Ishara-Lipi: The First Complete MultipurposeOpen Access Dataset of Isolated Characters forBangla Sign Language

Md. Sanzidul Islam
Dept. of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
sanzidul15-5223@diu.edu.bd

Sadia Sultana Sharmin Mousumi
Dept. of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
sadia15-5191@diu.edu.bd

Nazmul A. Jessan
Dept. of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
nazmul15-4668@diu.edu.bd

AKM Shahariar Azad Rabby
Dept. of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
azad15-5424@diu.edu.bd

Sayed Akhter Hossain
Dept. of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
aktarhossain@daffodilvarsity.edu.bd

***Abstract*—Collecting hand gesture data for sign language is too much difficult to researchers. Ishara-Lipi, the first complete isolated characters dataset of Bangla Sign Language (BdSL) is conducted in this article. It will help to increase interaction between hearing impaired community and general people. The dataset contains 50 sets of 36 Bangla basic sign characters, collected by the help of different deaf and general volunteers. In Bangla Sign Language sign characters there have 6 vowels and 30 consonants by which they can finger spell all Bangla words. In Ishara-Lipi dataset, after discarding mistakes and preprocessing, 1800 character images of Bangla Sign Language were included in the final state. This dataset could be used to develop computer vision based or any kind of system that approves users to search the meaning of BdSL sign.**

***Index Terms*—Bangla Sign Language, Computer vision, Open Source, Sign language characters, BdSL, NLP, Machine Learning, Pattern Recognition, Sign Language Dataset**

I. Introduction

Bangladesh is 8th populated country in the world where almost 16cr people are live. Among them almost 2.6 million people are deaf and mute. Bangla is our mother tongue by-born but deaf and mute people canFLt understand our language what general people use. Sign language is widely used medium among deaf and mute people to communicate with general people. When a baby is born he/she learn his/her first word from surroundings. At first he/she hear the word than memorize it and then try to express it. But a deaf and mute baby cant hear anything. For this reason deaf and dumb people using different types of body gestures to create different signs. There are few books from where deaf and mute people can learn Bangla sign language. One of them is Bengali Sign Language Dictionary published by National Centre for Special Education Ministry of Social Welfare in 1974. There are some originations who are working for deaf and mute people in Bangladesh. Centre for Disability in Development (CDD) which was established in 1996 to develop a more friendly society for disabled persons.

Ishara Bhasay Jogajog a book of Bangla Sign Language was printed by CDD in 2005 and reprinted in 2015 for learning basic gesture of BdSL. According to CDD standard Bangla Sign Language, there are 36 isolated characters and 10 digits in Bangla Sign Language.

In Bangladesh there is no open access complete dataset for Bangla Sign characters for research work and development. In this mindset we are working to develop an open access dataset of isolated characters for Bangla Sign Language. The Ishara-Lipi dataset will help deaf and hearing impaired community by ensuring them to develop education tools.

II. Literature review

There are many datasets for sign language character recognition. Researchers have been working on sign language recognition systems using datasets. Because of characters datasets

978-1-5386-8207-4/18/$31.00 2018 IEEE

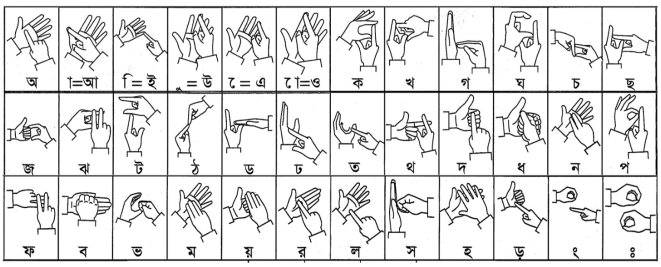

Fig. 1. Bangla sign language digits.

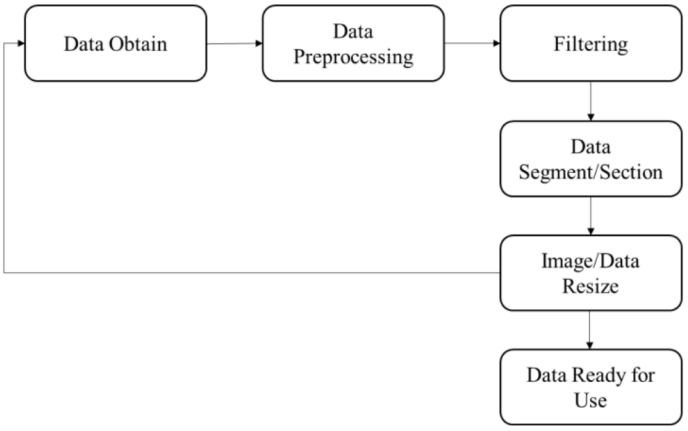

Fig. 2. Working flow of whole process.

is associated with sign language recognition for any research. As various datasets remain in the world, researchers attempt to gather their own dataset with their personal standard and environmental state to expand and act SL recognition on their dataset.

Research in SL recognition has been used for different sign languages in different countries. There are many research, in SL character recognition only described the output of process and did not contain the wide dataset in the works.

A significant volume of action has been done to create SL character recognition. The American Sign Language Lexicon Video Dataset [1] (Neidle and Vogler, 2012) forms such a lexicon for American Sign Language (ASL), containing more than 3000 signs in multiple video views. The Argentinian Sign Language (LSA) paper offering a dataset of 64 signs. (LSA64, Franco and Facundo) This dataset contains 3200 videos of 64 different LSA signs recorded by 10 subjects.

The Arabic Sign Language recognition paper [3] contains 30 manual alphabets signs (Omar Al-Jarrah and Alaa Halawani,2001). In Chinese SLR [2] has 120 signs dataset (2012, Yun Li and Xiang Chen).

An ASL dataset has more complex for recognition because of these data are video files so that these sign have different motions. In sign language recognition one of the authentic part is to develop a Sign Language dataset. Many countries dataset of the SL exits and a BdSL dataset till now not present. This dataset contains 50 sets of 36 Bangla characters sign. There is remarkable benefit for research in BdSL character recognition with the presence of an open dataset.

III. Data collection and preprocessing

Proposed Data collection and preprocessing methods has been stated in figure 2.There are six different states assigned to perform the entire process.

A. Capturing image

A comparatively wide sized BdSL image dataset has been arranged in this paper. The dataset contains total 1800 image. We had collect data from many Deaf School Community and we took images of uncovered hands in white backgrounds. A conducive resolution camera keeping a stable resolution is conduct to capture the images.

B. Labeling data

This move is essential to minimize the noise achieve in the capturing process and to improve the images quality. Some of BdSL character sign vary little from another. Images of this kind of characters seem intimate and enhance the exception of experiment. So we categorized the characters individual. Different characters has different classes/folders.

Bangla Language has about 49 characters but in sign language its not. In Bangla Sign Language there has only 36 characters at all. So we kept 36 characters by naming with numeric convention from 1 to 36 (1, 2, 3...9). The naming convention of all characters are given below in a chart.

C. Cropping images

The captured image cannot be used without cropping for any character recognition purpose. Cropping is necessary to make the image for further experiment. The images are cropped observing the rate of height and width for future processing. Images are cropping to show the hand region.

Fig. 3. Cropping image.

D. Resizing image and convert to gray scale

For making Ishara-Lipi dataset usable in machine learning, deep learning or computer vision based works it is resized in a standard format by python cv2 script. We made a script what will go in every folders containing images and take the same action in every images. The script resize the image by 128 * 128 pixels first and then convert RGB to gray scale.

978-1-5386-8207-4/18/$31.00 2018 IEEE

Fig. 4. Cropped images.

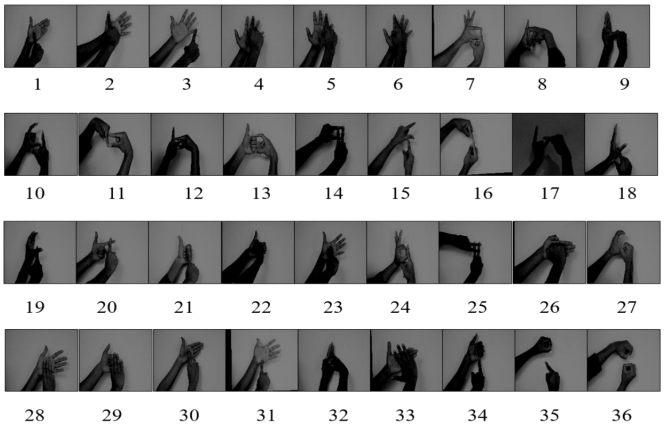

Fig. 5. Final gray scale images.

IV. Dataset properties

- Final dataset contains total 1800 images (36 * 50 =1 800).
- Here in 36 characters- 6 Bangla vowels and 30 consonants.
- Every image is resized by 128*128 pixels.
- The images is captured then resized and finally converted to gray scale from RGB.
- Ishara-Lipi dataset contains 36 folders of images labeled by numbers (1, 2, 3. . . 36) as sequentially presented as Bangla sign characters sequence.
- All images are kept in .jpg formatted datatype.

V. Possible usage of Ishara-Lipi dataset

- It could be broadly used in data science area for making artificial models.
- Ishara-Lipi dataset may keep a significant part in research and development works in BdSL.
- People may take Ishara-Lipi as a virtual standard format of BdSL dataset as its the first one.

VI. Naming convention and data repository

Naming convention is an important fact for using the data properly. So made Ishara-Lipi dataset in simple and friendly standard naming convention. There has 36 folders labeled in numbers from 1 to 36.

1- অ	7- ক	13- জ	19- ত	25- ফ	31- ল
2- আ	8- খ	14- ঝ	20- থ	26- ব	32- শ
3- ই	9- গ	15- ঞ	21- দ	27- ভ	33- হ
4- ঈ	10- ঘ	16- ঠ	22- ধ	28- ম	34- ড়
5- এ	11- চ	17- ড	23- ন	29- য়	35- ং
6- ও	12- ছ	18- ঢ	24- প	30- র়	36- ঃ

Fig. 6. Numeric representation of BdSL characters.

The folder name is as following numbers and every folder contains 50 images or each character. Different images of characters naming in .jpg extension of file type, like- 1_01.jpg, 1_02.jpg, 1_03.jpg. 1_04.jpg and so on.

VII. Model Construction

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.

Algorithm 1:
1: ADAM (Learning Rate)
2: For 30 iterations in all batch do:
3: Convolution 1 (Filter, Kernel Size, Stride, Padding, Activation)
4: Convolution 2 (Filter, Kernel Size, Stride, Padding, Activation)
5: MaxPool 1 (Pool Size)
6: Dropout (Rate)
7: Convolution 3 (Filter, Kernel Size, Stride, Padding, Activation)
8: Convolution 4 (Filter, Kernel Size, Stride, Padding, Activation)
9: MaxPool 2 (Pool Size)
10: Dropout (Rate)
11: Dense (Units, Activation, Kernel initializer, Bias Initializer)
12: Dropout (Rate)
13: Dense (Units, Activation, Kernel initializer, Bias Initializer)
14: end for

Proposed Model in this paper use ADAM optimizer with a learning rate of 0.001. The model has 9-layer CNN. For convolution 1 and 2, where filter size is 32, kernel size is (5x5), Stride is (1x1), same padding with ReLU (1) activation. Followed by a 5 x 5 max-pooling layer. Then used 25% dropout to reduce overfitting.

$$ReLu(x) = Max(0, x) \quad (1)$$

For convolution 3 and 4, the filter is 64, kernel size is (3x3), Stride is (1x1), same padding with ReLU activation. Followed

978-1-5386-8207-4/18/$31.00 2018 IEEE

by a 2 x 2 max-pooling layer. Then used 25% dropout. Then flatten the layer and use a Dense layer with 256 units with ReLU activation and 50 % dropout. At final output layer, used 36 units with SoftMax (2) activation. Fig 3 is showing the neural network architecture.

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

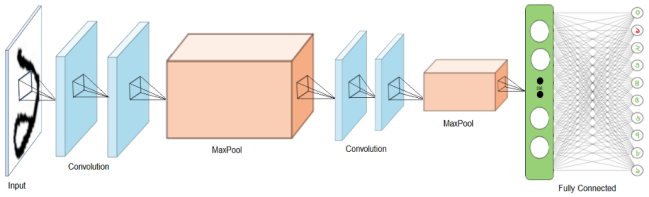

Fig. 7. Bangla sign language processed data.

A. 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 dont get smaller and smaller and so training isnt 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.

VIII. Model Evaluation

For Ishara-Lipi sign character database, through data augmentation 15% data kept for testing and 85% data kept fortraining. For Ishara-Lipi sign character database, after 50 epoch model gets92.65% accuracy on the training set and 94.74% accuracy on the validation set.Fig 4 shows the loss value and accuracy of the training set and the validation.

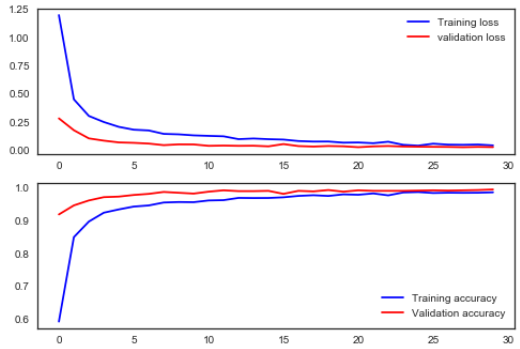

Fig. 8. Bangla sign language processed data.

IX. Conclusion and Future Work

In this paper we have presented a complete dataset named Eshara-Lipi containing all characters. This dataset is publicly obtainable. By providing all characters that could be used for training and testing in data science. We presume that this dataset will be a great resource for researchers in Bangla Sign Language recognition. At the same time, this dataset can be useful for computer vision and machine learning methods designed for learning signs and approaches for analyzing gesture. We believe that this dataset will be an effective resource for all users, learners, researchers of Bangla Sign Language and we prospect that the compatibility of the BdSL dataset will inspire and aid other researchers to study the difficulty of sign language recognition, gesture recognition.
As working in data science here data is the most major fact for working and making effective model. So increasing the dataset size will be the future work for this project. As it is an open source project anyone can contribute us voluntarily.

References

[1] Athitsos, Vassilis, et al. "The american sign language lexicon video dataset." Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008.
[2] Li, Yun, et al. "A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data." IEEE transactions on biomedical engineering 59.10 (2012): 2695-2704.
[3] Al-Jarrah, Omar, and Alaa Halawani. "Recognition of gestures in Arabic sign language using neuro-fuzzy systems." Artificial Intelligence 133.1-2 (2001): 117-138.
[4] Forster, Jens, et al. "Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather." LREC. 2014.
[5] Nanivadekar, Purva A., and Vaishali Kulkarni. "Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation." Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on. IEEE, 2014.

978-1-5386-8207-4/18/$31.00 2018 IEEE