ESM3NY

Listen To Me

SIGN LANGUAGE TRANSLATION SYSTEM

2023

REPORT BY:
AYMAN AL-ATTILI

ABSTRACT

Deaf individuals rely on sign language, which combines gestures, movements, postures, and facial expressions to convey alphabets and words from spoken language. Tagging in Arabic refers to placing a distinctive mark in the active activity.

The Arabic Sign Language Translation System helps to convert the gestures and gestures that indicate the alphabet into a text written in the Arabic language using the desktop application with a simple and easy-to-use user interface.

The yolov5 algorithm was used for training and testing and used in the final deployment system. The creation of the system went through several stages, where the first stage was researching the details of the problem, then the stage of designing an appropriate solution using artificial intelligence and computer vision techniques, the third was data collection and preparation, then the software work began and finally the testing stage.

Real images were collected over a period of several days, and approximately 2,000 new images were added to a roboflow stack of approximately 7,000 images, to improve the model's capabilities, accuracy, and efficiency.

Finally, future work will be a system that supports a larger number of sentences, expressions, and gestures and is able to convert Arabic sign language into Arabic text and audio.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 Research problem

There is a huge communication gap between normal and deaf people, and we want to make it small with the computer vision system, but it is a long way to go, so the best way is to study the topic from scratch and grasp its foundation.

So, we must learn in detail the signs of the Arabic alphabet to reduce the hurdles for sign language learners, but it is not an easy thing for all learners. Many of them will be confused when they are learning a new field and may find this problematic. For this reason, we intend to build a model that recognizes the alphabet sign from Arabic Sign Language speakers and then interprets it into text.

A meeting was held with a group of deaf students at the University of Jordan as part of the process of researching this problem and identifying the major problems and obstacles that they face within their daily routine activities. Also, problems and obstacles were discussed with one of the sign language interpreters, who is an employee at the University of Jordan at the Deanship of Student Affairs, and this is the result of meeting the students and the translator:

- Deaf students face a problem in communicating with other students, as they are the vast majority of students.

- Deaf students find it evidently difficult in daily dealings during university hours, such as ordering food from the restaurant, also when reviewing the menu sections on student affairs and dealing with staff.

- They face a daily problem in using transportation.

- Difficulty communicating in situations such as going to the hospital or security center without an accompanying interpreter.

1.2 Significance and potential contributions

After examining the problem and obstacles, and then looking at the available traditional solutions such as translators or non-traditional ones that use modern technology, we did not find any of these technological solutions that support the Arabic language. However, the accompanying translator is a high-cost solution that some injured or their families may not be able to benefit from, also limited The interpreter accompanies the injured person constantly which may this work it stressful for the interpreter and the injured person, also eliminates the privacy of the injured person.

From this standpoint, we started the idea of developing a system capable of converting Arabic sign language into a text written in Arabic, a system available 24/7 throughout the week that guarantees the user freedom of communication and high privacy in many scenarios.

The importance of this project lies in bridging a societal gap that many people suffer from, which reduces their quality of life and limits their creativity and contribution to building and developing society. We believe that through this computer vision system and artificial intelligence techniques, the gap will be smaller and smaller over time and these problems solved, also empower this group within the various activities of society.

Work and research have been carried out in this project as it is an extension of research that was carried out in the Faculty of Engineering at the University of Jordan, where this part will be discussed in detail throughout the report.

1.3 Scope and limitations

This system works to solve a very difficult problem, as this problem faces a challenge represented by the difficulty of the Arabic language which is the biggest hard aspect of the system, as the Arabic linguistic system consists of letters, simple expressions, complex expressions, and sentences, and the process of translating from sign language into Arabic is a difficult and accurate process It contains a lot of details, including facial expressions while speaking with gestures and signals. Also, there are many expressions and sentences that are not included in the data set that we worked on and that they are not available and need to be collected and prepared.

On this basis, work has been done on this system to recognize all the letters of the Arabic language and some simple and complex expressions, which allows the formation of several sentences within different scenarios and writing them on the screen using an easy-to-use user interface.

Chapter 2: Background and literature review

Background

According to the World Health Organization (WHO), there is a significant global prevalence of hearing loss, with approximately 466 million individuals affected, including 34 million children. This number is projected to surpass 900 million by the year 2050 [1]. Individuals with hearing loss may possess some degree of auditory capacity but often struggle to discern the presence of hearing aids. Conversely, individuals who are deaf experience complete or near-complete hearing impairment due to factors such as head trauma, exposure to excessive noise, disease, or genetic conditions [2].

Sign language serves as the primary mode of communication among the deaf community and between deaf and hearing individuals, with each country having its unique sign language. In the Arabic regions, Arabic Sign Language (ArSL) is utilized. ArSL was officially recognized in 2001 by the Arab Federation of the Deaf (AFOD) [3]. It relies on hand movements and gestures to convey meaning. Different countries may have variations in their dialects of ArSL, but they generally share a common set of 28 letters. The signs representing the Arabic alphabet can be observed in Figure 1-1.

Figure 2-1: Signs of Arabic alphabets [4]

literature review

In[5], Muhammad Al-Barham and Ahmad Jamal used the pretrained model ResNet-18 which is a pretrained network trained on more than a million images from the ImageNet database. The model used pictures of real hands for the training, validation and testing its efficiency, which made it easier for the model to simulate real experiences. The built model achieved an accuracy of 99.36%, which was the highest percentage compared to the models tested in the research.

Chapter 3: Solution Design

3.1 Methodology

Many research steps were taken during the process of building the system, as it was characterized by the characteristics of the experimental approach, where it was observed carefully the phenomenon of the difficulty of communication between the deaf and students within the university's environment, then a group of deaf students were met at the University of Jordan and their experiences were scrutinized and these experiences were tested in a way that supports Research. Where theories have been concluded that include:

- Define the problem
- Causes of the problem
- Available and suggested solutions
- Linguistic boundaries
- Technological infrastructure
- The target group of the proposed system in its first version

Data gathering

We faced challenges in finding data that represented the gestures and movements of Arabic sign language, but a data set consisting of approximately 7,000 images representing the gestures and movements of signs of the Arabic alphabet letters was used [6], as it has the diversity required to train the computer vision model. Also, the images were annotated which is suited to our detection model.

However, this set of data contained only signs of the Arabic alphabet only, but sign language is characterized by the presence of gestures and signs indicating complex sentences and expressions that must be considered during the process of building the translated system. So, using those expressions make the translation process becomes more realistic and efficient in daily use.

Accordingly, work has been done to identify a group of well-known signs and signals that indicate complex expressions, so that the process of constructing sentences is easier and closer to real experiences. The complex expressions are as follows:

- أنا (I am)
- أحبك (love you)
- عمري (my age)
- اسمي (my name)

The team began to collect the required data using several methods such as : 1- Google Form which was published on social networking sites to collect the required images from the communities on Facebook and Linkedin, 2- and self-collection of images through smartphone photography. Operation spanned several days to collect more than 500 images, including the four categories identified. During the data collection process, several factors were taken in account for the system results accuracy and highly efficient, namely:

- Balance the number of images for each category
- Diversity in the images in terms of place, time, skin tones, gender
- Resolution of captured images
- Reliability where there are no duplicated images

In addition, auxiliary tools such as Roboflow were used in the image annotation process. To ensure efficiency and accuracy, an augmentation of the images was done, as this process included doubling the number of images and enlarging the size of the data set through several operations such as: dictating images, changing colors or brightness, adding some distractions and other operations. After these operations, the size of the data set reached 1770 images, and each of the two data sets was merged using RoboFlow [7], bringing the number of images to 9264 images[8].

عمري (my age)　　أنا (I am)　　اسمي (my name)　　أحبك (love you)

Figure 3-1 Signs of Arabic experiences

3.2 Computer Vision Technique

This section provides an overview of the technology used to build the translation system. Where the YOLOv5 algorithm was used for training the model weights and torch.hub.load was used to access the model in Inference. The purpose is to provide a comprehensive understanding of the YOLOv5 algorithm and its practical application in computer vision tasks.

The YOLO (You Only Look Once) algorithm is a real-time object detection system that aims to localize and classify objects in images. YOLO treats object detection as a regression problem and divides the input image into a grid. Each grid cell predicts a fixed number of bounding boxes and corresponding class probabilities. The algorithm operates in the following steps:

1. Input Image Division:

The input image is divided into an S x S grid, where each grid cell predicts B bounding boxes and C class probabilities.

2. Bounding Box Prediction:

Each bounding box prediction consists of five attributes: (x, y) coordinates, width, height, and a confidence score representing the accuracy of the prediction.

3. Class Prediction:

For each bounding box, the algorithm assigns a class probability for the object contained within it.

4. Non-Maximum Suppression:

To eliminate redundant bounding box detections, non-maximum suppression is applied to retain only the most confident bounding box predictions. By decaying the scores according to Intersection Over Union (IoU) recursively, higher IoU detections will be eliminated with a minimum score threshold.

Training the model

The stage of training the model on a large set of large-sized data such as images needs a lot of computing power, so I used cloud services such as Google Colab which is Jupyter notebook that runs in the cloud and is highly integrated with Google Drive, making them easy to set up access and share.

Also, the most important service that this cloud allows us is using powerful GBU sources that speed up the training process and make it smoother and more efficient in terms of time.[9]

In the training process, we use YOLOv5 Small for several reasons. YOLOv5 Small is a variant of the YOLOv5 architecture that is specifically designed for efficient and lightweight object detection. It provides a good balance between accuracy, computational requirements, and detection speed, which makes it suitable for real-time sign language recognition system requirements.

On the other hand, the weights file format has been converted from PyTorch (pt) to Open Neural Network Exchange(onnx) for many aspects:

1- Flexibility:

.onnx: It offers a higher degree of flexibility and portability due to its wide framework compatibility. It enables deployment on a variety of devices, including edge devices, mobile devices, and specialized hardware accelerators.

.pt: It is more focused on the PyTorch ecosystem and is well-integrated with PyTorch's training and deployment workflows. It provides fine-grained control and customization options within the PyTorch framework.

2- Model Size:

.onnx: The serialized model in the ONNX format tends to be smaller in size compared to the .pt format. This can be beneficial and faster when transferring models over networks or storing them in memory-limited environments.

.pt: The PyTorch serialized model can be larger in size because it includes additional information such as the training state and optimizer parameters. However, PyTorch provides options to save only the model's state dictionary without these additional details, reducing the file size.

To load the YOLOv5 Small model for inference, the torch.hub.load function is utilized. The torch.hub module provides a simple interface for downloading and using pre-trained models from various repositories.

3.3 Python Frameworks

Python frameworks and libraries are pre-built sets of code that provide developers with tools and functionality to simplify and accelerate the development process. These resources enhance productivity, foster code reusability, and enable developers to focus on application-specific logic.

1- Numpy

A powerful library for numerical computing in Python, providing support for large, multi-dimensional arrays and a collection of mathematical functions to manipulate and analyze data efficiently.

2- PIL

(Python Imaging Library): A library for image processing tasks, enabling developers to open, manipulate, and save various image file formats. It offers a wide range of functions for image enhancement, filtering, and transformation.

3- Imutils

A library that simplifies common image processing tasks, such as resizing, rotating, and translating images. It provides convenient functions to streamline the development of computer vision applications.

4- Torch

An open-source machine learning framework that facilitates the creation and training of deep learning models. It offers a dynamic computational graph and a collection of pre-built modules, making it popular for tasks like computer vision and natural language processing.

5- OpenCV

(Open Source Computer Vision Library):A comprehensive library for computer vision and image processing tasks. It provides a vast collection of functions and algorithms to handle tasks such as object detection, image filtering, feature extraction, and more.

6- PyQt5

is a Python binding for the Qt framework, providing developers with the ability to create cross-platform graphical user interfaces (GUIs). It offers an extensive set of tools and widgets for building interactive applications, including buttons, menus, windows, and more. PyQt5 enables the development of visually appealing and feature-rich applications with ease, making it a popular choice for GUI development in Python.

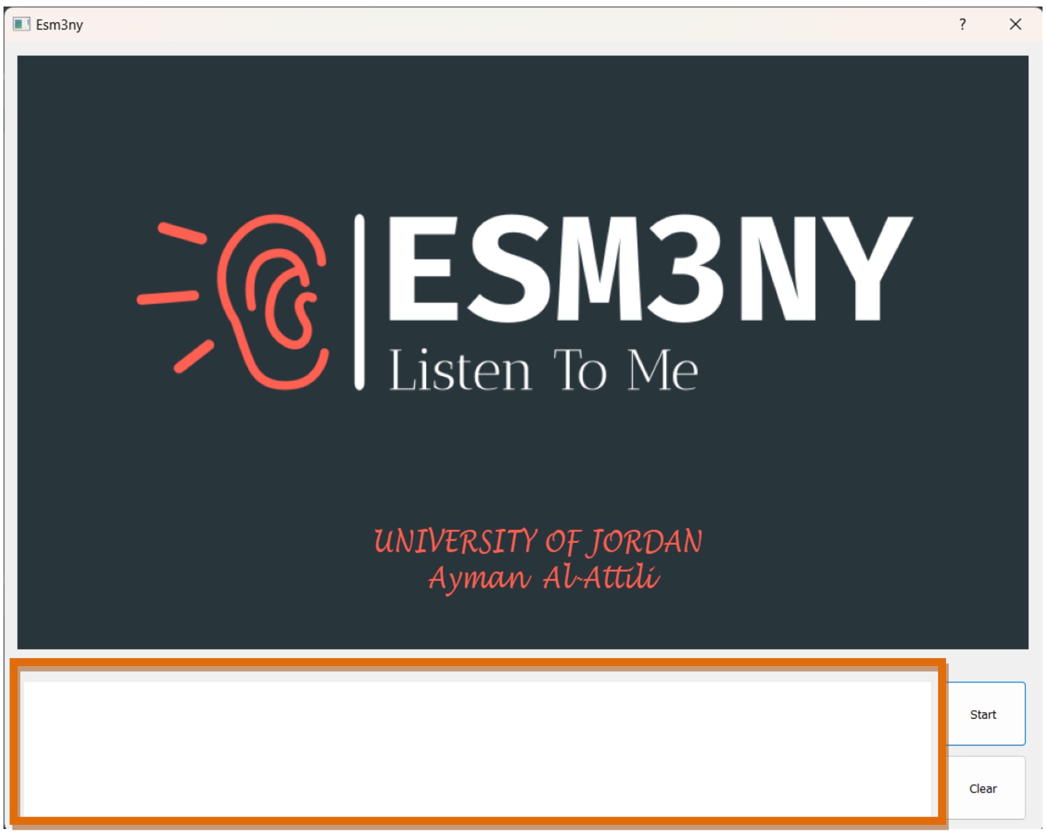

Figure 3-2 The User Interface of The Application

Displaying Section: Display the translated signs to the Arabic statements, words or letters.

Start Button : To run the real-time video capturing and start detecting signs then translate to Arabic written statements.

Clear Button : To erase all the words and statements on displaying section

Chapter 4: Data Analysis and Result

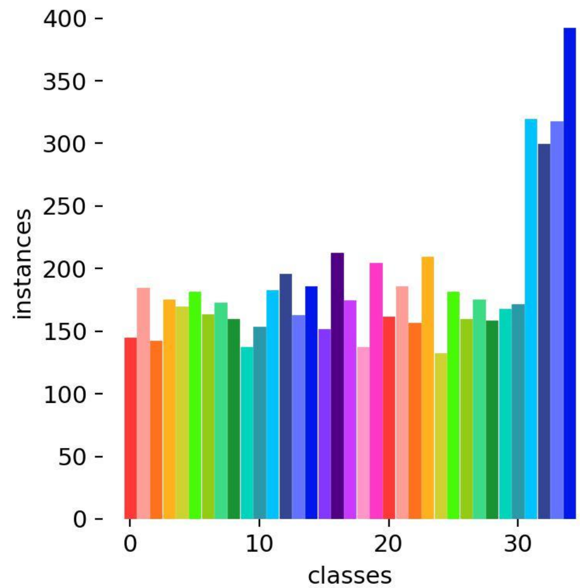

Figure 4-1: the number of images for each class in the dataset

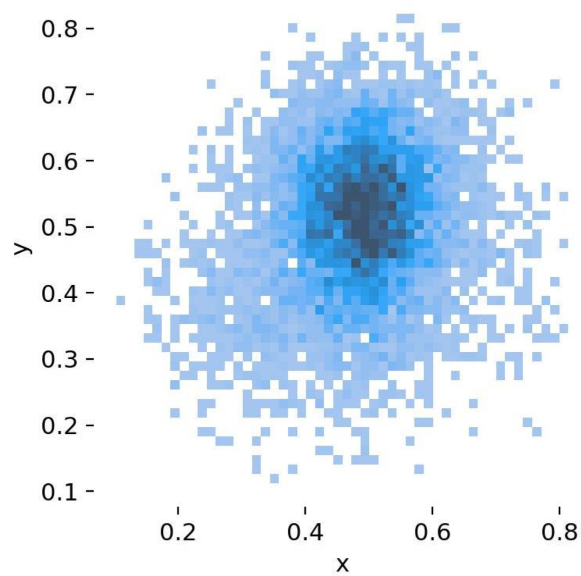

Figure 4-2: normalized values of points that make up the bounding boxes

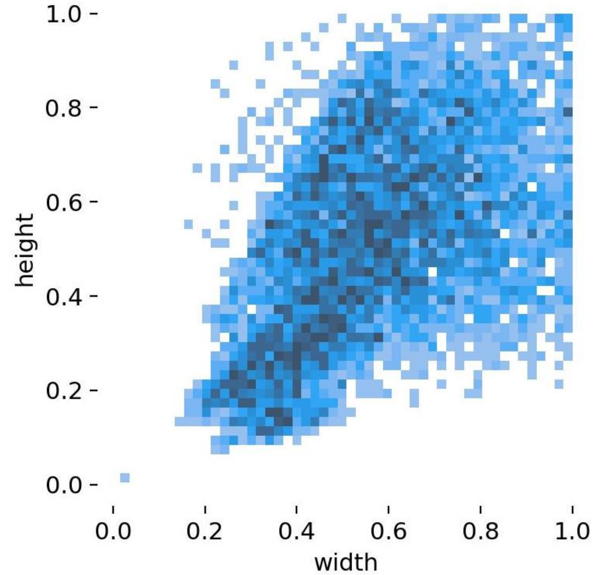

Figure 4-3: The normalized values of Height and Width that used to determined bounding boxes

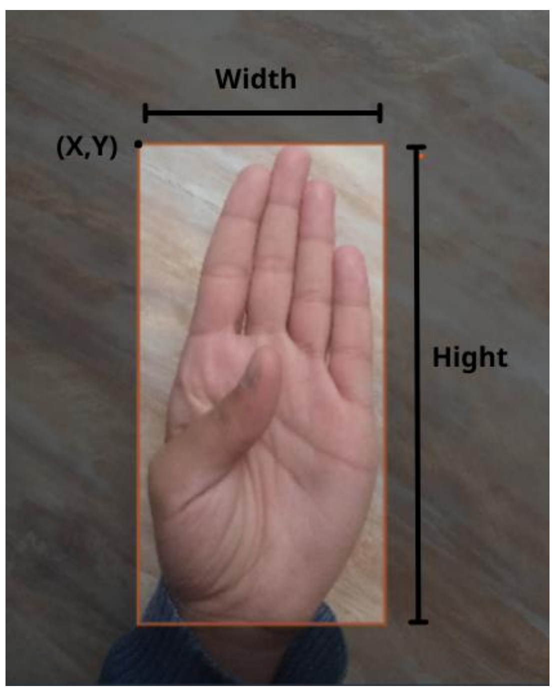

Figure 4-4: Explanation for Figure 4- 3,2

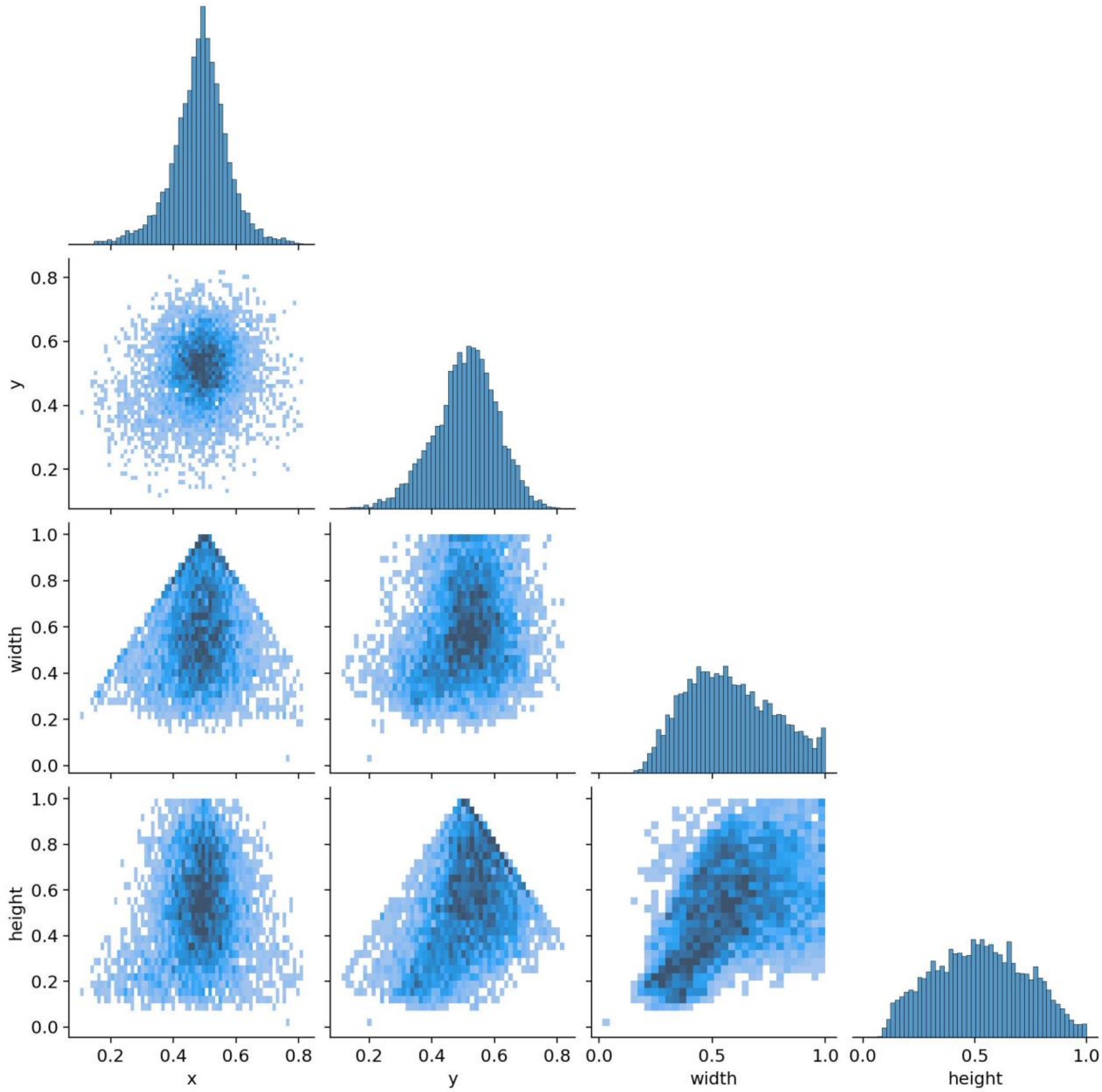

Figure 4-5: 2d histograms showing each axis of our data against each other axis

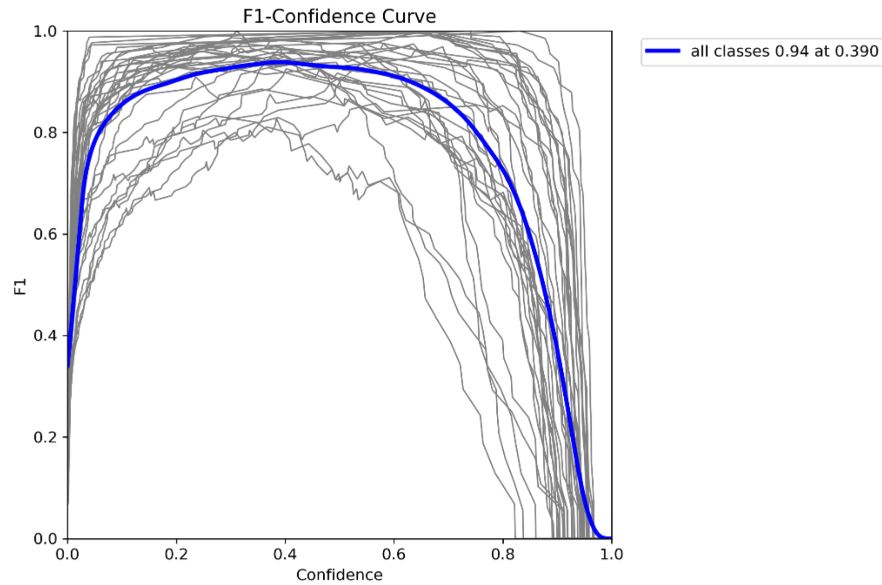

Figure 4-6: F1-Confidence Curve

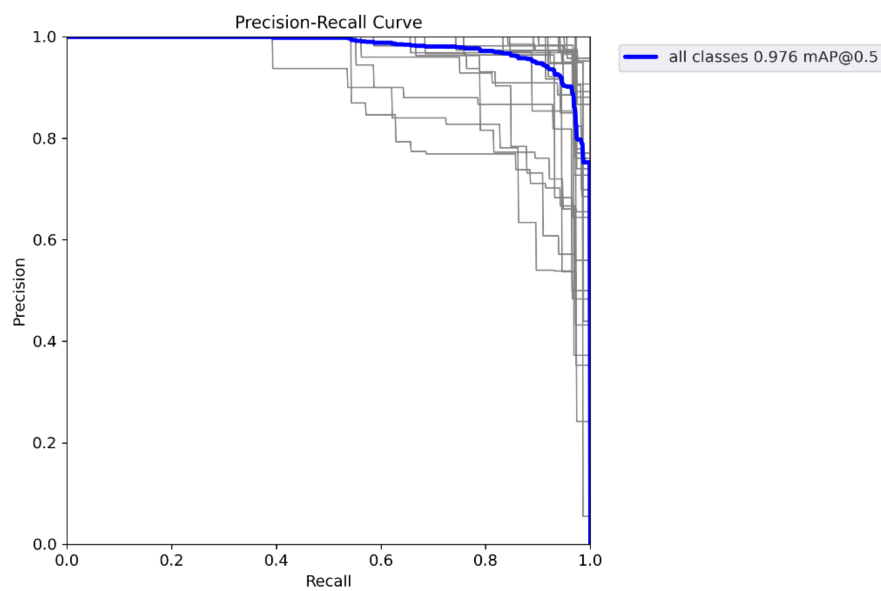

Figure 4-7: Precision-Recall Curve

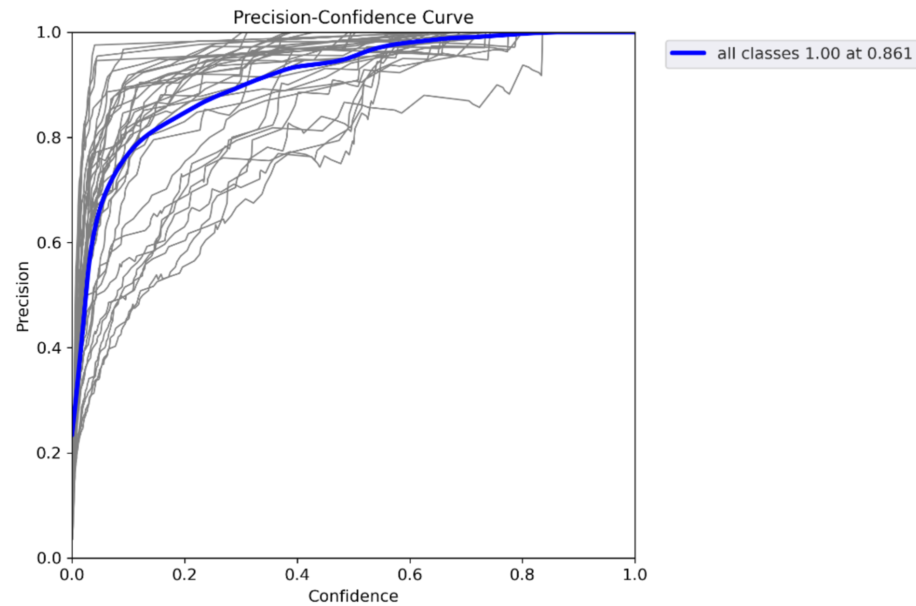

Figure 4-8: Precision Confidence Curve

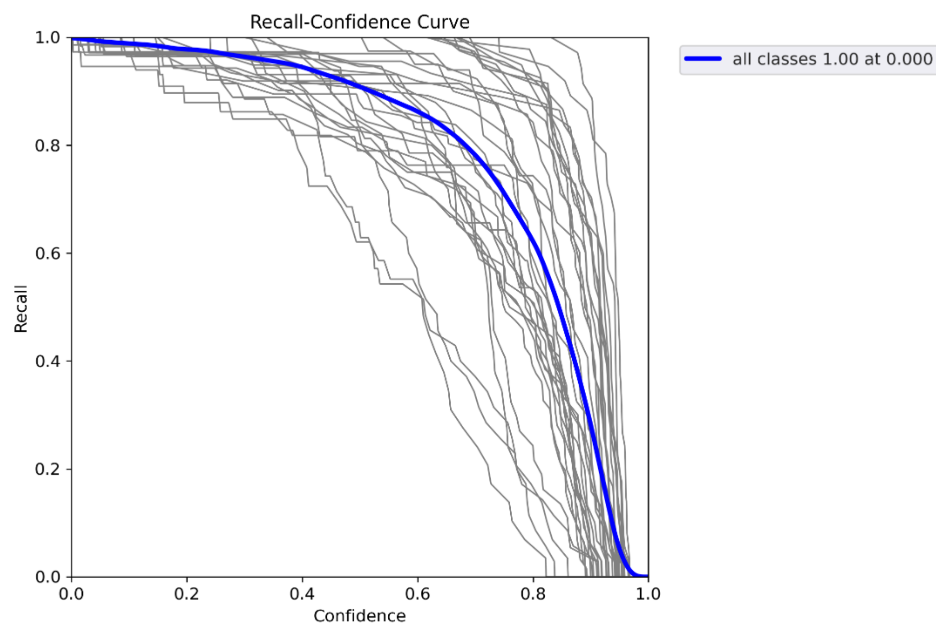

Figure 4-9: Recall Confidence Curve

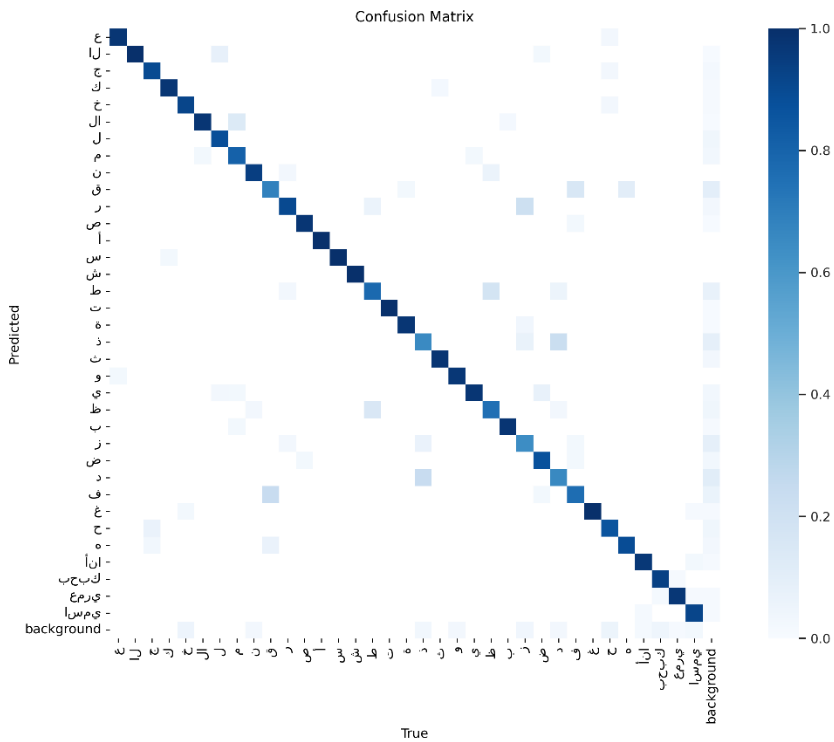

Figure 4-10: Confusion Matrix of Yolov5 model Shows the Matching Between the Predicted Values with True Values at The Testing Stage.

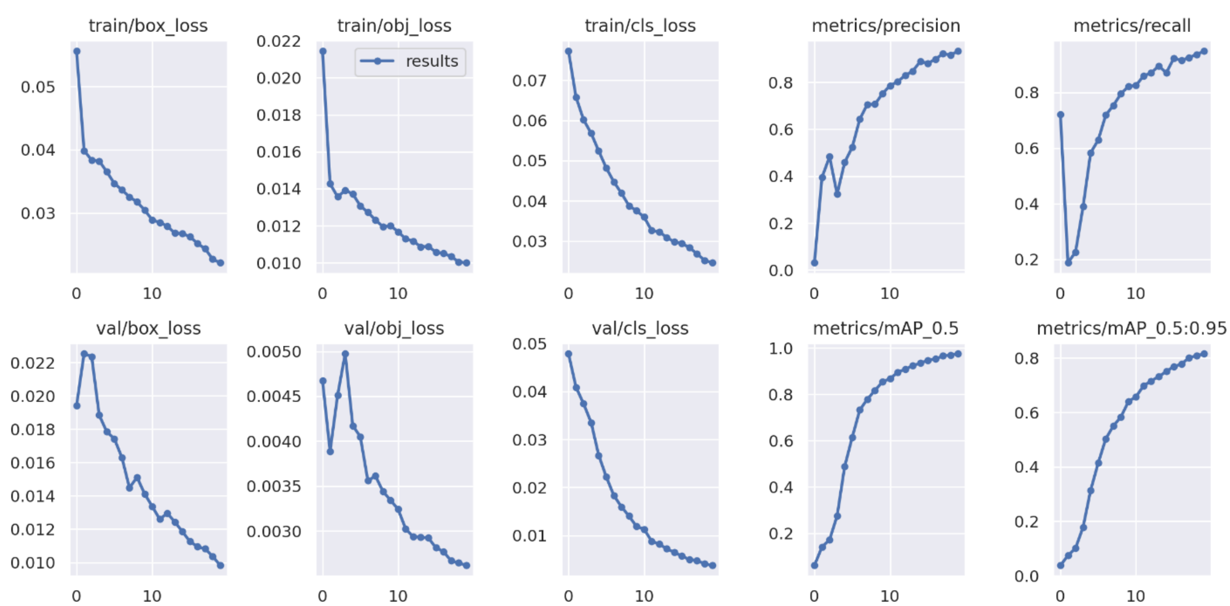

Figure 4-8 : Results of Object Detection Evaluation Metrics

Chapter 5: Conclusions and Future Work

5.1 Conclusions

This project was designed to be the second step and the extension of previous research " Design of Arabic Sign Language Recognition Model " to develop an approach that can help the deaf community using computer vision techniques to convert the ArSL alphabets into readable Arabic statements by real-time and high-performance system.

The Yolov5 model is used in training the model for the inference and deploy the model in desktop application.

5.2 Recommendation For Future Work

We will expand the project to include:

- Contribute to collecting and gathering new instances for the dataset to include dynamic alphabets, complex expressions and more words.
- Deploying the model on the mobile application.
- Integration with cloud computing technologies.
- Converting the translated written text from sign language into audible speech.
- Add translation from written or audio text to sign language with an avatar.

REFERENCES

[1] "Deafness and hearing loss." https://www.who.int/news-room/fact-sheets/detail/deafness-andhearing-losss (accessed Oct. 29, 2020).

[2] "The types, causes and treatment of hearing loss."https://www.starkey.com/blog/articles/2018/01/The-human-hearing-system-and-hearing-loss (accessed Oct. 29, 2020)

[3] M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers," J. Ambient Intell. Humaniz. Comput., no. 0123456789, 2020, doi: 10.1007/s12652-020-01790-w.

[4] sign language, Wikipedia " https://www.google.com/url?sa=i&url=https%3A%2F%2Far.wikipedia.org%2Fwiki%2F%25D9%2584%25D8%25BA%25D8%25A9_%25D8%25A5%25D8%25B4%25D8%25A7%25D8%25B1%25D8%25A9&psig=AOvVaw2VbsaGKTqswAygdhN6ls4E&ust=1687089808624000&source=images&cd=vfe&ved=0CBMQjhxqFwoTCJD9qLihyv8CFQAAAAAdAAAAABAE " , June 2, 2023

[5] M. Al-Barham, A. Jamal, 2021, "Design of Arabic Sign Language Recognition Model"

[6] "RGB Arabic Alphabet Sign Language (AASL) dataset Image Dataset", https://universe.roboflow.com/rgb-arsl/rgb-arabic-alphabet-sign-language-aasl-dataset/dataset/12

[7] "Roboflow", https://universe.roboflow.com/

[8] "ArSLW4EX_2023 Image Dataset", https://universe.roboflow.com/capston-project/arslw4ex_2023/dataset/1

[9] MasterNote: Colab notebook that has training and exporting operations, https://colab.research.google.com/drive/1u_CrbXAQP9NfCZpHMfT3b4qqV3rQtPa2?usp=sharing