BANGLA HANDWRITTEN CHARACTER RECOGNITION USING DENSE CONVOLUTION NETWORK

MEHADI HASAN REAZ
Dept. of Computer Science and Engineering
American International University Bangladesh
Email: Mehadireaz@gmail.com

ARIFA AKTER EVA
Dept. of Computer Science and Engineering
American International University Bangladesh
Email: arifaeva2019@gmail.com

MD.ABRAR HAMIM
Dept. of Computer Science and Engineering
American International University Bangladesh
Email: abrarhamin1@gmail.com

TASMIA KAHN
Dept. of Computer Science and Engineering
American International University Bangladesh
Email: khantasmia17@gmail.com

Abstract

Over the past few decades, there has been a lot of research done on the handwritten character recognition problem, with varied degrees of success. Researchers had to move away from using artificial neural networks to formulate the HCR problem of various languages (including Bangla) in order to use hand-crafted features-based classification methods due to low computational speed and other computational resource constraints. Recent advancements in deep learning algorithms and major advancements in parallel computing technology have created a strong foundation for researchers to produce cutting-edge, dependable performance in a variety of domains. A few projects have recently been made to create extensive datasets of Bangla handwritten characters and make them accessible to the general public. In this study, we suggest modifying DenseNet, a convolutional neural network architecture, to recognize Bangla handwritten letters. Isolated handwritten Bangla datasets that were recently developed are used to test the suggested technique. The employed datasets are sizable and useful for applying deep learning. With the suggested approach, we attain cutting-edge recognition performance. (Data set link- Link: https://drive.google.com/drive/folders/1P0YwAlNg8GVZH9jqu3dCCXXnBX4jcXNt?usp=share_link
GitHub: https://github.com/MehadiReaz/CVPR/tree/main/FINAL/Final_Project)

1. Introduction:

Bangla a language known as Bengali, is an Indo-Aryan language spoken mostly in South Asia's Bengal area, which includes Bangladesh and the Indian state of West Bengal. It is the second most commonly spoken language in the Indian subcontinent. Bangla is the native and primary tongue of Bangladesh, as well as one of the spoken languages of several Indian states. There are around 235 million native speakers and another 40 million second-language speakers. Bengali is the world's sixth most-spoken primary language and the seventh most-spoken language overall [1]. Bengali is the fifth most widely spoken Indo-European language. The 21st of February has been marked by UNESCO as International Mother Language Day in honor of the language martyrs who died defending it in Bangladesh in the year 1952. It is a significant language having a rich legacy. Handwritten identification of characters is becoming increasingly important in this age of digitization, and its use is common in computer vision. Governments are attempting to keep up with the advancement of computer technology. Computerize their data store, which contains many handwritten scripts. The old way involves manually retyping everything, which demands a large number of people and time.

A number of algorithms based on neural networks with artificial intelligence are examined for handwriting English character identification using different feature extraction techniques. Like other languages, recently Bangla language or reorganization of characters is a popular research topic. Researchers used different types of deep learning technology to recognize it. Because of the difference in shape, size, and details of Bangla text, including the presence of many confusing characters, working with characters written by hand or recognizing digits is always a major difficulty. However, for several years, few works have been completed for Bengali handwritten digit recognition utilizing CNN, SVM, RESNET, and ImageNet, with limited success in categorizing those sorts of recognition to be studied with educational and economic fascinating uses.

Bangla writing scripts differ from English or other languages writing scripts because they are derived from Sanskrit script, which is completely distinct, and it includes alinement, and some characters are identical to other characters as some of them differ from small dots and line. Bangla script, also known as Bengali script or Bangla lipi. Bangla script consists of 50 basic letters 11 are “Swarabarno” and 39 are “Banjanbarno”. For writing numbers in Bangla script, a collection of numerals called "Bangla numerals" is used and also has 10 modifiers, 10 numerals, and more than 300 compound characters. After all of these and other factors, it is challenging to use Bangla Handwriting Character Identification to provide good results. We can enhance many different handwritten recognition-based applications, such as Picture Text Speech, Optical Character Recognition, Number Plate identification, ID card reading, etc. if we are able to overcome all these difficulties and create a model.

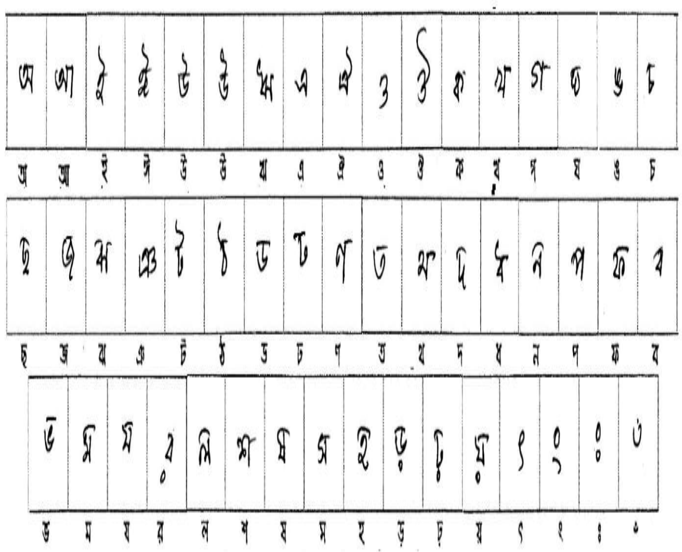

Figure 1: Some Examples of Bangla Characters

Focusing on that, we proposed a Deep Learning method for detecting Bangla Handwriting Characters using image processing. Bangla Handwriting can be determined using a variety of techniques. There is some research on this topic, although several techniques have been applied in the past to detect Bangla

Handwriting. so, we're proposing Bangla Handwriting identification using Dense Convolutional Network architecture

2. Related work:

Bangla Handwriting Characters identification is recently a very popular research topic. After studying few papers, we got most of the papers that applied CNN architecture and a few papers used ImageNet and Resnet. Previous research on the classification of Bangla characters has mostly concentrated on the ten-digit Bangla digit. A handful of the works are accessible in Bangla for the handwritten identification of characters.

The first paper we studied was "Bangla Handwritten Character Recognition using Convolutional Neural Network". In this paper, they used written characters with different shapes and sizes approximately 2000 images. The architecture they used was CNN. Their accuracy rate was 94.55%. They used Bangla and English for both characters and English performed well than the Bangla character. In our paper, we used only Bangla characters, numbers, and compound character

The next paper was "BornoNet: Bangla Handwritten Characters Recognition Using Convolutional Neural Network". They used CNN architecture because it is lightweight. The data set they used was BanglaLkha-isolated, CMATERdb, and ISI. The accuracy they got was 95.71%,96.81%, and 98% respectively. The limitation they faced was overwriting characters. The advantage was their large dataset. We picked a small data set only BanglaLkha-isolated and as well as the model is different.

The next paper was "Isolated Bangla Handwritten Character Recognition with Convolutional Neural Network". They worked with handwritten character recognition. They used ResNet-18 architecture and Adam optimizer, RMSProp, and SGD optimizers. Their data set was BanglaLekha-Isolated and CMATERdb dataset. The accuracy they got was 94.52%.

The next paper was "Bangla Handwritten Characters Recognition Using Convolutional Neural Network". They used CNN for the recognition handwritten bangle alphabet. The dataset they used was Bangalekha isolated and the Ekush dataset. The accuracy they got was 90.22% validation accuracy for Bangalekha isolated dataset and 93.22% validation accuracy for the Ekush dataset.

From those papers, the difference is our data set is a small dataset and our architecture is DenseNet. We're hoping the accuracy will be better from those.

3. Proposed model:

The proposed method focuses on utilizing the capabilities of deep learning, particularly the DenseNet architecture, to deal with the difficult issue of recognizing Bangla handwritten characters. By dealing with the vanishing gradient issue and enhancing the gradient flow during training, DenseNet has shown outstanding performance in a variety of computer vision tasks. We want to recognize handwritten Bangla letters with high accuracy and resilience using DenseNet.

- **DenseNet Architecture**

Dense nets are convolutional networks that are highly connected. In 2016, Huang et al. presented DenseNet, a deep learning architecture for convolutional neural networks (CNNs). It is intended to solve the disappearing gradient issue and encourage use of features in CNNs, enhancing the network's effectiveness and performance. In conventional CNN architectures, each layer takes input from the layer before it and outputs something that is then passed on to the layer after [2]. However, in DenseNet, each layer receives inputs in a feed-forward manner from all preceding levels. In order to create a "dense" connection pattern, each layer's output is combined with the outputs of all preceding levels.

The primary idea of DenseNet is that the network has direct access to the map features of all previous levels since layers are connected densely. The network can move information more easily thanks to this dense connectivity, which also promotes feature reuse [3]. As a result, as compared to conventional CNN designs, DenseNet models are smaller, have fewer parameters, and have better gradient flow. Consider a convolutional network being applied to a single image, x0. The network consists of L layers, where ‘indexes the layer and each layer perform a non-linear transformation H'(). H'() is a composite function that can include operations like Batch Normalization (BN) [4], Rectified Linear Units (ReLU) [5], Pooling [6], or Convolution (Conv). The layer's output is designated as x'.

DenseNet is better because it addresses the vanishing gradient problem, parameter efficiency, enhanced feature reuse, gradient flow, and information flow, combating overfitting, compact model size, strong performance, etc. Despite being utilized as a replacement for batch normalization [7], other works, like this one, explained that batch normalization as well as they provided us with some inspiration, and we included the aforementioned alteration in the suggested architecture. We maintain a maximum pooling of 32x32 with the stride 2x2 as it is described in the architecture. We use SoftMax after fully connected layers as default. we used Adam optimizer. Based on adaptive estimations of lower-order moments, the Adam optimizer is a memory-efficient and quick computation optimization method.

- **Input processing:**

For the purpose of the network's generic performance, we pre-process input photos by inverting, removing noise with the median filter, thickening the edges of the image, and resizing the image to a square shape with the proper paddings by default [8]. By including elastic distortions, we made sure that the diversity of our input photographs was likewise guaranteed. The records were enhanced utilizing elastic distortions by width and height shifting, as suggested by [9]. 0.6 was maintained in the range for this move. Data augmentation makes datasets more varied, ensuring that the network is viewing various samples during the training process.

4. Experimental details:

After experiment we’re presenting the performance and result. We conduct experiments on an windows-11 containing Intel 11th Generation Core i7-11700k CPU with 16 GB RAM and Intel Arc A750 Limited Edition 8GB GDDR6 Graphics Card. The proposed modified DenseNet architecture is implemented in Keras [9] with TensorFlow [10] backend.

- **Dataset:**

The largest isolated Bangla character dataset, BanglaLekha-Isolated, was used for our investigation of handwritten character recognition in Bangla. The most recent dataset of Bangla handwriting characters that is freely available is called BanglaLekha-Isolated. It contains 84 classes, of which 24 classes are frequently used conjunct characters and 50 classes are vowel and consonant classes. There are 166,105 total photos in this dataset; 132,884 of those are in the training set and 33, 221 are in the test set. In particular, it comprises 19,748 numerals, 47,407 frequently occurring conjunct consonants, and 98,950 simple vowels and consonants. The image size of this dataset varies from 110 x 110 to 220 x220 pixels.

- **Experiments:**

The BanglaLekha-Isolated dataset contains a wide range of image sizes, therefore choosing the appropriate size for the input image is essential for attaining the best classification performance. We experiment to determine the ideal input size. The experimental findings are presented. As can be observed, the proposed modified DenseNeT architecture performs well with the greatest 98.60% classification accuracy for images measuring 32 x 32 pixels.

Two hyperparameters are used to optimize the suggested method's performance. We first test the impact of several optimizers on categorization performance. In this experiment, we evaluate the performance of Adam and SGD, two cutting-edge optimizers, on 110x110 input photos. We increase performance by 0.4% and 0.1% using Adam Optimizer. We choose to employ the Adam optimizer in the remaining experiments in light of this study. The performance of the suggested technique is then examined using various dropout rates.

For the compound class, the accuracy we got was 97.6%.

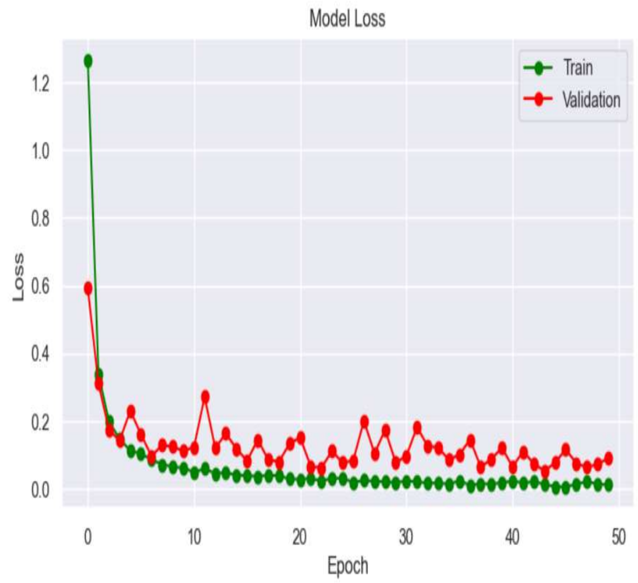

Figure 2: Compound Class Model Loss

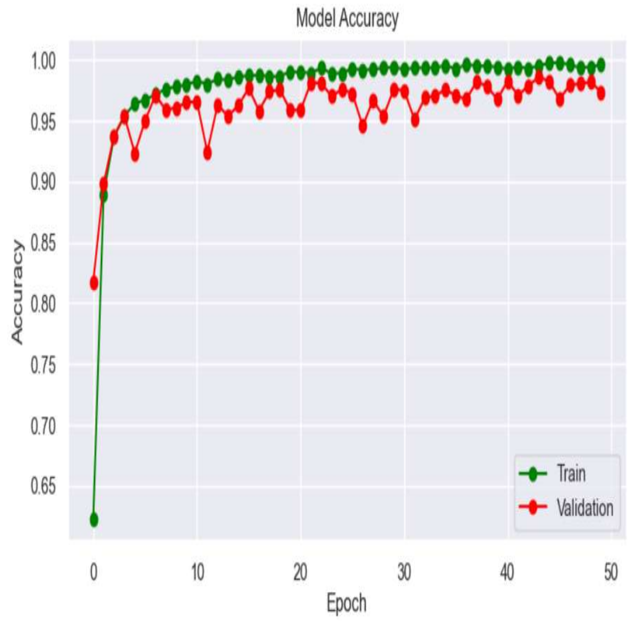

Figure 3: Compound Class Model Accuracy

The result we got for the Bangla digits was 98.2%.

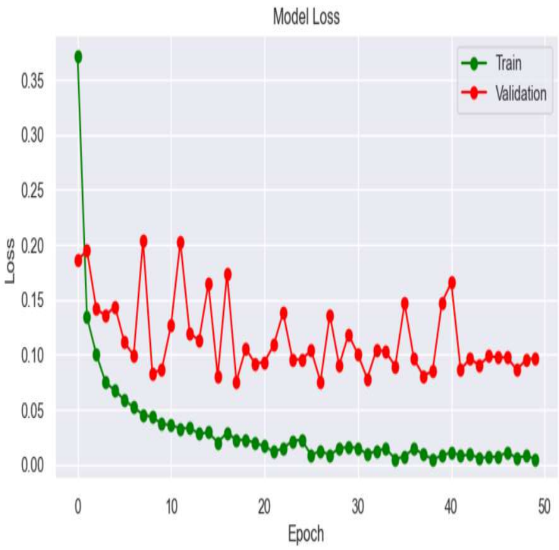

Figure 4: Number Class Model Loss

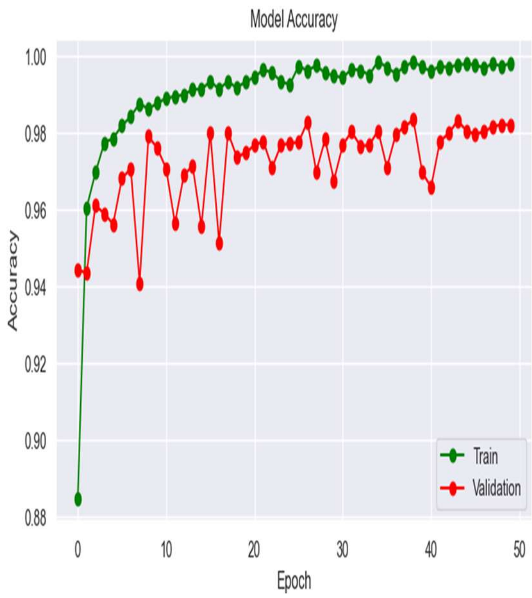

Figure 5: Number Class Model Accuracy

The result we got for Vowels and consonants was 91.2%

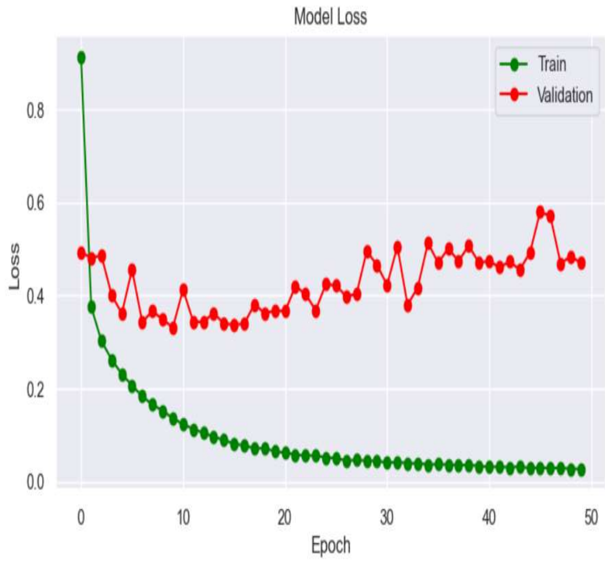

Figure 6: Simple Class Model Loss

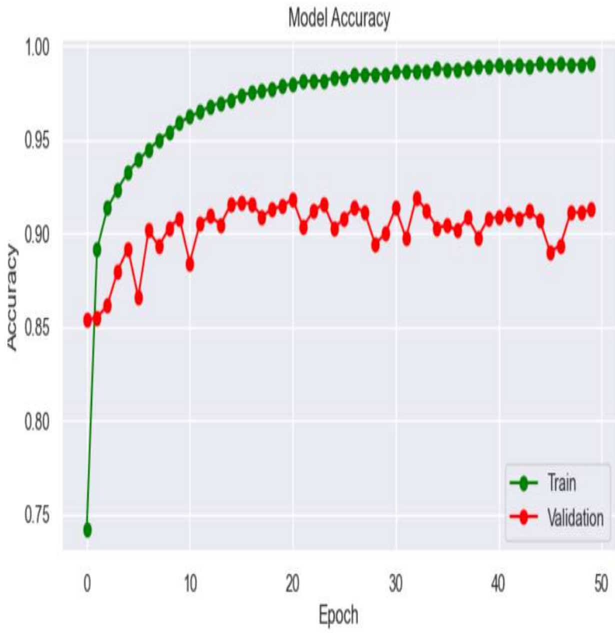

Figure 7: Simple Class Model Accuracy

An evaluation tool for performance in machine learning and classification tasks is a confusion matrix, sometimes referred to as an error matrix. It is a tabular summary of the classification model's predictions made in relation to the dataset's actual labels. The confusion matrix of the BanglaLekha-Isolated dataset's classification performance using the suggested approach

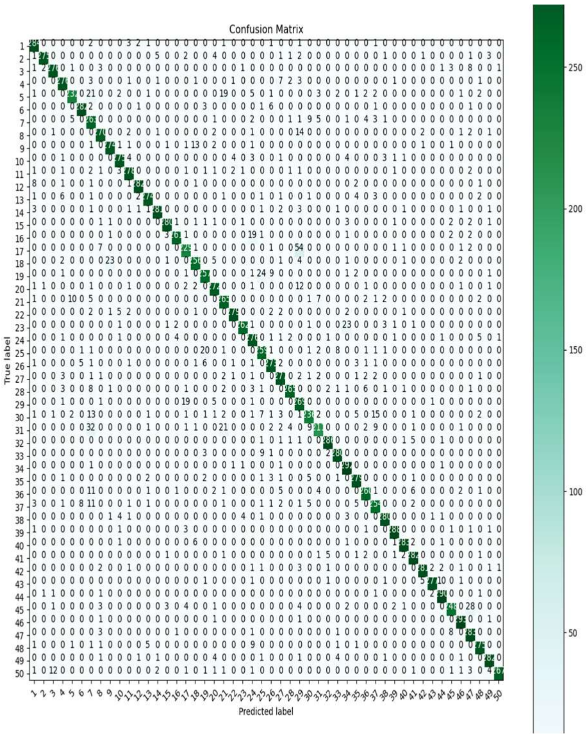

Figure 8: Simple confusing Matrix

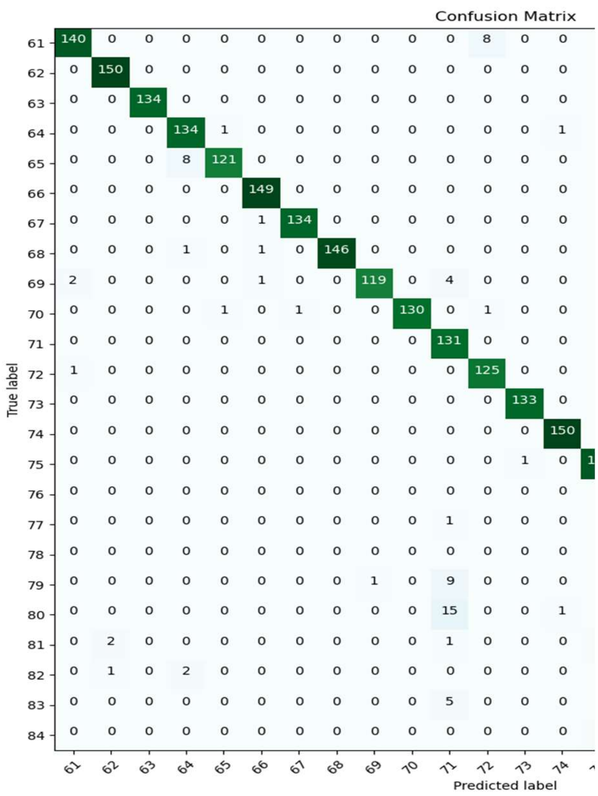

Figure 9: Compound Confuse Matrix

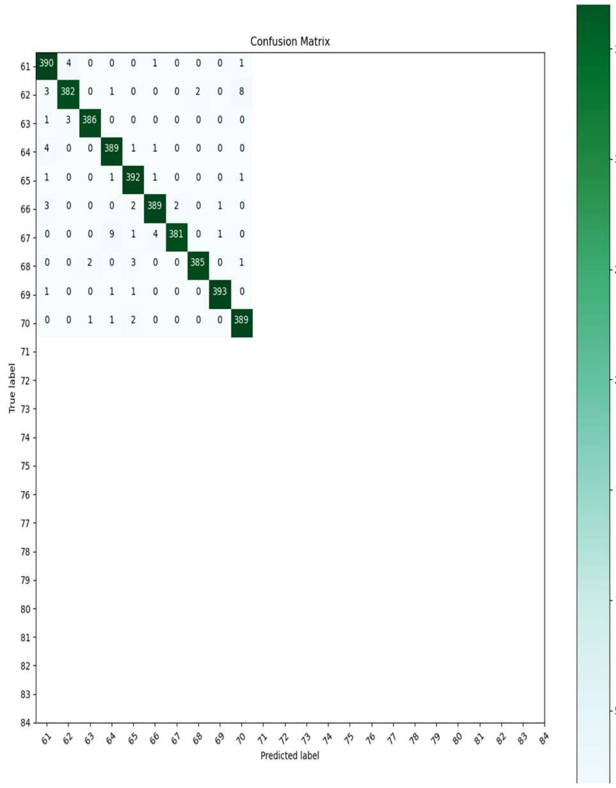

Figure 10: Number Confusing Matrix

So, the final result is for the research,

Category	Accuracy
Vowels and consonant	91.2%
Bangla digits	98.2%.
compound class	97.6%.

Table 1: Accuracy for all categories

5. Discussion and Conclusion:

In this research, we suggested a desneNet architecture that can identify Bangla handwritten characters on one newly proposed large Bangla handwritten character dataset with state-of-the-art accuracy. Each desneNet module in the proposed method includes a dropout layer that expands the functionality of the desneNet architecture in Bangla HCR. Evidently, adding dropout layers as a suggested change to desneNet significantly improved classification performance. The robustness of this approach for Bangla HCR is demonstrated by experiments on sizable datasets of Bangla handwritten characters. We need to further examine the issue in order to provide a better solution by creating an entirely new architecture for Bangla HCR in order to enhance the performance already achieved. We left that section for later work.

Reference

1. Gary F. Simons, Abbey L. Thomas, and Chad K. White. Proceedings of the 29th International Conference on Computational Linguistics, 12–17 October 2022
2. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.
3. Q. Liao and T. Poggio. Bridging the gaps between residual learning, recurrent neural networks and visual cortex. arXiv preprint arXiv:1604.03640, 2016.
4. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.
5. X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In AISTATS, 2011
6. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient based learning applied to document recognition. Proceedings of the IEEE.
7. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in International Conference on Machine Learning, 2015.
8. M. Biswas, R. Islam, G. K. Shom, M. Shopon, N. Mohammed, S. Momen, and M. A. Abedin, “Banglalekha-isolated: A comprehensive bangla handwritten character dataset,” arXiv preprint arXiv:1703.10661, 2017.
9. S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, “Understanding data augmentation for classification: when to warp?” in Digital Image Computing: Techniques and Applications (DICTA), 2016 International Conference on. IEEE, 2016
10. F. Chollet et al., “Keras,” https://github.com/fchollet/keras, 2015.
11. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, J. Dean, M. Devin et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:1603.04467, 2016