Filipino Sign Language Alphabet Interpreter

Alexander James Enriquez
College of Computing and Information Technology
National University
Manila, Philippines
enriquezap@student.national-u.edu.ph

Jul Seth Mari Mercado
College of Computing and Information Technology
National University
Manila, Philippines
mercadoj@student.national-u.edu.ph

Francis Stephen Lee
College of Computing and Information Technology
National University
Manila, Philippines
Leefr@students.national-u-edu.ph

Sherdon Rappah Esquejo
College of Computing and Information Technology
National University
Manila, Philippines
esquejosp@students.national-u.edu.ph

***Abstract*—Filipinos who are hearing impaired or completely deaf rely on sign language for communication. For the average vocal Filipino, communicating with them requires considerable effort, compounded by the scarcity of sign language education opportunities. This study focuses on utilizing object detection and machine learning to facilitate Filipino Sign Language (FSL) learning, starting with alphabetic signs. This approach aims to overcome the initial hurdle learners face in understanding sign language, paving the way for a machine learning model that simplifies FSL acquisition. By harnessing Machine learning techniques, this research aims to streamline the learning process and make it more accessible for the hearing-impaired community. This study modified a saved model from TensorFlow and utilized transfer learning when training the model that's focused on image classification.**

Keywords: object detection, machine learning, Filipino sign language, hearing-impaired, saved model, image classification, transfer learning

I. Introduction

Communication is a vital aspect of daily life, enabling people to express their thoughts, needs, and emotions. For the deaf community, this essential interaction is facilitated through sign language However, the limited use and teaching of FSL outside the deaf community pose significant challenges, creating barriers for effective communication and social integration. Traditional methods of learning sign language are often time-consuming, requiring extensive resources and effort. In the Philippines the lack of proper tools and regulated teachings when it comes to sign language creates bigger gaps.

The Filipino deaf community encounters substantial communication obstacles due to the limited dissemination and instruction of Filipino Sign Language (FSL) beyond their immediate circles. This deficiency in widespread understanding erects barriers for deaf individuals and those trying to engage with them. Conventional methods of learning sign language often prove arduous and time-intensive, necessitating more streamlined and captivating alternatives. Recent advancements in computer vision and deep learning technologies present promising avenues to address these challenges. These technologies offer innovative tools to enhance FSL learning, potentially fostering improved communication and inclusivity within the community. Moreover, while some technological advancements have been made in sign language recognition using computer vision and machine learning, there is still a lack of accessible, effective tools designed for FSL. These tools are crucial in providing more efficient and engaging methods for learning and using FSL, thereby promoting better communication and inclusivity.

This study aims to address this gap by developing a Filipino Sign Language alphabet interpreter using computer vision and machine learning techniques. By focusing on the recognition of static alphabet signs, the study seeks to provide an accessible and efficient tool for beginners to learn FSL, ultimately fostering better communication and inclusivity. This research will contribute to the broader field of educational technology and promote social integration by bridging the communication gap between hearing and deaf communities in the Philippines.

II. Related Work

There is a daily increase in the number of people affected by speech impairments and hearing loss, which are becoming more common. The World Health Organization (WHO) estimates that 430 million individuals worldwide, or 5% of the total population, suffer from speech impairment, and by 2050, that figure is predicted to increase to 1 in 4 people. Hearing loss has highly negative effects.[6] With the current rate of people who get affected by hearing loss it presents more desire for a sign language model that can accommodate deaf people better.

The study by Murillo et al. (2021) presents "Speak the Sign," a real-time sign language-to-text converter application focused on basic Filipino words and phrases. The research addresses the need for accessible communication tools for the Filipino deaf community. The authors developed an innovative system that utilizes deep learning techniques to recognize Filipino Sign Language gestures from images, enabling real-time translation into text. This application can enhance communication accessibility and inclusivity for the Filipino deaf community, providing a valuable tool for everyday communication and interaction [1].

The study of sign language recognition can provide a new horizon for the deaf community, signifying hope to express themselves better and communicate with people who aren't knowledgeable about sign language. The integration of deep learning with sign language recognition also provides better results and high potential, especially for applications. Reddy et al. (2024) highlighted the potential of machine learning and deep learning in sign language recognition (SLR) technology, emphasizing its role in improving communication and accessibility for the hearing impaired. They noted that traditional methods often struggle with the diverse vocabulary and grammar of sign languages, making deep learning approaches, such as convolutional neural networks (CNNs), more suitable for SLR tasks [2]. Several studies have proved that the use of deep learning provides remarkable success when it comes to sign language recognition. Especially with the variety of sign languages around the world, the use of deep learning and computer vision may just be the key to closing the gap in enhancing accessibility for the hearing impaired.

Adeyanju et al. (2021) conducted a comprehensive review of artificial intelligence methods applied in sign language recognition (SLR) systems, focusing on decision support and intelligent systems. They analyzed 649 publications extracted from the Scopus database over the past two decades to understand research progress and publication patterns in intelligent-based SLR. The study aimed to provide insights into the application of machine learning and deep learning in SLR and to highlight open issues and research areas for future consideration. The review highlighted the challenges posed by the diversity of sign languages, with over 7000 present-day sign languages exhibiting variability in motion position, hand shape, and body part positions. Despite these challenges, researchers have demonstrated remarkable success in developing intelligent solutions for automatic sign language recognition (ASLR) systems [3].

Recent research on Continuous Sign Language Recognition (CSLR) by Alyami et al. (2023) has made significant advancements in deep learning techniques to enhance the recognition of sign language sequences. Their comparative study evaluates various CSLR models, such as Visual Alignment Constraint (VAC) and Self-Mutual Distillation Learning (SMDL), across datasets like RWTH-PHOENIX-Weather-2014 and ArabSign. The study addresses challenges such as recognizing gestures without clear boundaries and the variability in performance by different signers. By establishing new benchmarks, Alyami and Luqman contribute valuable insights into the development of more accurate and generalizable CSLR systems [4].

Thakar et al. (2022) contributes to this body of research by addressing accuracy limitations in existing SLR models. Their use of transfer learning on the American Sign Language (ASL) dataset to improve predictions demonstrates the applicability and effectiveness of leveraging existing data for training new models, aligning with the trend towards utilizing deep learning techniques for better performance in sign language recognition tasks [5]. While significant advancements exist, challenges such as handling variations in signing styles, different sign languages, and background noise persist. Researchers continue to explore new methods and technologies to improve the accuracy and usability of sign language recognition systems.

A wide range of information and communication technologies (ICTs) are employed in the identification and interpretation of various sign languages utilized by individuals with speech disorders. However, many people with speech impairments find that some of these technologies are either too expensive or socially undesirable. Iftikhar Alam, Abdul Hameed, Riaz Ahmad Ziar (2024) [6]

This study reveals that even with the existing systems that provide sign language translations their reach is limited to their nativity and the climb to reach a system that provides for everyone is still a few steps far to grab.

III. METHODOLOGY

A. *Dataset*

The data used within this test has been collected manually by the researchers themselves. The images acquired in this model were the 26 letters of the alphabet from the letter A to Z that are signed in the Filipino Sign Language and manually sectioned into their respective labels. The basis of these letters referenced images of hand signs labelled on A to Z. Due to the object detection method using still images, the researchers had to use different signs for the letters J and Z, both needing the use of writing the sign out in the air.
The 3640 images were split evenly between all letters with 2912 for Training and 728 for Validation, each letter receiving 112 used for Training and 28 images for Validation with half of the images are from the right hand and the other half are from the left hand. To ensure the model’s learning rate, the background of said images are also kept as clean and flat as possible to keep focus on the hands.

B. *Preprocessing*

Images that were blurry were removed from the data set as it makes it harder for the model to identify the hand. The images used within the model were in different image sizes and quality and were standardized before being processed to a height of 384 with a width of 384.

C. *Model*

This section will show the different approaches taken with fine tuning a saved model to cater for making a Filipino sign language model. The saved model is called: Efficientnet it is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. Unlike conventional practice that arbitrary scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients. The choice of optimizer also plays a critical role in the training process, influencing the convergence rate and final accuracy of the model. The optimizers that were tested were Stochastic Gradient Descent (SGD) and Adam.

SGD is a variant of the Gradient Descent algorithm used for optimizing machine learning models. It addresses the computational inefficiency of traditional Gradient Descent methods when dealing with large datasets in machine learning projects. The researchers defined the model with an initial of 0.2 dropout followed with a dense layer of 64 neurons and ReLU as its activation function followed by another 0.2 dropout with a final layer of 26 neurons of which will be the final layer for all succeeding trials with the model. Categorical Cross entropy was used for the loss function, SGD for the optimizer and a learning rate of 0.005, batch size of 10. This model was trained for 25 epochs.

The next test defined the model with a dense layer of 128 ReLU as its activation function a dropout of 0.2 and another layer which consists of the same specifications followed by a dense layer of 256 neurons with ReLU as its activation function and a drop rate of 0.2. This model was trained with a batch size of 32 and the same specifications for the optimizer, learning rate, and epochs as the previous model.

The final test defined the model with almost the same layouts as the first two models, but all three dense layers consisted of 128 neurons each. The biggest change from the three is this final test was tested with 30 epochs instead of the regular 25. This model was tested with a batch size64.

Moving on to the Adam optimizer, the first test defined the model a dense layer of 256 neurons followed by a dropout of 0.2 then another dense layer of 256 with the same dropout value and with the third layer of 128 with a drop rate of 0.2. this model had a batch size of 10, 30 epochs and a learning rate of 0.005

For the following iterations of the Adam optimizer all specifications were the same. The only factor that was changed was the batch sizes, which were 32 and 64 respectively.

IV. Results and Discussion

Chapter three identified the different methods that were used in training a model for Filipino sign language. This chapter explores the outcomes of said training. Each optimizer was given different batch sizes 10, 32 and 64 respectively. With SGD as an optimizer, the model's outcome was stable, and the accuracy of each was close to each other. While closely monitoring each run to stop it if there were signs of overfitting the SGD model showed no sign of any overfitting as with each epoch the validation accuracy steadily went up while the validation loss slowly descended. The only models that the researchers stopped early were the ones that use the optimizer Adam with batch sizes 10 and 32 as the validation accuracy simple did not go up and the loss was getting worse and worse which would indicate that there is over fitting when it comes to those sizes. while with a batch size of 64 it yielded a high accuracy score of 92%

When comparing both optimizers with their respective batch sizes the accuracy difference is significant especially when looking at SGD with a batch size of (10) and Adam (10). The outcome of most recall values has close and almost no differences among all the results except for Adam (10).

SGD (32) has an accuracy of 0.90 but a precision score of 0.18 which indicates that the model has a high chance of predicting them with a false positive result. While Adam (32) has better precision but when it comes to its accuracy it falls behind greatly as there is a .21 difference with them.

The best performing optimizer among everything that was tested is Adam (64) with high precision, recall, F1 score and decent accuracy. It outperforms all the runs that the researchers tested in training the model. While its counterpart SGD (64) has the second worst precision among the results in all metrics

Table 1.

Metric Scores

Optimizer	Precision	Recall	Accuracy	F1 Score
SGD(10)	0.20	0.99	0.95	0.33
SGD(32)	0.18	0.99	0.90	0.30
SGD(64)	0.10	0.98	0.82	0.18
Adam(10)	0.05	0.79	0.69	0.09
Adam(32)	0.32	0.96	0.79	0.48
Adam(64)	0.32	0.99	0.92	0.48

All of the results were trained using the aforementioned pretrained model. The varying results among the different batch sizes and optimizer reveal that the model is very good at identifying positive instances (high recall) but is also predicting a lot of false positives. This may suggest a class imbalance and not enough data augmentation done within the model.

V. CONCLUSION

In this paper, the approach towards a more effective neural network was done to a successful degree with each iteration of the model improving upon the previously tested neural network. The way this was done, however, has had its flaws. The use of datasets that split the hands between left and right might prove to be useful on a larger data set for training, however it poses some issues that could still be worked on in its current state. The importance of using this neural network in this manner remains as a step towards the improvement of our communication towards those who are incapable of speaking.

One great limiting factor of this research is the digital power and timeliness of it as the ability to train models does not come simple nor quick. Training models also require larger scale data sets. One recommendation is to use more proper preprocessing methods, to clearly define the shape of a hand making signs by isolating the hand itself as well as using an even larger data set to better improve the model's accuracy when predicting.

REFERENCES

[1] *] S. C. M. Murillo, M. C. A. E. Villanueva, K. I. M. Tamayo, M. J. V. Apolinario, M. J. D. Lopez, and Edd, "Speak the Sign: A Real-Time Sign Language to Text Converter Application for Basic Filipino Words and Phrases," Central Asian Journal of Mathematical Theory and Computer Sciences, vol. 2, no. 1, pp. 74-92, 2021.* ***URL****:https://cajmtcs.centralasianstudies.org/index.php/CAJMTCS/article/view/92/74.*

[2] *] V. Reddy P., V. V. Reddy B., and Sukriti, "Sign Language Recognition based on YOLOv5 Algorithm for \the Telugu Sign Language,"arXiv,Apr.2024.****URL****:https://arxiv.org/abs/2406.10231.*

[3] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine Learning Methods for Sign Language Recognition: A Critical Review and Analysis," Computers & AI Journal, vol. 12, no. 11, pp. 200056, Nov. 2021.URL:https://www.sciencedirect.com/science/article/pii/S2667305321000454.

[4] S. Alyami and H. Luqman, "A Comparative Study of Continuous Sign Language Recognition Techniques,"arXiv,Jun.2023.[Online]. **URL**: https://arxiv.org/pdf/2406.12369.

[5]] S. Thakar, S. Shah, B. Shah, and A. V. Nimkar, "Sign Language to Text Conversion in Real Time using TransferLearning,"arXiv,Nov.2022.URL: https://arxiv.org/pdf/2211.14446.

[6] Iftikhar Alam, Abdul Hameed, Riaz Ahmad Ziar" Exploring Sign Language Detection on Smartphones: A Systematic Review of Machine and Deep Learning Approaches" URL: Exploring Sign Language Detection on Smartphones: A Systematic Review of Machine and Deep Learning Approaches - Alam - 2024 - Advances