Rabea Ahmed

1711894042

Week 1 - Paper Review

CSE499B

**Bangla sign language recognition using concatenated BdSL network**

July 2021

**Abstract**

Sign language is the only medium of communication for the hearing impaired and the deaf and dumb community. Communication with the general mass is thus always a challenge for this minority group. Especially in Bangla sign language (BdSL), there are 38 alphabets with some having nearly identical symbols. As a result, in BdSL recognition, the posture of hand is an important factor in addition to visual features extracted from traditional Convolutional Neural Network (CNN). In this paper, a novel architecture "Concatenated BdSL Network" is proposed which consists of a CNN based image network and a pose estimation network.

**Introduction**

Visual and manual modes are used in Sign Language in order to communicate. It is a complex mechanism, which is why with the progress in the field of computer vision, a lot of research has been done to detect Sign Language. However, the amount is rather small in case of detection of BdSl. In BdSL there are 38 symbols of which 9 are vowels and 27 are consonants... Machine Learning classifier, Support Vector Machine (SVM) was used by a lot of researchers to detect each alphabet from the hand gestures. But image classifiers, like CNN, are now popular among scientists to detect Sign Languages from images. In this paper, the researchers proposed a novel model which is a combination of Image Network (CNN) and Pose Estimation Network.

**Literature Review**

Bangla Sign language recognition is not yet much explored among the researchers. BdSL is actually quite unique. It is the modified form of American SL, British SL and Australian SL. Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) were previously the common techniques used for BdSL recognition. But in the recent times CNN is the most popular among researchers for this job.

**Methods**

The Bengali Sign Language Dataset is used which was collected from the students of Bangladesh National Federation of the Deaf (BNFD). It contains 11061 still images for training and 1520 for testing. All images are in their own 38 differently labeled folder. The dataset is splitted into training and validation set with 9959 and 1102 images respectively following previous works. The experiments have been conducted under identical computational environment.

In this architecture, there are two separate inputs for image network and pose estimation network. For the image network input, the images are converted into grayscale. The images for pose estimation network input are in BGR format. Both the image inputs are normalized by the highest level (255) and resized into 64 x 64 for the task.

This paper refers to the proposed architecture as “Concatenated BdSL Network”. Framework: Tensorflow 2.2, Platform: Google Colaboratory having a 1-core allocated Intel Xeon processor with 2.2GHz and 12.72GB RAM, GPU: None.

**Results**

Most of existing BdSL recognition models only extract features from CNN, which is termed as “Image Network” in the proposed architecture. In addition, the proposed “Concatenated BdSL Network” also extracts hand key point values from the images. As a result, to analyses the effectiveness of the additional features, a comparison is made between the proposed model with respect to existing CNN based model like modified VGG-19 Image Network using the same dataset. Only the “Image Network” part of this proposed architecture has also been trained separately and can be considered as another CNN based model. This model is referred to as “Only Image Network Model” in this paper and is used as the second model for comparison with the Concatenated BdSL Network. Both the “Only Image Network Model” and Concatenated BdSL Network has been trained for 30 epochs due to lack of computational resources like GPU. The proposed Concatenated BdSL Network architecture achieved validation score of 98.28 and overall test score of 91.51 whereas our “Only Image Network Model” got only 90 in test score.

**Discussion**

Some of the alphabet symbols in Bangla sign language have very subtle difference with respect to others which makes the posture of hands very important for recognizing BdSL alphabets. With this view in mind, one of the highlights of the proposed method is to extract additional hand key point features from the pose estimation network to deal with the difficulty of recognizing Bangla sign language. By comparing with deep learning based models that only extract visual features from CNN for Bangla sign language, the significance of hand pose estimation features is analyzed. Even though the modified VGG-19 Image Network was trained on higher resolution images and for higher number of epochs, the Concatenated BdSL Network shows more promising result. It is also compared with “Only Image Network Model” which is trained on the Image Network part of the proposed architecture only, excluding the Pose Estimation Network. This is done to get a better understanding of the effectiveness of extracting hand key points.

**Conclusion**

Despite being a very good classification model, the confusion matrix is not completely diagonal as some misclassification is still prevalent. Even for general people, some signs are indistinguishable. They appear to be the same in terms of looks. The signs suffered mostly from recognizing these two symbols - [উ-ঢ]. The model predicted 'ঢ' instead of 'উ' 13 times out of 40 cases. Other signs that suffered bad predictions are [ঙ-ঞ], [ত-শ], [ঘ-ং]. Some hand gestures are nearly identical to each other. Therefore, there are still plenty of scope for improvement for this paper.

**Reference**

[(PDF) Bangla sign language recognition using concatenated DBs network (researchgate.net)](https://www.researchgate.net/publication/353479114_Bangla_sign_language_recognition_using_concatenated_BdSL_network)