# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

**BELAGAVI-590018, KARNATAKA**

**PROJECT REPORT**

ON

***"Vision Assisted Anti-Braking System for Two Wheelers (VAABS)"***

**Submitted by**


**Vipul Kumar Gupta 1CR16EC193**

**Sanjhi Pandey 1CR16EC152**

**Yash Khandelwal 1CR16EC197**

**NM Nikhil 1CR16EC089**


**Under the guidance of**


| | |
|---|---|
| **Prof. Sridevi** | **Prof. Mahesh G.** |
| Assistant Professor | Associate Professor |

Department of Electronics and Communication Engineering

**Department Of Electronics and Communication Engineering**

**CMR INSTITUTE OF TECHNOLOGY**

#132, AECS LAYOUT, IT PARK ROAD, KUNDALAHALLI,

BENGALURU-560037

**Department Of Electronics and Communication Engineering**

# CERTIFICATE

This is to certify that the dissertation work **"Vision Assisted Anti-Braking System For Two Wheelers (VABS)"** carried out by **NM Nikhil**, **Sanjhi Pandey, Vipul Kumar Gupta, Yash Khandelwal**, holding USN: **1CR16EC089, 1CR16EC152, 1CR16EC193, 1CR16EC197** bonafide students of **CMRIT** in partial fulfillment for the award of **Bachelor of Engineering** in **Electronics and Communication Engineering** of the **Visvesvaraya Technological University, Belagavi,** during the academic year **2019-20**. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said degree.

Signature of Guide | Signature of HOD | Signature of Principal

**External Viva**

Name of Examiners Signature & date

1.

2

## ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without mentioning the people whose proper guidance and encouragement has served as a beacon and crowned my efforts with success. We take an opportunity to thank all the distinguished personalities for their enormous and precious support and encouragement throughout the duration of this seminar.

We take this opportunity to express my sincere gratitude and respect to **CMR Institute of Technology, Bengaluru** for providing me an opportunity to present my technical seminar.

We have a great pleasure in expressing my deep sense of gratitude to **Dr. Sanjay Jain,** Principal, CMRIT, Bangalore, for his constant encouragement.

We would like to thank **Dr. R Elumalai,** HoD, Department of Electronics and Communication Engineering, CMRIT, Bangalore, who shared his opinion and experience through which I received the required information crucial for the project.

We consider it a privilege and honor to express my sincere gratitude to my guide **Prof. Sridevi,** Assistant Professor, Department of Electronics and Communication Engineering, for the valuable guidance throughout the tenure of this project.

We express our gratitude to **Prof. Mahesh G.**, Associate Professor, Department of Electronics and Communication, for his immense support and guidance during the project.

I also extend my thanks to the faculties of the Electronics and Communication Engineering Department who directly or indirectly encouraged me.

Finally, I would like to thank my parents and friends for all their moral support they have given me during the completion of this work.

# ABSTRACT

The two-wheeler has a camera mounted on it and moving objects are detected by the camera which include pedestrians, moving cars, etc. The camera is mounted on a prototype that is used to measure the distance of the detected or tracked objects from the prototype. Detection, tracking and related distance calculations were performed using the integration of results and is done electronically and brake is applied automatically to the prototype. ABS then continuously and repeatedly applies optimum braking pressure to each wheel, meaning the system will brake just enough to not lock the wheels. It reduces the risk of skidding even when undertaking excessive evasive maneuvers. The servo motor connects the output signal of the microcontroller which in turn controls the brakes of the vehicles.Once the relative distances are calculated it is then compared with a predefined threshold value, the threshold value can depend on various factors (slow traffic conditions, night time, wet weather,etc.).After the comparisons are made in the next step the microcontroller will send a signal to the control system of the brakes only if the object in question is nearer i.e within the threshold values.If it's not then no such signals are passed.

# CONTENTS

# LIST OF FIGURES AND TABLES

Chapter 1

# INTRODUCTION

## 1.1 Lack of Safety

This section gives an overview of the present scenario of accidents taking place among two wheeler vehicles

Two-wheeler riders accounted for 40% of all traffic deaths in India in 2018, according to WHO's Global Status Report on Road Safety 2019, which puts the total traffic deaths at 1.5 lakh that year.

The actual number of deaths, however, could be higher as per the report, which estimates that there were 2.99 lakh traffic fatalities. This means, an estimated 22.6 people per 100,000 population died of traffic accidents in India in 2016.

"The highest number of casualties are reported among people on two and three wheelers, and let me add pedestrians, because they do not have a protective exterior exposing them to more injuries. Also, the kind of speed and traffic mix in India means that cars with higher speed limits drive on the same road with two-wheelers with lower speed limit, making them more accident prone," said Dr Rakhi Dandona who heads the Global Burden of Disease –Road Injuries group for the State-level Disease Burden initiative.

Globally, 1.35 million people died in traffic crashes the same year.

Protocols and measures taken by the government to aid this issue may not be sufficient. Even the automobile industry is not entirely focused on providing safety measures for two-wheelers.

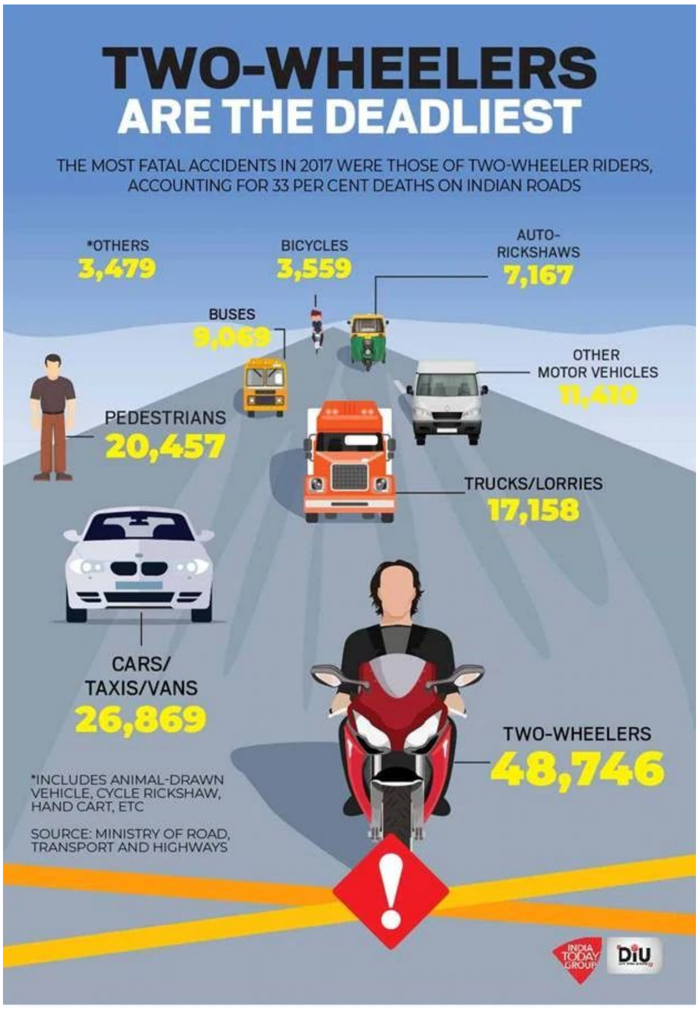


Fig 1.1 Accident rates of two wheeler

### 1.2 Proposed system

This section talks about the solution we proposed to the existing problem of traffic hazard on two wheeler vehicles.

Even if you're an experienced driver, unexpected trouble sometimes can get in the way. In an attempt to avoid imminent collision or danger on the road ahead, you may find yourself jumping on the brake.

Some of the safety measures provided in the two wheelers include the conventional ABS which basically prevents the rear tyre from locking during sudden brakes. Other measures include traction control for better handling in wet weather conditions. Majority of the accidents that occur are caused due to driver behavior and brake reaction time is one of the major factors. Average brake reaction time for a driver is about 2.3 seconds but this reaction time can get affected due to drunk driving, driver fatigue, tailgating, etc.

These issues are easily resolved by a driver assistant system that employs image or video processing algorithms for detection of moving vehicles. These systems are predominantly made for four wheelers or heavy vehicles.

Hence we have proposed a method using vision assistance to determine the distance between vehicle and the user in an urban traffic and automate the braking system in order to avoid accidents. The purpose of the proposed method is to implement machine learning algorithms like YOLO object detection model in computer vision and mono depth model perception. This proposed system is called Automatic Braking System.

### 1.2.1 What is an Automatic Braking System?

The two-wheeler has a camera mounted on it and moving objects are detected by the camera which include pedestrians, moving cars, etc.

The camera is mounted on a prototype that is used to measure the distance of the detected or tracked objects from the prototype. Detection, tracking and related distance calculations were performed using the integration of results and is done electronically and brake is applied automatically to the prototype.

ABS then continuously and repeatedly applies optimum braking pressure to each wheel, meaning the system will brake just enough to not lock the wheels.

It reduces the risk of skidding even when undertaking excessive evasive maneuvers.

### 1.2.2 How effective is it?

Certain older models can be bought without ABS, and some people do prefer not having it. But it's a very effective safety feature. This can revolutionize safety for two-wheelers by providing a Vision Assisted Automated Braking System in real-time assistance to the rider which ensures drivers get optimal riding speed based on surrounding vehicles and prompt for overtaking vehicles.

- Bikes fitted with ABS are less likely to be involved in a fatal crash.
- It decreases the chance of frontal collision on wet and dry roads.
- Bikes with ABS rarely stray from the road ahead.

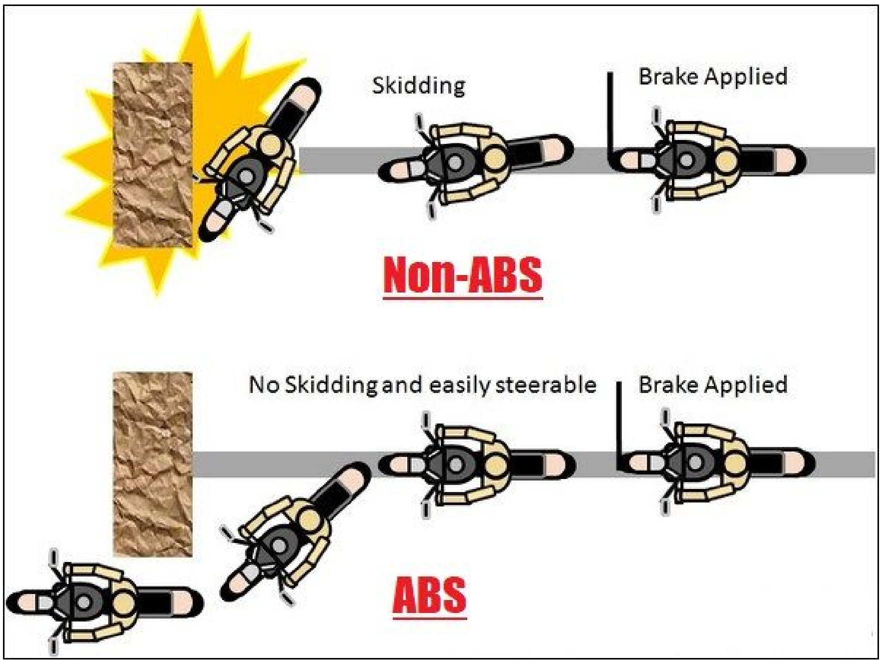


Fig. 1.2 Benefit of ABS

**1.3 Steps involved**

There are five steps in the autonomous braking system:

1. Collecting real-time data frames through a camera.

2. Detection and classification of different types of vehicles.

3. Calculating the relative distances from the vehicles using Depth perception.

4. Taking a decision based on the distance calculation and
Further feeding it into the brake control system.

This section gives details about the problems occurring among two wheeler vehicles due to reduced safety on road and the system proposed by us to solve the problem.

# Chapter 2

## LITERATURE SURVEY

There are many risks involved with night driving including night vision, fatigue, rush hour and dangerous driving. To minimize these risks, numerous ways to assist drivers have been developed. One such system is the advancements in the vision based classification **[1]** the purpose of the first part is to make sure that the driver can be in sufficient condition to receive warnings; this is part of the driver acquisition experience. The purpose of the second part is to notify the driver when the driver's vehicle is traveling in heavy traffic during the evening or at night.

**Driver Assistance System ``EyeSight"** [2] the Autonomous emergency braking system (AEB).This system is mainly developed by automobile manufacturers Subaru is refined by two points. The first is to speed up the reaction of crossing pedestrians coming out after obstacles to increase the effect of slowing down. Another point is to improve pedestrian response at night. Efficiency is increased by suppressing a decrease in the recognition performance caused by the blackness of the entry or the color of pedestrian clothing. This upgrade will provide a higher level of security.

**Jiawei Mo and Junaed Sattar**[3] presented a paper on Safe Drive that was aimed at improving the visibility of self-driving and autonomous driving marks, especially under poor visibility. The system detects additional marking of traffic images in the vehicle area and also uses the unique 3D model of the surroundings. When measuring the geometric relationship between this 3D model and the current view, the road markings are displayed in a vertical position; any type of traffic detection algorithm can be used later to find the paths in the resulting image. SafeDrive requires no additional sensors except for peripherals and data. SafeDrive is based on the idea that some images show the current location and looks better quality; therefore, road marks can be found in those images and later reflected in the present, distorted image.In SafeDrive, routes are expected to use geometric transcripts between the current view and sparse reconstruction Current view taken from other photos,using Structure from Motion method. Detail photographs, identified by location, searches for images very similar to the current scenario, with the 3D model the road has been rebuilt in these images. Current viewis now registered with this 3D model.This method attests to the availability of other images in the current driving environment and the ability to perform street detection with those images, eventually incorporating roads into the original camera image. Since we have solid detection algorithms, accurate detection detection, and robust methods of linking the previous image to a live frame, we believe this algorithm can greatly improve driver safety. The ultimate goal of our job is to create an efficient system, while simultaneously improving the quality of autonomous transport and the safety that exists for road vehicles. Ongoing research focuses on consistent road pixels, compressed data management and improved performance efficiency, and comprehensive evaluation of data collected from various spatial frameworks.

In a paper presented by Yusuke Nakamura and Toshiyuku Murakami [4] they have underlined the importance of the **Advanced Emergency Braking System (AEBS)** in cars. The AEBS exposure is as follows: when a camera or radar attached to a car catches the same obstacle as another car or traveler before, the risk of collision breaks down from relative position and relative speed, and the system deteriorates. The car. According to a vehicle manufacturer survey, the number of accidents has decreased by 61% compared to unarmed vehicles, and the number of back-accidents has dropped by 84%. Further, the paper also suggests that the system takes action against pedestrians which in turn reduces the chance of collisions. Over the years, the Autonomous Emergency Braking System (AEBS) has been attracting attention as a technology to reduce the number of road accidents. The AEBS framework is like a sequence of signs: when a camera or radar mounted on a car detects a collision risk being judged from a relative position and at the same speed, and the system lowers the vehicle. According to a survey by car manufacturers, the number of accidents dropped by 61% compared to unmanned vehicles, while the number of rear accidents dropped by 84% (4). It proves the importance and help of AEBS.The outline of the proposed route is as follows. First, pedestrian safety is measured from the current pedestrian location with the Kalman filter. The predictive nature of pedestrians in the future is predicted from the pedestrian wheel, and the number of maximum power orders is still calculated by the MPC. Here, as the MPC is most affected by interference and parameter error, a high power response team is recommended to improve control performance.In addition, the driving rating is included in the proposed system. By using the proposed method, it is possible to predict the location of pedestrians in the future and to determine the risk of a collision.

Pedestrian detection [5] with video processing using active and night vision systems is an increase in vehicle safety, especially during night when conditions are tricky to drive. -IR at night detecting an object with the help of **IR LED and photodiode pair**, this camera has the potential to detect an object up to 100m. The thermal camera detects the heat emitted by any objects such as cars, Human Animals etc. By using these two cameras attached to the car it helps the driver to drive safely.Pedestrian detection obtained by Video Processing using Night thermal Vision System where Matlab 2014 (B) version for image acquisition and Computer matlab package is used for making image processing.First step is to upload a photo or video frame. As it is a night camera image the photo will look like it image is measured in gray but will not. So the system changes the image and gray scale does the processing on it. After that it Converts a large gray image to block the binary image used to find the corner using the FAST / object Corner algorithm detects and identifies all updated angles Then an Extract histogram for oriental gradients (HOG) features to get the right item in the picture.In the pre-processing stage the blur caused by the movement of the vehicle-related object is removed. This provides a viable process. After this step we have to prepare our region of interest (ROI). This is a very important step, if the selected region is incomplete or unsuitable the outcome of that framework will not be correct. In other words, if at this stage of the district there is a shortage of pedestrian objects then this framework is a waste. The first step in ROI is segmentation; separation is the process of removing the beloved region from the back of the image. Frequently used operating system maps and split maps are used in stereo view systems. In this work a modified dres-threshold direction is used. The algorithm converts a gray-like image into a binary image, where white objects are collectors and the background is black. It works flexibly under a variety of lighting conditions as well as a comparison phase.The final stage that allows the object to be classifier. The most common classifiers are: vector support machine (SVM) as an example of a supervised learning method, neural networks, self-organizing maps (SOM), and matrices of neurons. The most useful algorithm at the time of division is a scalable algorithm.

In this system, a **HOG (Histogram of oriented gradients) was created for the algorithm and support vector machine (SVM)** [6].

One approach has been the integration of the Vision based driver assist system [7] with computational compilation.The proposed system monitors vehicles in front of a vehicle and notifies the driver of any hazard caused by the other vehicle's movement. The system uses a smartphone that is mounted on a car, captures photos and sends the necessary information to the cloud for processing. A notice is sent to detect any dangerous situations.The system is a useful driver-assisted driver based on the FCW concept, which warns the driver of collisions based on the use of current Dynamics in front vehicles. the workload varies between a smartphone placed on a dashboard or an air car ego display and an EC server on Cloudlet. Frame detection and preparation steps require minimal memory and processing power so it can be executed in real time on a smartphone platform. Frames are extracted and sent to the EC server for future processing. With the EC server to detect and track algorithms, process the uploaded frames and detect any dangerous situation, send a notification message to the smartphone app to notify the driver. and it has been shown that implementing a DAS-based concept is possible with a possible 5G network.

# Chapter 3

## HARDWARE

This project focuses on automotive safety systems. The two wheeler has a camera mounted on it and moving objects are detected by the camera which include pedestrians, moving cars, etc.

A camera is mounted on the prototype which is used to measure the distance of the detected or tracked object from the prototype model. The detection, tracking and relative distance calculation of the moving object by the camera are done by using YOLO and Monodepth algorithm. After the simulation,

Integration of results is done with the computer and brake is applied automatically to the prototype. An overview of the automated system is shown in fig.

A schematic overview of the hardware is shown in Fig1. The camera head is mounted on two wheelers and is also connected to an on-board computer. All general-purpose tasks or algorithms such as YOLO or MONODEPTH are run on the main computer. Then the main computer is interfaced with a servo driver. First the series of images are fed to the algorithm using Jetson nano and all the computations are done using using Nvidia Maxwell GPUs which performs calculations parallely and gives the output to the servo driver which generates PWM signals to activate the servos.

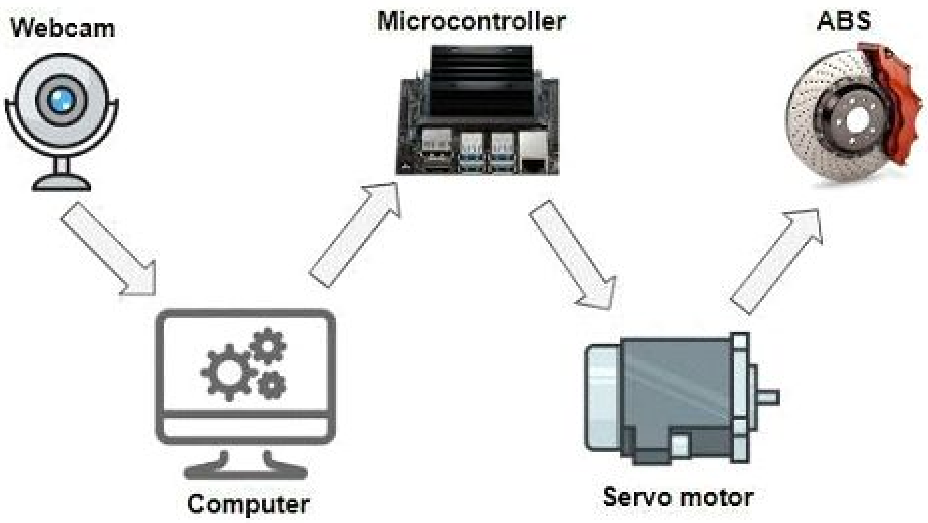

Fig. 3.1 Basic design of System

## 3.1 Webcam

The camera is mounted on a prototype that is used to measure the distance of the detected or tracked objects from the prototype. Here we have used Logitech c270 camera which captures all the image frames from the surroundings.

Fig 3.2 Logitech C270 Camera

## 3.2 Onboard Computer

All the input outputs are controlled and managed by the onboard computer present in Jetson nano. It handles all the interfaces and connection of other devices to it. It has a ARM based CPU which processes all the functionalities in the control system.

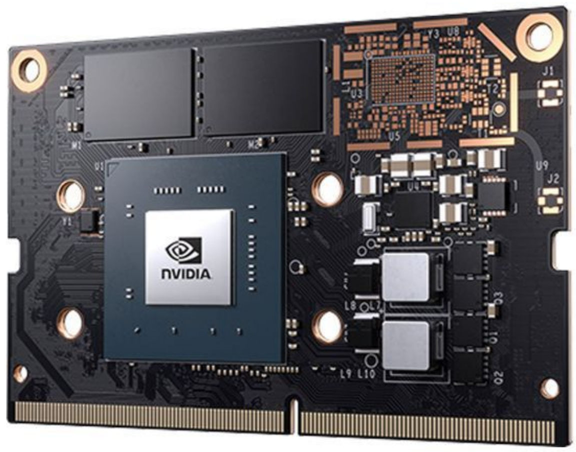


Fig 3.3 Quad-core ARM® Cortex®-A57 MPCore processor

## 3.3 Computational Devices

All the work related to image processing is done using two different computational devices. In the first version, Intel neural compute stick is used which uses Intel's OpenVINO toolkit and the second version uses another computational device i.e Nvidia Maxwell GPUs pre-installed with Jetson Nano.

### 3.3.1 Intel Neural Compute Stick

Intel has launched the next version of its Neural Compute Stick, a simple USB device that accelerates AI processing and deep learning inference on consumer PCs.The new Neural Compute Stick 2 uses the Intel Movidius Myriad X VPU.

The Neural Compute Stick 2 is slightly larger than a thumb drive and plugs into a standard USB 3.0 port. It can be used with pre-trained neural networks for smart cameras, IoT devices, robotics, drones, and VR hardware.

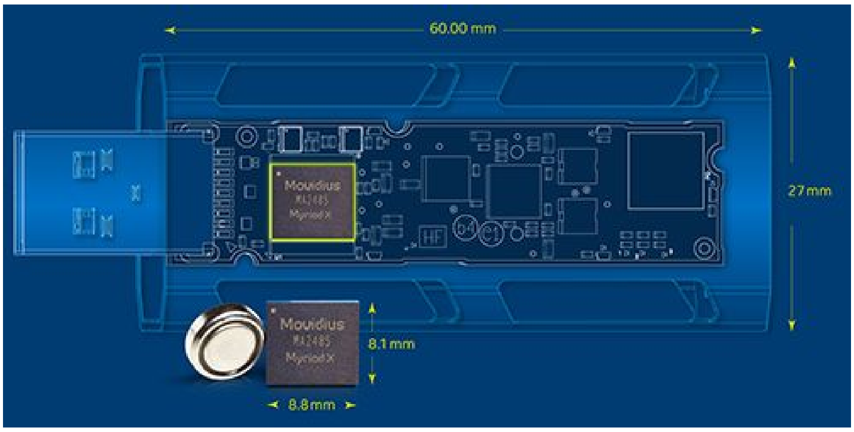


Fig 3.4 Intel Neural Compute Stick with MyriadX VPUs

### 3.3.2 Nvidia Jetson Nano

NVIDIA Jetson Nano Developer Kit is a small, powerful computer. It has a memory of 4GB. It lets you run multiple neural networks in parallel. It is used for applications like image classification, object detection, segmentation, and speech processing. It is easy to use platform that runs on minimum 5 watts. It also has a built in camera interface.

Fig 3.5 NVIDIA Maxwell™ architecture with 128 NVIDIA CUDA® cores 0.5 TFLOPs (FP16)

## 3.4 ABS

At the end feedback from the system is given to the control brakes which further stops the vehicle. Here a redesigned ABS braking system is used where servo motors controls the braking action of the vehicle.


Fig 3.6 ABS

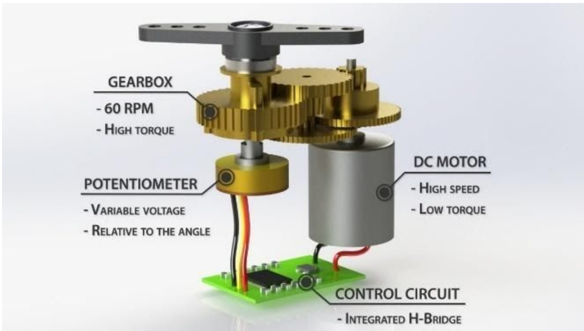


Fig 3.7 Braking system

## 3.5 Proposed Hardware Design

The camera head is mounted on two wheelers and is also connected to an on-board computer. All general-purpose tasks or algorithms such as YOLO or MONO DEPTH are run on the main computer. Then the main computer is interfaced with a servo driver. First the series of images are fed to the algorithm using Jetson nano and all the computations are done using Nvidia Maxwell GPUs which performs calculations parallelly and gives the output to the servo driver which generates PWM signals to activate the servos.

# Chapter 4

## SOFTWARE

Intel OpenVINO toolkit [8] is used for computer vision applications to get higher performance.The YOLO[9] object detection model and monodepth model[10] are converted using a model optimizer, which generates two different files having extensions .bin and .xml. The inference engine processes the files giving us the results.The results before and after using the monodepth model is shown in Fig 3 and Fig 4 respectively.

For improving system performance the same algorithms are run on an improved hardware system i.e Nvidia Jetson nano[11].

## 4.1 Implementation Steps

The steps in the automated braking system as shown in Fig 5

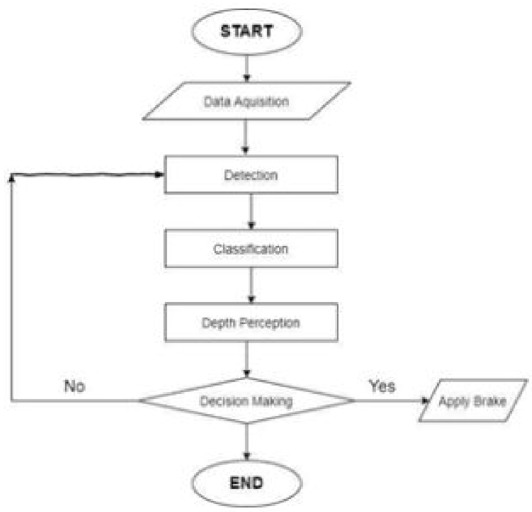


Fig: 4.1 Flowchart of whole process

### 4.1.1 Environment Monitoring

The first crucial step involves monitoring of the environment around the vehicle in this system; it is done by a relatively small camera(Logitech C270 ) which continuously captures the surroundings and then these frames are feeded to the microcontroller(jetsan nano) in real-time.

### 4.1.2 Overview

Detection and classifications are obtained from the microcontroller as the real time video is fed, it performs object detection and classification models on those video inputs.The algorithms that are used are fast enough to do the processing in real time. The different objects present in each of the frames are detected by the object detection model and the detection boundaries around the vehicles or any other objects are marked out.This process is usually done by a YOLO object detection model with the help of CNNs in different layers.

### 4.1.3 Depth Perception

The next steps involves calculating the relative distances and velocities of the driver and the traffic ahead or him/her and this can be done with the help of mono depth perception algorithm.This algorithm is based on tensorflow, first a color map of the frames is created which is then compared with the standard scale of color- distance.The results of the comparisons are obtained using which the relative distances between the object and the camera is calculated.

### 4.1.4 Decision making

The servo motor connects the output signal of the microcontroller which in turn controls the brakes of the vehicles.Once the relative distances are calculated it is then compared with a predefined threshold value, the threshold value can depend on various factors (slow traffic conditions, night time, wet weather,etc.).After the comparisons are made in the next step the microcontroller will send a signal to the control system of the brakes only if the object in question is nearer i.e within the threshold values.If it's not then no such signals are passed.

## 4.2 Yolo Algorithm

“You Only Look Once,” or YOLO, family of models are a series of end-to-end deep learning models designed for fast object detection, developed by Joseph Redmon.This methods is extremely fast as compared to other methods.Fast YOLO can process up to 155 frames per second and hence is used in real time application.

It is based on regression i.e the model easily predicts clases and bounding boxes in the first run. Implementation steps are as follows:

- A single neural network is applied to the full image.
- This gives us a division of the image into regions with bounding boxes and probabilities for each region
- Then the high scoring detections are selected.
- These high scoring detections are then classified as different objects in a frame

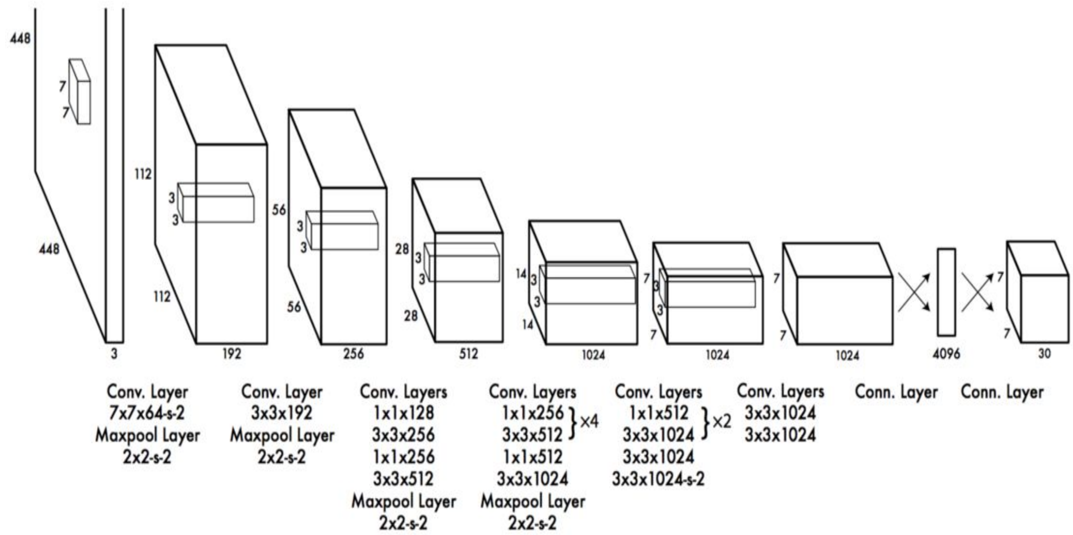


Fig 4.2 - Convolutional neural network with layers

YOLO has 24 convolutional layers followed by 2 fully connected layers (FC). Some convolution layers use 1 × 1 reduction layers alternatively to reduce the depth of the features maps. For the last convolution layer, it outputs a tensor with shape (7, 7, and 1024). The tensor is then flattened. Using 2 fully connected layers as a form of linear regression, it outputs 7×7×30 parameters and then reshapes to (7, 7, 30), i.e. 2 boundary box predictions per location.

A faster but less accurate version of YOLO, called Fast YOLO, uses only 9 convolutional layers with shallower feature maps.

YOLO predicts multiple bounding boxes per grid cell. To compute the loss for the true positive, we only want one of them to be **responsible** for the object. For this purpose, we select the one with the highest IoU (intersection over union) with the ground truth. This strategy leads to specialization among the bounding box predictions. Each prediction gets better at predicting certain sizes and aspect ratios.
YOLO uses sum-squared error between the predictions and the ground truth to calculate loss. The loss function comprises of:

- The **classification loss**.
- The **localization loss** (errors between the predicted boundary box and the ground truth).
- The **confidence loss** (the objectness of the box).

### 4.2.1 Class Prediction

Most classifiers assume output labels are mutually exclusive. It is true if the output is mutually exclusive object classes. Therefore, YOLO applies a softmax function to convert scores into probabilities that sum up to one. YOLOv3 uses multi-label classification. For example, the output labels may be “pedestrian” and “child” which are not non-exclusive. (the sum of output can be greater than 1 now.) YOLOv3 replaces the softmax function with independent logistic classifiers to calculate the likeliness of the input belonging to a specific label. Instead of using mean square error in calculating the classification loss, YOLOv3 uses binary cross-entropy loss for each label. This also reduces the computation complexity by avoiding the softmax function.

### 4.2.2 Bounding box prediction & cost function calculation

YOLOv3 predicts an objectness score for each bounding box using logistic regression. YOLOv3 changes the way in calculating the cost function. If the bounding box prior (anchor) overlaps a ground truth object more than others, the corresponding objectness score should be 1. For other priors with overlap greater than a predefined threshold (default 0.5), they incur no cost. Each ground truth object is associated with one boundary box prior only. If a bounding box prior is not assigned, it incurs no classification and localization lost, just confidence loss on objectness. We use tx and ty (instead of bx and by) to compute the loss.

### 4.2.3 Feature Pyramid Networks (FPN) like Feature Pyramid

YOLOv3 makes 3 predictions per location. Each prediction composes of a boundary box, a objectness and 80 class scores, i.e. $N \times N \times [3 \times (4 + 1 + 80)]$ predictions.
YOLOv3 makes predictions at 3 different scales (similar to the FPN):

1. In the last feature map layer.
2. Then it goes back 2 layers back and up samples it by 2. YOLOv3 then takes a feature map with higher resolution and merges it with the upsampled feature map using element-wise addition. YOLOv3 applies convolutional filters on the merged map to make the second set of predictions.
3. Repeat 2 again so the resulting feature map layer has good high-level structure (semantic) information and good resolution spatial information on object locations.

To determine the priors, YOLOv3 applies a k-means cluster. Then it pre-select 9 clusters. For COCO, the width and height of the anchors are (10×13),(16×30),(33×23),(30×61),(62×45),(59× 119),(116 × 90),(156 × 198),(373 × 326). These 9 priors are grouped into 3 different groups according to their scale. Each group is assigned to a specific feature map above in detecting objects.

### 4.2.4 Feature extractor

A new 53-layer Darknet-53 is used to replace the Darknet-19 as the feature extractor. Darknet-53 mainly comprises 3 × 3 and 1× 1 filters with skip connections like the residual network in ResNet. Darknet-53 has less BFLOP (billion floating point operations) than ResNet-152, but achieves the same classification accuracy at 2x faster.

### 4.2.5 YOLOv3 performance

YOLOv3's COCO AP metric is on par with SSD but 3x faster. But YOLOv3's AP is still behind RetinaNet. In particular, AP@IoU=.75 drops significantly comparing with RetinaNet which suggests YOLOv3 has higher localization error. YOLOv3 also shows significant improvement in detecting small objects. YOLOv3 performs very well in the fast detector category when speed is important.

| | backbone | AP | $AP_{50}$ | $AP_{75}$ | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|---|---|---|
| *Two-stage methods* | | | | | | | |
| Faster R-CNN+++ [3] | ResNet-101-C4 | 34.9 | 55.7 | 37.4 | 15.6 | 38.7 | 50.9 |
| Faster R-CNN w FPN [6] | ResNet-101-FPN | 36.2 | 59.1 | 39.0 | 18.2 | 39.0 | 48.2 |
| Faster R-CNN by G-RMI [4] | Inception-ResNet-v2 [19] | 34.7 | 55.5 | 36.7 | 13.5 | 38.1 | 52.0 |
| Faster R-CNN w TDM [18] | Inception-ResNet-v2-TDM | 36.8 | 57.7 | 39.2 | 16.2 | 39.8 | **52.1** |
| *One-stage methods* | | | | | | | |
| YOLOv2 [13] | DarkNet-19 [13] | 21.6 | 44.0 | 19.2 | 5.0 | 22.4 | 35.5 |
| SSD513 [9, 2] | ResNet-101-SSD | 31.2 | 50.4 | 33.3 | 10.2 | 34.5 | 49.8 |
| DSSD513 [2] | ResNet-101-DSSD | 33.2 | 53.3 | 35.2 | 13.0 | 35.4 | 51.1 |
| RetinaNet [7] | ResNet-101-FPN | 39.1 | 59.1 | 42.3 | 21.8 | 42.7 | 50.2 |
| RetinaNet [7] | ResNeXt-101-FPN | **40.8** | **61.1** | **44.1** | **24.1** | **44.2** | 51.2 |
| YOLOv3 608 × 608 | Darknet-53 | 33.0 | 57.9 | 34.4 | 18.3 | 35.4 | 41.9 |

Fig 4.3 Different model comparision

## 4.3 Mono Depth perception algorithm

Monodepth perception model solely focuses on depth perception problems on a single frame basis.

This can be divided into four steps:

- background depth generation
- motion analysis
- moving object extraction
- depth fusion

In the first step the background depth generation (BDG) marks out the geometric perspective and sky detection. Vanish lines and Vanish points are created. These give an idea about the depth in image.

In the next Step, motion analysis is done between static frames and the frames from the camera in motion.This is carried out by calculating the sum of absolute difference (SAD) of the four edges between two consecutive frames.

Once the frames with motion have been detected then the moving objects in those frames are extracted.This is done by employing morphological operations on them(like holes

filling or closing operations).These morphological operations are more efficient than the convolutional operations as they are shape dependent.
The last step involves the merging the depth of the moving object and the depth of the background. The depth of the bottom-most part of the moving object is assigned the same value as that of the ground.Post this a filter is applied to the resulting image to smoothen it out.

### 4.3.1 How mono-depth models are trained.

Generally in order to determine the depth or distance of an object two cameras are required but in our experiment we use only one camera. So, in our case the images obtained by the camera are first inverted. By using this technique two images are obtained one is the original image and other is the mirrored image. So both the images are given as an input to the mono-depth model.

### 4.3.2 Supervised Single Image Depth Estimation

Single-view, or monocular, depth estimation refers to the problem setup where only a single image is available at test time. Saxena et al. proposed a patch-based model known as Make3D that first over-segments the input image into patches and then estimates the 3D location and orientation of local planes to explain each patch. The predictions of the plane parameters are made using a linear model trained offline on a dataset of laser scans, and the predictions are then combined together using an MRF. The disadvantage of this method, and other planar based approximations, e.g., is that they can have difficulty modeling thin structures and, as predictions are made locally, lack the global context required to generate realistic outputs. Instead of hand-tuning the unary and pairwise terms, Liu et al. use a convolutional neural network (CNN) to learn them. In another local approach, Ladicky et al. Incorporate semantics into their model to improve their per pixel depth estimation. Karsch et al. attempt to produce more consistent image level predictions by copying whole depth images from a training set. A drawback of this approach is that it requires the entire training set to be available at test time. Eigen et al. showed that it was possible to produce dense pixel depth estimates using a two scale deep network trained on images and their corresponding depth values. Unlike most other previous work in single image depth estimation, they do not rely on hand crafted features or an initial over segmentation and instead learn a representation directly from the raw pixel values. Several works have built upon the success of this approach using techniques such as CRFs to improve accuracy, changing the loss from regression to classification, using other more robust loss functions, and incorporating strong scene priors in the case of the related problem of surface normal estimation. Again, like the previous stereo methods, these approaches rely on having high quality, pixel aligned, ground truth depth at training time. We too perform single depth image estimation, but train with an added binocular color image, instead of requiring ground truth depth.

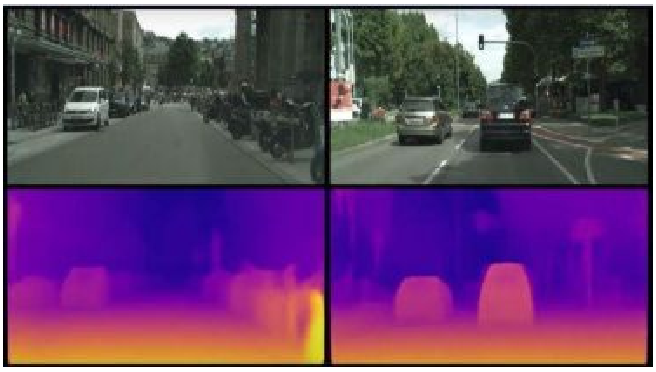

Fig. 4.3.2 Mono Depth Perception output image

The training of the model is based on supervised learning where it is trained on the two images original and the inverted image.

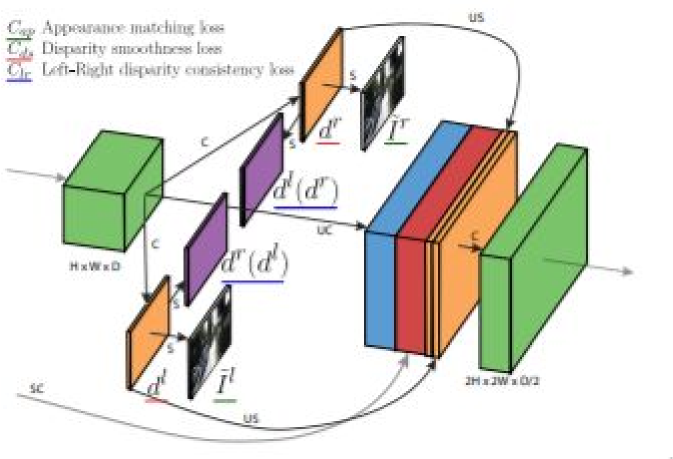


Fig 4.3.2 (b) Real and mirror image

### 4.3.3 Depth Estimation as Image Reconstruction

Given a single image I at test time, our goal is to learn a function f that can predict the per-pixel scene depth, ˆd=f (I). Most existing learning based approaches treat this as a supervised learning problem, where they have color input images and their corresponding target depth values at training. It is presently not practical to acquire such ground truth depth data for a large variety of scenes. Even expensive hardware, such as laser scanners, can be imprecise in natural scenes featuring movement and reflections. As an alternative, we instead pose depth estimation as an image reconstruction problem during training. The intuition here is that, given a calibrated pair of binocular cameras, if we can learn a function that is able to reconstruct one image from the other, then we have learned something about the 3D shape of the scene that is being imaged. Specifically, at training time, we have access to two images I l and I r, corresponding to the left and right color images from a calibrated stereo pair, captured at the same moment in time. Instead of trying to directly predict the depth, we attempt to find the dense correspondence field d r that, when applied to the left image, would enable us to reconstruct the right image. We will refer to the reconstructed image I l (d r) as ˜I r. Similarly, we can also estimate the left image given the right one, ˜I l=I r (d l). Assuming that the images are rectified, d corresponds to the image disparity - a scalar value per pixel that our model will learn to predict. Given the baseline distance b between the cameras and the camera focal length f, we can then trivially recover the depth ˆd from the predicted disparity, ˆd=bf/d.

The mono-depth model returns a heatmap version of the image as shown in Fig 4.3.2. The color intensity in the image represents different distances. A mean distance is also returned from the mono-depth model known as pixel mean. Which is considered as distance from a whole object in this experiment it is the distance from the vehicle which is detected.

**Table 4.1- Original data acquisition from camera**

| | |
|---|---|
| Pixel_max | 201 |
| Window | [276 213 373 265] |

| | |
|---|---|
| position | -14.395207300589744 |
| Relative speed | -1.2038563745875308e-06 |
| Pixel_mean | 81.6304347826087 |

| | |
|---|---|
| Pixel_max | 184 |
| Window | [274 286 372 332] |
| position | -21.32931816642413 |
| Relative speed | -1.78374882866431835e-06 |
| Pixel_mean | 56.301886792452834 |

| | |
|---|---|
| Pixel_max | 146 |
| Window | [274 286 372 332] |
| position | -21.32931816642413 |
| Relative speed | -2.6384101303733587e-06 |
| Pixel_mean | 73.5384804261756 |
| Braking signal: -1 | low |

The final output shown below which shows how brakes are applied indicating ABS Activated when the distance is less than or equal to the predefined threshold value i.e the distance between bus and two-wheeler as shown in Fig.5.2.

The performance analysis is given in the below Table 4.2

**Table 4.2 Performance analysis of proposed work**

| S.No | Resource | Model | Latency |
|---|---|---|---|
| 1. | PC | YOLOV2+ monodepth | 2sec |
| | | YOLOV3+ monodepth | 2sec |
| | | YOLO tiny V3}+ monodepth | Less than 1 sec |
| 2. | Neural compute stick | YOLO V2+ | 1-2 sec |
| 3. | Nvidia Jetson Nano | YOLOV2+ monodepth | 1sec |
| | | YOLOV3+ monodepth | 1 sec |
| | | YOLO tiny V3}+ monodepth | Less than 0.3 sec |

# Chapter 5

# RESULTS

In this the window consists of two output modules. The first module shows the original data acquisitioned live video captured by the camera as shown in Fig 5.1.

Fig: 5.1- Output view of Relative speed and distance

The image shows various data output from the mono-depth perception and open yolo algorithm i.e data obtained from live video captured from the camera. From the received data: window matrix is a 1x4 matrix which is the coordinates of the window which appears over the detected vehicles. Pixel mean is the mean distance calculated through the mono-depth perception output and the distance is based on color obtained from monodepth. Pixel max is the maximum distance obtained from the mono-depth model.

Pixel/metre data is a standard value which remains constant and from the mono-depth values in pixels are being obtained, so pixel mean and pixel max are standardised using pixel/metre value.

The 4x3 matrix is the transformation matrix which is generated in order to transform the perspective of the road.

Second module shows the relative distance between the vehicles and their relative velocities as shown in Fig 5.2. So, finally the two main output obtained distance from the mono-depth model and relative velocity from perspective transformation is used as threshold values for applying the brakes.

Here is the final output shown below which shows how brakes are applied indicating **ABS Activated** when the distance is less than or equal to the predefined threshold value i.e the distance between bus and two wheeler as shown in Fig 5.2.

Fig 5.2 VAABS in action

Chapter 6

# APPLICATIONS AND ADVANTAGES

Automated brake systems are one of the more advanced systems that go into newer vehicles. Studies reveal they have resulted in a significant improvement in crashes, because vehicles no longer have to rely on the driver's instincts and fast reflexes.

If you own your own business, you might want to consider the various benefits of automated braking systems. In fact, according to a report by USA Today, there are 1.9 million crashes less a year due to the new features of vehicles, which include these braking systems. General Motors also reported that only about 40 percent of people brake when in an accident, showing just how important they can be.

## 6.1 Protect drivers

One of the big benefits to an automated braking system in your fleet vehicles is that your drivers are protected. They either get no injuries or less injuries thanks to the system. This not only is protecting the safety of your drivers, but is also decreasing how much you are paying for workers' compensation claims.

## 6.2 Insurance Premium Discounts

Speaking of insurance, what if you had lower insurance premiums with automated braking systems? This might include lower insurance premiums for workers' compensation insurance and your commercial auto insurance policies.

## 6.3 Less Damage to vehicles

You are saving money on having vehicles repaired since you are reducing how many potential accidents your drivers are getting into.

Chapter 7

## CONCLUSIONS AND SCOPE FOR FUTURE WORK

### 7.1 Conclusion

This project presents a vision based deceleration control for two wheelers.The system is designed to assist the driver to safe braking maneuvers.The exhaustive experimentation shows that the system leads to a more consistent and safer braking. This automated braking system is particularly valuable when the driver is distracted or visibility is poor.With an improved hardware or a dedicated microcontroller this system can be easily incorporated into the vehicles as an important safety feature.

### 7.2 Future Scope

We in the future want to bring out our project as an inbuilt commercial product for the two wheeler vehicles. This will be carried out initially on a bicycle prototype and then further expanded to higher applications.

Many different adaptations, tests, and experiments have been left for the future due to lack of time .Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods, or simply curiosity.

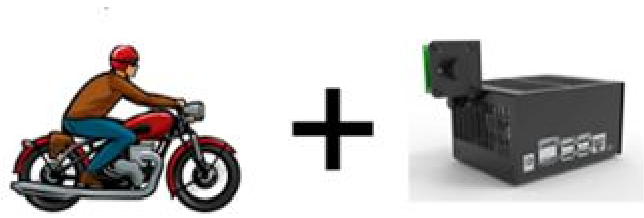

Fig 7.1 Vabs as a product

### 7.2.1 Vibration enabled gloves

This comes as an accessory for the betterment of user experience. It is connected through bluetooth and by fetching data from google maps it provides signals to the rider. This signal can be generated using vibrations or visible led lights. The signal is provided at a certain time gap prior to the direction which is to be followed by the user.

Fig 7.2 Vibration enabled gloves

### 7.2.2 Helmet with inbuilt direction Voice assistant

This is an optional external assistant that can be used by the rider to help make their driving experience easier and safer. This is implemented through an inbuilt speaker.

The speaker and the entire hardware system is controlled through a microcontroller.

This speaker uses google map direction assistant to play out the voice commands to the rider within the helmet and helps with factors like vehicle speed, distance gap from surrounding vehicles etc.

Fig 7.3 Helmet with inbuilt speakers

These are just a limited number of possible enhancements to the present work done and there are a wide range of possibilities that could be considered for future development of this product. This section gives an idea about the future products and applications of our project.

# REFERENCES


1. For textbooks – A.V. Oppenheim and R.W. Schafer, Digital Signal Processing, Englewood, N.J., Prentice Hall, 3 Edition, 1975.
2. For papers – Devid, Insulation design to combat pollution problem, Proc of IEEE, PAS, Vol 71, Aug 1981, pp 1901-1907.
3. Chien, J.-C., Chen, Y.-S., & Lee, J.-D “Improving Night Time Driving Safety Using Vision-Based Classification Techniques Sensors” 17(10), 2199. doi:10.3390/s17102199,2017.

1. S Usui, N. Nomura, H. Kumagai, and H. Sekine: “Development of Improvements to Driver Assistance System “EyeSight” for reduction of traffic Accidents”, Journal of Japan Society of Applied Electromagnetics and Mechanics, Vol.25, No.4,pp.383-389 ,2017.

1. Jiawei Mo and Junaed Sattar “SafeDrive: Enhancing Lane Appearance for Autonomous and Assisted Driving Under Limited Visibility” arXiv:1807.11575v1 [cs.CV] 24 Jul 2018.

1. Nakamura, Y., & Murakami, T. (2019). Advanced Deceleration Control Considering Driving Resistance by Predicting the Position of Pedestrians. IEEJ Journal of Industry Applications, 8(2), 334–341. doi:10.1541/ieejjia.8.334

[5] Ankit dilip yawale, V. B. Raskar. “ Pedestrian detection by video processing using thermal and night vision systems”. International journal of engineering sciences & research technology, 6(1), 29–38,2017

[6] Mishra,MS, and Anup girdhar. "Vehicle detection approach based support vector machine and histogram of oriented gradients."

[7] A. Keivani, f. Ghayoor and j. Tapamo, "A vision-based driver assistance system using collaborative edge computing," 2017 global wireless summit *(gws)*, cape town, pp. 160-164,2017.

[8] Intel open vino software manual

[9] J.Redmon, s. Divvala, r. Girshick and a. Farhadi, "you only look once: unified, real-time object detection",IEEE conference on computer vision and pattern recognition (cvpr), las vegas, nv, pp. 779-788,2016.

[10] F. Tosi, f. Aleotti, m. Poggi and s. Mattoccia, "learning monocular depth estimation infusing traditional stereo knowledge," IEEE /cvf conference on computer vision and pattern recognition (cvpr), long beach, CA, USA,9, pp. 9791-9801,2019.

[11] https://www.nvidia.com/en-in/autonomous-machines/ embedded-systems/jetson-nano/