VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

DIGITAL ASSIGNMENT -1

CSE1901: Technical Answers for Real Problems (TARP)

Submitted to: Dr. Manoov R

REGISTRATION NUMBER	NAME
20BCI0090	Vandit Gabani
20BCI0128	Aditi Nitin Tagalpallewar
20BCI0176	Yash Bobde
20BCI0271	Harsh Rajpal
20BCE2759	Payal Maheshwari
20BCI0138	Bagade Shaunak Rahul
20BCI0169	Konark Patel
20BCI0159	Nikhil Harshwardhan

February 2023

I. TEAM INTERACTION

The division of the work is done in the following order amongst all the team members of the project:

Team Members of this project are:

1. Vandit Gabani
2. Aditi Nitin Tagalpallewar
3. Yash Bobde
4. Harsh Rajpal
5. Payal Maheshwari
6. Bagade Shaunak Rahul
7. Konark Patel
8. Nikhil Harshwardhan

Our project will be divided into three parts:

Part 1: Preprocess the pose classification training data into a CSV file, specifying the landmarks (body key points) and ground truth pose labels recognized by the MoveNet model.

The preprocessing of pose classification training data into a CSV file involves several steps, which will be divided amongst the team members.

We will be working together to ensure that the annotated data is accurate and consistent, and to validate the extracted key points and the created CSV file. We will also consider data augmentation techniques to increase the size of the training data, such as flipping, rotation, or scaling.

Overall, the preprocessing of pose classification training data into a CSV file is a crucial step in training a pose estimation model and will be completed by the entire team.

Part 2: Build and train a pose classification model that takes landmark coordinates from a CSV file as input and outputs predicted labels.

Building and training a pose classification model involves several steps, which will be divided amongst the team members.

Everyone will work together to implement and train the model, ensuring that the model architecture and training parameters are correctly specified and will

collaborate to evaluate the model's performance on the validation set and make any necessary adjustments to improve its accuracy.

Overall, building and training a pose classification model requires a strong understanding of deep learning and computer vision, as well as a well-structured approach to model selection, implementation, and training.

Part 3: Convert the pose classification model to TFLite.

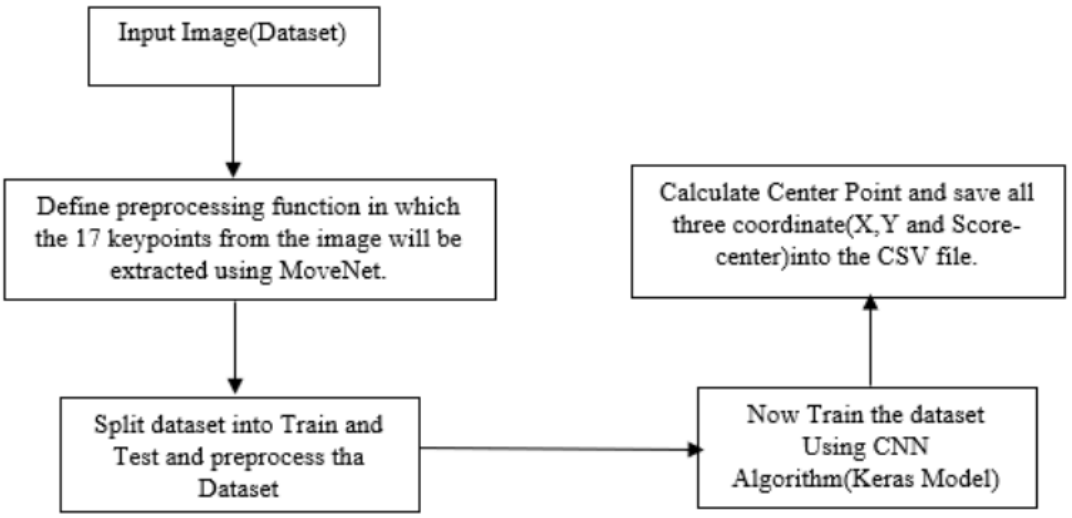
Converting a pose classification model to TFLite involves several steps, which can be divided into tasks amongst all the team members.

Everyone will work together to ensure that the conversion process is carried out smoothly and that the TFLite model is optimised and validated correctly.

Overview: As there is no Dataset available on the internet so we are going to use our own dataset which consist of 7 different traffic police /pose images.Our dataset consist of almost around 5000-6000 different images of the shortlisted important 7 traffic poses.

<u>Movenet Model:</u>

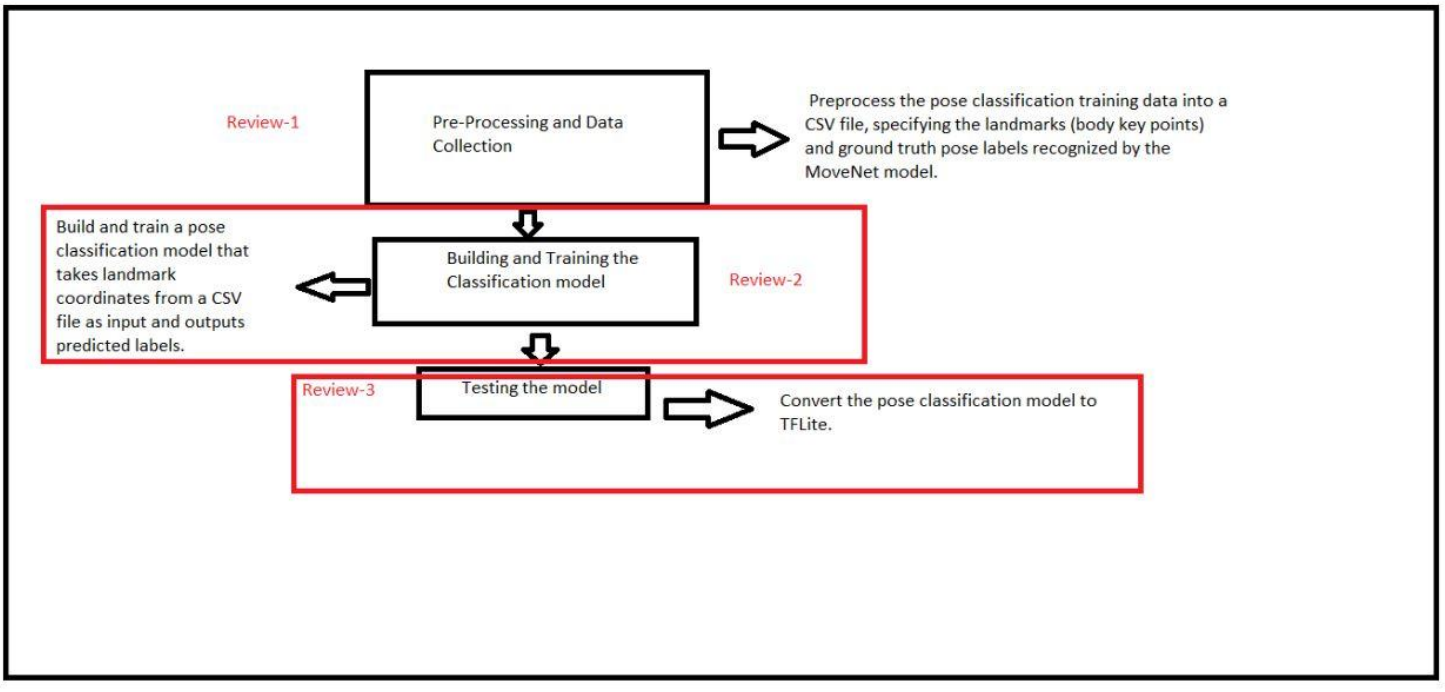
Training Dataset:



Timeline - WorkFlow

Tasks	Review - I			Review - II			Review - III	
Problem Statement Identification								
Project Ideation & Planning								
Risk & Requirment Analysis								
Dataset Collection								
Dataset Filtering and Creation								
Model Designing								
Training of Model								
Development of Model								
Testing								
Deployment								

Project WorkFlow



Abstract

Gesture understanding is one of the most challenging problems in computer vision. Among them, traffic police hand signal recognition requires the consideration of speed and the validity of the commanding signal. The lack of available datasets is also a serious problem. Most classifiers approach these problems using the skeletons of target actors in an image. Extracting the three-dimensional coordinates of skeletons is simplified when depth information accompanies the images. However, depth cameras cost significantly more than RGB cameras. Furthermore, the extraction of the skeleton needs to be performed in prior. Here, we show a hand signal detection algorithm without skeletons. Instead of skeletons, we use simple object detectors trained to acquire hand directions.

In recent years, self-driving cars have gradually entered people's field of vision. Therefore, driverless cars must be able to not only recognize traffic lights but also quickly and correctly respond to and process traffic police's flexible gestures. Thus, traffic police gesture recognition is crucial in driver assistance systems and intelligent vehicles. Traffic police gesture recognition is important in automatic driving. Most existing traffic police gesture recognition methods extract pixel-level features from RGB images which are uninterpretable because of a lack of gesture skeleton features and may result in inaccurate recognition due to background noise.

There are some object detection algorithms available which can detect objects like car,tree,person ,vehicle , bicycle , animal etc – (YOLO).It can not able to detect Any traffic police hand gesture.So In this project we are going to Use CNN algorithm(Deep Learning) to detect traffic police hand gesture.There are no dataset available so we wil try to make our own dataset.it will be developed on mediapipe.

Introduction

Today artificial intelligent used in many real time application and there are numerous developments in deep learning techniques implemented on the area of computer vision which has grown immensely in the field of: Video surveillance, Industrial automation, Self-driving vehicle,military, medical industry etc these development accomplished excellent results. Using this process object can be localized, predicted and classified based on the object that is detected.In daily traffic, traffic signals are important for ensuring the smooth flow of road traffic and increasing roadway traffic security. Traffic signals include not only signal lamps, signs, and markings but also traffic police commands. In the event of special situations, such as traffic light failure, bad weather, traffic congestion, and so on, traffic police typically control traffic and guide drivers using command gestures. In recent years, self-driving cars have gradually entered people's field of vision.

However, current traffic police gesture recognition methods pose certain difficulties, and the recognition task generally faces two challenges. First, most existing traffic police gesture recognition methods extract pixel-level features from RGB images which are uninterpretable because of the lack of gesture skeleton features and may result in inaccurate recognition due to background noise. Appropriate and effective features representing traffic police gestures should be chosen and extracted. However, traffic police typically work in complex and unpredictable environments, which can introduce interference and render features uninterpretable.One method is a gesture skeleton extractor (GSE),which can extract and improve interpretable skeleton coordinate information. Compared with extracted pixel-level features, skeleton information can eliminate background interference and make features interpretable through coordinates and the proposed attention mechanism. Second, existing deep learning methods are not suitable for handling gesture skeleton features. These methods ignore the inevitable connection between skeleton joint coordinate feature and gestures. Several works extracted traffic police skeleton data and proved that this method is effective.

However, some problems exist in prior works. For example, previous studies generally relied on handcrafted components or pixel-level information to extract skeleton features. This method cannot determine the relationship between skeleton joint coordinates and gestures, misses interpretable topologic features, and demonstrates weak generalization capability and poor recognition performance self-driving vehicles must follow the road traffic law, they must also be able to understand the hand signals from the traffic controller. Because self-driving cars move at high speeds, the recognition of hand signals must be performed in real time. Many methods that rely on the computation of skeletons require high computational load and are not suitable for real-time processing. The use of depth sensors can reduce this computational load, but this approach requires expensive devices.

Previous works have employed skeleton-based action recognition. Input videos were processed for skeletons that identify the joints and limbs. Subsequently, the movement of the joints was applied to identify hand signals. However, these methods required the preprocessing of video

streams to extract skeletons, and this extra burden reduced the overall processing time. Furthermore, previous works were applied to videos taken indoors or videos with a limited number of backgrounds. Because we cannot expect that the recognition of hand signals will be con- ducted in a controlled environment, these datasets are not suit- able to generalize the trained neural network to real-world problems.The following three critical aspects need to be considered to understand hand signals. First, it is essential to distinguish between police officers giving appropriate hand signals and those not delivering meaningful hand signals. The detection accuracy is significantly affected by the ability to distinguish between situations in which signals are given and situations when no intentional signals are made. Second, it is vital to know the intended designation of a hand signal. To avoid classification failure, the classifier must understand whether the police officer is giving the signal to them or to another driver in other directions. Last, the classifier must be able to infer continuous changes in hand motions.

Literature Survey

Sr.No	Title/Author	Techniques	Limitation/Future Work
1.	**Object Detection in Self Driving Cars Using Deep Learning** **Author :** Prajwal P, Prajwal D, Harish D H, Gajanana R, Jayasri B S and S. Lokesh (IEEE - 2021)	• Take video input form the camera into car and detect object. • SSD(Single Shot MultiBox Detector) • DNN(Deep Neural network)	• It can only able to detect car,person ,animal,trees,Bus,divider,bicycle etc.
2.	**Traffic Police Gesture Recognition Based on Gesture Skeleton Extractor and Multichannel Dilated Graph Convolution Network** **Author :** Xin Xiong , Haoyuan Wu , Weidong Min , Jianqiang Xu , Qiyan Fu and Chunjiang Peng (IEEE - 2021)	• GSE - extracts traffic police skeleton seq. From a video.(skeleton coordinate) • MD-GCN - take gesture skeleton seq. as input and construct a graph convolution.	• Due to the differences and changes in the angle of view, the “left turn waiting” might be misclassified as “stop” and “slow down” might be misclassified as “left turn”
3.	**Pothole and Object Detection for an Autonomous Vehicle Using YOLO** **Author :** Kavitha R , Nivetha S (IEEE - 2020)	• YOLOv3 • Camera capture the object as the input image.	• Detect the object classes like: car, person, truck, bus, pothole, wetland, trafficlight, motorcycle.
4.	**The Real-Time Detection of Traffic Participants Using YOLO Algorithm** **Author :** Aleksa Ćorović, Velibor Ilić, Siniša Đurić, Mališa Marijan, and Bogdan Pavković (IEEE - 2022)	• YOLO • First extract single image from video stream and resized.. • Image goes to CNN and bounding box is o/p.	• It can only detect object like car, truck, pedestrian, traffic signs, and lights .
5.	**Object Detection for Autonomous Vehicle using Single Camera with YOLOv4 and Mapping Algorithm** **Author :** Mochammad Sahal,Ade Oktavianus Kurniawan (IEEE - 2021)	• YOLOv4 with CSPDarknet-53. • Mapping algorithm for location.	• It can only able to detect object like animal,people,tree,Vehicle , tv.

Sr.No	Title / Author	Techniques	Limitations / Future Work
6.	**On-road object detection using Deep Neural Network** **Author :** Huieun Kim, Youngwan Lee, Byeounghak Yim, Eunsoo Park, Hakil Kim (IEEE - 2018)	• SSD YOLO • RL(Reinforcement Learning)	• If data is obtained from traditional methods such as loop detectors, they will not provide accurate on-time predictions. • lack of clear policies, resistance to adopting new technologies. • Lack of the establishment or ethical regulations.
7.	**American Sign Language Recognition using Convolutional Neural Network** **Author:** Sadaf Ikram,Namrata Dhanda (IEEE – 2021)	• Deep Learning (Keras Model)	• we can also use text to speech to hear what the character, word or sentence is showing on the output screen.
8.	**Application of Artificial Intelligence for Traffic Data Analysis, Simulations and Adaptation** **Author :** Daniela Koltovska Nechoska , Renata Petrevska Nechkoska and Renata Duma (ICEST- 22)	• ANN(Artificial Neural Network) • FL(Fuzzy Logic) • RL(Reinforcement Learning)	• Focused their analysis on the current transport fields that benefit from AI-based technologies and especially on automated traffic data collection.
9.	**Traffic control hand signal recognition using recurrent neural networks** **Author :** Taeseung Baek and Yong-Gu Lee (Journal of Computational Design and Engineering, 2022)	• RNN • The police officer is localized, and the pose of the arm is detected. • The sequence generator concatenated the directions of the poses into a sequence and sent to the RNN for classification. • A stream of flags denoting whether the gazing direction of the traffic controller is facing toward the camera is also generated. The two sequences, are compared to classify the hand signal.	• various adverse weather conditions, such as fog, rain, and snow, can degrade the image quality.
10.	**Prediction of Metacarpophalangeal Joint Angles and Classification of Hand Configurations Based on Ultrasound Imaging of the Forearm** **Author:** Keshav Bimbraw, Christopher J. Nycz, Matthew J. Schueler, Ziming Zhang and Haichong K. Zhang (IEEE - 2022)	• Hand Configuration Classification • MCP Joint Angle Estimation • SVC and CNN Model	• It can inspire research in the promising domain of utilizing ultrasound for predicting both continuous and discrete hand movements,which can be useful for intuitive and adaptable control of physical robots and non-physical digital and AR/VR interfaces.

Sr.No	Title / Author	Techniques	Limitations / Future Work
11.	**Traffic Sign Detection using Clara and Yolo in Python** **Author:** Yogesh Valeja, Shubham Pathare , Dipen patel , Mohandas Pawar (ICACCS - 2021)	• Feature Extraction Techniques for Object Detection • Clara • YOLO	• This paper's proposed algorithm senses and monitors one or more moving objects in a variable context simultaneously. The experimental results show that the use of two object dimension features and their intensity distribution solved the data association problem very efficiently during monitoring.
12.	**Deep Learning based Traffic Analysis of Motor Cycles in Urban City** **Author :** Abirami T, Nivas C, Naveen R, Nithishkumar T G (IEEE - 2022)	• YOLO v3 • SORT tracker • R-CNN (better than CNN)	• The horizontal scaling would allow for even more effective traffic-flow optimization across the metropolis.
13.	**ODAR: A Lightweight Object Detection Framework for Autonomous Driving Robots** **Author :** Le Hoang Duong , Huynh Thanh Trung , Pham Minh Tam , Gwangzeen Ko , Jung Ick Moon , Jun Jo1 , Nguyen Quoc Viet Hung (IEEE - 2021)	• Residual CNN Network - aggregates and forms image features. • series of layers to mix and combine image features to pass them forward to prediction	• Can not used to detect traffic police gesture.
14.	**Yolo Target Detection Algorithm in Road Scene Based on Computer Vision** **Author :** Haomin He (IEEE - 2022)	• YOLOv4.	• It can detect person, vehicle, bicycle ,motorbike, MAP FPS.
15.	**Low Cost Hand Gesture Control in Complex Environment Using Raspberry Pi** **Author**: Chana Chansri, Jakkree Srinonchat, Eng Gee Lim and Ka Lok Man (IEEE-2019)	• hand gesture • raspberry pi • radian fingertip analysis	• This work presented a system for contactless HCI with hand gesture. The wireless controlling system using the Raspberry Pi as well as RGB camera has been completed and tested.

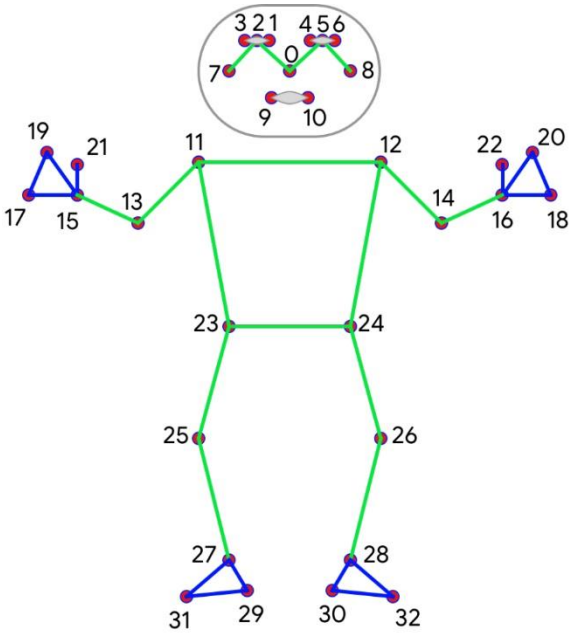
Proposed Work

1. In this model we have used YOLO and CNN to identify the traffic police.
2. Then we have used Googles MEDIAPIPE library to extract the skeleton points.
3. We have udes sea custom made dataset to train the model as there is no dataset available. Our model will aid self driving cars to become 100% autonomous in Indian Roads.

Mediapipe Library:

MediaPipe is a Framework for building machine learning pipelines for processing time-series data like video, audio, etc. This cross-platform Framework works in Desktop/Server, Android, iOS, and embedded devices like Raspberry Pi and Jetson Nano.

MediaPipe Toolkit comprises the Framework and the Solutions. Handpose recognition is a deep learning technique that allows you to detect different points on your hand. These points on your hand are commonly referred to as landmarks. These landmarks consist of joints, tips, and bases of your fingers.



0. nose
1. right eye inner
2. right eye
3. right eye outer
4. left eye inner
5. left eye
6. left eye outer
7. right ear
8. left ear
9. mouth right
10. mouth left
11. right shoulder
12. left shoulder
13. right elbow
14. left elbow
15. right wrist
16. left wrist
17. right pinky knuckle #1
18. left pinky knuckle #1
19. right index knuclke #1
20. left index knuckle #1
21. right thumb knuckle #2
22. left thumb knuckle #2
23. right hip
24. left hip
25. right knee
26. left knee
27. right ankle
28. left ankle
29. right heel
30. left heel
31. right foot index
32. left foot index

- After extracting skeleton points we try to figure out / find out the angle for all the points for hands and then for different poses we will make some range
- i.e there are some hand gesture for traffic police like VIPsalute , TurnRight , TurnLeft , RightTurnWaiting , LeftTurnWaiting , MoveStraight , Stop etc.