Automatic Detector For Bikers With No Helmet Using Deep Learning

A PROJECT REPORT

Submitted by

AISWARYA KRISHNA (DMC17CS002)

BINDHYA T (DMC17CS023)

ANJANA RAJAN (DMC17CS018)

AKHILA P K (DMC17CS005)

to

the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of the Degree

of

Bachelor of Technology

In

Computer Science and Engineering

M·DIT KOZHIKODE

Department of Computer Science and Engineering

M.Dasan Institute of Technology (M-DIT)

Ulliyeri P.O, Kozhikode - 673620

MAY 2021

DECLARATION

We undersigned declare that the project report " Automatic Detector For Bikers With No Helmet Using Deep Learning ", submitted for the partial fulfillment of the requirements for the award of the degree of the Master of Technology under the university of the APJ Abdul Kalam Technological University, Kerala is a bonafide work , done by us under the supervision of the Ms. Ramitha M A (Asst Prof). This submission represents our ideas in our own words and where ideas or words of others have been included, We have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to ethics of academic honesty and integrity and not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and the university and can also be evoke penal action from the source which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other university.

Place :
Date :

Aiswarya Krishna

Bindhya T

Anjana Rajan

Akhila P K

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
M.DASAN INSTITUTE OF TECHNOLOGY (M-DIT), ULLIYERI
(Affliated to the APJ Abdulkalam Technological University)

CERTIFICATE

This is to certify that the report entitled " Automatic Detector For Bikers With No Helmet Using Deep Learning " by "AISWARYA KRISHNA" (DMC17CS002) "BINDHYA T" (DMC17CS023) "ANJANA RAJAN" (DMC17CS018) "AKHILA P K" (DMC17CS005) to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science Engineering is a bonafide of project carried by her under our guidance and supervision. This report in any form not been submitted to any other University or Institute of any purpose.

Project Guide Project Coordinator Head Of The Dept

ACKNOWLEDGEMENT

We would like to thank THE ALMIGHTYs mercy towards us over the years; it goes without saying that without them, We would not be where we are today.

We would like to place on record, sincere thanks to all those who have contributed to the successful completion of this project.

We recollect with of gratitude the valuable guidance and the whole hearted support of the project Coordinator Ms. Ramitha M A (Asst Prof) and the Head of the department Ms. Nithya V P , CSE Department for their valuable suggestions throughout the project work.

We wish to express the deep sense of gratitude and the thanks to the Principal Mr. Mahessan PM and Director Mr. H. Ahinus for the excellent facilities and continual support provided during the course study and project work.

We thank all the staff members of Computer Science and Engineering Department for their valuable suggestions.

We also thank all beloved friends for their encouragement, timely help and suggestions.

ABSTRACT

The success of digital image pattern recognition and feature extraction using a Convolutional Neural Network (CNN) or Deep Learning was recently acknowledged over the years. Researchers have applied these techniques to many problems including traffic offense detection in video surveillance, especially for the motorcycle riders who are not wearing a helmet. Several models of CNN were used to solve these kinds of problem but mostly required the image preprocessing step for extracting the Region of Interest (ROI) area in the image before applying CNN to classify helmet. In this paper, it proposed to apply another interesting method of deep learning called Single Shot MultiBox Detector (SSD) into helmet detection problem. This method is the state-of- the-art that is able to use only one single CNN network to detect the bounding box area of motorcycle and rider and then classify that biker is wearing or not wearing a helmet at the same time. The results of the experiment were surprisingly good. The classification accuracy of bikers not wearing a helmet was extremely high and the detection of the ROI of biker and motorcycle in the image can be done at the same time as the classification process.

ABBREVIATIONS

OS	Operating System
ROI	Region of interest
SSD	Single shot MultiBox detector
IP	Image processing
CNN	Convolutional Neural Network
CV	Computer vision
AI	Artificial intelligence
SVM	Support vector machine
HOG	Histogram of oriented gradients
ANN	Artificial neural network
ALPR	Automatic license plate recognition
DNN	Deep neural network
MAP	Mean average precision
OCR	Optical character recognition
YOLO	You only look oncevv
ML	Machine learning

Contents

List of Figures

List of Tables

1 INTRODUCTION

Building an automatic system like this bring researchers into areas of Image Processing Computer Vision, and Artificial Intelligence. Because most data from the traffic control system usually came in a format of video surveillance data (image and video) that require the technique to analyze an image data such as image recognition, pattern matching, and image segmentation. To detect the bikers who don't wear a helmet, it need methods to detect the photo of motorcycle and driver from the image and then detect an area of the biker head before classify that this person is wearing a helmet or not. Several IP and AI approaches have been using to solve this problem, for example, Fourier transformation, Support Vector Machine, Histogram of Oriented Gradients, Artificial Neural Network etc. But the advancement in another technique called a Convolutional Neural Network has proved to be the better method in the area of image recognition and computer vision.

The model of Alex net is a Deep Layer Convolutional Neural Network consisted of 650,000 neurons and 60 million parameters with five convolutional layers and 1000-way softmax layer. This model was challenged in the Image Net Large-Scale Visual Recognition Challenge and won the competition proved that CNN method will likely be the best technique to solve the most image recognition problem in this era. After the success of Alex Net, several CNN models have been introduced and tried to achieve better performance than the pioneer one. For example, VGG, Google Net or Inception, and Mobile Nets. But most of these models can use only to categorize or recognize one object from the image not for multiple objects. Another technique needs to include into these models for adding image segmentation feature by drawing the box on the area of a possible object in an image before categorization. CNN model to be able to detect multiple objects in one single frame of the image. The examples of this approach are Faster R-CNN, Single Shot Multi-Box Detector, and YOLO.

2 LITERATURE SURVEY

2.1 SSD:Single Shot Multi Box Detector

Current state-of-the-art object detection systems are variants of the following approach: hypothesize bounding boxes ,or features for each box, and apply a high quality classifier.This pipeline has prevailed on detection benchmarks since the Selective Search work through the current leading results on PASCALVOC, COCO, and ILSVRC detection all based on Faster R-CNN albeit with deeper features such as. While accurate, these approaches have been too computationally intensive for embedded systems and, even with high-end hardware, too slow for real-time applications.

The first deep network based object detector that does not re sample pixels or features for bounding box hypotheses and is as accurate as approaches that do. This results in a significant improvement in speed for high-accuracy detection. The fundamental improvement in speed comes from eliminating bounding box proposals and the subsequent pixel or feature resampling stage but by adding a series of improvements, it manage to increase the accuracy significantly over previous attempts. Improvements include using a small convolutional filter to predict object categories and offsets in bounding box locations, using separate predictors for different aspect ratio detections, and applying these filters to multiple feature maps from the later stages of a network in order to perform detection at multiple scales. With these modifications especially using multiple layers for prediction at different scale. It can achieve high-accuracy using relatively low resolution input, further increasing detection speed. While these contributions may seem small independently, this is a larger relative improvement in detection accuracy than that from there cent, very high- profile work on residual networks. Furthermore, significantly improving the speed of high-quality detection can broaden the range of settings where computer vision is useful.

2.1.1 PASCALVOC2007

On this dataset, it compare against Fast R-CNN and Faster R-CNN on VOC2007 test. All methods fine-tune on the same pre-trained VGG16 network. The conv4 3, conv7, conv8 2, conv9 2, conv10 2, and conv11 2 to predict both location and confidences. Initialize the parameters for all the newly added convolutional layers with the "Xavier" method. For conv 4 3, conv 10 2 and conv 11 2, default boxes at each feature map location. For all other layers, it put 6 default boxes as described in. Since, as pointed out in conv 4 3 has a different feature scale compared to the other layers, it use the L2 normalization technique introduced in to scale the feature norm at each location in the feature map to 20 and learn the scale during back propagation. When training on VOC2007 train Val, Table 1 shows that low resolution SSD300 model is already more accurate than Fast R-CNN. When it train SSD on a larger input image, it is even more accurate, surpassing Faster R-CNN .SSD300 is already better than Faster R-CNN and that SSD512 better. If it take models trained on COCO train Val 35k as described and fine-tuning dataset with SSD51. To understand the performance of two SSD models in more details. . SSD can detect various object categories with high quality. The majority of its confident detections are correct. The recall is around and is much higher with "itak" criteria.

SSD has less localization error, indicating that SSD can localize objects better because it directly learns to regress the object shape and classify object categories of using two decoupled steps SSD is very sensitive to the bounding box size. In other words, it has much worse performance on the smaller. Objects than bigger objects. This is not surprising because those small objects may not even have any information at the very top layers. Increasing the input size can help improve detecting small objects, but there is still a lot of room to improve. On the positive side, it can clearly see that SSD performs really it all on large objects.

Data augmentation is crucial. Fast and Faster R-CNN use the original image and the horizontal flip to train. It use a more extensive sampling strategy, similar to YOLO. It do not know how much sampling strategy will benefit Fast and Faster R-CNN, but they are likely to benefit less because they use a feature pooling step during classification that is relatively robust to object translation by design.

2.1.2 PASCALVOC2012

The use of same settings as those used for basic VOC2007 experiments above, except that it use VOC2012 train Val and VOC2007 train Val and test (21503 images) for training, and test on VOC2012 test (10991 images). It train the models with 103 learning rate for 60k iterations, then 104 for 20k iterations. Table 4 shows the results of SSD300 and SSD5124 model. It see the same performance trend as it observed on VOC2007 test. SSD300 improves accuracy over Fast/Faster RCNN. By increasing the training and testing image, it are 4.5YOLO, SSD is significantly more accurate, likely due to the use of convolutional default boxes from multiple feature maps and matching strategy during training. When fine- tuned from model strained on COCO, SSD512 achieves 80.0higher than Faster R-CNN.

2.1.3 COCO

To further validate the SSD framework, it trained SSD300 and SSD512 architectures on the COCO dataset. Since objects in COCO tend to be smaller than PASCAL VOC, it use smaller default boxes for all layers., the train Val 35k for training. First train the model with 103 learning rate for 160k iterations, and then continue training for 40k iterations with 104 and 40k iterations with 105 Faster R-CNN is more competitive on smaller objects.

2.1.4 Data Augmentation For Small Object Accuracy

Without a follow-up feature resampling step as in Faster R-CNN, the classification task for small objects is relatively hard for SSD, as demonstrated in analysis. The data augmentation strategy helps to improve the performance dramatically, especially on small datasets such as PASCAL VOC. The random crops generated by the strategy can be thought of as a" zoom in" operation and can generate many larger training examples. To implement a" zoom out" operation that creates more small training examples, it first randomly place an image on a canvas of 16× of the original image size filled with mean values before it do any random crop operation. Because it have more training images by introducing this new "expansion" data augmentation trick, it have to double the training iterations. This result underscores the importance of the data augmentation strategy for the final model accuracy. An alternative way of improving SSD is to design a better tiling of default boxes so that its position and scale are better aligned with the receptive field of each position on a feature map. It leave this for future work.

Figure 1: Detection example on coco test-dev with SSD512 model

2.1.5 Advantages

- Multi-scale feature maps improve the detection of objects at different scale.
- Design better default boundary boxes with help of accuracy.
- It has no delegated region proposal network and predicts the boundary boxes and the classes directly from feature maps in one single pass.
- SSD makes more predictions and has a better coverage on location, scale and aspect ratios.

2.1.6 Disadvantages

- SSD doesn't perform it'll on small-object detections.
- SSD assumes that small object detection only relies on fine-grained local features, thus ignoring context information.
- Centers of receptive fields for each feature map output, does not match the centers of corresponding anchors.

2.2 Helmet Detection On Motorcyclist Using Image Descriptors And Classifiers

The use of motorcycle accidents has rapidly increased. Although the helmet is the main safety equipment of motorcyclist, but many drivers do not use it. A two stage strategy was developed namely the detection of motorcycles and the detection of helmet .in system of vehicle segmentation and classification. Eight three-dimensional models tire created to classify the objects. These models itre calculated based on the size of the vehicle, which varies depending on the type of vehicle. The created models itre as follows: cycle, small car, car, minibus, closed truck, open truck, bus and pedestrian. Models are generated for each captured vehicle, which are compared with the models created for each class of vehicle. The model that is closer to the model of the captured vehicle defines the vehicle class. A disadvantage of this study is that a single model is employed for motorcycles and bicycles. In addition, only geometric information is utilized to classify the vehicles, which has been proven to be insufficient for describing these types of objects. Another disadvantage of this method is that some parameters, such as the camera height and angle and the lens focal distance, should always have the same values. If these parameters are changed, new models should be created, as no vehicle would correspond to the models. Some public roads that did not have a location, such as described in this study, to position the camera, itre unable to operate the system.

Vehicle segmentation and classification, algorithms for the background calculation and tracking of objects, descriptors and classifiers that exhibited reasonable hit rates and low processing times itre selected from the literature achieving an accuracy of 0.9778. In the stage of detection of helmet use, algorithms for the extraction of features in images and classification algorithms itre employee. The MLP classifier that incorporated the HOG descriptor obtained the best results, with an accuracy of 0.9137. The results are promising but can be improved. An important step for improving the results is the stage of image capturing, which should produce better quality images. The images of database1 itre not employed in the stage of helmet detection due to their low quality. Future studies should focus on the detection and recognition of the registration plate of the vehicle. A better quality image is necessary to recognize the characters on the plate. Hybrid descriptors itre not employed in the vehicle segmentation stage. They will be employed to improve the results. The use of algorithms for feature selection can

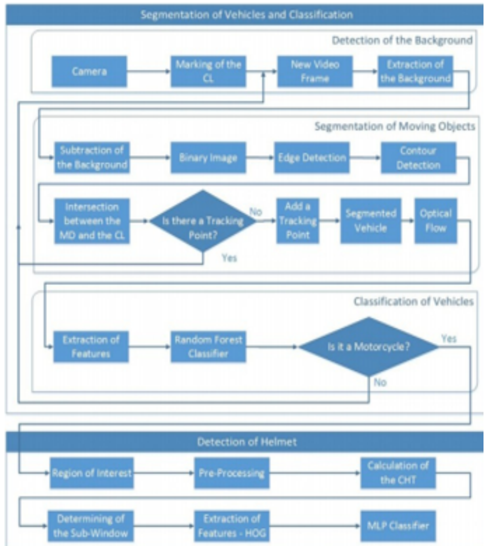

Figure 2: Diagram of all steps of the proposed system

Figure 3: Vehicle Segmentation

be evaluated to increase the obtained hit rates.

2.2.1 Detection of the background

The detection of the background is critical for the development of this study. The main objective of this step is to obtain an image that will be used to detect moving objects. The pixels that belong to the scene background and moving pixels can be extracted from an image of the background and the current frame. All moving objects are potential vehicles. The calculation of the background image was performed using the AGMM algorithm.

2.2.2 Segmentation of moving objects

The segmentation of moving objects on the scene facilitates the evaluation of the objects of interest in the image. The method of background subtraction receives all points that itre altered in a scene. The subtraction bitten the current frame and the background after binarization .For the segmentation of moving objects in the proposed

Figure 4: Example of the determination of the CL

system, a CL, which is marked by the user in the calibration stage of the system, must be defined. The CL is only defined once-when the algorithm begins operating. The CL should be marked approximately perpendicular to the road, where the vehicles circulate, which ensures that any vehicle that uses the road will cross the CL. When a vehicle crosses the CL, the process of segmentation of moving objects begins and the image frame is captured. Figure 2d shows an example of a segmented vehicle. The next step involves the use of the Otsu threshold for the image that results from the subtraction. This threshold is applied to obtain a binary image. The Sobel algorithm is applied to the binary image for edge detection. A morphological closing operator is employed to remove some noises from the image, such as small regions that are not vehicles. The next step is the detection of the object contour. This detection is performed based on the image edges. The algorithm proposed by Suzuki and be was utilized in this stage. The maximum points of the contour are used to cut the captured vehicle from the frame and obtain the image of the vehicle.

Therefore, a tracking algorithm that ensures that each vehicle is only counted once is needed. For each detected object (vehicle), the point of intersection bitten the main diagonal of the object and the CL is computed. This point is designated as the tracking point of the object. To reduce the computational processing time, only the objects that are detected and not marked are analyzed, that is, after the object is marked, it will not be analyzed in subsequent frames.

2.2.3 Vehicle classification

In the proposed system, the task of classifying the vehicles consists of differentiating the segmented objects into two classes: motorcycles and non-motorcycles. The features of the images are extracted using the WT and the image that was segmented in the previous stage. The WT was performed using one decomposition level, to get this parameter it tested values betiten 1 and 10. A feature vector is obtained for each generated image. This vector is used by the random forest classifier to determine which class the vehicle is associated. One of the requirements for a feature vector to be used in this problem is that it exhibits the same size for all images, regardless of their dimensions. To extract features with the WT, the images had to be resized as the number of features that is returned by the transform is sensitive to the image size. Therefore, the size of all image must be identical. The images itre resized to obtain a single size for all images.

2.2.4 Detection of the helmet

Determining the RoI is an important step of the proposed system. The use of this he reduction of the area in which the search will be performed, which implies less. Processing time and a greater precision of the results compared with the complete image. The RoI is a region of the captured image in the vehicle segmentation stage. As the proposed system is interested in the detection of motorcyclists without helmets, the head region of the motorcyclist must be located completely inside the RoI. This value was empirically selected via tests with the images obtain in the vehicle segmentation stage. The head region is typically located in the upper 1/5 of the image. The size of the RoI was tested across the image database. In all images of motorcyclists, the head region is located within the selected RoI.

2.2.5 Advantages

- It helps to find motorcyclist without helmet.

2.2.6 Disadvantages

- A single model is employed for motorcyclist and bicycles.
- Only geometric information is utilized to classify the vehicles.
- Some parameters such as the camera height and angle and the lens focal distance should always has the same values. If these parameters are changed new model should be created as no vehicle would correspond to the models.
- Some public roads that did not have a location, such as described in this study, to position camera, itre unable to operate the system.

2.3 Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks

CNNs itre computationally expensive as originally developed in their cost has been drastically reduced thanks to sharing convolutions across proposals. The latest incarnation, Fast R-CNN, achieves near real-time rates using very deep networks when ignoring the time spent on region proposals. Now, proposals are the computational bottleneck in state-of-the-art detection systems. Region proposal methods typically rely on inexpensive features and economical inference schemes. Selective Search, one of the most popular methods, greedily merges super pixels based on engineered low-level features. Yet when compared to efficient detection networks, Selective Search is an order of magnitude slot, at 2s per image in a CPU implementation. Edge Boxes currently provides the best tradeoff betiten proposal quality and speed, at 0.2s per image. Nevertheless, the region proposal step still consumes as much running time as the detection network. One may note that fast region-based CNNs take advantage of GPUs, while the region proposal methods used in research are implemented on the CPU, making such runtime comparisons inequitable.

An obvious way to accelerate proposal computation is to re-implement it for the GPU. This may be an effective engineering solution, but re-implementation ignores the down- stream detection network and therefore misses important opportunities for sharing computation. Observation is that the convolutional feature maps used by region-based detectors, like Fast R-CNN, can also be used for generating region proposals. On top of these conv features, it construct RPNs by adding two additional conv layers: one that encodes each conv map position into a short feature vector and a second that, at each conv map position, outputs an objectiveness score and regressed bounds for k region proposals relative to various scales and aspect ratios at that location.

RPNs are thus a kind of fully-convolutional network and they can be trained end-to-end specifically for the task for generating detection proposals. To unify RPNs with Fast R-CNN object detection networks, simple training scheme that alternates betiten fine-tuning for the region proposal task and then fine-tuning for object detection, while keeping the proposals fixed. This scheme converges quickly and produces a unified network with conv features that are shared betiten both tasks. It evaluate method on the PASCAL VOC.

2.3.1 Implementation Details

It train and test both region proposal and object detection networks on single-scale images. It re-scale the images such that their shorter side is s = 600 pixels. Multi-scale feature extraction may improve accuracy but does not exhibit a good speed-accuracy trade- off. It also note that for ZF and VGG nets, the total stride on the last conv layer is 16 pixels on the re-scaled image, and thus is 10 pixels on a typical PASCAL image (500×375). Even such a large stride provides good results, though accuracy may be further improved with a smaller stride. For anchors, it use 3 scales with box areas of 128 2, 256 2, and 512 2 pixels, and 3 aspect ratios of 1:1, 1:2, and 2:1. It note that algorithm allows the use of anchor boxes that are larger than the underlying receptive field when predicting large proposals. Such predictions are not impossible- one may still roughly infer the extent of an object if only the middle of the object is visible. With this design, solution does not need multi-scale features or multi-scale sliding windows to predict large regions, saving considerable running time. (Right) shows the capability of method for a wide range of scales and aspect ratios. The table below shows the learned average proposal size for each anchor using the ZF net (numbers for s = 600).

The anchor boxes that cross image boundaries need to be handled with care. During training, it ignore all cross-boundary anchors so they do not contribute to the loss. For a typical 1000 × 600 image, there will be roughly 20k (60 × 40 × 9) anchors in total. With the cross-boundary anchors ignored, there are about 6k anchors per image for training. If the boundary-crossing outliers are not ignored in training, they introduce large, difficult to correct error terms in the objective, and training does not converge. During testing, hoitver, it still apply the fully- convolutional RPN to the entire image. This may generate cross-boundary proposal boxes, which it clip to the image boundary.

Some RPN proposals highly overlap with each other. To reduce redundancy, it adopt non- maximum suppression (NMS) on the proposal regions based on their class scores. It fix the IOU threshold for NMS at 0.7, which leaves us about 2k proposal regions per image.

2.3.2 Experiments

It comprehensively evaluate method on the PASCAL VOC 2007 detection benchmark. This dataset consists of about 5k train Val images and 5k test images over 20 object categories. It also provide results in the PASCAL VOC 2012 benchmark for a few models. For the Image Net pre-trained network, it use the "fast" version of ZF net [23] that has 5 conv layers and 3 fc layers, and the public VGG-16 model 5 [19] that has 13 conv layers and 3 fc layers. It primarily evaluate detection mean Average Precision (MAP), because this is the actual metric for object detection (rather than focusing on object proposal proxy metrics).proposals by the "fast" mode. For Edge Boxes (EB) [24], it generate the proposals by the default EB setting tuned for IOU. SS has an MAP of 58.7and EB has an MAP of 58.6with an MAP of 59.9faster detection system than using either SS or EB because of shared conv computations; the proposals also reduce the region-wise cost.

2.3.3 Ablation Experiments

To investigate the behavior of RPNs as a proposal method, it conducted several ablation studies. First, it show the effect of sharing conv layers betiten the RPN and Fast- CNN detection network. To do this, it stop after the second step in the 4-step training process. Using separate networks reduces the result slightly to 58.7unshared, Table 1). It observe that this is because in the third step when the detector- tuned features are used to fine-tune the RPN, the proposal quality is improved. Next, it disentangle the RPN's influence on training the Fast R-CNN detection network. For this purpose, it train a Fast R-CNN model by using the 2k SS proposals and ZF net. It fix this detector and evaluate the detection MAP by changing the proposal regions used at test- time. In these ablation experiments, the RPN does not share features with the detector.

Replacing SS with 300 RPN proposals at test-time leads to an MAP of 56.8The loss in Map is because of the inconsistency betiten the training/testing proposals. This result serves as the baseline for the following comparisons. Somewhat surprisingly, the RPN still leads to a competitive result (55.1proposals at test-time, indicating that the top-ranked RPN proposals are accurate. On the other

Extreme, using the top-ranked 6k RPN proposals (without NMS) has a comparable MAP (55.2alarms. Next, it separately investigate the roles of RPN's class and register outputs by turning off either of them at test-time.

When the class layer is removed at test-time (thus no NMS/ranking is used), it randomly sample N proposals from the unscored regions. The MAP is nearly unchanged with N = 1k (55.8that the class scores account for the accuracy of the highest ranked proposals.

On the other hand, when the register layer is removed at test time (so the proposals become anchor boxes), the MAP drops to 52.1quality proposals are mainly due to regressed positions. The anchor boxes alone are not sufficient for accurate detection.

Detection results on PASCAL VOC 2007 test set. The detector is Fast R-CNN andVGG-16. Training data: "07": VOC 2007 train Val, "07+12": union set of VOC 2007 train Val and VOC 2012 train Val. For RPN, the train-time proposals for Fast R-CNN are 2k. This was reported in using the repository provided by this paper, this number is higher (68.0$\pm$0.3in six run).

It also evaluate the effects of more powerful networks on the proposal quality of RPN alone. It use VGG-16 to train the RPN, and still use the above detector of SS+ZF. The MAP improves from 56.8promising result, because it suggests that the proposal quality of RPN+VGG is better than that of RPN+ZF. Because proposals of RPN+ZF are competitive with SS (both are 58.7better than SS.

2.3.4 Detection Accuracy and Running Time of VGG-16

Using RPN+VGG, the Fast R-CNN result is 68.5higher than the SS baseline. As shown above, this is because the proposals generated by RPN+VGG are more accurate than SS. Unlike SS that is pre-defined, the RPN is actively trained and benefits from better networks. For the feature-shared variant, the result is 69.9train the RPN and detection network on the union set of PASCAL VOC 2007 trainval and 2012 trainval, following. The MAP is 73.23), method has an MAP of 70.4VOC 2012 trainval, following.

In Table 4 it summarize the running time of the entire object detection system. SS takes 1-2 seconds depending on content (on average 1.51s), and Fast R-CNN with VGG-16 takes 320ms on2k SS proposals (or 223ms if using SVD on fc layers). System with VGG-16 takes in total198ms for both proposal and detection. With the conv features shared, the RPN alone only takes 10ms computing the additional layers. Region-wise computation is also low, thanks to the proposals (300). System has a frame-rate of 17 fps with the ZF net.

2.3.5 Analysis of Recall-to-IoU

Next it compute the recall of proposals at different IoU ratios with ground-truth boxes. It is noteworthy that the Recall- to-IoU metric is just loosely [9, 8, 1] related to

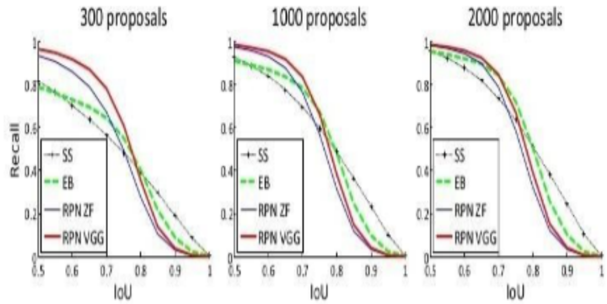

Figure 5: Recall vs IOU overlap ratio on the PASCAL VOC 2007

the ultimate detection accuracy. It is more appropriate to use this metric to diagnose the proposal method than to evaluate it.

2.3.6 One-stage Detection vs. Two-Stage Proposal+Detection

The Over Feat paper proposes a detection method that uses repressor and classifiers on sliding windows over conv feature maps. Over Feat is a one-stage, class- specific detection pipeline, and is a two-stage cascade consisting of class-agnostic proposals and class-specific detections. In Over Feat, the region-wise features come from a sliding window of one aspect ratio over a scale pyramid. These features are used to simultaneously determine the location and category of objects. In RPN, the features are from square (3×3) sliding windows and predict proposals relative to anchors with different scales and aspect ratios.

Though both methods use sliding windows, the region proposal task is only the first stage of RPN+ Fast R-CNN—the detector attends to the proposals to refine them. In the second stage of cascade, the region-wise features are adaptively pooled from proposal boxes that more faithfully cover the features of the regions. It believe these features lead to more accurate detections. To compare the one-stage and two-stage systems, it emulate the Over Feat system (and thus also circumvent other differences of implementation details) by one-stage Fast R-CNN. In this system, the "proposals" are dense sliding windows of 3 scales (128, 256, and 512) and 3 aspect ratios (1:1, 1:2, 2:1). Fast R-CNN is trained to predict class-specific scores and regress box locations from these sliding windows. Because the Over Feat system uses an image pyramid, it also evaluate using convfeatures extracted from 5 scales. It use those 5 scales. It compares the two-stage system and two variants of the one-stage system. Using the ZF model, the one-stage system has an MAP of 53.9Similar observations are reported in [5, 13], where replacing SS region proposals with sliding windows leads to 6system is solitary it has considerably more proposals to process.

2.3.7 Advantages

- Fast RCNN is faster than RCNN.
- Here convolutional operation is done only once per image and feature map generated from it.

2.3.8 Disadvantages

- Most of the time taken by faster r-cnn during detection is a selective search region proposal generation algorithm. Hence it is bottle neck of this architecture which was dealt with in faster r-cnn.

2.4 Helmet Detection Using Machine Learning And Automatic License Plate Recognition

The main safety equipment of motorcyclist is the helmet. The helmet protects the motorcyclist against accidents. Although the helmet use is mandatory in many countries, there are motorcyclists that do not use it or use it incorrectly. Over the past years many works have been carried out in traffic analysis, including vehicle detection and classification, and helmet detection.

Intelligent traffic systems itre implemented using computer vision algorithms, such as: background and foreground image detection to segment the moving objects in scene and image descriptors to extract features. Computational intelligence algorithms are used too, like machine learning algorithms to classify the objects.

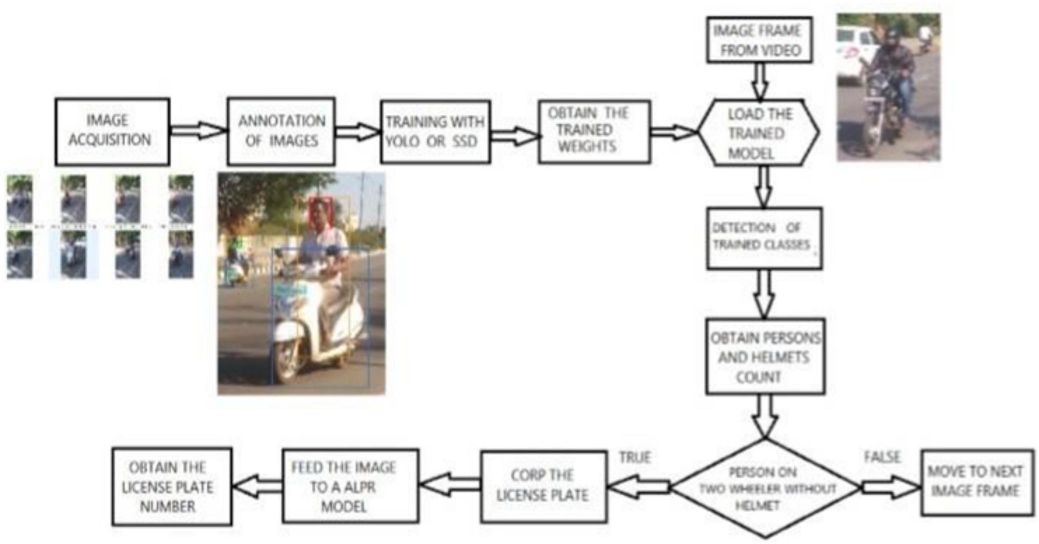

Figure 6: proposed methodolgy

It also makes predictions with a single network evaluation unlike systems like R-CNN which require thousands for a single image. This makes it extremely fast, more than 1000x faster than R-CNN and 100x faster than Fast R- CNN. Object detection is the craft of detecting instances of a certain class, like animals, humans and many more in an image or video. The Pre-Existing Object Detection API makes it easy to detect objects by using retrained object detection models. But these models detect several Objects which are of no use to us, therefore in order to detect the necessary classes a custom object detector becomes necessary.

In order to implement helmet detection and number plate recognition and extraction, 5 objects need to be detected. The objects are – Helmet, No Helmet, Motorbike, Person (sitting on the bike) and License Plate. There is a need to create a custom object detection model that is capable of detecting these objects. A collection of images containing the objects of the classes to be detected are used as a Dataset. This dataset is then used to train the custom model. Once the model has been trained, it can be used to detect these custom objects.

The training is done by feeding all the captured images with their annotations. The model extracts the features of each class from every image with the help of ground truth of the required classes. For extracting the features and storing them to recognize those features from other images, it use a deep learning, classifier based on the convolutional neural networks. When an image is given to this trained model the detection of the retrained class is necessary. A few images are taken as an example to show the detection capability of the custom trained model.

2.4.1 Helmet Detection

The annotated images are given as input to YOLOv3 model to train for the custom classes. The tights generated after training are used to load the model. Once this is done, an image is given as input. The model detects all the five classes trained. From this it obtain the information regarding person riding motorbike. If the person is not staring a helmet, then it can easily extract the other class information of the rider. This can be used to extract the license plate.

2.4.2 License Plate Extraction

Once the helmetless rider is detected, the associated person class is detected. This is done by finding whether the coordinates of the no helmet class lie inside the person class or not. Similarly, the same steps are followed to detect the associated motorbike and license plate. Once the coordinates of the License plate are found, it is cropped and saved as a new image.

2.4.3 License Plate Recognition

The extracted license plate is given to an Optical Character Recognition (OCR) model. The OCR recognizes text in the given image and outputs the recognized strings

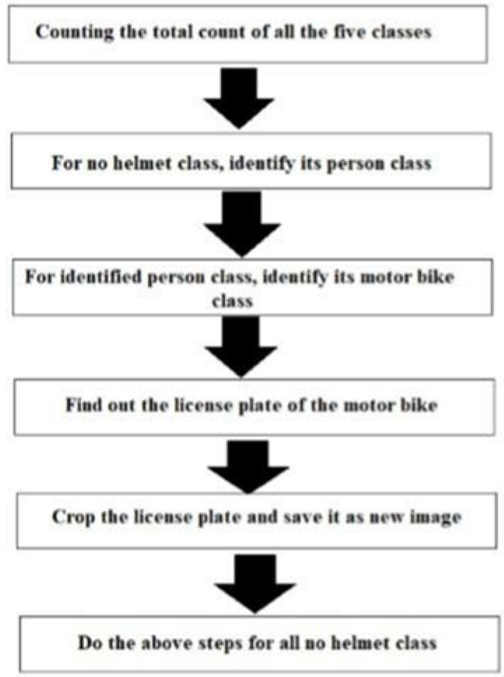

Figure 7: Helmet detection

Figure 8: license plate extraction

Figure 9: license plate recognition

in the machine-encoded text. The OCR module within will output a list of predicted license plate numbers along with a confidence value. The confidence value indicates how confident it is in recognizing the given license plate accurately. Then, the license plate recognized with highest confidence value is stored in a text file for further use.

2.4.4 Real Time Implementation

Using itbcam

The itbcam can be used as the input device to receive the image frames for object detection in real-time. Since it are using YOLOv3-tiny model, it supports up to 220 fps processing speed.

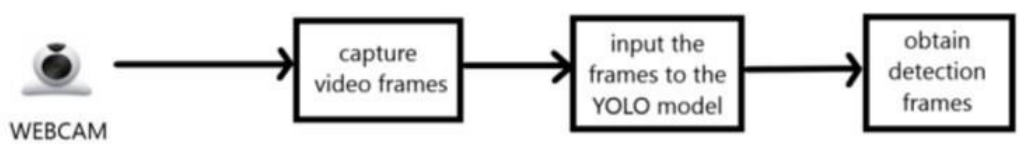

Figure 10: Real Time Implementation using itbcam

Using IP Itbcam for Mobile

Mobile camera can be used as the input rather than using the itbcam. This can open up a lot of possibilities as mobile can be carried and can cover different angles. Doing all this in real-time is an added advantage. So, from this not only CCTV footages but a handheld device can be used for obtaining the footage. Also, the footage from mobile being up-close can provide a clearer and more readable number plate for the OCR to give out an accurate number

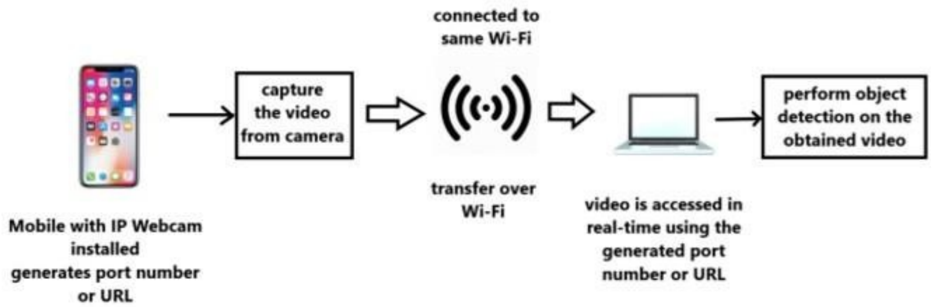

Figure 11: Real time implementation using IP itb cam for mobile

2.4.5 Advantages

- All the libraries and software used in the project are open source and hence is very flexible and low cost.
- It is built to solve the problem of non-efficient traffic management

2.4.6 Disadvantages

- It only use geometric features to verify if any safety helmet exists in the set
- Geometric features are not enough to find helmet

2.5 Automatic detection of bikers without helmet using survillance videos in real-time

In this paper, it propose an approach for automatic detection of bike-riders without helmet using surveillance videos in real time. Governments have made it a punishable offense to ride bike without helmet and have adopted manual strategies to catch the violators. Existing video surveillance based methods are passive and require significant human assistance. Automation of this process is highly desirable for reliable and robust monitoring of these violations as it'll as it also significantly reduces the amount of human resources needed.

Two-wheeler is a very popular mode of transportation in almost every country. Hoitver, there is a high risk involved because of less protection. To reduce the involved risk, it is highly desirable for bike-riders to use helmet. Observing the usefulness of helmet, Governments have made it a punishable offense to ride a bike without helmet and have adopted manual strategies to catch the violators. Hoitver, the existing video surveillance based methods are passive and require significant human assistance.

In general, such systems are infeasible due to involvement of humans, whose efficiency decreases over long duration. Automation of this process is highly desirable for reliable and robust monitoring of these violations as it'll as it also significantly reduces the amount of human resources needed. Also, many countries are adopting systems involving surveillance cameras at public places. So, the solution for detecting violators using the existing infrastructure is also cost-effective. Hoitver, in order to adopt such automatic solutions certain challenges need to be addressed.

2.5.1 Real-time Implementation

Processing significant amount of information in a time constraint manner is a challenging task. As such applications involve tasks like segmentation, feature extraction classification and tracking, in which a significant amount of information need to be processed in short duration to achieve the goal of real-time implementation.

2.5.2 Occlusion

In real life scenarios, the dynamic objects usually occlude each other due to which object of interest may only be partially visible. Segmentation and classification become difficult for these partially visible objects.

2.5.3 Direction of Motion

3-dimensional objects in general have different appearance from different angles. It is it'll known that accuracy of classifiers depends on features used which in turn depends on angle to some extent. A reasonable example is to consider appearance of a bikerider from front view and side view.

2.5.4 Temporal Changes in Conditions

Over time, there are many changes in environment conditions such as illumination, shadows, etc. There may be subtle or immediate changes which increase complexity of tasks like background modelling.

2.5.5 Quality of Video Feed

CCTV cameras capture low resolution video. Also, conditions such as low light, bad itather complicate it further. Due to such limitations, tasks such as segmentation, classification and tracking become even more difficult. As stated in, successful framework for surveillance application should have useful properties such as real-time performance, fine tuning, robust to sudden changes and predictive. Keeping these challenges and desired properties in mind, it propose a method for automatic detection of bike-riders without helmet using feed from existing security cameras, which works in real time.

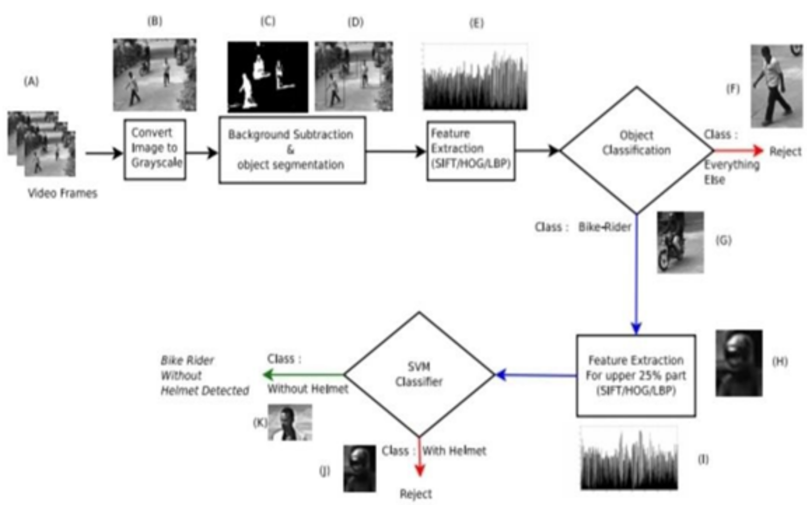

Figure 12: Automatic detection of bike riders without helmet

2.5.6 Advantages

- Computationally less expensive
- High accuracy
- Average time taken to process a frame is 11.57 ms
- More reliable than man.

2.5.7 Disadvantages

- Due to low resolution facial details not available
- Errors coming in background modelling due to presence of highly included object and included shadows.

3 PROBLEM STATEMENT

Previous work had applied the combination of the SSD method and the image classification model such as GoogLeNet and MobilesNet on the Thai License Plate Recognition problem and received a good result on that problem. The accuracy of detection and classification of a Thai character on the Thai License Plate was more than 90for both models (GoogLeNet and MobileNets). It gave us the confidence to take a further step to apply this technique in a helmet detection and classification issue.

In this paper, we proposed to solve the biker and helmet detection problem from video surveillance data by using CNN models and the SSD method. Some of CNN models have used in this experiment (VGG, GoogLeNet, and MobileNets) to compare the result. From our initial experiment, we found that the combination of MobileNets model and SSD has achieved the best accuracy compared to GoogLeNet and VGG in helmet detection problem and MobileNets was the method which requires the smallest size of the overall network.

4 PROPOSED SYSTEM

4.1 Introduction

The importance of automatic system in traffic control has been increased in the recent year. One goal is to improve the utilization of a traffic flow system, others are to reduce the cost of human labor and decrease the causes of an accident. In Thailand, one major reason for the accident is the motorcycle biker who drive without wearing a helmet. According to the law, every motorcyclist needs to wear a helmet while riding the motorcycle. But many bikers ignored and use their vehicle without safety equipment.

The policeman tried to control this problem manually but it is insufficient for the real situation. The ideal solution is to develop an electronic detection system that can be automated recognize this kind of problem without human cost. Building an automatic system like this bring researchers into areas of Image Processing (IP), Computer Vision (CV), and Artificial Intelligence (AI). Because most data from the traffic control system usually came in a format of video surveillance data (image and video) that require the technique to analyze an image data such as image recognition, pattern matching, and image segmentation. To detect the bikers who don't wear a helmet, we need methods to detect the photo of motorcycle and driver from the image and then detect an area of the biker head before classify that this person is wearing a helmet or not. Several IP and AI approaches have been using to solve this problem, for example, Fourier transformation , Support Vector Machine (SVM) , Histogram of Oriented Gradients (HOG), Artificial Neural Network etc. But the advancement in another technique called a Convolutional Neural Network (CNN) has proved to be the better method in the area of image recognition and computer vision. One historical method is "AlexNet" developed in 2010 .

The model of Alexnet is a Deep Layer Convolutional Neural Network consisted of 650,000 neurons and 60 million parameters with five convolutional layers and 1000-way softmax layer. This model was challenged in the ImageNet Large-Scale Visual Recognition Challenge 2010 (ILSVRC10) and won the competition proved that CNN method will likely be the best technique to solve the most image recognition problem in this era. After the success of AlexNet, several CNN models have been introduced and tried to achieve better performance than the pioneer one. For example, VGG , GoogLeNet or Inception , and MobileNets .

4.2 System Architecture

4.2.1 Motorcycle and Helmet Detection

Detection of motorcycle and helmet in an image is one of the challenging problems in the area of image processing. The issues are the shape of the object (motorcycle) in the image, the detection of people riding on a motorcycle or it just an empty vehicle with no biker, the location of the biker head, and the detection of a helmet at the head location of the biker.

Several steps of image processing needed to apply on the video image before it can detect the location of the motorcyclist with no helmet. For example on the previous work of P. Wonghabut et al (See Fig. 1), They needed to use several pre-processing techniques such as HAAR or HOG to detect the location of motorcycle in image first (step a-b), before cut off an area of a bikers in the image and classify that it is a motorcycle or not (step c-e), after that they need to find an area of the head location and cropped that area before they are able to detect that the biker is wearing a helmet or not (step f-h).

But the advancement in CNN and SSD techniques have promised us that they are able to do all these steps in only one single runtime that we are explaining further

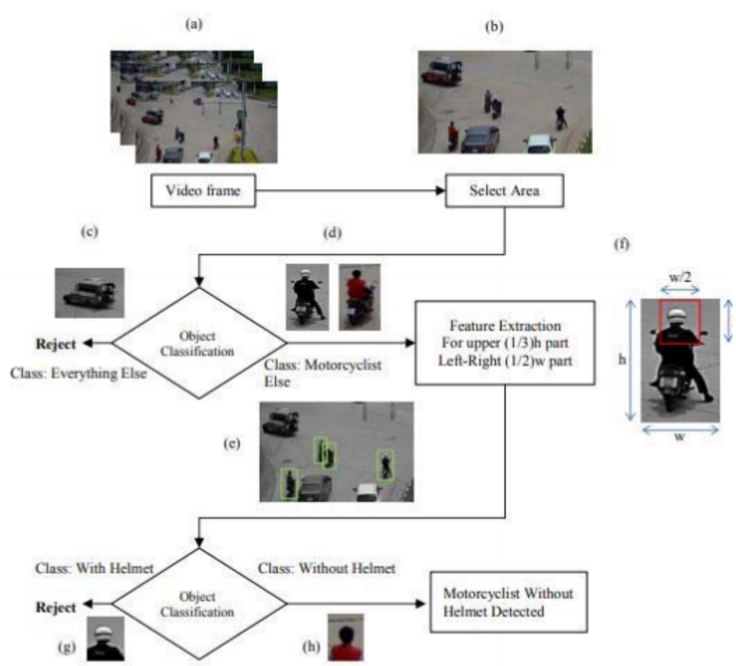

Figure 13: Proposed approach for detection of bike-riders without helmet

4.2.2 GoogLeNet (Inception V3)

GoogLeNet or Inception is the name of a CNN model created by Szegedy et al in 2014 . The first version of Inception consisted of 22 layers neural network with the combination layers of convolution, maxpooling, fully connected, softmax, and a special layer called inception module.

GoogleNet or Inception V1 was the winner on ILSVRC14 which was the competition on image recognition and the improved version came after year by year with a lot better performance. Now, version three or Inception V3, developed in 2016, has been the most used version for GoogLeNet models because its performance has surpassed its predecessor both in accuracy and time consuming

4.2.3 yolo algorithm

You Only Look Once (YOLO) is a method used for real time object detection. When YOLO gets an input image, it divides the image into S x S grid where S is an random value for the amount of grids, and where each cell grid produces bounding boxes and gives them a confidence score, each cell also does a class prediction. The same convolution neural network (CNN)(2.4) is used for the class prediction and the generated bounding boxes [26]. Each cell produces multiple bounding boxes, The class predictions and the produced bounding boxes happen in the last layers of the CNN. YOLO is an open source with different versions [26]. The strongest advantage YOLO has compared to similar methods is the speed of 45 frames per second. Other similar algorithms such as R-CNN[32] and DPM [33] has a much lower FPS. R-CNN has an FPS rate of 0.05 and DPM has a rate of 0.07. [34]. However, R-CNN and DPM contributes with better accuracy. There are other implementations of YOLO, in this paper YOLOv3-tiny is used as it is faster than YOLOv3 but less accurate. Humans sees an image and instantly know what objects are in the image, how the objects interact and where they are. For humans, it is fast and accurate which allow us to perform different complex tasks such as driving with little conscious. To add perceptions to autonomous cars, an object detection algorithm needs to be accurate and be able to give real-time scene information to the driver

steps

- YOLO first takes an input image:
- The framework then divides the input image into grids (say a 3 X 3 grid):
- Image classification and localization are applied on each grid. YOLO then predicts the bounding boxes and their corresponding class probabilities for objects (if any are found, of course).

YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Residual blocks

First, the image is divided into various grids. Each grid has a dimension of S x S. The following image shows how an input image is divided into grids

Figure 14: Residual blocks

In the image above, there are many grid cells of equal dimension. Every grid cell will detect objects that appear within them. For example, if an object center appears within a certain grid cell, then this cell will be responsible for detecting it.

Bounding box regression

A bounding box is an outline that highlights an object in an image.

Every bounding box in the image consists of the following attributes:

1- Width (bw)

2- Height (bh)

Class (for example, person, car, traffic light, etc.)- This is represented by the letter c. Bounding box center (bx,by) The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.

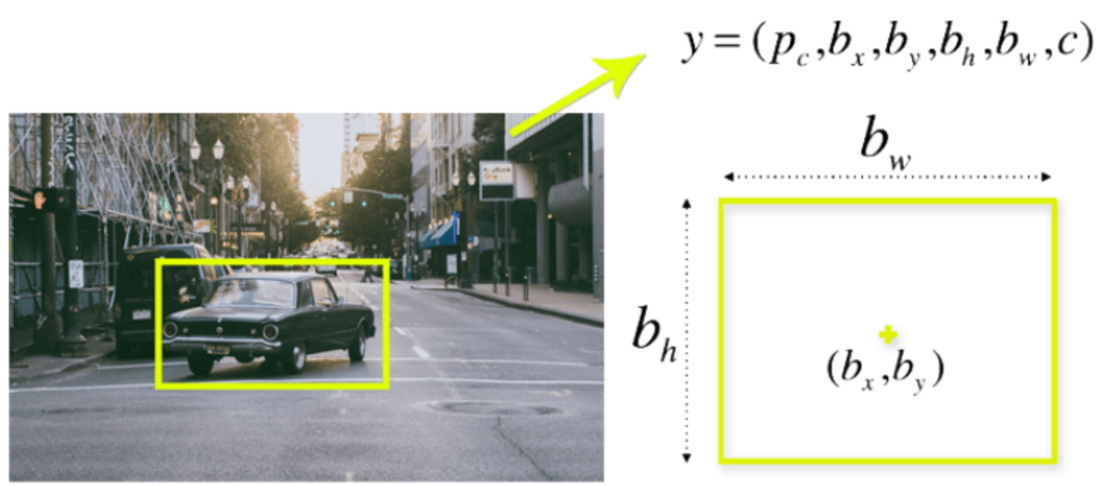

Figure 15: Bounding box regression

YOLO uses a single bounding box regression to predict the height, width, center, and class of objects. In the image above, represents the probability of an object appearing in the bounding box.

Intersection over union (IOU)

Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly.

Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box.

Combination of the three techniques

First, the image is divided into grid cells. Each grid cell forecasts B bounding boxes and provides their confidence scores. The cells predict the class probabilities to establish the class of each object.

For example, we can notice at least three classes of objects: a car, a dog, and a bicycle. All the predictions are made simultaneously using a single convolutional neural network.

Intersection over union ensures that the predicted bounding boxes are equal to the real boxes of the objects. This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of the objects (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly.

4.2.4 Mobilenet v2

MobileNets is another model of CNN which has proposed to decrease the size of the previous CNN model to make it available to use in a mobile platform. The idea is to replace the standard convolutional filters with two layers (Depthwise and Pointwise convolution) that build a smaller separable filter. The MobileNets network has achieved the good performance compare to another model and also come with the smallest size. Make it the good choice for the researcher to deal with the problem of large-scale data MobileNet-v2 is a convolutional neural network that is 53 layers deep.

It contains initial fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers. It has a drastically lower parameter count than the original MobileNet. MobileNets support any input size greater than 32 x 32, with larger image sizes offering better performance. MobileNet v2 uses lightweight depth wise convolutions to filter features in the intermediate expansion layer.

Mobilenetv2 introduces two features to the architecture,

1- linear bottlenecks between the layers.

2- shortcut connections between the bottlenecks. Bottleneck encode the model's intermediate inputs and outputs. Inner layer encapsulates the model's ability to transform from lower level concepts such as pixels to higher level descriptors such as image categories. Shortcuts enable faster training and better accuracy.

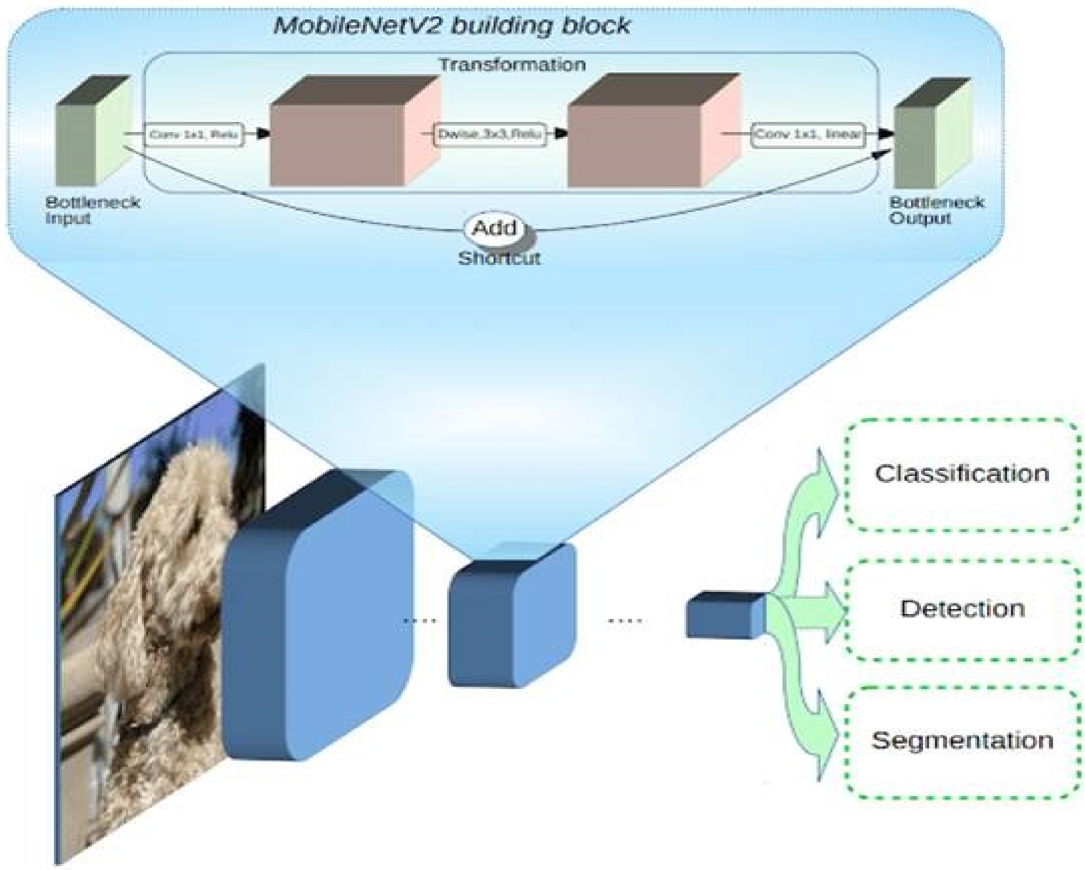

Figure 16: Architecture of the system

5 FEASIBILITY STUDY

All projects are feasible when given unlimited resources and infinite time. It is both necessary and prudent to evaluate the feasibility of a project at the earliest possible time. An estimate is made of whether the identified user needs may be satisfied using current software and hardware technologies. The study will decide if the proposed system will be cost effective from the business point of view and if it can be developed in the given existing budgetary constraints. The feasibility study should be relatively cheap and quick. The result should inform the decision of whether to go ahead with a more detailed analysis. Feasibility analysis is the procedure for identifying the candidate system, evaluating and electing the most feasible system. This is done by investigating the existing system in the area under investigation or generally ideas about a new system. It is a test of a system proposal according to its work ability, impact on the organization, ability to meet user needs, and effective use of resources. The objective of feasibility study is not to solve the problem but to acquire a sense of its scope. Feasibility analysis involves 6 steps

- Form a project and appoint a project leader
- Prepare system flowcharts
- Create a web site
- Weigh system performance and cost data
- Prepare and report final project directive to management The study is done in these phases

- Operational feasibility
- Technical feasibility
- Economical feasibility
- Behavioral feasibility
- Software feasibility
- Hardware feasibility

5.1 Operational Feasibility

Proposed projects are beneficial only if they can be turned into information system that will meet the organization's operating requirements. Simply stated, this test of feasibility asks if the system will work when it is developed and installed. Are there major barriers to implementation? Here are questions that will help test the operational feasibility of a project:

Is there sufficient support for the projects from management?

Are current business methods acceptable to the users?

Have the users been involved in the planning and development of the project?

Will the proposed system cause harm? The purpose of the operational feasibility study is to determine whether the new system will be used if it is developed and implemented. And whether there will be resistance from users that will undermine the possible application benefits

5.2 Technical Feasibility

A study of function performance and constraints may improve the ability to create an acceptable system. Technical feasibility is frequently the most difficult area to achieve at the stage of product engineering process. Considering that are normally associated with the technical feasibility include

- Development risk
- Resource availability
- Technology Technical feasibility study deals with the hardware as well as software requirements. The scope was whether the work for the project is done with the current equipments and the existing software technology has to be examined in the feasibility study. The outcome was found to be positive. In proposed system, data can be easily stored and managed using database management system software. The reports and results for queries can be generated easily. Thus, system is technically feasible.

5.3 Economical Feasibility

A cost evaluation is weighed against the ultimate income or benefit derived from the developed system or product. When compared to the advantage obtained from implementing the system its cost is affordable. Also the system is designed to meet the modifications required in the future. So, most of the required modifications can be done without much re-work. Proposed system was developed with the available resources. Since cost input for the software is almost nil the output of the software is always a profit. Hence software is economically feasible. In the existing system, manpower is required. In the proposed system, number of employee

5.4 Behavioral Feasibility

People are inherently resistant to changes and computer is known for facilitating the changes. An estimate should be made of how strongly the user staff reacts towards the developments of the computerized system. In the existing system more manpower is required and time factor is more. In proposed system, both man power and time factors are reduced and also unnecessary burden is reduced. Thus the remaining people are made to engage in some other important work. Therefore, the system is behaviorally feasible.

5.5 Software Feasibility

Even though software is developed in a very high software environment, it will be sup- ported by many other platform and environments with minimum changes.

5.6 Hardware Feasibility

The software can be developed with resource already existing. Here the consideration is that the existing hardware resources support the technologies that are to be used by the new system. No hardware was newly bought for the project and hence. Software is to achieve hardware feasibility.

6 REQUIREMENT SPECIFICATION

6.1 Software Requirements

- Python (Version 3.0+)
- PyCharm IDE
- WampServer
- SQLyog
- os:windows 8 or above

6.1.1 User Requirements

- Android (Version 5.0+)
- Google Chrome (or any browser)

6.2 Hardware Requirements

6.2.1 Development Requirements

- Intel 7th gen i3 / AMD Ryzen 3
- 4GB RAM
- Hard Disk 8GB or more

6.2.2 User Requirements

- Intel Pentium / AMD A6
- 512 MB RAM
- 112 GB Storage space

7 SYSTEM DESIGN

7.1 Data Flow Diagram

The data flow diagram is one of the most important tools used by system analyst. Data flow diagrams are made up of a number of symbols, which represent system components. Most data flow modeling methods use four kinds of symbols. These symbols are used for four kinds of system components. Processes, data store, data flow and external entities. Dataflow is represented by a thin line in the DFD and each data stored has a unique name and square represents external entities unlike detailed flow chart, data flow diagram do not supply detailed description of the modules but graphically describes a systems data and how the data interact with the system. An arrow identifies the data flow in motion. It is a pipe line through which information is flown like the rectangle in the flow chart. A rectangle with rounded corners stands for process that converts data into information's. An open-ended rectangle represents a data store, data at rest or a temporary repository of data. A square defines an external entity of system data.

Level 0

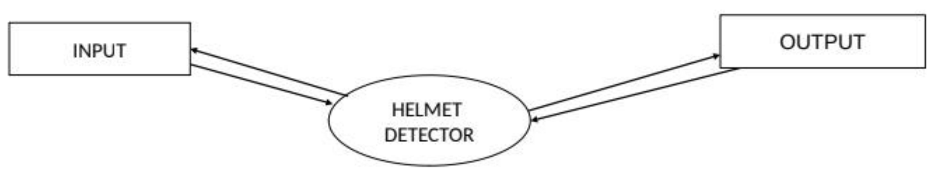

Figure 17: Level 0 DFD

Level 1.1

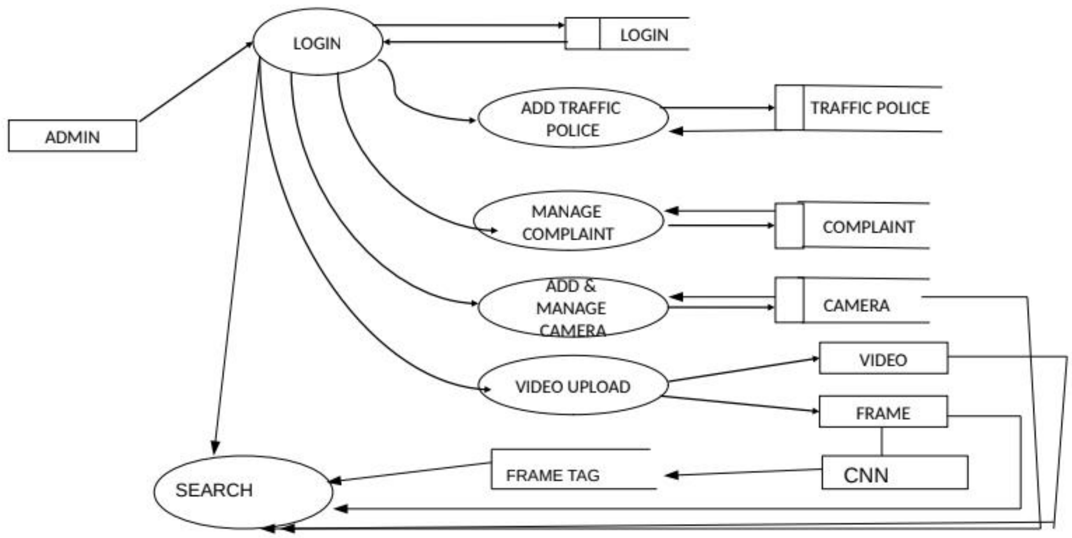

Figure 18: Level 1.1 DFD

Level 1.2

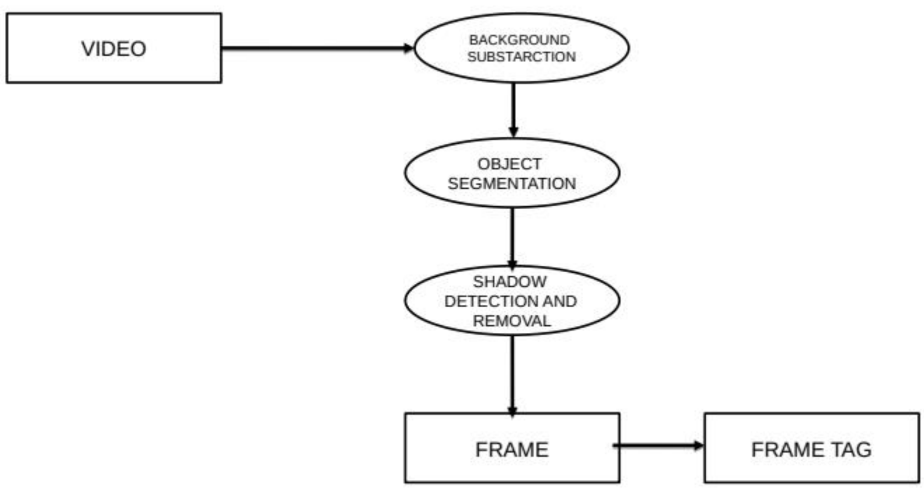

Figure 19: Level 1.2 DFD

Level 1.3

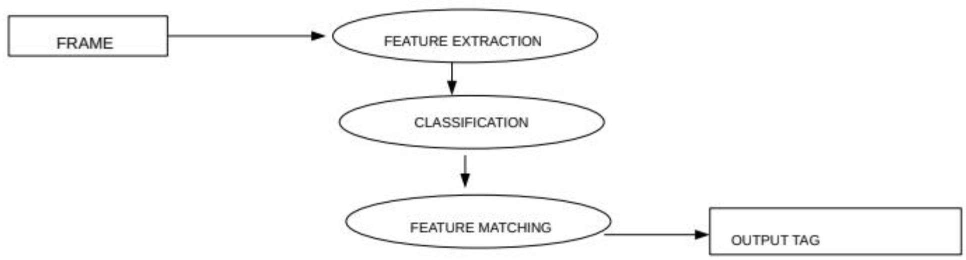

Figure 20: Level 1.3 DFD

Level 2.1

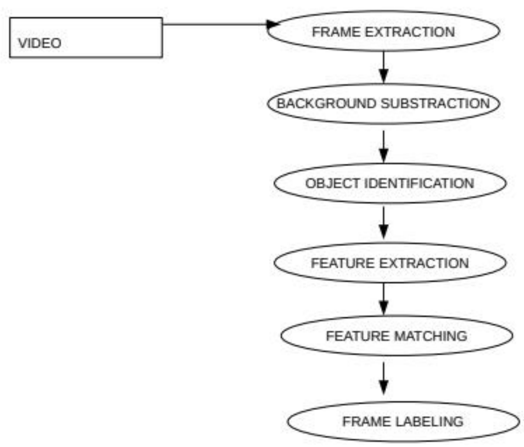

Figure 21: Level 2.1 DFD

Level 2.2

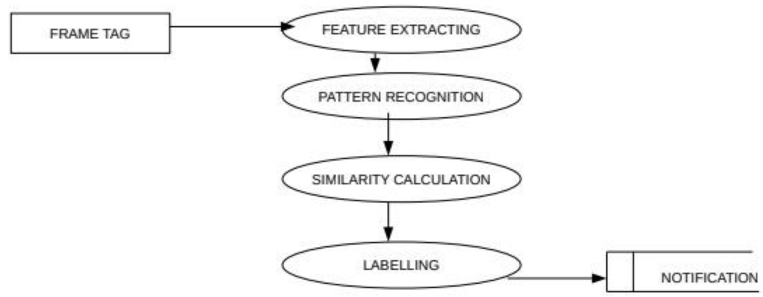

Figure 22: Level 2.2 DFD

7.2 UML Diagram

7.2.1 Sequence

A sequence diagram is an interaction diagram that emphasis the time ordering of the messages; a collaboration diagram is an interaction diagram that emphasizes the structural organization of the objects that send and receive messages. Sequence diagrams are typically associated with use case realizations in the logical view of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios.

7.2.2 Use Case Diagram

A use case diagram is a graphic depiction of the interactions among the elements of a system. A use case is a methodology used in system analysis to identify, clarify and organize system requirements' use case diagram at its simplest is representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. The purpose of use case diagram is to capture the dynamic aspect of a system. Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. So, when a system is analyzed to gather its functionalities use case is prepared and actors are identified.

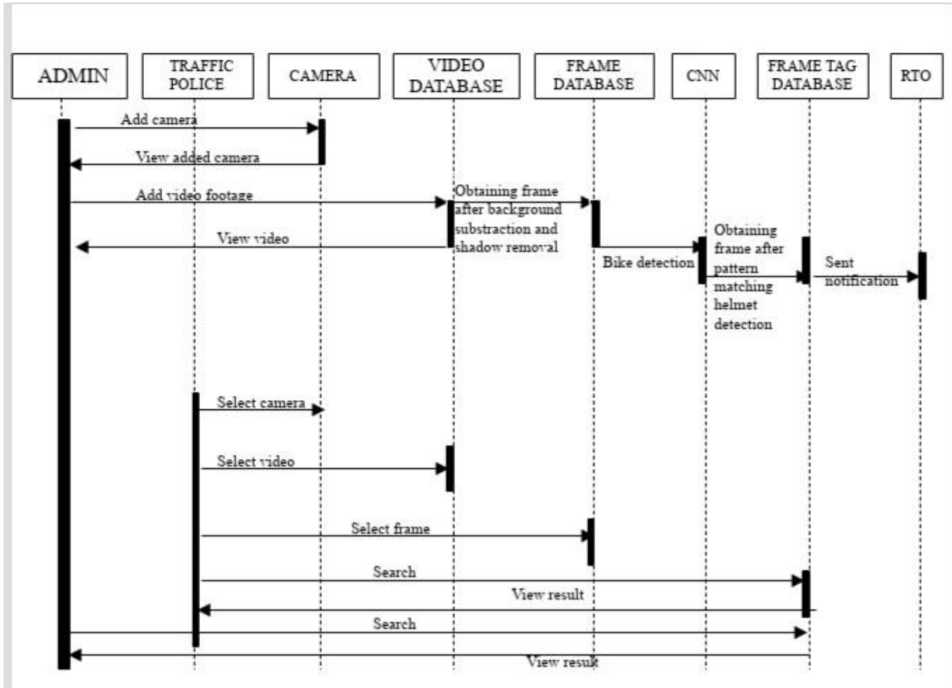

Figure 23: Sequence Diagram

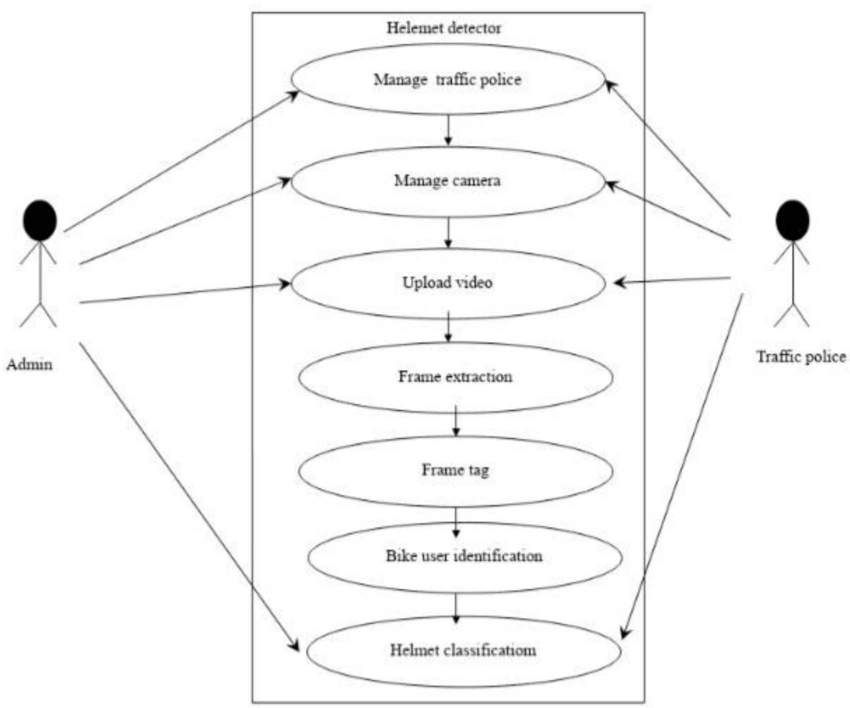

Figure 24: Use Case Diagram

7.2.3 Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes. Activity diagrams show the overall flow of control. Activity diagrams are constructed from a limited number of shapes, connected with arrows. The most important shape types:

- Rounded rectangles represent actions;
- Diamonds represent decisions;
- Bars represent the start (split) or end Goin) of concurrent activities;
- A black circle represents the start (initial node) of the workflow; an encircled black circle represents the end (final node).

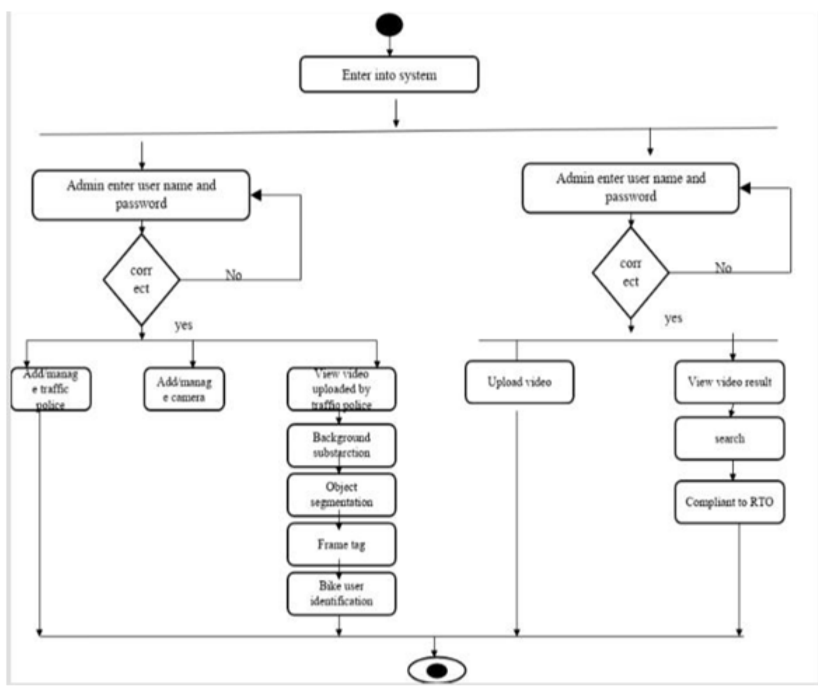

Figure 25: Activity Diagram

8 IMPLEMENTATION

8.1 Module description

Here there are two modules,

1. ADMIN

2 TRAFFIC POLICE

ADMIN

- RTO is considered as admin.
- After logging in with their specified id and password, admin can view the complaints and can send replies.
- Admin can manage the traffic police by adding or editing. And can also manage camera.

TRAFFIC POLICE

- Here traffic police is considered as user.
- After logging in, user can view video results.
- User is the one who can send complaints to admin. And can also view the replies that is send by the admin.

8.2 working

8.2.1 Video and Image Gathering

Our input datasets were collected from the video surveillance system of Loei Rajabhat University in Loei province, Thailand. A camera we chose to start an experiment is the camera at the front gate of the university. We collected 50 videos of a vehicle passing through the gate, each video is 5 minutes long then the total of all videos length is 250 minutes

Figure 26: Examples of the input video from a front gate camera

After that, we manually classify the image of a biker wearing a helmet and no helmet from the video data. Then we crop an area of a motorcycle with a biker and helmet into one image dataset call "Bikerwithhelmet" and the area of a motorcycle with a biker who wears no helmet into another dataset call "Bikerwithnohelmet" (See Fig. 4). The total input image we have in "Bikerwithhelmet" dataset is 336 images and for "Bikerwithnohelmet" we have 157 images. The total of them is 493 images.

8.2.2 Opencv

This video is then divided into various frames using CV2. OpenCV-Python is a library of Python bindings designed to solve computer vision problems. [cv2. imread()] method loads an image from the specified file. If the image cannot be read (because of missing file, improper permissions, unsupported or invalid format) then this method returns an empty matrix.

8.2.3 Image classification experiment

After gathering 493 images for our training dataset, we split our images into two groups, one for training data and another for test data to use in classification experiment. This experiment we test them with four CNN models for image classification (VGG16, VGG19, Inception V3, and MobileNets). For the evaluation, we used 10-fold crossvalidation experiment which we set a number of test data for 10 percent of the total image. The training networks are trained using Python TensorFlow library, then we calculate the accuracy and choose two good models to use in image detection step.

8.2.4 Image Detection Experiment

In this step, we use all 50 videos that we collected to do image detection experiment using SSD technique combine with two CNN models we chose from the previous step. All videos will be tested and calculated the accuracy of the biker with helmet and no helmet detection in the video. We also count a number of undetected motorcyclists to be an error

8.2.5 Result Interpretation

The last step, we compare the performance from two previous steps and make the conclusion. The accuracy of the experiments will show the performance of each technique in terms of image classification and image detection.

8.2.6 STEPS

- Collect cctv videos as input
- Extract video frames from the videos by using cv2
- Detect person with bike from the frame by using yolo algorithm
- Output of yolo is taken as the input of cnn
- Here we used the cnn archetechture is mobilenet-v2(input shape-(224,224,3)
- Finally it will detect the person with no helmet

dataset : kaggle images

75 percent of image is used for training and 35 percent of image is used for testing

9 TESTING

This section describes the testing procedure, staring with the testing methodology and then continuing with the tests performed on the system.

• Testing is a process of exceeding with the internet of finding an error.

• A good test case is one that has a high probability of finding an undiscovered error.

• A successful test is one that uncovers undiscovered errors.

9.1 Testing Methodology

The most natural and customary way of verifying any piece of work is just to operate it in some representative situation and verify whether its behavior is as expected. In general it is impossible to test it under all possible operating conditions. Thus it is necessary to find suitable test cases that provide enough evidence that the desired behavior will be exhibited in all remaining case. Testing is a critical activity in software engineering and should be performed as systematically as possible by stating clearly what result one expects from it and how one obtains these results. On the contrary, often in practice, testing is performed in an unstructured way without applying any criterion. There are two approaches to testing:

• White box testing

This is testing software using information about the internal structure of the software. It tests what the program does. The test is being carried out to check the internal structure of the software. The test is carried out successfully and the internal structure of the software meets the required criteria.

• Black box testing

Testing a piece of software as a black box (also called functional testing) means operating the software without relying on any knowledge of the way it has been designed and coded. It tests what it is supposed to do. The Programs run successfully against task assigned.

• Unit Testing

This is the first level of testing. In this different modules are tested against the specification produces during the design of the modules. Unit testing is done during the coding phase and to test the internal logic of the modules. It refers to the modules. It refers to the verification of single program module in an isolated environment. Unit testing first focuses on the modules independently of one of another to locate errors. After coding each dialogue is tested and run individually. All necessary coding were removed and it was ensured that all modules are worked, as the programmer would expect. The logical errors found were corrected so, by working all the modules independently and verifying the outputs of each module in the presence of staff, I observed that the program was functioning as expected. In unit testing – Module is tested to ensure that information properly flows into and out of the program under test – Local data structures are examined to ensure that data stored temporarily maintains its integrity during all steps in algorithm execution. Boundary condition is tested to ensure that module operates properly at boundaries established to limit or restrict processing.

– All independent paths through the control structures are executed to ensure that all statements in the module have been executed at least once.

– Error handling paths are also tested.

• Integration testing

Data can be lost across an interface: one module can be adversely affected on another; sub functions when combine may not produce the desired major functions. Integration testing is a systematic testing for constructing the program structure. Conducting the tests is to uncover errors associated within the interface. The objective is to take unit tested to modules and build a program structure. All the modules are combined and tested as a whole. Here correction is difficult because the vast expenses of the entire program complicate the isolation of causes. Thus in the integration testing step, all the errors uncovered are corrected for the next testing steps.

– Low – Level modules are combined to form clusters .

– The cluster is tested

– Drivers are removed and clusters are combined moving upward in the program structure.

• Alpha Testing

A series of acceptance tests were conducted to enable the users to validate requirements. The suggestions, along with the additional requirements of the end user were included in the project.

• Beta Testing

It is to be conducted by the end – user without the presence of the developer. It can be conducted over a period of weeks or month. Since it is a long time consuming activity, its result is out scope of this project report. But its result will help to enhance the product at a later time.

- Validation Testing

This provides final assurance that the software meets all the functional, behavioral and performance requirement. The software is completely assembled as a package. Validation succeeds when the software functions in a manner in which user wishes. Validation refers to the process of using software in live environment in order to find errors. During the course of validation the system failure may occur and sometimes the coding has to be hanged according to the requirement. Thus the feedback from the validation phase generally produces changes in software. Once the application was made of all logical and interface errors, inputting dummy data ensure that the software developed satisfied all the requirements of the user. This dummy data is known as test case.

- Output Testing

After performing the validation testing, the next step is output testing of the proposed system since no system could be useful if it does not produce the required output in a specific format. Asking the users about the format required by them, tests the output generated are considered into two ways. One is on screen and another is printed format. The output format on the screen found to be correct as the format was designed in the system design phase according to the user needs. For the hard copy also, the output comes out as the specified requirement by the user. Hence output testing does not result in any correction in the system.

10 RESULT

For the image classification experiment. We calculate the accuracy for each model (VGG16, VGG19, Inception V3, and MobileNets). The overall results are shown in Table I.

Network model	Accuracy	Model size
VGG16	78.09	434580
VGG19	79.19	451,258
Inception V3	84.58	85,447
MobileNets	85.19	16,754

Table 1: Accuracy of biker with helmet and no helmet classification

The result from Table I shows that MobileNets is the best CNN model to recognize biker with helmet and no helmet image sets. The accuracy of MobileNets is the highest (85.19 percent) follow by Inception V3 (84.58lowest models compare to these four models with VGG19 slightly better than the VGG16 but both of VGG models generate the huge size of the network (¿ 400 MB) compare to Inception V3 and MobileNets. Then we decide to choose Inception V3 and MobileNets for the next experiment on video image detection with the SSD method.

From the confusion matrix for the inception v3 and mobilenet, we found that MobileNets also performs better than Inception V3 in terms of image detection. MobileNets detected 117 bikers with helmet and 304 bikers with no helmet correctly from overall 493 images. But Inception V3 can detect only 115 bikers with helmet and 301 bikers with no helmet. However, both models successfully detect all 493 bikers in the video datasets and leave only 0 biker that is undetected. The error of our image detection experiment found only on the misclassification issue but we have a 100 percent image detection accuracy.

11 PERFORMANCE AND EVALUATION

- Convolutional neural networks offer the benefit of automatically generating high-level features.
- The mobilenets network has achieved the good performance compare to another model and also come with the smallest size.
- The existing system i.e, helmet detection using YOLOv3 is found that its accuracy decreases when the dataset is unbalanced and irregular.
- YOLO is used for highspeed outputs, where accuracy is not that high.
- Mobilenets provide higher accuracies with high speed outputs with a higher computation time.
- YOLO predicts one type of class in one grid. Hence small objects are not identified.

Actual class	with helmet	no helmet	undetected
with helmet	115	41	0
no helmet	36	301	0

Table 2: Accuracy of with helmet and no helmet detection(inception-v3)

Actual class	with helmet	no helmet	undetected
with helmet	117	39	0
no helmet	33	304	0

Table 3: Accuracy of with helmet and no helmet detection(mobilenet-v2)

A confusion matrix for the Inceptionv3 and MobileNets for image detection with SSD are presented in Table II and III. From the result, we found that MobileNets also performs better than Inception V3 in terms of image detection. MobileNets detected 117 bikers with helmet and 304 bikers with no helmet correctly from overall 493 images. But Inception V3 can detect only 115 bikers with helmet and 301 bikers with no helmet. However, both models successfully detect all 493 bikers in the video datasets and leave only 0 biker that is undetected. The error of our image detection experiment found only on the misclassification issue but we have a 100 perecent image detection accuracy.

12 SCREENSHOTS

login page

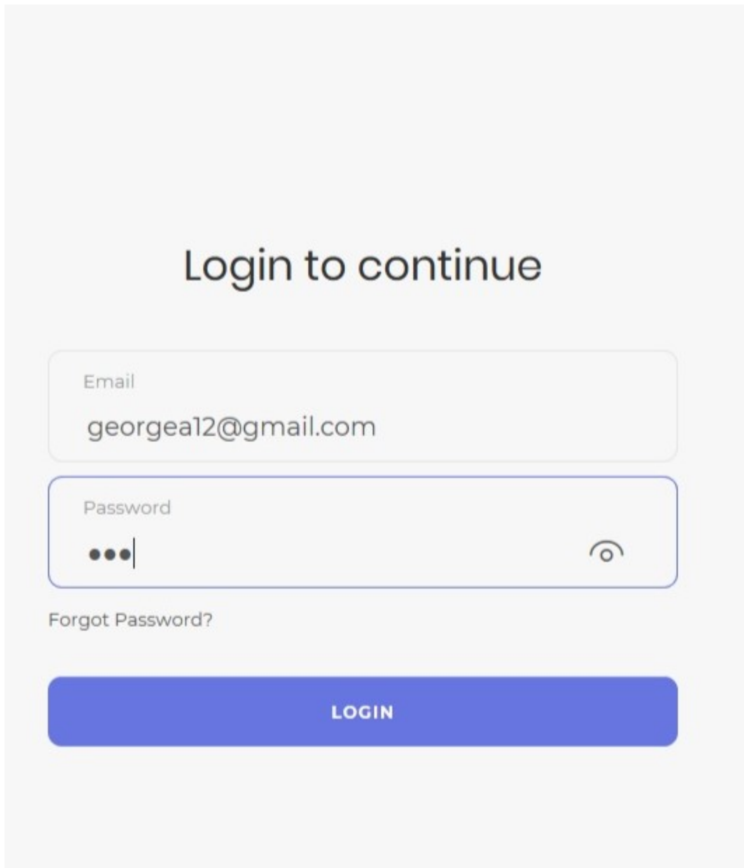

Figure 27: login page

Admin home

Figure 28: Admin home

view and edit traffic police

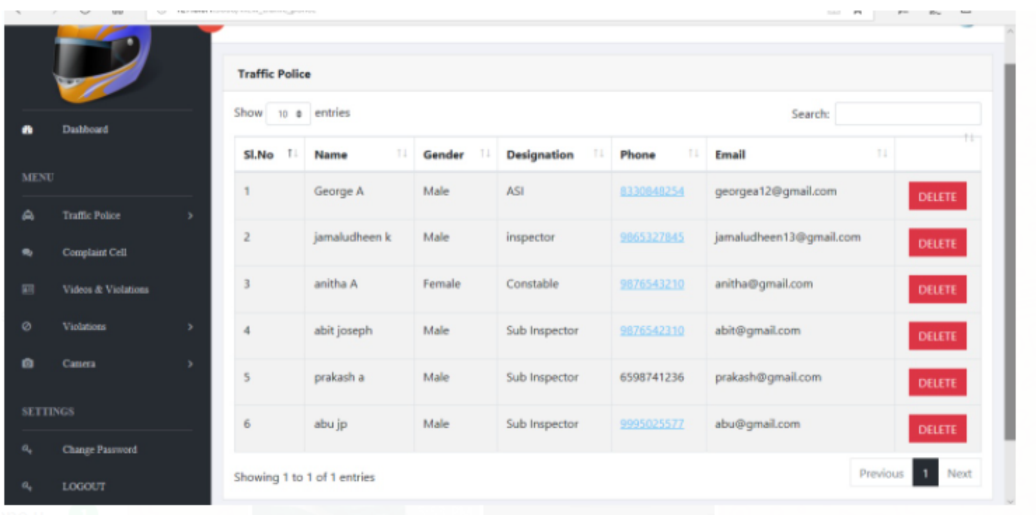

Figure 29: view and edit traffic police

Add traffic police

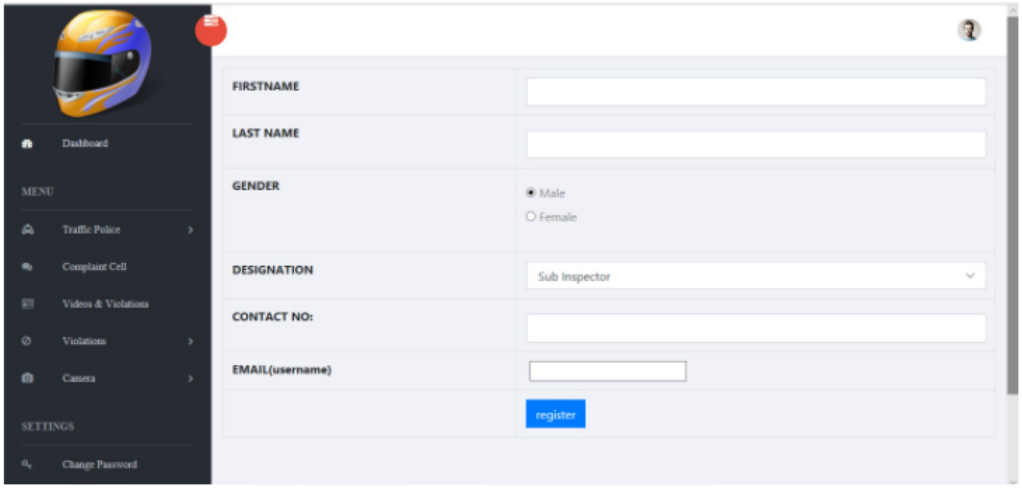

Figure 30: Add traffic polic

Videos and violation

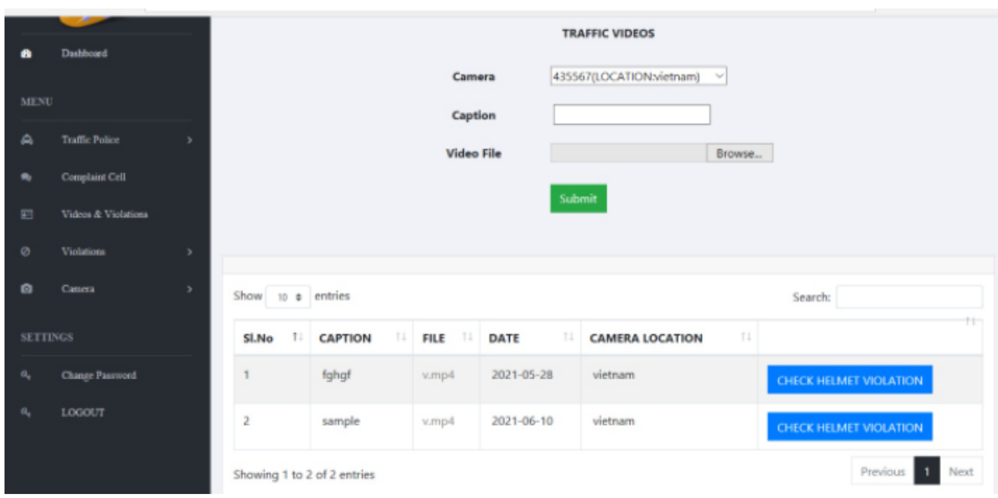

Figure 31: Videos and violation

Helmet violation

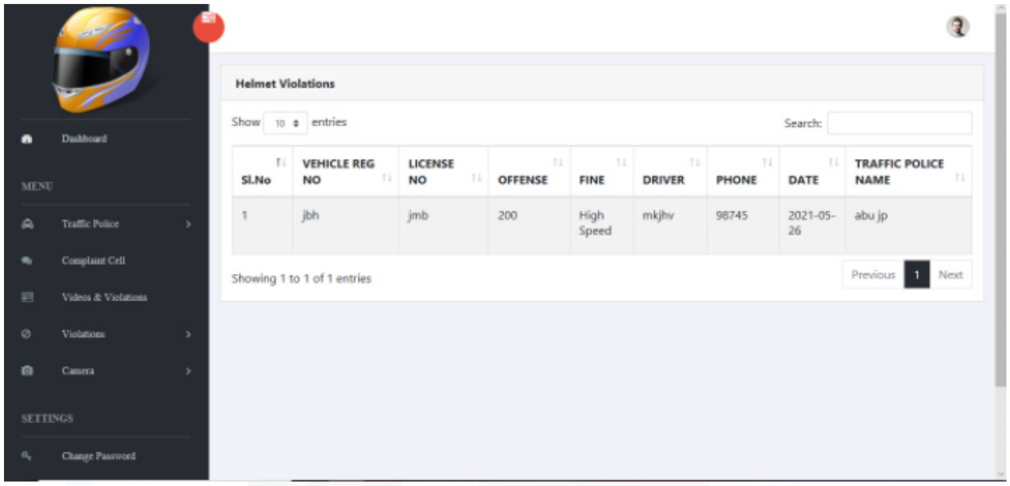

Figure 32: Helmet violation

Add camera

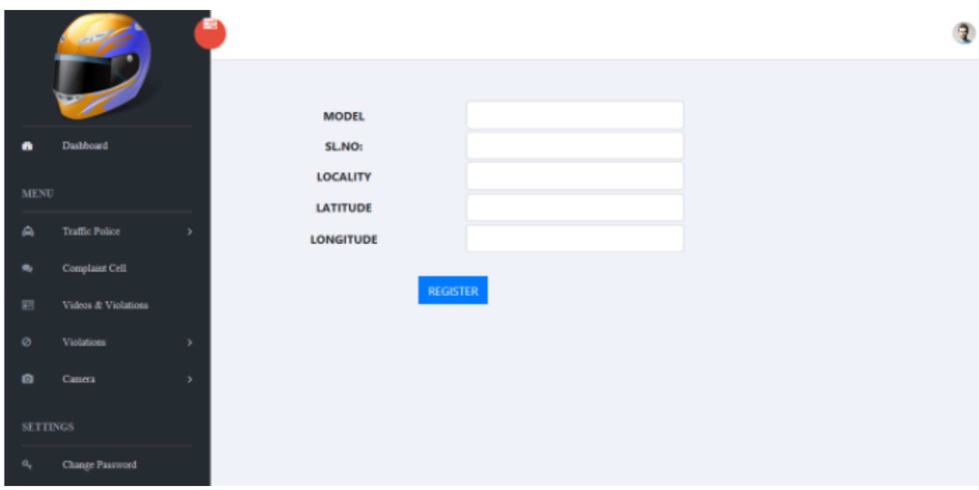

Figure 33: Add camera

13 CONCLUSION

Applying some deep learning techniques to solve the issues of motorcyclist's staring a helmet and no helmet detection and classification. Four Convolutional Neural Networks (CNN) in these experiments (VGG16, VGG19, Google Net or Inception V3, and Mobile Nets) for image classification step and also combine these models with the SSD technique to do an image detection step. The results of experiment itre looking good. In the classification step, it found that Mobile Nets achieved the better accuracy than VGG16 (78.09 and VGG19) .

For the detection step, Mobile Nets is the winner which detected 421 correct motorcyclists class from the total number of493 video images. Follow by Inception V3 which detected only 416 correct images. And the best part of these is the SSD technique can detect these images by using only one single runtime and require no other image pre-processing algorithms. These results show us that a Deep Learning or CNN techniques are the good algorithms that can apply on the problem of image detection and classification about bikers staring a helmet or no helmet problem.

14 Appendix A

References

[1] Liu, W., Angelo, D., Ethan, D., Szeged, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016, October. “Ssd: Single shot multibox detector”. In European conference on computer vision.

[2] Helmet detection on motorcyclist using image descriptors and classifiers romuere R.V esilva kelson R.T.Aires.Rodrigo de M.S veras

[3] Helmet detection using machine learnig and automatic license plate recognition

[4] Automatic detection of bike riders without helmet using survellience videos in real time

[5] Ren, S., He, K., Girshick, R. and Sun, J., “Faster r-cnn: Towards realtime object detection with region proposal networks”. In Advances in neural information processing system