Isolated Bangla Handwritten Character & Digit Recognition using Convolutional Neural Network

Submitted by

Name	ID
Mujadded Al Rabbani Alif	13301066
Sabbir Ahmed	13301109
Aleo Aninda	13301113
Tanoy Kumar Das	13301123

Supervisor:

Moin Mostakim

BRAC UNIVERSITY

Inspiring Excellence

Department of Computer Science and Engineering

BRAC University

Declaration

We, hereby declare that this is an original report written by us with our findings, and has not been published or presented in parts or as a whole for any other previous degree Resources and materials by other researchers used as guidelines for our research are carefully mentioned in reference citations.

Signature of the Authors:

Mujadded Al Rabbani Alif

Sabbir Ahmed

Aleo Aninda

Tanoy Kumar Das

Signature of Supervisor:

Moin Mostakim
Lecturer
Dept. of Computer Science and Engineering
BRAC University

Acknowledgment

We would like to sincerely thank our supervisor Mr. Moin Mostakim, Lecturer, Department of Computer Science and Engineering, BRAC University, for his valuable guidance, suggestions and encouragement to complete our work. Also, we would like to thank Dr. Sifat Momen, Assistant Professor and Head, Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, for kindly allowing us to utilize the 'BanglaLekha-Isolated' dataset for the purpose of this thesis, which has helped us achieve our objectives. Finally, we thank our friends and family members, who have been our source for inspiration and encouragement during the last four months of thesis research, without whom this work would not have been possible.

Contents

List of Figures

List of Tables

Outlines

This paper consists of total seven chapters, which has been outlined below. In each of these chapters, there exists one or more sections that explains the specific part of that chapter, which has also been described below.

1. **Chapter 1**

   Describes the source of our inspiration and motivation, the challenges that we face, and the goals the we set to achieve from the beginning of this venture.

   1.1. Provides a background on Bangla language, it's demography, and its linguistic domain.

   1.2. Explains the challenges and difficulties in converting handwritten Bangla characters and digits to digitized formats.

   1.3. Explores the opportunities that would be available with the implementation of paper

   1.4. Clearly stating our goals for this thesis and what we want to achieve

2. **Chapter 2**

   Analysis of all the related works in the field that intersects with our objectives

   2.1. Describing previous implementations and their methods

   2.2. Analysis of latest state of the art systems and it's key component

3. **Chapter 3**

   Detailed description of the architecture and its sub components for the model

   3.1. Description the dataset that was used to train the model

3.2. Outlining the proposed strategy on how to implement it

3.3. Describing how data is pre processed before giving feed into the model

3.4. Showcasing Resnet architecture and how it can help us achieve our goal

3.5. Explaining Adam Optimization and its importance

4. Chapter 4

Elaborate description of the experimental procedures before and after training

4.1. Detailing the experimental process for our model

4.2. Analysis of experimental results with relevant graphs

5. Chapter 5

Comparisons and analysis of experimental results between different models and verifying its efficiency

5.1. Step by step comparisons of results of accuracy, obtained from models with different layers and modifications.

5.2. Evaluation of Confusion Matrix to determine if the model gives uniform accuracy for all classes.

6. Chapter 6

Concluding our thesis paper

7. Chapter 7

Outlining our limitations and expectations for future

7.1. Description of the obstacles and limitations that we faced during thesis

7.2. Our future endeavours on the current thesis

ABSTRACT

This paper proposes a mechanism of Handwritten Letter and Digit Recognition (HLDR) to decipher images of Bangla handwritten characters into electronically editable format, which holds an important role in augmenting and digitalizing many analog application, which will not only paves the way to further research but also have many practical applications in current times.The mechanisms of HLDR has been studied broadly in the last half century, moreover, the rapid growth of computational power and main memory breaks the barrier and gives the opportunity for the implementation of more efficient and complex HLDR methodologies, which creates an increasing demand on many forthcoming application domains. In the field of pattern recognition one of the most productive way of achieving higher accuracy or lower error rate is to adopt an architecture that is deep, optimized and can process a large number of data. Therefore, this paper propose that using deeper residual network [1](ResNet) architecture and recently released Bangla-lekha dataset [2], we can achieve a result which is higher than any research that has been done before.

Chapter 1

Introduction

1.1 Overview

Bangla is the second-most popular language in the Indian subcontinent, and being the language of 220 million people of this region and 300 million from all over the world, Bangla holds the position of being in the fourth place among the world's most common language. Moreover, it is the first language of the people of Bangladesh, which is also part of the greater Indo-European Language family, originating its primary roots from Sanskrit and kept evolving by the absorption of foreign words over thousands of years since its existence. The modern version of the language holds a total 50 basic alphabets, and complementing them are 24 compound characters, along with 10 distinct digits.

1.2 Problem Definition

Over the last several decades, distribution of information has shifted from handwritten hard copy documents to digital file formats, which is more reliable and durable. However, with the transition to new forms of document handling, a large portion of the older documents are still archived in handwritten forms. This is especially the case with our government trying to convert their existing archives, written in Bangla, into digital formats. The problem lies in attempting to convert them, where traditional methods relies on manual typing to copy an existing archive. This slow and unimaginative process can require a lot of time to analyze the documents, and a considerable manpower to make accurate copies of each and every document. The problem reaches to peak when trying to comprehend the writing style of the handwritten documents, since everyone has a unique approach to Bangla handwriting. Furthermore, Bangla characters have a complex arrangement of curvatures with distinct forms of compound characters that complements other basic characters.

1.3 Motivation

This problem has inspired us to create a solution for identifying handwritten Bangla texts, by using a model that is trained with machine learning algorithms, which can classify handwritten Bangla alphabets from images of documents. This system offers an alternative solution to the traditional method of transcripting handwritten Bangla documents, and reduces the manpower, associated costs, and the time required for the whole process. Furthermore, there are large numbers of potential applications of Bangla HLDR, such as, Bangla traffic number plate recognition, automatic postal code identification,extracting data from hardcopy forms, automatic ID card reading, automatic reading of bank cheques and digitalization of documents etc. We believe that this system will ultimately complement other systems that will allow governments and organizations alike to increase their efficiency of document handling, saving storage space, and ensuring the security of documents.

1.4 Objectives

Our goal is to classify and recognise handwritten Bangla digits and characters from sample images of handwritten isolated words, for which we have decided to classify all the 50 simple characters, 24 compound characters and 10 digits, with 84 classes in total. Covering the entire domain of Bangla language family is a herculean task that requires the proper utilization of complex and efficient machine learning algorithms, and top tier hardwares to train a model within the least amount of time. For us, it is vital that the model achieves the minimum expected value accuracy compared to the last generation of state of the art systems on Bangla handwritten character and digit recognition.

Chapter 2

Literature Review

2.1 Early Works

In the field of classification and identification using machine learning algorithms, several state of the art implementations on Bangla handwritten character and digit classification have been presented by various scholars, where most of the early systems relied on standard shallow learning methods like feature extraction and Multilayer Perception (MLP) techniques. Among them, some of the most prominent work includes the works of A. Roy, Bhattacharya, B. Chaudhuri and U. Pal, who are the pioneers in the field of Bangla handwritten character and digit classification, and set a high standard for future scholarly works and implementation.

Bhattacharya et al. [1] proposed a two-stage recognition scheme, which consisted of 50 classes of basic Bangla characters, where a rectangular grid consisting of regularly spaced horizontal and vertical lines is overlaid on the character bounding box and feature vector for the first classifier is computed, where the response of this first classifier is analyzed to identify its confusion between a pair of similar shaped characters. A second stage of classification resolves the confusion, and another rectangular grid is overlaid over the character bounding box to compute the feature vector, but this time the rectangular grid consists of irregularly spaced with horizontal and vertical lines over the character bounding box. In both the stages, they used Modified Quadratic Discriminant Function (MQDF) classifier and MLP as classifiers respectively. Basu et al. [2] proposed a different architecture, which uses a hierarchical approach to segment characters from words and MLP is used for classification. In segmentation stage, they reduced character patterns into 36 classes merging similar characters in a single class and used three different feature extraction techniques. Bhowmik et al. [3] proposed a fusion classifier for 45 classes, using Multilayer Perceptron (MLP), RBF network and SVM, which uses wavelet transformation for feature extraction from character images, while in classification, they considered some similar characters as a single pattern and trained the classifier for 45 classes.

2.2 Convolutional Neural Network

The advent of Convolutional Neural Network (CNN) ushered new opportunities in the field of machine learning for high precision classification, which is helping numerous researchers to implement their state of the art system in solving various real world problems with optimum accuracy. Following the path of the pioneers, several researchers like S. Roy, N. Das, M. M. Rahman, M. A. H. Akhand have proposed their state of the art systems for Bangla handwritten digit and character classification using CNN, which have surpassed the value accuracy of previous non-CNN based systems on Bangla handwritten digit and character classification, and proved the effectiveness of the Convolutional Neural Network.

S. Roy et al. [4] proposed a deep learning technique for the recognition of isolated Bangla handwritten compound characters, with a greedy layerwise training of Deep Convolutional Neural Networks (DCNN) in a supervised fashion, while augmenting the training process with the RMSProp algorithm to achieve faster convergence. The results were then compared with those obtained from standard shallow learning methods with predefined features, as well as standard DCNNs, and showed that supervised layers trained with DCNNs outperforms previous standard shallow learning models such as SVMs, as well as regular DCNNs of similar architecture of its generation. M. M. Rahman et al. [12] proposed another novel CNN based Bangla handwritten character recognition, which first normalizes the written character images and then employs CNN to classify individual characters, and also does not employ any feature extraction method like early related works. The model employs two convolutional layers and two subsampling layers in between, and the experimental results revealed that their proposed CNN based method outperformed previous non-CNN based systems.

Chapter 3

Architecture

3.1 Dataset:

Bangla handwriting recognition is becoming a very important issue nowadays. It is potentially a very important task specially for Bangla speaking population. BanglaLekha-Isolated is a extensive Bangla handwritten character dataset [5]. This dataset contains Bangla handwritten numerals, basic characters and compound characters.Bangla Lekha isolated is a huge dataset with 84 classes where 50 class of basic letters ,10 class of numeral and 24 classes are for oftenly used compound letters(Figure 3.1).

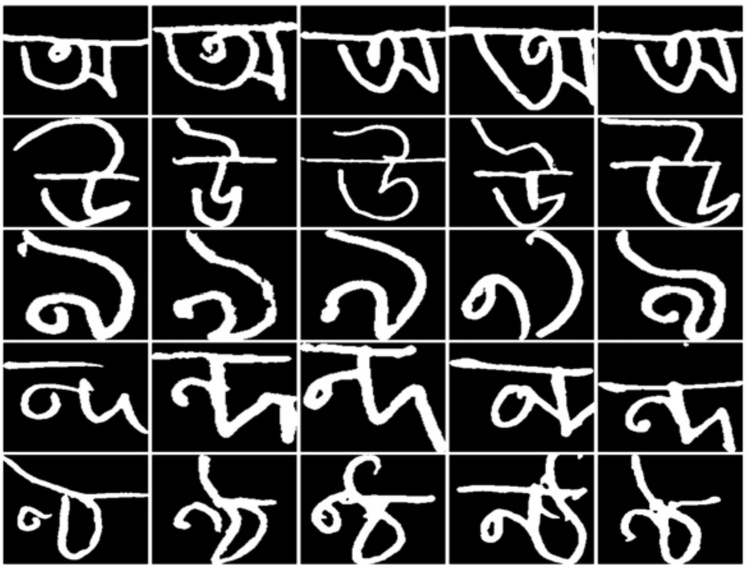

Figure: 3.1 Classification of Dataset

This dataset collects from various region with age range 4 to 27. This dataset is very much different from other dataset.

Dataset	Basic letter	Numeric	Compound letter
CMATERdb 3.1.3.3[1]	15,103	6000	42248
ISI [3]	30966	23299	None
Bangla Lekha	98950	19748	47407

Table: 3.1 The chart of results we found after validating dataset

Moreover, there has been a lot of success in the automatic recognition of handwritten English content [6], the state of automatic Bangla handwriting recognition task is lagging far behind. Recent trends have shown that using machine learning, preciously deep learning techniques, can be very effective in tackling handwriting recognition tasks. However, such learning mechanisms usually require large quantities of labelled data. BanglaLekha-Isolated, the dataset[5] provide such opportunity to improve the lacking.

3.2 Proposed Strategy:

A Convolutional Neural Network (CNN) is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image. This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. Another benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. For example, the current best error rate on the MNIST digit-recognition task (<0.3%) approaches human performance [7]. However, recognition of Bangla Handwritten characters are far to different from that because of large number of classification and different shapes of writing. To properly recognize Bengali characters we had to choose a proper Convolutional model that is efficient and can easily be optimized. Moreover, the dataset had to be slightly preprocessed and to achieve our desired result.On the other augmented hand, the dataset was also while being trained narrow down overfitting and achieve an optimized result. The following subsection gives brief description of each steps

3.3 Data Preprocessing:

The bangla lekha-isolated dataset is already prepossessed with its inverted foreground and background ,removal of noise with median filter, edge thickening filter and also being resized in square in shape with appropriate paddings [5]. The forms are scanned and each handwritten character is extracted automatically and the extraction is verified manually As the dataset is envisioned to be used for machine learning/pattern recognition tasks, the background was converted to black and the characters samples were converted to white. A median filter was employed to reduce image noise, which was followed by the application of an edge thickening operation to bring clarity to the images (Figure 3.2).

Figure: 3.2 Character Extraction

However, to achieve a more certain level of accuracy we normalized the dataset by dividing all images with 255:

$$\mathbf{Out(i,j)=In(i,j)/255.0}$$

The dataset is also augmented during the training in real time.The augmentation was done in dataspace using elastic distortions [8] by width and height shifting. The range was kept 0.4 for this shifting. Once again an underscore separates the two parts. Thus, the filenames can be used to infer the character whose image the file contains as well as the age and gender of the person who wrote it in the form. For convenience, the image files have been organized by character in folders, i.e. 84 folders, one folder per character.

3.4 Resnet Architecture:

Unlike traditional sequential network architectures such as AlexNet, OverFeat, and VGG, ResNet is instead a form of extraordinary architecture, that relies on micro-architecture modules (also called “network-in-network architectures”). The term micro-architecture refers to the set of blocks used to construct the network. A collection of micro architecture blocks leads to the macro-architecture. Resnet of 2015 [9] used an exceptional architectural procedure, where unlike AlexNet or VGGNet it used a very small architecture refers to “building-blocks” to construct the network. Although ResNet is much deeper than VGGNET16 and VGGNET19, the network size is smaller as it’s usage of global-average-pooling rather than fully-connected layers. The ResNet model is one of the best CNN architecture that we currently have and is a great innovation for the idea of residual learnin. We have choose the resnet-34 to be the architecture for our network, and achieved our desired result with it.

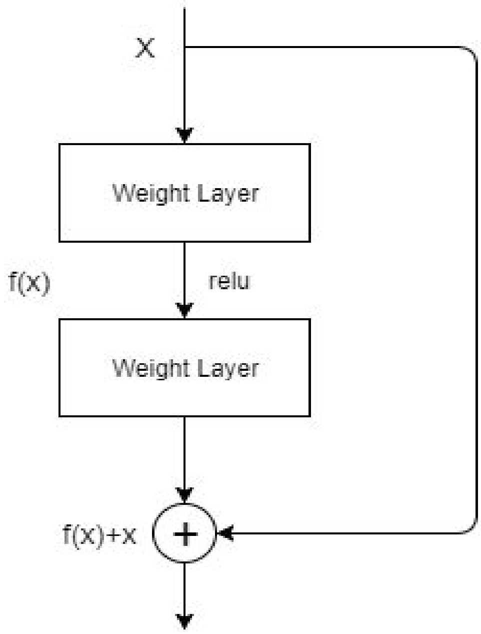

Figure: 3.3 Resnet Architecture

In this paper, we address the degradation problem by introducing a deep residual learning framework. However we apply underlying mapping with a nonlinear H(x) function from input to output. Lets we stac nonlinear layers fit another mapping of H(x), instead of function F(x) which defined as H(x)-x. At the output of the second weight layer we arithmetically add x to the F(x) and pass F(x)+x (Figure 03) through Rectified Linear Unit (ReLU). This enables us to carry important information in the previous layer to the next layers. By doing so we can prevent vanishing gradient problem. Even if this connection looks like an addition to standard CNN approach, surprisingly, it fastens the training of the network. we used batch-normalization only on top of the traditional layers. we get experimental evidence that training with residual connections accelerates the training of networks significantly[10].

3.5 Adam Optimisation:

Stochastic gradient-based optimization is of core practical importance in many fields of science and engineering. Many problems in these fields can be cast as the optimization of some scalar parameterized objective function requiring maximization or minimization with respect to its parameters. We propose Adam, a method for efficient stochastic optimization that only requires first-order gradients with little memory requirement. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients; the name Adam is derived from adaptive moment estimation. Adam optimisation is a memory efficient and faster computing optimization function [8], which is based on adaptive estimates of lower-order moments.Adam is short for Adaptive Moment Estimation which is an update of RMSProp optimiser. The adam optimisation function is a sum of errors over training examples, and training can be done on individual examples or mini batches. It is in some way as same as in stochastic gradient descent (SGD), which is often more efficient than batch training because parameter updates are more frequent [11]. Gradient descent is a way to minimize an objective function J(θ) parameterized by a model's parameters θ∈Rd by updating the parameters in the opposite direction of the gradient of the objective function ∇θJ(θ) w.r.t. to the parameters. The learning rate ηη determines the size of the steps we take to reach a (local) minimum. Stochastic gradient descent (SGD) in contrast performs a parameter update for each training example x(i) and label y(i):

$$\theta=\theta-\eta\cdot\nabla\theta J\ (\theta;x(i);y(i))$$

Batch gradient descent performs redundant computations for large datasets, as it recomputes gradients for similar examples before each parameter update. SGD does away with this redundancy by performing one update at a time. It is therefore usually much faster and can also be used to learn online. SGD performs frequent updates with a high variance that cause the objective function to fluctuate heavily.

Chapter 4

Experiment

4.1 Experimental Process:

Recognition of Bangla Handwritten characters are far too different from other language. To solve this complex problem we address the degradation problem by introducing a deep residual learning framework. Our purpose is to classify Bengali alphabets, numeral and compound characters with a single classifier. We have experimented with some famous and also custom architectures to achieve best possible solution. We used 80:20 as the train and test dataset ratio on 84 classes.We completed our experiment on core i3(gen) with 12 gb RAM and 4gb Vram on linux machine. This method is implemented in keras written in python with theano-gpu.

4.2 Experimental Results:

We have experimented several network on our system to optimize results. Experimental results of the proposed recognition scheme have been collected based on the samples of the prepared dataset from Bangla-Lekha. We have performed batch wise training in this study due to large sized training set. The number of Batch Size (BS) is considered as a user defined parameter. On the other hand, Learning Rate (LR) is also an element that influences learning. At first, we tried with Resnet 34 with the dataset image size of 32x32 pixel values which gave us the loss of 0.59 approximately on 30 epocs.

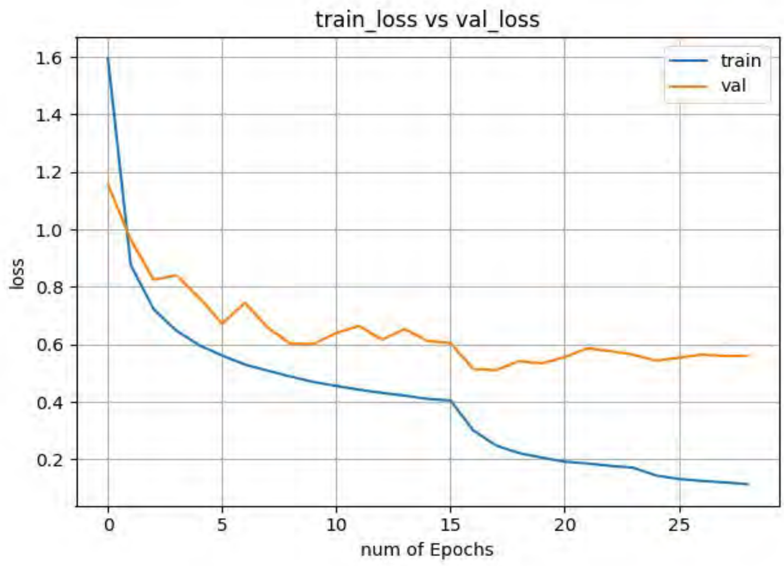

Figure: 4.1 A graph to show value loss in Resnet 34

Then we tried the augmentation of dataspace using elastic distortions [8] which decreased the loss to 0.4.

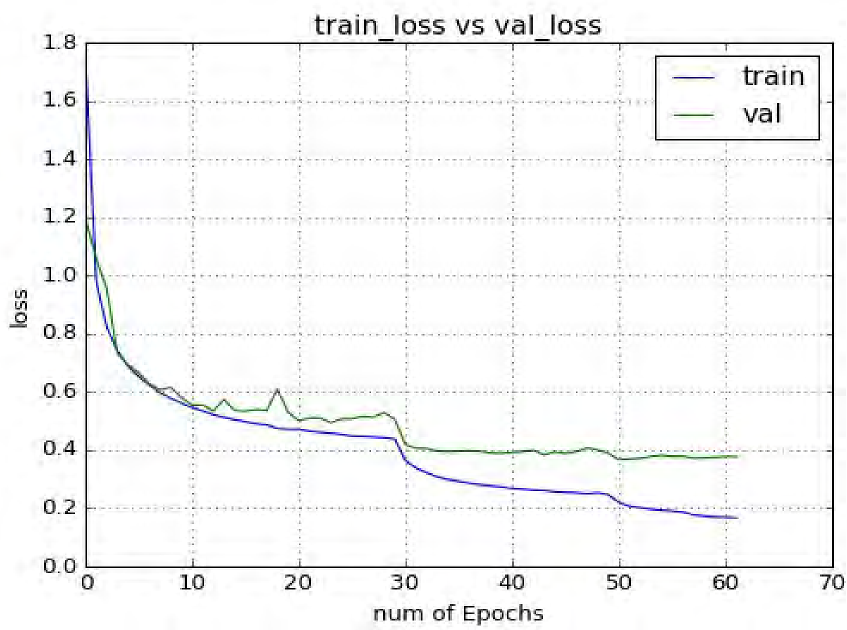

Figure: 4.2 A graph to show the decrease of loss using augmentation

After that we tried more augmentation on the real time training but using a smaller network of resnet 18. Although it the network size was small it gave us a improved result in test loss of 0.35.

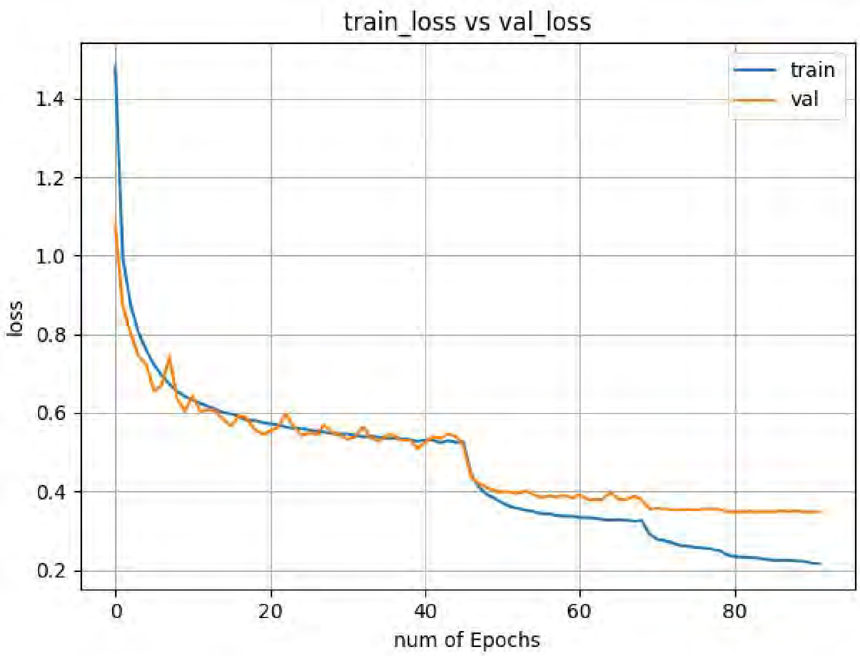

Figure: 4.3 A graph to show achieving efficiency in resnet 18

Then we tried decreasing maxpool to 2x2 from the default value of resnet 3x3 but we kept the default stride of 2x2. This network architect gave us 0.36 test loss. We tried removing the stride too which gave the same loss but increase in accuracy. After we updated the image size to 64x64 which gave us the decreased test loss of 0.29.

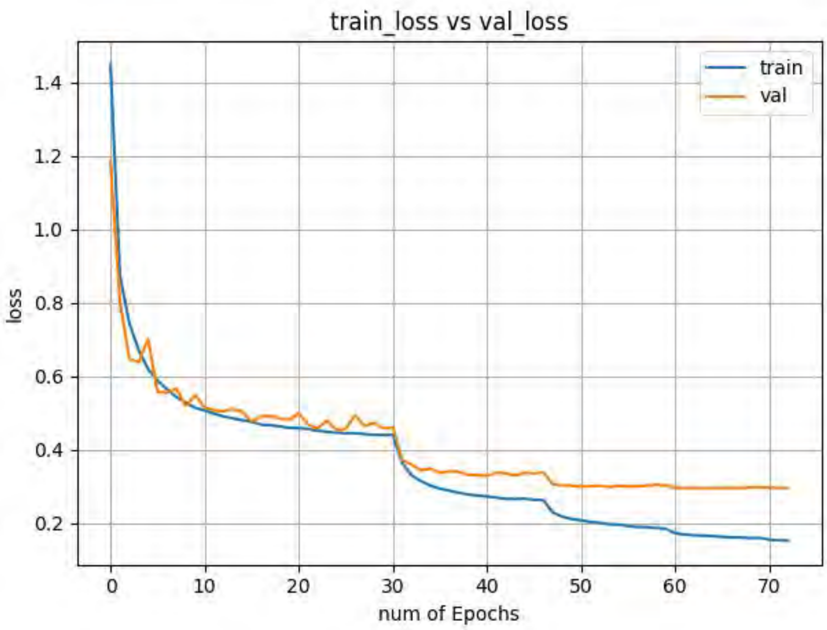

Figure: 4.4 A graph to show achieving efficiency in resnet 18 using maxpool

Further, we tried excluding the maxpool layer from resnet which gave a slight increase in accuracy and test loss decreased to 0.28. Moreover, we tried increasing the the image size even more to 110x110 and it gave a improved result of 0.27 loss. Previously we only use 32*32 pixel image size, it provides us accuracy around 93%. But after increasing the image size to 64*64 pixel it shows much better accuracy around 94.3%-94.4%. Again we increase image size. It increased the computational time of the program. We set the image size as 110*110 pixel and again switch to resnet-34, the it provides us the best result of 94.8%,

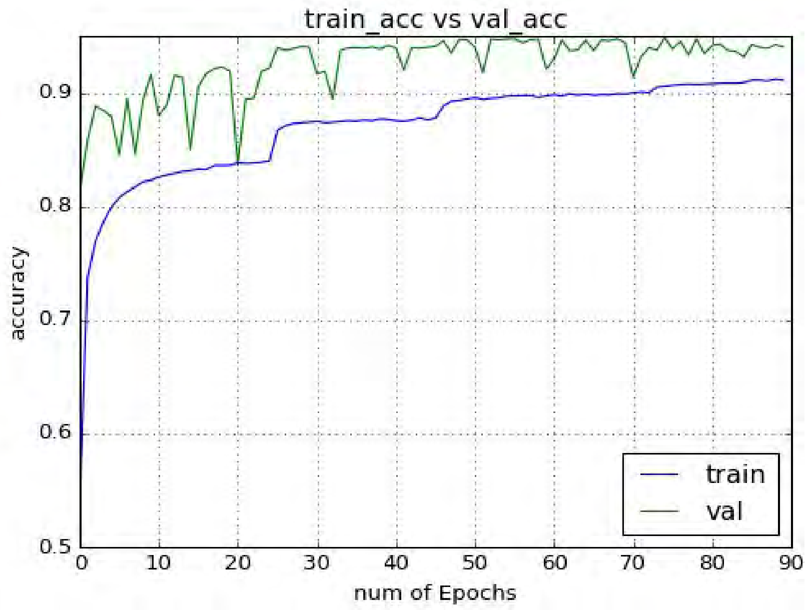

Figure: 4.5 A graph to show achieving better accuracy using Resnet 34

Finally we moved to Resnet 18 with slightly increasing image size**.** Al this time we set the image size as 112*112 pixel which reduces our valu loss.

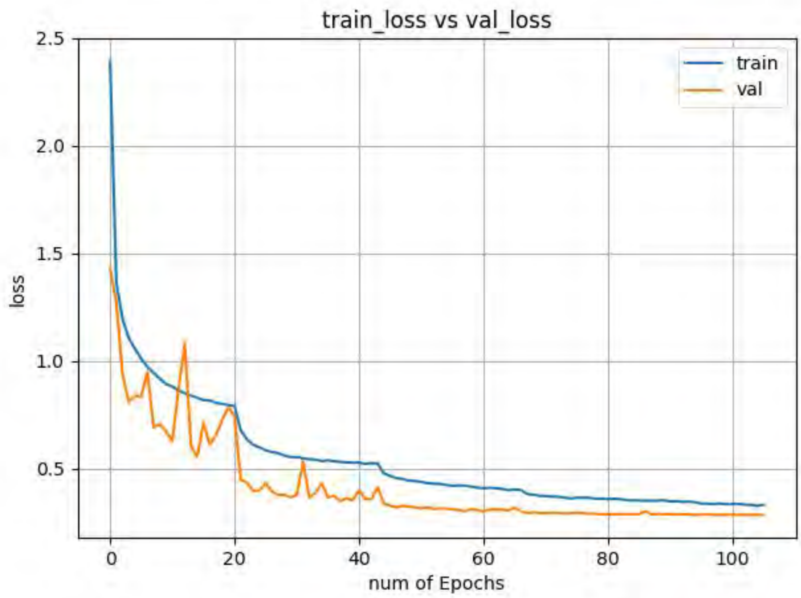

Figure: 4.6 A graph to show achieve lest value loss using Resnet 18

As we have successfully decreased our value loss, then our value accuracy increase in this time in our desired level. By increasing image size with Resnet 18 we achieved accuracy rate 95.10%, which is our best outcome so far.

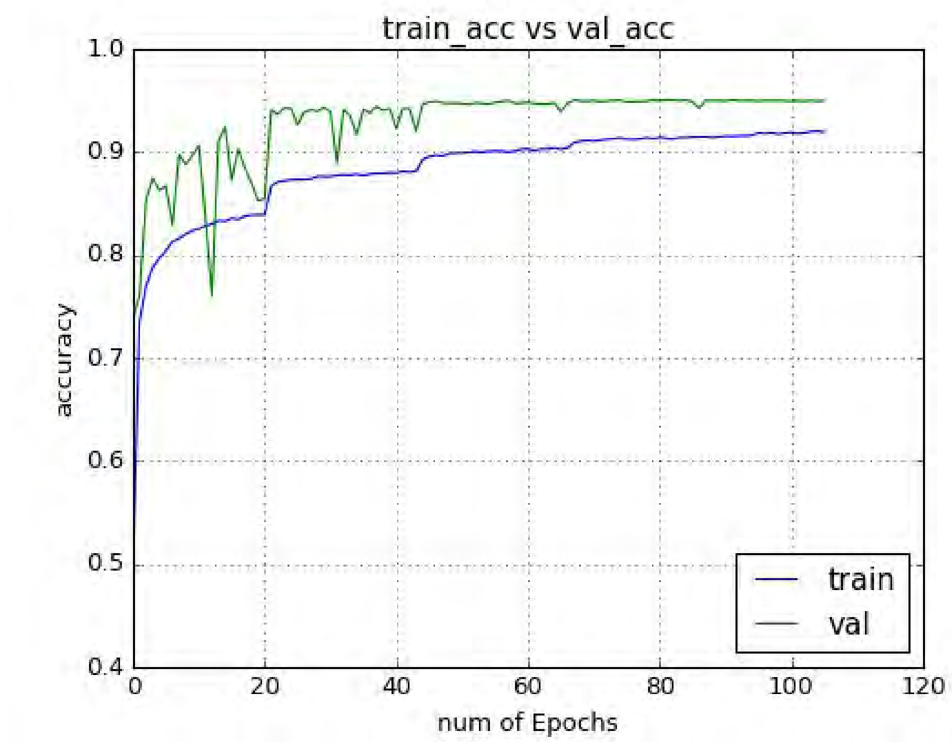

Figure: 4.7 A graph to show achieve best accuracy using Resnet 18 by increasing image size.

Chapter 5

Result

5.1 Model Wise Result Comparison

Each of the models trained with different image size and modified layers produced some of the most singular results that were not anticipated. However, a pattern followed that established the fact that the image size of the training dataset is correlated to the value accuracy that can obtained by the model. Unfortunately, increasing the image size comes with an additional cost of computational requirements and training time. Thus, we had to be economical with our available resources and model architecture, and gradually increase the size of images and change other parameters to obtain the optimum accuracy.

As shown in Table 5.1, we started our training with the image size from 32x32 to 118x118 and applied Adam optimization and augmentation of height and width shift for most of the models, where the learning rate for all the models started with 0.01 decayed learning rate to the factor of square root of 0.1 to 0.5e-6, when value loss was not decreasing. Resnet-18 and Resnet-34 were used with varying degrees, by adding or modifying layers to reduce overfitting of data and increase value accuracy for that particular model as much as possible.

Our first model Resnet-34 with 32x32 image size gave 91.30 % accuracy without augmentation and increased accuracy to 93.10% when augmentation was applied. From here on forwards, we decided to apply augmentation to all the models that we tested to achieve maximum accuracy. We applied the same parameters to Resnet-18 model, and the accuracy slightly increased by 0.1% to 93.20%. With additional layers of maxpool and stride, the value accuracy still remained at 93.20%, however when removing stride, the accuracy increased by 0.1%. Then we increased our training image size to 64x64 and applied to the the previous model with the same parameters, and as expected, accuracy increased to 94.30%. Surprisingly, without any maxpool and strides, the same model gave a higher accuracy of 94.40%, which is slightly better than before.

After acquiring better hardwares like CPU, GPU and RAM, we began to test our model with even larger size of images to achieve greater accuracy. We started with image size of 118x118 on the earlier Resnet-18 model and obtained 94.10% accuracy, which is surprisingly lower than the previous model with image size of 64x64, despite using dataset with larger image size. Then we reduced the image size to 110x110 and applied it to Resnet-18 with and without augmentation, obtaining the results of 94.50% and 93.40% respectively. Later, our experiment to train the model with Rmsprop optimization could not surpass the value accuracy of previous model with Adam optimization. Then we shifted to Resnet-34 with the same image size, and our accuracy increased to 94.60%, which later increased by 0.2% when we applied augmentation with shift of 0.4 height and width. Finally, we again tested on Resnet-18, however this time, we applied dropout optimization to each of the Resnet cells, with image size of 112x112. This model made a breakthrough in value accuracy, which reached the sweet spot of 95.10%, and as of writing this paper, this is the highest accuracy ever reached by any previous state of the art systems on Bangla handwritten character and digit recognition.

Sl	Used Model	Image Size	Optimizer	Augmentation	Accuracy
1	Resnet 34	32x32	Adam	No	91.30%
2	Resnet 34	32x32	Adam	Yes Shift(width(.1),height(.1))	93.10%
3	Resnet 18	32x32	Adam	Yes Shift(width(.1),height(.1)) and More image per step	93.20%
4	Resnet 18 with maxpool 2x2 and stride 2,2	32x32	Adam	Yes Shift(width(.1),height(.1))	93.20%
5	Resnet 18 with maxpool 2x2 and no stride	32x32	Adam	Yes Shift(width(.1),height(.1))	93.30%
6	Resnet 18 with maxpool 2x2 and no stride	64x64	Adam	Yes Shift(width(.1),height(.1))	94.30%
7	Resnet 18 with no maxpool	64x64	Adam	Yes Shift(width(.1),height(.1))	94.40%
8	Resnet 18	118x118	Adam	Yes Shift(width(.1),height(.1))	94.10%
9	Resnet 18	110x110	Adam	No	93.40%
10	Resnet 18	110x110	Adam	Yes Shift(width(.1),height(.1))	94.50%
11	Resnet 18	110x110	Rmsprop	Yes Shift(width(.1),height(.1))	94.10%
12	Resnet 34	110x110	Adam	Yes Shift(width(.1),height(.1))	94.60%
13	Resnet 34	110x110	Adam	Yes Shift(width(.4),height(.4))	94.80%
14	Renset 18 (drop .2 out added in every renset cell)	112x112	Adam	Yes Shift(width(.4),height(.4))	**95.10%**

Table 5.1 Model wise accuracy

5.2 Confusion Matrix Evaluation

In order to test the performance of our classifier, we applied confusion matrix over the 84 classes against a set of test data, for which the values are known. Below in Figure 5.1, the result of convergence is shown.

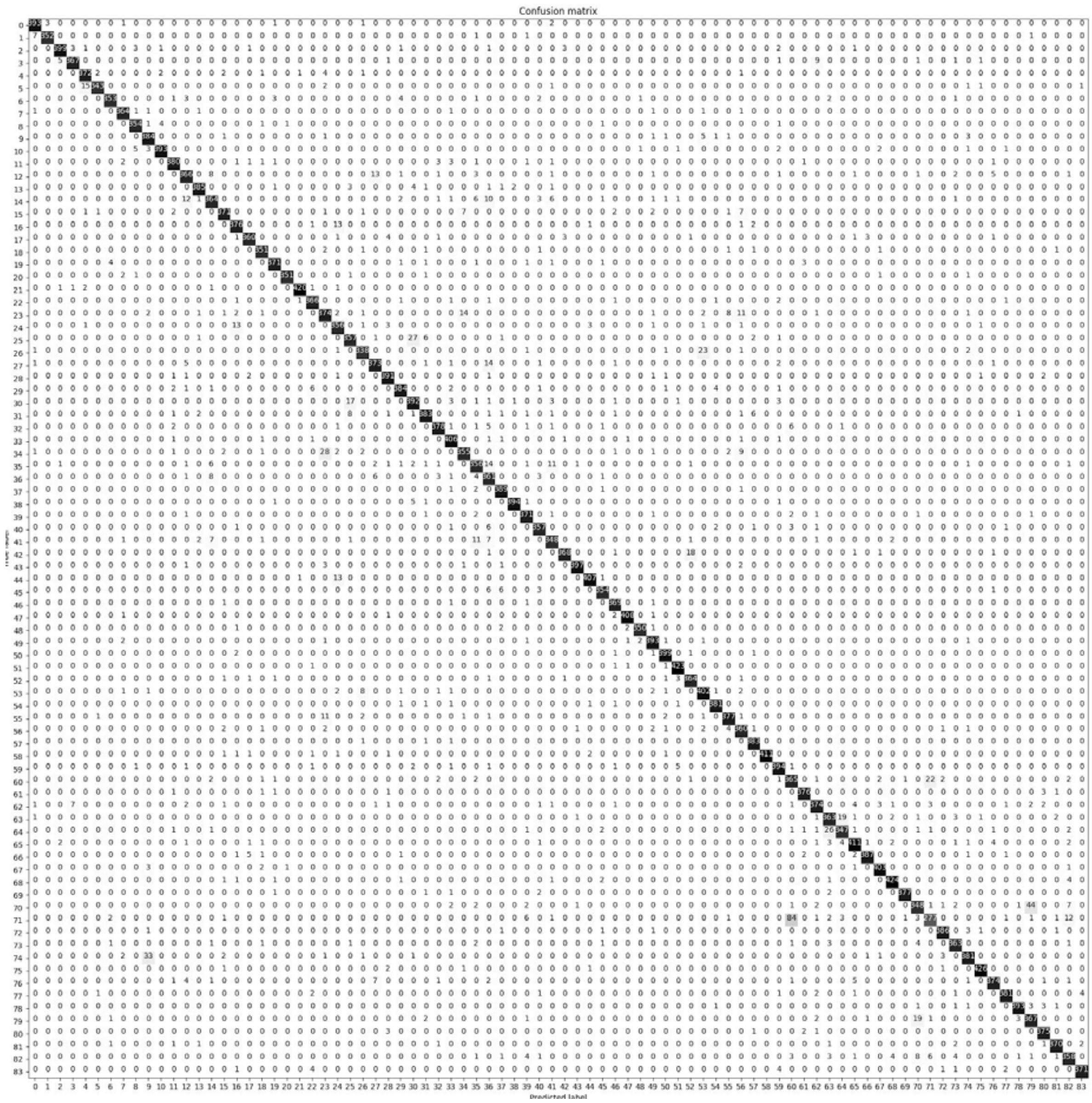

Figure 5.1 Confusion Matrix

As we can see, each of the classes have scored highest in its own test case and shows that the classifier successfully matched each of the classes against its correct test case, proving that the performance of the classifier is uniform for all the classes. Later, we scrutinized the scores achieved by the classes, where some of them are presented in Table 5.2. We discovered that the Bangla character ‘ক্ষ’ scored the lowest precision score of 0.78 and recall of 0.86, while the character ‘ং’ scored the highest precision score with exact 1 and recall of 0.99. These results can be attributed to the fact that complexity of characters and its degree of cursive strokes can reduce the classifier’s ability to correctly predict that character’s class.

Class	Precision	Recall	F1-Score	Support
অ	0.98	0.98	0.98	401
ড	0.87	0.88	0.88	422
ষ	0.95	0.94	0.94	373
ং	1.00	0.99	0.99	412
৮	0.99	0.98	0.98	420
৯	0.97	0.98	0.97	408
ক্ষ	0.78	0.86	0.82	405

Table 5.2 Confusion matrix of sample classes

Chapter 5

Conclusion

We showed that effective utilization of Convolutional Neural Network can achieve better performance to classify and recognize Bangla handwritten characters and digits into digitally readable formats, than the more shallow learning methods like MLP and SVM. If we had more hardware support, then the accuracy would have been better as this method gave a promising result. The results are analogous with previous related work, however, they did not test on a large amount of handwritten character dataset like Bangla-Lekha, we helped us to surpass the previous highest value accuracy of systems based on CNN. Experiments on a large dataset showed the robustness of this model for Bangla handwritten character recognition. In future with more resource and bigger CNN network architecture, we can achieve a better result and improve the state of the art scale for Bangla handwritten letter and digit recognition.

Chapter 6

Future Study

6.1 Limitations

Several factors hindered the progress of the implementation, that either consumed our valuable time, or created restrictions that we were forced to oblige. For example, due to the unavailability of sufficient RAM, the efficiency of the model was bottlenecked to a certain degree, such that memory restrictions were limiting the accuracy that can be achieved. Since our GPU and CPU were of ordinary specification, increasing input parameters like image size would greatly increase the training time needed for the model, which forced us to compromise and work on a simple network instead of a complex network, thus sacrificing any opportunities to further improve the efficiency of the model.

6.2 Future Works

Our future endeavours for this system is to further improve the accuracy that can obtained, by using high end hardwares that will give us the edge in creating a more complex network. Also, we would like to increase the number of classes to include more complex Bangla characters, with word structures that includes basic and compound characters combined within a single character. Furthermore, we would like to expand the scope of this system to include isolated Bangla handwritten words, which will be first segmented to isolate the characters, and then the core classifier will perform its task separately. This can be further extended to include handwritten sentence classifier, where the previous word segmentation along with isolated character classifier can be merged into a single system.

References

[1] U. Bhattacharya, M. Shridhar, S. K. Parui, P. K. Sen and B. B. Chaudhuri, Offline recognition of handwritten Bangla characters: an efficient two-stage approach, Pattern Analysis and Applications, vol. 15, no. 4 , pp. 445-458, 2012

[2] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri and D. K. Basu, A hierarchical approach to recognition of handwritten Bangla characters, Pattern Recognition, vol. 42, pp. 1467–1484, 2009

[3] T. K. Bhowmik, P. Ghanty, A. Roy and S. K. Parui, SVM based hierarchical architecture for handwritten Bangla character recognition, International Journal on Document Analysis and Recognition, vol. 12, no. 2, pp. 97-108, 2009.

[4] S. Roy, N. Das, M. Kundu, M. Nasipuri, Handwritten Isolated Bangla Compound Character Recognition: a new benchmark using a novel deep learning approach.

[5] M. Biswas, R. Islam, G. K. Shom, M. Shopon, N. Mohammed, S. Momen, M. A. Abedin. BanglaLekha-Isolated: A Comprehensive Bangla Handwritten Character Dataset.

[6] D. Wan, M. Zeiler, S. Zhang, S., Y. LeCun., & R. Fergus. Regularization of neural networks using dropconnect. in: Proceedings of the 30th International Conference on Machine Learning (ICML-13), pp. 1058–1066.

[7] D. Cire¸san, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. Arxiv preprint arXiv:1202.2745, 2012.

[8] Sebastien C. Wong ; Adam Gatt ; Victor Stamatescu ; Mark D. McDonnell. Understanding Data Augmentation for Classification: When to Warp?

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Deep Residual Learning for Image Recognition. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[10] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

[11] Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization. arXiv:1412.6980v9 [cs.LG] 30 Jan 2017

[12] M. M. Rahman, M. A. H. Akhand, S. Islam, P. C. Shill, M. M. H. Rahman. Bangla Handwritten Character Recognition using Convolutional Neural Network.