BENGALI FONT IDENTIFICATION USING DEEP CONVOLUTIONAL NEURAL NETWORK

BY

KH. TANZILA RAHMAN
ID: 152-15-5737
AND

ROKEYA ISLAM RIYA
ID: 152-15-5559

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Md Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Md. Riazur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

MAY 2019

APPROVAL

This Research titled **"Bengali Font Identification Using Deep Convolutional Neural Network"**, submitted by Kh. Tanzila Rahman, ID No: 152-15-5737 and Rokeya Islam Riya, ID No: 152-15-5559 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 3rd May 2019.

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazmun Nessa Moon **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Abdus Sattar **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I declare that, this research has been done by me under the supervision of **Md Zahid Hasan, Assistant Professor, Department of CSE,** Daffodil International University. I also declare that neither this research nor any part of this research has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Zahid Hasan 15/04/19

Md Zahid Hasan
Assistant Professor and Coordinator of MIS
Department of CSE
Daffodil International University

Co- Supervised by:

Mr. Md. Riazur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Kh. Tanzila

Kh Tanzila Rahman
ID: -152-15-5737
Department of CSE
Daffodil International University

Riya

Rokeya Islam Riya
ID: -152-15-5559
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First of all, our heartiest thanks and gratefulness to Almighty Allah for His divine blessing that makes us capable to complete this research successfully.

We would like to thanks to our honorable teacher & project supervisor **Md Zahid Hasan, Assistant Professor, Department of CSE,** Daffodil International University for his endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Syed Akhter Hossain**, Head, Department of CSE, for his kind help to finish our research and we are also thankful to all the other faculty and staff members of our department for their co-operation and help.

We must acknowledge with due respect the constant support and patients of our parents.

Finally, we would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

ABSTRACT

Automatic font recognition or similar font suggestions from an image or picture are the core design works for many designers. This paper proposes a framework based on Convolutional Neural Network (CNNs) to the widely neglected problem of bangla font recognition by the vision community. First of all, we build up the available large-scale dataset consisting of both labeled synthetic data by Adobe and partly labeled real-world data. Next to CNN is trained to classify images into predefined font classes. Global average pooling layer is proposed instead of fully connected layers over feature maps in the classification layer to correspondence between feature maps and output. Thus the feature maps can be easily interpreted as font categories confidence maps. We show that our method achieves state-of-the-art performance on a challenging dataset of 10 selected bangla computer fonts with 96% line level accuracy. Large-scale experiments show that our approach is exceptionally viable on our synthetic test images and achieves promising results on real world test images.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Identification of Bangla font types continues to flourish in many sectors such as wedding invitations, event invitations, and birthday cards, logo design, cut stone inscriptions, memorial documents, testimonials, and maps and in other written works. Font identification from image can significantly enhance the efficiency of numerous individuals' work. Most importantly, it helps individuals to recognize what their most loved text style. On the other hand, bangla is the mother tongue by Bangladeshi people and uses partly different countries in the world and it also recognized as 7th widely spoken language of the world [1]. Therefore identifying and classifying Bangla font from images is becoming more essential.

Bangla font classification has a wide range of applications. The terms “typography” and “font” are pretty much same. Typography is the visual component of the written word. It holds an essential part in graphic design. Designers should have envisioned how the utilization of textual styles and styling will influence the people [2]. If any designer sees any different kind of font styles anywhere, they would click the picture of it for knowing what kind of font is it which they can use anywhere later. As an example of restaurant, the menu card is made of different font styles which make the menu card attractive for the customers. In advertisements, font identification is greatly needed so that it can easily grab the attention of readers. Newspaper provides the information of the news of the whole world. For heading on the newspaper, different types of fonts are used. Calligraphy is the art of creating decorative handwriting and lettering with a pen or brush. Calligraphers also need to know about different font types so that they can create beautiful and attractive texts.

According to the necessity of Bengali font recognition, the point of this analysis is to assess the execution of deep learning strategies effectively ordering Bengali text styles dependent on visual textual style datasets. The Deep convolutional neural system (DCNN) thought utilized crude data worried on preparing tests. These days, profound learning models are utilized to work out various types of challenges in picture grouping assignments over customary machine learning approach.

A deep convolutional neural network (DCNN) model is suggested in view of our examination to anticipate correctly classify the 10 different Bengali fonts such as sutonyMJ, Adorsholipi, Shonartori, etc. To best our knowledge, this is the first research conducted on visual bangla font recognition utilizing Deep Learning approaches. At last, to assess the execution of models, we utilized an extensive arrangement of standard execution measurements for the most part utilized in the structure of order issues.

1.2 Objectives

The fundamental goal of this examination is to evaluate the execution of deep learning procedures correctly classifying Bengali fonts based on visual font datasets according to the necessity of Bengali font recognition.

The goals of our research are:

- Automatically classify different Bengali fonts by using DCNN model.
- The Proposed Algorithm crops the characters from the taken image and then matches the cropped character with the pre-trained by deep convolution neural network.
- Global average pooling layer is proposed inside the network instead of fully connected layers over feature maps in the classification layer to correspondence between feature maps and output.
- Deep learning models are used to work out different kinds of difficulties in image classification.
- To avoid huge number of computations, we used RGB to Gray color conversion method applied in images for further processing.
- The DCNN network follows 2 Convolutional layers with global average pooling layer. Feature is extracted by the convolution layers through the convolution operation on the input image.
- ReLU is used for no-linearity which is several times faster than their equivalents with tanh units.
- Max pooling layers controls over fitting by minimizing the structual amount of the input features.
- Global average pooling (GAP) layers minimizes over fitting and lessens the complete number of parameters.

- The supervised machine learning technique, deep convolution neural network (DCNN) was employed to classify the visual bangla fonts.

1.3 Motivation

Bangle is the mother tongue by Bangladeshi people and uses partly different countries in the world and it also recognized as 7th widely spoken language of the world. A huge amount of peoples are speaking in bangle but there is a widely neglected problem to identifying bangle font for various purposes like wedding invitations, event invitations, birthday cards, logo design, cut stone inscriptions, memorial documents, testimonials, maps and in other written works. Graphics designers, Calligraphers, Advertisements maker, bangle newspaper Writers, Banner makers are facing difficulties in their work place because of proper knowledge about font Recognition. If any designer sees any different kind of font styles anywhere, they would click the picture of it for knowing what kind of font is it which they can use anywhere later. Therefore identifying and classifying Bangla font from images is becoming more essential for this kind of work place. So our system /proposed method will help them to recognize their most loved text style and font.

1.4 Expected Outcome

In this research we proposed visual Font classification model by using DCNN model which can automatically classify different Bengali fonts. The proposed algorithm crops the characters from the taken image and then matches the cropped character. The model achieves 96% line level accuracy for different styles. DCNN model has improved the precision level from the earliest starting point to the more elevated amount with the increasing of epoch's and finally has reached at the accuracy 96% after fifteen epochs. The performance of deep convolutional neural network (DCNN) was evaluated for the classification of bangla fonts. The DCNN model was trained by 800 (80%) and tested by 200 (20 %) of the total samples images of dataset for the classification task. Substantial scale tests demonstrate that our methodology is especially reasonable on our manufactured test pictures and accomplishes promising outcomes on certifiable test pictures. We have got 96% accuracy in 10 selected bangla computer fonts.

1.5 Report Layout

First chapter contains the introduction, Objectives, Motivation, Expected outcome and Report layout of our thesis. Then second chapter contains Introduction, Related works, Comparative studies and Scope of the problems. Our third chapter contains Materials and methods, Image processing, cropping characters from text and training the network using deep learning. Fourth chapter describes our classification performance assessment and results. Our fifth chapter contains conclusion and future scope. This report contains all about our research work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Text style identification from an image is a difficult issue in many sectors such as wedding invitations, event invitations, and birthday cards, logo design, cut stone inscriptions, memorial documents, testimonials, and maps and in other written works. Programmed textual style acknowledgment can extraordinarily improve the productivity of numerous individual's work. So we emphasize on identifying and classifying Bangla font from images by CNN. We have done the comparative studies of different works that were done in many languages but this is the first work on Bangla language. We have to face many challenges for this work. The machine was needed to be learnt about the font previously, then by taking real world images and cropping characters from those text images; our system could identify the font.

2.2. Related Works

A simple framework based on CNN is presented in [3] on Arabic letters. The author has used CNNs for classifying font or vast document images into script. In this structure, they have prepared a CNN to characterize little picture patches into textual style classes. Utilizing CNNs, they classified extensive pictures into content or text style classes. They compared two CNN architectures of AlexNet composed of 5 convolution layers followed by 3 fully connected layers [4]. The other architecture is the state-of-the-art ResNet-50 [5], which was used to win the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). They achieved 98.8% text line accuracy on the King Fahd University Arabic Font Database (KAFD) for 40 type faces in 4 styles and 10 different sizes.

An epic technique dependent on deep learning for perceiving the textual styles of writings in common pictures of Chinese language is proposed in [6]. From a lot of common pictures containing writings in different font and styles, and by synthesizing those texts, they are utilized to prepare the deep neural system for text style acknowledgment. They have modified two famous CNN models AlexNet and VGG16.

The DeepFont system by studying the Visual Font Recognition (VFR) problem is developed in [7] based on English font. They developed large-scale VFR dataset, named AdobeVFR, which consists of both labeled synthetic data and partially labeled real-world data. The introduction of SCAE-based domain adaption helped their trained model to achieve a higher than 80% top-5 accuracy.

K-Nearest Neighbor (kNN) classification model is used as a baseline in [8] for Chinese language and VLFeat library for feature extraction. In training part they used SIFT descriptors for extract and labeling the character. They kept each neuron active with probability using dropout which has been shown to be an effective anti-overfitting technique so that features can be automatically selected by the network. They focused on both large-scale and small-scale features and tried to make different number of Convolutional layers within each block.

The English word level font transfer problem is tried to be solved in [9] by developing convolution Recurrent Generative model. They mainly focused on font to font translation of document images of printed words using cGAN (Conditional Generative adversarial networks which is more appropriate .They develop a novel Convolutional Recurrent Generative Adversarial Network Architecture for Image to Image translation problem .They showed the potential of their structure in word level Font-to-Font translation problem and used base line methods for font-to-font translation task. They applied batch normalization after each layer in the Encoder and the Decoder network and used leaky ReLU instead of normal ReLU in the network architectures. The main success of their work is to indicate the ability in the font translation task.

Support vector machines (SVM) is used to identify various fonts of English Language in [10]. They used Gabor filters to extract the featured vectors and the proposed SVM is prepared utilizing these highlights. They used six frequently used English fonts such as Times New Roman, Arial, Comic Sans MS, Courier New, Algerian, and Tahoma were combined with four styles regular, bold, italic and bold italic were considered for developing the model. The value of parameters Gamma and C in the SVM formulation are also changed to achieve the best results. They used Radial Basis function Kernel for feature mapping. And the SVM classifier shows an average accuracy of 93.54%.

From a given set of familiar fonts, typeface, weight, slope and size of the text taken from English language is identified in [11]. They called the system ApOFIS (A priori

Optical Font Identification System) and extracted the typo-graphical features from text images by using multivariate Bayesian classifier. They performed the classification among 280 font models. Eight global features have been used by this prototype which is extracted from horizontal and vertical profiles of text lines. And their classifier performed a very good result that is the recognition rate was almost 97%.

A method is exhibited for distinguishing proof and acknowledgment of manually written and typewritten content from archive pictures utilizing hidden Markov models (HMMs) in [12]. It is demonstrated that the logical limitations from the HMM significantly improve the recognizable proof execution over the regular Gaussian mixture model (GMM)-based technique. This sort subordinate way to deal with process the casing test rate and casing width shows noteworthy enhancement in OCR precision over type-independent approaches.

A manually written and printed acknowledgment framework utilizing morphological tasks was proposed in [13]. A lot of conspicuous auxiliary highlights are extracted to decisively recognize one character from the other. The arrangement procedure settles on utilization of a decision tree classifier where at every hub the decision principles are characterized by some morphological tasks. It achieves 95% accuracy for manually written text and for printed text, the accuracy is 99%.

Multi font and multisize Arabic acknowledgment framework was represented in [14] which depends on Hidden Markov Model Toolkit (HTK) and Bernoulli HMMs(BHMMs), that is, HMMs in which regular Gaussian blend thickness capacities are supplanted with Bernoulli mixture probability functions. An accuracy of 98.3% is obtained.

Another skeletonization calculation utilizing changed Block Adjacency Graph (BAG) structure to perceive characters is suggested in [15]. Computational execution on three prevalent text styles and sizes of an Indian content, Telugu, is tried and it is demonstrated that the strategy in fact amplifies the difference between various characters and similitude between same characters in various textual styles. The test results affirm the acknowledgment accuracy as of 97.7% for vowels and 98.92% for numerals.

Multilingual programmed recognizable proof of Arabic and Latin in both manually written and stamped content was suggested in [16]. This method depends on worldwide surface investigation, by separating fractal multi dimensions highlights.

The proposed framework has been tried for 1000 models with different three text style types and sizes. The exactness segregation rate is about 96.64% by utilizing KNN and 98.72% by utilizing RBF.

2.3 Comparative Studies

Author's name	Working Principle	Used ML Algorithm	Language	Classification Accuracy	Year
Tensmeyer. C et al	Classifying font or vast document images into script	Deep Convolution Neural Network	Arabic letters	98.8%	2017
Wang. C et al	Developed DeepFont system by studying the Visual Font Recognition (VFR) problem	Convolution Neural Network based on a Stacked Convolutional Auto-Encoder (SCAE)	English font	80%	2015
Ramanthan. R et al	Identification of various fonts of English Language by using Gabor filters	Support vector machines (SVM)	English Language	93.54%	2009
Zramdini. A et al	Identification of familiar fonts, typeface, weight, slope and size of the text taken from English language	multivariate Bayesian classifier	English Language	97%	1998
Wang. Y et al	recognizing the font styles of texts in natural images	CNN models AlexNet and VGG16	Chinese language	N/A	2018
Huaigu. C et al	Identification of written and typewritten content from archive pictures	Hidden Markov models (HMMs)	English language	N/A	2011

Sukhija. S et al	A manually written and printed framework utilizing morphological tasks	decision tree classifier	English language	95%	2013
Slimane. F	Multi font and multisize Arabic acknowledgment framework	Hidden Markov Model Toolkit (HTK) and Bernoulli HMMs(BHMMs)	Arabic language	98.3%	2011
Lakshmi. C	Finding difference between various characters and similitude between same characters in various textual styles	Block AdjacencyGraph (BAG)	Indian, telugu language	97.7%	2009
Ben Moussa. S	Multilingual programmed recognizable proof in both manually written and printed content	KNN	Arabic and Latin Language	96.64%	2008

2.4 Scope of the Problems

- **Noise removal:**

When a picture is clicked, it gets attached to many unavoidable things which distort the image. All this needs to be removed from the image. And this is removed by using morphological image operation techniques.

- **Blurring:**

Here, we lessen the edge substance and makes the change structure from one color to the other.

- **Reducing Overfitting:**

Convolution layers remove low-level features. Top of the features of the framework, pooling layer is considered as larger amount highlights features in the preparation information that is to progressively lessen the spatial size of the portrayal to decrease the measure of parameters and calculation in the system, and henceforth to likewise control overfting.

- **Cropping Characters from Text:**

The function of content extraction from scene pictures can be partitioned into three stages : text detection to extract text regions from camera-caught scene pictures, content division to binarize content body inside content locales, and content acknowledgment to uncover the content data. Regionprops is a very essential function in Matlab for measuring a variety of image quantities and features in binary image. In particular, it can automatically recognize the properties of each contiguous white region which is 8-connected given a black and white image.

- **Font Characters training:**

First the characters were needed to be trained by DCNN. After that, they were matched with the given fonts.

- **Max pooling layer:**

Max pooling layers controls overfitting by minimizing the structural size of the input features. It uses MAX operation for reducing size spatially.

- **Gap layer:**

Global average pooling (GAP) layers minimizes overfitting and lessens the complete number of parameters in the model which is similar to max pooling layers. This layer reduces the spatial dimensions.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 Materials and Methods

The idea of deep convolution neural system (DCNN) utilized for classifying the textual styles from the features Figure 3.1 Single vector(X) is utilized in our framework and the information layer of the DCNN is considered and pursued by convolution layers. Each neuron acquires some contribution to perform a dot multiplication in convolution layer. Convolution layers in a system are noted as a vigorous example marker of local features. Thus the input vector (X) is fed into the image preprocessing block for removing noises from image.

In the Figure 3.1 shows that one of the dimensional pooling layers trails the preprocessed images of two convolution layers. Convolution layers remove low- level features. Top of the features of the framework, pooling layer is considered as larger amount highlights features in the preparation information that is to progressively lessen the spatial size of the portrayal to decrease the measure of parameters and calculation in the system, and henceforth to likewise control overfitting.

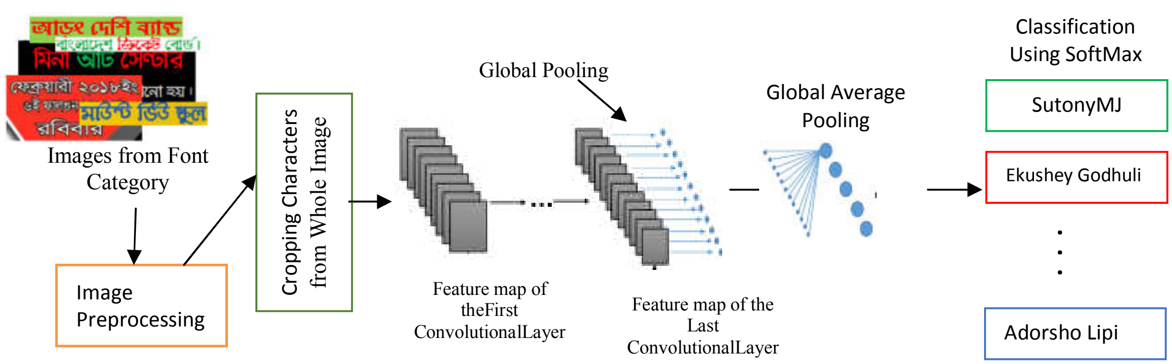

Figure 3.1: Proposed architecture of Feature Extraction and Classification of Bengali Font using DCNN

3.2 Image processing

To avoid huge number of computations, we used RGB to Gray color conversion method applied in images for further processing. In preprocessing, we used the binarization method of the gray scale image which helps in segmentation of text lines. Then Gray scale image is converted into binary image. The conversion of gray scale image $I_{Grey}(x, y)$to binary image $I_{bin}(x, y)$ is conceded out using equation (1).

$$I_{Bin} = \begin{Bmatrix} 1, if\ I_{Grey}(x,y) > Tresold \\ 0, if\ I_{Grey}(x,y) \leq Tresold \end{Bmatrix} \quad (1)$$

The threshold is calculated by using the information of histogram of grayscale image [17].

Noise: Image distortion is most pleasance problems in image processing. When clicking any image, sometimes unavoidable things get attached to it. We have removed the noise using morphological image operation techniques from image during preprocessing.

Blur: We have used Gaussian blur to reduce image noise and reduce detail.

3.3 Cropping Characters from Text

This binary image comes from the preprocessed block contains a group of objects that are divided into different regions. Pixels that fit in to an object are declared as true (1) while those pixels that are the background are false (0). In matlab, bwlabel function provides the membership of each pixel where it will return an integer map of every regions or objects in an image. This membership function helps to understand the background and foreground image region of an entire image. Figure 3.2 explained the procedure of how the sub-image formed from the entire image (Adorsholipi fonts is used for explanation).

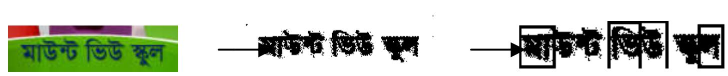

Figure 3.2: Cropped Characters from Text

Regionprops is a very essential function in Matlab for measuring a variety of image quantities and features in binary image. In particular, it can automatically recognize the properties of each contiguous white region which is 8-connected given a black and white image. The properties centroid is used to find out center of mass means the middle of the object is located.

3.4 Training the network using Deep Learning

The DCNN network follows 2 Convolutional layers with global average pooling layer. Feature is extracted by the convolution layers through the convolution operation on the input image (Figure 3.3) size of 32 X 32 of every character in individual fonts (as for Example, AdorshoLipi font). Let us consider at m^{th} plane of the x^{th} layer,the activation of a node at co-ordinate (n,q) be denoted by $Z^{n}{}_{(m,n,q)}$. At plane m of layer x,we denoted the set of input planes as B_m^x. The weight of the connection from $Z^{x-1}_{(h,nT_x+a,qT_x+b)}$ to $Z^{n}{}_{(m,n,q)}$ is denoted by $v^{x}_{(h,m,a,b)}$where $0 \leq a, b \leq Dx \quad 1$, here D_x connects layers x-1 and x.

$$Z^{n}{}_{(m,n,q)} = f\big(\textstyle\sum_{n \in B_m^x} \sum_{0 \leq a,b \leq Dx-1} (V^{x}_{(n,m,a,b)} \times Z^{x-1}_{(h,nT_x+a,qT_x+b)} + {}^{x}_{m}\,\big) \qquad (2)$$

The output can be calculated by the above equation where f is an activation function and ${}^{x}_{m}$ is a bias.

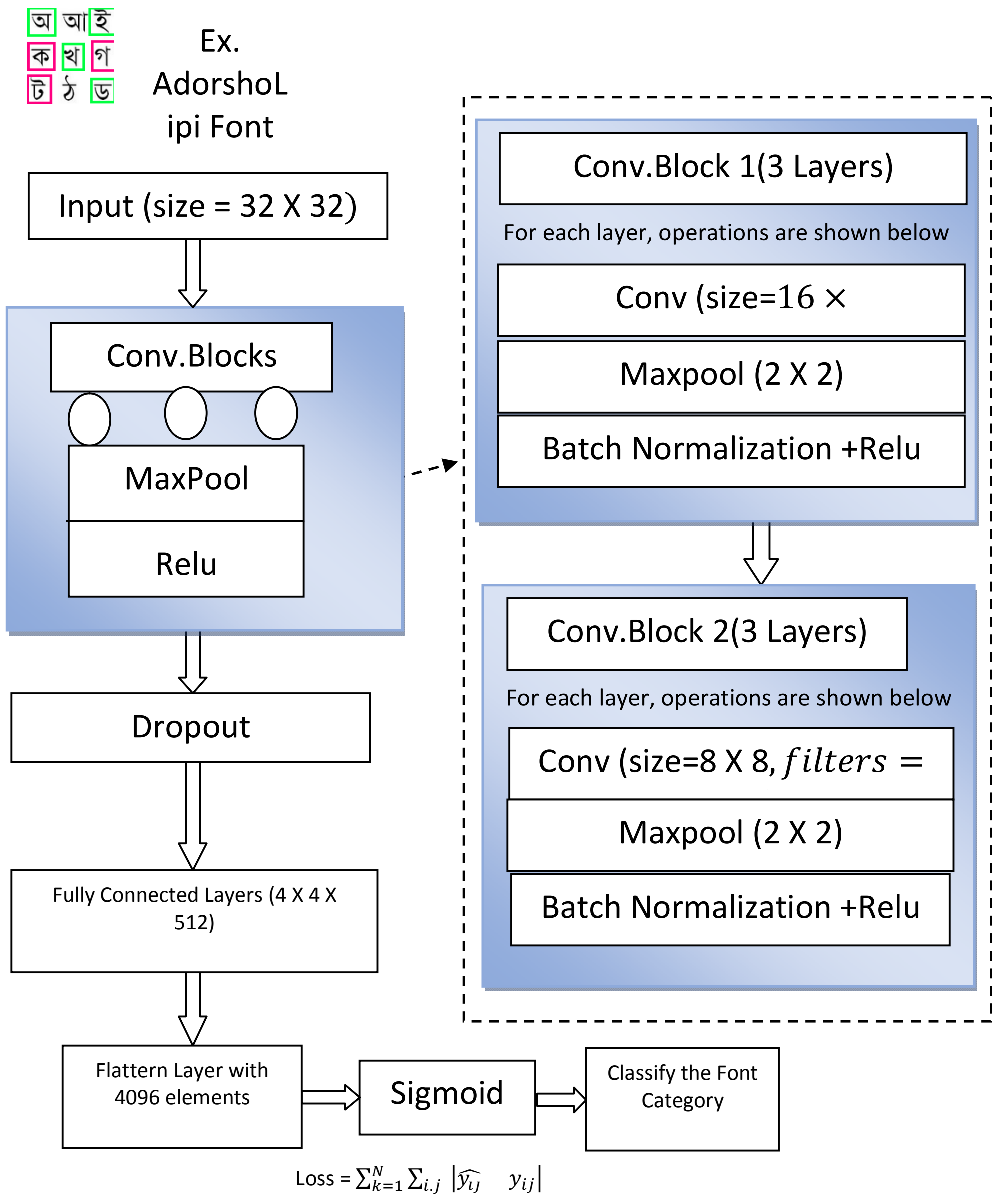

Figure 3.3: Font Characters training by Convolution Neural Network

By equation(2), we can convolute multiple input feature maps in this manner empowering the convolution layers to extricate more elevated amount features from different lower-level features. Here no features are missed if the position is shifted also. We can derive the size of the output feature map by the following equation.

$$widt^{\ x} = (widt^{\ x-1} \quad D_x + 1)/T_x \quad \text{and}$$
$$eig\ t^{x} = (\ eig\ t^{x-1} \quad D_x + 1)/T_x \qquad (3)$$

The standard way to model a neuron's output *f* as a function of its input *p* is with

$$f(p) = \tanh(p) \text{ or } f(x) = (1+ e^{-p})^{-1} \quad (4)$$

In terms of training time with gradient descent, we refer to neurons with this non-linearity by following a batch normalization followed by a ReLU for non-linearity which is several times faster than their equivalents with tanh units. Batch normalization applies this equation to the input.

Max pooling layers controls over fitting by reducing the spatial size of the input features. It uses MAX operation for reducing size spatially. A node at (n,q) is connected to the input nodes in an M*M window whose upper left corner is at (n^T,q^T). Eq(3) indicates that maximum value is selected by each node among the input nodes and Eq(3) does not need any weight.

$$Z^{n}{}_{(m,n,q)} = f\left(\max_{0\le a,b\le D_x} Z^{x-1}_{(h,nT_x+a,qT_x+b)} \right) \quad (5)$$

Global average pooling (GAP) layers minimizes over fitting and diminishes the total numeral of specification in the model which is similar to max pooling layers. This layer reduces the spatial dimensions. GAP layer was used in PatternNet in [18] to discard the location information which leaded to a quicker and more powerful example mining calculation as the component after global pooling is more reduced and can successfully abridge input patches from any area.

A few Convolutional and pooling layers are typically stacked over one another to separate more extract more abstract representations in traveling through the network. The fully connected layers that take after these layers perform the function of high-level reasoning. We can compute the output of each node by Eq (4) where we omit the 2D coordinate(n,q).

$$Z_m^x = f\left(\sum_n \quad v^x_{(h,m)} z_n^{x-1} + \theta_m^x\right) \quad (6)$$

The final layer of sigmoid function in the softmax operation works around by implementing a soft arg max:

$$f_j(Z) = \frac{e^z\ k}{\sum_k e^z\ k} \quad (7)$$

To understand this better, we have trained a network to recognize and classify bangla font from images

CHAPTER 4

RESULTS & DISCUSSION

4.1 Classification Performance Assessment

The supervised machine learning technique, deep convolution neural network (DCNN) was employed to classify the visual Bengali fonts. The performance of the classification model is assessed by the nine evaluation metrics. Here,

- True Positive (TP): Correctly identified the font.
- False Positive (FP): Erroneously identified the font.
- True Negative (TN): Correctly discarded the font.
- False Negative (FN): Erroneously discarded the font.

The evaluation metrics are defined as below:

- Classification Accuracy = (TP + TN) / (TP + FN + FP + TN)
- Error Rate = (FP + FN) / (TP + FN + FP + TN)
- Sensitivity = TP / (TP + FN)
- Specificity = TN / (TN + FP)
- False Positive Rate = FP / (FP +TN)
- False Negative Rate = FN / (FN + TP)
- Precision or Positive Predictive Value = TP / (TP + FP)
- Negative Predictive Value = TN / (TN + FN)
- F1-Score=(2 ×(Sensitivity × Precision)) / (Sensitivity × Precision)

4.2 Results

The performance of deep convolutional neural network (DCNN) was evaluated for the classification of bangla fonts. The DCNN model was trained by 800 (80%) and tested by 200 (20%) of the total samples images of dataset for the classification task. The confusion matrix of DCNN is tabulated in Table 1.

From the Table 1, the diagonal values of the particular font class is considered as true positive (TP). False negative (FN) is the sum of the total values in the corresponding row excluding the TPs and false positive (FP) is the sum of the total values in the corresponding column without the TPs. True negative (TN) is the sum of the total values of all the rows and columns excluding the row and column of specific fonts class respectively.

Table 2 shows the classification result that is calculated from the confusion matrix of DCNN. The indication from the classification results that Adorsho Lipi and Siyam Rupali fonts are forecasted the highest number of true positive (TP) and Adorsho Lipi also forecasted the highest number of true negatives. The highest number of false positive (FP) and false negative (FN) are predicted by Mitra Mono and Sonar Bangla fonts using DCNN.

From the Table 2, the total number of TP, FN, FP and TN are 818,182, 182, 8818 respectively therefore the classification accuracy of DCCN is calculated by the macro-average method is 96.4%.

Table 4.1. The confusion matrix of deep convolutional neural network (DCNN).

	Adorsho Lipi	Amar Bangla	Apona Lohit	Charu Chandan	Ekushe y Godhu li	Ekush ey Sumit	Mitra Mono	Rajon Shoily	Siyam Rupali	Sonar Bangl a
Adorsho Lipi	**90**	1	1	1	1	1	2	0	1	2
Amar Bangla	1	**87**	0	1	3	2	4	0	1	1
Apona Lohit	1	2	**76**	2	2	3	2	4	3	5
Charu Chandan	0	4	3	**71**	4	3	6	2	5	2
Ekushey Godhuli	0	2	3	2	**86**	0	2	0	2	3
Ekushey Sumit	2	4	1	3	2	**79**	2	4	2	1
Mitra Mono	1	4	0	3	6	2	**77**	2	5	0
Rajon Shoily	2	0	5	0	1	0	0	**87**	3	2
Siyam Rupali	0	2	3	1	1	1	2	0	**90**	0
Sonar Bangla	2	0	4	3	2	5	3	4	2	**75**

Table 4.2. The classification results using deep convolutional neural network (DCNN).

	True Positive (TP)	False Negative (FN)	False Positive (FP)	True Negative (TN)
Adorsho Lipi	90	10	9	891
Amar Bangla	87	13	19	881
Apona Lohit	76	24	20	880
Charu Chandan	71	29	16	884
Ekushey Godhuli	86	14	22	878
Ekushey Sumit	79	21	17	883
Mitra Mono	77	23	23	877
Rajon Shoily	87	13	16	884
Siyam Rupali	90	10	24	876
Sonar Bangla	75	25	16	884
Total	818	182	182	8818

Table 4.3. The classification performance of deep convolutional neural network (DCNN).

	CA	ER	Sens.	FNR	Spec.	FPR	PPV	NPV	F1
Adorsho Lipi	.98	.02	.90	.10	.99	.01	.91	.99	.91
Amar Bangla	.97	.03	.87	.13	.98	.02	.82	.99	.84
Apona Lohit	.96	.04	.76	.24	.98	.02	.79	.97	.77
Charu Chandan	.96	.04	.71	.29	.98	.02	.82	.97	.76
Ekushey Godhuli	.96	.04	.86	.14	.97	.03	.80	.98	.83
Ekushey Sumit	.96	.04	.79	.21	.98	.02	.82	.98	.80
Mitra Mono	.95	.05	.77	.23	.97	.03	.77	.97	.77
Rajon Shoily	.97	.03	.87	.13	.98	.02	.84	.99	.85
Siyam Rupali	.97	.03	.90	.10	.97	.03	.79	.99	.84
Sonar Bangla	.96	.04	.75	.25	.98	.02	.82	.97	.78

Classification accuracy (CA), error rate (Error Rate), sensitivity (Sens.), false negative rate (FNR), false positive rate (FPR), positive preditive value (PPV), negative preditive value (NPV), f1-score (F1)Table 3 specifies the detailed classification performance of DCCN model for the recognition of 10 bangla fonts is measured by the 9 evaluation metrics according to micro-average method and classification performance is plotted in Figure 4.1.

Adorsho Lipi is achieved the highest classification accuracy of 98% while second highest accuracy of 97% are achieved by the Amar Bangla, Rajon Shoily and Siyam

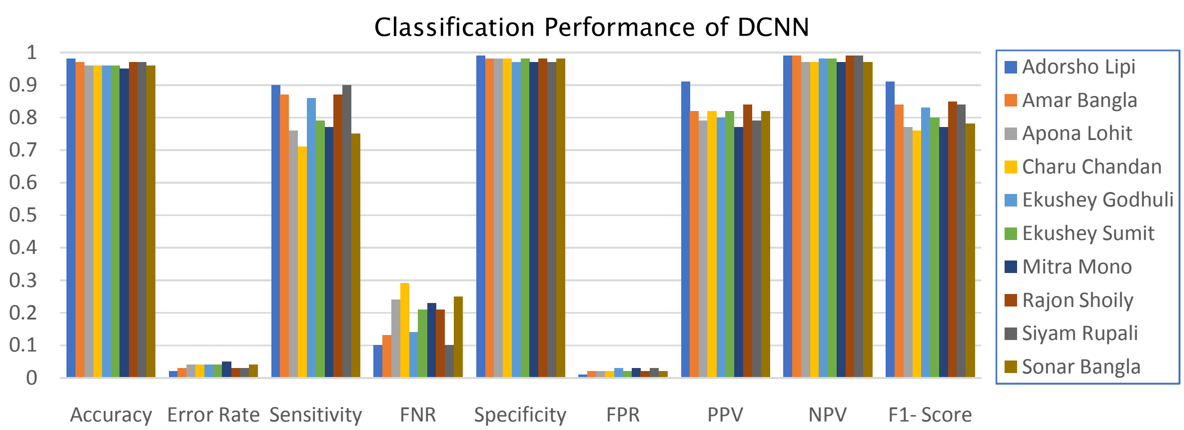

Figure 4.1. The classification performance assessment using of DCNN

Rupali fonts. According to the classification accuracy, Adorsho Lipi is accomplished the highest level of sensitivity, specificity, positive predictive value, negative predictive value and f1-score of 90%, 99%, 91%, 99% and 91% correspondingly. The lowest level of error rate 2%, false negative rate 10% and false positive rate 1% is also performed by the Adorsho Lipi font using the implementation of DCNN model.

From the graph in Figure 4.2, it shows that the requirement of the minimum number of iterations needed for achieving the best classification accuracy to recognize the fonts. DCNN model has improved the level of correction from the very beginning to the higher level with the increasing of epochs and finally has reached at the accuracy 96% after fifteen epochs. Figure 4.3 displays that the DCNN model has decreased the loss level from the higher to the lowest level with the increasing number of the epochs.

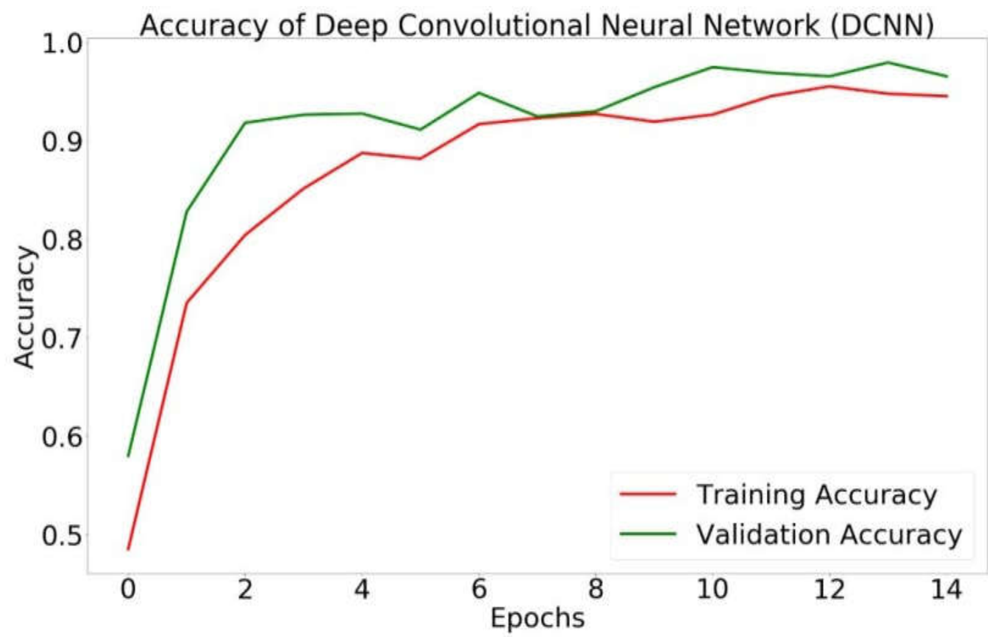

Figure 4.2. Training Accuracy vs Validation Accuracy of DCNN

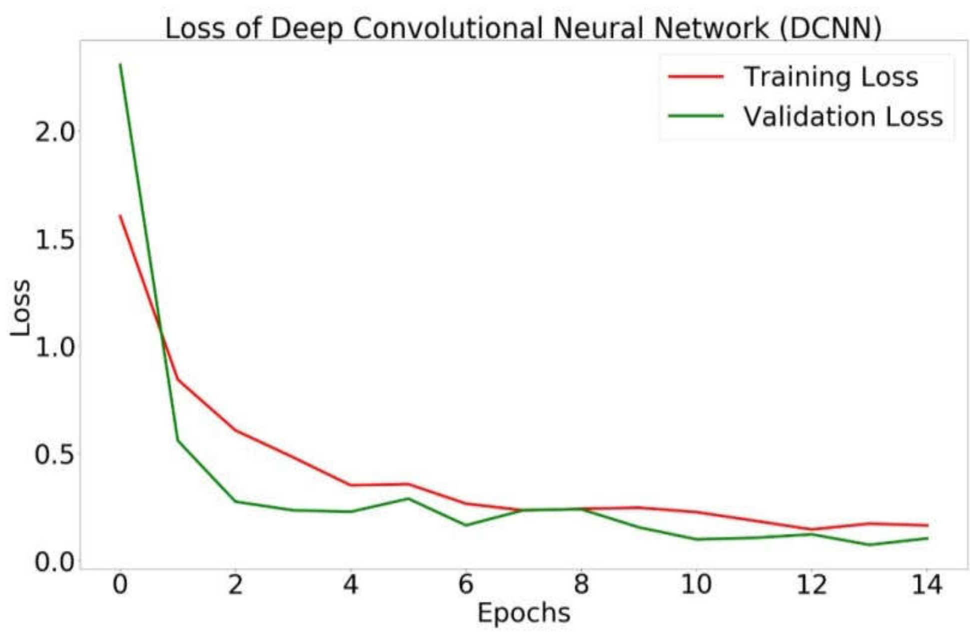

Figure 4.3. Training Loss vs Validation Loss of DCNN

CHAPTER 5
CONCLUSION & FUTURE SCOPE

5.1 Conclusion

In this paper, we propose the visual Font classification model by using DCNN model which can automatically classify different Bengali fonts. The proposed algorithm crops the characters from the taken image and then matches the cropped character with the pre-trained network which is trained by the deep convolution neural network. Global average pooling layer is proposed inside the network instead of fully connected layers over feature maps in the classification layer to correspondence between feature maps and output. The model achieves 96% line level accuracy for different styles.

5.2 Future Scope

The future of image processing will include checking the sky for other smart life out in space. Additionally new keen, computerized species made totally by researchers in different countries of the world will incorporate advances in picture handling applications. Because of advances in picture handling and related innovations there will be a huge number of robots on the planet in a couple of decade's time, changing the manner in which the world is overseen. Advances in picture preparing and counterfeit intelligence will include spoken directions, foreseeing the data necessities of governments, interpreting dialects, perceiving and following individuals and things, diagnosing ailments, performing medical procedure, reconstructing abandons in human DNA, and programmed driving all types of transport. With expanding force and refinement of current registering, the idea of calculation can go past as far as possible and in future, picture handling innovation will progress and the visual arrangement of man can be repeated. Our future work depends on text style comparability. A convolutional neural system (CNN) is prepared for text style acknowledgment and textual style comparability learning. In a preparation stage, content pictures with textual style names are incorporated by acquainting fluctuations with limit the hole between the preparation pictures and genuine content pictures. It may also be used in text style recommendation, text style perusing, or textual style acknowledgment applications.

REFERENCE:

[1] Most Spoken Languages in the world, goo.gl/fhTq2S. Last accessed 20 Aug 2018

[2] Chen, G., Yang, J., Jin, H., Brandt, J., Shechtman, E., Agarwala, A., Han, T.X.: Large-Scale Visual Font Recognition. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 3598{3605 (2014)

[3] Tensmeyer, C., Saunders, D., Martinez, T.: Convolutional Neural Networks for Font Classication. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 985{990 (2017)

[4] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classi_cation with Deep Convolutional Neural Networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems.Volume 1, pp. 1097-1105. Curran Associates Inc., Lake Tahoe, Nevada (2012)

[5] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,770{778 (2016)

[6] Wang, Y., Lian, Z., Tang, Y., Xiao, J.: Font Recognition in Natural Images via Transfer Learning. In: Yingmin Tang., Jianguo Xiao. (eds.) International Conference on Multimedia Modeling 2018, LNCS, vol. 10704, pp. 229-240.Springer, cham (2018).

[7] Wang, Z., Yang, J., Jin, H., Shechtman, E.: DeepFont: Identify Your Font from An Image. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 451{459. ACM, Brisbane, Australia (2015)

[8] Deng, L., Wang, L., Ren, Z.: Chinese Calligraphy Font Classi_cation and Transformation (1999)

[9] Bhunia, A.K., Bhunia, A.K., Banerjee, P., Konwer, A., Bhowmick, A., Roy, P.P,Pal, U.: Word Level Font-to-Font Image Translation using Convolutional Recurrent Generative Adversarial Networks. CoRR, abs/1801.07156, (2018)

[10] Ramanthan, R., Thaneshwaran, L., Viknesh, V., Arunkumar, T., Yuvaraj, P., Soman, D.K.P.: A Novel Technique for English Font Recognition Using Support Vector Machines. 2009 International Conference on Advances in Recent Technologies in Communication and Computing, 766{769 (2009)

[11] Zramdini, A., Ingold, R.: Optical Font Recognition Using Typographical Features. IEEE Trans. Pattern Anal. Mach. Intell. 20(8), 877{882 (1998)

[12] Huaigu, C., Prasad, R., Natarajan, P., 2011. Handwritten and typewritten text identification and recognition using hidden Markov models. In:2011 International Conference on Document Analysis and Recognition (ICDAR), 18–21 September 2011, pp. 744–748

[13] Sukhija, S., Panwar, S., Nain, N., 2013. CRAMM: character recognition aided by mathematical morphology. In: 2013 IEEE Second InternationalConference on Image Information Processing (ICIIP), 9–11 December 2013, pp. 323–328

[14] Slimane, F., Kanoun, S., Abed, H.E., Alimi, A.M., Ingold, R., Hennebert, J., 2011. ICDAR 2011 – Arabic recognition competition: multi-fontmulti-size digitally represented text. In: 2011 International Conference on Document Analysis and Recognition (ICDAR), 18–21 September2011, pp. 1449–1453

[15] Lakshmi, C.V., Singh, S., Jain, R., Patvardhan, C., 2009. A novel approach to skeletonization for multi-font OCR applications. In: Chaudhury, S.,Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S. (Eds.), Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science, vol.5909. Springer, Berlin, Heidelberg, pp. 393–399

[16] Ben Moussa, S., Zahour, A., Benabdelhafid, A., Alimi, A.M., 2008. Fractal-based system for Arabic/Latin, printed/handwritten script iden-tification. In: 19th International Conference on Pattern Recognition, 2008. ICPR 2008, 8–11 December 2008, pp. 1–4

[17] Shejwal, M.A, Bharkad, S.D.: Segmentation and extraction of text from curved text lines using image processing approach. In: 2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC), pp. 1{5. IEEE, Indore, India (2017)

[18] Li, H., Ellis, J.G., Zhang, L., Chang, S.-F.: PatternNet: Visual Pattern Mining with Deep Neural Network. CoRR (2017)

Bengali Font Identification

ORIGINALITY REPORT

15%	4%	15%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Md. Zahid Hasan, K.M. Zubair Hasan, Abdus Sattar. "Burst Header Packet Flood Detection in Optical Burst Switching Network Using Deep Learning Model", Procedia Computer Science, 2018 Publication	3%
2	Najmeh Samadiani, Hamid Hassanpour. "A neural network-based approach for recognizing multi-font printed English characters", Journal of Electrical Systems and Information Technology, 2015 Publication	2%
3	Ankan Kumar Bhunia, Ayan Kumar Bhunia, Prithaj Banerjee, Aishik Konwer, Abir Bhowmick, Partha Pratim Roy, Umapada Pal. "Word Level Font-to-Font Image Translation using Convolutional Recurrent Generative Adversarial Networks", 2018 24th International Conference on Pattern Recognition (ICPR), 2018 Publication	1%