Machine Learning Approach for Real Time Translation of Sinhala Sign Language into Text

S.D. Hettiarachchi
Apple Research and Development Centre, Department of Computer Science
Faculty of Applied Sciences, University of Sri Jayewardenepura
Nugegoda, Sri Lanka
shanuka.d.hettiarachchi@gmail.com

R.G.N.Meegama
Apple Research and Development Centre, Department of Computer Science
Faculty of Applied Sciences, University of Sri Jayewardenepura
Nugegoda, Sri Lanka
rgn@sci.sjp.ac.lk

***Abstract* — An effective communication bridge has to be adopted between deaf people and the rest of the society to make deaf and mute people feel involved and respected. This research is aimed at creating a real time Sinhala sign language translator by identifying letter-based signs using image processing and machine learning techniques. It involves creating a digital image database of hand gestures for the 26 static signs. These images are processed, recognized and classified by a Convolutional Neural Network (CNN) based machine learning technique. The proposed solution is able to identify 26 hand gestures by using the CNN network with 91.23% validation and 89.44% training accuracy.**

Keywords* — *Sinhala sign language, Convolutional Neural Network, Digital image processing, Real time translator

I. INTRODUCTION

Development of language as a communication medium was a huge achievement in evolution, and there is no human community without it. Humans have a natural tendency for language in two different modalities: vocal-auditory and manual-visual. Speech is the predominant medium for transmission vocal-auditory language and it seems that spoken languages themselves are either also very old or are descended from other languages with a long history. On the other hand, sign languages do not have the same histories as spoken languages because special conditions are required for them to arise and persevere.

Many natural languages have created their own sign language system with different grammar, syntax, and vocabulary where each displays the kinds of structural differences from the country's spoken language that show it to be a language in its own right. Among those, the Sinhala Sign Language is a visual language used by the deaf people in Sri Lanka which currently consists of more than 2000 sign based words. In any sign language, there are signs allocated for particular nouns, verbs and phrases and are frequently used and highly standardized. These are known as established signs.

This research is aimed at creating a real time Sinhala sign language translator based on letter based signs using image processing and machine learning with the intention of producing an effective communication platform for people with auditory and verbal impairments.

At first, a database of hand gestures for 26 categories is created and those digital images were processed, recognized and classified by a CNN. Then, we identify the most suitable architecture and the implementation platform to develop the system to translate the Sinhalese signs into text through recognition of static alphabet based signs.

A device that translates sign language of deaf-mute person to synthesized text and voice for communication is revealed in [6]. In [1], a new way of communication called artificial speaking mouth is introduced. Because there are drawbacks in the haptic-based approach, work on gesture recognition of sign language is often done by using vision-based approaches as they provide a simple and instinctive communication between computer and a human .[2]. The model proposed in [3] is used to recognize hand gestures captured using a webcam where the feature extraction is done efficiently using SIFT computer vision algorithm. Herath [4] presents a real time Sinhala sign language recognition application by using a low cost image processing method by capturing images having a green background. Vision-based approaches have also been studies in further literature [5, 7].

II. METHODOLOGY

A. The Dataset

In this study, we have only considered 26 letters which have static hand gestures having green as the background color. There are 34 images in one category and the total number of 884 images in the training dataset. Our testing data set consists of 11 images in one category and a total number of 286 images.

B. Preprocessing

In the proposed research, the images are taken under identical parameters such as background color, same side of the hand, etc. The selected images have a width and height of 255 pixels and a scaling factor 1./255 on either side. The proposed CNN model is shown in the below Fig. 1.

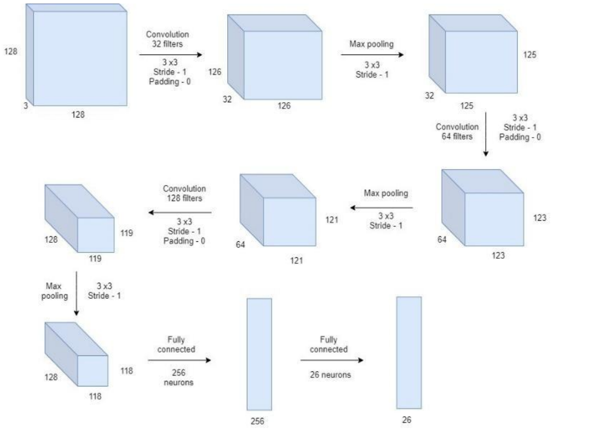

Fig. 1: The CNN architecture

We used a 2D convolutional layer as it provides a better validation accuracy than 3D convolutions. The main task of the convolution stage is to extract high level features such as edges of an input image. After inserting a 128 x 128 image with 3 colors into the convolutional layer, it produces a 126 x 126 3 color image. Starting with a 3x3 filer, we gradually increase the filter sizes while adding more convolutional layers. To classify the dataset, we add an artificial neural network to the convolutional neural network. Basically, a fully connected layer looks at what high level features most strongly correlate to a particular class to produce an output.

We used 256 units which is the number of nodes that should be present in a hidden layer and also leaky relu activation function to achieve non-linearity in the fully connected layer. We have 26 nodes in the output layer because there are 26 categories to reflect the alphabet letters. The Softmax function is used for the activation in the output layer [8]. Subsequently, ooptimizers update the weights to minimize the loss function at each iteration [9].

G. Desktop Application

When the user shows a sign from the right hand to the web camera window in the computer, it processes 200 frames and the final frame will be captured to be used for further tasks. Then, the location of the image is transmitted to the web server where the CNN is deployed. Finally, the relevant letter, which is predicted from the CNN model, is considered as the response. The relevant letter and the cropped image is displayed in the desktop application as in Figure 2.

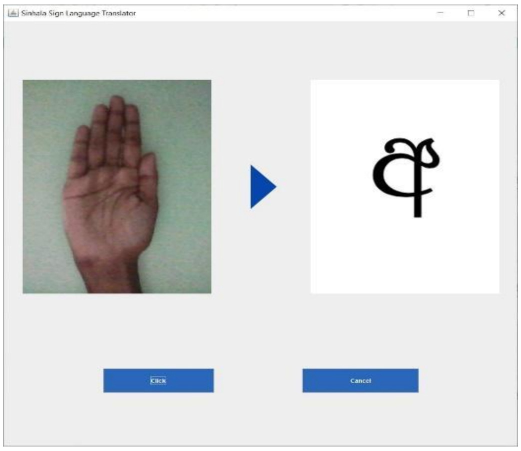

Fig. 2: final output view of the desktop application

III. Results and Discussion

A) Results of CNN model

Training loss and training accuracy: According to Figure 3 the training accuracy of the proposed CNN model is 89.44%. It is pretty much a good performance when we consider the amount of data in the dataset. The training data fit into the model well as the training loss of the proposed CNN model is 0.2647. As in Figure 4, the loss of training set is gradually decreasing with respect to each epoch.

The validation accuracy of the proposed model is 91.23% while the loss is 0.2651 as depicted in Figures 3 and 4. According to these figures although the graph fluctuates at certain points, the validation accuracy is increased.

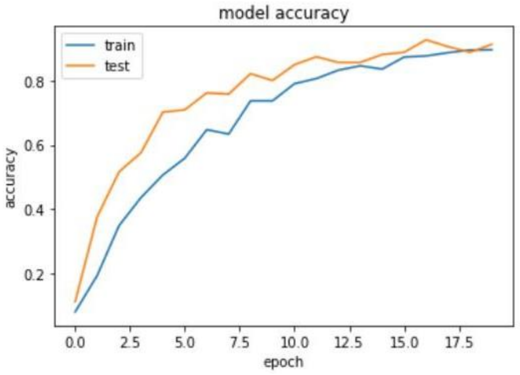

Fig. 3: Accuracy vs epochs of the model

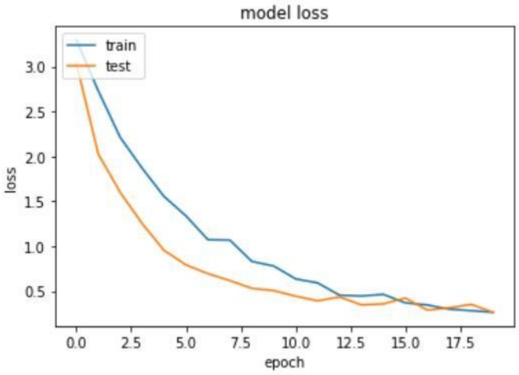

Fig. 4: loss vs epochs of the model loss

IV. Conclusion

We proposed a model for a Sinhala sign language translator, which can be embedded in an application to give a real-time experience to the user. It was able to identify 26 hand gestures using a convolutional neural network with 91.23% validation accuracy and 89.44% training accuracy. The application is able to generate the relevant letter by getting an input of a hand gesture within 1.75 seconds of average time. Additionally, it is capable of tracking the hand gestures of Sinhala sign language for letters and printing it in a text field on a user's device.

References

[1] V. Padmanabhan and M. Sornalatha, Hand gesture recognition and voice conversion system for dumb people," vol. 5, no. 5, pp. 5, 2014.

[2] M. Punchimudiyanse and R.G.N. Meegama, "Unicode Sinhala and phonetic English bi-directional conversion for Sinhala speech recognizer", IEEE International Conference on Industrial and Information Systems 2015.

[3] S. Masood, H. C. Thuwal, and A. Srivastava, "American Sign Language Character Recognition Using Convolution Neural Network,["in Smart Computing and Informatics, S. C. Satapathy, V. Bhateja, and S. Das, Eds. Singapore: Springer Singapore, 2018, vol. 78, pp. 403–412. [Online]. Available: http://link.springer.com/10.1007/978-981-10-5547-842

[4] S. P. More and A. Sattar, "HAND GESTURE RECOGNITION SYSTEM FOR DUMB PEOPLE, " International Journal Of Engineering, vol. 3, no. 2, p. 4

[5] H. C. M. Herath, "IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE," p. 5, 2013.

[6] N. Kulaveerasingam, S. Wellage, H. M. P. Samarawickrama, W. M. C. Perera, and J. Yasas, ""The Rhythm of Silence" - Gesture Based Intercommunication

Platform for Hearing- impaired People (Nihanda Ridma)," Dec. 2014. [Online]. Available : http://dspace.sliit.lk:8080/dspace/handle/123456789/279

[7] A.-A. Bhuiyan, "Recognition of ASL for Human-robot Interaction," p. 6, 2017.

[8] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv:1207.0580 [cs], Jul. 2012, arXiv: 1207.0580. [Online]. Available: http://arxiv.org/abs/1207.0580.

[9] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," arXiv:1811.03378 [cs], Nov. 2018, arXiv: 1811.03378. [Online]. Available: http://arxiv.org/abs/1811.03