A Project Report On

DYNAMIC GESTURE RECOGNITION

Submitted in partial fulfillment of the requirement for the 8^{th} semester

Bachelor of Engineering

in

Computer Science and Engineering

DAYANANDA SAGAR COLLEGE OF ENGINEERING

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560111

Submitted By

Mownika Reddy K A 1DS19CS728

Prajna Harish 1DS19CS732

Sambhrama K 1DS19CS743

Vinayak Bhat 1DS19CS757

Under the guidance of

Dr. Deepak G
(Internal Guide)
Associate Prof. CSE, DSCE

Mr. Yash Basrani
Co-guide

2022 - 2023

Department of Computer Science and Engineering

DAYANANDA SAGAR COLLEGE OF ENGINEERING

Bangalore - 560111

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Dayananda Sagar College of Engineering

(An Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE & ISO 9001:2008 Certified)

Accredited by National Assessment & Accreditation Council (NAAC) with 'A' grade

Shavige Malleshwara Hills, Kumaraswamy Layout, Bengaluru-560111

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the project entitled **Dynamic Gesture Recognition** is a bonafide work carried out by **Mownika Reddy K A [1DS19CS728]**, **Prajna Harish [1DS19CS732]**, **Sambhrama K [1DS19CS743]** and **Vinayak Bhat [1DS19CS757]** in partial fulfillment of 8th semester, Bachelor of Engineering in Computer Science and Engineering under Visvesvaraya Technological University, Belgaum during the year 2022-23.

Dr. Deepak G	**Dr. Ramesh Babu D R**	**Dr. B G Prasad**
Assoc. Prof.	Vice Principal & HOD	Principal
CSE, DSCE	CSE, DSCE	CSE, DSCE
Signature:.....	Signature:.....	Signature:.....

Name of the Examiners: Signature with date:

1..... .....

2..... .....

ACKNOWLEDGEMENT

We are pleased to have successfully completed the project **Dynamic Gesture Recognition**. We thoroughly enjoyed the process of working on this project and gained a lot of knowledge doing so.

We would like to take this opportunity to express our gratitude to **Dr. B G Prasad**, Principal of DSCE, for permitting us to utilize all the necessary facilities of the institution.

We also thank our respected Vice Principal, HOD of Computer Science & Engineering, DSCE, Bangalore, **Dr. Ramesh Babu D R**, for his support and encouragement throughout the process.

We are immensely grateful to our respected and learned guide, **Dr. Deepak G**, Associate Professor CSE, DSCE and our co-guide **Mr. Yash Basrani** for their valuable help and guidance. We are indebted to them for their invaluable guidance throughout the process and their useful inputs at all stages of the process.

We also thank all the faculty and support staff of Department of Computer Science, DSCE. Without their support over the years, this work would not have been possible.

Lastly, we would like to express our deep appreciation towards our classmates and our family for providing us with constant moral support and encouragement. They have stood by us in the most difficult of times.

Mownika Reddy K A 1DS19CS728

Prajna Harish 1DS19CS732

Sambhrama K 1DS19CS743

Vinayak Bhat 1DS19CS57

Abstract

Body language, which includes hand gestures, arm movements, and facial emotions, is an essential component of nonverbal communication. These modes of expression are essential to human connection and are utilized to communicate information. Hand gesture recognition systems have made an appearance in the field of human-computer interaction (HCI) as a creative, easy, and natural method of communicating with computers.

Systems for hand gesture recognition have two main uses: sign language recognition (SLR) and gesture-based control. Gesture-based control eliminates the requirement for conventional input devices like keyboards or mice by allowing users to interact with computers or other devices using hand motions. SLR, on the other hand, uses computer algorithms to automatically translate sign languages in order to narrow the communication gap between the hearing and the deaf communities.

The realization that sign language is a highly organized and largely symbolic collection of human movements led to the development of universal gesture-based HCI. Computers can recognise and decipher these motions by utilizing image processing techniques (IPT) and sophisticated algorithms, facilitating communication between those who use sign language and those who do not.

The essential terminology used to describe this area of study are sign language recognition (SLR), hand gesture recognition (HGR), and image processing techniques (IPT). Systems like SLR and HGR have the capacity to revolutionize accessibility and communication, empowering people with disabilities and boosting human-computer connection in a variety of fields.

In conclusion, systems for recognising hand gestures and sign language have grown to be crucial areas of study in the field of HCI. These devices can improve communication between sign language users and the general public by utilizing image processing techniques and sophisticated algorithms. Computers can close the communication gap and enable more inclusive and intuitive interactions between humans and machines by deciphering and interpreting body language.

Table of Contents

List of Figures

List of Tables

Chapter I

1 Introduction

1.1 The Problem

Sign language is a rich and expressive form of communication used by many deaf individuals as their primary means of interacting with the world. It encompasses a combination of hand gestures, body movements, eye contact, and facial expressions, all of which work together to convey meaning and convey complex messages. However, this reliance on sign language can create challenges for deaf individuals when they interact with non-deaf individuals who may not understand or be fluent in sign language.To overcome this barrier and promote inclusive communication, there is a need for a medium that can recognize and interpret sign language motions into spoken or written words that can be easily understood by non-deaf individuals. This is where a hand gesture recognition system for Sign Language Recognition (SLR) becomes crucial.The recognition of hand gestures is particularly essential in SLR because it is the primary means by which sign language users convey information. Hand shapes, movements, and placements carry specific meanings, and accurately capturing and understanding these gestures is crucial for effective communication.Overall, the development of a hand gesture recognition system for Sign Language Recognition is instrumental in breaking down communication barriers faced by deaf individuals. By enabling the interpretation of sign language motions into spoken or written words, it promotes inclusivity, understanding, and effective communication between deaf and hearing individuals, fostering a more inclusive and accessible society.

For a development of accurate and efficient systems, a number of issues related to gesture recognition must be solved. Some major problems with gesture recognition include

1.1.1 Variability and Complexity

It can be challenging to accurately record and categorise human gestures due to the wide range of hand shapes, movements, speeds, and styles. The complexity of gestures makes it even more challenging to create reliable recognition systems.

1.1.2 Occlusion and Partial Visibility

It can be tricky to effectively identify and track gestures when they are partially or completely covered. Obstruction can happen when hands, objects, or other body parts are present, which reduces the accuracy of gesture identification.

1.1.3 Noise and interference

Noise and Interference can be present in the gesture data recorded by sensors or cameras, such as sensor errors, modifications to lighting, or backdrop clutter. These background noise sources may jeopardise the system's dependability and interfere with precise gesture recognition.

1.1.4 Ambiguity in Gesture

Some motions might display similar visual appearances or minor differences, making them difficult to comprehend. For accurate gesture recognition, it can be difficult to distinguish between similar movements or to resolve ambiguity.

1.1.5 Real-time processing

Real time Processing is frequently essential to gesture recognition systems in order to provide seamless user-machine interaction. A significant challenge in gesture recognition is achieving minimal latency and high-speed processing while preserving accuracy.

1.1.6 Generalisation and Adaptability

Gesture detection systems should be able to adapt to diverse users, taking into account differences in hand sizes, shapes, and movement patterns. Furthermore, the system must adapt well to novel gestures without the need for expensive retraining or calibration.

1.1.7 Limited Training Data

Substantial and diverse datasets are often needed to train accurate gesture recognition models. However, gathering and annotating such datasets for gesture identification can be labor- and time-intensive, which limits the amount of training data that is accessible.

1.1.8 User-Dependent Variability

Due to individual behaviours, cultural influences, or physical limitations, people may perform gestures differently. To ensure reliable recognition, the system has to take account for individual variances and adapt based on various users' gesture patterns.

1.2 Real World Application

The project's goal was to reduce the reliance on software by using an existing Gesture Recognition system to solve a real-world issue. Human-robot interaction, augmented and virtual reality experiences, accessibility and assistive technology, sign language interpretation, gaming, and entertainment are prominent real-world uses.

These applications show how beneficial and influential gesture recognition technology can be in a variety of sectors.

1.2.1 User Interfaces and Wearable Technology

By facilitating interaction through hand gestures, gesture recognition technology revolutionizes user interfaces. It improves the accessibility of smartwatches, smartphones, and tablets by offering a more immersive and intuitive way to use digital interfaces.

1.2.2 Security and surveillance

Gesture recognition is essential to both of these applications. It additionally makes it possible to analyze human gestures in order to spot odd conduct and spot potential threats. Security measures in public places, airports, and high-security situations are enhanced by gesture-based authentication and access control.

1.2.3 Healthcare and rehabilitation

These are two fields in which gesture recognition is extensively used. During treatment or rehabilitation sessions, it enables the observation and evaluation of patients' gestures and motions. Progress may be tracked and rehabilitation courses' effectiveness is increased with the use of real-time feedback and advice based on gesture analysis.

1.2.4 Education and Training

Interactive learning environments are improved with gesture recognition technological advances. Students can perform experiments, manipulate virtual objects, and have immersive experiences by using hand gestures. With the help of this technology, students can learn more actively, become more engaged, and comprehend difficult concepts.

1.2.5 Car Interfaces

To make driving safer and more convenient, gesture recognition has been included into car interfaces. Hand gestures enable drivers to operate a variety of features, such as volume control, phone answering, and navigation system access, minimising distractions and enhancing driver focus while driving.

1.2.6 Retail and marketing

Gesture recognition enables companies to investigate consumer involvement and conduct. Retailers can gain insight about customer preferences, improve store layouts, and create individualised marketing campaigns by monitoring and analysing gestures. By enhancing the entire shopping experience, this technology promotes

customer engagement.

1.3 Organization of Project Report

The project report is organized as follows:

In Chapter (2) we discuss the problem statement and the proposed solution. We also take a look at the systems that exist today and the drawbacks they face.

Chapter (3) takes a more in-depth look at various hardware and software based solutions that exist, with a survey on existing literature available.

Chapter (4) looks at the architecture of the proposed solution with an overview of the system design, utilizing system block diagrams and data flow diagrams.

Chapter (5) dives into the Implementation of the solution, by describing the hardware and software requirements, along with dataset descriptions and im- plementation details.

Chapter (6) describes our testing process, while

Chapter (7) looks at our ex- perimentation process and the obtained results.

Chapter (8) summarizes our findings and concludes the paper.

Chapter II

2 Problem Statement and Proposed Solution

2.1 Problem Statement

The problem statement is how to create and put into use a real-time dynamic gesture recognition system that can accurately identify and interpret gestures in real-time using hand movements. Users should be able to fluidly engage through speech and text formats, and the system should be able to deliver results with high accuracy and efficiency. The major goal is to develop a solid and dependable system that can recognise hand gestures and convert them into actionable instructions or actions, improving user experience and enabling easy and intuitive human-computer interaction.

2.2 Existing Systems

The immediate hand configuration is frequently the focus of current hand recognition technologies. They might make use of methods like hand segmentation using skin colour models or coloured gloves, or they might demand that the user wear sensors or wearable technology. These methods have a number of drawbacks, though.

First off, prolonged use of wearable technology or sensors can make for a bad user experience. In some circumstances, wearing additional gear may be cumbersome, uncomfortable, or unpractical. Users would find it uncomfortable to wear such devices all the time, which would prevent these systems from being widely adopted.

Second, many of the current hand recognition systems don't have intuitive user interfaces. A segmented hand region may represent the extent of the user's involvement, which can be laborious and counterproductive. To improve the user experience overall and make the interaction more intuitive and natural, a better user interface design is required.

Furthermore, these systems frequently struggle with real-time gesture recognition. Real-time, high-speed hand gesture recognition is essential for a variety of applications, including virtual reality, gaming, and human-computer interaction. Existing systems can have trouble keeping up with how quickly hands move, which could cause latency or delayed replies.

Future hand recognition systems should concentrate on enhancing user experience by reducing the requirement for extra devices, offering user-friendly interfaces, and enabling real-time gesture recognition in order to get around these restrictions. The

versatility and usability of these systems in many applications would also be improved by the integration of both text and speech output capabilities.

2.2.1 Deep Learning and Computer Vision based systems

Dr. Odubanjo conducted a study to tackle the challenges associated with accurately detecting and identifying hand gestures in individuals with darker skin tones. Her innovative approach combined skin color analysis with other characteristics such as hand shape and motion. By integrating these factors, she aimed to enhance the overall performance of gesture recognition systems specifically for people with darker skin, providing a solution to the difficulties faced in this domain.

In their research, Mohamed Elmezain and Ebtisam S. Elsadig also addressed the issue of hand gesture recognition in individuals with darker skin tones. They proposed an innovative technique that incorporated a diverse set of features, including skin color, hand shape, and contour analysis. To ensure the effectiveness of their approach, they conducted extensive experiments using a comprehensive dataset that represented a wide range of skin colors. This enabled them to obtain robust and inclusive results, thus overcoming the challenges associated with accurately recognizing hand gestures in people with darker skin.

Although Dr. Malik's expertise extends beyond skin color-based hand gesture recognition and small datasets, he has made significant contributions in the field of computer vision. His research primarily focuses on leveraging machine learning algorithms, such as deep learning and convolutional neural networks (CNNs), to analyze visual data effectively. These methodologies are applicable to various scenarios, including situations involving skin color-based gesture recognition. Dr. Malik's work aims to achieve precise and reliable gesture recognition through advanced machine learning techniques, benefiting a wide range of applications.

2.3 Proposed Solution

One proposed solution for achieving fast, reliable, and robust object detection using deep learning involves combining the strengths of YOLOv5 and RCNN Multibox architectures. To begin, it is essential to create an extensive and diverse dataset by collecting and annotating images that specifically focus on the target object classes. This dataset will serve as the foundation for training the deep learning model.

YOLOv5, renowned for its real-time performance and accuracy, can be employed as the primary detector in this approach. It utilizes a single-stage architecture that

divides the input image into a grid and predicts bounding boxes and class probabilities directly. This allows for quick initial predictions, enabling fast detection of objects in real-time scenarios.

However, to further improve the accuracy of object detection, the RCNN Multibox approach can be utilized as a secondary detector in the system. RCNN Multibox employs a two-stage architecture, where the first stage proposes regions of interest (RoIs), and the second stage classifies and refines these proposals. By incorporating RCNN Multibox into the solution, the results from YOLOv5 can be refined and false positives can be reduced, enhancing the overall reliability of the object detection system.

To facilitate the implementation and customization of the model, both YOLOv5 and RCNN Multibox architectures offer streamlined processes for training and inference. Users can easily fine-tune the model on their specific dataset, adjust hyperparameters to optimize performance, and monitor the system's accuracy during the training process.

By combining the strengths of YOLOv5 and RCNN Multibox architectures, the proposed solution aims to achieve a fast, reliable, and robust object detection system. It leverages the real-time performance and accuracy of YOLOv5 for quick initial predictions, while incorporating the refinement capabilities of RCNN Multibox to improve accuracy and reduce false positives. Such a system would be capable of accurately identifying objects in real-time scenarios, making it suitable for a wide range of applications, including autonomous driving, surveillance systems, and object recognition tasks.

2.3.1 Dataset Preparation

A smart technique to train a hand gesture recognition model is to recreate a dataset of hand gesture photos or video frames with bounding box annotations and accompanying labels. You have successfully produced 20 photos with 100 samples for each sign by manually clipping each sign from the hand using bounding box coordinates and applying the proper class labels.

Building a strong and reliable hand gesture recognition system requires a wide dataset that includes a range of ethnicities, colors, lighting circumstances, clothing preferences, and age groupings. Your model can learn from and generalize across a wide range of data thanks to which it is better equipped to recognise hand movements in real-world circumstances.

A smart technique to train a hand gesture recognition model is to recreate a dataset of hand gesture photos or video frames with bounding box annotations and accompanying labels. You have successfully produced 20 photos with 100 samples for each sign by manually clipping each sign from the hand using bounding box coordinates and applying the proper class labels.
Using a different set of validation or test photos, you can assess the model's performance once it has been trained on the dataset. This assessment procedure enables you to determine how well the model generalizes to unseen data and provides insights into its accuracy and robustness.

2.3.2 Train YOLO

You Only Look Once (YOLO) must be trained to recognise and detect particular hand motions, which necessitates customising the YOLO algorithm for the task at hand. An summary of the procedures is provided below:

Compile a sizable labelled collection of hand motions in pictures or video frames along with the appropriate bounding box comments. Each hand gesture in the photos should have its location and class label specified in the annotations.

Convolutional neural networks (CNNs) are commonly used by YOLO as the foundation of its network design. You might need to change the network architecture for YOLO to be suitable for hand gesture recognition. Convolutional layer size and number may need to be changed, spatial attention methods may need to be added, or new layers may need to be added to capture fine-grained hand gesture information. To manage object detection at various scales and aspect ratios, YOLO uses anchor boxes. It's possible that the anchor boxes need to be changed to account for the fact that hand motions can differ in size and shape. You can identify the ideal set of anchor boxes to cover the variety of sizes and aspect ratios included in the data by examining the distribution of hand gesture sizes in your dataset.

To train the model, YOLO combines localization loss and classification loss. While the classification loss assesses the accuracy of the predicted hand gesture classes, the localization loss gauges the precision of the predicted bounding boxes. The loss function may need to be modified to prioritise the recognition of hand gestures that are important to you.

2.3.3 Real-time object detection

Real-time gesture detection using a trained YOLO (You Only Look Once) model allows for efficient and accurate detection and localization of objects within video frames or image sequences. In the case of hand gesture recognition, the YOLO model can be applied to each frame of a video stream to detect and locate hand gestures in real time. To perform real-time sign detection using YOLO, the following steps are typically involved:

Preparing the trained YOLO model: Before performing sign detection, you need to have a YOLO model trained on a hand gesture dataset. The training process involves feeding the model with labeled images of hand gestures, optimizing its parameters through backpropagation, and adjusting the network's weights to minimize the detection errors.

Configuring the YOLO model: The YOLO model consists of a deep convolutional neural network (CNN) architecture. It is important to configure the model with the appropriate network architecture, layer sizes, and hyperparameters to achieve accurate and efficient detection. The YOLO model is usually designed to output bounding boxes, class probabilities, and confidence scores for each detected object.

Preprocessing the video frames: Before passing each frame to the YOLO model for detection, you may need to preprocess the frames to prepare them for input. This preprocessing step can involve resizing the frames to the input size expected by the YOLO model, normalizing pixel values, and converting the frames to the appropriate color space if needed. Running the YOLO model on video frames: Once the frames

are preprocessed, they can be passed through the YOLO model. The model performs forward propagation to generate predictions for each frame, including bounding box coordinates, class labels, and confidence scores for detected hand gestures.

Post-processing the detections: After obtaining the model's predictions, post-processing is typically done to refine the detected hand gestures. This may involve filtering out low-confidence detections, applying non-maximum suppression to eliminate overlapping bounding boxes, and adjusting the coordinates to match the original frame's size.

Visualizing the results: Finally, the detected hand gestures can be visualized by

drawing bounding boxes or overlaying masks on the original video frames. This step helps visualize the real-time object detection output and can be used for further analysis or integration with other applications.

2.3.4 Gesture Classification

Hand gesture recognition systems must include gesture classification. The next stage is to categorise hand gestures into various gesture categories when they have been identified. A gesture classification algorithm is commonly used for this classification.

The hand gesture regions that are identified are analysed by the gesture classification algorithm in order to categorise or label them with one of many predetermined gesture labels. A Hello, Thank you, I Love You, Please, or any other motion that the system is built to recognise are represented by these labels.

A gesture classification algorithm can be implemented using a variety of techniques; the best strategy to use will depend on the application's unique needs and limitations. Here are a few typical methods for categorisation of gestures:

Create a database of template pictures or other attributes to represent each predefined category of gestures using the template matching method. The system assigns the detected hand gesture to the label that most closely matches the features it gathered from the library of template hand gestures.

A model can be trained to categorise hand gestures using machine learning approaches, such as supervised learning algorithms. This method trains the model using a labelled dataset of hand gesture photos or characteristics. The model picks up on the patterns and characteristics that set various movements apart. The model can identify fresh hand gestures into the categories it has learned after being trained.

Convolutional neural networks (CNNs), in particular, have demonstrated notable effectiveness in gesture categorization tasks. Deep learning. With the help of automatic feature extraction from the input data, CNNs are able to develop hierarchical representations of hand motions. These networks can recognise and classify movements properly because they have been trained on enormous datasets of labelled hand gesture photos. HMMs (Hidden Markov Models) are statistical models that may be applied to sequential data, such a series of hand motion frames. An HMM predicts the temporal connections between succeeding frames and represents each gesture cat-

egory. The method compares the sequence of hand gesture frames that were observed with those from the various HMMs and assigns the gesture label that best matches. The choice of method to utilise is influenced by a number of variables, including the complexity of the gestures, the amount and quality of the dataset, the need for real-time processing, and the computer resources available. It is usual practise to test out various algorithms and choose the one that works best for the particular application. The overall hand gesture recognition system can incorporate the trained or created gesture categorization algorithm. The algorithm receives the hand gesture regions that have been recognised, analyses the data, and assigns a corresponding gesture label. Additional uses for this classification information include manipulating user interfaces, interacting with virtual objects, and giving input for gaming applications.

2.4 YOLO v5 Architecture

The You Only Look Once (YOLO) object detection method, which is renowned for its real-time performance, is the foundation of the YOLO-V5 architecture. In order to improve the network's feature representation and increase the precision of object identification, YOLO-V5 includes the PANet (Path Aggregation Network) module.

The YOLO architecture's core is in charge of removing features from the input image. It analyses the image and finds pertinent patterns and structures to serve as the building blocks for object detection. A convolutional neural network (CNN) serves as the backbone in most cases, processing the image at various sizes and capturing features at various degrees of abstraction. A feature pyramid network that improves the feature representation of the backbone is the PANet module in YOLO-V5. This is accomplished by adding top-down and bottom-up paths to collect both low-level and high-level features at various sizes. The backbone network is used to process the input image in the bottom-up pathway, creating feature maps at various granularities. The top-down pathway then combines these feature maps via upsampling and lateral connections to produce a more thorough understanding of things at various scales.

The PANet module combines the feature maps from various backbone network layers so that the YOLO-V5 network may represent the input image in a deeper and more detailed manner. This combination of qualities enhances the network's ability to handle objects of various sizes and improve the overall detection performance A collection of finely tuned feature maps with in-depth semantic data about the items

seen in the input image make up the PANet module's output. Following layers for bounding box regression and object classification receive these feature maps. These feature maps are used by the YOLO-V5 network to produce precise predictions of item positions and class labels.

YOLO-V5 considerably increases the network's capacity to detect objects of different sizes and improves its overall performance in terms of accuracy and speed by integrating the PANet module into the YOLO architecture. With the help of the PANet module, the network is able to record both low-level and high-level information, enabling a more thorough analysis of the objects in the image and more accurate object detection outcomes.

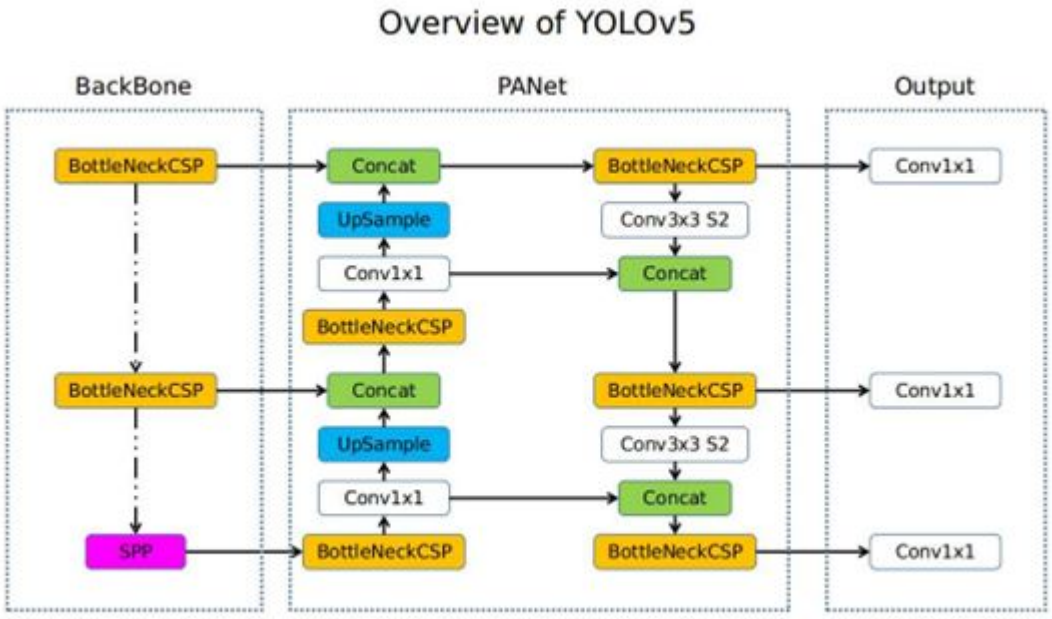


Figure 2.1: YOLO v5 Architecture

2.5 System Requirements

- Processor : 4 Core, 8 Thread Processor
- Memory : Minimum of 16 GB ram for computation
- Video Memory : Minimum of 3 GB of VRAM for Object Detection and another 2 GB for feature extraction and the ANN.

Chapter III

3 Literature Survey

Gesture recognition technology provides numerous advantages, such as intuitive interaction, hands-free operation, and immersive experiences. Nevertheless, it is important to acknowledge the drawbacks associated with this technology. These drawbacks encompass the absence of standardization, limited accuracy, susceptibility to environmental factors, and challenges in interpreting gestures. Furthermore, the hardware solutions employed for gesture recognition, which typically involve sensors, cameras, or a combination of both, can be expensive, complex, have restricted viewing capabilities, encounter compatibility issues, necessitate calibration and setup procedures, and are influenced by environmental conditions.

Noraini Mohamed et al.(2020) aimed to assess the progress made in the field of vision-based hand gesture recognition and identify areas that require further attention. They conducted a comprehensive review of 98 relevant articles sourced from reputable online databases using specific keywords. The review revealed that vision-based hand gesture recognition research is a thriving and active area, with a significant number of studies and publications in reputable journals and conference proceedings each year. The majority of these articles focus on three key aspects: data acquisition, data environment, and hand gesture representation. The researchers also evaluated the performance of these systems in terms of recognition accuracy.

In terms of signer-dependent recognition, the selected studies reported accuracy levels ranging from 69% to 98%, with an average of 88.8%. For signer-independent recognition, the reported accuracy varied from 48% to 97%, with an average of 78.2%. However, the lack of notable progress in continuous gesture recognition indicates that further work is needed to develop a practical vision-based gesture recognition system. These findings highlight the ongoing advancements in the field while emphasizing the need for continued research and development to overcome the challenges in achieving accurate and continuous gesture recognition.

Minhyuk Lee et al.(2019) present an innovative method for real-time dynamic finger gesture recognition using a data glove embedded with soft sensors. The primary objective of their research is to address the challenge of accurately identifying meaningful dynamic gestures from a continuous data stream, which is further complicated

by unconscious hand movements and variations among different users. To overcome this challenge, the authors propose a deep learning-based gesture spotting algorithm that effectively detects the start and end points of a gesture sequence. They introduce a scalar value called gesture progress sequence (GPS) to quantify the progress of the gesture. Furthermore, they propose a sequence simplification algorithm and a gesture recognition algorithm based on deep learning to handle variations in gestures. The developed real-time gesture recognition system, which incorporates these algorithms, was tested with 11 dynamic finger gestures and exhibited impressive performance. It successfully estimated the GPS in a mere 6 milliseconds and recognized completed gestures in no more than 12 milliseconds. This study highlights the advantages of data glove-based approaches over vision-based methods for hand gesture recognition, emphasizing their significance due to the limited availability and restrictions of commercial data gloves.

They introduce an advanced approach for real-time recognition of dynamic finger gestures using a data glove embedded with soft sensors. Their main focus is on overcoming the challenges associated with accurately distinguishing meaningful gestures from a continuous stream of hand movements, which is further complicated by unconscious actions and user variations. To address this issue, the researchers propose a cutting-edge gesture spotting algorithm based on deep learning techniques. This algorithm effectively identifies the beginning and end points of gesture sequences. Additionally, they introduce a metric called gesture progress sequence (GPS) to quantify the progress of gestures. To handle variations in gestures, the authors also present a sequence simplification algorithm and a deep learning-based gesture recognition algorithm. The system developed using these algorithms achieves remarkable real-time performance. In the experimental evaluation, it successfully estimated the GPS within a mere 6 milliseconds and recognized completed gestures in no more than 12 milliseconds. Importantly, this study emphasizes the advantages of data glove-based approaches compared to vision-based methods, highlighting the significance of such techniques due to the limited availability and constraints of commercial data gloves.

JUN XU et al. introduce a robust method based on RGB-D data for recognizing static and dynamic hand gestures. The main objective of their research is to facilitate natural interaction with virtual environments. To achieve this, the authors employ various algorithms such as Distance Transform (DT) and K-Curvature-Convex De-

fects Detection (K-CCD) to extract the contour of hand gestures, locate the center of the palm, and identify the fingertips for static gesture recognition. Additional features, including distances between contour pixels and the palm center and the angle between fingertips, are utilized to create a multimodal feature vector. A recognition algorithm is then proposed to accurately identify static hand gestures.

For dynamic gesture recognition, the authors combine the Euclidean distance between hand joints and the shoulder center joint with modulus ratios of skeleton features. This combination generates a unified feature descriptor for each dynamic gesture. To obtain recognition results for dynamic gestures, an improved version of the Dynamic Time Warping (DTW) algorithm called IDTW is introduced. Extensive experiments are conducted to validate the effectiveness and reliability of the proposed algorithms for static and dynamic hand gesture recognition. The paper also highlights the implementation of a real-time, low-cost application that enables natural interaction with virtual environments using hand gestures.

In conclusion, this paper presents a comprehensive method that utilizes RGB-D data for accurate recognition of both static and dynamic hand gestures. The proposed algorithms demonstrate robustness and effectiveness, indicating their potential for enabling natural interaction in virtual environments.

CHANGJIANG LIU et al., explore the possibility of utilizing radar sensors for gesture recognition in remote control of electronic devices. The authors investigate the radar measurement parameters' spectra as a means to identify human gestures. By combining radar theory and classification methods, they find that the frequencies of different gesture parameters can serve as distinctive features for gesture recognition. To simplify the process, they design six types of periodic dynamic gestures that eliminate the need to define and extract the start and end points of each gesture.

In addition to frequency ratios, the authors extract features related to motion range and detection coherence to address interferences caused by unintended gestures. They develop a decision tree classifier based on experimental observations, ensuring effective classification between different gestures. Overall, each gesture achieves a correct recognition rate exceeding 90%. The experiment employs a W-band millimeter wave radar to collect hand position and Doppler velocity information for classification, confirming the feasibility of the proposed radar-based gesture recognition method.

In conclusion, this paper showcases the potential of radar sensors for gesture recog-

nition in remote control applications. By analyzing radar measurement parameters' spectra and employing a decision tree classifier, the authors achieve accurate recognition of various gestures with a high success rate. The experiment conducted using a millimeter wave radar validates the effectiveness of the proposed method.

JOSIAH W. SMITH et al. concentrate on enhancing the accuracy of classifying non-moving (static) hand gestures by utilizing a convolutional neural network (CNN) in conjunction with frequency-modulated-continuous-wave (FMCW) millimeter-wave (mmWave) radars. Traditional approaches for non-contact hand pose and static gesture recognition rely on optical or depth cameras, which are limited by ideal lighting and temperature conditions. In contrast, mmWave radar devices offer a cost-effective solution, providing precise spatial information even in non-ideal imaging conditions.

While deep convolutional neural networks have shown success in image recognition tasks, their application to static gesture recognition using mmWave radars has posed challenges due to the difficulty in extracting meaningful features from radar return signals. Additionally, the performance of static gesture classification using mmWave radars and CNNs has been inferior compared to dynamic gesture classification. This paper introduces an effective data collection approach and a novel training technique for CNNs by incorporating "sterile" images. These sterile images assist in distinguishing distinct features among static gestures, leading to improved classification accuracy.

By employing the proposed data collection and training methods, the classification rate for static hand gestures is enhanced from 85% to 93% for range profiles and from 90% to 95% for range-angle profiles. These results demonstrate the effectiveness of the approach in improving the recognition of static hand gestures using mmWave radars and CNNs.

Q. Wang et al., proposed a deep model for traffic sign detection and identification. There are a few challenges in traffic sign identification jobs, such as fewer recognisable signs, tiny target sizes, easily leading to recognised failure, and so forth. First, we introduce Coordinate Attention (CA) for failure detection; second, to accelerate the regression of prediction box, the angle loss into objective function is introduced; third, for the overlapping and occlusion phenomenon of ground truth, a dynamic label assignment strategy- simple Optimal Transport Assignment (SimOTA) is used during label assignment; and finally, for the target size problem, a hierarchical feature fusion

network is used.Experiments were conducted on two publicly available data sets, and the results show that their improved model surpassed the basic model, YOLOv5s, and other renowned algorithms in terms of accuracy, recall, and mAP.

X. Yuan et al., proposed a similar approach for Traffic Sign Detection using YOLO. Traffic sign recognition technologies have been integrated into advanced driving assistance and autonomous driving systems to assist drivers in obtaining accurate route information. The current mainstream detection approaches perform well in this job, but the number of model parameters is huge, and detection performance is poor. Based on YOLOv5s as the foundation, this research offers YOLOv5S-A2, which may enhance detection speed and minimize model size at the expense of detection accuracy. To begin, a data augmentation technique based on combining several processes is presented to address the issue of imbalanced class instances. Second, to create additional horizontal connections, a path aggregation module for the Feature Pyramid Network (FPN) is presented. It can improve multi-scale feature representation capabilities and compensate for feature information loss. Third, an attention detection head module is presented to address the aliasing issue in cross-scale fusion and improve predictive feature representation.

Y. Li et al., performed a research where a YOLOv5-SLL sign language letters recognition model is suggested to address the challenges of low recognition accuracy and poor resilience of existing sign language letters recognition models in scenarios with complicated background interference and overlapping hands. First, a Sign Language Letters (SLL) dataset with 3373 annotated sign language letters pictures is created. Second, the Convolutional Block Attention Module (CBAM) is included into the foundation of the YOLOv5 network structure, allowing the network to focus more on the extraction of hand characteristics while reducing background noise interference. Finally, the redundant bounding boxes are optimized using the soft-Non Maximum Suppression (soft-NMS) method, which eliminates the problem of missed detection, which happens often when the hands overlap.

T. F. Dima et al., proposed a method for detecting the alphabet and numbers presented by each motion. Sign language is a communication tool that involves visual gestures and signals and is widely used by those who have trouble speaking or hearing. To incorporate such people into the society of verbal communicators, it is

vital to comprehend the communication gestures they utilise. People who do not use the gesture in everyday life may not understand what it represents. Deep learning methods for sign language recognition have already been developed. However, the utility of these models is limited. A YOLOv5-based technique is suggested since it is lightweight, fast, and accurate. A. Puchakayala et al., believes communication is

essential in everyday life, but imagine two people who are unable to speak with one another because one of them can not understand what the other is attempting to say. The majority of the deaf-mute community encounters this when interacting with non-deaf people. Normal people do not know or comprehend sign language since it is used by people with disabilities. This communication chasm must be closed. As a result, a model has been designed to let normal people and deaf-mute persons interact with one another. The sign language identification system is one such model, which uses a deep learning method to recognise American Sign Language (ASL) movements and output the associated alphabet in text format. A CNN model and a YOLOv5 model were developed and compared. The YOLO model had an accuracy of 84.96%, whereas the CNN model had an accuracy of 80.59%.

Chapter IV

4 Architecture and System Design

The overview of the system is represented in Fig.. It shows the modules involved in building the system i.e

- WEB CAMERA
- DATASET
- INBUILT SPEAKER
- SYSTEM INTERFACE

4.1 Software Overview

The first step in gesture recognition is image capturing, which includes snapping pictures with lenses or other imaging gear. The accuracy of gesture recognition is greatly influenced by the clarity and resolution of the obtained images.

An image is processed beforehand after it is taken to improve its quality and prepare it for processing subsequently. To maintain consistency in the input shots, techniques for pre-processing may include noise removal, contrast enhancement, and scaling.

The filtered image is next fed into the YOLOv5 algorithm, which stands for "You Only Look Once." The object detection algorithm YOLOv5 renders use of deep learning to locate and identify things within the image. By splitting the input image into a grid and instantly predicting bounding box coordinates and class probabilities, it employs a single-stage architecture to operate.

The region of interest (ROI) is extracted from the image using the bounding box coordinates provided by YOLOv5. The hand or motion that has to be acknowledged is contained in the ROI. To concentrate on the information needed for gesture detection, this region is segregated and processed individually.

The ROI is then subjected to feature extraction computations in order to obtain pertinent elements that distinguish various hand gestures. Extraction of color, texture, shape, or motion elements from the image is one of these techniques. Feature extraction aids in developing a more condensed and detailed representation of the gesture.

A neural network that has been trained on a database of hand gestures is then fed the extracted features. By comparing the retrieved features with predetermined classes of gestures, the neural network performs classification. The recognised gesture,

which can be associated with a particular action or meaning, is the result of the classification process. It is crucial to remember that the availability of a broad and representative training dataset is crucial to the effectiveness of gesture recognition. A vast collection of labeled photos containing a variety of movements and different lighting, backgrounds, and hand orientations are needed to train the neural network. To sum up, the gesture identification procedure entails taking a picture, enhancing its quality through preprocessing, utilizing object detection to find the hand or motion, extracting pertinent information, and then categorizing the gesture using a trained neural network. The ability to recognise and interpret hand movements through this method creates opportunities for a variety of applications, including sign language interpretation, human-computer interaction, and more

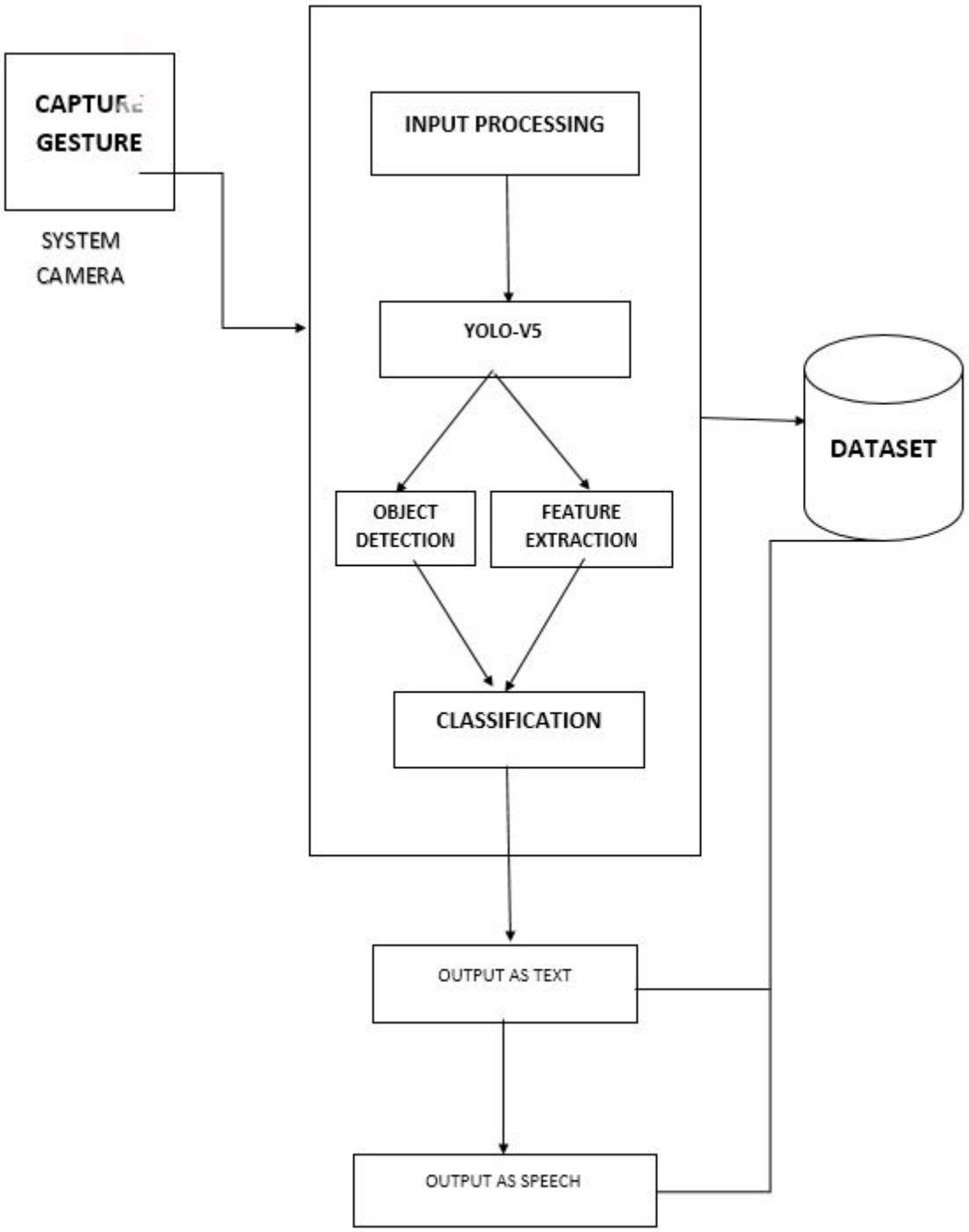


Figure 4.1: System Architecture

4.2 Data Flow Diagram

The follwing Figure 4.2 shows a brief Data Flow diagram of the project.

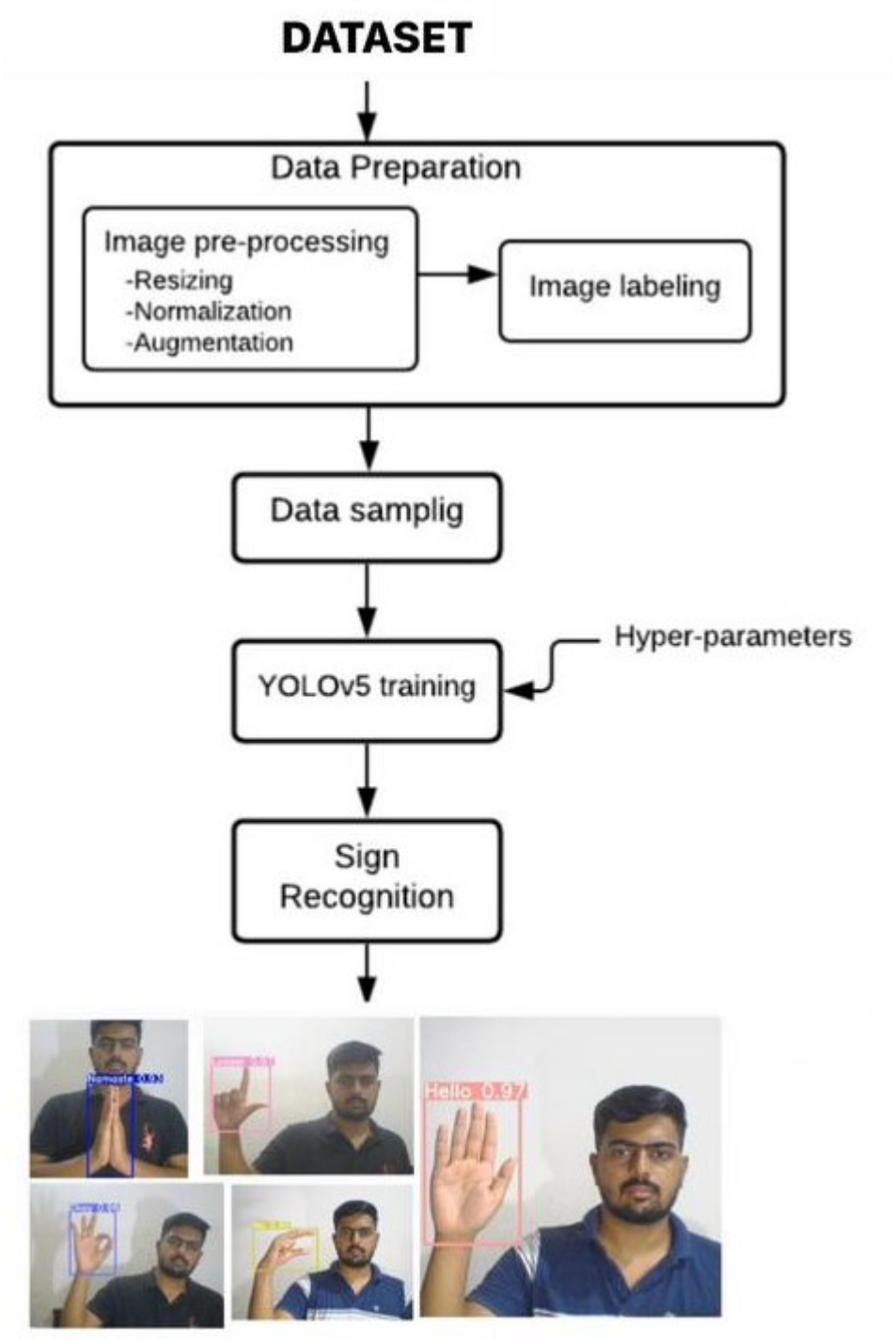


Figure 4.2: Data Flow Diagram

4.2.1 Extracting Key Point Values using PYTROCH

For computer vision tasks like posture estimation, the PYTROCH framework provides pre-built models. Developers can take important point values out of picture or video frames by using PYTROCH posture estimation model. The model can recognise and follow human poses, offering coordinates for significant body parts such the head, shoulders, elbows, wrists, hips, knees, and ankles. The programme precisely determines and tracks human poses within real-time by analysing the spatial correlations between these key points. The key point values that were recovered allow for use in motion analysis, augmented reality, gesture recognition, and other applications. To ensure accuracy and dependability, PYTROCH posture estimation algorithm is honed on extensive data sets. PYTROCH makes it easier to extract key point information and makes it easier to design new applications because of its personalization choices and ease of utilisation.

4.2.2 Collect Key Point Values for Training and Testing

There are various stages you must take in order to evaluate the effectiveness of a gesture recognition model using MediaPipe. Beginning by compiling a dataset of pictures or videos that show the gestures you wish to recognize. A varied range of variants, such as various backgrounds, lighting setups, and gesture performers, should all be represented in the collection. The key point values for each frame in the dataset should then be extracted using MediaPipe's posture estimation technique. These key point values correspond to the body keypoints' coordinates. According to the precise gesture being made in each frame, indicate the key point values. This stage entails manually annotating the dataset to identify the appropriate gesture for each group of important point values.Make training and test sets from the dataset after it has been labelled. The training set is used to develop the gesture recognition model, while the testing set is used to gauge how well the model performs on untrained data. The performance of the model may be reliably evaluated on new data thanks to this split. The collection, preprocessing, labelling, and splitting of a dataset to train and test a gesture recognition model using MediaPipe can be done successfully by adhering to these stages.

4.2.3 Pre-process Data and Create Labels and Features:

Pre-processing steps are required to get the retrieved key point values ready for the YOLO neural network's gesture detection training. These actions entail scaling the values to a standard range and normalising the coordinates. For supervised learning, labels that indicate whether a gesture is present or not are essential. To improve training, additional transformations and data augmentation methods might be used, such as flipping, rotating, or translating images. These pre-processing stages enable accurate gesture detection and categorization by preparing the critical point values for YOLO neural network training. As a result, the gesture recognition model becomes more reliable and powerful.

4.2.4 Build and Train YOLO Neural Network

The structure of a YOLO neural network model for gesture detection is determined based on the input size and gesture classes. For transfer learning, weights can be pre-trained or initialized. Classification and regression losses are combined in the loss function. It is selected to use optimisation methods like SGD or Adam, with parameters for learning rate, batch size, and epochs. Pre-processed key point values, labels, and features are used in training. Precision, recall, and mAP evaluation metrics are tracked, and hyperparameter modifications are made as needed. On the testing set, the model is assessed for generalizability. YOLO weights that have been trained are retained for use in real-time gesture detection applications. Architecture design, training, and evaluation are the main steps in creating a YOLO model, which produces a reliable and accurate gesture recognition system.

4.2.5 Make Predictions and Save Weights

Once trained, the YOLO model is capable of recognising gestures by having pre-processed key point values flow through the forward function within the model. Predicted bounding boxes and associated gesture labels are generated by the model, enabling the localization and depiction of detected gestures. In order to guarantee the preservation of learnt information, the training weights of the YOLO model should be preserved for later usage. The save weights function offers convenient file path preservation of the model's characteristics and weights.Reliability is facilitated by conserving the weights, which enables the model to be reloaded for interpretation or further training jobs without having to start from scratch. This permits the simple integration with the YOLO model within real-time gesture detection programmes or

manufacturing environments.

4.2.6 Detecting Gestures

Leveraging an optimised YOLO model, perform real-time gesture recognition by using MediaPipe to extract critical points from each frame or image. Normalise and scale the key point values that you have extracted as a preparatory step. To acquire predicted bounding boxes and gesture labels, run the YOLO model on the pre-processed key point values. Examine the bounding box coordinates and confidence scores to identify and categorise the gestures. Update the display to reflect the bounding boxes and gesture labels to see the outcomes. To recognise gestures in real-time, repeat this procedure for every successive frame or image. Applications like real-time gesture recognition systems and interactive user interfaces are made possible by this.

4.3 Use Case Diagram

The following Figure 4.3 shows use case diagram for the project.

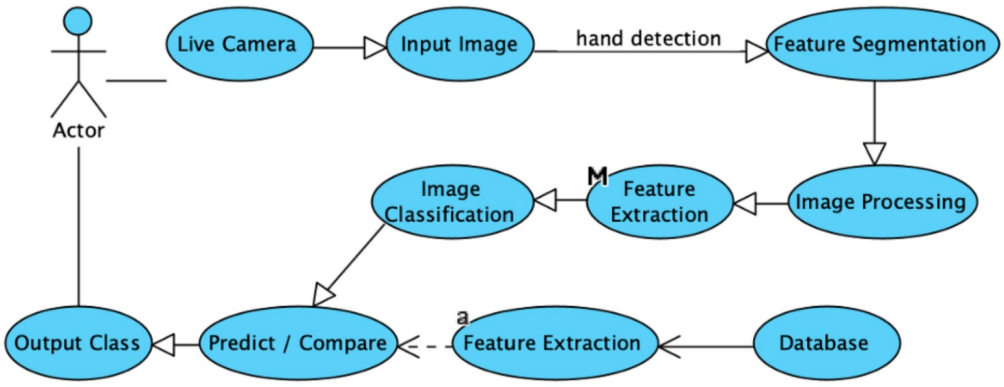


Figure 4.3: Use Case Diagram

4.3.1 Use Cases

1. Capture Gesture: The user initiates the gesture capture process by performing a hand gesture in front of the camera or sensor.

2. Preprocessing: The system pre-processes the captured data, which may include image filtering, noise reduction, or normalization.

3. Gesture Recognition: The pre-processed data is analyzed and classified using machine learning algorithms or computer vision techniques to recognize the specific hand gesture.

4. Gesture Feedback: The recognized gesture is used to trigger an action or provide feedback to the user, such as controlling a device, executing a command, or displaying a corresponding message.

5. Gesture Customization: Users have the option to customize or define their own gestures, associating them with specific commands or actions.

6. Gesture Training: In order to improve recognition accuracy, the system allows users to train the model with additional gesture samples, facilitating model refinement.
7. System Configuration: Users can configure the system settings, such as sensitivity, threshold, or gesture library management.

8. Gesture Logging: The system logs information about recognized gestures, which can be used for analysis, evaluation, or system improvement purposes.

4.3.2 Actors

1. User: Initiates the gesture capture and interacts with the system.
2. System: Recognizes gestures, provides feedback, and manages gesture-related functionalities.

4.3.3 Relationships

Capture Gesture, Preprocessing, Gesture Recognition, and Gesture Feedback are associated with the User and the System.
Gesture Customization and Gesture Training are associated with the User and the System for user-specific customization and model refinement.
System Configuration is associated with the User and the System for adjusting system settings. Gesture Logging is associated with the System for capturing and storing gesture-related information.

Chapter V

5 Implementation

5.1 Implementation Platform

5.1.1 Hardware

i)Processor: INTEL i7 11th gen processor

ii) RAM: 16GB DDR4 3533 MT/s

iii)GPU: INTEL IRIS Xe graphics

iv) Storage platform: 512GB NVMe SSD

5.1.2 Software

i)Operating System: Windows 64 bit (Docker container)

ii) Software Used : sklearn, OpenCV, matplotlib.pyplot,numpy

iii)Programming Languages : Python 3

iv) Server: Python HTTP server

5.2 Implementation Details

5.2.1 Implementation Workflow

For effective gesture identification, systems frequently use complex artifacts like Fast R-CNN, YOLOv5, or specially trained data models. The useful features are extracted from the input data by these artifacts using deep learning algorithms, and predictions are then based on those features.

By training them on gesture datasets, well-known object detection frameworks like Fast R-CNN and YOLOv5 can be modified to recognise gestures. These frameworks analyze input photos or video frames and employ convolutional neural networks (CNNs) to pinpoint the presence and placement of objects, including hand motions. To teach them the precise visual patterns and attributes connected to various motions, they can be trained on labeled gesture data.

Typically, these object detection systems require the following stages to be trained:

Data Preparation: The labeled dataset, which consists of picture or video clips and the associated gesture labels, is ready for training. To achieve reliable recognition, this dataset should include a wide range of gesture variations, lighting situations, backgrounds, and camera angles.

Model Setup: The chosen artifact, such as YOLOv5 or Fast R-CNN, is set up to recognise gestures. The network design, hyperparameters, and optimisation settings are all specified in this.

Training Method: The artifact is trained using the prepared dataset. The artifact gains the ability to identify pertinent elements in the input data and generate precise predictions based on the labeled gesture classes during training. Batches of data are iteratively fed into the model during the training process, and the loss (the (the discrepancy between predicted and true labels), and updating the model's parameters to minimize the loss.

Evaluation and fine-tuning: After training, a separate testing dataset is used to gauge the model's performance. The model's accuracy, precision, recall, and other metrics are discussed in this evaluation. The model can be fine-tuned by changing hyperparameters or adding more training data based on the evaluation findings to increase performance.

Alternatively, you can train your own unique models utilizing deep learning architectures customized to the particular needs of gesture recognition instead of using pre-existing object detection frameworks. This strategy gives you more freedom and power over the model design. Popular deep learning frameworks like TensorFlow or PyTorch may be used to create custom models that let you customize the network architecture, loss function, and training procedure to meet your unique requirements.

Data preparation, model configuration, training, and evaluation are phases in custom model training that are comparable to those in training with pre-existing frameworks. The bespoke network architecture's design and execution, which can be adapted to capture the particular traits and complexity of gesture recognition tasks, constitute the primary distinction.

Overall, the decision between an artifact and a bespoke model is influenced by the particular requirements, the resources that are available, and the required amount of customization for the gesture detection system. Both methods can be successful in establishing precise and effective gesture recognition, and each has advantages.

5.3 Data set

Creating a data set of hand gesture photos or video frames with bounding box annotations and labels is a recommended technique for training a hand gesture recognition model. The manual extraction of signs from the hand using bounding box coordinates, along with the application of appropriate class labels, facilitates the preparation of a data set suitable for model training. The effectiveness of a hand gesture recognition system relies heavily on the diversity of the data set. It is crucial to incorporate variations in ethnicity, colors, lighting conditions, clothing preferences, and age groups to ensure robustness and generalizability of the model. By encompassing a wide range of data points, the model can learn from different scenarios and improve its ability to recognize hand movements in real-world environments.

The dataset mentioned in this context consists of 20signs,100 samples each. Although this size is reasonably adequate for training purposes, it is important to emphasize that larger and more diverse data sets tend to yield better results. Increasing the data set size and diversity can enhance the model's capacity to handle variations and generalize effectively.

After training the model with the data set, rigorous evaluation is necessary using a separate validation or test set of photos. This evaluation process enables the assessment of the model's performance on unseen data and provides valuable insights into its accuracy and robustness. Evaluating the model on diverse images or video frames helps measure its proficiency in correctly classifying and recognizing hand gestures.

To ensure a comprehensive evaluation, the validation or test set should encompass a diverse range of factors, including ethnicity, colors, lighting conditions, clothing preferences, and age groups. This diversity enables an assessment of the model's performance across various real-world scenarios, thereby determining its effectiveness in different contexts. Adopting an iterative approach of training, evaluation, and refinement allows for progressive enhancements in the model's performance. This process facilitates the development of a reliable hand gesture recognition system capable of accurately identifying gestures across diverse situations.

5.4 Advantages of Hand Gesture Recognition System

The Hand Gesture Recognition System offers several advantages that contribute to its effectiveness and versatility. Let's elaborate on each of the mentioned advantages:

5.4.1 Custom data set

• Custom data set captures video frames dynamically without color or lighting restrictions: The ability to capture video frames dynamically without being constrained by color or lighting restrictions ensures that the system can accurately recognize hand gestures in various environments. This flexibility allows for reliable performance in different lighting conditions, making it suitable for real-world applications.

• Data set encompasses a wide range of scenarios for robustness and real-world handling: The data set used for training the Hand Gesture Recognition System covers a diverse range of scenarios. This inclusivity ensures that the system is robust and can handle various situations encountered in the real world. By training on a comprehensive data set, the system becomes more capable of accurately recognizing gestures across different contexts.

• Extensive data set enables rigorous training and generalization across individuals and hand variations: The availability of an extensive data set facilitates rigorous training of the system. By incorporating a wide range of hand variations and gestures from different individuals, the system can generalize well, meaning it can accurately recognize gestures from different people with varying hand shapes, sizes, and movements.

5.5 Advantages of YOLO compared to CNN

Convolutional neural network has max pooling layers which lead to slow processing. CNN has many layers for training, so the computer takes a lot of time to train the model. CNN requires a lot of data points for training the model. In contrast to CNN, coordinate frames can't be used. These coordinate frames are part of computer vision. These frames are used to keep track of the orientation and different features of an object. In real time detection we need to define the frame for detection of objects. It will detect images only in constrained areas. So this is the main disadvantage of CNN. YOLO can detect images at any position with fast processing. So this is the main reason why we are choosing YOLO. The same comparison can be noticed in Table 1.

Parameters	**CNN**	**YOLO**
Accuracy	Less	More
Accuracy Score	0.53	0.88
Processing Time	Slow	Fast
Realtime detection	Slow	Fast

Table 5.1: CNN vs YOLO

Chapter VI

6 Testing

For the end-to-end testing process, we utilized a test image set of 10 images, captured in varied conditions. These images were not included in the training dataset, to ensure the most accurate testing scenarios. The Deep Learning model's performance was tested based on our data set. The data set consisted of 2000 images. . The data set was split into train, test and evaluation data at 80 percent for training, 20 percent for testing and and the remaining 20 percent for evaluation of the models performance. This testing resulted in an accuracy of percent of the deep learning models in classification of images. Figure 6.1 shows the classes the data set has been trained for.

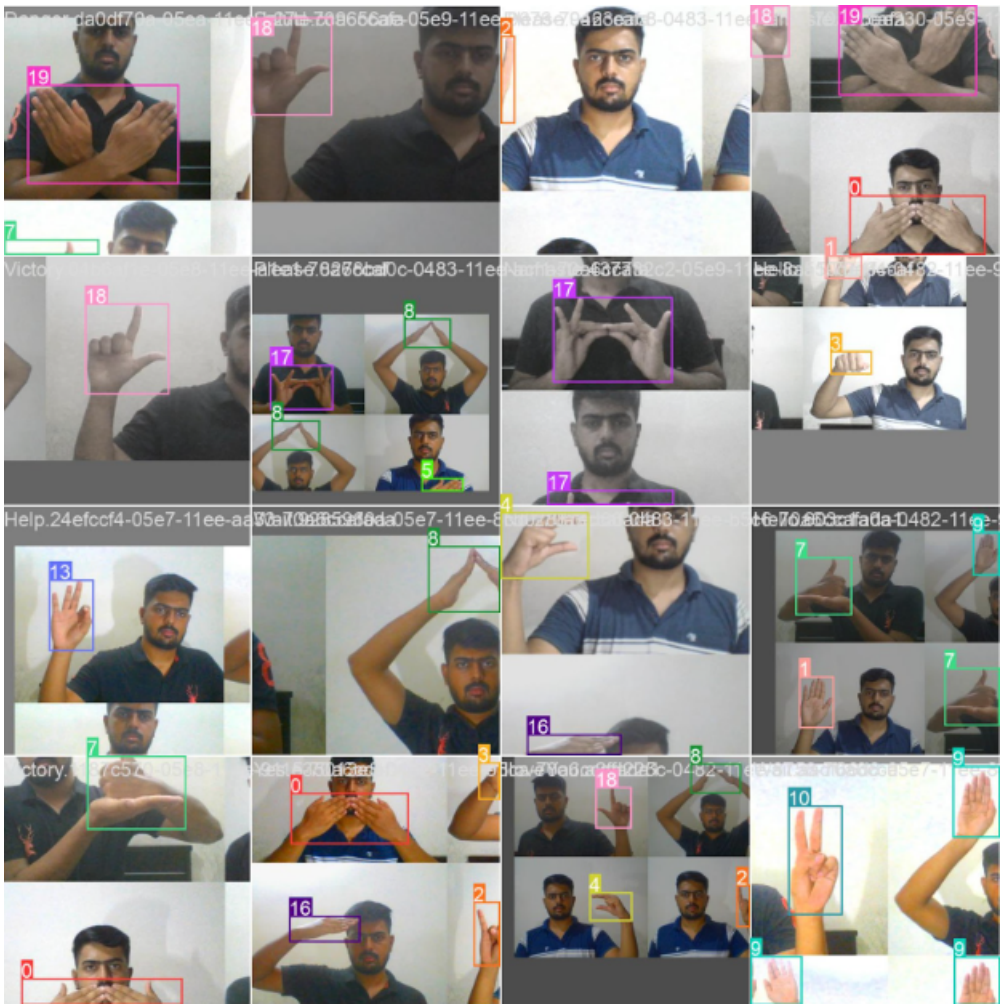


Figure 6.1: Image Classes

6.1 Formulae

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Precision is calculated as the ratio of true positives to the number of total guesses of positives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Accuracy is defined as the fraction of total predictions that the model gets correct.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Recall is defined as the measurement of the proportion of actual positives that was identified correctly.

$$F1\text{-}mean = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

F-1 scores is the harmonic mean of both recall and precision.

6.1.1 Mean Average Precision

The mean average precision (mAP) is a measure of an object detection model's performance. It is determined by averaging the average accuracy (AP) values for each class at a certain intersection over union (IoU) threshold. The IoU threshold specifies how much overlap must exist between the anticipated and ground truth bounding boxes for a detection to be judged accurate. The mAP@0.5:0.95 measure computes the mAP using an IoU threshold range of 0.5 to 0.95 in 0.05 increments. mAP is given by the following formula.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i$$

In this report, we described a model for sign language recognition based on the YOLOv5. The sign gesture recognition model can recognise real-time motions from video in real time and convert them to voice with a 99.5% accuracy. Furthermore, we evaluated the performance and execution timings of YOLOV5 with other models

and discovered that our suggested model was more effective at extracting needed characteristics from hand signs and recognised hand gestures with 99.5% accuracy, 98.2% precision,87.1% f1 score, and 100% recall.

```
custom_YOLOv5s summary: 182 layers, 7297761 parameters, 0 gradients
                 Class     Images  Instances          P          R      mAP50   mAP50-95: 100% 13/13 [00:04<00:00,  2.74it/s]
                   all        399        399      0.996          1      0.995      0.896
                Thanks        399         20      0.996          1      0.995      0.888
                 Hello        399         20      0.994          1      0.995      0.916
              IloveYou        399         20      0.997          1      0.995       0.87
                   Yes        399         20      0.996          1      0.995      0.825
                    No        399         20      0.997          1      0.995      0.888
                Please        399         20      0.996          1      0.995      0.878
                   Ask        399         20      0.996          1      0.995        0.9
                  Help        399         20      0.995          1      0.995      0.964
                  Home        399         19      0.996          1      0.995      0.923
                  Wait        399         20      0.996          1      0.995      0.853
               Victory        399         20      0.996          1      0.995      0.931
              Thumbsup        399         20      0.996          1      0.995      0.881
            Thumbsdown        399         20      0.995          1      0.995       0.93
                  Okay        399         20      0.995          1      0.995      0.954
               Namaste        399         20      0.996          1      0.995      0.867
                Callme        399         20      0.997          1      0.995      0.826
                Salute        399         20      0.996          1      0.995      0.836
                 Dance        399         20      0.995          1      0.995      0.908
                Looser        399         20      0.995          1      0.995       0.94
                Danger        399         20      0.996          1      0.995      0.933
Results saved to runs/train/yolov5s_results
CPU times: user 1min 28s, sys: 8.93 s, total: 1min 37s
Wall time: 2h 15min 29s
```

Figure 6.2: mAP Values for various signs

Chapter VII

7 Experimentation and Results

7.1 Experimentation

Extensive experimentation was carried out when working with this particular gesture recognition system. When approaching the problem of gesture recognition, various models were experimented with to compare and contrast between their accuracy and efficiency. In terms of gesture recognition as a whole, Single Stage detectors outperform all other methods of gesture recognition and YOLO-v5S seems to outperform any other single stage detector. Within yolo itself, there exist variations of the model for further fine tuning of accuracy and efficiency but the base YOLOv5S seemed to have a good balance between both of the metrics and we proceeded with this instead.

Figure 7.1: Ground Truth

7.2 Results

The model has been trained to accurately recognize a specific set of signs, precisely 20 signs. However, it is important to note that the results achieved by the model can vary depending on the individual characteristics of the signer. For example, differences in hand shape, size, and movement can impact the performance of the system. Despite these potential variations, the current implementation of the model has demonstrated impressive accuracy levels, providing reliable sign recognition. Additionally, the model operates with very low latency, allowing for fast and responsive sign recognition in real-time. Overall, the successful training and performance of the model is a significant achievement and provides a strong foundation for further development and improvement. Figure 7.1 and Figure 7.2 show the results for signs "Hello" and "Namaste" respectively.

Figure 7.2: Result for sign Hello

Figure 7.3: Result for sign Namaste

7.2.1 Confusion Matrix

A confusion matrix is a table that is frequently used to describe the performance of a classification model on a collection of data with known true values. As shown in Figure 7.3, it compares the anticipated and reality values and shows the findings

as a matrix. The number of true positives, false positives, true negatives, and false negatives is displayed in the matrix. This data may be used to compute performance measures including accuracy, precision, recall, and F1-score.

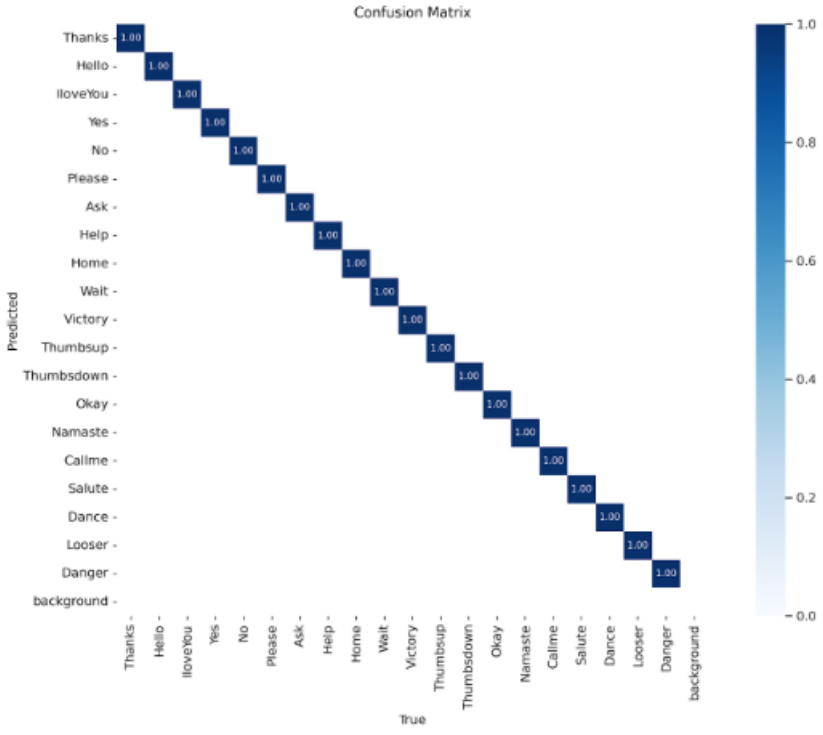


Figure 7.4: Confusion matrix

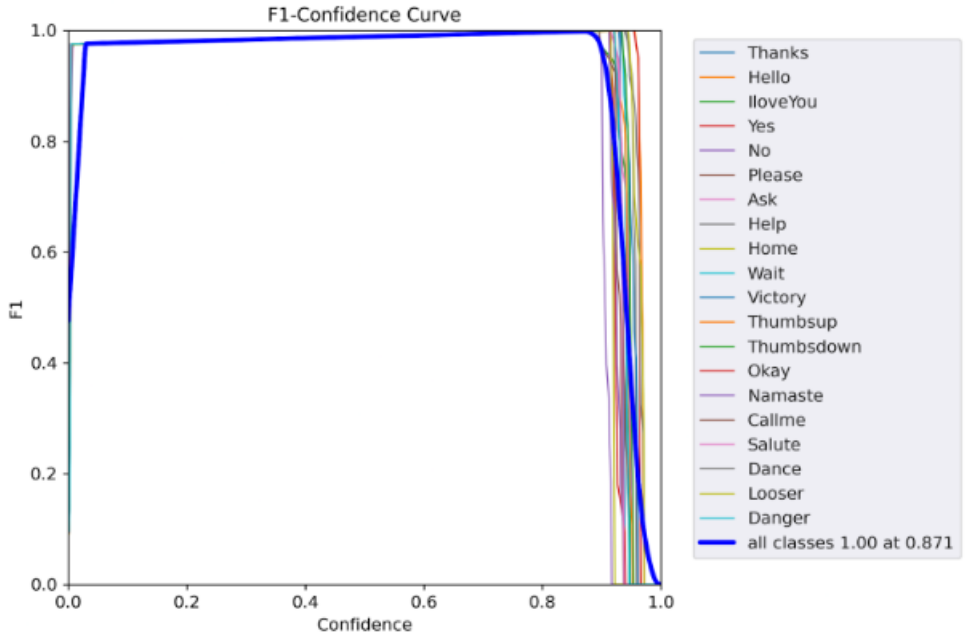


Figure 7.5: F1 Confidence Curve

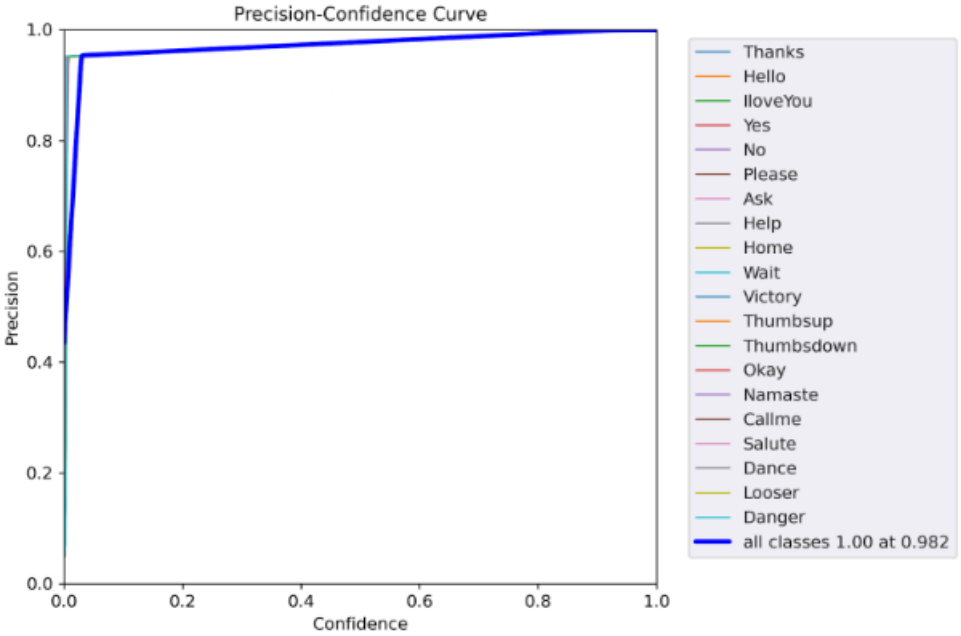


Figure 7.6: Precision Confidence Curve

7.2.2 F1 Confidence Curve

Figure 7.4 shows the F1 Confidence Curve. An F1-score is a measure of a model's accuracy that takes precision and recall into account. It is the harmonic mean of accuracy and recall and has a value between 0 and 1, with 1 being the greatest attainable result. An F1 confidence curve illustrates the F1-score vs a set of confidence levels. This shows how the F1-score varies as the confidence level is changed.

7.2.3 Precision Confidence Curve

Figure 7.5 shows the Precision Confidence Curve. Precision is a metric that measures how successfully a model recognises true positives. It is determined by dividing the number of true positives by the total number of true positives and false positives. A precision confidence curve illustrates accuracy vs a set of confidence levels. This shows how the accuracy varies as the confidence threshold is changed.

7.2.4 Precision Recall Curve

A precision-recall curve shown in Figure 7.6 depicts the connection between accuracy and recall in a binary classification model. It is generated by displaying accuracy vs recall at various levels. The area under the precision-recall curve (AUC) can be used to assess the performance of the model. The AUC of a model with perfect

accuracy and recall would be 1.

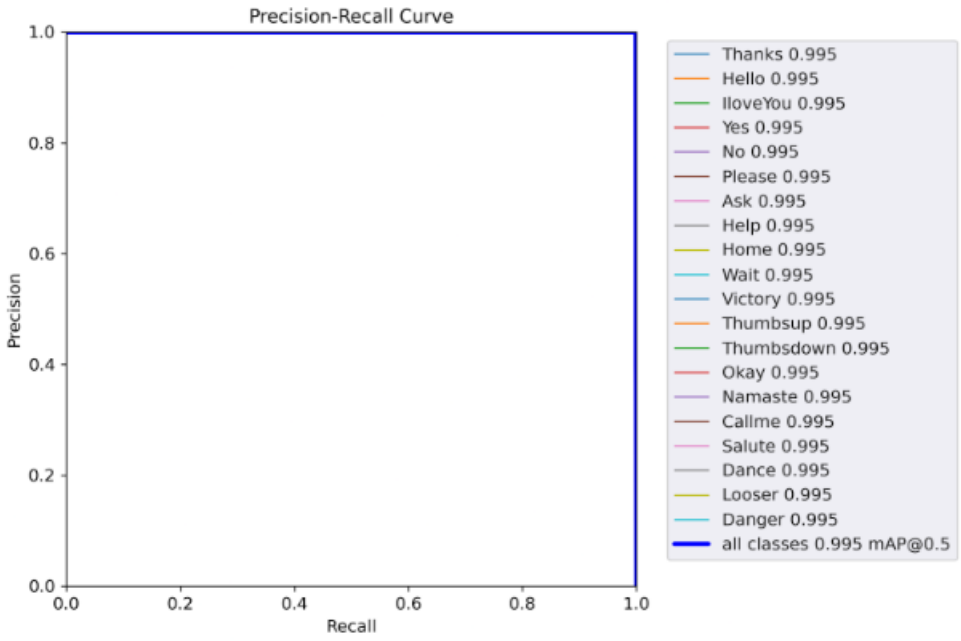


Figure 7.7: Precision Recall Curve

7.3 Plot Results

7.3.1 train/box_loss

The box loss is a component of the training loss that assesses how effectively an object detection model can localize items in training pictures. It compares the predicted bounding boxes to the ground truth bounding boxes and computes a loss number that indicates the model's performance. During training, the aim is to minimize box loss so that the model can correctly predict the placement of objects in fresh photos.

7.3.2 train/obj_loss

The object loss is a component of the training loss that indicates how successfully an object recognition model recognises objects in training pictures. It compares the predicted objectness scores to the ground truth labels and computes a loss value to determine how well the model is doing. During training, the aim is to minimize object loss so that the model can correctly predict the existence of objects in fresh pictures.

7.3.3 train/cls_loss

The classification loss (cls_loss) is a component of the training loss that assesses how successfully an object detection model classifies objects in training pictures. It

compares projected class probabilities to ground truth labels and computes a loss number that represents how well the model performs. During training, the aim is to minimize classification loss such that the model can correctly predict the class of objects in fresh pictures.

7.3.4 metrics/precision

A precision curve shows a binary classification model's accuracy versus a range of confidence criteria. This shows how the accuracy varies as the confidence threshold is changed. The area under the accuracy curve (AUC) may be used to assess the performance of the model. The AUC of a model with perfect accuracy would be 1.

7.3.5 val/box_loss

The box loss is a validation loss component that gauges how effectively an object detection model can accurately localize items in validation pictures. It compares the predicted and true bounding boxes and generates a loss number that shows how effectively the model generalizes to new data. The purpose of validation is to minimize box loss so that the model can predict the placement of objects in fresh pictures accurately.

7.3.6 val/obj_loss

The object loss is a component of the validation loss that assesses how effectively an object detection model recognises objects in validation pictures. It compares the projected objectivity scores to the ground truth labels and computes a loss value that indicates how effectively the model generalizes to new data. The purpose of validation is to minimize object loss so that the model can predict the existence of objects in fresh photos accurately.

7.3.7 val/cls_loss

The classification loss (cls_loss) is a validation loss component that analyzes how effectively an object detection model can accurately categorize objects in validation pictures. It compares projected class probabilities to ground truth labels and computes a loss number to indicate how effectively the model generalizes to new data. During validation, the aim is to minimize classification loss such that the model can correctly predict the class of objects in fresh pictures. Figure 7.7 shows the above criteria in graph form.

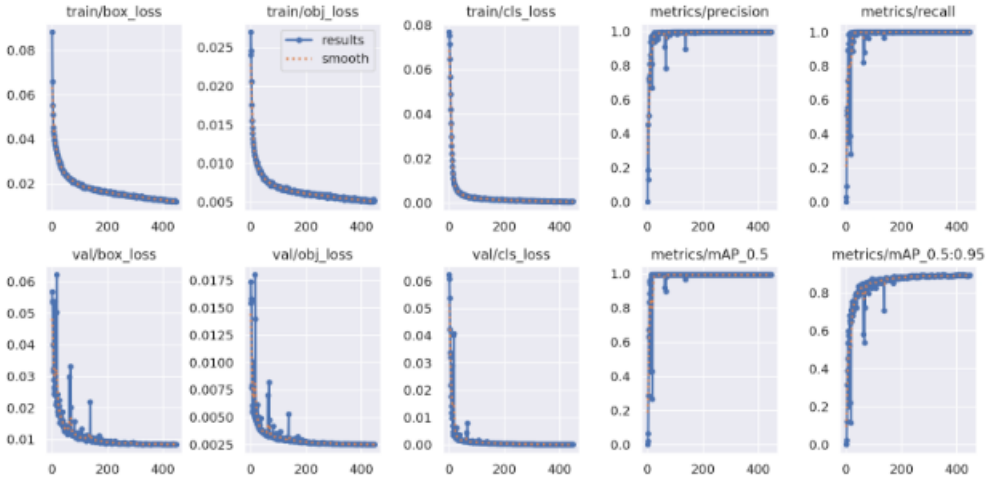


Figure 7.8: Plot Results

Chapter VIII

8 Conclusion and Future Enhancement

8.1 Conclusion

• Numerous benefits can be realised by implementing the YOLO (You Only Look Once) algorithm in gesture recognition systems. As a result of YOLO's real-time object identification capabilities, hand motions may be efficiently and accurately detected and tracked in real-time, enabling instantaneous and seamless interpretation of gestures as well as responsive and interactive applications.

• The YOLO algorithm's ability to process envision or footage in a single pass is one of its main advantages. This feature assures low latency and high throughput, making it suitable for workloads requiring gesture recognition. By quickly evaluating the input data and classifying hand motions, YOLO can provide real-time feedback and enable quick system responses. In applications like gaming, interactive interfaces, or virtual reality systems, where prompt identification and response to gestures are essential, this real-time capacity is very significant.

• Additionally, YOLO's the ability to determine numerous items at once is quite beneficial for circumstances when various hand or gesture movements need to be acknowledged simultaneously. Users may make several gestures simultaneously or quickly in applications like sign language recognition or interactive virtual environments. The accuracy with which these motions may be identified and tracked thanks to YOLO's handling of multiple object detection improves both system efficiency and user experience.

• The effectiveness of gesture recognition systems can be improved by the precise detection and monitoring of hand gestures provided by YOLO. Smooth and intuitive interactions are made possible by the system's ability to understand gestures precisely and reliably. This accuracy is especially useful in applications like robotics, smart home automation, or assistive technologies that call for exact control or command execution.

• In conclusion, the YOLO algorithm's incorporation into gesture recognition systems has several positive effects. The efficient and accurate identification and

tracking of hand gestures is made possible by its real-time processing capabilities, effective detection, and capacity to handle many objects at once. This in turn enables rapid and seamless interpretation of gestures, resulting in applications that are responsive and engaging across a range of areas.

8.2 Future Enhancement

This dynamic gesture recognition project can be explored for a number of potential future improvements.

8.2.1 Multi-modal Integration

Investigate the incorporation of several sensing modalities, such as adding depth sensors, wearable tech, or other sensors to record various gesture data points. The accuracy and robustness of the gesture detection system can be improved by combining input from various modalities.

8.2.2 Gesture Synthesis

Create methods for creating synthetic gestures that resemble real-world gestures based on recognised gestures. To make the user experience more interactive and immersive, this can entail making realistic hand animations or synthesising noises that are associated with gestures.

8.2.3 Gesture-based Interaction with Smart Devices

The project's application can be expanded to support gesture-based communication with a variety of smart devices, including smartphones, smart TVs, and home automation systems. Users would thereafter be able to operate equipment and carry out tasks using simple hand gestures.

8.2.4 Deep Learning and Transfer Learning Techniques

Look into using deep learning or transfer learning techniques to boost the performance of the model. Convolutional neural networks (CNNs), for example, can be adjusted or tailored to the unique task of gesture detection, potentially improving accuracy and efficiency.

8.2.5 Gesture Recognition in Difficult Situations

Increase the system's resistance to adverse situations like dim lighting, occlusions, or background noise. This can entail creating methods or algorithms that are more resistant to these environmental influences.

8.2.6 Gesture-based Accessibility

Investigate the ways in which the gesture recognition system can be used in the field of accessibility to help people with disabilities or impairments engage with digital devices or access information through gestures.

8.2.7 Constant Learning and Adaptation

Create the system so that it is constantly learning from new gestures or evolving user behaviour. To increase the system's performance and adaptability, this may entail implementing adaptive algorithms or online learning approaches.

These are but a few potential upgrades for a project on dynamic gesture detection. The particular improvements selected will be determined by the project's objectives, target audience, resources available, and desired level of complexity.

Chapter VIII

9 References

1. Mownika Reddy K A, Prajna Harish, Sambhrama K, Vinayak Bhat, "A Review on Dynamic Hand Gesture Recognition Techniques", International Journal for Research in Applied Science and Engineering Technology, 2023, doi: 10.22214/ijraset.2023.49256.

2. Li, Min, Zhenjiang Miao, and Cong Ma. "Dance movement learning for labanotation generation based on motion-captured data." IEEE Access 7 (2019): 161561-161572.

3. Natarajan, B., et al. "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, andVideo Generation." IEEE Access 10 (2022): 104358-104374.

4. Zong, Yan, et al. "Robust Synchronized Data Acquisition for Biometric Authentication." IEEE Transactions onIndustrial Informatics 18.12 (2022): 9072-9082.

5. DelPreto, Joseph, et al. "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification." IEEE Robotics and Automation Letters 7.4 (2022): 10589- 10596.

6. Alwaely, Basheer, and Charith Abhayaratne. "Graph spectral domain feature learning with application to in-air hand- drawn number and shape recognition." IEEE Access 7 (2019): 159661-159673. Elshenawy, Abdelghafar R., and Shawkat, K. Guirguis. "On-Air Hand-Drawn Doodles for IoT Devices Authentication During COVID-19." IEEE Access 9 (2021): 161723- 161744.

7. Arsalan, Muhammad, Avik Santra, and Vadim Issakov. "RadarSNN: A Resource Efficient Gesture Sensing System Based on mm-Wave Radar." IEEE Transactions on Microwave Theory and Techniques 70.4 (2022): 2451-2461.

8. Bencherif, Mohamed A., et al. "Arabic sign language recognition system using 2D hands and body skeleton data." IEEE Access 9 (2021): 59612-59627.

9. Luqman, Hamzah. "An Efficient Two- Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion." IEEE Access 10 (2022): 93785- 93798.

10. Xu, Jun, et al. "Robust Hand Gesture Recognition Based on RGB-D Data for Natural Human-Computer Interaction." IEEE Access (2022).

11. Al-Hammadi, Muneer, et al. "Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation." IEEE Access 8 (2020): 192527-192542.

12. Lee, Minhyuk, and Joonbum Bae. "Deeplearning based real-time recognition of dynamic finger gestures using a data glove." IEEE Access 8 (2020): 219923-219933.

13. Y. Li, R. Cheng, C. Zhang, M. Chen, J. Ma and X. Shi, "Sign language letters recognition model based on improved YOLOv5," 2022 9th International Conference on Digital Home (ICDH), Guangzhou, China, 2022, pp. 188-193, doi: 10.1109/ICDH57206.2022.00036.

14. T. F. Dima and M. E. Ahmed, "Using YOLOv5 Algorithm to Detect and Recognize American Sign Language," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 603-607, doi: 10.1109/ICIT52682.2021.9491672.

15. A. Puchakayala, S. Nalla and P. K, "American Sign language Recognition using Deep Learning," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 151-155, doi: 10.1109/ICCMC56507.2023.10084015.

16. G. Dai, L. Hu, J. Fan, S. Yan and R. Li, "A Deep Learning-Based Object Detection Scheme by Improving YOLOv5 for Sprouted Potatoes Datasets," in IEEE Access, vol. 10, pp. 85416-85428, 2022, doi: 10.1109/ACCESS.2022.3192406.

17. X. Yuan, A. Kuerban, Y. Chen and W. Lin, "Faster Light Detection Algorithm of Traffic Signs Based on YOLOv5s-A2," in IEEE Access, vol. 11, pp. 19395-19404, 2023, doi: 10.1109/ACCESS.2022.3204818.

18. G. Dai, L. Hu, J. Fan, S. Yan and R. Li, "A Deep Learning-Based Object Detection Scheme by Improving YOLOv5 for Sprouted Potatoes Datasets," in IEEE Access, vol. 10, pp. 85416-85428, 2022, doi: 10.1109/ACCESS.2022.3192406.

19. Q. Wang, X. Li and M. Lu, "An Improved Traffic Sign Detection and Recognition Deep Model Based on YOLOv5," in IEEE Access, vol. 11, pp. 54679-54691, 2023, doi: 10.1109/ACCESS.2023.3281551.

20. X. Yuan, A. Kuerban, Y. Chen and W. Lin, "Faster Light Detection Algorithm of Traffic Signs Based on YOLOv5s-A2," in IEEE Access, vol. 11, pp. 19395-19404, 2023, doi: 10.1109/ACCESS.2022.3204818.

DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

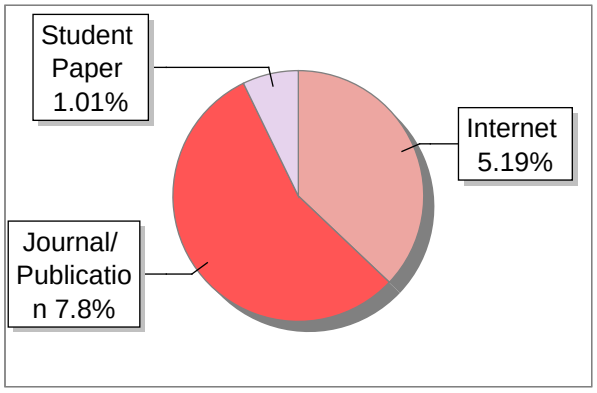
Author Name	Mownika Reddy K A
Title	DYNAMIC GESTURE RECOGNITION
Paper/Submission ID	771738
Submission Date	2023-06-12 13:41:39
Total Pages	56
Document type	Project Work

Result Information

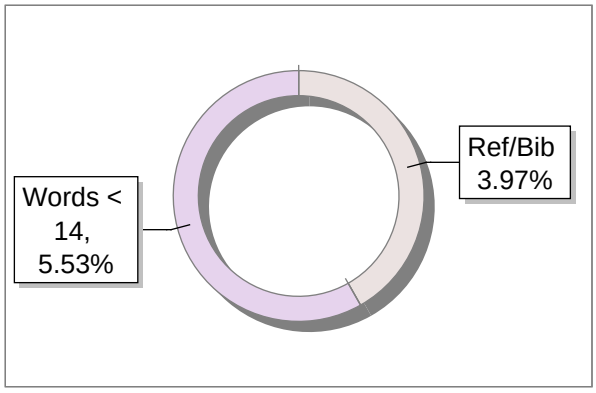
Similarity **14 %**

1 10 20 30 40 50 60 70 80 90

Sources Type



Report Content



Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Sources: Less than 14 Words Similarity	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

14	110	B
SIMILARITY %	MATCHED SOURCES	GRADE

A-Satisfactory (0-10%)
B-Upgrade (11-40%)
C-Poor (41-60%)
D-Unacceptable (61-100%)

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	www.dx.doi.org	1	Publication
2	www.mdpi.com	1	Internet Data
3	Submitted to Visvesvaraya Technological University, Belagavi	1	Student Paper
4	Dynamic hand gesture recognition usingRGB-D data for natural human-computer int by Linqin-2017	1	Publication
5	www.ijcttjournal.org	<1	Publication
6	www.isca.co.in	<1	Publication
7	www.researchgate.net	<1	Internet Data
8	coek.info	<1	Internet Data
9	Thesis Submitted to Shodhganga Repository	<1	Publication
10	IEEE 2018 IEEE Nuclear Science Symposium and Medical Imaging Confere	<1	Publication
11	Submitted to Visvesvaraya Technological University, Belagavi	<1	Student Paper
12	www.sapub.org	<1	Publication
13	www.ijettjournal.org	<1	Publication

14	dblp.org	<1	Internet Data
15	dsatm.edu.in	<1	Publication
16	www.ncbi.nlm.nih.gov	<1	Internet Data
17	intbrazjurol.com.br	<1	Publication
18	mdpi.com	<1	Internet Data
19	www.dx.doi.org	<1	Publication
20	dspace.nwu.ac.za	<1	Publication
21	moam.info	<1	Internet Data
22	beei.org	<1	Internet Data
23	Condition-based maintenance methods for marine renewable energy by Mrigaud-2016	<1	Publication
24	ir-library.ku.ac.ke	<1	Publication
25	worldwidescience.org	<1	Internet Data
26	www.dx.doi.org	<1	Publication
27	bmcbioinformatics.biomedcentral.com	<1	Internet Data
28	coek.info	<1	Internet Data
29	eproceeding.undiksha.ac.id	<1	Internet Data
30	religiondocbox.com	<1	Internet Data
31	uir.unisa.ac.za	<1	Publication
32	www.dx.doi.org	<1	Publication

33	www.dx.doi.org	<1	Publication
34	www.scribd.com	<1	Internet Data
35	A Multi-Scale Approach for Remote Sensing Scene Classification Based on Feature by Zhang-2019	<1	Publication
36	Thesis Submitted to Shodhganga Repository	<1	Publication
37	www.mdpi.com	<1	Internet Data
38	qdoc.tips	<1	Internet Data
39	citeseerx.ist.psu.edu	<1	Internet Data
40	Submitted to Visvesvaraya Technological University, Belagavi	<1	Student Paper
41	Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in	<1	Publication
42	Thesis Submitted to Shodhganga Repository	<1	Publication
43	mdpi.com	<1	Internet Data
44	Thesis Submitted to Shodhganga Repository	<1	Publication
45	www.ijsrd.com	<1	Publication
46	IEEE 2014 15th Conference of Open Innovations Association FRUCT and by	<1	Publication
47	IEEE 2018 9th International Conference on Computing, Communication , by Warrier, Sreejith R- 2018	<1	Publication
48	adeshpande3.github.io	<1	Internet Data
49	pure.royalholloway.ac.uk	<1	Publication
50	Submitted to Visvesvaraya Technological University, Belagavi	<1	Student Paper

51	www.dx.doi.org	<1	Publication
52	www.dx.doi.org	<1	Publication
53	www.ncbi.nlm.nih.gov	<1	Internet Data
54	arxiv.org	<1	Publication
55	dochero.tips	<1	Internet Data
56	Thesis Submitted to Shodhganga Repository	<1	Publication
57	Attention-Based Siamese Region Proposals Network for Visual Tracking by Wang-2020	<1	Publication
58	Formulation of a three-dimensional shelf edge model and its application to inter by Jiuxin-1998	<1	Publication
59	Thesis submitted to dspace.mit.edu	<1	Publication
60	Towards Smart Notifications using Research in the Large, by D. Weber A. S. Shi- 2015	<1	Publication
61	citeseerx.ist.psu.edu	<1	Publication
62	Edge computing Enabling technologies, applications, and services by Gao-2020	<1	Publication
63	inba.info	<1	Internet Data
64	mdpi.com	<1	Internet Data
65	moam.info	<1	Internet Data
66	privacynewshighlights.wordpress.com	<1	Internet Data
67	Student Paper Published	<1	Internet Data

68	The growth of e-government in municipal planning by Evans-Cowley-2006	<1	Publication
69	uir.unisa.ac.za	<1	Publication
70	worldwidescience.org	<1	Internet Data
71	www.ncbi.nlm.nih.gov	<1	Internet Data
72	www.scribd.com	<1	Internet Data
73	downloads.hindawi.com	<1	Publication
74	etd.aau.edu.et	<1	Publication
75	mdpi.com	<1	Internet Data
76	moam.info	<1	Internet Data
77	www.dx.doi.org	<1	Publication
78	www.jaas.ai	<1	Internet Data
79	An architecture for adaptive task planning in support of IoT-based machine learn by Sacco-2020	<1	Publication
80	Design of a real-time gesture recognition system high performance through algor by Ozer-2005	<1	Publication
81	ijstr.org	<1	Internet Data
82	repositorio.uam.es	<1	Publication
83	IEEE 2018 Fourth International Conference on Research in Computation	<1	Publication
84	animeturkey.com	<1	Internet Data
85	avxlive.icu	<1	Internet Data

86	ayomenulisfisip.files.wordpress.com	<1	Publication
87	beei.org	<1	Internet Data
88	Deep Learning Based Traffic Signs Boundary Estimation by Hrustic-2020	<1	Publication
89	docplayer.net	<1	Internet Data
90	drait.edu.in	<1	Publication
91	IEEE 2018 Fourth International Conference on Computing Communicatio, by Kulkarni, Omkar- 2018	<1	Publication
92	ijitee.org	<1	Internet Data
93	Lecture Notes in Computer Science Computer Vision ECCV 2020 Volume 12364 (	<1	Publication
94	mdpi.com	<1	Internet Data
95	mdpi.com	<1	Internet Data
96	mdpi.com	<1	Internet Data
97	moam.info	<1	Internet Data
98	moam.info	<1	Internet Data
99	moam.info	<1	Internet Data
100	moam.info	<1	Internet Data
101	smec.ac.in	<1	Publication
102	Thesis Submitted to Shodhganga, shodhganga.inflibnet.ac.in	<1	Publication
103	Thesis Submitted to Shodhganga Repository	<1	Publication

104	The Penalties of Physical Disability by Hislop-1976	<1	Publication
105	vdocuments.mx	<1	Internet Data
106	When school-based, in-service teacher training sharpens pedagogical awareness by Lund-2018	<1	Publication
107	www.igd.fraunhofer.de	<1	Internet Data
108	www.ijeat.org	<1	Publication
109	www.mdpi.com	<1	Internet Data
110	IEEE 2019 3rd International conference on Electronics, Communication	<1	Publication

iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49256

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

A Review on Dynamic Hand Gesture Recognition Techniques

Mownika Reddy K A[1], Prajna Harish[2], Sambhrama K[3], Vinayak Bhat[4], Dr. Deepak. G[5], Dr. Harish Kumar[6]
[1, 2, 3, 4]*Student,* [5,6]*Associate Professor, Dept of CSE, Dayananda Sagar College of Engineering Bangalore, India*

Abstract: ***Body language is one of the nonverbal methods of communication, and it comprises hand gestures, arm movements, posturing, and gestures and facial expressions. One way to communicate information through the movement of the body is through gestures.***
HGR is a smart, intuitive, and easy method of human-computer interaction (HCI). HGR systems have two key applications: SLR and GBC. To help the deaf communicate with the hearing community, SLR tries to automatically interpret SLs via a computer. The idea that SL is a highly ordered and primarily symbolic collection of human gestures is what led to the development of universal gesture-based HCI.
Keywords: HGR, ASL, IPT, Sign language Recognition (SLR), Image processing techniques[IPT], gesture-based control[GBC], Sign language [SL], British Sign Language (BSL), Human-computer interaction[HCI], Australian Sign Language (ASL), Gesture Spotting Algorithm(GSA), Hand gesture recognition[HGR], American Sign Language (ASL), Deep Learning(DL), skeletal CNN network(SCN),Gesture Progress Sequence (GPS), Index Inference Score (IES), Sequence Simplification Algorithm(SSA), Improved Dynamic Time Warping (IDTW), K- Curvature-Convex Defects Detection (K- CCDD), Gesture RecognitionAlgorithm(GRA)

I. INTRODUCTION

Typically, gestures entail moving the hands, face, or other body parts. It is a type of non- verbal communication in which specific directly observable actions are used in place ofor in addition to words to convey certain information.
In fact, according to some researchers, communication in Homo sapiens originated from a manual gesture-based form of communication in the past. Gestural Theory,which has its origins in the works of the priest and philosopher Abbé de Condillac from the 18th century, was resurrected in 1973 as a partof a discussion on the genesis of language. Gestures can be static or dynamic. An appealing technique for enabling realistichuman-computer interaction is direct hand input (HCI). While overcoming this restriction, vision-based solutions still have to deal with other issues caused by the user's body being partially obscured. Vision-based techniques can vary from one another depending on a variety of factors, including the quantity of sensors used, their delay and responsiveness, the organization of the surroundings, any user needs, the low-level characteristics used, and whether two-Dimensional or three-Dimensional representation is used.

A. Types of Gestures

Hand and Arm gestures: SL, HGR, entertainment applications. Head and Face gestures: Shaking/nodding one's head, eye gaze direction, arching one's brow, opening one's lips to speak, winking, flared nostrils, and expressions of surprise, as well as feelings likehappiness, fear, disgust, wrath, sorrow, anddisdain.
Body gestures: engagement of complete body movement, like recording the outside interactions of two individuals, assessing the dancer's motions to create music and images that fit, and identifying human gaits for athleticand medical training.

1) *Sign Language:* Deaf people frequently communicate through SL, a form of manual communication. The movements or symbols are organized linguistically in SL. A sign is a gesture that exists on its own. The hand shape,hand placement, and hand movement are the three distinguishing components of each sign. Conversation between people is transformed into HCI through a SLR system. The SL recognition system aims to provide deaf and hearing persons with a precise and efficient method of understanding text and voice makingit easier for the deaf and hearing people to communicate. There is no globally recognized SL for the deaf. Furthermore, just like spoken languages, SLs vary from location to region and are not universal. It's interesting to note that most nations with identical spoken languages do not necessarily share the same SL. For instance, there are three types of English:ASL, BSL, and ASL. Despite the widespread assumption that gestures are not "genuine languages," they are just as linguistically complex and rich as any spoken language.

Various SLs have been examined bylinguists in the field, and they have discoveredthat they include dual patterning and recursionin common with other languages. Languages are made up of fewer, useless components that may be merged to form bigger, meaningful ones. This is known as the duality of patterning.Recursion refers to the fact that languages havegrammatical rules, and that a rule's output mayalso serve as its input.

II. LITERATURE SURVEY

With the help of motion-captured data, they hope to provide Labanotation scores. They startoff by offering a reliable feature to obtain a helpful depiction of dance motions from the motion data after processing. The dancing movement segments are then used as input to instruct an HMM to match the Labanotation reference symbols for each type of lower leg motions. In particular, they suggest using a highly random tree technique to extract arm movements from upper limb data. In order to determine the notations in both the support columns and the arm columns, they employ the motion data. Finally, a dance piece's Labanotation score can be produced. Studies reveal that when compared to earlier methods, their system generates symbols with greateraccuracy. This could lighten the workload. [1] The analysis proceeds the evolution of a DL system for diverse SL union, conversion, and identification. They decrypt the affair that accompanied by preceding SL characterization and recording synthesis methods by utilizing the recommended H-DNA architecture. Using the How2Sign, ISL-CSLTR, and RWTH- PHOENIXWeather 2014T datasets, they quantitatively and subjectively evaluated the model's performance. The suggested H-DNA structure is also qualitatively rated using a variety of variety criteria. Produced recording sequences demonstrate the high caliber of theirwork. They outperformed preceding approaches in terms of identification rate and initiate interpretation. Produced video segments demonstrate the high caliber of their work. They performed better than earlier methods in terms of detection rightness and output. Suggest method has impressive human evaluation scores, an average BLEU score of 38.56, an mean FID2vid tally of 3.46, an mean SSIM value of 0.921, an mean Inception Scoreof 8.4, an mean PSNR tally of 29.73, an mean FID tally of 14.06, and an mean TCM tally of 0.715. It also has a classification accuracy for SLR of over 95%. These results show a significant advancement over earlier models. Human assessors are used to assess the realism,relevance, and coherence aspects, and the outcomes are good in real-world circumstances.[2]

In order to enable simultaneous biosignal acquisition for biometric authentication across scattered electrode arrays, they synchronized the data sample timer in an anatomy detector network using the D-PkCOs procedure. By requesting just one collide package for every synchronization cycle, the D-PkCOs protocol decreases communication overhead. In order toreliably survey all BSN nodes susceptible to drifting clock frequency and changing processing latency, they employed a dynamic controller to alter the clock offset and skew for decreasing the sampling jitters. With this approach, the impacts of variable processing latency are automatically removed, and a more precise clock skew prediction is made for readjustment. Additionally, they employed the H control approach to create the D-PkCOs synchronization protocol's parameters. . Additionally, they employed the H control approach to create the D-PkCOs synchronization protocol's parameters. Because all nodes' sample errors are kept to a minimum in the BSN, the drifting clock and varying processing latency have no effect on the sampling jitters. With their D-PkCOs procedure, sampling jitters may be kept to under 1s in a 10-node IEEE 802.15.4 network, according to experimental results. It is shown that when DO-PkCOs are applied to the BSN, an HD-sEMG signal with a high SNR value develops, which improves the performance of gesture categorization.[3]

An acceptable smart glove that measures motion with an accelerometer and posture with a strain-sensitive resistive knit. In order to read data, extract features, and execute a pre-trained computational model, a microcontroller, a tiny customized PCB, and sensors are utilized. Real-time classification of SL stances and gestures is done using the system.This work highlights the potential of fusing cutting-edge microcontrollers, machine learning, and innovative soft sensors. Future research, however, will be able to describe the learning pipeline's capabilities, constraints, and architecture in greater detail.Future research should broaden the range of subjects studied to assess robustness and generalizability, including if the network can be set up and used by additional users immediately or if a tuning technique needs to be included.

They might as well examine elements like the user's level of hand skin conductance, ASL experience, size, etc. Results of cross validation may provide guide changes to the structure to increase robustness.

Adjusting Filtering or categorization windows may also cut down on latency. Examining the learning pipeline's capability and trade-offs is also important. Gestures might be added with very little effort.Memory or speed by by setting the softmax only up to the point where the LSTM's learning capacity is exhausted. How many scales depend on scaling the LSTM layer or adding layers? the microcontroller must carry out in order to analyze the network. How much is impacted by network size and gesture count aswell. Data for training is necessary. To explainneural networks, these trade-offs betweenscale, velocity, precision, and training could be complex and application-specific.[4]

Novel graph spectral characteristics have been developed for dynamic form recognition. The proposed technique starts by pre-processing theinput, such as movement of the hands, to generate a fully connected graph. Next, thesuggested method analyzes the EV of the standardized Laplacian of the network adjacency matrix to produce the representativefeatures. They used the eigenvector u0 because it serves as the primary representative feature and captures the specifics of the graph's structure. The suggested method outperformedthe existing methods for three separate datasets, with accuracy values for numbers andsymbols of 99.56% and 99.44%, respectively. The appropriate approach also has the characteristics of quick operation and rotational and flipping invariance.[5]

Brand-new air-handwritten password-based authentication mechanism for IoT devices is proposed. This method is an application of computer vision where it identifies the line which is drawn on the air by making use of a camera, pair light-weighted deep CNN models and a Kalman signal processing filter. This combination serves as the main differentiation between this framework and the alternatives.The findings demonstrated the acceptability and significance of the suggested authentication approach for accessibility metrics such user happiness, precision, and speed. The suggested approach is safe and resistant to dangers from physical observation. There are no equipment, wearable sensors, or depth cameras required for this procedure. Future iterations of the intended design will be quick, straightforward, and appropriate for managing devices like smart television, smart watches, smart refrigerators, and smart air conditioners. Suggested method's drawback is that it is ineffective in low light.[6]

A SNN-based FMCW radar-based gesture recognition system. They recommend taking the bandwidth, Doppler, and angle spectral analysis from video of RDIs and feeding them as characteristic pictures into their proposed SNN architecture. SNNs are intriguing for various applications involving human-machine interaction because they offer embedded minimal latency and power solutions. A significant problem while training the SNNs is improving the algorithms to learn multimodal peak trains. They demonstrate that the recommended peaking network, which has better learning principles and is significantly smaller in size, can match for eight motions. [7]Framework for automatic recognition of Arabic SL that makes use of a fresh ArSLDataset that was captured on the grounds of their university. Three cameras—a Sony hand-held camera, a Kinect V1 camera, and a KinectV2 camera—were used to record the dataset. Each 80-sign set was captured by the 40 petitioners five times, creating a sample with multiple modes that comprises depthphotographs, RGB pictures, and bodily bones framework by V2 camera. In this study, they solely examined RGB pictures of the V2 camera by suggesting a sequential connection of both analogous networks—a 2-Dimensionalconvolutional grid and a second 1-dimensionalconvolutional skeletal grid. Their ideal networkconfiguration picked up 88.89% of static indications and 98.39% of dynamic indications, which is highly encouraging for the outcomes of the automatic ArSL. The trial effectiveness for the signer-reliant approachwas 89.62%, while the trial effectiveness for the signer-free mode was 88.09% when the exact system was conditioned using those staticand dynamic signals. The usage of the reverse efficiency revealed accuracy-speed trade-offcould have been made better with the appropriate amount if such models were used in production. [8]

For the recognition of SL, these deep neural models—SRN, DMN, and AMN—are suggested. The vital positions of the sign's principal poses are taught to the DMN stream. In this study, they suggest a method for addressing the differences between the sign samples by extracting key positions. This method makes use of the dominating postures that stand in for the important motion shifts in the symbol. They also recommendedcombining the sign's motions into a single picture using the AVM technique. The second suggested network, AMN, uses this image as its input. The characteristics out from DMN and AMN channels were pooled and utilized as input by the third network that was proposed, known as SRN. Two datasets were utilized to study these networks: the number of signers needed for histogram matching, bounding box computing, skin color separation, and region expansion; and signer free recognition is moredifficult than signer-dependent recognition. Correlation-based comparison and feature point comparison are two gesture comparison techniques. Other aspects of the applicationinclude word to motion translation and word tovoice out of word.[9]

A comprehensive hand gesture identification algorithm based on RGB-D by certain realistichand motion communication with the digital world. After obtaining the hand gesture outline,the Distance Transform (DT) technique estimates the palm center for static hand motion recognition.

The fingernails areidentified using the K-CCDD method. As supplementary features, the pixel spacing on the hand motion outline and the inclination between the fingers are utilized to generate a heterogeneous feature vector, and a specialized program is then introduced to accuratelyclassify static hand movements. Furthermore, acohesive explanation of each dynamic handmotion offers a IDTW approach to get dynamic hand gesture identification results by integrating Euclidean distance with bone modulus ratios separating the shoulder's centerjoints from the arm joints. They also develop an inexpensive live implementation of organichand gesture interaction with the digital world.Finally, thorough trials are carried out to confirm and validate both the still and moving hand motion detection algorithms.[10]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

III. METHODOLOGY

A. Input Preprocessing

A DL-Based Method for Recognizing SLR With Effective Hand Sign Representation. Using the Viola and Jones technique, the signer's face is located in the first method.

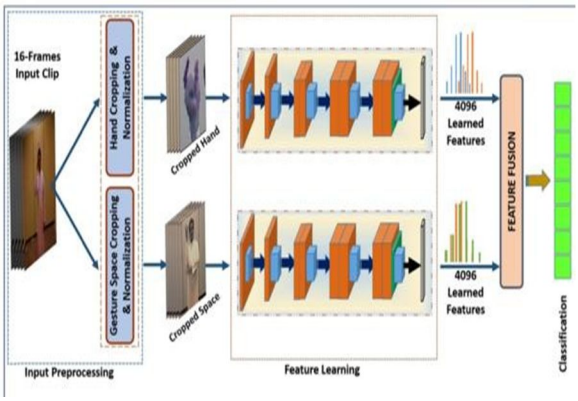
Trimming and spatial balancing lessen the influence of irrelevant characteristics in each frame. Contrarily, the second procedure trims and regulates the palm area to highlight more of the fingers arrangement. Two volumes per sample, each dimension 112 by 112 by 3 by 16,are produced during the pre - processing stage.The feature learning phase receives these two volumes, out of which one of them is dedicatedto the hand region, another represents the whole gesture area.[11]

B. Feature Learning

Hand configuration's precise spatial and temporal properties are learned by the maiden C3D instance. Each input segment for this occasion favors The Hand. The second C3D instance, on the other hand, picks up on the coarse spatial and temporal properties of the whole-body setup. The result of this step is twofeature vectors, each with 4096 size.[11]

C. Feature Fusion and Classification

After dimension reduction, we may acquire a precise representation of integrated features,resulting in reduced computing difficulty and higher face identification accuracy. Feature fusion aids in the complete learning of picture characteristics for the description of their rich internal information. Combining trainingpicture features vector from the common weight network layer with extracted features made up of other numerical data allows the proposed model to use as many features as feasible for the subsequent classification.[11]



Frame work

D. Gesture Spotting Algorithm

In a flow of frequently refreshed data, GSA locates the onset and termination of motions. The majority of earlier studies have suggested specific features that apply physical quantities like a tempo, an increase in tempo, or a recurrence in order to detect gestures. However, There are drawbacks to only using these "designed characteristics" for movement identification because There are too few of them to distinguish between genuine motions and othercommon fist movements like non-gestures or unintentional fist movements at their onset andtermination. As a result, they suggested thenew GSA. It employs a DL framework that canfind more characteristics while being educated.The DL-based gesture identification methodthat has been developed measures the GPS, anunique notion described in this work. The scalar value between 0 and 1 represents the GPThe gS. This is used to indicate where a movement starts and finishes. For instance, When a gesture starts, the GPS score is virtually nil. On the contrary, when the movement is going to stop, the GPS score is very near to one. The sum of the speed standards for each time step is basically how the GPS is determined. The GPS may be mathematically described as follows.

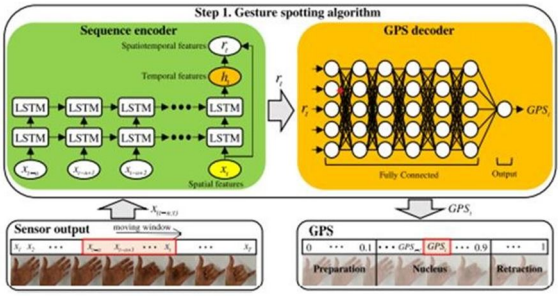
$$jv(st) = |mj(st) - mj(st-1)|, (st = 2, 3, \ldots, ST , jv(1) = 0)$$

$$Vsum(st) = Xst\ i=1\ jv(i)$$

$$GPS = \{ Vsum(1)\ Vsum(ST) =0,\ Vsum(2)\ Vsum(ST) , \ldots , Vsum(ST)\ Vsum(ST) =1\}$$

In (1), mj(st) ∈ R 10 is a metacarpophalangealjoint vector at time step t,

jv(st) ∈ R the second criterion of the disparity between the metacarpophalangeal joint vector at time steps st and st − 1. [12]



Sequence Simplification Algorithm: This introduces the SSA, which eliminates pace fluctuation in a succession of movements. A straightforward feature extractor, the SSA looks for significant modifications to themovement sequence. The SSA works asfollows. Beginning with the first sensor out of ten, the algorithm determines which sensor hasthe highest disparity between the maximum and minimum points of the movement sequence being measured by the sensors.

The mathematical expression:

i ∗ = argmax i (max ts SS i (ts) − min ts SS i(ts))

i is the SSI, which lies between 1 to 10. In the movement sequence, SS is a collection of sensor readings, so SS i (ts) is the sensor outputat time step ts measured by sensor i.

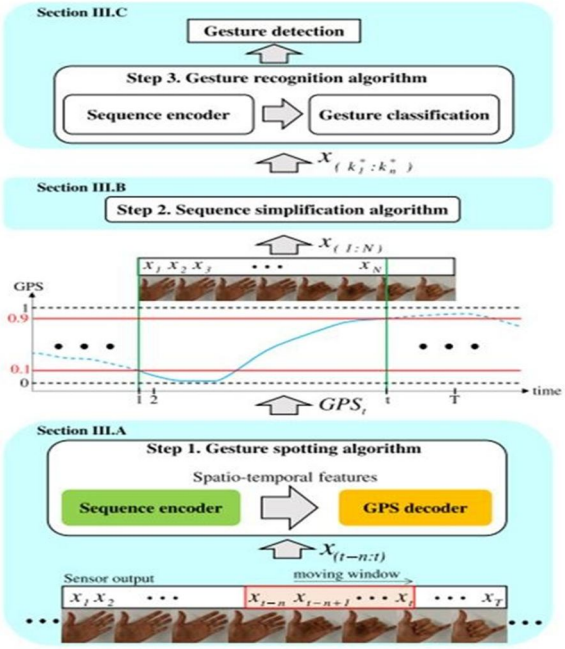
Second, It is defined as a line that runs across SS i's beginning and finishing locations. Third,The position of the data point with the greatestseparation from the line is looked up. The following is a mathematical formulation for this:

k1 = argmax d(SS i ∗ (k), SS i ∗ (1)SS i ∗

(TE)) 1<k<TE

TE represents the end of the sensor assessment,and The range between position a and line b isdetermined by the function d(a, b).[12]

Gesture Recognition Algorithm: The SSA produces an abbreviated design, which the GRA uses to classify the movement. To increase the GRA's resistance to changes in thesame movement that could not be completely ruled out after sequence reduction, a deep neural framework was used. The GRA consistsof the following: an output layer of outputaspect 11, three fully linked layers with 64 hidden units each, two LSTM layers with 64 hidden units, and three layers with 64 hidden units each. The output layer and the other threecompletely connected layers are activated by ReLU. Concatenation is not employed in the LSTM layer, which makes it different from themovement detecting method in certain ways. Analyzing the gesture sequence's context iscrucial for the GRA. Therefore, the currenthand shape does not need to be concatenated. The entire gesture sequence is not taken into account by the GRA. Since there are no more changes to the hand shapes after the GPS valuecrosses 0.9, it instead takes into account the gesture sequence from its inception to that point.[12]



CNN WITH IMPROVED PARALLELSKELETAL

A SCN supplied by 3D data, which was initially inspired by the three dimension Intel Real sense SDK's coordinates X, Y, and Z. This kind of camera creates the hands' 3D important points within a 3D environment. The research findings revealed around 91.28% effectivenesscovering around fourteen gestures in the hand gesture dataset, yet they also revealed certain shortcomings because of a decrease by a factor of 7% while there were twice as many classes totaling to 28. Their efforts to the redesigned SCN network focused on these key areas: First,they made the network architecture more straightforward by using 2D points rather than3D, which required one-third smaller computefor each network. As a next step, we increasedall input layers to include all 48 important areason the body.[8]

1) Performance Benchmarks Following are the metrics we will use:

a) Precision = True Positive rate/(True Positive rate + False Positive rate)

b) Recall = True Positive rate/(True Positive rate + False Negative rate)

c) F1 score = 2 *(precision* recall)/(precision + recall)

d) Accuracy = True Positive rate/(TruePositive rate + False Negative rate) [8]

2) IES

Providing the system several frames increases delay during the video detection procedure, while providing the system so very few frameslowers accuracy. The compromise between speed and precision has also been explored, and it has been determined that the IES may be applied in some circumstances. This following section is the most accurate amount of frames to transmit to our system. The following equation can be used to calculate

$$\text{IES} = \text{Response Time } 1 - \text{Percentage Error}$$
$$= \text{Response Time PC}$$

where PC is the percentage of correct answers.In this case, they suggest mapping RT to the frame count the system chooses. We are building our analysis on the assumption that frames propagate comparably fast from the camera to the initial OPL network. Once the fundamental prerequisite of a strong GPU is provided, according to the OPL design criteria for the optimum fps, the true aim may be attained. These various phases are added together to get the total pipeline latency:

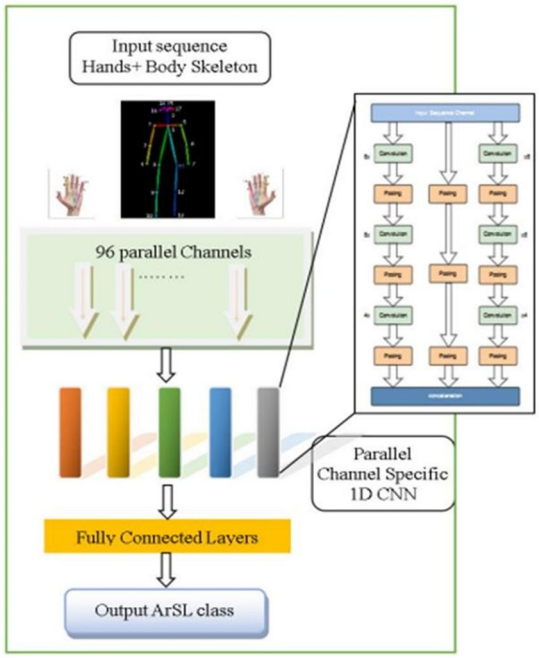
a) Frame capturing

b) Elongation

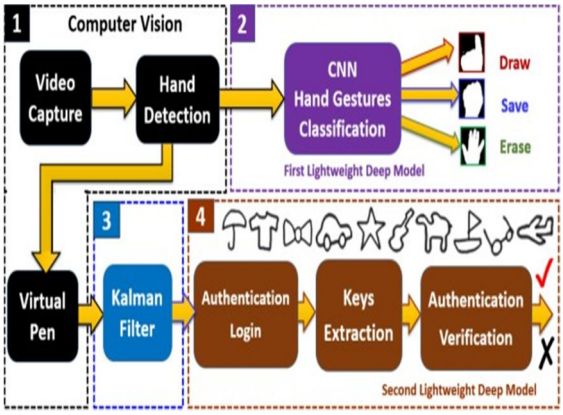
c) OPL

d) Frame queuing for key locations

e) SKN infrastructure sign decision



A Gaussian distribution is a round-the-clock likeliness allocation which is totally narrated by pair of conditions:
the mean (a) and variance (b2).
Gaussian distribution formula is expressed as Equation
f (x,a,b) = 1 b √ 2π e −1 2 (x−a)^ 2 .b ^2
For instance, incase we acquire the Y position of a certain point 30 times/second, and every time it returns a number between 17 and 23.Hence, established on basis of Gaussian distribution
The Kalman filter algorithm recapitulate the forecasting phase and inform phase for every
latest data points, like following phases: Forecast phase: The latest points forecast based on preceding computation point and motion phase, like in Equation
Forecast Point approximation = Preceding Point + Motion Phase
Though, the points and motion phase are established on the basis of Gaussian distribution, which has a mean and variance (error rate). Therefore, the latest point is the grand total of duplet Gaussians, as in Equation: GD(A1, B1) = GD(Ai, Bi) + GD(Am, Bm) = GD(Ai + Am, Bi + Bm)
where Ai and Bi are the mean and variance (error rate) of the preceding point, discreetly. Where Am and Bm are the mean and variance (error rate) of the Motion Phase, discreetly. The Forecast Point is a Gaussian distribution with the mean that equivalent (Ai +Am) and the variance that equivalent (Bi + Bm).[6]
Inform phase: The inform phase is used to extract the point by simply increasing the duo Gaussian distributions, Forecast Point point approximation GD (A1, B1), and the calculation or present position GD (A2, B2). The outcome of the amplification is Gaussian or foremost position approximation GD (A, B), and its mean is the nearest position of the accurate position. Then, the foremost position approximationGD (A, B) is used in the next Forecast Phase. [6]



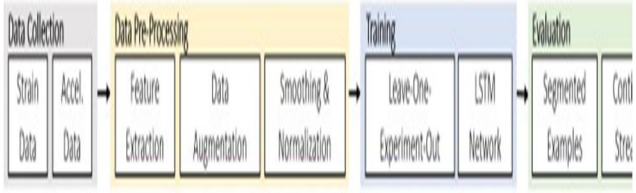
A neural network that can categorize segmented samples or streaming real-time datais created using the pipeline's processing of gathered training data.

- TRAINING DATA COLLECTION
- DATA PROCESSING AND FEATURE

EXTRACTION:

- Segmentation Into LabeledExample.
- Data Augmentation
- Feature Extraction: Smoothingand Normalization

- NEURAL NETWORK TRAINING
- RESULTS AND DISCUSSION[4]

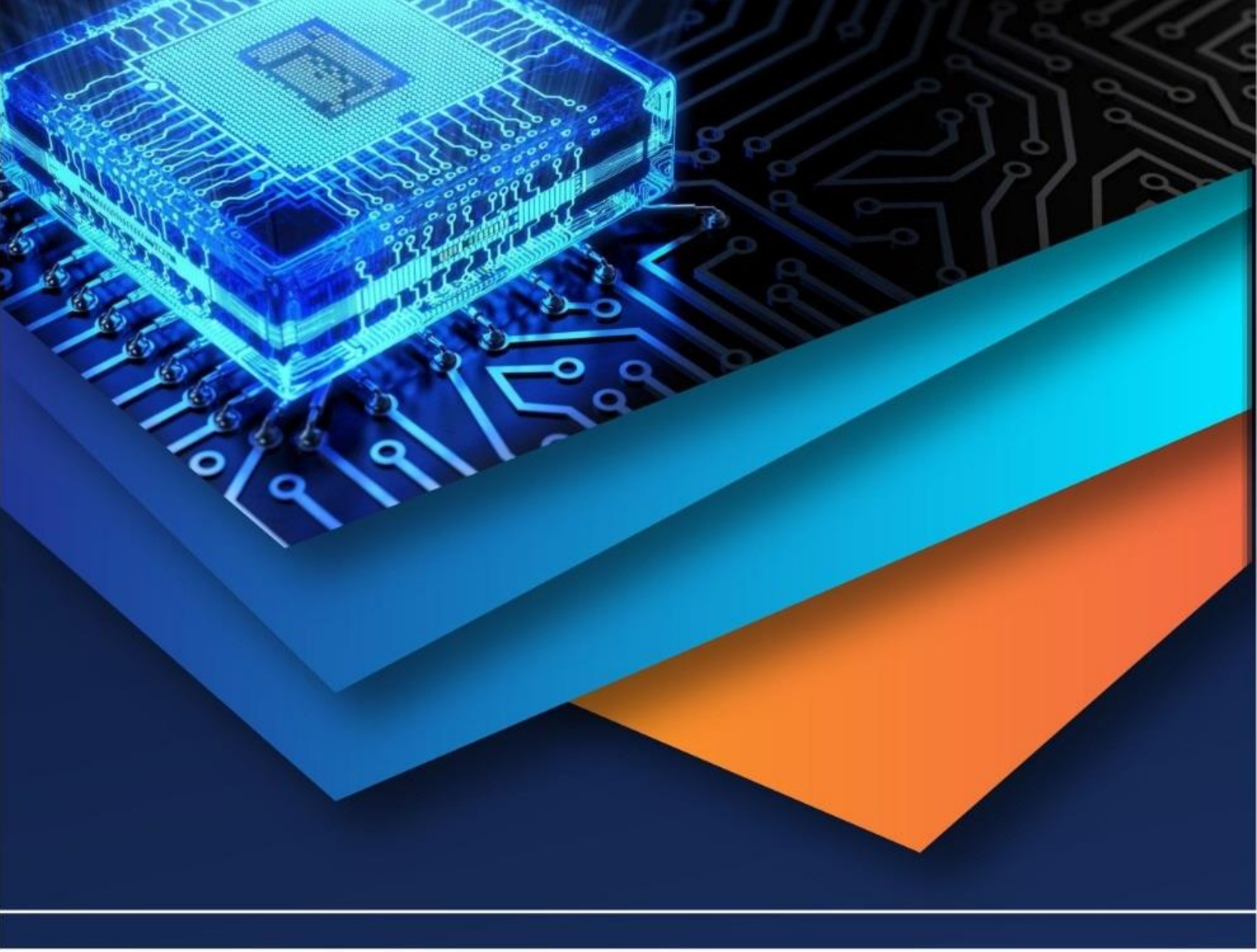


IV. CONCLUSION

Here we discuss various hand gesture recognition algorithms and methods. The use of hand gesture recognition systems is thoughtto lead to more effective and intuitive tools forhuman-computer interaction. Applications span from sign language interpretation to virtual prototyping to medical education. One means of communication for those who are physically disabled, deaf, or dumb is sign language. The aforementioned analysis shows that the field of hand gesture identification has advanced significantly thanks to vision-based hand gesture recognition.

REFERENCES

[1] Li, Min, Zhenjiang Miao, and Cong Ma. "Dance movement learning for labanotation generation based on motion-captured data." IEEE Access 7 (2019): 161561-161572.
[2] Natarajan, B., et al. "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation." IEEE Access 10 (2022): 104358-104374.
[3] Zong, Yan, et al. "Robust Synchronized Data Acquisition for Biometric Authentication." IEEE Transactions onIndustrial Informatics 18.12 (2022): 9072-9082.
[4] DelPreto, Joseph, et al. "A Wearable Smart Glove and Its Application of Pose and Gesture Detection to Sign Language Classification." IEEE Robotics and Automation Letters 7.4 (2022): 10589- 10596.
[5] Alwaely, Basheer, and Charith Abhayaratne. "Graph spectral domain feature learning with application to in-air hand- drawn number and shape recognition." IEEE Access 7 (2019): 159661-159673.
[6] Elshenawy, Abdelghafar R., and Shawkat
[7] K. Guirguis. "On-Air Hand-Drawn Doodles for IoT Devices Authentication During COVID-19." IEEE Access 9 (2021): 161723- 161744.
[8]
[9] Arsalan, Muhammad, Avik Santra, and Vadim Issakov. "RadarSNN: A Resource Efficient Gesture Sensing System Based on mm-Wave Radar." IEEE Transactions on Microwave Theory and Techniques 70.4 (2022): 2451-2461.
[10] Bencherif, Mohamed A., et al. "Arabic sign language recognition system using 2D hands and body skeleton data." IEEE Access 9 (2021): 59612-59627.
[11] Luqman, Hamzah. "An Efficient Two- Stream Network for Isolated Sign Language Recognition Using Accumulative Video Motion." IEEE Access 10 (2022): 93785- 93798.
[12] Xu, Jun, et al. "Robust Hand Gesture Recognition Based on RGB-D Data for Natural Human-Computer Interaction." IEEE Access (2022).
[13] Al-Hammadi, Muneer, et al. "Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation." IEEE Access 8 (2020): 192527-192542.
[14] Lee, Minhyuk, and Joonbum Bae. "Deeplearning based real-time recognition of dynamic finger gestures using a data glove." IEEE Access 8 (2020): 219923-219933.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET49256, entitled

A Review on Dynamic Hand Gesture Recognition Techniques

by

Mownika Reddy K A

after review is found suitable and has been published in Volume 11, Issue II, February 2023

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET49256, entitled

A Review on Dynamic Hand Gesture Recognition Techniques

by

Prajna Harish

after review is found suitable and has been published in Volume 11, Issue II, February 2023

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET49256, entitled

A Review on Dynamic Hand Gesture Recognition Techniques

by

Sambhrama K

after review is found suitable and has been published in Volume 11, Issue II, February 2023

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET

ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISRA Journal Impact Factor: 7.429

45.98 INDEX COPERNICUS

THOMSON REUTERS Researcher ID: N-9681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL SJIF 7.429

Certificate

It is here by certified that the paper ID : IJRASET49256, entitled

A Review on Dynamic Hand Gesture Recognition Techniques

by

Vinayak Bhat

after review is found suitable and has been published in Volume 11, Issue II, February 2023

in

International Journal for Research in Applied Science & Engineering Technology

(International Peer Reviewed and Refereed Journal)

Good luck for your future endeavors

Editor in Chief, iJRASET