INTERNATIONAL SCIENCE REVIEWS

AIU

No. 4 (1) 2020

Natural Sciences and Technologies series

www.aiu.kz

ISSN: 2707-4862

ASTANA INTERNATIONAL UNIVERSITY

INTERNATIONAL SCIENCE REVIEWS
Natural Sciences and Technologies series

Has been published since 2020

№4 (1) 2020

Nur-Sultan

Editorial address: 8, Kabanbay Batyr avenue, of.316, Nur-Sultan,
Kazakhstan, 010000
Tel.: (7172) 24-18-52 (ext. 316)
E-mail: natural-sciences@aiu.kz

International Sciense Reviews NST - 76153
International Science Reviews
Natural Sciences and Technologies series
Owner: Astana International University
Periodicity: quarterly
Circulation: 500 copies

CONTENT

KAZAKH SIGN LANGUAGE INTERPRETER USING DEEP LEARNING

Meruyert Zhuman
Astana International University, Kazakhstan
zh022mm@gmail.com

Abstract. Everyone can see, listen, and respond to his/her surroundings. There are some people who do not see, listen, and respond to his/her surroundings. Such individuals are mainly the dumb and deaf people. These people depend on sign language to interact with others. However, communication with normal people is a major problem for them because majority of normal people unable to understand their sign language. This will cause a problem for the dumb and deaf people to communicate with others, particularly when they are in social, educational, and work environments. This proposed system was developed in order to assist the hearing or speech impaired people to communicate with normal people. The main goal of this project is to develop sign language translation system that can translate the sign language into text using Convolutional Neural Networks. This uses the property of convolution, mainly devised for analyzing visual imagery. Segmented RGB hand gestures were fed to three layered Convolutional Neural Networks for training and testing in real time. The image dataset, for each gesture, was created using simple image of the hand taken with a personal device such as a laptop webcam.

Keywords: Sign language translation, Convolutional Neural Networks, Visual imagery, Hand gesture.

INTRODUCTION

Every ordinary individual sees, tunes in, and responds to encompassing. There are some people who do not see, listen, and respond to his/her surroundings. Such individuals, mainly deaf and dumb, depend on sign language to communicate with others. Statistics shows that about 9 billion people in this world are deaf and dumb [1, 7]. Interactions between deaf-dumb people and normal people has always been a troublesome assignment. Generally, not every ordinary person can comprehend the communication through sign language utilized by the weakened. This makes it very difficult for them, as communication is one of the most important necessities in life. Furthermore, this will cause a problem for the deaf and dumb communities to interact with others, particularly when they are attempting to coordinate into instructive, social and workplaces. To conquer this issue, a sign language recognition system must be developed with a specific end goal to kill the imperative between the ordinary and debilitated individual.

The main goal of this project is to develop sign language translation system that can translate the sign language into text using Convolution Neural Networks. Since not every typical person is being educated with communication through signing, this system will help them to comprehend the language of deaf and dumb people.

RELATED WORK

Hardik and Vishal Dixit worked on Automated Sign Language Interpreter project. They developed a system which automatically converts sign language into audio output [2]. It consists of various movements and gesture of the hands to train and therefore getting the right accuracy at a low-cost.

G. Geethu and V. Anu worked an Embedded sign language interpreter system for deaf and dumb people. The system consists two main parts [4]. The first part is in the sign recognition of hand and the second one is speech recognition.

Riyanto worked on Sign language interpreter hand using data glove. The method used in this research is used to detect hand and body movements as the object with the data glove as a marker [1]. Hand gestures identification process was used in this project to identify the position pattern of the hand and body.

METHODOLOGY

Sign language is mainly taught to deaf people. Figure 1 shows all the signs present in Kazakh Sign Language.

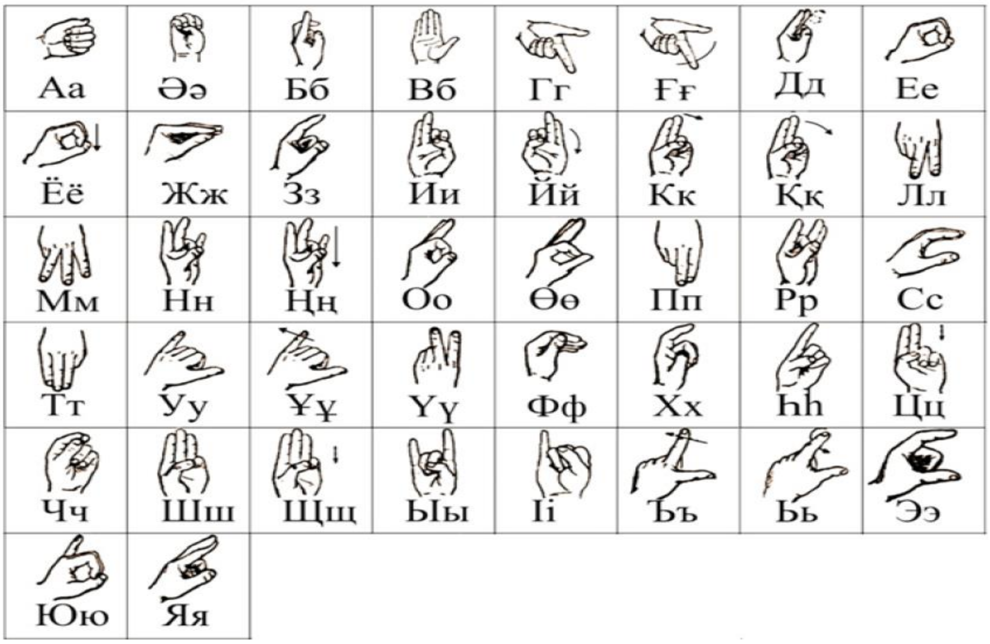

Fig. 1. Kazakh Sign Language Chart

In proposed system Convolutional Neural Networks model was used to classify sign images.

Convolutional Neural Networks are deep neural networks used to process data that have a grid like topology, e.g. images that can be represented as a two-dimensional array of pixels. A CNN model consists of four main operations [1, 8]. They are Convolution, Poling, Flattening and Classification (Fully Connected layer).

1. Convolution: The purpose of the convolution operation is to extract all the features from the input image. It stores the spatial relationship between all the pixels of image by learning image features. It is usually followed by Rectified Linear Unit.

2. Pooling: Pooling reduces the dimensionality of each feature map but retains important data. This process also called as down sampling.

3. Flattening: It is literally going to flatten our pooled feature map into a column. Once the flattening process is done, you end up with a long vector of data to pass through the artificial neural network for further processing.

4. Fully Connected layer: It is a multi-layer perceptron which uses SoftMax function in the output layer. Its purpose is to use features from previous layers for classifying the input image into various classes based on training data.

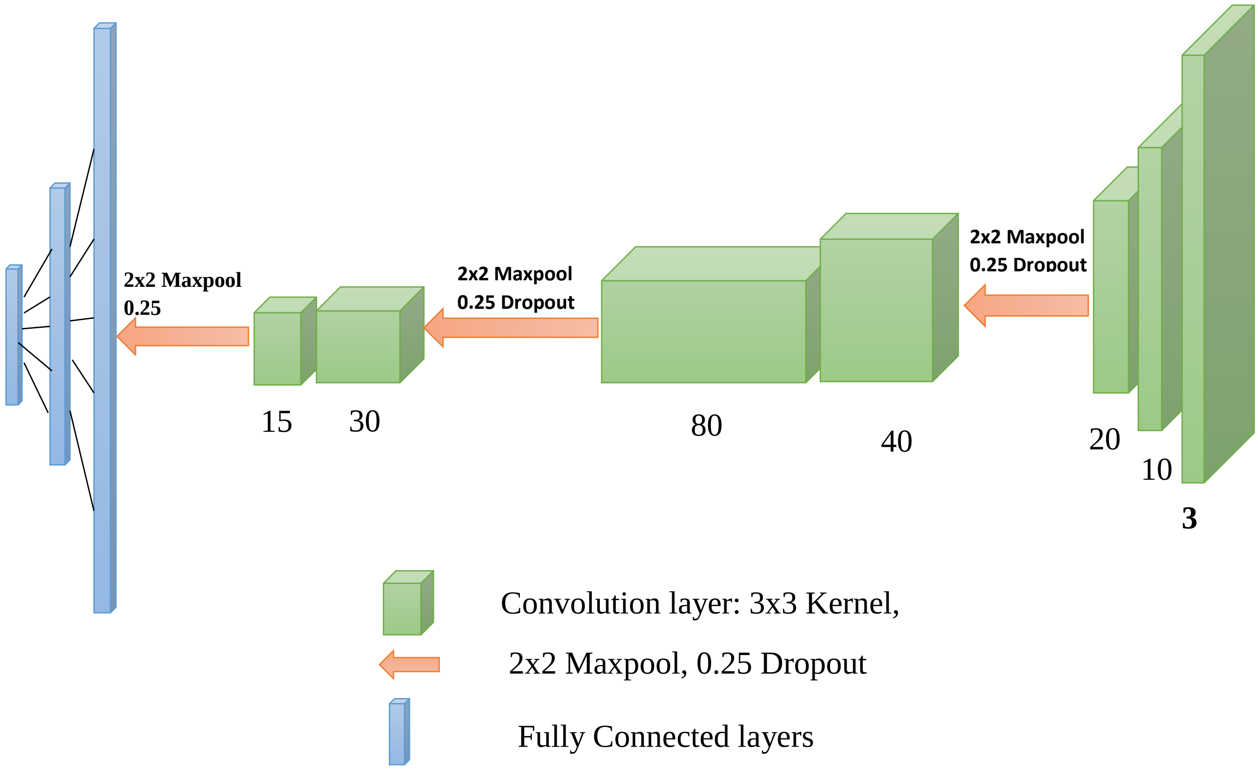

Fig. 2. CNN Model for Sign Language Recognition

The above Figure 2 shows the Sign Language Recognition Model using Convolutional Neural Networks.

Proposed system consists of convolutional blocks containing two 2D Convolutional Layers with Rectified Linear Unit (ReLU) activation, followed by Max Pooling, SoftMax and Dropout layers. These convolutional blocks are repeated three times and followed by Fully Connected layers.

ReLU: Rectified Linear Unit is an element wise operation. The purpose of ReLU is to introduce non-linearity in a convolution networks [1]. In this operation al negative pixel values are changed to zero in the feature map.

SoftMax: SoftMax is used to normalize neural networks output to fit between zero and one. SoftMax helps in converting the output of the last year into probability distribution in neural networks [3].

Dropout: During training process of proposed system, at each iteration, a neuron is temporarily dropped or disabled with probability p. At this step all the inputs and outputs to this neuron will be stopped at the current iteration [5]. The dropped-out neurons are recollected with probability p at every training step. The hyperparameter p gives dropout rate and it is around 0.5.

SYSTEM ARCHITECTURE

The proposed system is structured into three distinct functional blocks. They are

1. Data Processing
2. Training
3. Classify Gesture

1. **Data Processing:** The data.py code contains functions to load the Raw Image Data. It contains code for preprocessing of the image by resizing or rescaling and applying filters to the image to enhance features.
2. **Training:** The proposed system is trained with the hyperparameters obtained from the config file that lists the learning rate and image filtering. The training and validation datasets are loaded as Data-loaders and the model is trained with Cross Entropy Loss. Once training phase is done, the training and validation error is saved to the disk, along with a plot of error.
3. **Classify Gesture:** Once the model has been trained, we can use this model to classify a new sign language gesture. To do that we have to give the file path of the input gesture image and the test data.py script will pass the file path to predict the hand gesture.

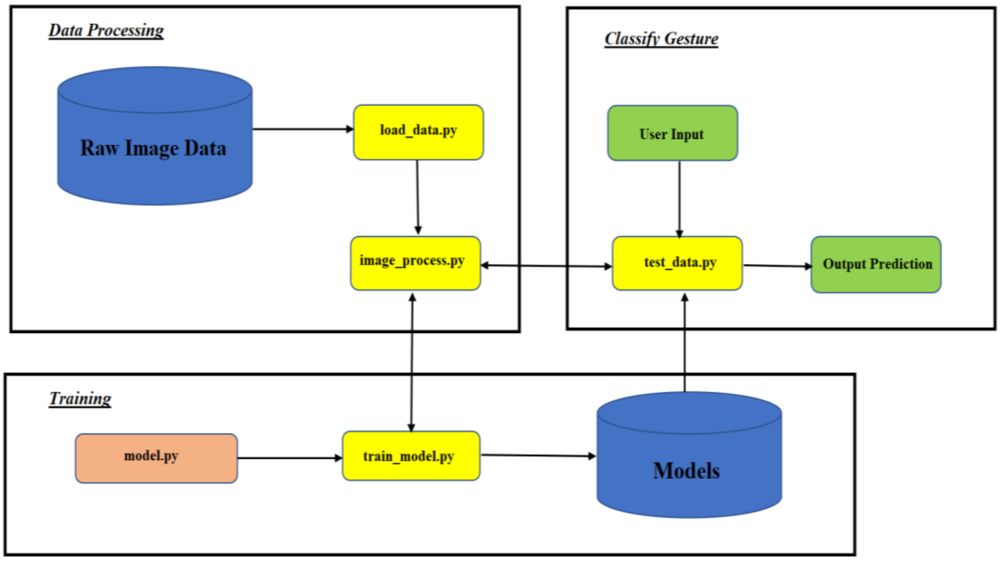

Fig. 3. System Architecture

Figure 3 shows the system architecture of the Sign Language Recognition Model. As shown in above Figure 3 the proposed system uses convolutional Neural Networks to build a model to predict the Sign Language hand gestures.

IMPLEMENTATION

The following are the steps involved in the implementation process of the proposed systems.

1. Importing modules and libraries required for the project.
2. Setting up of hand histogram for creating gestures.

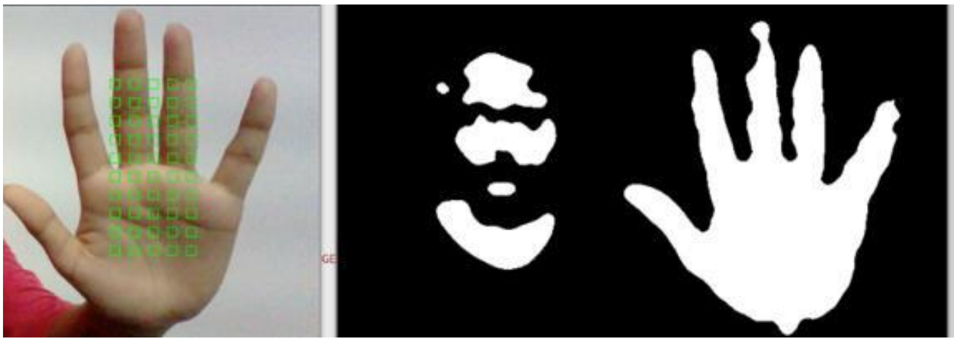

Fig.4. Setting up of hand histogram

3. Once a good histogram is achieved, saving it in the code folder.
4. Adding gestures and labelling them, using OpenCV which uses webcam feed.

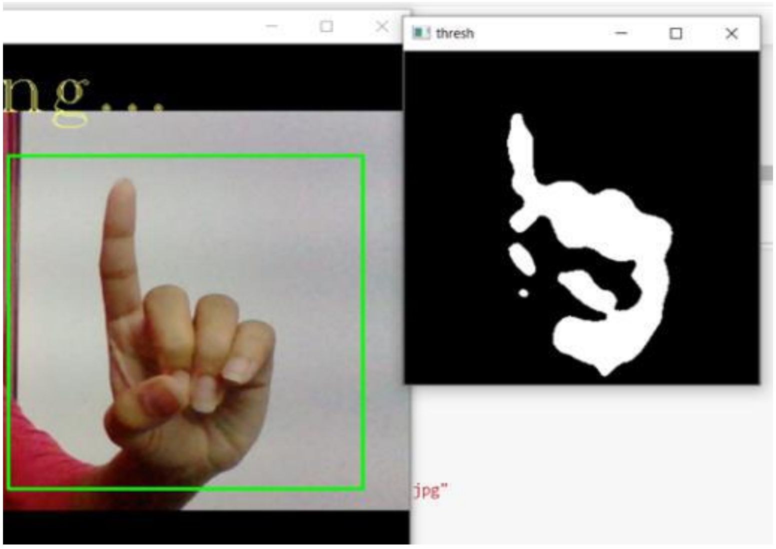

Fig. 5. Capturing multiple images to train the model

5. Storing them in a database.
6. Adding different variations to the captured gestures by flipping all the images.

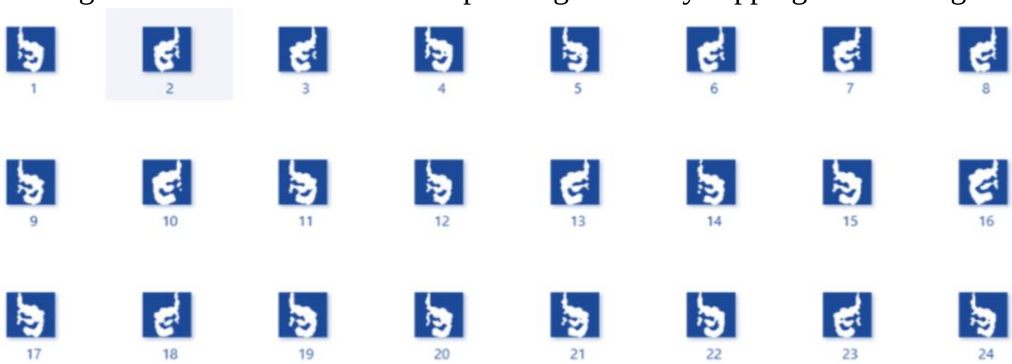

Fig. 6. Different variations of the captured gestures

7. Splitting all the captured gestures into training, validation and test set.
8. View all gestures.
9. Train the Convolutional Neural Networks model using Keras.

10. Open the gesture recognition window which will use the webcam to interpret the trained Sign Language gestures.

RESULTS

In proposed system histogram analysis was used, which is used to separate out the hand from the background image. It focuses on the oriented gradients.

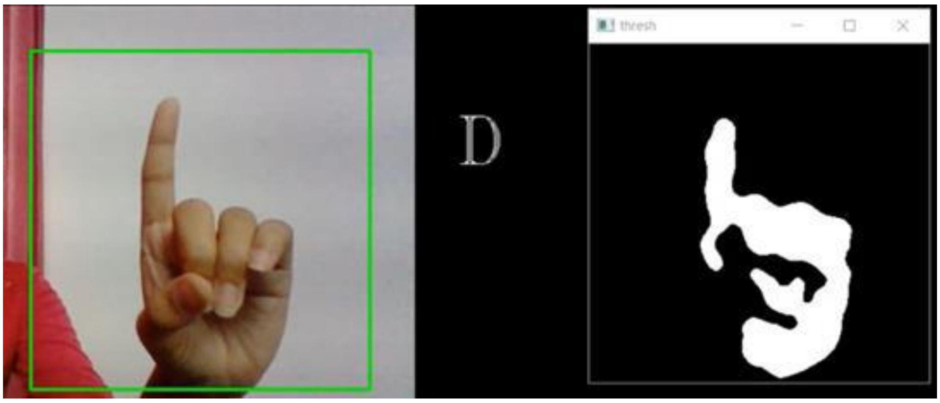

Fig. 7. Output of the Sign Language Recognition Model.

My image dataset is around 300 images, for each gesture, which were divided into two parts in the ratio 80:20 for training and testing. The training images are duplicated and added to the training set. These images are randomly synthesized and train each set for 80 to 100 epochs, one after the other[6]. Artificial synthesis was performed in order to learn the image patterns better so that it will be able to classify the images better in real time. Figure 7 shows the Output of Sign Language Recognition Model.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 61, 91, 32)	896
activation_1 (Activation)	(None, 61, 91, 32)	0
max_pooling2d_1 (MaxPooling2	(None, 30, 45, 32)	0
conv2d_2 (Conv2D)	(None, 28, 43, 32)	9248
activation_2 (Activation)	(None, 28, 43, 32)	0
max_pooling2d_2 (MaxPooling2	(None, 14, 21, 32)	0
conv2d_3 (Conv2D)	(None, 12, 19, 32)	9248
activation_3 (Activation)	(None, 12, 19, 32)	0
dropout_1 (Dropout)	(None, 12, 19, 32)	0
flatten_1 (Flatten)	(None, 7296)	0
dense_1 (Dense)	(None, 128)	934016
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 36)	4644
activation_4 (Activation)	(None, 36)	0

Total params: 958,052
Trainable params: 958,052
Non-trainable params: 0

Fig. 8. Summary of Sign Language Recognition Model

Figure 8 shows the Summary of Sign Language Recognition Model.

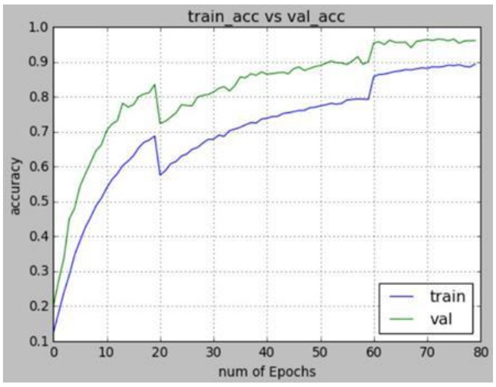

Fig. 9. Accuracy vs Number of Epochs.

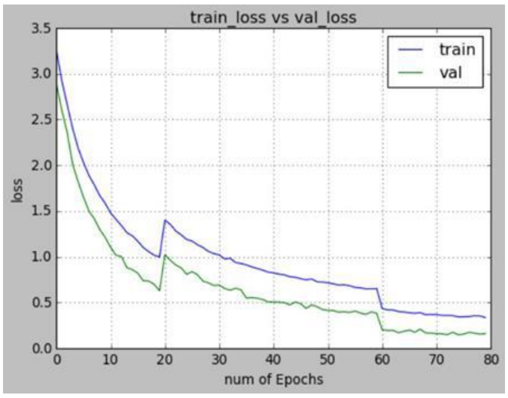

Fig. 10. Loss vs Number of Epochs.

Figure 9 shows the variation of Accuracy with Number of Epochs in proposed system. Figure 10 shows the variation of Loss with Number of Epochs in proposed system. Proposed system gives training accuracy of about 89 percentage with validation accuracy of about 96 percentage.

CONCLUSION

The main goal of this project is to develop sign language translation system that can translate the sign language into text using Convolutional Neural Network. The above work can be solved with better accuracy when we actually consider the segmented hand-gestures. The proposed Sign Recognition Model was developed using Convolutional Neural Networks. Proposed system gives training accuracy of about 85 percentage with validation accuracy of about 90 percentage. It has been observed that the proposed

REFERENCES

1. Xum Pei, “A real time hand gesture recognition and human computer interaction”, Department of Electrical and Computer Engineering University of Minnesota, 2017, pp. 1-8.
2. Hussain, Saxena, Han, J. A. Khan, “Hand gesture recognition using deep learning”, 2017 International SoC Design Conference (ISOCC), Seoul, pp 1-6.
3. Pigou and Lionel, “Sign languages recognition using convolutional neural networks.” Workshop at the European Conference on Computer Vision. Springer International Publishing ,2014.
4. Kuznetsova, “Real-time sign language recognition using a consumer depth camera.” Proceedings of the IEEE International Conference on Computer Vision. Springer International Publishing, 2014.
5. Xum Pei, “A real time hand gesture recognition and human computer interaction”, Department of Electrical and Computer Engineering University of Minnesota, 2017, pp. 1-8.
6. http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf
7. Arman Sabyrov , Medet Mukushev, Alfarabi Imashev, Kenessary Koishybay, Anara Sandygulova, “Towards Real-time Sign Language Interpreting Robot: Evaluation of Non-manual Components on Recognition Accuracy” IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
8. Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan, Celia Shahnaz, Shaikh Anowarul Fattah “Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning” Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference, 2018.