated cheque processing, forms processing etc. Other applications include biometric and criminal identification systems (Kale et al., 2013). It is a successful solution for printed text but not for multi-writer handwritten documents. Poor writing style, diversity in writing styles as well as cursive nature of text make it difficult for handwriting recognition systems to achieve 100% accuracy.

The leading research in text recognition is being carried out on Latin scripts (Mhiri et al., 2018; Wu et al., 2019). The research on Arabic scripts including Urdu, Arabic, and Persian etc. is comparatively very less. One reason for this is the non-availability of publically available handwritten datasets. Although this limitation is offset for Arabic after the availability of datasets like KFUPM Handwritten Arabic Text (KHATT) (Mahmoud et al.,2014) and the dataset Institut für Nachrichtentechnik/ Ecole Nationale d'Ingénieurs de Tunis (IFN/ENIT) (Pechwitz et al.,2002). However, for Urdu, such large datasets are still not available. In fact, the first handwriting dataset for Urdu script appeared in Ahmed (2017). The only online dataset available of Urdu handwriting is Urdu Nastalik Handwritten Dataset (UNHD) (Ahmed et al., 2019b). It has around 4300 segmented lines. Still, to the best of our knowledge, there are three research works on

Urdu handwritten script. Therefore, this is an open area of research and more people from research community are now focusing on research in Urdu script.

This research work focuses on Urdu handwritten text recognition. Urdu is the national and official language of Pakistan. It is also one of the constitutionally recognized "scheduled languages" of India. The speakers of Urdu language are also found in the Gulf States, Bangladesh, Europe and America in large numbers.

1.1 Background

In this section, a brief introduction to the terms is presented which are frequently mentioned in the coming sections.

1.1.1 Supervised Learning

In supervised learning, the learning system is first trained on labelled data Allred and Kelly (1990). The training enables the system to recognize similar unseen examples. CNNs and Recurrent Neural Networks (RNN) are examples of supervised learning algorithm.

1.1.2 Unsupervised Learning

The system is trained on unlabelled data. The unsupervised algorithm learns useful properties of the structure of the dataset (Goodfellow et al., 2016). It models the probability density of the inputs. Examples are deep belief networks (DBN),

Convolutional Deep Belief Networks (CDBN), Auto Encoders (AE) and Generative Adversarial Networks (GAN).

1.1.3 Convolutional Neural Networks

A convolutional neural network (CNN) as depicted in Figure 1.1 comprises of convolutional and fully connected layers (LeCun et al., 1998). Operations like pooling and batch normalization are also a significant part of CNNs.

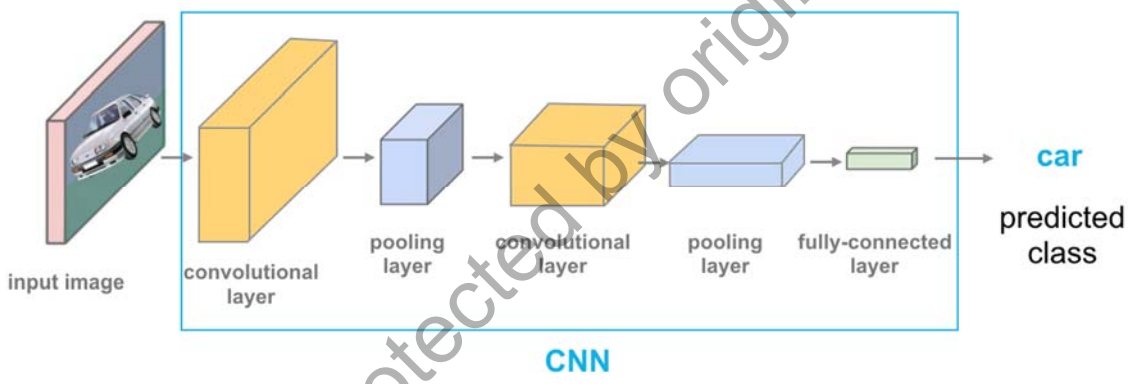


Figure 1.1 Convolutional Neural Networks (Camacho, 2018)

1.1.4 Residual Neural Networks (ResNets)

The accuracy of a CNN is increased by adding more layers. However, this addition of layers at a point may introduce the vanishing gradient problem and decrease the recognition eventually. To eliminate this problem in very deep CNNs, residual connections are introduced (He et al., 2016) and the resulting networks are called ResNets. Figure 1.2 presents a Residual block in ResNet.

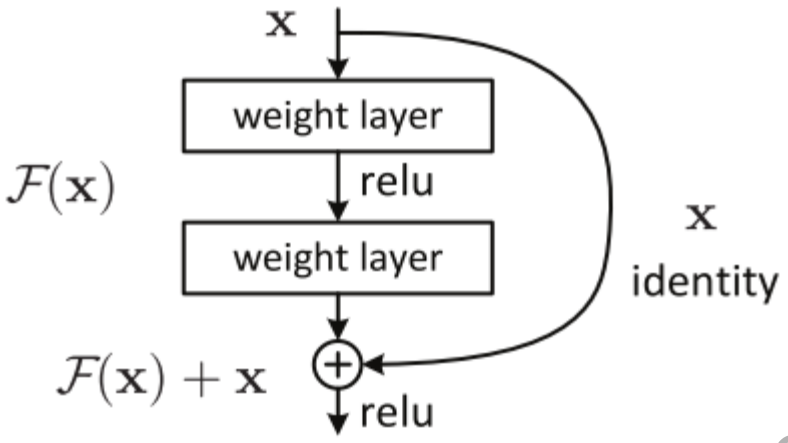


Figure 1.2 Residual block in ResNet (Fung, 2017)

1.1.5 RNNs and LSTMs

RNNs compute their output predictions for the current input taking into account the input before it in time Sherstinsky (2020). RNNs also suffer from the problem of vanishing gradient. To mitigate this problem, the RNNs are equipped with long short-term memory (LSTM) cells. In the document analysis, they are used for segmentation free text recognition or word spotting. An RNN with LSTM cells is depicted in Figure 1.3.

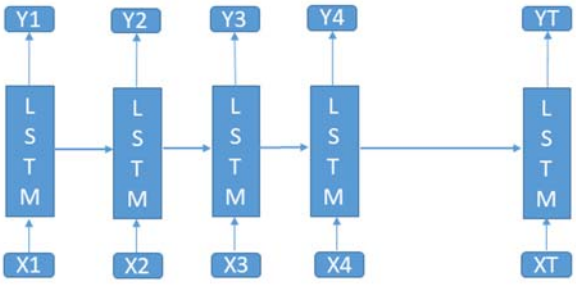


Figure 1.3 RNN with LSTM cells

1.2 Introduction to Urdu script and its challenges

Urdu script is derived from Arabic script. It is also written from right to left direction. Its alphabet includes all the letters of Arabic script. It has 58 letters in its alphabet. There are 18 different diacritic marks and dots arrangement in Urdu. Urdu is always written in cursive form, i.e. it is written in joining even in printed form. The shape of a letter may change according to its position in the word. Some letters like Bay (ب) have 4 shapes according to their position in the word. Figure 1.4 shows examples of such letters.

Letter	Isolated form	Beginning	Middle	End
Bay				
Fay				

Figure 1.4 Shapes of letters at different positions in the word

The letters can be a combination of main body, dots and diacritics. The dots or diacritics may appear above or below a letter. The letters may have 0 to 3 dots (چ،ح،ج،ق) but can have a maximum of two diacritic (آ،ٹ،گ). Two letters, *Ttay* (ٹ) and *Ddal* (ڈ) have a special diacritic called *Tua* (ط). The letter *Kaf* (ک) has a single sloped line connected to it at its upper right position while the letter *Gaf* (گ) has two such lines, one over the other.

No Urdu word begins with these letters: *Noon Ghunna*(ں), two eyed *Hay* (ھ), *Array* (ڑ) and Big *Yay* (ے).

Some letters cannot be joined to the letters next to them. They are *Waaw* (و), *Alif* (ا), *Dhaal* (ڈ), *Daal* (د), *Thaal* (ذ), *Raa* (ر), *Array* (ڑ), *Zaay* (ز) and *Hamza* (ء). This is the reason, an Urdu word may have more than one parts. These parts can be individual letters or combinations of two or more letters. The parts of a word (POWs) which are composed of two or more letters are called "ligatures". E.g. in the word Muea-malat (معاملات), the POW *Muea* (معا) is a *ligature* while (ت) is the Urdu letter *Tay*.

The main body of a letter or ligature is called primary connected component or primary component (PC) and the dots and diacritics are called secondary connected components or secondary components (SCs). There are spaces between words and also inside the words. Ideally the inter word spaces should be greater than the intra word spaces. Although there are many fonts in which printed Urdu text is written. But the Urdu handwriting is written in Nastalik font. Also the Urdu text is written and read from right to left, but the numbers are written and read from left to right direction.

There are many challenges in the segmentation and recognition of handwritten Urdu text. First is that, the writers may not write in a straight line. The writers don't care about the inter word and intra word space rules. Different writers may write with different slants or tilts. This gives rise to overlapping POWs of same word or even POWs of neighbouring words. In addition, the writers don't care too much about the position of dots and diacritics. The dots and diacritics of a letter or ligature may be written closer to the neighbours than itself. Pen lifting is another major issue. A ligature must be written

in its entirety without lifting the pen. Violating this rule gives rise to broken ligatures which are the cause of errors in segmentation systems. The readers of Urdu are accustomed to these issues and can deduce the meaning from the context (Parvez & Mahmoud, 2013). However, these difficulties make handwritten Urdu text recognition a challenging task.

The existing works on Urdu handwritten text recognition have used segmentation free recognition approach (Ahmed et al., 2017; Ahmed et al., 2019a; Ahmed et al., 2019b). In this approach, a classifier attempts to recognize the text present in the text line image without segmenting the line. While the latest research on scripts other than Urdu is heavily focused on the recognition of words or part of words which are obtained by segmenting a text line image into words or (POWs) (Wu et al., 2019; Krishnan & Jawahar, 2018b; Rusakov et al., 2018b). These researches have produced state of the art results in Arabic as well as Latin scripts. In the segmentation based handwritten text recognition, first, the text lines are segmented into words using a segmentation algorithm. Then the words are recognized using a deep learning classifier. There is no research work on segmentation based Urdu handwritten text recognition. One of the reasons for this is the non-availability of any segmentation algorithm, which can segment the Urdu handwritten text lines into Part of Words (POWs). Figure 1.5 shows a handwritten text line segmented into its constituent parts POWs.

Figure 1.5 A text line segmented into POWs

1.3 Problem statement

The recognition of Urdu handwritten text starts with taking the images of text lines as input. The text line in Urdu is built on basic units called POWs. First, the text line has to be segmented into POWs. This process is very critical because if the segmentation is erroneous, the recognition performance is also degraded. There have been POW segmentation algorithms proposed for Urdu printed text (Lehal, 2013; Ahmad et al., 2017a) but they are not capable of coping the challenges posed by multi-writer handwriting recognition. Overlapping POWs, writing in different styles, smaller size of some PCs like *Dal, Wao, Ray* (د،ر،و) which may be confused with SCs, and no clear word boundaries make the POW segmentation algorithms for printed text impractical for handwritten text (Ghaleb et al., 2016). So the first problem is: a segmentation algorithm is required to segment the Urdu handwritten text line images into POWs.

Although the next stage after segmentation is the recognition of POWs, but the Urdu handwritten dataset available for experiments has too few samples of many POWs. Therefore, the second problem is the scarcity of data. As a sufficient amount of data is required to train a learning based classifier, a data augmentation technique needs to be

designed to overcome this scarcity of data. There are certain data augmentations techniques e.g. affine transformations, data synthesis etc. proposed in literature that need to be investigated for Urdu handwritten script to overcome this problem Elanwar (2013).

The highest recognition accuracy achieved on Urdu handwritten text recognition is 93% (Ahmed et al., 2019b). There is a lot of room to improve this accuracy. So the third problem is to achieve high accuracy. For recognition, the selection of a classifier is very crucial. The deep learning classifiers have achieved state-of-the-art performance in Latin as well as Arabic handwritten scripts (Sudholt & Fink, 2018; Rusakov et al. (2018a)). The success of these deep learning models e.g. Convolutional Neural Networks (CNN) in other scripts provides a motivation to analyse them for the task of Urdu handwritten POW recognition.

1.4 Objectives

The objective of this research is to achieve state of the art accuracy in handwritten Urdu script recognition. To achieve this objective, a segmentation based recognition approach is to be proposed. Therefore, the main goals of this work are:

1. Propose a segmentation algorithm to segment handwritten Urdu text line images into POWs.

2. Design method for data augmentation to overcome the scarcity of available POWs.

3. Develop method to recognize the segmented POWs with high accuracy.

1.5 Research scope

This research focuses on segmentation based hand written Urdu text recognition. This is a closed vocabulary system, meaning that it can recognize those POWs which present in the database under study. UNHD dataset mentioned previously is used to conduct the experiments. This dataset consists of around 4000 grayscale Urdu handwritten text line images. The images are of ".png" type. For segmentation, a novel algorithm is to be developed. The output of the segmentation algorithm will be images of POWs also of type ".png". Data augmentation techniques that are based on deep learning generative models and affine transformations are to be used to generate new images of the same type as mentioned above. POW recognition is to be carried out using deep learning classifiers.

1.6 Thesis organization

The thesis is split into 5 chapters. It is organized as follows:

1. Second chapter discusses the state of the art in the literature which is related to our objectives. It has three sections. The first section presents research on Urdu and Arabic POW segmentation. The second section reviews data augmentation methods with focus on generative modelling. The third section presents latest research on hand writing recognition.

2. Third chapter explains the methodology adopted in this work. The first section presents the proposed POW segmentation algorithm. The second section presents the details of the Auxiliary Classifier Generative Adversarial Network (ACGAN). This is used to augment the data to overcome the scarcity of data. Also it introduces the affine transformations used for data augmentation in this research. The third section explains the deep learning classifiers used for recognition of POWs.

3. Fourth chapter starts with presenting the results of segmentation algorithm. It also analyzes the results of segmentation. The second section of this chapter presents and discusses the results of POW recognition.

4. The last chapter concludes this work and suggests future directions of research.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

In this thesis, a segmentation-based approach to recognize POWs of Urdu words is proposed. The adopted methodology consists of following phases: 1) POW segmentation 2) Data augmentation and 3) POW recognition. The following sections explore the research literature in all these areas. It is to be noted that as there are too few works available in the areas of segmentation and data augmentation, the sections related to these areas are smaller.

The literature is overwhelmed with the work on recognition. Most of the available works are on already segmented data. This task is usually termed as word recognition for Latin scripts. However, for Arabic scripts Part of Word recognition (POW) strategy is adopted. These works use databases like IFN/ENIT (Pechwitz et al.,2002) of Arabic script and Institute of Informatics and Applied Mathematics dataset (IAM) (Marti & Bunke, 2002) and George Washington datasets (Lavrenko et al. , 2004) of English script. These databases consist of handwritten words. A concept close to word recognition is word spotting. In word spotting, the user provides a query to be searched in the database. The word spotting system matches the query with the word images and retrieves a list of matching words using a matching algorithm. The literature on word spotting is also significant to look for more directions for word recognition. In fact, many works on the above-mentioned datasets address both word recognition and word spotting. This is the reason that this review also considers works on word spotting.

As segmentation of handwritten documents, is a challenging task, some approaches attempt to recognize the documents without segmentation. The previous works on Urdu script follow this approach. A review of these techniques is also included as they use the same dataset as is used in this thesis.

The research on Urdu script is in its infancy. There are some works available on Urdu script. Most of them are on printed script. There are only three works available on handwritten Urdu script. In addition, these works use very similar approaches. Therefore other state of the art approaches need to be investigated. These approaches are mostly presented for Latin scripts. These works also apply their approach on the IFN/ENIT dataset. Due to the similarity of Arabic and Urdu scripts, works on Arabic are the most relevant to this work. This is the reason; literature on Latin and Arabic scripts is also investigated.

This chapter is divided into following sections: 1) Review of segmentation techniques for handwritten Arabic scripts 2) data Augmentation techniques 3) word and POW recognition.

2.2 Segmentation of handwritten Arabic scripts into POWs

The work on segmentation in printed and handwritten scripts can be divided into different levels, namely: Line level segmentation, Word level and POW level. Figure 2.1 presents different levels of segmentation.

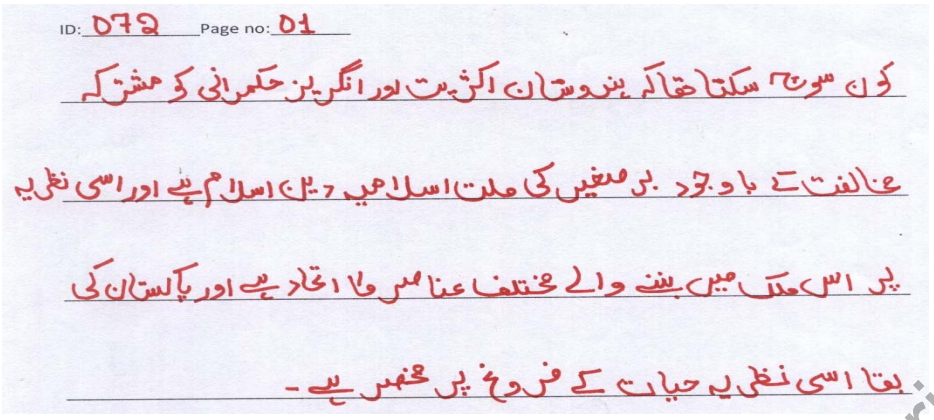
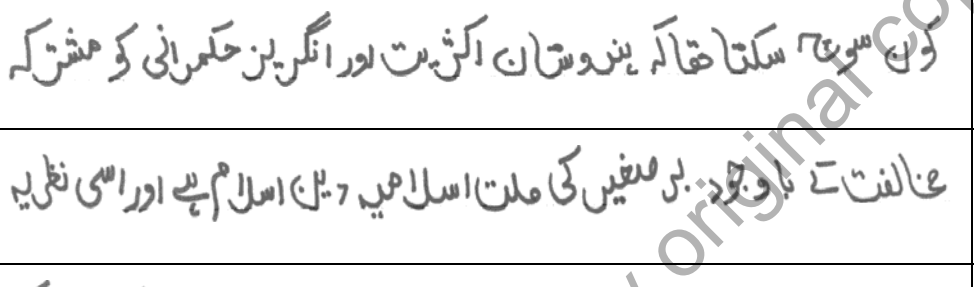


Figure 2.1 (a) Image of an unsegmented page containing four handwritten Urdu Text lines from UNHD dataset



First segemnted line

Second segmented line

Third segented line

Fourth Segmented line

Figure 2.1 (b) Images of segmented lines of the page (Line Level Segmentation)

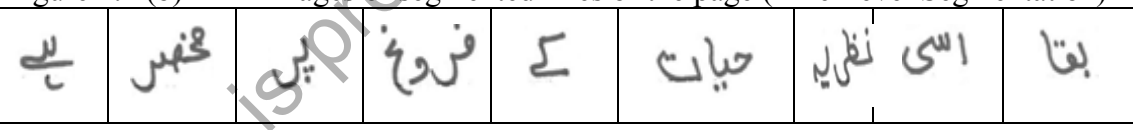


Figure 2.1 (c) Images of words after segmentation of fourth line into Words (Word level segmentation)

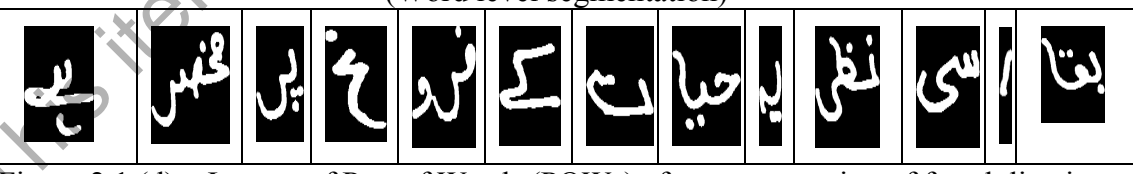


Figure 2.1 (d) Images of Part of Words (POWs) after segmentation of fourth line into POWs
(POW level segmentation)

Figure 2.1 Levels of Segmentation

For Latin scripts, word level segmentation is considered the most favourable. The word level segmentation in Latin scripts is considered easier as there are no intra word

spaces. While in Arabic scripts, a word may have more than one disjoint parts in the form of letters and ligatures. There should be some space between these parts. In addition, there should be space between the words. As mentioned earlier, the inter word spaces should be greater than intraword spaces. Figure 2.2 indicates interword spaces with arrows with dotted texture and intraword spaces with arrows with diagonal texture.

Figure 2.2 A text line with arrows indicating interword and intraword spaces

This rule is followed in printed scripts somewhat, but in handwriting; the writers do not usually follow this rule. This results in overlapping of the POWs, which belong to the same word. The POWs in the neighbouring words may also overlap. This makes segmentation of languages having Arabic scripts into words or POWs, a challenging task. Further segmentation of words or POWs into characters is found to be highly erroneous in Latin as well as Arabic scripts. This is the reason; state of the art segmentation based recognition of Latin scripts is focused on segmented words. For Arabic scripts, it is based on POWs. Therefore, in this work, segmentation of handwritten line images into POWs is carried out.

There are few works available on segmentation of handwritten Arabic scripts into POWs. AlKhateeb et al. (2009) assume that intra word distance is smaller than the inter word distance. They determined a distance threshold. If the distance between two connected components (CCs) were lesser than this threshold, they were merged to form a complete word; otherwise, each one of them is considered a separate word. They claim to achieve 85% success in segmentation.

Jamal et al. (2014) use a learning classifier to segment the word into POWs. The classifier was trained to recognize the isolated letters. In addition, it was trained to recognize ending letters in ligatures. No letter can be attached at the left of ending letters. Examples of these letters include RAA (ر), ALIF (ا), DAAL (د) etc. They used images of city names from IFN/ENIT dataset in experiments. At the start, the connected components (CC) were extracted. The classifier recognized the isolated characters and the ligatures having ending shapes. They were segmented and saved as individual images. They conduct experiments on 440 line images having two to three words only. They achieve a detection rate of 86%.

Ghaleb et al. (2016) use text lines' images from different Arabic handwritten datasets notably IFN/ENIT. Connected components were first extracted from the text line. Then starting from the right direction, the sub-words were extracted based on the rule that any CCs overlapped completely by other CCs should be merged with the larger one. Lastly, the misplaced dots were assigned to the sub-words using a refinement procedure. Interestingly their method did not use the baseline information. Due to this, they had many problems of disassociations and over segmentations. They claim to achieve 94.47% correct segmentation on 150 lines of IFN/ENIT dataset.

Maddouri et al. (2014) do not provide a clear methodology on POW segmentation. They conducted experiments on 30 images from IFN/ENIT dataset and achieved 85% correct segmentation.

Al-Dmour and Zitar (2016) extracted CCs from the handwritten text line image. Each CC was categorized into a letter, ligature or word by calculating its length and using

a clustering algorithm. The distance between two elements was calculated. The distance was categorized into intra word (D2) and inter word distance (D1). A complete word had D1 at its starting and end. An isolated character had D1 before and after it. A ligature has a D2 at one side. All these were segmented as separate units. They conducted experiments on 30 forms of AHDB database. They achieved 86.3% correct POW segmentation.

To the best of our knowledge, no work on segmentation of Urdu handwritten documents into POWs is available. However, there are researches on POW segmentation conducted on printed Urdu documents. Lehal (2013) proposed a POW segmentation algorithm for text lines from the Urdu Printed Text dataset named (UPTI) (Sabbour & Shafait, 2013). He first determined the base line. He assumed that almost all primary connected components (PCs) lie on the baseline but some secondary components (SCs) may also touch the baseline. He proposed 6 heuristics to detect such CCs. After the identification of all CCs, they were assigned to the PCs. He achieved 99.02% correct segmentation. (Ahmad et al., 2017a) followed a similar approach. After the extraction of connected components, they were classified as PCs and SCs using some heuristics. The SCs were then assigned to the PCs. They achieved 98.80% correct segmentation. Table 2.1 outlines the research on POW segmentation of Urdu and Arabic script.

Table 2.1 Research on POW segmentation in Arabic scripts

Reference	Script	Database	Data samples	Printed/ Handwritten	Segmenta-tion level	
Jamal et al. (2014)	Arabic	IFN/ENIT	440	Handwritten	POW	86%
Alkhateeb et al. (2009)	Arabic	IFN/ENIT	200	Handwritten	POW	85%
Al-Dmour et al. (2016)	Arabic	AHDB	30	Handwritten	POW	86.30%
Maddouri et al. (2014)	Arabic	IFN/ENIT	30	Handwritten	POW	85%
(Ghaleb et al. 2016)	Arabic	IFN/ENIT	150	Handwritten	POW	94.47%
Lehal (2013)	Urdu	UPTI	Not mentioned	Printed	POW	99.02%
Ahmad et al., (2017a)	Urdu	UPTI	Not Mentioned	Printed	POW	98.80%

2.3 Review of Data Augmentation techniques used in text recognition

When the transcribed data is available in limited amount, it may not be enough to successfully train a learning classifier. Especially the deep learning classifiers consist of millions of trainable parameters. If the data is not available in sufficient amount, they may not generalize well and overfit. Hence, the performance of the recognition system is not raised to required levels. A solution to this problem is data augmentation (Krishnan & Jawahar, 2018b). In data augmentation, new images are generated which are called "synthesized images". Data augmentation techniques can be used to enhance the amount

of data in the dataset. Data augmentation in handwritten images can be done by many different ways e.g. 1) Using deep learning generative models e.g. GANs (Goodfellow et al., 2014) 2) applying affine transformations like rotation and scaling on the original data and generating new images (Krishnan & Jawahar, 2016; Elanwar, 2013) 3) Using different fonts of the given script to generate new images (Krishnan & Jawahar, 2018a).

As is emphasized in the results and discussions chapter, after segmenting line images into POWs, too few samples are found for some POW classes. To increase the size of dataset, data augmentation is used. In this work, the first two techniques mentioned above are used for data augmentation.

The first technique, i.e. generating images using deep learning generative models, is different from the other techniques. The generative models have the ability to synthesize images, which look very similar to handwritten images. This application of synthesizing images is new but is rapidly improving and getting popularity. These networks have the ability to generate new training data that may result in better performing classification models (Shorten & Khoshgoftaar, 2019). Auto encoders and Generative adversarial Networks (GANs) are examples of generative models. Auto encoders suffer from the drawback of generating blurry images while vanilla GANs (based on Multilayer perceptron) generate noisy and incomprehensible images (Radford et al., 2015). On the other hand, the deep convolutional GANs (DCGANs) and their variants proposed later on, have continuously improved the quality of generated images. As for now, to the best of our knowledge, there are three works on handwriting synthesis using GANs. Alonso et al. (2019) propose a GAN with the aim of generating string images. Their system architecture consisted of a generator (G) and a discriminator (D)

network as in any GAN network. They introduced two new networks in the GAN: 1) a bidirectional LSTM network, which they denoted by (φ) 2) a gated convolutional neural network having convolutional layers followed by LSTM layers denoted by (R). The input to generator is random noise as well as embeddings of the word string generated by the Network (φ). The generator (G) generated the image of the input word string. Moreover, the discriminator recognized it as fake or original. The network (R) recognized the image generated by the network (G). They generated Arabic as well as French handwritten image strings. They included these images in the original datasets and reported improved accuracy. However, recognizer's performance solely on the GAN generated images was not found competitive.

Qian et al. (2019) propose a generative adversarial classifier to generate high-resolution character images from low-resolution images. They proposed an addition of a classifier network in the traditional GAN architecture. The generator network took a low-resolution image as input and generated a high-resolution image. The discriminator was trained to distinguish between real and high resolution images. The classifier network recognized the generated image. They achieved an accuracy of 93.69% on generated images of MNIST dataset (LeCun et al., 1998).

Chang et al. (2018) propose a version of GANs with an objective to generate character images of a font using character images from another font. The input character image and the output image did not need to be same. Their proposed network could also generate handwritten character images when given character images from a typed font as input. The network architecture consisted of an encoder network, which generated a low dimensional representation of the input image character. This representation was fed to a

transfer module that generated the feature representation in the output font style. This output representation was then fed into a decoder module, which generated a character in the target font style. Then the discriminator identified that whether the generated image was from the target font style or not. The generated images were then fed to HCCRGoogleNet (Zhong et al., 2015) classifier for recognition. An accuracy of 98.02% was achieved on generated handwriting characters.

2.4 Literature review on handwriting Recognition

This section reviews handwriting recognition literature. These techniques are mostly based on deep learning architectures. In traditional machine learning, features are first extracted from the input, and then these features are passed through a classifier to "train" it on that input. The classifier is trained on several labelled or known examples before it is able to classify or recognize the unseen examples. The classifier is a shallow network, which means having only one layer. Feature engineering plays a vital role in the performance of these systems. It requires deep expertise of the domain and tremendous human effort. Some of the examples of such feature extraction techniques, which achieved state-of-the-art results in handwriting recognition, include Scale Invariant Feature Transform (SIFT) Lowe (2004), Histogram of Oriented Gradient (HOG) (Dalal & Triggs, 2005), Local Binary Pattern (LBP) (Ojala et al., 1996) features etc. An advantage of these approaches is that they do not need high computational power and can be run on CPUs.

Deep learning is a sub field of machine learning. Deep learning algorithms involve multiple neural network layers. This is the reason they need a large-scale

annotated data as well as high performance parallel computing systems for training. Although (LeCun et al., 1998) introduced deep neural networks in 1998, due to the scarcity of large-scale publically available data, limited computation power, vanishing gradient problem and above all inferior performance compared with other machine learning techniques, it could not win the attention of the researchers.

Deep learning gained its popularity in 2006 by the ground breaking work of (Hinton et al., 2006) in which they introduced Deep Belief Networks (DBN). From there on, deep learning has overcome the field of computer vision, speech processing, natural language processing etc. This is made possible due to the following reasons:

1) Automatic feature extraction

2) Parallel processing using Graphics Processing Units (GPU)

3) Large scale publically available data

4) Availability of Pre-trained models

5) Deep learning frameworks like Tensorflow, Keras, PyTorch, MXNET etc. for rapid development of deep learning applications

6) Handling of problems like vanishing gradient using ReLu activation function, overfitting using dropout and efficient training using batch normalization

Although most of the research is carried out using Long Short Term Memory Networks (LSTM) and Convolutional Neural Networks (CNN), the latest works also use state-of-the-art techniques like attention mechanism, embedded attributes and reinforcement learning. In addition to that, there have been efforts to use unsupervised learning technologies, as there is a lack of availability of labelled handwritten datasets.

This section is organized as follows: Section 2.4.1 presents works done in Arabic, 2.4.2 presents works done in Urdu and 2.4.3 in Latin scripts.

2.4.1 Handwriting recognition in Urdu

This section presents the literature review on Urdu text. It is divided into sections based on the recognition techniques used in the mentioned works.

2.4.1.1 LSTM and MDLSTM with CTC

Urdu printed script recognition has been extensively carried out using different variations of Recurrent Neural Networks Bidirectional Long short term memory (RNN BLSTM) networks on different databases. Ul-Hasan et al. (2013) used text lines from Urdu Printed Text Image (UPTI) database. The database contains synthetically generated data. It has 10,063 lines. They used 80% of the database for training and validation while 20% for testing. They performed two types of experiments. In the first experiment, they labelled a character without considering its shape variations, which occur due to its locations in a ligature. That is to say, a character like Jeem (ج) which can appear in four different shapes depending on its position in the ligature is labelled by one class rather

than 4 classes. While in the second experiment, the same character is labelled by 4 different classes corresponding to its shape variations due to its location in the word. These resulted in 99 classes in the first experiment while 191 classes in the second experiment. They employed a BLSTM network with CTC as the output layer for the recognition task. Unsegmented text lines were first resized to a fixed height. A 30 x 1 window was traversed on the line image, which extracted the raw pixels from the image. This generated a one-dimensional vector, which was fed to the classifier along with the ground truth of the text line for training.

In (Ahmed et al., 2019), the authors introduce their Urdu handwritten dataset named Urdu Nastaleeq Handwritten Database (UNHD). They performed character recognition using 1-Dimensional LSTM RNNs and achieve an error rate of 7.93 % at character level recognition. A similar work (Ahmed et al., 2016), used printed Urdu text dataset named Urdu-Jang. They employed a BLSTM based classifier with CTC and tested their system on the text line images achieving a character level recognition accuracy of 88.94%.

Naz et al. (2017) use statistical features instead of raw pixels for the recognition of printed text lines using a BLSTM RNN based classifier. They also performed their experiments on UPTI database. In the output layer, they used CTC loss function. They used 13 different features to build the feature vector. They actually trained 8 different classifiers on 8 different combinations of features. A sliding window of size 4 X 48 was used to traverse the line image from right to left and top to bottom and then calculate the proposed features on them. They performed their experiments on 43 characters, which also included the blank space. They transcribed the different shape variations of a