A Simple and Mighty Arrowhead Detection Technique of Bangla Sign Language Characters with CNN

Chapter · July 2019

DOI: 10.1007/978-981-13-9181-1_38

4 authors:

Md. Sanzidul Islam
Daffodil International University

Sadia Sultana Sharmin
Daffodil International University

Akm Shahariar Azad Rabby
Daffodil International University

Syed Akhter Hossain
Daffodil International University

A Simple and Mighty Arrowhead Detection Technique of Bangla Sign Language Characters with CNN

Md. Sanzidul Islam(✉), Sadia Sultana Sharmin Mousumi, AKM Shahariar Azad Rabby, and Syed Akhter Hossain

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
{sanzidul15-5223,sadia15-5191,azad15-5424}@diu.edu.bd, aktarhossain@daffodilvarsity.edu.bd

Abstract. Sign Language is argued as the first Language for hearing impaired people. It is the most physical and obvious way for the deaf and dumb people who have speech and hearing problems to convey themselves and general people. So, an interpreter is wanted whereas a general people needs to communicate with a deaf and dumb person. In respect to Bangladesh, 2.4 million people uses sign language but the works are extremely few for Bangladeshi Sign Language (BdSL). In this paper, we attempt to represent a BdSL recognition model which are constructed using of 50 sets of hand sign images. Bangla Sign alphabets are identified by resolving its shape and assimilating its structures that abstract each sign. In proposed model, we used multi-layered Convolutional Neural Network (CNN). CNNs are able to automate the method of structure formulation. Finally the model gained 92% accuracy on our dataset.

Keywords: Bangla Sign Language · NLP · Computer vision · Machine learning · Image processing · Sign language characters · BdSL · BSL · CNN · Pattern recognition

1 Introduction

Deaf is an incapability that emasculate their hearing and establish them disable to listen [1], while mute is an incapability that emasculate their speaking and establish them disable to talk [2]. Both of them can't just speak and hear but can do much all other things. One thing that has separated them from ordinary people is communications. The hearing impaired people live like a normal people if there was any way to communicate. Sign Language is the only way for deaf and mute to communicate.

A sign Language is a language which is represented by alliance of gesture or movement of the hands. Sign Language is the visual language because of these sign language and spoken language both is different. In real world, people face

K. C. Santosh and R. S. Hegadi (Eds.): RTIP2R 2018, CCIS 1035, pp. 429–437, 2019.
https://doi.org/10.1007/978-981-13-9181-1_38

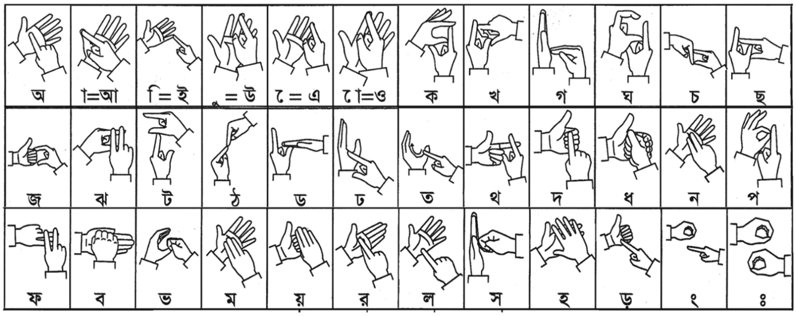

Fig. 1. Bangla sign language characters sign.

different gestures, Fig. 1 has some example. Different country has different sign languages rely on their alphabets and native expression. There are various sign language for example American, Arabic, French, Spanish, Chinese, and Indian etc.

In Bangladesh where around 2.4 million people use Bengali Sign Language. But normal people are not accustomed with their sign. For effective communication speech and hearing impaired people and normal people must have the similar set of knowledge for an individual gesture. It is difficult for the deaf and mute people to learn their sign as there is no appropriate model that work as a communication method. For this, a distance has been created in the society. Bangla Sign Language that creates sign more difficult to realize. So it is essential for construct a model which is convert the sign language to text that supported the mute people to communicate with general people and each other. Now a days, Bangladeshi Sign Language Recognition (BdSL) becomes one of the challenging topics in the area of machine learning and computer vision.

In this paper, a CNN based Sign Language Recognition is offer to acquire highest recognition rate. Here we have focused on static hand gesture in Bangla Sign Language (BSL) which is still challenging because of it visually same yet several sign. So we receive advantage of convolutional neural networks to fulfil a real time and appropriate sign language recognition system. It is mentioning that we can eliminate the obstacle of moving hands from background for hands because CNNs have the ability to learn structures automatically from raw data without any prior knowledge [3].

2 Literature Review

Convolutional Neural Networks have been really effective in image recognition and classification problems, and have been effectively executed for human sign recognition in recent years [4]. Automatic Sign Language Finger Spelling uses CNN architecture from Kinect Depth images. The system trained CNNs for the classification of 24 alphabets and 0–9 numbers using 33000 images and trained the classifier with different parameter configurations [5].

Kang et al. take the extremely well-organized primary step of automatic fingerspelling recognition system using convolutional neural networks (CNNs) from depth maps. In this work, they consider comparatively larger number of classes related with the forgoing literature. They train CNNs for the classification of 31 alphabets and numbers using a subset of collected depth data from multiple subjects [6]. In Deep Convolutional Neural Networks for Sign Language Recognition paper they proposed a CNN architecture for classifying selfie sign language gestures. A stochastic pooling method is applied which pools the benefits of both max and mean pooling techniques. They generates the selfie sign language database of 200 ISL sign with 5 signers in 5 user dependent viewing angles for 2 sec each at 30 fps generating a total of 300000 sign video frames [7].

Hosoe et al. demonstrated a structure for recognition of static finger spellings on images. This recognition of hand gestures is done using a convolutional neural network, which has been trained using physical images. They recorded 5000 images with static finger spellings from Japanese Sign Language [8]. Huang et al. developed a 3D CNN model for sign language recognition that acquires and removes temporal features by performing 3D convolutions. They use multilayer perceptron classifier to classify these feature demonstrations [9]. A voice/text format architecture is being proposed using the neural networks identification to translate the sign language and introduce the Point of Interest (POI) and trajectory idea delivers originality and cuts the storage memory condition in Real-time Sign Language Recognition based on Neural Network Architecture paper [10].

Pigou(B) et al. contribute a recognition system using the Microsoft Kinect, convolutional neural networks (CNNs) and GPU acceleration and making complex handcrafted features. They were able to recognize 20 Italian gestures with 91.7% accuracy [11]

Tsai and Huang use Support Vector Machine (SVM) to recognize the static sign and put on HMM model to classify the dynamic signs and they expended the finite state machine to confirm the correctness of the grammar of the recognized TSL sentence [12].

Yasir et al. measured leap motion controller to take the continuous frame and preprocessed structure and they take out the vital features from hand and fingers using LMC. They presented segmented HMM to discrete sign of expression from the constant frame by transition states. Next fetching the expression, they executed all the features as an input layer and accepted all of them as the limit to the convolutional layer [13].

In this paper, we develop CNN based recognition system which is significant algorithm for object recognition.

3 Proposed Methodology

A neural net is used in this system to recognize hand signs which is Convolution 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 characters from 0 to 35 (0, 1, 2 . . 36) (Fig. 2).

The dataset has following properties.

- Every class has 50 different images of different people's hand.
- Ishara-Lipi Dataset has total 1800 ($36 * 50 = 1800$) images.
- All sign images is cropped and resized by $128 * 128$ pixels.
- Dataset images is formatted in .JPG format.

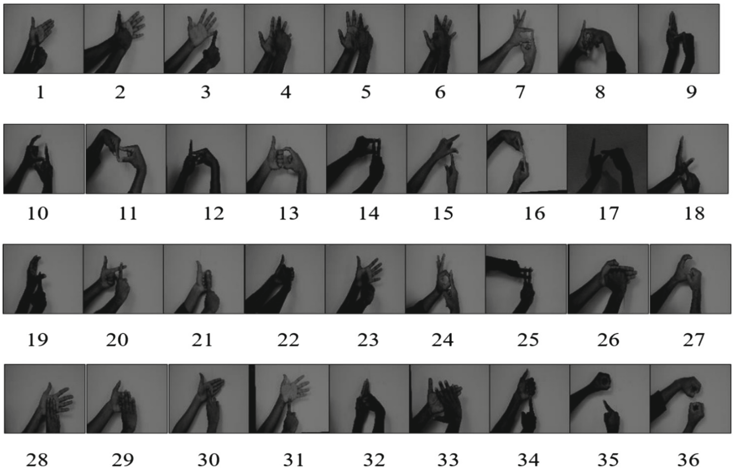

Fig. 2. Bangla sign language characters dataset samples.

3.2 Data Preprocessing

The Eshara-Lipi dataset provides $128 * 128$ pixels grayscale images. For making this model did some reprocessing works like - convert grayscale image to binary and threshold. The method we used determines the threshold automatically from the image using Otsu's method.

3.3 Model Preparation

Algorithm 1:

Convolution 1 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 2 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 3 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 4 (Filter, Kernel Size, Stride, Padding, Activation)

```
Convolution 5 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 6 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 7 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 8 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 9 (Filter, Kernel Size, Stride, Padding, Activation)
Convolution 10 (Filter, Kernel Size, Stride, Padding, Activation)
Flatten (data format)
Dense (Units, Activation, Kernel initializer, Bias Initializer)
Dropout (Rate)
Dense (Units, Activation, Kernel initializer, Bias Initializer)
Dropout (Rate)
Dense (Units, Activation, Kernel initializer, Bias Initializer)
end for
```

Proposed Model in this paper use ADAM optimizer with a learning rate of 0.001. The model has multi layered CNN. For convolution 1 and 2, where filter size is 30, kernel size is (3×3), Stride is (1×1), "same" padding with ReLU (1) activation. Followed 20, 60 filter size and 3, 5, 7 kernel size in other conv layers. Then used 25% dropout to reduce overfitting.

$$ReLu(x) = Max(0, x) \tag{1}$$

For convolution 3, 4 and 5, the filter is 20, kernel size is (3×3), (5×5) and (7×7), Stride is (1×1), "same" padding with ReLU activation. Then used 25% dropout. Then flatten the layer and use a Dense layer with 2560 units with ReLU activation and 50 % dropout. At final output layer, used 36 units with SoftMax (2) activation.

$$S(y_i) = \frac{\mathrm{e}^{y_i}}{\sum_j \mathrm{e}^{y_i}} \tag{2}$$

Densed layer is actually the linear operation on the layer's input vector. It works as below (Fig. 3).

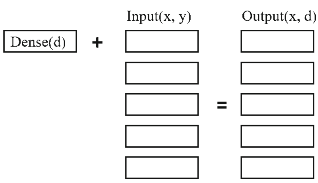

Fig. 3. Densed layer working method.

The flattening step is needed so that we can make use of fully connected layers after some convolutional layers (Fig. 4).

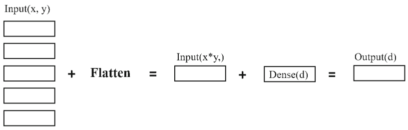

Fig. 4. Flattening layer working method.

Then finally the whole model architecture could be shown in a picture. Figure 5 is showing the neural network architecture.

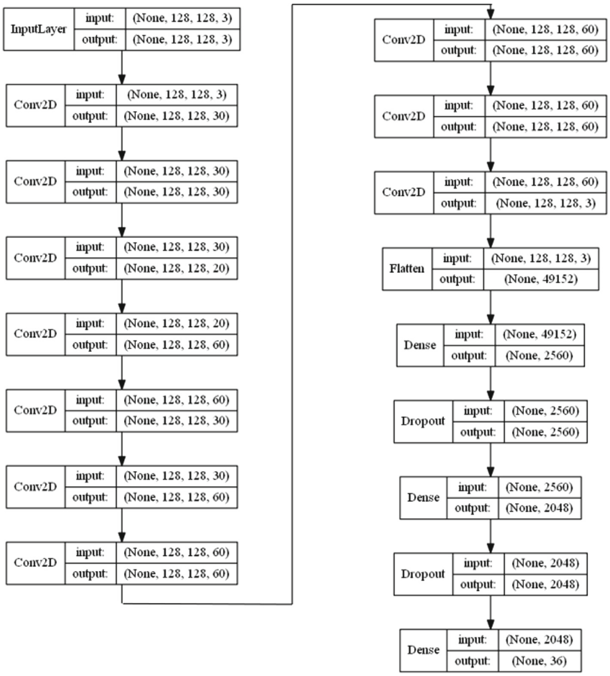

Fig. 5. The whole model architecture.

3.4 Model Optimization and Learning Rate

The choice of optimization algorithm can make a sufficient change for the result in Deep Learning and computer vision work. The Adam paper says, "...many objective functions are composed of a sum of subfunctions evaluated at different subsamples of data; in this case, optimization can be made more efficient by taking gradient steps w.r.t. individual sub-functions ...". The Adam optimization algorithm is an extension to stochastic gradient descent that recently adopting most of the computer vision and natural language processing application. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. Proposed method used ADAM Optimizer with learning rate = 0.001.

When using a neural network to perform classification and prediction task. A recent study shows that cross entropy function performs better than classification error and mean square error. Cross-entropy error, the weight changes don't get smaller and smaller and so training isn't s likely to stall out. Proposed method used categorical cross entropy (3) as loss function.

$$L_i = \sum_j t_{i,j} \log(p_{i,j}) \tag{3}$$

To make the optimizer converge faster and closer to the global minimum of the loss function, using an automatic Learning Rate reduction method. Learning rate is the step by which walks through the minimum loss. If higher learning rate use it will quickly converge and stuck in a local minimum instead of global minima. To keep the advantage of the fast computation time with a high Learning Rate, after each epoch model dynamically decreases the learning rate by monitoring the validation accuracy.

4 Model Evaluation

The dataset was divided into two portions - training data and test data. The model was tarined with the training data and then validated with the validation data. For Ishara-Lipi sign character database, after 30 epoch model gets 92.65% accuracy on the training set and 92.74% accuracy on the validation set. Figure 6 shows the loss value and accuracy of the training set and the validation.

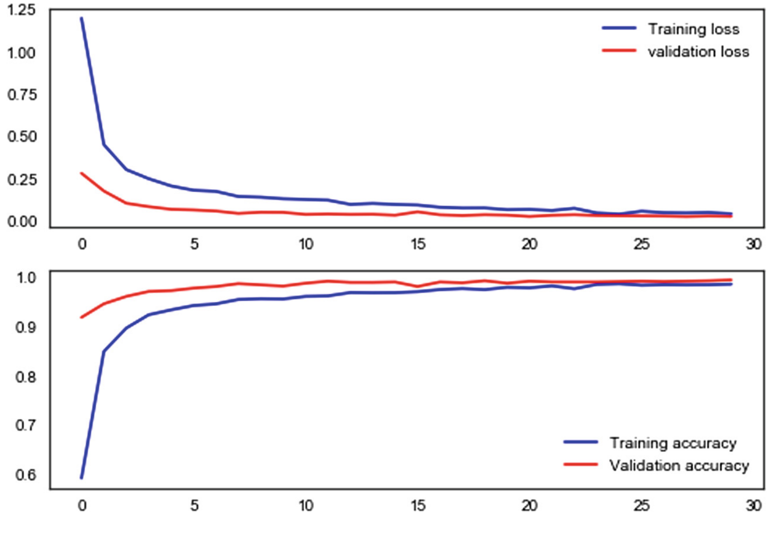

Fig. 6. Model evaluation graph.

5 Conclusion

Developing models that recognize sign from images is a challenging task. The capability of automatically recognize sign language could have a great impression on the lives of hearing impaired people. This will help them in their daily life communication.

In this paper, we represented a convolutional neural network (CNN) approach for a classification algorithm of Bangla Sign Language. The CNN have four convolutional layer which increases the speed and accurateness in recognition. CNN can create outcome in real-time manner and able to recognizing static sign language gesture. Here, we introduced a self-made large dataset that includes 1800 images of 36 alphabets for the Bangla Sign Language. This dataset is open for all researcher. We were capable of get an accuracy of 88% for our CNN classifier. By contributing to the arena of automatic sign language recognition the goal of our model is to reduce the difficulty of communication between hearing impaired people and normal people.

6 Future Work

Studying the boundary of this completed method like structure classification, a more exact sign recognition system can be exhibited. We will try to establish our model more efficient in future. We experiment for 36 Bengali alphabets and we will extent the accuracy for all the Bengali alphabets. In future, additional feature like body movements and facial expressions will proposed in BdSL. Enhance the vocabulary can also be computed as a future work. Our final destination, to build model for identify sign of the BdSL and to interpret them to Bangla text. We would like to conduct this model as a standard platform.

Acknowledgement. I would like to express my heartiest appreciation to all those who provided us the possibility to complete this research under the Daffodil International University. A special gratitude we give to Daffodil International University NLP and Machine Learning Research LAB for their instructions and support. Furthermore, I would also like to acknowledge that, this research partially supported by Bijaynagar Deaf School, Mirpur Deaf School, Mymensing Deaf School, CDD (Centre for Disability in Development) and all of the volunteers team who gave permission to collect valuable data. Any errors are our own and should not tarnish the reputations of these esteemed persons.

References

1. Press CU. Cambridge Dictionary (2017). https://dictionary.cambridge.org/dictionary/english/deaf
2. Press CU. Cambridge Dictionary (2017). https://dictionary.cambridge.org/dictionary/english/mute
3. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)

4. Agarwal, A., Thakur, M.: Sign language recognition using Microsoft Kinect. In: IEEE International Conference on Contemporary Computing (2013)
5. Beena, M.V., Namboodiri, M.N.A.: Automatic sign language finger spelling using convolutional neural network: analysis. Int. J. Pure Appl. Math. **177**(20), 9–15 (2017)
6. Kang, B., Tripathi, S., Nguyen, T.Q.: Real time sign language finger-spelling recognition using convolutional neural network from depth map. In: 3rd IAPR Asian Conference on Pattern Recognition (2015)
7. Rao, G.A., Syamala, K., Kishore, P.V.V., Sastry, A.S.C.S.: Deep Convolutional Neural Networks for Sign Language Recognition, Department of ECE, KL Deemed to be UNIVERSITY, SPACES-2018 (2018)
8. Hosoe, H., Sako, S., Kwolek, B.: Recognition of JSL finger spelling using convolutional neural networks. In: 15th IAPR International Conference on Machine Vision Application (MVA). Nagoya University, Nagoya, 8–12 May 2017
9. Huang, J., Zhou, W., Li, H., Li, W.: Sign language recognition using 3D convolutional neural networks, University of Science and Technology of China, Hefei, China (2015)
10. Mekala, P., Gao, Y., Fan, J., Davari, A.: Real-time sign language recognition based on neural network architecture. IEEE Conference, April 2011
11. Pigou, L., Dieleman, S., Kindermans, P.-J., Schrauwen, B.: Sign language recognition using convolutional neural networks. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8925, pp. 572–578. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16178-5_40
12. Tsai, B.-L., Huang, C.-L.: A vision-based Taiwanese sign language recognition system. In: International Conference of Pattern Recognition (2010)
13. Yasir, F., Prasad, P.W.C., Alsadoon, A., Elchouemi, A.: Bangla sign language recognition using convolutional neural network. In: International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (2017)