Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 143 (2018) 611–618

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Computing and Communication (ICACC-2018)

A Potent Model to Recognize Bangla Sign Language Digits Using Convolutional Neural Network

Md. Sanzidul Islam[a,*], Sadia Sultana Sharmin Mousumi[a], AKM Shahariar Azad Rabby[a], Sayed Akhter Hossain[a], Sheikh Abujar[a]

[a]*Dept. of Computer Science & Engineering, Daffodil International University, Dhaka- 1205, Bangladesh*

Abstract

Hearing impaired people have own language called Sign Language but it is difficult for understanding to general people. Sign language is the basic method of communication for deaf people during their everyday of life. Sign digits are also a major part of sign language. So machine translator is necessary to allow them to communicate with general people. For making their language understandable to general people, computer vision based solutions are well known nowadays. In this research work we aims at constructing a model in deep learning approach to recognize Bangla Sign Language (BdSL) digits. In this approach there used Convolutional Neural Network (CNN) to train particular signs with a respective training dataset (Eshara-Lipi) for acquiring our aim. The model trained and tested with respectively 860 training images and 215 (20%) test images of tent classes of digits. Finally, the training model gained about 95% accuracy at recognition of Bangla sign language digits. This model will contribute for moving one step forward to make BdSL machine translator.

Keywords: BdSL; Bangla Sign Language; CNN; Machine Learning; Deep Learning; NLP; Computer Vision; Pattern Recognition; Sign Digits; Sign Language

1. Introduction

Deaf-mute is a term which was used historically to identify a person who was either deaf using a sign language or both deaf and could not speak [1]. Both are only incapacitate at their hearing or speaking, hence they can do much several things. Communication with the general people which is the only matter that distinct them. The hearing impaired people can simply live like a general person if there is a way for communication between normal people and deaf people. Sign Language is the only way to communication between them. Although hearing impaired people who have sense of sign language, can talk and hear completely. Sign digits are also useful for daily accounting and

* Md. Sanzidul Islam
E-mail address: sanzidul15-5223@diu.edu.bd

1877-0509
10.1016/j.procs.2018.10.438

for communicating the general people and deaf community.
Sign language is a visual language that uses hand shapes, facial expression, gestures and body language [2]. Deaf people share their feeling with various hand shapes and movement in general. A huge amount of research has been done in the field of recognizing Sign Language using different techniques like Hidden Markov Models, skeleton detection, Principal Component analysis (PCA) etc. [3] [4] [5]. Other great techniques involve fulfill the motion history report associated with gestures, motion capturing gloves and computer vision connected with various colored gloves [6][7].
In our approach, there used CNN for data classification. A Convolutional Neural Network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied for analyzing visual imagery [8]. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered.

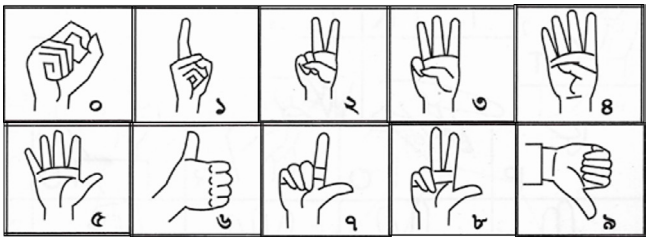

Fig. 1. Bangla sign language digits.

In our work- first, we speak literature review, then describe our model preparation, then discussion about model optimization and finally the evaluation of model.

2. Literature Review

This research goals to construct a model that will identify numbers of BdSL. For recognizing various sign multiple approaches have been used by several researchers which were accomplished in different area.

A New Approach of Sign Language Recognition System for Bilingual Users [9] can recognize 11 Bengali digits and 16 words. There they proposed an universal interpreter software for skin detection & feature extraction. Their system using a database of (27x10x20) images.
Numbers have been recognized effectively in Indian Sign Language Recognition [10]. They represented a framework for a HCI capable of recognizing signs from Indian sign language with PCA (Principle Component Analysis).
In Sign Language Recognition using Microsoft Kinect [11] paper, they used computer vision algorithms and build a characteristics depth and motion profile for each sign language digits 0-9. The feature matrix they generated was trained with SVM classifier. But this approach has a dependency on specific camera device.
Fine Hand Segmentation using Convolutional Neural Networks [12] proposed a method for recognition very accurate hands gesture views based on Deep Learning architecture. In their model they mapped convolution layers directly to a segmentation mask with a fully connected layer. They tried to implement it as efficient in real time as possible.
A recent work was done for recognizing Nigeria indigenous sign language. There they introduced an Yoruba Sign Language recognition system [13] using image processing and Artificial Neural Networks (ANN).

3. Proposed Methodology

A neural net is used in this system to recognize hand signs which is Convolutional Neural Network. The neural net layer explanation, dataset properties, data process, model training and many other methodology is discussed in this section.

3.1. Dataset properties

The Eshara-Lipi dataset which was collected for this project we used to train the model. Eshara-Lipi dataset contains Bangla Sign Language digits from 0 to 9 (0, 1, 2 . . . 9). The dataset has following properties-

- Every class has 100 different images of different peoples hand.
- Ishara-Lipi Dataset has total 1000 (10 * 100 = 1000) images.
- All sign images is cropped and resized by 128 x 128 pixels.
- Dataset images is formatted in.JPG format.
- Images are gray scale and binary coloured then did some preprocessing works.

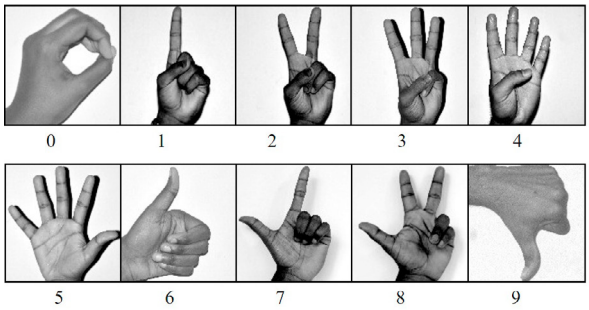

Fig. 2. Eshara-Lipi dataset samples.

3.2. Data Preprocessing

The Eshara-Lipi dataset provides 128 x 128 pixels gray scale images. Some preprocessing works were done for making it usable to train model. Firstly all images were resized by 28 x 28 pixel size. The images were converted into gray scale, then binary coloured image and given the correct labels. Finally saved the image pixels into a CSV file to reduce needed computation power. The method we used determines the threshold automatically from the image using Otsu's method.

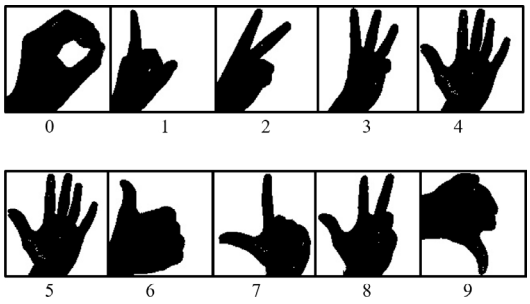

Fig. 3. Data preprocessing by Otsu's method.

3.3. Designing The Model

To recognize these digits here used multi-layer convolutional neural networks which are connected each other. The model is represented by multi layered CNN with two sub layers. First two layers are same, there have two

convolution layers with same padding and swish (3) activation using 32 filters and 5:5 kernel. Then also added a max-pooling layer there. The max-pooling layer has 2x2 followed by 25% dropout layer. All dropout layers used here is for reducing overfitting. The model also use ADAM optimizer [14].

The previous two conv layers generate output and then the output from this two layers goes as an input of two sub-layers. The both sublayers contain same 2 convolutional layers with the same swish activation, padding and 64 filters with a 5x5 kernel, followed by another convolutional layer with a 3x3 kernel. The output of last 2 sub convolutional layers added together and go through a Max-Pooling layer. This Max-pool has 20% dropout [15] layer.
Then flatten the layers and used a fully connected dense layer with 2048 hidden nodes. Final output layer has 50 nodes with SoftMax (1) activation. Using softmax activation means playing with the logistic regression on the feature extraction before the finally connected layer.

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_i}} \tag{1}$$

In this stage of model, a flatten function is used for shape optimization. The basic concept of applying flatten and dense layer function and its output pattern is shown below (Fig.4 and Fig.5).

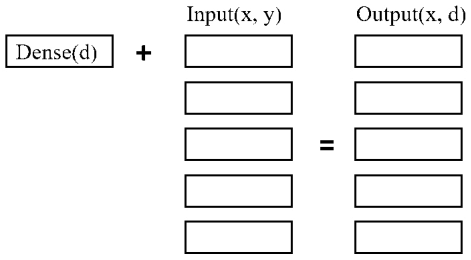

Fig. 4. Dense layer effect.

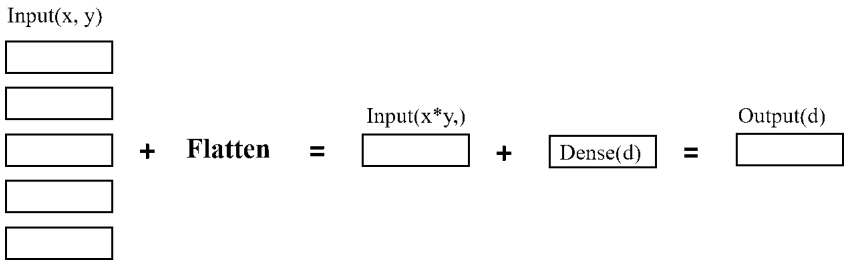

Fig. 5. Applying flatten on dataset.

3.4. Activation

Nowadays the most commonly used activation function is ReLU (2) by default. The ReLU function is defined by equation is-

$$ReLu(x) = Max(0, x) \tag{2}$$

The ReLU activation assigns the parameter back to itself. It creates the problem of ”dead neurons”. There has some better proposed alternative, such as the ELU, SELU and others. Another activation function is used nowadays for efficiency is named Swish activation (3). It’s very simple in equation-

$$Swish(x) = x\sigma(x) \tag{3}$$

Fig.6 is showing the architecture of the whole neural network-

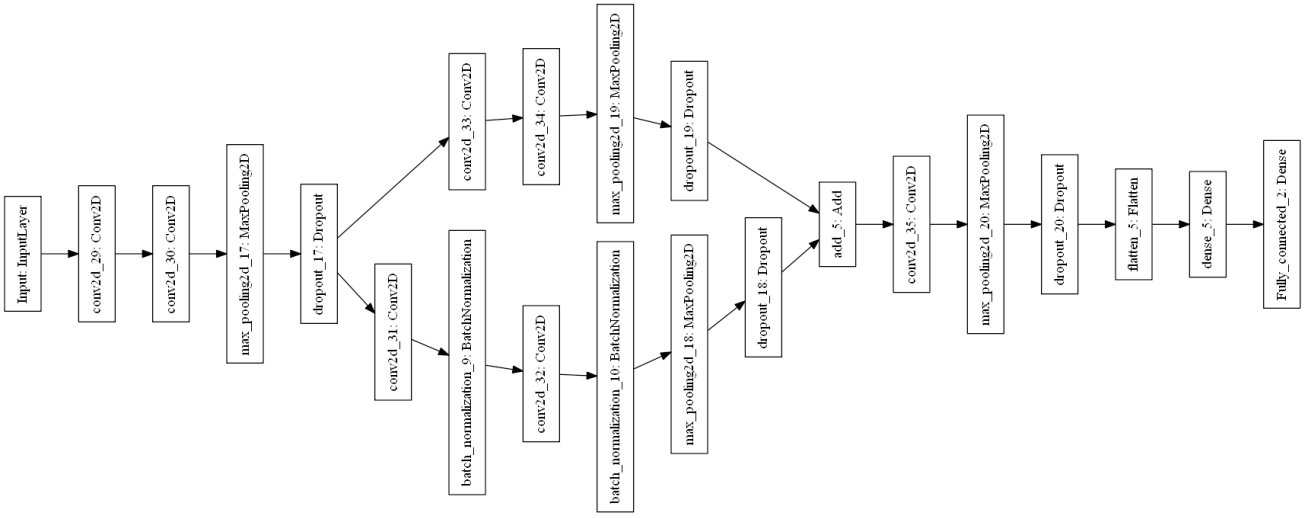

Fig. 6. CNN model architecture.

3.5. Model Optimization

Model optimization is used for making the model more efficient and reliable to input data. In this deep learning model here also applied some optimization techniques. Used SGD for compiling model as optimizer. Stochastic Gradient Descent (sgd) performs as a parameter for each training example. It is a much faster technique. It usually performs single update at a single time.

The cross-entropy is a better choice for cost function optimization. It is known as cross-entropy cost function; also called regularization method. For making better classification and prediction in neural network this function is used widely. Here used a categorical cross entropy as loss function (4).

$$L_i = \sum_j t_{i,j} \log(p_{i,j}) \tag{4}$$

3.6. Model Summery

Table 1. (Summery of model) All layers, their output shape, parameters and the layer with connected.

Layer No. (type)	Output Shape	Param	Connected to
1 Input (InputLayer)	(None, 28, 28, 1)	0	-
2 conv2d_1 (Conv2D)	(None, 28, 28, 32)	832	Input[0][0]
3 conv2d_2 (Conv2D)	(None, 28, 28, 32)	25632	conv2d_1[0][0]
4 max_pooling2d_1 (MaxPooling2D)	(None, 14, 14, 32)	0	conv2d_2[0][0]
5 dropout_1 (Dropout)	(None, 14, 14, 32)	0	max_pooling2d_1[0][0]
6 conv2d_3 (Conv2D)	(None, 14, 14, 64)	51264	dropout_1[0][0]
7 batch_normalization_1 (BatchNor)	(None, 14, 14, 64)	256	conv2d_3[0][0]
8 conv2d_4 (Conv2D)	(None, 14, 14, 64)	36928	batch_normalization_1[0][0]
9 conv2d_5 (Conv2D)	(None, 14, 14, 64)	51264	dropout_1[0][0]
10 batch_normalization_2 (BatchNor	(None, 14, 14, 64)	256	conv2d_4[0][0]
11 conv2d_6 (Conv2D)	(None, 14, 14, 64)	36928	conv2d_5[0][0]

Layer No. (type)	Output Shape	Param	Connected to
12 max_pooling2d_2 (MaxPooling2D)	(None, 7, 7, 64)	0	batch_normalization_2[0][0]
13 max_pooling2d_3 (MaxPooling2D)	(None, 7, 7, 64)	0	conv2d_6[0][0]
14 dropout_2 (Dropout)	(None, 7, 7, 64)	0	max_pooling2d_2[0][0]
15 dropout_3 (Dropout)	(None, 7, 7, 64)	0	max_pooling2d_3[0][0]
16 add_1 (Add)	(None, 7, 7, 64)	0	dropout_2[0][0] dropout_3[0][0]
17 conv2d_7 (Conv2D)	(None, 7, 7, 64)	36928	add_1[0][0]
18 max_pooling2d_4 (MaxPooling2D)	(None, 3, 3, 64)	0	conv2d_7[0][0]
19 dropout_4 (Dropout)	(None, 3, 3, 64)	0	max_pooling2d_4[0][0]
20 flatten_1 (Flatten)	(None, 576)	0	dropout_4[0][0]
21 dense_1 (Dense)	(None, 2048)	1181696	flatten_1[0][0]
22 Fully_connected_2 (Dense)	(None, 10)	20490	dense_1[0][0]

Total params: 1,442,474
Trainable params: 1,442,218
Non-trainable params: 256

4. Model Evaluation

The model developed with Ishara-Lipi dataset performed 94.88% validation accuracy and 95.35% training accuracy. As occurred the training loss and validation loss is shown below in table and graphical representation.

Table 2. Training and Validation.

Evaluation	Rate
Training Loss	12.38%
Validation Loss	26.13%
Training Accuracy	95.35%
Validation Accuracy	94.88%

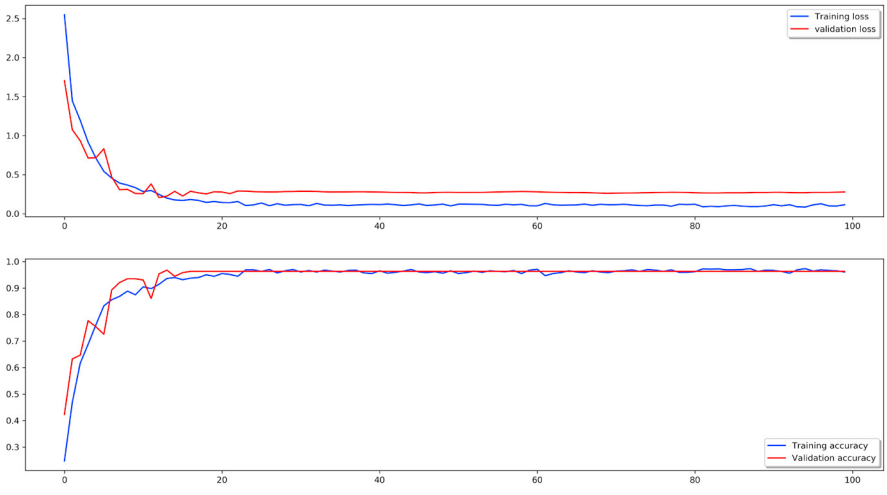

Fig. 7. Graphical view for accuracy and loss.

A confusion matrix or error matrix according to model is given below in fig.8-

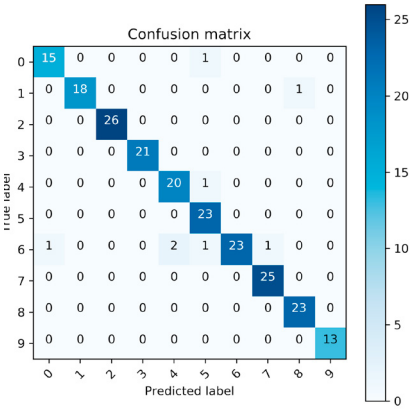

Fig. 8. Confusion matrix output graph.

5. Conclusion and Future Work

This paper represent a deep learning based Bengali Sign Language Digit Recognition System. For sign recognition methods, vision-based models and digit identification methods, convolutional neural network proves a strong candidature. The proposed models deliver output in text form which support to remove the communication interruption between hearing impaired and general people. For standardization of the Bangla Sign Language, we want to use our dataset and the model as a platform. However, everyone cant understand sign language, in future we will change conversation to sign for pleasant communication between different users. In Future we will reach our database & recognize more characters, even to recognize gesture of the Bangla Sign Language and to convert them to Bangla text.

References

[1] $https://en.wikipedia.org/wiki/Deaf-mute$ [Last accessed in 30th April 2018].
[2] $https://www.ndcs.org.uk/family_support/communication/sign_language/what_is_sign.html$ [Last accessed in 30th April 2018].
[3] Oliveira, V. A., and A. Conci. "Skin Detection using HSV color space." H. Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi. 2009.
[4] B. D. Zarit, B. J. Super and F. K. H. Quek, "Comparison of five color models in skin pixel classification," Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999. Proceedings. International Workshop on, Corfu, 1999, pp. 58-63.
[5] S. N. Sawant and M. S. Kumbhar, "Real time Sign Language Recognition using PCA," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, 2014, pp. 1412-1415.
[6] Zhang, Hao, Wen Xiao Du, and Haoran Li. "Kinect gesture recognition for interactive system." Stanford University Term Paper for CS 299 (2012).
[7] Parton, Becky Sue. "Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence." Journal of deaf studies and deaf education 11.1 (2005): 94-101.
[8] $https://en.wikipedia.org/wiki/Convolutional_neural_network$ [Last accessed in 30th April 2018].
[9] F S. M. K. Hasan and M. Ahmad, "A new approach of sign language recognition system for bilingual users," 2015 International Conference on Electrical & Electronic Engineering (ICEEE), Rajshahi, 2015, pp. 33-36.
[10] D. Deora and N. Bajaj, "Indian sign language recognition," 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking, Surat, Gujarat, India, 2012, pp. 1-5.
[11] A. Agarwal and M. K. Thakur, "Sign language recognition using Microsoft Kinect," 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, 2013, pp. 181-185.
[12] Vodopivec, Tadej, Vincent Lepetit, and Peter Peer. "Fine hand segmentation using convolutional neural networks." arXiv preprint arXiv:1608.07454 (2016).
[13] Oyewole, Ogunsanwo Gbenga, et al. "Bridging Communication Gap Among People with Hearing Impairment: An Application of Image Processing and Artificial Neural Network." International Journal of Information and Communication Sciences 3.1 (2018): 11.
[14] Kingma, Diederik P. and Ba, Jimmy. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG], December 2014.

[15] Srivastava, Nitish & Hinton, Geoffrey & Krizhevsky, Alex & Sutskever, Ilya & Salakhutdinov, Ruslan. (2014).Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research.15.1929-1958.