Development of an OCR-Based System for Reading Text in Dull Conditions

Kazi Al Refat Pranta
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
kazi.al.refat.pranta@g.bracu.ac.bd

Priya Saha
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
priya.saha@g.bracu.ac.bd

Md Shifatul Ahsan Apurba
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
md.shifatul.ahsan.apurba@g.bracu.ac.bd

Md Sabbir Hossain
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd

Md Mustakin Alam
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
md.mustakin.alam@g.bracu.ac.bd

Annajiat Alim Rasel
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—This article covers a brand-new optical character recognition (OCR) system designed to read the text in low-illuminated conditions. OCR systems are widely used for the purpose of text recognition; nevertheless, these systems may have difficulties identifying text under adverse lighting conditions, such as low illumination or dullness. We have developed a method that uses image enhancement techniques as a preprocessing step before delivering the photos to the OCR engine as a solution to this issue. In addition, the system employs a convolutional neural network, or CNN, in order to extract information from text and categorize it. CNN has been redesigned so that it can extract text information from boring photos more effectively. The OCR system will be assessed using many standard OCR datasets, and the results will be compared to those of other OCR systems. We have a plan to obtain a degree of precision that is more resilient than the current systems. Document processing, digital archiving, and word identification in low-light circumstances are but a few of the probable applications for the proposed system. The proposed OCR system has the potential to increase the precision and efficiency of data processing and document retrieval by making it simpler to read text in less-than-ideal lighting conditions. The goal of this study is to show that the OCR system we have presented is capable of overcoming the challenges provided by low-light circumstances and enhancing text recognition accuracy.**

***Index Terms*—OCR,**

I. Introduction

OCR systems have found widespread application in many areas, including document processing, digital archiving, and text recognition. Unfortunately, OCR technology can falter in poor lighting, such as when there is not enough light or the light is too dull. OCR systems can struggle to distinguish characters in noisy environments when the text is difficult to make out. So, there is a requirement for OCR systems that can accurately recognize text in low-light settings and improve readability by eliminating unwanted noise.

In this research, we present an OCR-based system that can read text in low-light settings. To prepare images for the proposed system, image-enhancing techniques are used, and a convolutional neural network (CNN) is used to extract features from the text and classify it. To improve text feature extraction from otherwise uninteresting photos, the CNN architecture is undergoing optimization. The suggested method is being tested on several industry-standard OCR datasets and compared against the best OCR solutions currently available. In terms of precision and sturdiness, we anticipate that our technology will shine.

Past studies have aimed to improve OCR systems' functionality in low-light or shadow settings. Nevertheless, there has been scant attention paid to improving OCR accuracy in low-light settings. Text can become difficult to discern from the background in low-light situations, making OCR systems work more than necessary.

The photos are preprocessed with image enhancement methods before being sent to the OCR engine, a solution proposed by our system. Contrast adjustment and noise reduction are two examples of image-enhancing techniques. In order to make the text in dull photographs stand out more, we will employ image-enhancing techniques to boost the contrast.

We will also use a convolutional neural network (CNN) to extract features from the text and classify it. Convolutional neural networks (CNNs) have seen widespread application in image recognition applications due to their demonstrated ability to successfully extract characteristics from images for text recognition. To improve text feature extraction from otherwise uninteresting photos, the CNN architecture is undergoing optimization.

Our suggested system will be measured against state-of-the-art OCR algorithms on many industry-standard OCR datasets. We believe that our proposed approach will be more accurate and reliable than current solutions. We believe our suggested system will outperform current systems and attain a respectable level of accuracy.

All in all, our suggested OCR-based method for reading text in dull situations has the potential to enhance the precision and efficacy of data processing and document retrieval in a number of disciplines, including document processing, digital archiving, and text recognition in low-light [1] settings. Our suggested approach can improve the overall quality of data processing and document retrieval by increasing the readability of text under low-light conditions, thus making OCR more reliable and effective.

II. Literature Review

[2]The paper introduced two knowledge distillation [3] methods namely Leveraging Isolated Letter Accumulations By Ordering Teacher Insights for Bangla Handwriting Recognition (LILA-BOTI). In their model, they trained a teacher model on separate printed Bangla characters and used the knowledge to teach the student model by using Convolutional Recurrent Neural Network (CRNN) to identify Bangla Handwritten words. This difference of imbalance in printed vs handwritten and character vs words in the teacher model and student model was addressed by Connectionist Temporal Classification (CTC) loss-based CRNN model. In [4] they used a different but simpler method of CNN to recognize handwritten Bangla characters. In their proposed model the cnn was incorporated with the Max pooling layer and fully connected dense layer. Then they regularized by using two methods batch normalization and dropout. [5] paper used the same model as Ekushnet as they also used Cnn with multiple layers of perceptrons accumulated with Maxpool layer and fully connected dense layer. They also used the same method of dropout to regularize the insight. An [6] used slightly different methods as they first trained the data on different CNN architectures namely Inception Version3, VGG 19, InceptionResnetV2 and Resnet50. Then they used a machine learning algorithm to the ensemble and created a single model to make a comparatively better selection. After extracting features it used stacked ensemble models named Random Forest, XGBoost, LightGBM, and ExtraTrees Classifier and used the soft voting algorithm to reach an effective predictive performance of the model.

For dataset, [2]the paper used BN-HTRd and BanglaWriting. Where Bangla writing data set has around 21234 handwritten Bangla words and BN-HTRd has 108,147 for the number. As they considered 213 graphemes in BN-HTRd and 177 graphemes from the BanglaWriting data set. [4]The paper used the data set [7] Ekush as it had handwritten characters of a count of 367,018 which was made with consonants, vowels, compound letters, graphemes, and numerical digits. [5] on the other hand used a data set consisting of 32400 images of handwritten Bangla digits and 44 types of mathematical symbols. The paper [6] used CMATERdb 3.1.3.3 dataset which includes 171 distinct character types including compound characters and modifiers.

The paper [2] used the F1-Macro score to evaluate and achieved a 3.5% of increase and for word recognition, it achieved a 4.5% increase compared with the base model of No Knowledge Distillation and conventional Knowledge Distillation. The paper [4] used their [7] Ekush data set and they cross-validated with the CMATERdb data set and got high accuracy on the test and validation set. The paper [6] used accuracy and loss performance And compared it with other existing cnn, and SVM models and outperformed them.

The paper [8] presented in detail using the standard deep CNN architecture for character recognition substantially boosts the number of parameters and needs more data to prevent over-fitting. These computer vision achievements help in recognizing simple text or objects and solving complex problems such as scene understanding and autonomous driving. [8] have proposed OCR by combining Convolutional Neural Network and Error Correcting Output Code (ECOC) classifier on NIST handwritten character image dataset and achieved an accuracy of 97. The CNN is used for feature extraction and the ECOC is used for classification. CNN architectures for image classification have two different types of layers: convolutional layers for extracting image features and fully connected layers for performing the classification task based on the features extracted by the preceding convolutional layers.

Again this research paper [9] used methods to handle OCR errors and deduplication in a collection of digitized books for NLP analysis. Their datasets are constructed from scanned texts, with manual correction in the case of the Gutenberg corpus. Besides, their HathiTrust datasets are constructed from a compilation of books from multiple libraries from universities and states. In their experiment, they test the baseline along with three language models based on BERT, RoBERTa, and GPT2. They experiment with these methods to handle errors, evaluated on a collection of 19,347 texts from the Project Gutenberg dataset and 96,635 texts from the HathiTrust Library.

Another paper [10] stated That Bangla is the second most spoken language in India and the fifth most spoken language in the world. The deep CNN models proposed perform better than the previous SVM or MLP models. But their used vanilla deep CNNs are not finetuned. The model is well scalable due to its lower complexity and faster convergence. They achieved an average accuracy of 99.82%, precision of 97.75%, recall of 97.63%, and F1-score of 97.62% using the widely used Mendeley BanglaLekha-Isolated 2 dataset.

This paper [11] achieves comparable accuracy by using a pure CNN architecture without ensemble architecture, which can reduce operational complexity and cost. computational redundancy can be avoided by (SGD) optimizers. deep learning has made great progress in the field of security, handwritten digit recognition, human action recognition, financial trading, and remote image processing. Ramos et al said that simple visual tasks like image classification of handwritten digits in the famous MNIST dataset, these models have surpassed the brain capabilities, performing better than human participants. The standard MNIST handwritten digit database has 60,000 and 10,000 normalized digit images in its training and testing datasets, respectively.

III. Dataset

Primarily we will choose public Bangla data set for our research. As our research is detecting optical character recognition in dull conditions, we will be providing synthetic dullness including different filters of noise and distortion. For this, we will be primarily using python's different libraries including numpy for distortion, and for filters we will be using other libraries such as PIL, and Pilgram, we will be using some other filters to make some parts of the images blurry or hazy. As we are targeting to recognize both handwritten and printed characters we will try to get the best of all the publicly available datasets including BN-HTRd, BanglaWriting , Ekush, CMATERdb 3.1.3.3 etc.

IV. Methodology

Our method utilized a Convolutional neural network with two convolutional layers (activated by relu) and a single fully connected output layer to generate classifiers. We used the Keras library to train a CNN model to perform image classification of handwritten characters. The model was initialized as a sequential object.

The model begins with a convolutional layer of 2 dimension. It employs padding and the ReLU activation function and contains sixteen filters with a kernel size of two by two. This layer receives its input based on the structure of the training dataset. The results of this first layer are fed into the next layer of the max pool, which also has a dimension of two by two.

The next layer is also a convolutional layer, however this time it's again a 2D layer. There are 32 filters in this layer, and the kernel size is 2 by 2. Again, with the right padding and the same activation function of ReLU. The layer's output is passed on to the following layer in the max pool, a 2 by 2 layer.

Third in this model's stack is the GlobalAveragePooling2D layer. Since this layer takes the mean of each feature map generated by the layers below it, the resulting vector has 32 components.

The last layer is a Dense one, and it uses a softmax activation function and includes 50 terminal nodes. There are 50 potential classes of picture, split evenly between vowels (11) and consonants (39).

After that, we use the RMSprop optimizer to construct the model using the loss function of categorical crossentropy. By feeding test images into a trained model and receiving an index of the class with the highest projected score, we are able to acquire predictions from the test data. Finally, we compare our best guess for the class index to the actual indices to ascertain the degree of accuracy.

The model was trained with a batch size of 32, and 100 epochs were found to be optimal for avoiding overfitting and underfitting. Again, we use a callback function during this training session to store the weights of the model with the best performance so far, as determined by the validation loss.

V. Conclusion

To conclude, optical character recognition (OCR)-based systems may prove useful in low-light settings. The reliability, however, could be jeopardized by low-resolution images. To get around this problem, OCR-based systems can improve the quality of the images by using techniques like image enhancement and preprocessing. OCR systems with more complex algorithms can perform better in low-light settings and produce more accurate results.

References

[1] C. M. Nguyen, E. R. Chan, A. W. Bergman, and G. Wetzstein, "Diffusion in the dark: A diffusion model for low-light text recognition," *arXiv preprint arXiv:2303.04291*, 2023.

[2] M. I. Hossain, M. Rakib, S. Mollah, F. Rahman, and N. Mohammed, "Lila-boti: Leveraging isolated letter accumulations by ordering teacher insights for bangla handwriting recognition," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 1770–1776.

[3] C. Li, W. Liu, R. Guo, X. Yin, K. Jiang, Y. Du, Y. Du, L. Zhu, B. Lai, X. Hu *et al.*, "Pp-ocrv3: More attempts for the improvement of ultra lightweight ocr system," *arXiv preprint arXiv:2206.03001*, 2022.

[4] A. S. A. Rabby, S. Haque, S. Abujar, and S. A. Hossain, "Ekushnet: Using the convolutional neural network for bangla handwritten recognition," *Procedia computer science*, vol. 143, pp. 603–610, 2018.

[5] S. N. Shuvo, F. Hasan, M. U. Ahmed, S. A. Hossain, and S. Abujar, "Mathnet: using cnn bangla handwritten digit, mathematical symbols, and trigonometric function recognition," in *Soft Computing Techniques and Applications: Proceeding of the International Conference on Computing and Communication (IC3 2020)*. Springer, 2021, pp. 515–523.

[6] T. B. Ovi and A. Ahnaf, "An ensemble based stacking architecture for improved bangla optical character recognition," 2022.

[7] A. S. A. Rabby, S. Haque, M. S. Islam, S. Abujar, and S. A. Hossain, "Ekush: A multipurpose and multitype comprehensive database for online off-line bangla handwritten characters," in *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part III 2*. Springer, 2019, pp. 149–158.

[8] M. B. Bora, D. Daimary, K. Amitab, and D. Kandar, "Handwritten character recognition from images using cnn-ecoc," *Procedia Computer Science*, vol. 167, pp. 2403–2409, 2020.

[9] A. Kim, C. Pethe, N. Inoue, and S. Skiena, "Cleaning dirty books: Post-ocr processing for previously scanned texts," *arXiv preprint arXiv:2110.11934*, 2021.

[10] M. M. Khan, M. S. Uddin, M. Z. Parvez, and L. Nahar, "A squeeze and excitation resnext-based deep learning model for bangla handwritten compound character recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3356–3364, 2022.

[11] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, "Improved handwritten digit recognition using convolutional neural networks (cnn)," *Sensors*, vol. 20, no. 12, p. 3344, 2020.