IOP Conf. Series: Materials Science and Engineering **732** (2020) 012082 doi:10.1088/1757-899X/732/1/012082

SIBI (Sistem Isyarat Bahasa Indonesia) translation using Convolutional Neural Network (CNN)

A R Syulistyo*, D S Hormansyah and P Y Saputra

State Polytechnic of Malang, Malang, Indonesia

*arie.rachmad.s@polinema.ac.id

Abstract. Deaf people are one of those with disabilities who cannot speak and hear. Deaf people using sign language for communication who use sign language with hand gestures, body and convey facial expressions using sign language. One important component in sign language is the alphabetical finger or the manual alphabet needed to complete communication. The alphabet of the finger is done by spelling words on spoken language, by spelling letter by letter using a finger. This method is used to spell the name or mention the word. However, not everyone understands sign language, so tools needed to bridge communication between people who are deaf with normal people. One solution that will be offered is to use computer technology as a tool to recognize sign language. The technology is an automatic language translator system that process input images using by using the Convolutional Neural Network (CNN). On this research consist of 3 class such us A, *assalamualaikum* and hallo which get the accuracy respectively 100% for each class.

1. Introduction

Equal rights for persons with disabilities are regulated in UU No. 8/2016 [1], so that people with disabilities get the right like other Indonesian citizens including getting higher education. In addition to the opportunity to get an education like other Indonesian citizens, disabilities have the right to participate in academic activities that support the educational process, one of the activities is participating in conference or workshop.

Deaf person communicates through sign language, there are two sign languages system which used in Indonesia, namely *Sistem Bahasa Isyarat Indonesia (SIBI)* and *Bahasa Isyarat Indonesia (BISINDO)*. However, the government recognize SIBI as the standard system to communicate.

An important component of sign language used by deaf is the finger alphabet or the manual alphabet needed to complete communication. The finger alphabet used for spelling words in spoken language, letter by letter using fingers which can be seen on figure 1 as the illustration. However, it faces problems cause of not everyone understands the sign language which used by deaf persons, so they will be get problem when attending a conference where most participants do not have knowledge of sign language for deaf persons.

To overcome this problem, we need a tool that bridges communication between the person with hearing impairment with participants in the conference or workshop. One technology that will be used as a tool is an automatic translator system that process input images using the Convolutional Neural Network (CNN). We use CNN which belong to neural network technique because has abilities to

automatically extract feature which differentiate among classes [2]. It easier compare with image processing method which should define the feature manually [3].

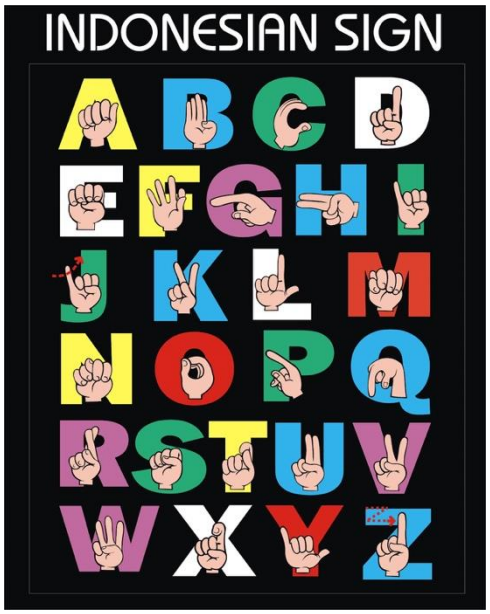

Figure 1. Sign language illustration [2].

2. Background study

2.1. Deaf person

According to Pernamarian Somad and Tati Herawati, stated: "Deaf is someone who has a deficiency or loss of hearing ability either in part or in whole which is caused due to the malfunction of part or all of the sense of hearing, so that he cannot use his sense of hearing in daily life which brings impacts on life in a complex way ".

Sardjono argues that: "Deaf children are children who lose their hearing before learning to speak or hearing loss due to a hearing loss when the child begins learning to speak, sound and language as if lost". According to Prof. Soewito cited by Sardjono in the book Orthopedic Deaf: "Deaf is someone who has severe to total deafness, who can no longer capture speech without reading the lip movements of the interlocutor".

From some of the above understanding it can be concluded that deaf children are children who experience loss of the ability to hear either partially or wholly due to damage to the sense of hearing function either in part or in whole so as to have a complex impact on their lives.

2.2. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a development of the Multilayer Perceptron (MLP) which is designed to process 2D data. CNN is one of the Deep Neural Network methods because it has hidden layer depth. The CNN concept was then matured by Yann LeCun, a researcher from AT&T Bell Laboratories in Holmdel, New Jersey, USA. The CNN model with the name LeNet was successfully applied by LeCun in his research on number recognition and handwriting [3]. In 2012, Alex Krizhevsky with his CNN implementation won the ImageNet Large Scale Visual Recognition Challenge 2012 competition [4]. This achievement become an evidence that the Deep Learning method, especially CNN successful in surpassing other Machine Learning methods such as SVM in the case of images classification.

CNN is now becoming very popular due to high its performance across many data type especially in analyzing image in large amounts of data. CNN can be divided in two parts of process such as the

Feature Extraction Layer and the Fully Connected Layer. The illustration of CNN architecture can be seen figure 2.

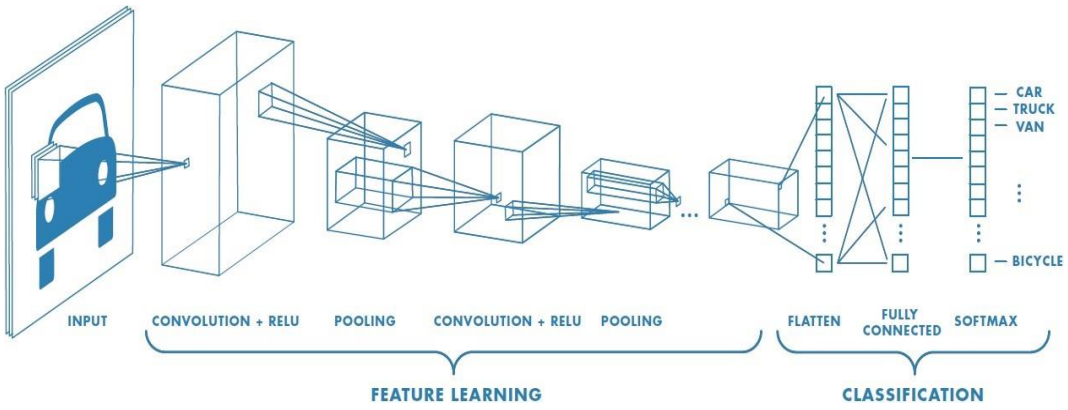

Figure 2. Architecture of CNN.

2.2.1. Feature extraction layer. This process is used to extract image feature in numbers format which represent the pixels of image. Feature Extraction Layer consist of two process such as Convolutional Layer and Pooling Layer. Image convolution is a technique for smoothing an image by replacing the pixel value with pixel values that are corresponding to the original pixel [5]. The convolution process multiply image filter which is certain size with input image. The illustration of convolution process can be seen on Figure 3.

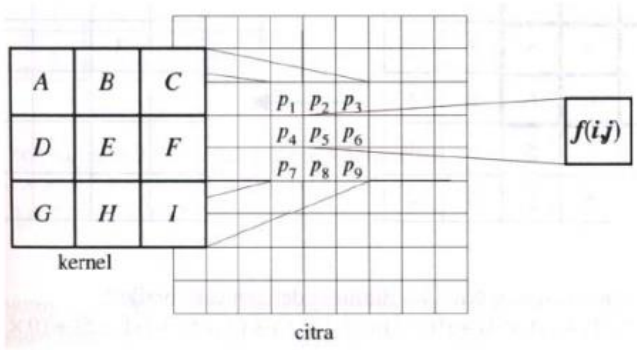

Figure 3. Illustration of convolution process.

The result of above illustration is $f(i,j) = Ap_1 + Bp_2 + Cp_3 + Dp_4 + Ep_5 + Fp_6 + Gp_7 + Hp_8 + Ip_9$. Convolution has 2 parameters, the first is stride which determine how far filter will move horizontally until the on of the image size then move vertically. The smallest stride value, we can get the information more detail however the computation cost higher than use large value of stride. The second parameter is padding which used to manipulate the image dimension by adding value on each side of input image.

The another layer of feature extraction layer is pooling layer which used to reduce feature map dimension. Generally, there are 3 types of pooling operation such as max pooling, min pooling and average pooling. The illustration of max pool with filter size 2x2 pixel and stride 2 can be seen on figure 4.

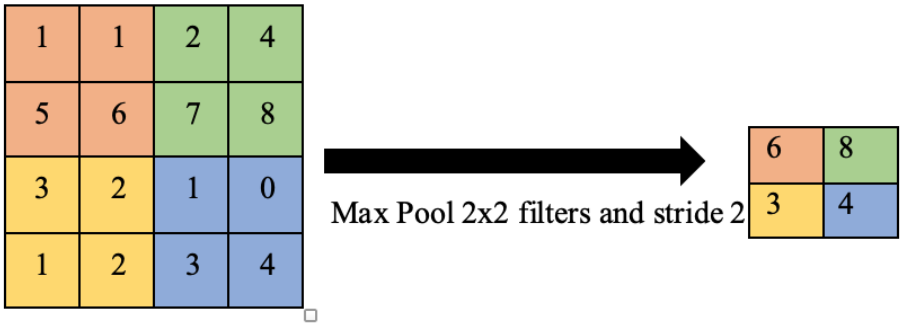

Figure 4. Illustration of max pooling with filter size 2x2 and stride 2.

2.2.2. Fully connected layer. Feature map result from feature extraction process layer is still in the multidimensional array, so it needs to process become one dimensional array by using flatten process which become input of fully connected layer. Fully connected layer has same mechanism of MLP and has the same hyper parameter, namely: hidden layer, activation function, output layer and loss function These layer consist of 2 big process, such as: forwards propagation and backpropagation.

3. Proposed method

3.1. Dataset

We have collected 3 class such as A, *assalamualaikum* and hallo which captured with hand phone camera. The dataset consists of 100 training data and 10 testing data for each class.

3.2. Convolutional Neural Network (CNN)

In this research, we use pre-trained model that provided by tensor flow framework, the list can be seen on GitHub account. We use faster_rcnn_inception_v2_coco model as retrained model in this research because this model has fast speed to detect the object.

3.2.1. Research design. Figure 5 is research design of this paper, the first step input data in the form of digital image will be re-scaled then will be processed by CNN which will eventually produce weight. The weight of training result will use to translate the input image into certain class in the testing process.

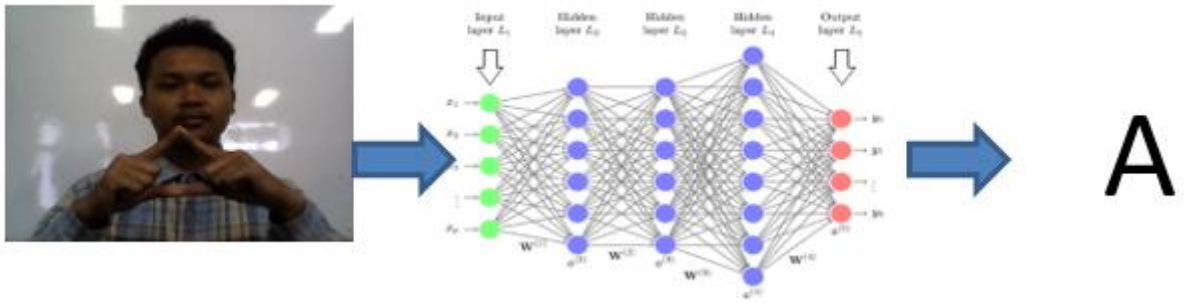

Figure 5. Illustration of sign language translation.

4. Experiment

4.1. Evaluation

The evaluation of this system uses accuracy formula which can be seen in equation 1. There are 4 condition should be measure namely true positive, true negative, false positive and false negative. True positive is the condition when system predict correctly compare with the desire label, the opponent of this condition is false negative. The another condition is false positive when the system detect image which actually do not belong to certain class however the system detect it as member of the class, the

opponent of this condition is true negative when the system correctly reject the input image which do not belong to class. This illustration can be seen on table 1.

$$Accuracy = \frac{true\ positive + true\ negative}{true\ positive + true\ negative + false\ positive + false\ negative} \quad [1]$$

Table 1. Metric of performance evaluation.

	Prediction of the model: Positive	Prediction of the model: Negative
Truth: positive	TP (True positive)	FN (False negative)
Truth: negative	FP (False positive)	TN (True negative)

4.2. Result

System can detect input image with accuracy 100% for each class with total 30 images test. The classification can be seen on figure 6-9.

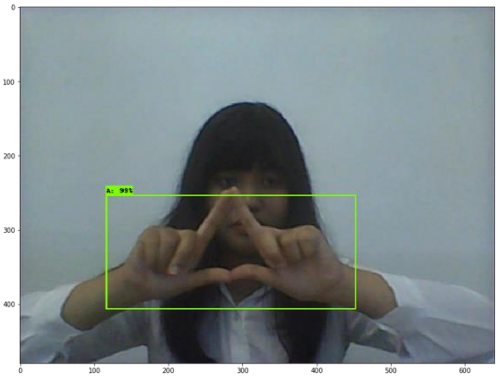

Figure 6. Illustration Class A detected as Class A

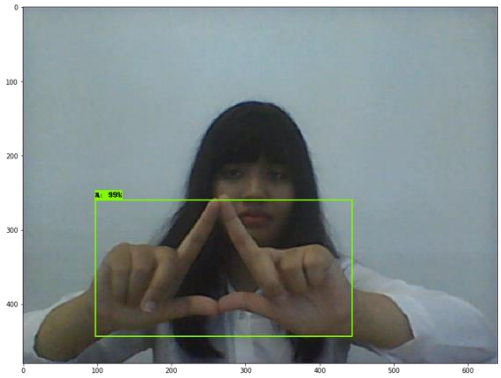

Figure 7. Illustration Class A detected as Class A

Figure 8. Illustration Class *assalamualaikum* detected as Class *assalamualaikum*

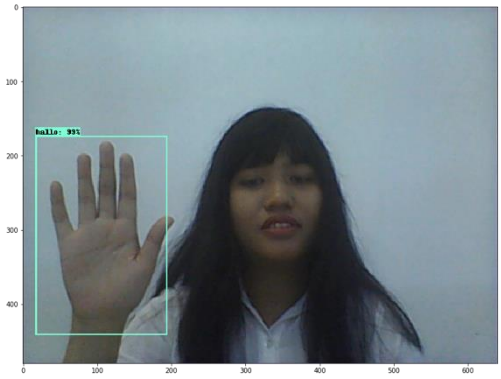

Figure 9. Illustration Class hallo detected as Class hallo

5. Conclusion

CNN (Convolutional Neural Network) use as main method to extract sign language features and predict class labels for each image. Based on the experiment CNN able to translate input image into an expected class label with accuracy 100% from 3 classes such as *assalamualaikum*, hallo and A. In future research, we will add class to translate and implement the model into minicomputer.

References

[1] Kementerian Hukum dan HAM RI [Online]. Available: http://peraturan.go.id/common/dokumen/ln/2016/uu8-2016bt.pdf. [Accessed 12 September 2019].
[2] Budiyono E 2012 *SLMBN Kota Tegal Mesemo* [Online]. Available: http://mesemo.blogspot.com/2012/06/sistem-isyarat-bahasa-indonesia-sibi.html. [Accessed 12 11 2019].
[3] LeCun Y, Boser B E, Denker J S, Henderson D, Howard R E, Hubbard W E and Jackel L D 1990 Handwritten digit recognition with a back-propagation network *In Advances in neural information processing systems* pp 396-404
[4] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet Classification with Deep Convolutional Neural Networks *in Advances in Neural Information Processing Systems, Stateline, Nevada*
[5] Putra I W S E 2016 *Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember)