Bangla Handwritten Character Recognition With Multilayer Convolutional Inception Module

***Abstract*—This Handwritten character detection and recognition from a natural image has a large set of difficulties. A systematic approach that can automatically recognize character from handwriting, printed books, road signs and also classifies text and nontext blocks from natural image has many significant applications. For instance, visual assistance for visually impaired people, image understanding, classification of text in image, implementing autonomous navigation system. Recent development of deep learning approach has strong capabilities to extract high level feature from a kernel (patch) of an Image. In this paper we will demonstrate a novel approach that integrates a multilayer convolutional neural network (CNN) with supervised feature learning. Unlike previous OCR methods, this CNN based approach does not focuses on some specific feature extraction. 166,105 Bangla handwritten character of different shapes and strokes has been used to train and evaluate the performance of the model. This approach allows a higher recall rate for the character in an image and outperforms some current methodologies.**

Keywords— Bangla; Handwritten Character; Neural Network; Inception Module

I. INTRODUCTION

RECOGNIZING handwritten text from document and scene-text have received a lot of attention as it allows any system using it to be robust and the use of such models in variable real world Scenarios.In recent times, the field of computer vision has seen great strides towards expert systems that can be trained using large image dataset due to the rise of high performance computing (HPC) systems. High performing parallel processing GPUs has enabled the computer vision researchers to turn the problem of image recognition into a problem of numerical optimizations problem. So now there exists a wide range of algorithms and architecture to tackle such problems. One such architecture to create models and solve those problems are neural networks. It has been designed on the basis such that it imitates some function of the neural pathways of human brain. And to maximize the performance of such architecture, numerous methodologies have been proposed and one such implementation of a system of detector and recognizer that uses machine learning techniques is convolutional neural network or CNN [2], [3],[4]. In CNN, the state of the art detection performance is 80% on F-measure with ICDAR 2011[4]. And for end-to-end recognition, the result is 57% accurate on SVT dataset [5]. The above mentioned works clearly focuses on detection and recognition of English alphabets.

In this paper we will be showing the results of experimentation on different types of convolutional neural network architecture which will be appropriately evaluated on the context of Bangla alphabet datasets like BanglaLekha-Isolated [6]. By analyzing this dataset we will establish a base ground truth for a superior CNN architecture that can be used in handwritten text recognition in Bangla. And ultimately with the chosen CNN architecture we will implement a Bangla character recognizer that will be able to localize and recognize text from any background image. With the experimentation results we determine which architecture is performing better for our current problem.

In order to improve the accuracy of a CNN-based architecture it is imperative to fine tune the hyper-parameters according to the specific needs of the problem domain. And using the power of hyper-parameters is one of the most powerful advantages the neural network architecture, making any problem set to be solved. This general purpose usage of neural networks makes it so powerful that a complex problems in any field can now be solved with ease. And one of the most popular CNN based architecture is called Inception or GooLeNet [7]. This general purpose architecture contains blocks of CNN-based module called Inception Module. We will use modified version of Inception architecture. Using combination of Inception module and CNN layers. Ultimately we will demonstrate the improvement in accuracy for Bangla Handwritten characters with convolutional.

II. BACKGROUND STUDY AND RELATED WORKS

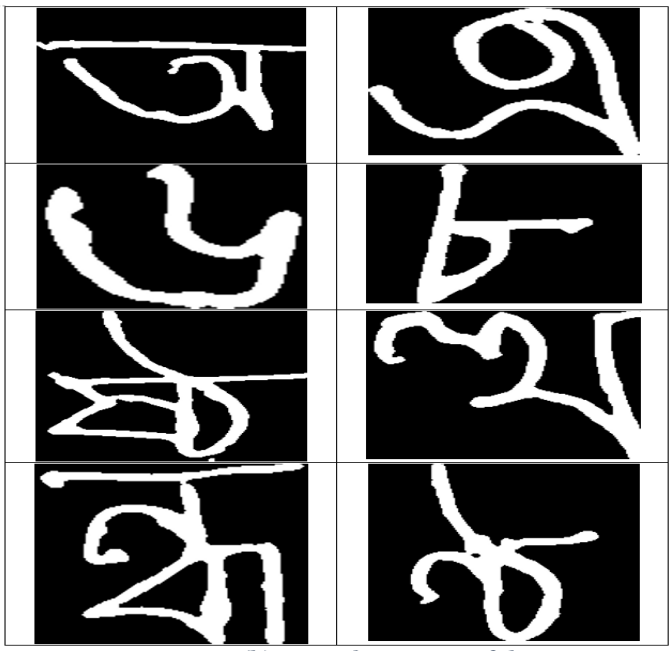
End-to-end text recognition in natural images has two key components (I) text spotting and (II) text recognizing [3]. The main objective is to locate the specific text oriented regions then recognize the actual character. Many research and work have been done for solving the character recognition problem [3]. Much approach and methods are implemented and showed higher performance. However in a complex image locating and recognizing character still remains a problem. A wide range of approaches are used for text segmentation and recognition. In an end-to-end system text detection system identify the text region which is the input of character or word recognizer [3].Then the recognizer recognizes each character in the word or the whole word. Deep convolutional neural network models has higher performance rate for recognizing and computing high level deep features [8]. LeNet a traditional CNN model, powerful for digit and hand-written character recognition [9].In [10] a CNN based approach BHCR-CNN proposed for recognizing Bangla character. Where CNN is employed to classify individual character. In [11] a deep belief network is introduced to recognize Bangla numerals and alphabet. Where an unsupervised feature learning followed by a supervised fine tuning of the network parameters is used.

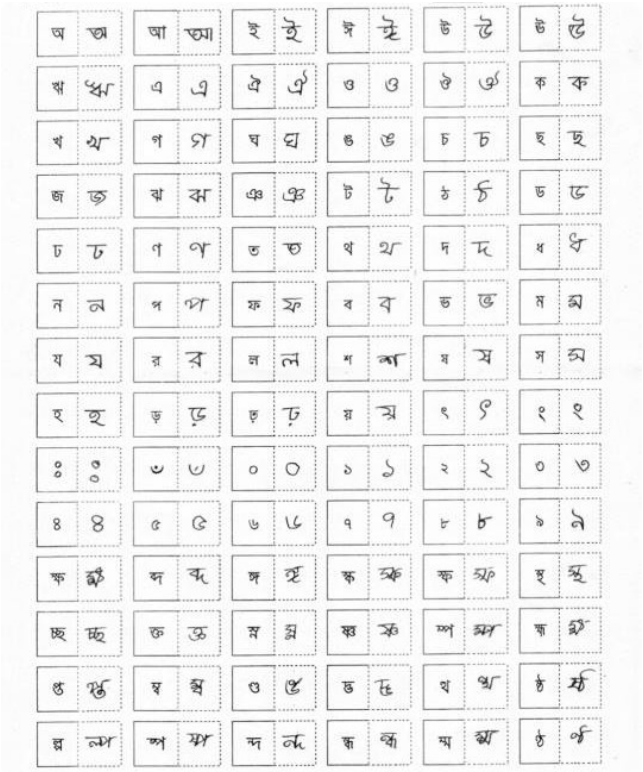
III. HANDWRITTEN BANGLA CHARACTER RECOGNITION USING CNN AND INCEPTION MODULE

Before Main objective of our work is to build a Bangla character recognizer which can automatically recognize word or character from any simple or complex image components. In this section our proposed multilayer inception based convolutional model explains in detail. Following subsection gives a brief idea of dataset, image preprocessing, system architecture and classifier.

A. Dataset

The text recognition subsystem is trained on character-level by using BanglaLekha-isolated [6] dataset. BanglaLekha-isolated has total 1, 66,105 handwritten character images. It consists sample of 84 different Bangla handwritten numerals, basic characters and compound characters [6] which collected from different geographical location of Bangladesh and different age group. In fig. 1(a) collected from [6], a sample of Bangla handwritten character of a particular age is shown. Where first 11 characters are vowel followed by 39 consonant characters then 10 characters are Bangla numerals and rest 24 are Bangla compound character. The dataset provides multiple





labels per character/character group. It contains a wide variation of frequently used in Bangla compound sentences which are very complex shaped. In fig. 1(b) last two character are compound character.

B. Image Preprocessing

As these data from dataset has different image size so it was a bit challenging to train them as we used fixed image size which is 28*28 pixel for each image. Therefore image pre-processing is needed for handling this big size of data. For resizing the images we used python imaging library PIL. Anti-Aliasing filter is used so that quality of the image doesn't degrade while resizing. After setup the train environment we convert images into numPy and save these images in separate class file. Then images of each class are shuffled to have random validation and training set. Also merge them into a single dataset. Duplicate data between training and test can skew the result. So, we need to remove all the overlap between training and test data. We measured the duplicates data and find 1081 samples overlap between valid and train data, 1278 samples overlap between test and train data, 185 samples overlap between test and valid data. Then we create a sanitized test and valid datasets. This data cleaning process gives better accuracy.

C. Model Architecture

Our system architecture consists of several layers of Convolutional Neural Network. As shown in fig.2, the pre-processed data with a size of 28x28 image is taken as an input into the 1st convolutional neural network layer with patch or kernel size of 3x3 and 'same' padding property. After Max Pooling, the output is again fed into a 2nd convolution layer

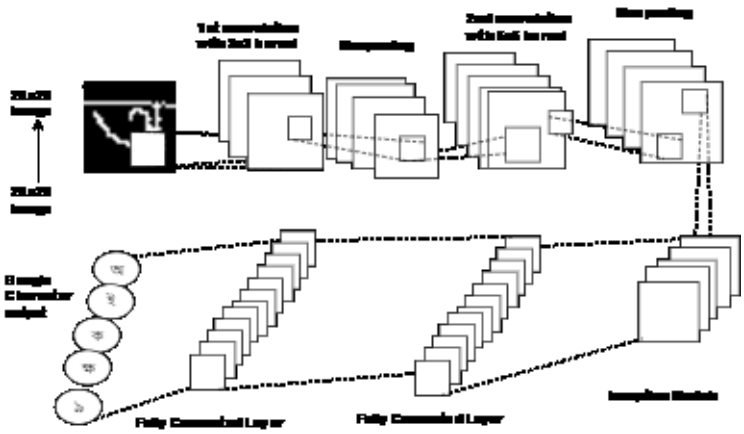
with patch size of 5x5 and 'same' padding. Its output again fed into Max Pooling layer of same configuration as before. The max pooling from the previous layer is used as the input of the Inception module, where CNN is not only stacked sequentially but also parallel. Then output of the inception module is fed into a fully connected network where every single node is connected with each other. This layer flattens the high level feature outputs from the inception module and converts it into a nonlinear combination of these features. After that, another fully connected layer is added to flatten the data again.

D. Classification Using CNN and Inception Module

Classifying Bangla handwritten character has high – dimensional complexity which requires a multilayer, hierarchical neural network [10]. For visual recognition three principle factors are very important for instance, local receptive fields, weight sharing, and subsampling layer which makes a difference in CNN from other simple feed forward neural network[3].In our proposed CNN based approach we used two convolution layer followed by one inception module. Unlike other traditional MLP based method, here CNN itself capture local features from the input image by forcing each filter of CNN layer depends on spatially local patch of the previous layer. Therefore a feature map is generated in next layer. In CNN a great advantage of weight sharing is it significantly reduces the number of free parameter [3] by sharing the same set of edge weights across all features in the hidden layers.
Subsampling or pooling layer is another powerful building block of CNN. The main objective of pooling layer is to reduce the dimensionality of the response map created by convolution layer [3]. And also allows translation invariance for instance, rotation, scale and other distortions into the model. In our approach we used an inception module on top of convolutional layers. This module works as filter inputs of multiple convolution which works on same input. For instance, a 1x1, 3x3, and 5x5 convolution layer works on same input and also employ pooling layer at the same time. This module improves the performance of overall model by extracting multi-level feature In fig.3 shows structure of convolutional neural network used for Bangla character recognition. Two convolutional layer with 3x3 and 5x5 receptive fields along with two max pooling and ReLu layer. For instance, a 28x28 pixels image is input to the CNN. Applying 1st CNN layer on input produced first level feature maps. These feature maps are produced such a way that it can extract a variation of local features where distinct patches with different weights and biases from other patches are used. .

IV. TRAINING EXPERIMENT ANALYSIS AND RESULTS

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly



created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

E. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) For author/s of only one affiliation (Heading 3): To change the default, adjust the template as follows.

a) Selection (Heading 4): Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Deletion: Delete the author and affiliation lines for the second affiliation.

2) For author/s of more than two affiliations: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the "Columns" icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Highlight author and affiliation lines of affiliation 1 and copy this selection.

d) Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

e) Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select "2 Columns". If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

F. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include ACKNOWLEDGMENTS and REFERENCES, and for these, the correct style to use is "Heading 5." Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract," will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1," "Heading 2," "Heading 3," and "Heading 4" are prescribed.

G. Figures and Tables

1) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1," even at the beginning of a sentence.

TABLE I. TABLE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a.] Sample of a Table footnote. *(Table footnote)*

[b.]

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi resolution TIFF or EPS file with all fonts embedded) because this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

Fig. 1. Example of a figure caption. *(figure caption)*

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization," or "Magnetization, M," not just "M." If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization (A (m(1)," not just "A/m." Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)," not "Temperature/K."

ACKNOWLEDGMENT *(Heading 5)*

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g." Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.