REAL TIME BANGLA DIGIT RECOGNITION THROUGH HAND GESTURES USING DEEP LEARNING AND OPEN CV

A PROJECT SYNOPSIS

Submitted by

CHAYTI SAHA
ROLL NO : 11608031

in partial fulfillment for the course Research Project (Part - 1)

of

BACHELOR OF SCIENCE(ENGG.)

IN

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COMILLA UNIVERSITY :: CUMILLA – 3506

FEBRUARY 2021

COMILLA UNIVERSITY :: CUMILLA – 3506

BONAFIDE CERTIFICATE

Certified that this project synopsis entitled, "**REAL TIME BANGLA DIGIT RECOGNITION THROUGH HAND GESTURES USING DEEP LEARNING AND OPEN CV**" is the bonafide work of "**CHAYTI SAHA**" who carried out this project synopsis under my supervision.

SIGNATURE	**SIGNATURE**
MD. Hasan Hafizur Rahman	**Khalil Ahammad**
CHAIRMAN, EXAM COMMITTEE	**SUPERVISOR**
Assistant Professor,	**Lecturer,**
Department of Computer Science and Engineering	Department of Computer Science and Engineering
Comilla University, Cumilla	Comilla University, Cumilla

Tentative Title:

Real time Bangla Digit Recognition through hand gestures using Deep Learning and Open CV.

1. Introduction

Bangla is recognized one of the most spoken languages which ranked fifth in the world and also spoken by more than 200 million people. It is the national language of Bangladesh along with official language. It is the second most popular language in India specially in Kolkata. In addition, Bangla has a rich heritage. UNESCO announced February 21st as the International Mother Language day to respect the language martyrs for the language Bangla in Bangladesh in the year of 1952. Bangla character involves a Sanskrit-based script which is inherently different from English- or Latin-based scripts, and it is relatively tough to achieve desired, highest accuracy on the recognition based tasks.

For hand gesture recognition, the vision-based technology is a very important part of human-computer interaction (HCI). Gesture can be defined as a symbol of any physical behavior or any emotional expression. A gesture can be used also as a tool of communication between computer and human. It is very much different from the traditional hardware based methods. Through gesture recognition, we can accomplish human-computer interaction. It determines the user intent through the recognition by the gesture or movement of the body or any body parts. In the past few decades, many researchers have already strived to improve the hand gesture recognition technology. This Hand gesture recognition has a significant value in many applications like augmented reality, sign language interpreters for the disabled, sign language recognition and robot control.

There are many research works related this language but none of them is done for real time Bangla digit recognition through hand gestures. Therefore, developing a recognition system for Bangla digits is of a great interest.

2. Motivation

The research work related Real time Bangla digit recognition is significant and has many academic and commercial interests because this research work leads the staircase and creating

the scope for further research works on the same field for different Bangla letters, even sentences. But the main concern and challenge for it is to deal with the enormous variety of handwriting styles provided by different writers. Furthermore, different complex handwriting scripts comprise different styles for writing numbers.

Bangla handwritten digit recognition from hand gesture is more challenging compared with the printed forms of character due to the following reasons: (1) numbers written by different persons are not only non-identical but also vary in different aspects such as size and shape; (2) there are numerous variations in writing styles of individual digit ,thus making the recognition task difficult; (3) the resemblances of different digit in shapes, the overlaps, also complicate the digit recognition problem. In summary, it can be said that a large variety of writing styles and the complex features of the handwritten digits make it a challenge to accurately classifying handwritten digit.

There are 10 digits in Bangla language. Moreover, Bangla consists of many alike shaped numbers. This makes it difficult to achieve a better performance with simple classification technique as well as hinders to the development of a reliable handwritten Bangla digits recognition system. However, the lack of such real time works in Bangla language along with gestures shows the potential to explore this exciting area.

3. Objectives of this research

This research aims to build a Bangla Digit Recognition model to predict the digits written using hand gesture on air. The primary objectives of this research can be stated as follows:

- Analyzing the hang gestures to detect the Bangla digit written on air and prepare them to feed in a CNN (Convolution Neural Network) model.
- Development of a CNN digit recognition model to predict the digit.

4. Literature Review:

In the area of digit recognition through hand gesture, many works have been done. Though they were mainly for English or for Bangla sign language. SR Kalbhor, AM Deshpande [1] implemented sign language recognition system for recognition of 0-9 digits using contour-SVM method having accuracy 69% and CNN based approach with 98.31% accuracy. F

Liu, W Zeng, C Yuan, Q Wang, Y Wang [2] developed a trajectory-based method for hand gesture recognition using Kinect. By using the twofold and tenfold cross-validation styles, the correct recognition rates for the system were Arabic numbers (0–9) and English alphabets (A–Z) are reported to be 95.83%,97.25%95.83%,97.25%, and 91.35%,92.63%91.35%,92.63%, respectively. G Li, H Tang, Y Sun, J Kong, G Jiang, D Jiang, B Tao, Shuang Xu & Honghai Liu [3] applied Convolution Neural Network for recognition of hand gesture and SVM is added to optimize the classification function of CNN to improve the validity and robustness of the model. Jain S, Chauhan R [4] implemented NN algorithms such as DNN, CNN, and RNN for the classification of handwritten digits where the best accuracy gained of CNN 99.6% model and the error rate of algorithms were ranges from 0.2–3%. In the area of vision-based hand gesture recognition Misra S, Singha J, Laskar RH [5] introduced 18 new ASCII printable characters along with some of the previously introduced characters [A–Z alphabets, 0–9 numbers and four arithmetic operators (add, minus, multiply, divide)] in their paper and carried out a comparative study for 58 gestures. For hand written digit recognition, Qiao J, Wang G, Li W, Chen M [6] proposed an adaptive deep Q-learning strategy to improve accuracy and shorten running time for handwritten digit recognition where the adaptive deep Q-learning strategy combines the feature-extracting capability of deep learning and the decision-making of reinforcement learning to form an adaptive Q-learning deep belief network (Q-ADBN). For Bengali sign language Md Shafiqul Islalm Md Moklesur Rahman, Md. Hafizur Rahman, Md Arifuzzaman, Roberto Sassi, Md Aktaruzzaman [7] presented a large dataset of Bangla sign language consisting of both alphabets and numerals, composed of 7052 samples representing 10 numerals and 23864 samples correspond to the 35 basic characters of the alphabet. convolutional neural network was used in the recognition of numerals and alphabet separately, and provided an average recognition accuracy of 99.83%, 100%, and 99.80%, respectively for basic characters, numerals, and for their combined usage of characters and numerals. AKM Shahriar, Sheikh Abujar, Sadeka Haque, Syed Akhter Hossain [8] developed and implemented a lightweight CNN model for classifying Bangla Handwriting Digits that outperforms any previous implemented method with fewer epochs and faster execution. Rabeya Basri, Mohammad Reduanul Haque, Morium Akter, Mohammad Shorif Uddin [9] investigated the performance of some state-of-the-art deep CNN techniques for the recognition of handwritten digits. They experimented four deep CNN architectures, such as AlexNet, MobileNet, GoogLeNet (Inception V3), and CapsuleNet models on NumtaDB and confirmed that the AlexNet

showed the best performance on the basis of accuracy and computation time. Chandrika Saha, Rahat Hossain Faisal, Md. Mostafijur Rahman [10] proposed a Deep Convolutional Neural Network (DCNN) based Bangla handwritten digits recognition scheme with a seven layered D-CNN, containing three convolution layers, three average pool layers and one fully connected layer for recognizing Bangla handwritten digits on CMATERdb dataset.

5. Methodology

This research work is divided into several parts and each module is developed in order. The steps to develop this system can be broadly divided into the following module:

1. Data Collection

2. Data Preparation

3. Model Development

4. Implementation

5. Evaluation

A brief overview of each module is discussed below:

5.1 Dataset

The dataset was obtained online from BanglaLekha-Isolated [12]. It is a Bangla handwritten isolated character dataset. This BanglaLekha-Isolated dataset includes 84 different characters having 50 Bangla basic characters, 10 Bangla numerals and 24 selected compound characters. There are 2000 handwriting samples were collected for each of the 84 characters were collected. Finally, 1,66,105 handwritten character images were included in the dataset after discarding mistakes. This dataset could be used both for optical handwriting recognition research work and to explore the influence of gender and age on handwriting. The image is inverted, resized and add some padding to be square with preserving the aspect ratio. About 15,000 and around 3,000 images respectively are taken as train-set and test-set.

5.2 Data Preparation

Initially we prepared dataset for building the model. First, we fixed some incorrect labeling images from the dataset, then deleted some incorrect images such as blank images. This dataset already containing inverted image is prepossessed with its inverted foreground and background, removal of noise with the median filter, edge thickening.

5.3 Model

Neural network is often used in pattern recognition and image processing problems. The most promising tool to carry out this is CNN, a deep learning technique which is inspired by the neuron connectivity pattern in animal visual cortex. It is mainly used in image recognition, object recognition etc. The CNN model for recognizing the digits is constructed using the KERAS library of python and TensorFlow as a backend. Sequential Model is used as classifier which consists of linear stack of layers.

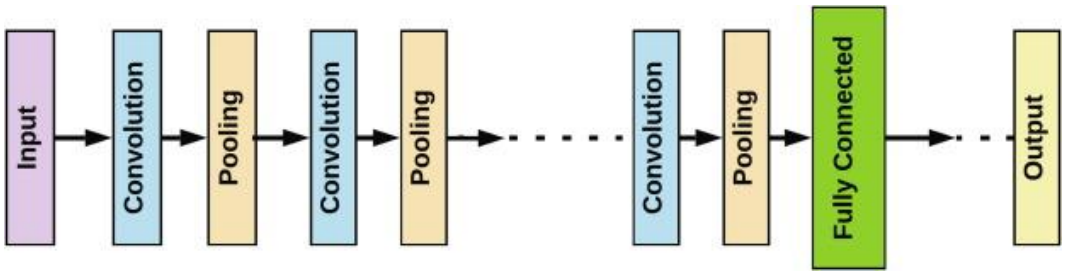

Fig - 1: Basic Building blocks of CNN

In CNN the neurons have learnable biases and weights. Compared to others CNN requires minimum preprocessing. The input is always a vector in neural network but in CNN the input will be a multi channeled image. The CNN contains an input layer, hidden layers and an output layer. The hidden layer constitutes the convolutional layer, Rectified layer unit (ReLU) i.e., activation function, pooling layers, normalized layers and fully connected layers.

The overall architecture of a CNN is shown in Figure 2, which includes two main parts as feature extractor and classifier. In the feature extraction unit, each layer (input layer) sends output to its intermediate next layer considering it as output layer for the previous layer and current output is passed as inputs to the immediate next layer, whereas the classification

part generates the predicted outputs associated with the input data. Convolution and pooling layers the two basic layers in CNN architecture. In the convolution layer, each node extracts the features from the input images using convolution operation, applied on the input nodes. The max-pooling layer helps to abstract the feature through average or maximum operation applying on input nodes. The outputs of the l – 1th layer are used as input for the lth layer. The inputs go through a set of kernels which are followed by nonlinear function called ReLU. Suppose, f refers to activation function of ReLU. For example, if x_i^{l-1} inputs from l – 1th layer, $k_{i,j}^{l}$ are kernels of lth layer. The biases of lth layer are represented with b_j^l. Then, convolution operation can be expressed as following:

$$x_j^l = f (x_i^{l-1} * k_{i,j}^l) + b_j^l$$

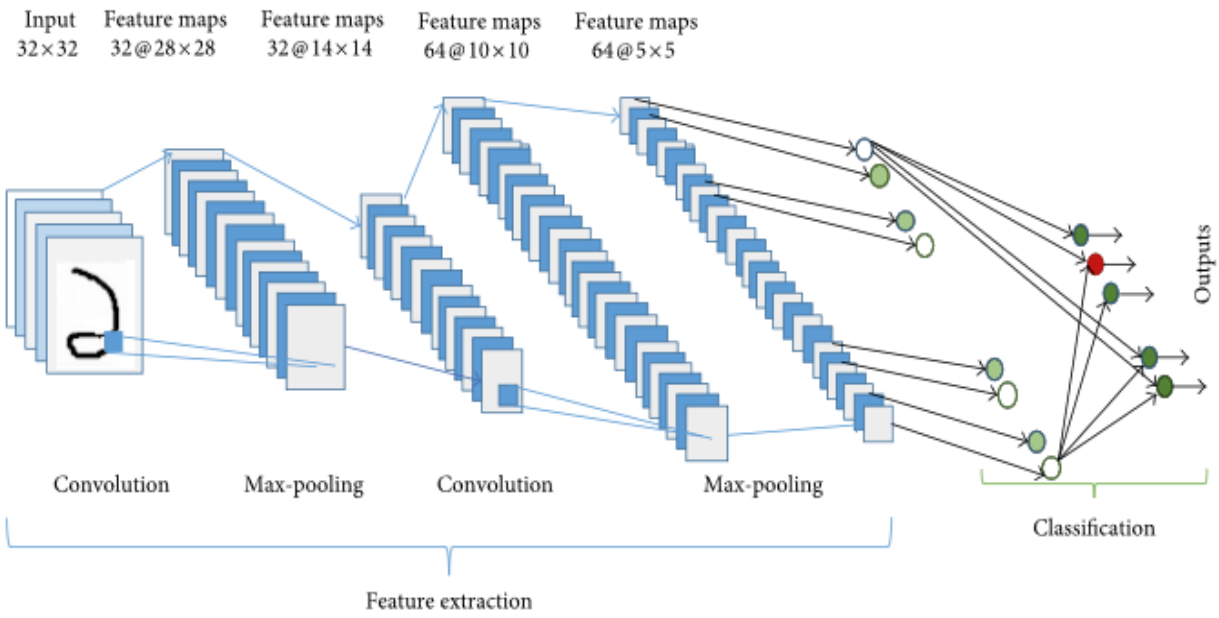

Fig – 2: CNN for digit recognition

The subsampling or pooling layer applies average or maximum operation on input nodes and abstracts the feature. For example, for a 2*2 down sampling kernel, each output dimension will be half of the corresponding input dimension of all the inputs. The pooling operation can be defined as:

$$x_j^l = \text{down} (x_i^{l-1})$$

CNN extracts low- to high-level features, in contrast to traditional neural networks(NN). The propagated feature of the lower-level layers generated the higher-level features. The dimension

of the feature can be reduced depending on the size of the convolution and pooling masks, as the features propagate to the highest layer. However, the number of feature mapping is usually increased to select or map the extreme suitable features of the input images for getting a better classification accuracy.

The outputs of the last layer of CNN are used as the inputs to the fully connected network, it produces the classification outputs and typically uses a Softmax operation. For any input sample x, any weight vector w, and distinct linear functions (k) , the Softmax operation can be defined for the ith class as follows:

$$P\left(y = i \middle| \mathbf{x}\right) = \frac{\exp\left(\mathbf{x}^T \mathbf{w}_i\right)}{\sum_{k=1}^{K} \exp\left(\mathbf{x}^T \mathbf{w}_k\right)}.$$

5.4 Implementation and Evaluation

The proposed model will be trained as described in 5.3 on the processed data. Several robust machines learning software and libraries (PyTorch, Keras, Tensorflow) will be used. The model is then evaluated by both human evaluators and machine generated metrics.

6. Conclusion

Bengali is the native language of Bangladesh which is spoken by 300 million people worldwide and this makes it the seventh most spoken language in the world. In spite of advances in object recognition technology, Real time Bangla Digit Recognition through Gestures (RBDRG) remains largely unsolved due to the presence of many complexities. Even many advanced existing methods are present there which do not lead to satisfactory performance in practice. This research work is expected to shade light on this field for future researchers working with machine learning techniques.

References:

1. SR Kalbhor, AM Deshpande “Digit Recognition Using Machine Learning and Convolutional Neural Network” in 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), May 2018.

2. F Liu, W Zeng, C Yuan, Q Wang, Y Wang," Kinect-based hand gesture recognition using trajectory information, hand motion dynamics and neural networks", April 2019.

3. G Li, H Tang, Y Sun, J Kong, G Jiang, D Jiang, B Tao, Shuang Xu & Honghai Liu ," Hand gesture recognition based on convolution neural network", December 29,2017.

4. Jain S, Chauhan R, "Recognition of handwritten digits using DNN, CNN, and RNN.", In: International conference on advances in computing and data sciences, 2018.

5. Misra S, Singha J, Laskar RH, "Vision-based hand gesture recognition of alphabets, numbers, arithmetic operators and ASCII characters in order to develop a virtual text-entry interface system.", 2018.

6. Qiao J, Wang G, Li W, Chen M, "An adaptive deep Q-learning strategy for handwritten digit recognition.", 2018.

7. Md Shafiqul Islalm Md Moklesur Rahman; Md. Hafizur Rahman; Md Arifuzzaman; Roberto Sassi; Md Aktaruzzaman, "Recognition Bangla Sign Language using Convolutional Neural Network.", published in 2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), September 2019.

8. AKM Shahriar, Sheikh Abujar, Sadeka Haque, Syed Akhter Hossain, "Bangla Handwritten Digit Recognition Using Convolution Neural Network.", December 2018.

9. Rabeya Basri, Mohammad Reduanul Haque, Morium Akter, Mohammad Shorif Uddin, "Bangla Handwritten Digit Recognition Using Deep Convolutional Neural Network.", January 2020.

10. Chandrika Saha, Rahat Hossain Faisal, Md. Mostafijur Rahman, "Bangla Handwritten Digit Recognition Using an Improved Deep Convolutional Neural Network Architecture.", published in: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), February 2019.

11. U. Pal and B. Chaudhuri, "Automatic recognition of unconstrained off-line Bangla handwritten numerals," in Proceedings of Advances in Multimodal Interfaces–ICMI 2000, pp. 371–378, Springer, Beijing, China, October 2000.

12. http://www.https.com//data.mendeley.com/datasets/hf6sf8zrkc/2