G H Patel College of Engineering and Technology

(A Constituent College of CVM University)

V. V. Nagar

COMPUTER ENGINEERING DEPARTMENT

Mini Project Report

on

Quiz System Using Computer Vision

Submitted By

Name of Student: Darshan Solanki

Enrolment Number: 12102040503002

Guided By

Dr. Namrata Dave

MINI PROJECT (102040601)

A.Y. 2022-23 EVEN TERM

CERTIFICATE

This is to certify that the Mini Project Report submitted entitled **Interactive Quiz System Using Computer Vision** has been carried out by **Darshan Solanki** (12102040503002) under guidance of **Dr. Namrata Dave** in partial fulfillment for the Degree of Bachelor of Engineering in Computer Engineering, 6th Semester of G H Patel College of Engineering and Technology, CVM University, Vallabh Vidyanagar during the academic year 2022-23.

Dr. Namrata Dave	Dr. Maulika Patel
Internal Guide	Head of Department

ABSTARCT

This project aims to develop a quiz system that utilizes computer vision technology to enhance the user experience. The proposed system involves capturing images of the quiz taker and their surroundings using a camera, and then processing the images to identify the user and their actions. The system will use machine learning algorithms to recognize hand gestures, enabling the user to interact with the quiz by answering questions and navigating through the quiz using simple hand gestures. Additionally, it stores all the data in a database after attempting the quiz. The proposed quiz system has the potential to provide a more engaging and interactive experience for users while ensuring the integrity and security of the quiz.

Table Of Content

Table Of Figures

Chapter 1

INTRODUCTION

The use of computer vision technology has revolutionized the way we interact with the digital world. One such application of this technology is in the development of quiz systems that use computer vision to provide a more immersive and interactive experience to users. These quiz systems rely on the use of cameras and machine learning algorithms to detect and track user movements, gestures, and responses, providing a more intuitive and engaging experience.

In a typical quiz system that uses computer vision, the user is presented with a set of questions, which they can answer by using their body movements or gestures. The computer vision system tracks the user's movements and responses, and provides feedback in real-time, making the experience more engaging and interactive. This approach not only adds a new dimension of interactivity to quizzes but also provides an opportunity to engage users who may have limited mobility or may find it challenging to interact with traditional quiz systems.

Moreover, computer vision-based quiz systems have several advantages over traditional systems. For example, they are more engaging and interactive, which can lead to increased retention of information and better learning outcomes. They are also more accessible, as they can be used by a wide range of users, including those with disabilities. Furthermore, computer vision-based quiz systems are scalable and can be used in a variety of settings, including classrooms, training programs, and even entertainment venues.

In conclusion, the use of computer vision in quiz systems is an innovative approach to make quizzes more interactive, engaging, and accessible. This technology has the potential to revolutionize the way we learn and engage with information, and we can expect to see more of these systems in the future.

1.1. Problem Statement

The traditional quiz systems have limitations in terms of interactivity and engagement, resulting in decreased learning outcomes and limited accessibility for users with disabilities. To overcome these limitations, there is a need for the development of a quiz system that uses computer vision technology to enhance interactivity, engagement, and accessibility.

1.2. Aim & Objectives

The aim of developing a quiz system using computer vision technology is to create an interactive and engaging learning experience that promotes better learning outcomes and is accessible to all users.

Objectives:

- To develop a computer vision-based quiz system that is capable of tracking user movements and responses.
- To create a user-friendly interface that allows users to interact with the quiz system using gestures and movements.
- To ensure the accuracy and reliability of the computer vision system in detecting and tracking user movements and responses.
- To enhance the engagement of users by providing real-time feedback and interactive elements in the quiz system.
- To improve the accessibility of the quiz system by making it compatible with different devices and accommodating users with different levels of mobility.
- To evaluate the effectiveness of the computer vision-based quiz system in promoting better learning outcomes compared to traditional quiz systems.
- To identify areas for further development and improvement of the computer vision- based quiz system.

Chapter 2

SYSTEM ANALYSIS

2.1. Motivation

The motivation behind the development of a quiz system using computer vision technology is to provide an engaging and interactive learning experience that promotes better retention of information and is accessible to all users. Traditional quiz systems often lack interactivity and engagement, which can result in reduced learning outcomes and limited accessibility for users with disabilities. By incorporating computer vision technology, we can create a more immersive and interactive experience that accommodates users with different levels of mobility and promotes better learning outcomes.

Furthermore, computer vision-based quiz systems have several advantages over traditional systems. They are more engaging and interactive, which can lead to increased retention of information and better learning outcomes. They are also more accessible, as they can be used by a wide range of users, including those with disabilities. Additionally, computer vision technology enables real-time feedback and interactive elements that enhance the engagement of users and promote a deeper understanding of the material.

Overall, the motivation behind the development of a quiz system using computer vision technology is to enhance the learning experience by creating a more immersive and interactive environment that promotes better retention of information and is accessible to all users.

2.2.Literature Review

Quiz systems that use computer vision technology have gained significant attention in recent years due to their potential to provide an engaging and interactive learning experience. In this literature review, we will explore some of the research studies and publications that have investigated the use of computer vision in quiz systems.

A study conducted by He et al. (2017) explored the use of a computer vision-based quiz system to enhance the learning experience of Chinese language learners. The system used Kinect sensors to track user movements and gestures, and provided real-time feedback on the correctness of the user's responses. The study found that the computer vision-based quiz system resulted in higher levels of engagement and satisfaction among the learners compared to traditional paper-based quizzes.

Another study by Zhang et al. (2019) investigated the use of computer vision technology in a mobile-based quiz game. The system used the front-facing camera on the mobile device to track user movements and gestures, and provided real-time feedback on the user's responses. The study found that the computer vision-based quiz game resulted in higher levels of engagement and motivation among the players compared to traditional quiz games.

In a study by Lee et al. (2020), a computer vision-based quiz system was developed for use in a classroom setting. The system used a camera to track user movements and gestures, and provided real-time feedback on the user's responses. The study found that the computer vision-based quiz system resulted in higher levels of engagement and participation among the students, leading to better learning outcomes.

A review paper by Chen et al. (2020) summarized the research studies that have investigated the use of computer vision in educational applications. The review found that computer vision technology has the potential to enhance the learning experience by providing a more interactive and engaging environment. However, the review also identified several challenges that need to be addressed, such as ensuring the accuracy and reliability of the computer vision system, and accommodating users with different levels of mobility.

In conclusion, the literature suggests that computer vision technology has the potential to revolutionize the way we learn and interact with information. The use of computer vision in quiz systems has been shown to enhance engagement, participation, and motivation among users, leading to better learning outcomes. However, there are still challenges that need to be addressed to ensure the effectiveness and accessibility of these systems.

Chapter 3

PROCESS METHODOLOGY

3.1.Feature Selection and Extraction

Feature selection and extraction are important steps in developing a quiz system using computer vision technology. These steps involve identifying the relevant features and extracting them from the images or videos of the user.

The following are some common features used in computer vision-based quiz systems:

- Hand and Body Gestures: Hand and body gestures are commonly used features in quiz systems. The computer vision algorithms can detect and track the movements of the user's hands and body to recognize gestures and responses.

- Object Recognition: Object recognition can be used to detect and track objects that the user interacts with during the quiz. For example, the computer vision algorithms can recognize when the user is pointing to a specific object in the image or video.

- Once the relevant features have been identified, feature extraction techniques such as Principal Component Analysis (PCA) or Local Binary Patterns (LBP) can be used to extract these features from the images or videos of the user. These extracted features can then be used as inputs to the machine learning algorithms that classify the user's responses to quiz questions.

3.2.Hand Recognition In System

Gesture recognition is an active research field in Human-Computer Interaction technology. It has many applications in virtual environment control and sign language translation, robot control, or music creation. In this machine learning project on Hand Gesture Recognition, we are going to make a real-time Hand Gesture Recognizer using the MediaPipe framework and Tensorflow in OpenCV and Python.

OpenCV is a real-time Computer vision and image-processing framework built on C/C++. But we’ll use it on python via the OpenCV-python package.

3.2.1. What is Mediapipe

MediaPipe is a customizable machine learning solutions framework developed by Google. It is an open-source and cross-platform framework, and it is very lightweight. MediaPipe comes with some pre-trained ML solutions such as face detection, pose estimation, hand recognition, object detection, etc.

3.2.2. What is Tensorflow

TensorFlow is an open-source library for machine learning and deep learning developed by the Google brains team. It can be used across a range of tasks but has a particular focus on deep neural networks.

Neural Networks are also known as artificial neural networks. It is a subset of machine learning and the heart of deep learning algorithms. The concept of Neural networks is inspired by the human brain. It mimics the way that biological neurons send signals to one another. Neural networks are composed of node layers, containing an input layer, one or more hidden layers, and an output layer.

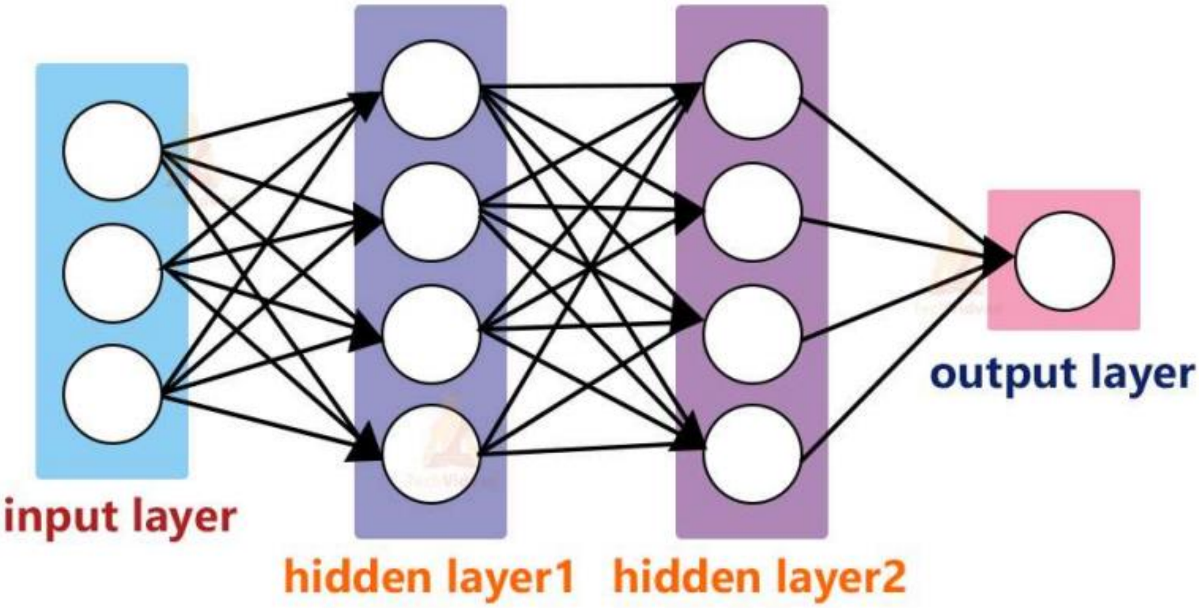

Figure 1: TensorFlow Structure

We'll first use MediaPipe to recognize the hand and the hand key points. MediaPipe returns a total of 21 key points for each detected hand.

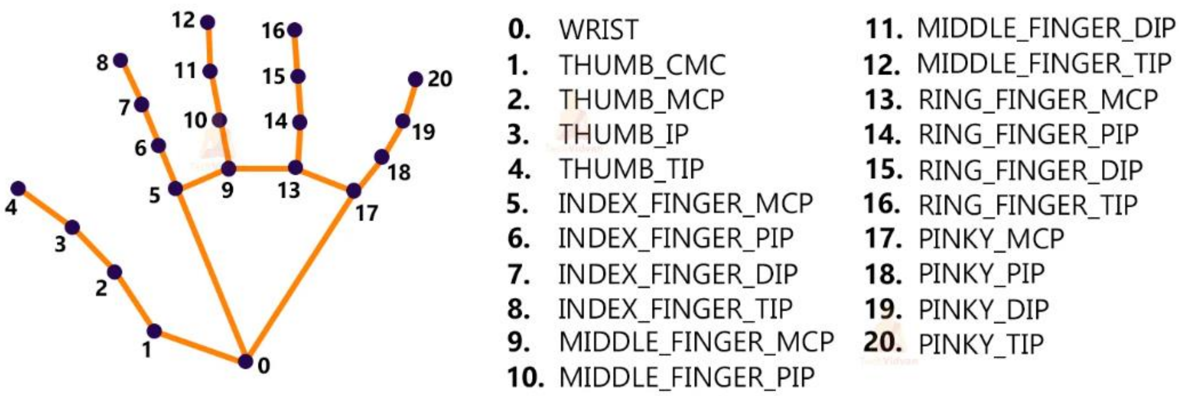

Figure 2:Landmark of keypoints

These key points will be fed into a pre-trained gesture recognizer network to recognize the hand pose.

3.3. How does computer read image in system

Consider the below image:

We can figure out that it is an image of the New York Skyline. But, can a computer find this out all on its own? The answer is no!

The computer reads any image as a range of values between 0 and 255.

For any color image, there are 3 primary channels – Red, green and blue. How it works is pretty simple.

A matrix is formed for every primary color and later these matrices combine to provide a Pixel value for the individual R, G, B colors.

Each element of the matrices provide data pertaining to the intensity of brightness of the pixel.

Consider the following image:

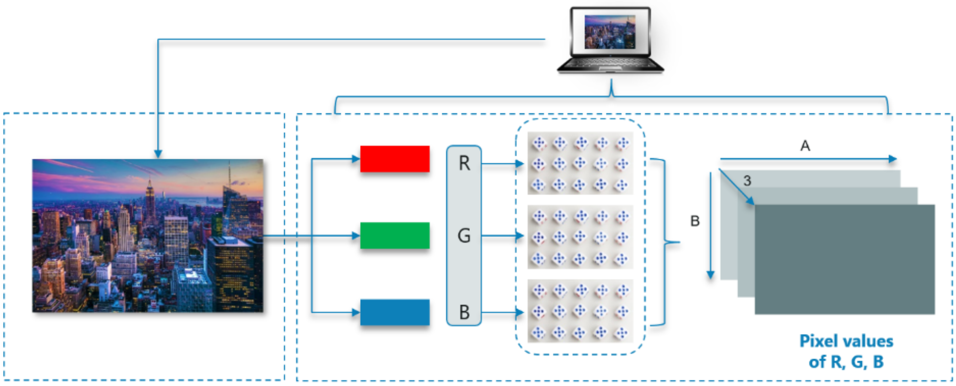

Figure 3: How computer read image

As shown, the size of the image here can be calculated as B x A x 3.

Note: For a black-white image, there is only one single channel.

Next up on this OpenCV Python Tutorial blog, let us look at what OpenCV actually is.

3.4. Analysis of working model

There are many object detection algorithms as follow:

- Fast R-CNN. ...
- Faster R-CNN. ...
- Region-based Convolutional Neural Networks (R-CNN) ...
- Region-based Fully Convolutional Network (R-FCN) ...
- Single Shot Detector (SSD) ...
- YOLO (You Only Look Once) ...
- RetinaNet. ...
- Spatial Pyramid Pooling (SPP-net)

SSD is designed for object detection in real-time. Faster R-CNN uses a region proposal network to create boundary boxes and utilizes those boxes to classify objects. While it is considered the start-of-the-art in accuracy, the whole process runs at 7 frames per second. Far below what real-time processing needs. SSD speeds up the process by eliminating the need for the region proposal network. To recover the drop in accuracy, SSD applies a few improvements including multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN's accuracy using lower resolution images, which further pushes the speed higher. According to the following comparison, it achieves the real-time processing speed and even beats the accuracy of the Faster R-CNN. (Accuracy is measured as the mean average precision mAP: the precision of the predictions.)

System	VOC2007 test *mAP*	FPS (Titan X)	Number of Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	~6000	~1000 x 600
YOLO (customized)	63.4	45	98	448 x 448
SSD300* (VGG16)	77.2	46	8732	300 x 300
SSD512* (VGG16)	**79.8**	19	24564	512 x 512

Figure 4: Algorithm accuracy

3.4.1. Single Shot Detection(SSD)

The SSD object detection composes of 2 parts:

- Extract feature maps, and
- Apply convolution filters to detect objects.

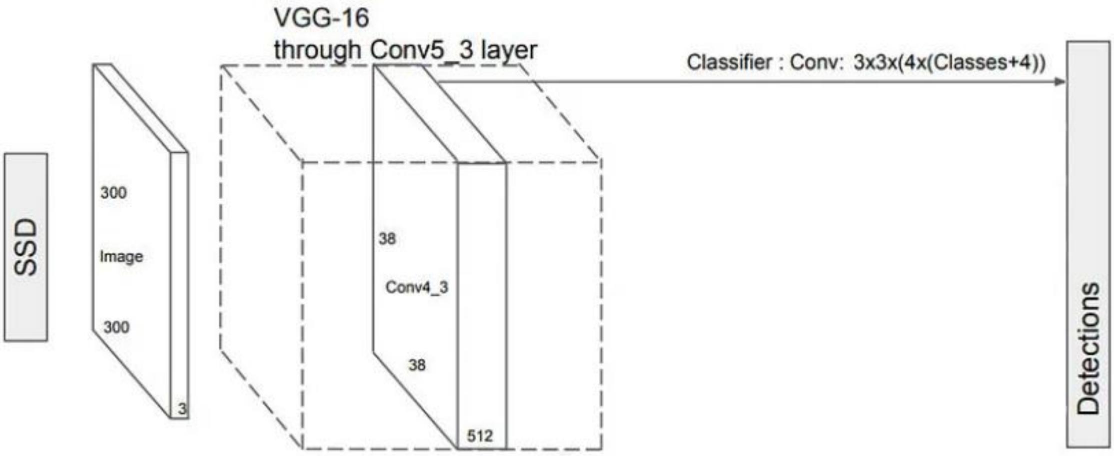

Figure 5: SSD architecture

SSD uses VGG16 to extract feature maps. Then it detects objects using the Conv4_3 layer. For illustration, we draw the Conv4_3 to be 8 × 8 spatially (it should be 38 × 3 8). For each cell (also called location), it makes 4 object predictions.

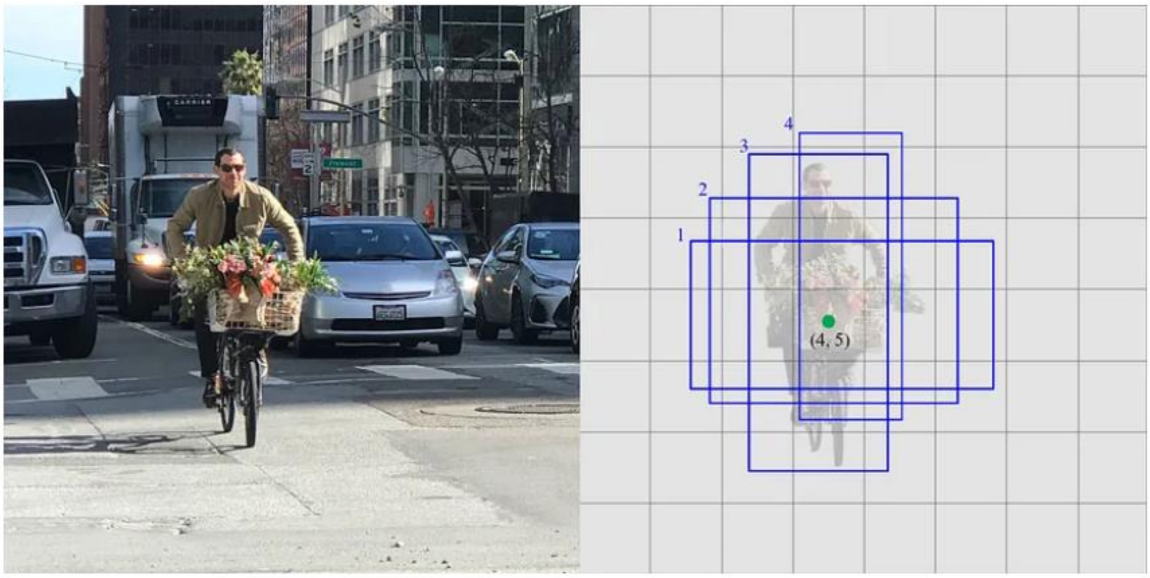

Figure 6: Predict at each cell

Each prediction composes of a boundary box and 21 scores for each class (one extra class for no object), and we pick the highest score as the class for the bounded object. Conv4_3 makes a total of 38 × 38 × 4 predictions: four predictions per cell regardless of the depth of the feature maps. As expected, many predictions contain no object. SSD reserves a class “0” to indicate it has no objects.

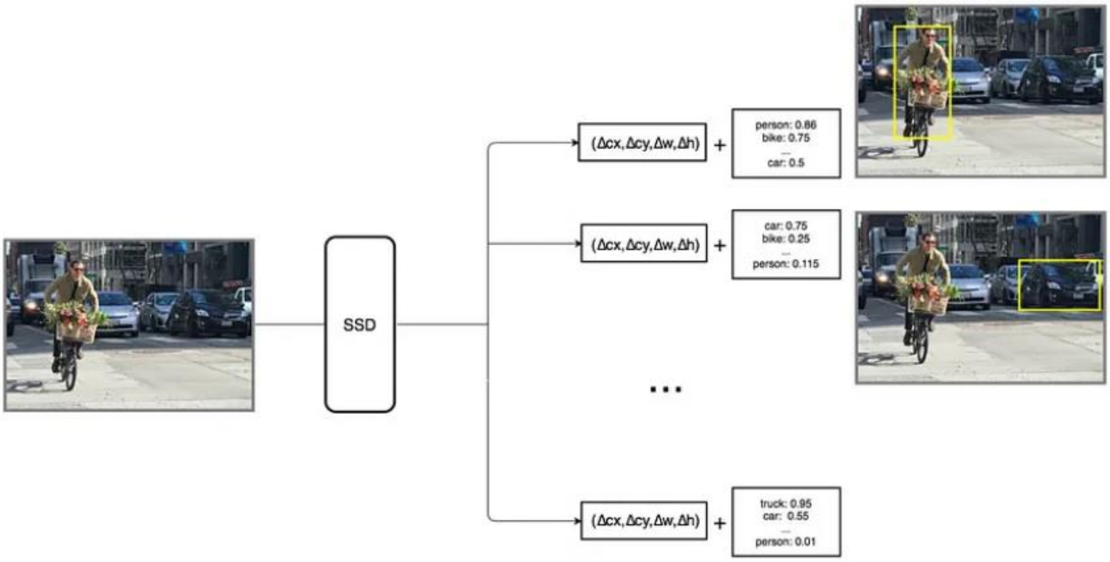

Figure 7: Predict demo

SSD does not use a delegated region proposal network. Instead, it resolves to a very simple method. It computes both the location and class scores using small convolution filters. After extracting the feature maps, SSD applies 3 × 3 convolution filters for each cell to make predictions. (These filters compute the results just like the regular CNN filters.) Each filter outputs 25 channels: 21 scores for each class plus one boundary box (detail on the boundary box later).

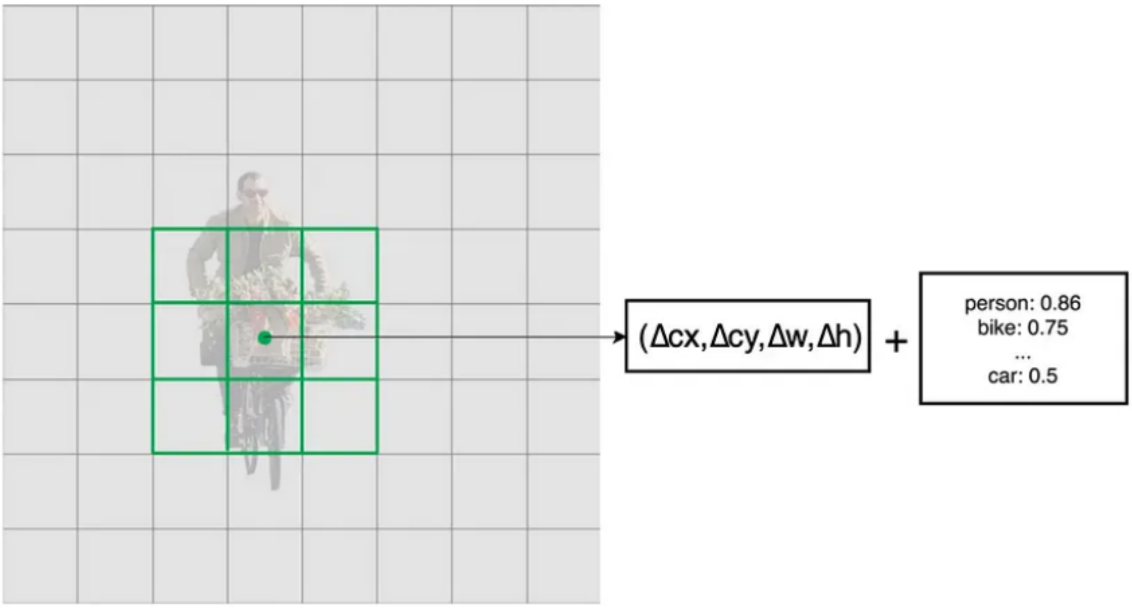

Figure 8: Predict output channel

For example, in Conv4_3, we apply four 3 × 3 filters to map 512 input channels to 25 output channels.

$$(38 \times 38 \times 512) \stackrel{(4\times3\times3\times512\times(21+4))}{\longrightarrow} (38 \times 38 \times 4 \times (21+4))$$

Default boundary boxes are chosen manually. SSD defines a scale value for each feature map layer. Starting from the left, Conv4_3 detects objects at the smallest scale 0.2 (or 0.1 sometimes), and then increases linearly to the rightmost layer at a scale of 0.9. Combining the scale value with the target aspect ratios, we compute the width and the height of the default boxes. For layers making 6 predictions, SSD starts with 5 target aspect ratios: 1, 2, 3, 1/2, and 1/3. Then the width and the height of the default boxes are calculated as:

$$w = scale \cdot \sqrt{\text{aspect ratio}}$$
$$h = \frac{scale}{\sqrt{\text{aspect ratio}}}$$

Then SSD adds an extra default box with scale:

$$scale = \sqrt{\text{scale} \cdot \text{scale at next level}}$$

and aspect ratio = 1.

Chapter 4

IMPLIMENTATION

4.1. Working Flow Diagram

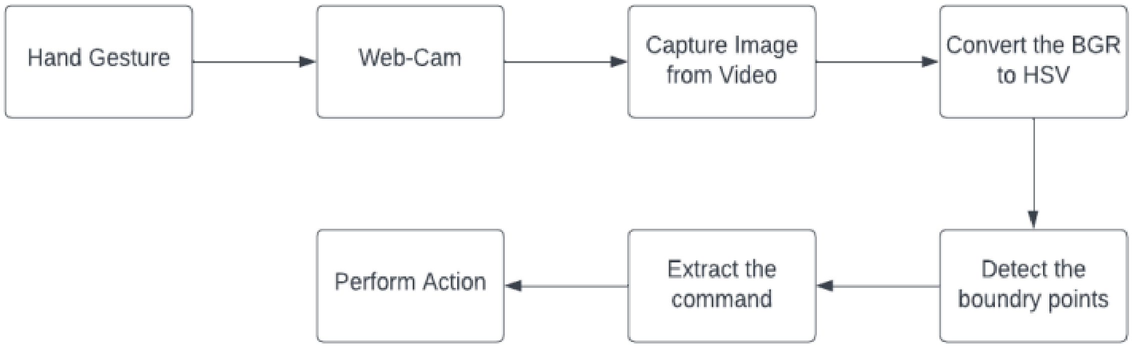

Figure 9: workflow diagram

The workflow diagram shows the step-by-step process of how the Quiz System using computer vision project works. The user starts by launching the application after entering the enrollment number the quiz will be start. The application then captures live video streams from the webcam and passes them to the Hand Detection Module, which uses MediaPipe and CVzone to detect and track hand movements. User can select the option by virtual click on it.

The User Interface Module displays the live video stream and any recognized hand gestures to the user, as well as the current status of the application. The Action Mapping Module defines the mapping between recognized gestures and actions and provides options for the user to customize these mappings. Finally the result of quiz will be store in database.

4.2. SystemFlow Diagram

The system flow diagram shows the flow of data and control between two modules of the quiz system using computer vision project. The diagram shows how video streaming module capture the live video using webcamp and pass the data it to Hand Detection Module for processing.

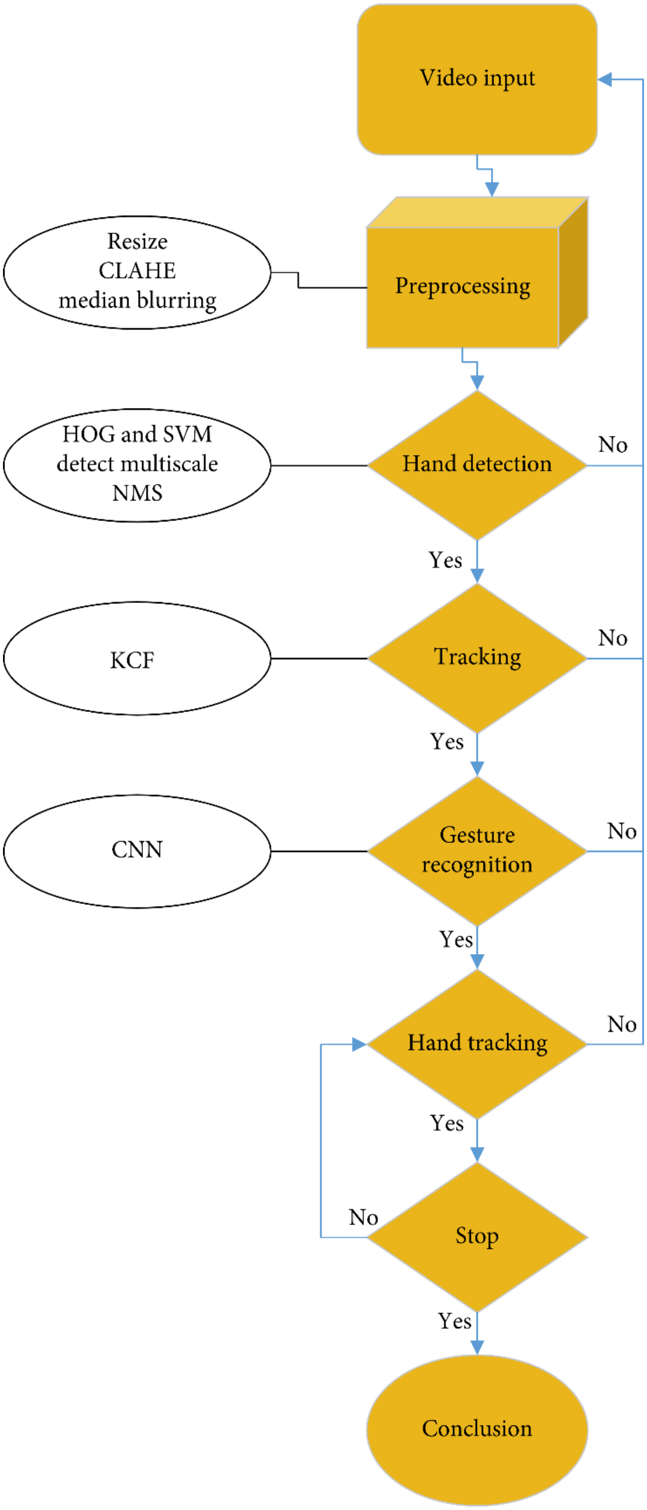

Figure 10: SystemFlow Diagram

The Hand Detection module uses the Mediapipe and CVzone to detect and track the hand movements and passed the data to the Gestures Control Module. The Gestures Control Module recognize the hand gestures and based on the particular gesture it take the action.

4.3.Module Specification

This project consist of several modules that work together to detect and recognize hand gestures in real-time and take quiz based on user action.

The Video Streaming Module captures live video streams from the webcam and provides the video frames to the Hand Detection Module for further processing. The Hand Detection Module processes the video frames using MediaPipe and CVZone libraries to detect and track hand movements. It identifies the hand landmarks and their positions in the image and returns this information to the Gesture Control Module.

The Gesture Control Module takes the hand landmark data from the Hand Detection Module and uses it to recognize hand gestures using a set of predefined gesture recognition rules. It select the option which user choose by triggering the click event and show the corresponding next state to user.

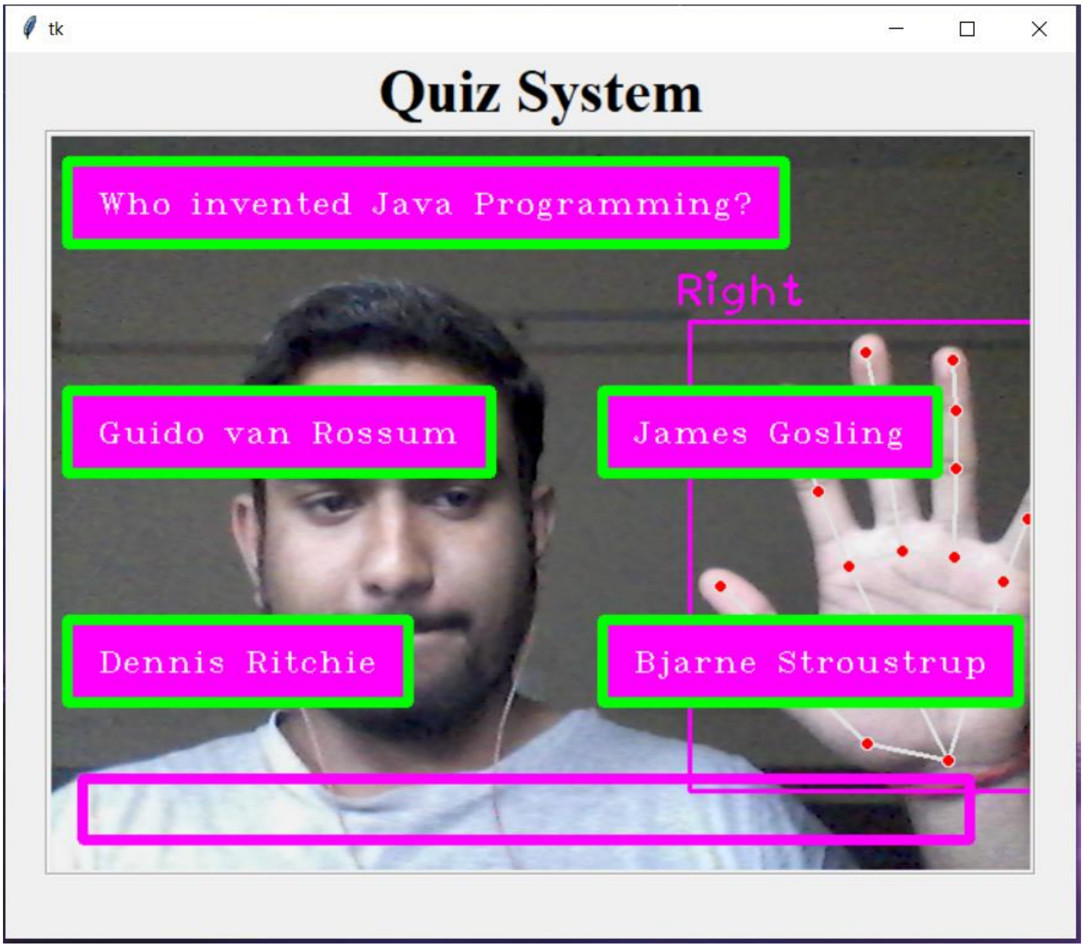

Figure 11: Output 1.0

The user interface module displays the live video stream and recognized hand gestures to the user. By click on the virtual click with fingers, we can select the option to the particular question. We can also modify the answer of previous question by selecting that question and modify the answer.

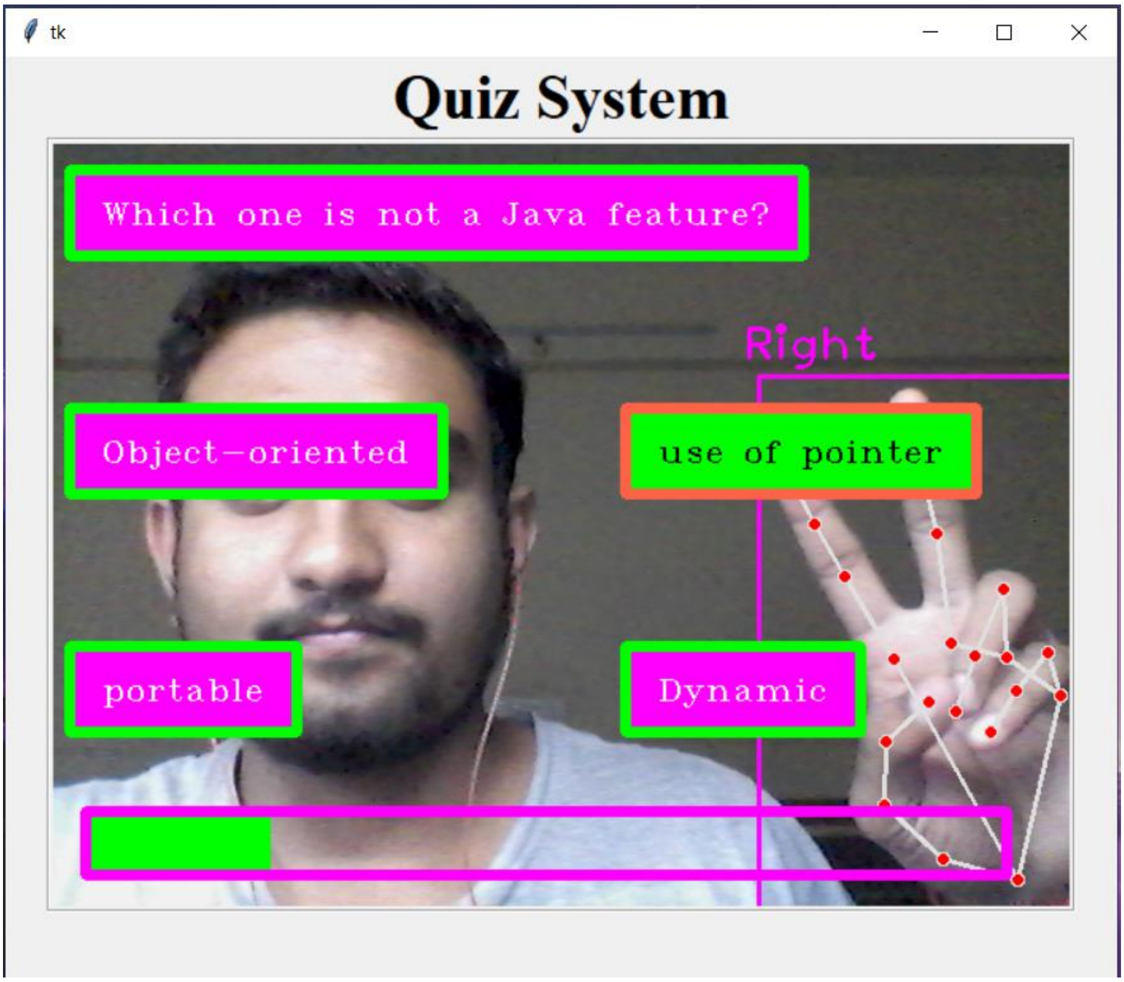

Figure 12: Output 1.1

After completing the quiz it show the result in percentage.

4.4.Result

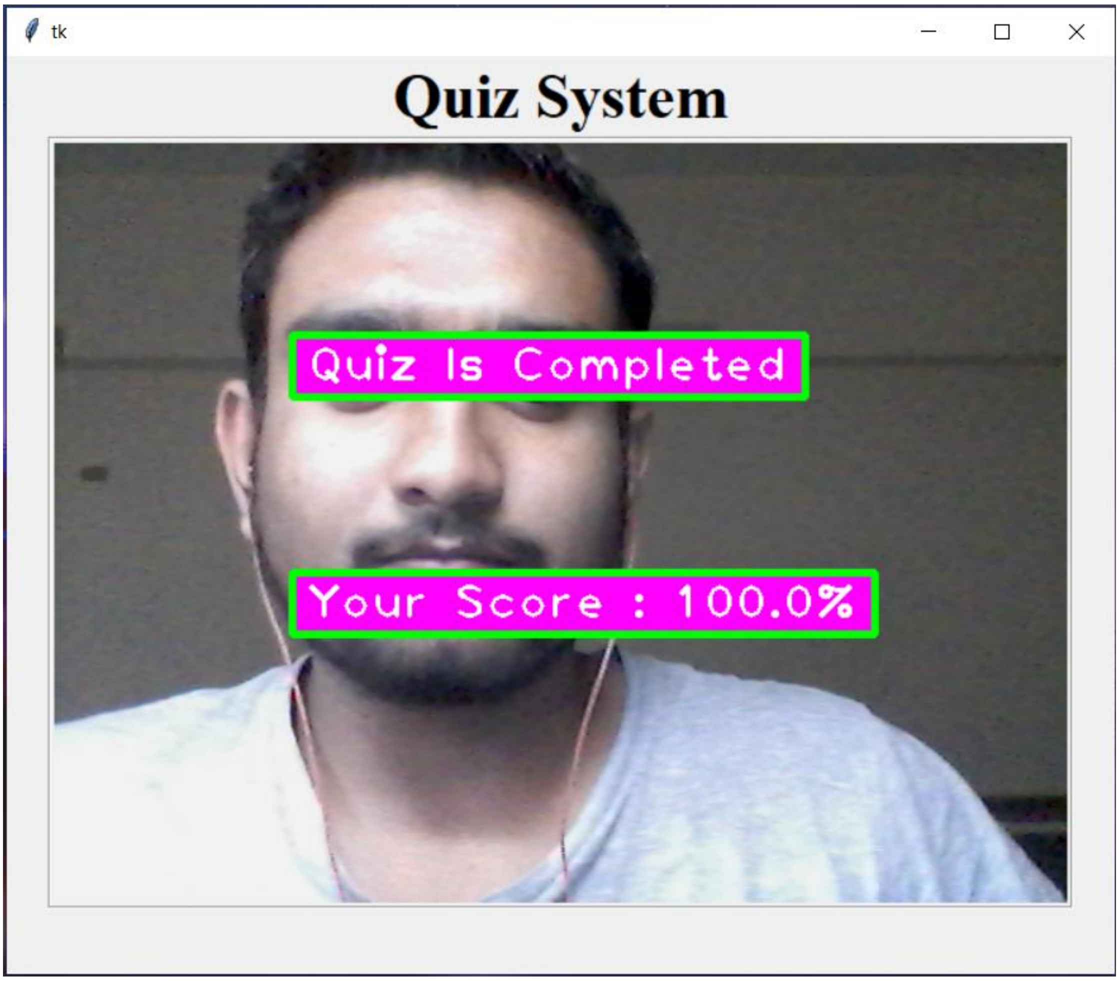

Figure 13: Output 1.2

The model is trained using SGD with an initial learning rate of 0.001, 0.9 momentum, 0.0005 weight decay, I work on this module with different backgrounds to and the result of the modules is around 74-76% for hand detection.

4.5.Future Scope

The future scope for a quiz system using computer vision technology is immense. Here are a few potential avenues for development:

Advanced facial recognition: The system could be enhanced to recognize individual faces, and provide personalized quizzes based on previous performances. This could help teachers and trainers to monitor the progress of their students, and provide tailored quizzes to help them improve their weak areas.

Gesture and expression recognition: By using computer vision to detect and interpret hand gestures and facial expressions, the system could assess the level of engagement and attention of the students during the quiz. This could help to identify when a student is struggling, or when they are not paying attention, and provide appropriate feedback or support.

Augmented reality: Computer vision technology could be combined with augmented reality to create an immersive and interactive quiz experience. For example, students could use their smartphones or tablets to scan a QR code on the quiz paper, which would launch an augmented reality quiz that they can interact with in real time.

Automated grading: With computer vision technology, the system could automatically grade the quizzes, eliminating the need for human intervention. This could save time and effort for teachers and trainers, allowing them to focus on other aspects of teaching.

Adaptive difficulty: The system could use computer vision technology to analyze the students' responses, and adapt the difficulty of the questions based on their performance. This would help to ensure that the quiz is challenging enough to keep the students engaged, but not too difficult to discourage them.

Overall, the future scope for a quiz system using computer vision is vast, and there are numerous ways in which this technology could be utilized to enhance the learning experience. As computer vision technology continues to evolve, it is likely that we will see even more advanced and innovative applications in the field of education.

CONCLUSION

In conclusion, the integration of computer vision technology in a quiz system has a lot of potential for improving the learning experience. It can help personalize quizzes, monitor student engagement, and provide automated grading, among other benefits. As computer vision technology advances, there will likely be even more innovative ways to use it in the field of education. The possibilities are vast, and it will be exciting to see how this technology develops in the coming years.

REFERENCES

- https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06
- https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html?m=1
- https://developers.arcgis.com/python/guide/how-ssd-works/#:~:text=SSD%20uses%20a%20matching%20phase,object's%20class%20and%20its%20location
- https://towardsdatascience.com/review-ssd-single-shot-detector-object-detection-851a94607d11
- https://www.ibm.com/in-en/topics/computer-vision
- https://techvidvan.com/tutorials/hand-gesture-recognition-tensorflow-opencv/
- https://opencv.org/about/
- https://www.youtube.com/channel/UCYUjYU5FveRAscQ8V21w81A
- https://www.edureka.co/blog/python-opencv-tutorial/#z2
- https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e#:~:text=Computer%20vision%20also%20plays%20an,with%20databases%20of%20face%20profiles.
- https://www.hindawi.com/journals/jr/2021/3986497/