American Sign Language Interpret using Web Camera and Deep Learning

Kurt Christian G. Fernandez, Brian Quintin A. Paredes, Janiño I. Perfecto and Joel C. De Goma
School of Information Technology
Mapúa University – Makati
Makati, Philippines
kcgfernandez@mymail.mapua.edu.ph, qbaparedes@mymail.mapua.edu.ph,
jiperfecto@mymail.mapua.edu.ph, jcdegoma@mapua.edu.ph

Abstract

In this paper, the proponents have utilized the deep learning models of 3dCNN and LSTM along with OpenCV for webcam functionalities and Google's MediaPipe Framework to develop the American Sign Language Interpreter System. The objective of this study is as follows: First is to test the proponents' own dataset on Tran's network model (Tran et al. 2020) and gather experimental results for comparison. Second, The proponents did a modification of Tran's model network by adding an LSTM layer to accommodate the temporal structure of the proponents' dataset. Overall, the metrics of Training Accuracy and F1-Score were used as the basis for the performance of each model network - Tran's network model performed well by achieving an 81.6% accuracy during training and 81.8% on the F1-Score in real-time. In comparison, the proponent model achieved an 89.9% accuracy on training and 89.8% in real-time. The most notable difference is during real-time, as the proponents' model classified the gestures more correctly. It uses the sequence prediction that is made possible by the LSTM layer.

Keywords
Deep Learning, Object Detection, MediaPipe, Computer Vision, OpenCV, 3dCNN, LSTM

1. Introduction
According to the World Health Organization (WHO), the total percentage of people suffering from hearing impairment is five percent (5%) (World Health Organization 2021). These are a collection of people born deaf and people who lost their hearing throughout their lives. That approximates to 430 million people worldwide having a hearing impairment. It is projected that this number can rise to 630 million by 2030 and 900 million by 2050 (World Health Organization 2021). That means five percent of the world's population communicates through sign language and gestures. One of the most potent means of communication among humans is through gestures (Oudah et al. 2020).

Sign Language uses the visual-manual modality to explain the meaning of a specific gesture (Quer et al. 2019). To date, there are a total of thirty-eight (138) different types of sign language that are known, and most of them have the same linguistics but differ in gestures (Media 2020). In addition, the researchers chose the most prominently used sign language in the world, which is the *American Sign Language* (ASL). This paper will focus on developing a system that can detect the mentioned sign languages to bridge the gap between the people who are unable to comprehend and those sign users.

1.1 Objectives
The study aims to use the previous deep learning model network towards the proponent's ASL gesture dataset. Specifically, it aims to determine the network model of (Tran et al, 2020) in distinguishing the ASL gesture dataset. Moreover, it aims to introduce and utilize the LSTM model when predicting the sequence of hand gestures.

2. Literature Review

According to Tran et al (2020), the traditional means of interacting between humans and computers are lacking flexibility and to address this concern various methods are implemented such as speech recognition and body-based language interaction. The latter is more suitable for most cases since this is a familiar method of communication.

According to Nguyen et al (2021), the research to extract the 3D human pose will be conducted by utilizing all the frames of the continuous gesture sequences. Through the RGB hand, ROI localizes from the 3D hand palm position, when the hand palm stands still and over the spine base joint. They extracted the 3D position of the finger joints using the 3D hand pose estimation algorithm. They also estimated the 3D hand pose by using the real-time 3D hand joints tracking network of OccludedHands. Lastly, with the use of the 3D_ResNet architecture, the researchers can classify the gestures.

A study by Al-Hammadi et al (2020) claimed that there are two main purposes of gesture recognition, these two purposes are: To help with the growing community of deaf and hard-of-hearing population. And it has extended the use of vision-based and touchless applications. The proposed system utilizes the 3DCNN architecture for spatiotemporal feature learning using two approaches. In the first approach, features from the entire video sample were extracted with the use of 3DCNN. In their second approach, with the use of the 3DCNN, they aimed to enhance the temporal dependency of the video frames. In addition, their first approach consists of three main phases: video preprocessing, feature learning, and classification. In the video preprocessing phases, the input video was converted into an RGB frame sequence. Linear sampling was applied to preserve the order of the selected frames and fix the length of 16 frames; the corresponding indices of these frames are calculated where len(input) is the length of the input sequence. Spatial dimension normalization was used to overcome variations in the heights and distances of the signers from the camera. The final step was to resize the cropped square frames into 112x112 pixels.

According to Pothanaicker (2019) The paper proposed the integration of a convolutional neural network and long short-term memory recurrent neural network for processing the video. The process for the convolutional is the given output produces the informative spatial features. The features that are extracted are directed to the long- term memory module to generate temporal features. The feature maps of the long short-term memory component are fed to the proposed attention element. This process captures the highly valuable informative features in the frame video.

According to Varga et al (2019) The purpose of no-reference video quality assessment algorithms that are based on the long-short term memory network is a pretrained convolutional neural network that is introduced. The study proposed the results of the experiments on KoNVID-1k demonstrate the proposed method outperforms the state-of-the-art algorithm. The results of the study are confirmed using the test on the LIVE Video Quality Assessment Database, which consists of artificially distorted videos. The study uses a reliable method of assessing the quality of digital videos through subjective evaluation. The paper also introduced an architecture for NR-VQA that utilizes deep features extracted from a pre-trained CNN and LSTM network for sequence-to-one regression The main purpose of the objective VQA is to design mathematical models that are able to predict the quality of the video assessed by humans.

3. Methods

3.1 Data Processing

In Processing, frames from the video dataset were extracted, then cropped and centralized using the interpolation method to estimate values and know points. In addition, the extracted frames were resized to have a better fit for the model (Dong, 2018). The image above is a sample of the captured frame from a video dataset with a resolution of 1920x1080 pixels that will undergo downscaling to fit the proponent's network model better. Part of the data processing was implementing the detection of hand movements from every extracted frame. As a solution, the proponents have utilized the camera function from OpenCV to test if the webcam functions correctly. In addition, the hand tracking library from MediaPipe was utilized as it uses two (2) backends that are essential for this study. First, Palm Detection crops images and focuses on the palm area. Next, the Landmark module finds the twenty-one (21) various indicators of the hand and is also used to identify the multiple positions of the hand GeeksforGeeks (2019) and Halder et al. (2021). Furthermore, the Hand class from the MediaPipe library was utilized as well to create hand objects that became the basis for the creation of the custom hand landmarks and points for the different ASL gestures. For the feature extraction part, the

initial step was to define the directory paths (11 gesture classes) on where the data would reside. This was done programmatically using the OS. Path module from python. In addition, 30 frames or sequences from every single input data were extracted using OpenCV and were also processed using Media Pipe to determine the hand landmarks. Next, the sequence length or number of frames for sequential patterns is set to 30 as seen in Tran's study - to represent each sequence or frame, a data transformation from image to the array was conducted. The newly created directory contains the processed frames for each gesture class that is now represented in an array format, which includes the image data (Height, Width, Color Channel). Furthermore, the proponent's network model will take in each image in *height, width & color channel* (Ramalingam 2022). There are 30 sequences that represent each iteration of the gesture per class.

3.2 Custom Model-Building

The proposed network model is as follows: four (4) ConvNets, two (2) Max-Pool layers with 50% dropout rate, one (1) LSTM layer, two (2) Fully-connected layers with 50% dropout and one (1) Softmax layer for output. The experiment was conducted fifteen (15) times to acquire varying outputs and obtain various performance results. In addition, an ensemble model was utilized as it provides better performance or higher accuracy compared to a single model (Brownlee 2021).

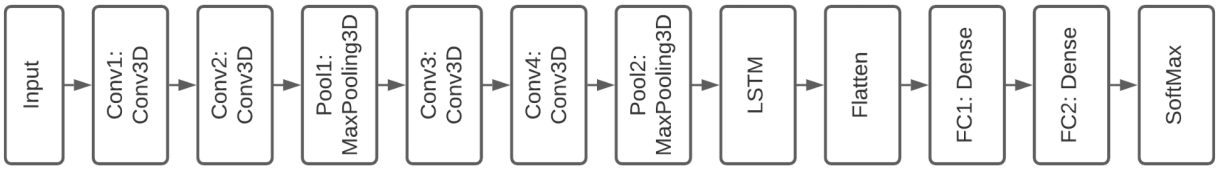


Figure 1. The Architecture of the Proposed Network Model

3.2.1 *Training the Model.* The dataset is composed of 6,270 videos and was randomly split using random shuffle to ensure all data are equally divided (GeeksforGeeks 2020) into training with 70% of data with 4,389 samples per class, testing with 20% of data with 1,254 samples per class with additional 99 samples of unseen data and validation with 10% of data with 627 samples per class (Chan 2021). The 3dCNN was used to extract features from each data input and the LSTM network was utilized for the sequential features or correct order of frames.

3.2.2 *Validating the Model.* The experiment was conducted fifteen (15) times to produce varying results in accordance to the study of Tran et al (2020). The proponent's model was compared to the result of Tran's model after using the same dataset - Overall, five (5) metrics were used to evaluate the proposed model. The first is *training accuracy*, which provides the correct predictions per input and the partial performance of the model. The second is the *confusion matrix* from the same library to oversee the classification score for each gesture and differentiate the actual and predicted values. Then, *Precision* and *Recall,* where the former gives the percentage of total predicted instances per class while the latter provides the percentage probability of correct types per sample. Subsequently, the *F-measure* is the last metric used by the proponents to ensure that the precision and recall metrics will always provide the true predictions during classification.

3.2.3 *Testing the Model.* Following the process of training and validation, the ensemble model from the proposed model was selected and tested in real-time using a webcam. This step ensures that the produced model is functioning as intended and the assessment was done by checking the accuracy levels or percentage of each gesture class, the testing was done by the subject standing in front of the webcam and doing the mentioned gestures. Furthermore, the proponents have included 11 similar hand gestures based on the original chosen gestures were added to ensure that the model is able to differentiate the correct predictions from the false ones - the addition of these dataset resulted to a better recognition as the model has been able to predict the true correct gestures.

4. Data Collection

The researchers have created and provided their own dataset containing one-handed ASL gestures using only the right hand for simplicity, and these words are *Abstain, Accident, Airplane, Badger, Commute, Daily, Dart, Glad, Glance, Police, Reason.* The inputs were taken using a phone camera tweaked to record at steady 24fps with a 1920 x 1080 resolution. The participants stood in front of the camera during the process of data gathering - They were recorded from the waist up or half body at their homes while having a simple with little to no obstructions or objects behind them. In addition, the4 participants are non-signers while their age group will fall under the ages of 17 to 50 to be standard as bones stop growing at the mentioned age and variations happen as the latter age is reached (Royal Osteoporosis Society, 2016). Gender was not a factor to consider, as long as the participant is within the range of the age parameters they are considered. Another factor that was considered was the lighting of the participants in the video, as long as their hands were visible. Moreover, each of the eleven (11) gestures was performed ten (10) times by the fifty-seven (57) participants resulting in six thousand two hundred-seventy (6270) videos in total. Videos that were shot lasted up to five (5) seconds.

Figure 2. Sample Procedure for Data Gathering

5. Results and Discussion

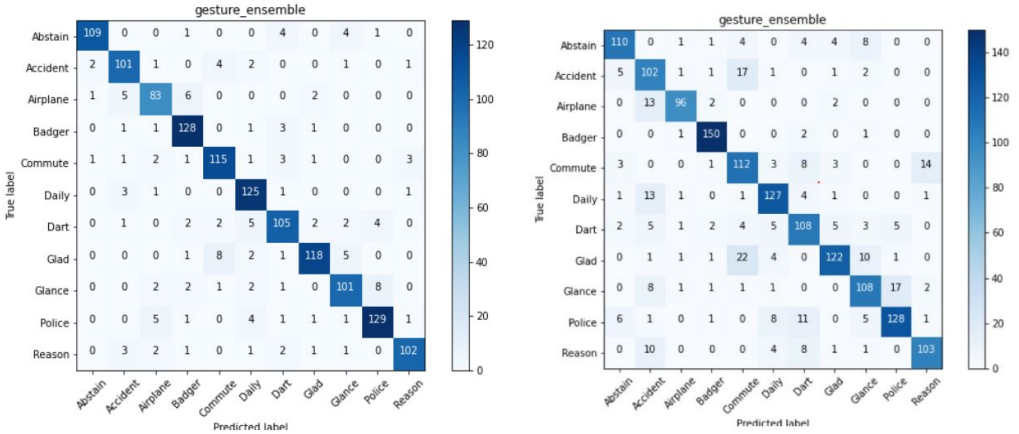
5.1 Analysis of Results

The purpose of Table 1 is to present the side-by-side results of the models after running the dataset on two (2) different networks (Tran's model and the Proponents model). Also, the result of the two (2) ensemble results are noticeable since the proponent's model performed better in classifying the different gesture classes in real-time.

Table 1. Comparison of evaluation score for Proponent's Model and Tran's Model

TRAINING ACCURACY COMPARISON		
Model Number	**Proponent's**	**Tran's**
1	88.90%	80.60%
2	85.20%	81.20%
3	86.40%	79.00%
4	87.20%	81.10%
5	87.30%	80.00%
6	87.70%	79.30%
7	87.50%	80.30%
8	87.90%	80.20%
9	87.10%	79.60%
10	87.10%	80.00%
11	88.10%	77.30%
12	85.90%	80.20%
13	86.30%	80.50%
14	88.10%	80.27%
15	86.30%	78.00%
Ensemble	89.87%	81.62%

Figure 3, represents the confusion matrix derived from the ensemble model from Tran's and the Proponent's model. As shown on the figure, the gestures that are not difficult to perform had the best performance in terms of accuracy, particularly the *Abstain, Badger* and *Reason.* In addition, the lowest accuracy came from gestures that are more sophisticated to perform, such as *Airplane* and *Dart* - these gestures are performed on the z-axis, which makes it difficult for the model to recognize immediately.



PROPONENT'S MODEL TRAN'S MODEL

Figure 3. Comparison using Confusion Matrix

The researchers used Recall, Precision, and Test accuracy metrics to interpret the data from the confusion matrix. As seen in Table 2, Precision is the ratio of correct classification towards predicted classification. On the other hand, Recall is the ratio of predicted classification towards the correct classification. Test accuracy is the correct prediction overall (Jayaswal 2021). Furthermore, both models achieved a high percentage value in the two metrics; this means that the base and proposed model have higher chances of correctly predicting and classifying the specific ASL gesture. On both models, the highest obtained value for precision is the gesture *Airplane* for the model of Tran and *Abstain* for the model of the proponents; this means, when the program predicts the gesture *Airplane*. It is 93.2% correct on Tran's model while the program predicts the gesture *Abstain* 96.5% of the time. This is also evident on Recall; when the true positive is *Badger* in Tran's model, it will predict the gesture 97.4% of the time. On the other hand, if *Daily* was the true positive of the proponents' model it would predict the gesture 95.4%.

Table 2. Interpreted Results from Pronent's and Tran's Model Network

	Tran's Model		Proponent's Model	
Gesture	**Recall**	**Precision**	**Recall**	**Precision**
Abstain	83.30%	86.60%	91.60%	96.50%
Accident	78.50%	66.70%	90.20%	87.80%
Airplane	85.00%	93.20%	85.60%	85.60%
Badger	97.40%	93.80%	94.80%	89.50%
Commute	77.80%	69.60%	89.80%	88.50%
Daily	85.20%	83.00%	95.40%	87.40%
Dart	77.10%	74.50%	85.40%	86.80%
Glad	87.40%	87.80%	75.30%	93.70%
Glance	78.80%	78.30%	86.30%	87.80%
Police	79.50%	84.80%	90.10%	90.80%
Reason	81.10%	86.60%	90.30%	94.40%

The F-score is harmony between the *Recall* and *Precision*, and is a measure of a model's accuracy on a dataset. Shown on Table 3, The f-measure is derived from the confusion matrix of each gesture from Tran's and the proponent's model. As evident, the highest achieved f1-score that Tran's model produced was in the gesture *Badger* which produced a 95.5% F1-score. While in the proponents' model the gesture *Reason* produces a 92.3 F1-score. All gestures from the proponent model apart from *Airplane* and *Badger* have a higher F1-score value than their counterpart on Tran's model.

Table 3. Comparison of Gesture Accuracy from Proponent's and Tran's Model

Gesture	Tran's Model F1-score	Proponent's Model F1-score
Abstain	84.90%	94.00%
Accident	72.10%	89.00%
Airplane	88.90%	85.60%
Badger	95.50%	92.10%
Commute	73.40%	89.10%
Daily	84.10%	91.20%
Dart	75.80%	86.10%
Glad	81.10%	90.40%
Glance	78.50%	87.10%
Police	82.10%	90.50%
Reason	83.70%	92.30%

In Table 4, it was clearly shown that the proponent's model had better performance compared to Tran's model, which was 8.3% lower than the 3dCNN+LSTM training accuracy and recall results. Moreover, the 1st iteration of the proposed model had the best performance above all by achieving a 88.9% accuracy, while the least performer was the 2nd iteration with 85.2% accuracy. Comparing this to the base model, the best performance was found on the 2nd iteration with 81.2%, and the least accuracy was found on the 11th iteration with 77.3%. The main advantage of the proposed model is the LSTM network that provides a description for the order of images or videos - it also supports the prediction of sequences that has accuracy when recognizing the hand gestures in a sequential manner (Brownlee 2019). In the case of precision, there was a 7.7% difference between the two while in recall and F1-score there was a 8% difference.

Table 4. Results from Ensemble from Proponent's and Tran's Model

Models	Training Accuracy	Recall	Precision	F1- score
Tran's Model	81.6%	81.7%	82.2%	81.8%
Proponent's Model	89.9%	89.7%	89.9%	89.8%

By utilizing TensorBoard, the proponents were able to view the proposed models' respective performances, showcasing both metrics of accuracy and losses. Indicated in Figure 4, the accuracy for training and validation are directly proportional towards a higher epoch. This means that as the epoch reaches a larger number - the value of accuracy for both increases as well and the mentioned model is a good fit for the problem. The graph above represents a single iteration from the proponent's model - 1st model that achieved an overall accuracy of 88.9%.

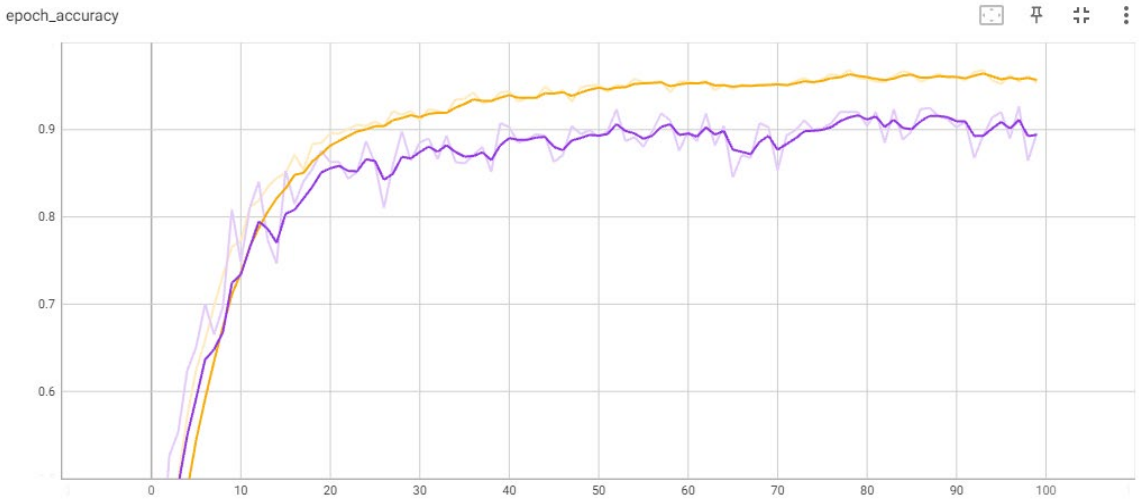


Figure 4. Graphical Result of Proponent's Model during Training

5.2 Model Comparison in Real-Time

The proponents tested the capabilities of the real time program, testing its ability to detect and classify the eleven (11) gestures. The proponents have tested each gesture ten (10) times to determine if the real time program can correctly detect and predict the gestures that were fed to it. Table 5 shows the *Precision*, *Recall*, and *F1-score* of the results. It is evident that in terms of real time detection and classification, the proposed model that utilizes LSTM had a higher percentage in all metrics. Only with the gestures *Airplane* and *Badger* does the model of Tran generate a higher percentage compared to its counterpart.

Table 5. Model Comparison of Real-Time Results between Proponent's and Tran's Model

Gesture	Tran's Model			Proponent's Model		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Abstain	76.20%	80.00%	76.20%	90.90%	100.00%	95.20%
Accident	72.70%	88.90%	72.70%	80.00%	80.00%	80.00%
Airplane	84.20%	80.00%	84.20%	77.80%	70.00%	73.70%
Badger	95.20%	100.00%	95.20%	75.00%	90.00%	81.80%
Commute	66.70%	60.00%	66.70%	80.00%	80.00%	80.00%
Daily	82.40%	70.00%	82.40%	88.90%	80.00%	84.20%
Dart	76.20%	80.00%	76.20%	88.90%	80.00%	84.20%
Glad	82.50%	87.80%	82.50%	75.00%	90.00%	81.80%
Glance	76.20%	80.00%	76.20%	77.80%	70.00%	73.70%
Police	82.40%	70.00%	82.40%	81.80%	90.00%	85.70%
Reason	82.40%	70.00%	82.40%	100.00%	80.00%	88.90%

6. Conclusion

6.1 Interpretations

The custom dataset provided by the proponents are composed of 6,270 videos that are classified and divided into the eleven (11) gestures have been tested on two (2) different networks - the first network is based from the study of Tran et al. (2020) which utilizes an all 3dCNN model while the second was derived from the same network with the addition of a single LSTM network before the output layer. Subsequently, the dataset had been used on the base model and achieved results close to the obtained outputs from the reference study - On Chapter IV the researchers it is evident that a single iteration from the base model reached an accuracy of 94.8% which is the highest among the others and considered as a good model score. In contrast, an iteration of the proposed model with the best performance had a slight increase in accuracy by achieving a 96.2% - this is due to the fact that an LSTM network supports the sequence prediction which is an essential part when recognizing hand gestures as it follows a systematic manner.

The proponent's model performed better in terms of classifying gestures in real-time - 89.8% was the F1 score for the proponents model while 81.8% was achieved using Tran's model. This method is a great tool to differentiate the performance of the models. Another reason for the difference in real time performance is that the proponents' model utilizes an LSTM layer that supports sequence prediction and better interpretation while Tran's model was not able to recognize the correct order of gestures immediately. This is due to the same movements used on the gestures that lead to misclassification when using the latter network model.

6.2 Future Works

For future works regarding this study, data is an integral part which means data used on future works should have more variety towards them - particularly towards the distance of the participant from the camera. As explained in the previous chapter, the real time detection has a harder time detecting gestures when the person isn't half body in the frame. Another is to go beyond the said eleven (11) gestures that were used in this study. For the program to detect and classify more gestures of ASL, not just gestures that use one hand but also gestures that utilize two hands. There are gestures that use two hands ra ther than one in ASL. This can be the basis program for a mobile application that helps non-signers understand hearing and speaking impaired individuals.

References

Al-Hammadi, M., Muhammad, G., & Abdul, W. Hand Gesture Recognition for Sign Language Using 3DCNN. Available: https://www.researchgate.net/publication/340972266_Hand_Gesture_Recognition_for_Sign_Language_Using_3DCNN, April 2020

Brownlee, J. CNN Long Short-Term Memory Networks. Machine Learning Mastery. Available: https://machinelearningmastery.com/cnn-long-short-term-memory-networks, August 14, 2019

Brownlee, J. Why Use Ensemble Learning? Machine Learning Mastery. Available: https://machinelearningmastery.com/why-use-ensemble-learning/ April 26, 2021

Chan, C. H. M. Step by Step Implementation: 3D convolutional neural network in Keras. Available: https://towardsdatascience.com/step-by-step-implementation-3d-convolutional-neural-network-in-keras-12efbdd7b130 July 26, 2021

Dong, W. What is OpenCV's INTER_AREA Actually Doing?. Available: https://medium.com/@wenrudong/what-is-opencvs-inter-area-actually-doing-282a626a09b3 June 26, 2018

GeeksforGeeks. numpy.random.shuffle() in python. Available: https://www.geeksforgeeks.org/numpy-random-shuffle-in-python/. August 19, 2020

GeeksforGeeks. Python OpenCV cv2.cvtColor() method. Available: https://www.geeksforgeeks.org/python-opencv-cv2-cvtcolor-method/. October 18, 2019

Halder et al. Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. Available: https://www.ijrpr.com/uploads/V2ISSUE5/IJRPR462.pdf. 2021

Jayaswal, V. Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score. Available: https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262 December 15, 2021

Media, A. I. Sign Language Alphabets From Around The World - Ai-Media, Available: https://www.ai-media.tv/ai-media-blog/sign-language-alphabets-from-around-the-world/ October 27, 2020

Nguyen, N.-H., Phan, T.-D., & Yang, H.-J. 3D Skeletal Joints-Based Hand Gesture Spotting and Classification. Available: https://www.researchgate.net/publication/351753745_3D_Skeletal_Joints Based_Hand_Gesture_Spotting_and_Classification. May 21, 2021

Oudah, M., Al-Naji, A., & Chahl, J.Hand Gesture Recognition Based on Computer Vision: A Review of Techniques, Available: https://www.mdpi.com/2313-433X/6/8/73. July 23, 2020

Pothanaicker, K. Human Action Recognition using CNN and LSTM-RNN with Attention Model. Available: https://www.researchgate.net/publication/334508614_Human_Action_Recognition_using_CNN_and_LSTM-RNN_with_Attention_Model June 2019

Quer, J, Steinbach, M, Handling Sign Language Data: The Impact of Modality. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00483/full February 19, 2019

Ramalingam, A. How to Pick the Optimal Image Size for Training a Convolutional Neural Network?. Available: https://medium.com/analytics-vidhya/how-to-pick-the-optimal-image-size-for-training-convolution-neural-network-65702b880f05. January 6, 2022

Royal Osteoporosis Society. Causes: Age and bone strength. Available: https://theros.org.uk/information-and-support/osteoporosis/causes/age-and-bone-strength/ July 21, 2016

Tran, D., Ho, N., & Yang, H. Real-Time Hand Gesture Spotting and Recognition Using RGB-D Camera and 3D Convolutional Neural Network. Available: https://www.researchgate.net/publication/338701325_Real-Time_Hand_Gesture_Spotting_and_Recognition_Using_RGB-D_Camera_and_3D_Convolutional_Neural_Network. January 20, 2020

Varga, D., Szirányi, T. No-reference video quality assessment via pretrained CNN and LSTM networks. Signal, Image and Video Processing. Available: https://www.researchgate.net/publication/333625481_No-reference_video_quality_assessment_via_pretrained_CNN_and_LSTM_networks. November 2019

World Health Organization, Deafness and hearing loss, Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss, April 1, 2021

World Health Organization, WHO: 1 in 4 people projected to have hearing problems by 2050, Available: https://www.who.int/news/item/02-03-2021-who-1-in-4-people-projected-to-have-hearing-problems-by-2050. March 1, 2021

Biographies

Kurt Christian G. Fernandez had his Bachelor of Science degree in Computer Science from Mapúa University specializing in Artificial Intelligence. During his college years, he had gained knowledge about different programming languages along with the practical use of each of them. During July 2021, he was accepted as an Intern Data Engineer at a company located at Taguig City, Philippines and his daily tasks were to manage and create different kinds of data frameworks.

Quintin Brian A. Paredes received his Bachelor of Science degree in Computer Science from Mapua University with a specialization in Application Development. During July of 2021, he was accepted as an Intern Quality Assurance in a company located in Taguig, Philippines. In January of 2022, he was accepted as an application developer.

Janino I. Perfecto took his Bachelor of Science degree in Computer Science from Mapúa University with a specialization in Artificial Intelligence. In his college years, he had gained knowledge about different programming languages along with the practical use of each of them. During July 2021, he was accepted as an Intern Business Intelligence Developer at an IT company located at Taguig City, Philippines - where most of his daily tasks was to process data and create recommendations for business actions.

Joel C. De Goma is an academic instructor from Mapúa University. Currently, he has three (3) academic degrees that he took from the same institution: *Master of Science in Computer Science, Master of Engineering in Electronics Engineering* and *Bachelor of Science in Electronics Engineering*. He also has publications that can be found at the

Scopus Index and one of his recent papers is called *"Detecting Red-Light Runners and Speeding Violations through Video Captures."*