VOL. 20, NO. 2-3, 2021, 70-72
www.elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

Development of a Malaysian Sign Language Interpreter using Image Recognition for the Community to Understand the Deaf

Nurul Ashikin binti Mohammad[1*], **Kian Sek Tee** [1], **Chin Fhong Soon**[1], **Toong Hai Sam**[2] and **Ruzairi Abdul Rahim**[3]

[1]Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia.
[2]Faculty of Business, Communications and Law (FOBCAL), INTI International University, Nilai, Negeri Sembilan, Malaysia.
[3]School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia.

[*]Corresponding author: ashikinmohammad11@gmail.com, Tel: 607-4537000 Ext:1405, Fax: 607-4564502

Abstract: In Malaysia, Person With Disabilities (PWD) with hearing problems or commonly known as the deaf, struggle to have a conversation with the communities who do not know or how to do sign language efficiently. The consequence is hearing PWD received unequal treatment in jobs and learning opportunities. This project aims to develop a Malaysian Sign Language interpreter to convert hand signs to texts, in order to facilitate the conversation between hearing PWD and the communities. The system would implement a camera and vision system to capture images of hand signs. Four hand signs have been selected. Stacks of these images were processed digitally using deep learning method, and eventually the trained network could recognize the hand signs successfully. This feasible study suggests that the proposed setup could be further implemented to train more hand signs and enrich the hand signs-to-text vocabulary.

Keywords: Malaysian Sign Language, Person With Disabilities, SSD, interpreter.

Article History: received 25 May 2021; accepted 12 June 2021; published 15 October 2021.

1. INTRODUCTION

Communication is a two-way process on how peoples exchange messages between each other in order to have the same understanding or subject. There are two types of communication; verbal communication and non-verbal communication. Verbal communication is about language; it can be either spoken or written. Meanwhile, nonverbal communication refers to communication that occurs through means over words [1], [2]. For example, sign language, body language and facial expression, in simple words it is a wordless signal. For community of hearing PWD, they use a non-verbal language to deliver messages to other parties. In 2008, Malaysia Sign Language (MSL) was acknowledged to be the main sign language for the community of hearing PWD in Malaysia [3]. Even though MSL is used widely by the hearing PWD in Malaysia, yet they still face difficulties of efficient communication in their daily life, especially with non-sign language users. This is because most verbal language users are not used to hand sign language as they can speak well to interact with others. This creates a communication barrier between the normal communities and hearing PWD.

Based on a recent yearly statistic report 2018 [4] (latest available at the time of writing this paper) published by Department of Social Welfare, Ministry of Women, Family and Community Development, Malaysia, the total population in Malaysia at 2018 was 32,382,300 people and the total of registered PWD were 497,390 persons, which consists of 1.53% out of total population. Categorized PWD include visual impaired, hearing, physical, learning disabilities, speech, mental and others. Statistically, there are 44,523 hearing PWD, contributing 8.95% of total PWD; 6,635 under 18 years old, 37,888 above 18 years old. Every year, a few hundreds of hearing aids which worth nearly 1 million Malaysia Ringgit were gifted. Hearing PWD population are found in all ages, ethnics and states.

The hearing PWD face obstacles in their daily life. Normal community has insufficient exposures and knowledge about MSL and some still held prejudices against PWD [5]. Prejudices and severe communication break-down cause them to gain or secure less opportunities in jobs and learning [6]. In Malaysia, a portion of hearing PWD leads a hardship in their life.

Hence a research gap is mitigated to be ineffective communication due to inconvenience to understand MSL. It is hypothesized that a communication aid could bridge the communication and improve the effectiveness of communication between the normal community and hearing PWD. Hence, this project would present the development procedure to create a communication aid, which is an application software working in real time to interpret MSL into text.

2. METHODOLOGY

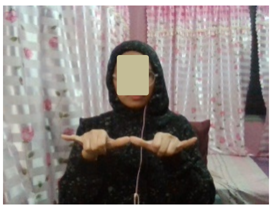
In recent years, deep learning architectures for object recognitions in computer vision have proven to be sufficiently fast for real-time applications with acceptable accuracy. One of it is Single Shot Multibox Detector, (SSD) [7], which claimed to have 74.3% mean Average Precision (mAP) and to be able to process images at 59 frame-per-second (FPS). This paper deploys SSD framework to recognize MSL and hence to achieve the aim as the communication interpreter. Detail working principles and algorithms of SSD was reported and elaborated by its creator, W. Liu et al [7].

2.1 Experiment Setup

A simple webcam is used to capture images during the collections of images. Pre-defined images of resolution 640 x 480 were saved. Four MSL messages (also known as the classes for SSD) as illustrated in Figure 1, are experimented, i.e. “Hello”, “Sorry”, “Stay”, “Home”. We have deployed open-source mobile inference box, SSD *MobileNet* v2 320x320, shared by TensorFlow 2 Detection Model Zoo [8], for the purpose of hand signal recognition. The setup is illustrated in Figure 2.


(a)


(b)


(c)


(d)

Figure 1. The classes of hand signs, (a) “Hello”, (b) “Sorry”, (c) “Stay”, (d) “Home”

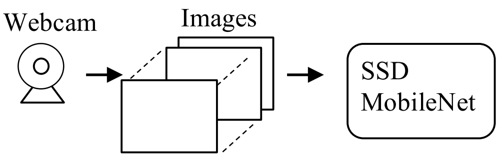


Figure 2. Experiment setup

2.2 Preparation for model training

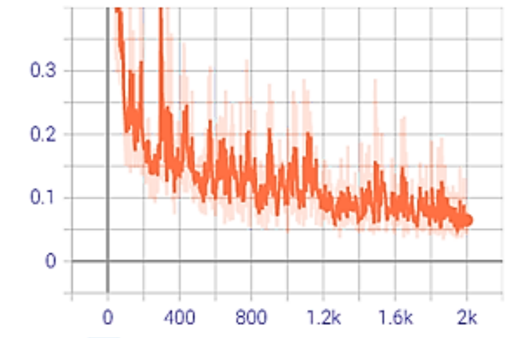
For the purposes of train and test, five images were taken for each class where four for train, one for test. A total of 4 classes x 5 images = 20 images, were taken. These images were labelled and annotated. For each image, its class was labelled, and a bounding box was drown manually at the location of the hand signal. Image annotation was conducted using *LabelImg* [9]. A metadata of each image annotation was generated and was saved as a XML file. The images and their image annotations were sorted and saved in known folders labelled as “train” and “test”. These data were fed into SSD for train and test.

3. THE RESULTS

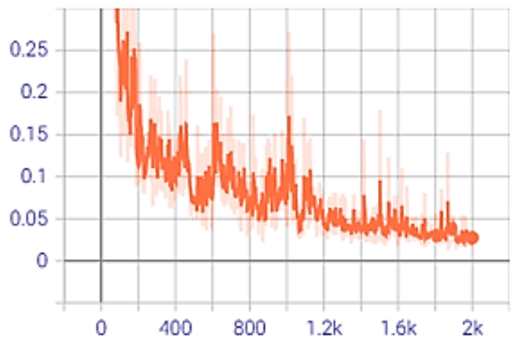
3.1 Training results

SSD training was conducted on HP Laptop 15s, running on AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10GHz, RAM 4GB. By default, Tensorflow V2 turns on GPU mode. Total training time took 1 hour and 45 minutes.

Two critical types of losses were monitored, i.e. Classification Loss and Localization loss. Classification Loss measures the loss between the classes with the predictions; Localization Loss measures the loss between the actual bounding boxes with the predictions. During training, SSD would update the network while the losses as shown in Figure 3, steps to an acceptable low value. Eventually, the training achieves 72.5% mPA.



(a)



(b)

Figure 3. Losses versus number of steps (a) Classification Loss, (b) Localization Loss

3.2 Trials

The trained model was tested, and a new image was trialed randomly as shown in Figure 4. The location of the hand sign was self-identified with a bounding box and its classes ("Home") was labelled with confidence percentage (100%). The hand sign was successfully being recognized. Other hand signs ("Hello", "Sorry" and "Stay") were tested and also were being successfully recognized. The results (the bounding box and the class) displayed almost instantaneously, no significant time delay to human. It indicates a good response for real-time application.

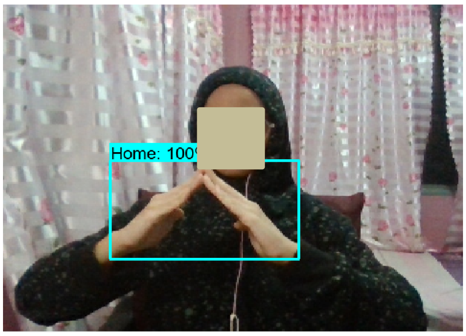


Figure 4. Recognized hand sign – "Home"

4. CONCLUSION

It was mitigated that the community of hearing PWD could not be understood properly due to MSL is not common among the normal community. A communication aid is proposed and being developed. It deploys low-cost webcam and an open-source deep learning algorithm, i.e. SSD, to interpret MSL into text. Once the model was properly trained, it could be re-used to recognize hand signs of the trained classes. The results are instantaneous with high accuracy. Future work could enrich the hand signs-to-text vocabulary and explore the effectiveness of communication using this application aid.

ACKNOWLEDGMENT

The work is supported by Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia.

REFERENCES

[1] M. Zuckerman, B. M. DePaulo, and R. Rosenthal, "Verbal and Nonverbal Communication of Deception," vol. 14, L. B. T.-A. in E. S. P. Berkowitz, Ed. Academic Press, 1981, pp. 1–59.

[2] R. Buck and C. A. VanLear, "Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior," *J. Commun.*, 2002, vol. 52, no. 3, pp. 522–541.

[3] S. M. M. Rashid, M. H. M. Yasin, and N. Sahari, "Penggunaan Bahasa Isyarat Malaysia (BIM) di dalam Terjemahan Maksud Surah Al-Fatihah," *GEMA Online® J. Lang. Stud.*, 2017, vol. 17, no. 4, pp. 209–224.

[4] K. P. W. D. M. Jabatan Kebajikan Masyarakat (JKM), "Laporan Statistik 2018."

[5] M. Isaruddin, M. W. Lokman, and A. S. Amin, "Isu Pekerjaan Orang Kurang Upaya Pendengaran: Kajian Kes di Selangor," *J. Wacana Sarj.*, 2019, vol. 2, no. 2, pp. 1–10.

[6] R. Mohd Amin, "Garis panduan pelaksanaan dasar inklusif orang kurang upaya di institusi pendidikan tinggi," *Kementerian Pendidikan* Malaysia, 2019.

[7] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision -- ECCV 2016*, 2016, pp. 21–37.

[8] "TensorFlow 2 Detection Model Zoo." https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md (accessed Mar. 18, 2021).

[9] "LabelImg." https://github.com/tzutalin/labelImg (accessed Mar. 20, 2021).