TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

A

PROPOSAL

ON

‘**DEBLURRING REAL TIME LICENSE PLATE IMAGES FOR THE IDENTIFICATION OF VEHICLES BY VARIATIONAL TECHNIQUE’**

By:

ASHISH PAUDEL—072BCT506,

JEEVAN THAPA—072BCT514,

SUMAN SAPKOTA—072BCT542,

SUSHMIT RAJAURE—072BCT545.

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

LALITPUR, NEPAL

FALGUN 2075

ACKNOWLEDGEMENTS

We would like to extend our sincerest gratitude to Dr. Dibakar Raj Pant for helping us refine our project idea and for providing us with guidance regarding the research we needed to do in order to bring forth a concrete project idea. We would also like to thank the department for providing this opportunity to employ our engineering skills in a major project that is bound to provide us with numerous insights and challenges as we embark upon this one year long journey.

ABSTRACT

This project is focused on recognizing license plates from real time CCTV footage in order to detect stolen vehicles on the road. Our approach is to first segment out image of vehicles from the frames of video footage at a constant frame rate. Then further image segmentation will be done to extract license plate from the vehicle's image. Such image may be blurred due to motion blur. A deblurring algorithm will be implemented to enhance the image. Then, an OCR system shall be used to record the license number and compare it against a database of stolen vehicles to identify whether the vehicle is stolen or not.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. OBJECTIVE

To develop a system to identify the license number of vehicles from traffic video feed.

1.2. OVERVIEW

Every vehicle has a unique license plate. The license plate number of the vehicle is used to identify the vehicle. Our main goal of the project is to identify the license plate number of the vehicle. The input for the project will be a real time video in traffic and the output will be the license plate for each vehicle present. The end to end system will consider different sub portions. The first job will be to segment out the vehicles in the road. Then, the number plates of the vehicles will be segmented out, which will be sent to a deblurring model to remove noise and reduce the effect of motion deblurring in the picture. The final job will be to identify the number in the license plate.

For the project, we will be using Machine Learning algorithms for segmentation and deblurring. Also a variational approach for image deblurring will be implemented and used for deblurring the other way. The input to the system will be a video. The frames in the video will be passed through 4 subsequent sub-systems: **Vehicle segmentation**, **License Plate extraction**, **Deblurring sub-system** and finally **OCR**. The subsystems are described below shortly.

- **Vehicle Detection:**

Segmentation of vehicles from videos is a popular Machine Learning problem. Many successful and highly efficient sophisticated solution attempts have been made in the past few years for this problem. One of them include **YOLO (You Only Look Once)** algorithm. YOLO is a massive Convolutional Neural network for object detection and classification. As a quote from its website explains "This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities." Thus, the problem can be tackled using a similar algorithm. The solution is to build up a vehicle segmentation using an encoder like fully convolutional neural network. Some popular datasets include those hosted in **Berkeley DeepDrive BDD100k** and **Baidu Apolloscapes.**

- **License Plate extraction:**

The images of vehicles will be cropped from the frames and then license plate will be extracted. For this a fully convolutional dataset will be implemented and trained. There are various datasets available for this job with labelled positions for license plates. Some of them include **OpenALPR benchmark and UCSD car dataset.**

- **Deblurring Sub-system:**

Two deblurring sub-systems will be implemented for the job. The first will be using Image processing. And the other will be using a fully convolutional neural network. This fully CNN will be similar to an autoencoder and will be trained to deblur motion blur. The dataset for the image deblurring will be generated synthetically by using proper convolutional kernels on the images.

- **OCR:**

The OCR portion will be used to extract and recognize the numbers and letters in de-blurred photo of license plate. The OCR will be developed such that it recognizes both letters and digits of English as well as Devanagiri fonts.

2. LITERATURE REVIEW

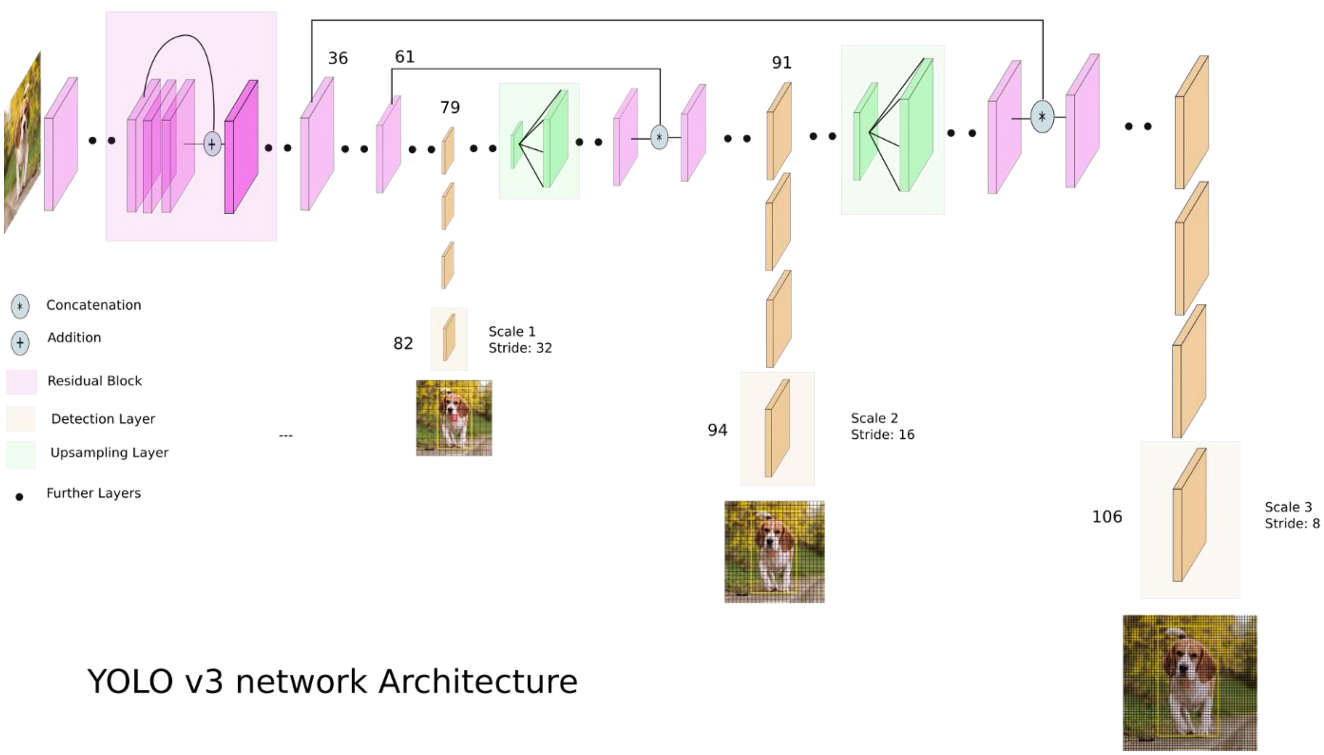
2.1. OBJECT DETECTION

We use YOLO v3 algorithm for the task of object detection. This architecture is based on Darknet, which was trained for object classification task. Darknet originally has 53 layer network trained on Imagenet. For the task of detection, 53 more layers are stacked onto it, giving us a 106 layer fully convolutional underlying architecture for YOLO v3.

YOLO v3 has pretrained weights available as well. The model was trained on the 'Common Objects in Context' Dataset, in short COCO. This model is able to recognise various frequently occouring objects like people, trees, traffic signs, animals, furnitures, etc.

Our usage of this dataset and model is to detect vechiles including bikes, car, truck, bus as well as pedestrian and traffic signs.

The complete architecture of YOLO v3 is as below:



YOLO v3 network Architecture

The most salient feature of YOLO v3 is that it makes detections at three different scales. YOLO is a fully convolutional network. The detection is done by applying 1 x 1 detection kernels on feature maps of three different sizes at three different places in the network.

The input to the model is:

$$H \times W \times C$$

where, H = 416 : Height of image

W = 416 : Width of image

C = 3 : Number of channels

The output at each scales is:

$$B \times (5 + C)$$

where, B = no of class sharing the same bounding box

5 → x, y, height, width of bounding box, confidence of the detection

C = no of classes the bounding box can predict (80 for COCO)

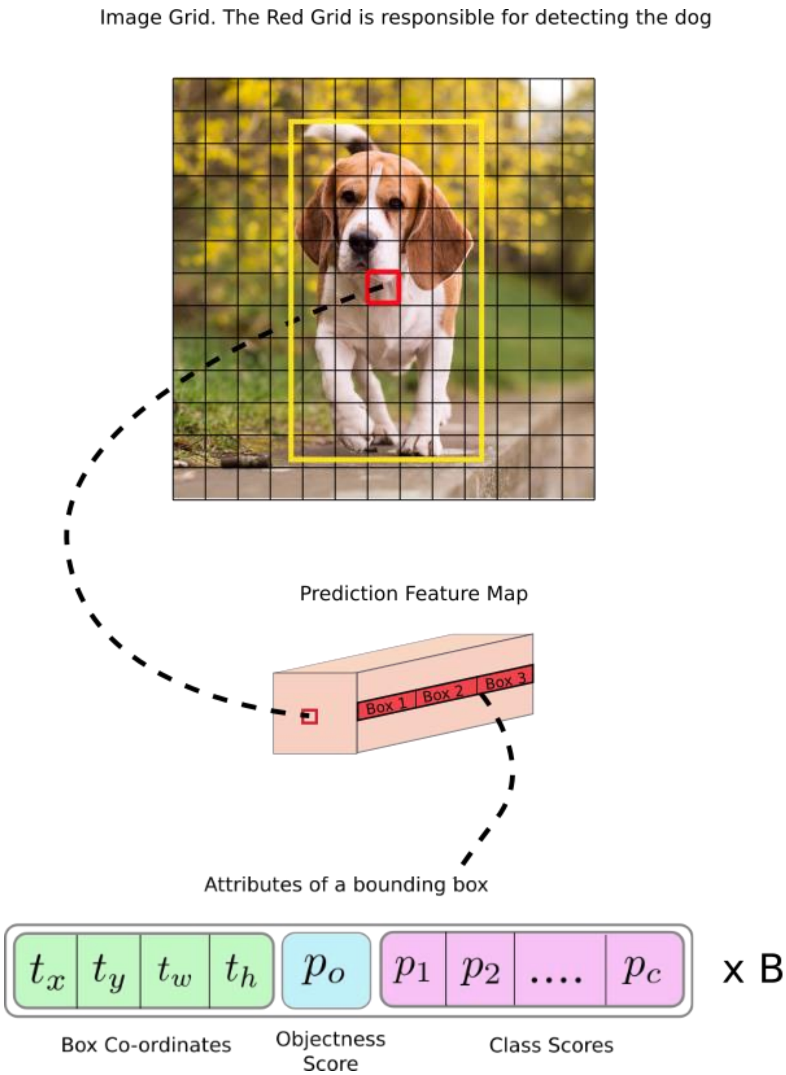


Image Grid. The Red Grid is responsible for detecting the dog
Prediction Feature Map
Box 1
Box 2
Box 3
Attributes of a bounding box
t_x t_y t_w t_h p_o p_1 p_2 .... p_c
x B
Box Co-ordinates
Objectness Score
Class Scores

2.2. IMAGE DEBLURRING WITH THE VARIATIONAL METHOD

An important problem in image analysis is the reconstruction of an original image f_0 describing a real scene from an observed image y. The transformation (or degradation) connecting f_0 to y is in general the result of two phenomena. The first phenomenon is deterministic and is related to the mode of image acquisition or to possible defects of the imaging system. The second phenomenon is random: the noise inherent degradation in any signal transmission. Suppose that the noise denoted by w is white, Gaussian, and additive. The simplest model accounting for both blur and noise is the linear degradation model: we suppose that f_0 is connected to y by an equation of the form

$$y = \Phi f_0 + w = h \star f_0 + w \tag{2.1}$$

The reconstruction problem, in its simplest form is to find f_0 from y. In general, the problem is ill-posed (in the sense of Hadamard) since the information provided by y (the observed image) and the model of the above equation is not sufficient to ensure the existence, uniqueness and the stability of the solution f_0. So, regularization is needed to find the numerical solution of 2.1.

We consider variational deconvolution methods, that finds a regularizer through a convex optimization:

$$f^\star \in \operatorname{argmin}_f \frac{1}{2}\|y - \Phi f\|^2 + \lambda J(f) \tag{2.2}$$

Where J(f) is a prior energy and λ is a constant parameter. 2.2 is a well-posed form of 2.1 in the discrete domain. We are planning to use the total variational regularization prior energy (assuming the images has edges of bounded perimeter):

$$J(f) = \sum_x \|\nabla f(x)\| \tag{2.3}$$

However, the TV norm of 1.3 is not smooth and needs to be smoothened to be used as a regularizer. So, the form we are going to use is given below:

$$J(f) = \sum_x \sqrt{\|\nabla f(x)\|^2 + \varepsilon^2} \tag{2.4}$$

Where ε is the regularization parameter for the TV norm.

The energy prior J(f) is non-quadratic (i.e. it doesn’t converge in a quadratic sense to the Fourier Transform) so, an iterative approach is needed to find the solution f from 2.2. We use the method of gradient descent to approximate the solution. An iteration in the gradient descent reads:

$$f^{(k+1)} = f^{(k)} - \tau \left(h \star (h \star f^{(k)} - y) + \lambda \mathrm{Grad} J(f^{(k)}) \right) \tag{2.5}$$

Where τ is the step size. τ must be smaller than twice the Lipschitz constant of the Gradient of the functional to be minimized. Hence,

$$\tau < \frac{2}{1 + \lambda 8/\varepsilon}. \tag{2.6}$$

2.3. OPTICAL CHARACTER RECOGNITION

Optical character recognition is a sequence of multiple processes – segmentation, feature extraction, and classification. Different models or techniques are proposed for character segmentation. These techniques can be categorized into three major strategies – dissection technique, recognition driven technique, and holistic methods. The use and selection of these techniques highly depends on the construct of script and language. Various feature extraction and classification techniques has been proposed by different researchers. The feature extraction algorithms may rely on morphology of characters for better classification. Classification in one of the major steps in OCR and design of good classifier is also a challenging task. Mostly supervised learning is used for the classification of characters.

2.3.1. DIFFERENT MODELS OF CHARACTER SEGMENTATION IN OCR SYSTEMS

Character segmentation is an operation that seeks to decompose an image of sequence of characters into sub-images of individual symbols. The difficulty of performing accurate segmentation is determined by the nature of the material to be read and by its quality. Segmentation is the initial step in a three-step procedure. (Casey & Lecolinet, 1996):
Given a starting point in a document image:

- Find the next character image.
- Extract distinguishing attributes of the character image.
- Find the member of a given symbol set whose attributes best match those of the input, and output its identity.

This sequence is repeated until no additional character images are found. A character is a pattern that resembles one of the symbols the system is designed to recognize. But to determine such a resemblance the pattern must be segmented from the document image. Casey & Lecolinet (Casey & Lecolinet, 1996) have classified the segmentation methods into 15 three pure strategies based on how segmentation and classification interact in the OCR process. The elemental strategies are:

- ***The classical approach***, in which segments are identified based on "character-like" properties. This process of cutting up the image into meaningful components is given a special name, "***dissection"***.

- ***Recognition-based segmentation***, in which the system searches the image for components that match classes in its alphabet.

- ***Holistic methods***, in which the system seeks to recognize words as a whole, thus, avoiding the need to segment into characters.

2.3.2 DISSECTION TECHNIQUES

By dissection means decomposition of image into a sequence of sub-images using general properties of the valid characters such as height, width, separation from neighboring components, disposition along a baseline etc. Dissection is an intelligent process in that an analysis of the image is carried out; however, classification into symbols is not involved at this point. The segmentation stage consisted of three steps:

- Detection of the start of a character.
- A decision to begin testing for the end of a character
- Detection of end-of-character.

The analysis of the projection of a line of print has been used as a basis for segmentation of non-cursive writing. When printed characters touch, or overlap horizontally, the projection often contains a minimum at the proper segmentation column (Casey & Lecolinet, 1996). A peak-to-valley function has been designed to improve this method. A minimum of the projection is located and the projection value noted. A vertical projection is less satisfactory for the slanted characters.

Analysis of projections or bounding boxes offers an efficient way to segment non-touching characters in hand- or machine-printed data. However, more detailed processing is necessary 16 in order to separate joined characters reliably. The intersection of two characters can give rise to special image features. Consequently dissection methods have been developed to detect these features and to use them in splitting a character string image into sub-images. Only image components failing certain dimensional tests are subjected to detailed examination.

2.3.3 RECOGNITION DRIVEN SEGMENTATION

This approach also segment words into individual characters which are usually letters. It is quite different from dissection based approach. Here, no feature-based dissection algorithm is employed. Rather, the image is divided systematically into many overlapping pieces without regard to content. These are classified as part of an attempt to find a coherent segmentation/recognition result. Letter segmentation is a by-product of letter recognition, which may itself be driven by contextual analysis. The main interest of this category of methods is that they bypass the segmentation problem: No complex “dissection" algorithm has to be built and recognition errors are basically due to failures in classification.

The basic principle is to use a mobile window of variable width to provide sequences of tentative segmentations which are confirmed (or not) by character recognition. Multiple sequences are obtained from the input image by varying the window placement and size. Each sequence is assessed as a whole based on recognition results. In recognition-based techniques, recognition can be performed by following either a serial or a parallel optimization scheme. In the first case, recognition is done iteratively in a left-to-right scan of words, searching for a "satisfactory" recognition result. The parallel method proceeds in a more global way. It generates a lattice of all (or many) possible feature-to-letter combinations. The final decision is found by choosing an optimal path through the lattice (Casey & Lecolinet, 1996).

Recognition-based segmentation consists of the following two steps:

1) Generation of segmentation hypotheses (e.g. windowing)

2) Choice of the best hypothesis (verification step)

2.1.3 HOLISTIC TECHNIQUE

Holistic technique is opposite of the classical dissection approach. This technique is used to recognize word as a whole. Thus skips the segmentation of words into characters. This involves comparison of features of unsegmented word image to the features or descriptions of words in database.

Since a holistic approach does not directly deal with characters or alphabets, a major drawback of this class of methods is that their use is usually limited to predefined words. A training stage is thus mandatory to expand or modify the scope of possible words. This property makes this kind of method more suitable for applications where the lexicon is statically defined, like check recognition. They can be used for specific user as well as to the particular vocabulary concerned. Holistic methods usually follow a two-step scheme:

1. The first step performs feature extraction.

2. The second step performs global recognition by comparing the representation of the unknown word with those of the references stored in the lexicon. (Chaudhuri & Pal, 1997)

3. METHODOLOGY

3.1. SYSTEM BLOCK DIAGRAM

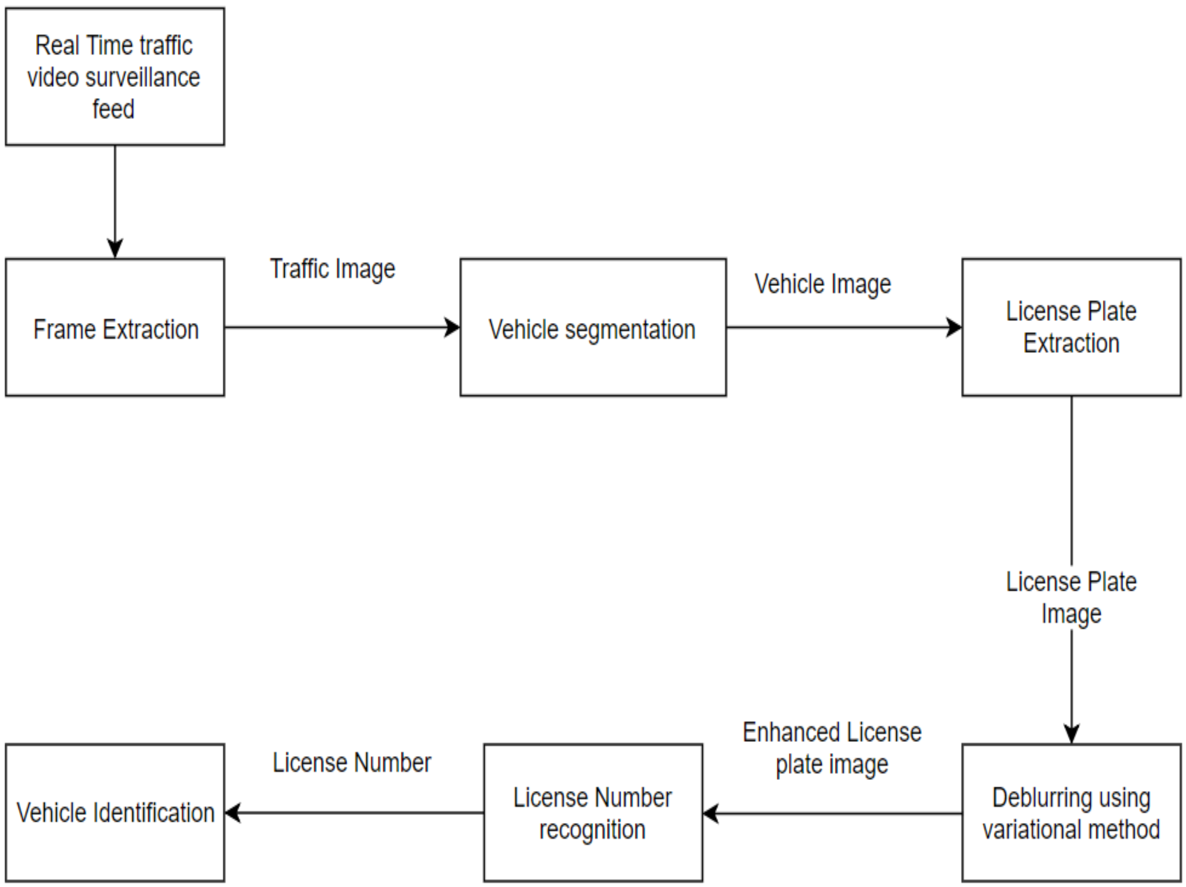


Fig. 3.1

3.2. SYSTEM ARCHITECTURE

3.2.1. TOP-LEVEL ARCHITECTURE

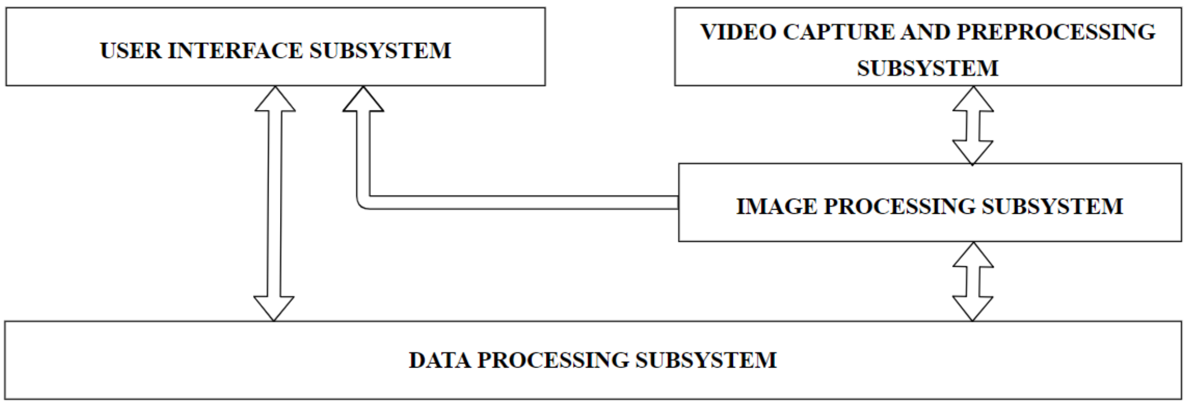


Fig. 3.2

3.2.2. USER INTERFACE SUBSYSTEM

The user interface for the project will provide the user with real-time information about the vehicles being recorded at the particular location (a junction of a road, for example). In the event that a stolen vehicle is spotted, the user will be notified of it immediately.

3.2.3. VIDEO CAPTURE AND PREPROCESSING SUBSYSTEM

The video capture and preprocessing subsystem will be composed of the following parts:

i) A high-definition digital camera with a WiFi or Bluetooth interface for real-time data transfer.
ii) A video processing software that extracts frames from the video and sends it to the image processing subsystem.

3.2.4. IMAGE PROCESSING SUBSYSTEM

The image processing subsystem is the core subsystem of the project and will have the following architecture:

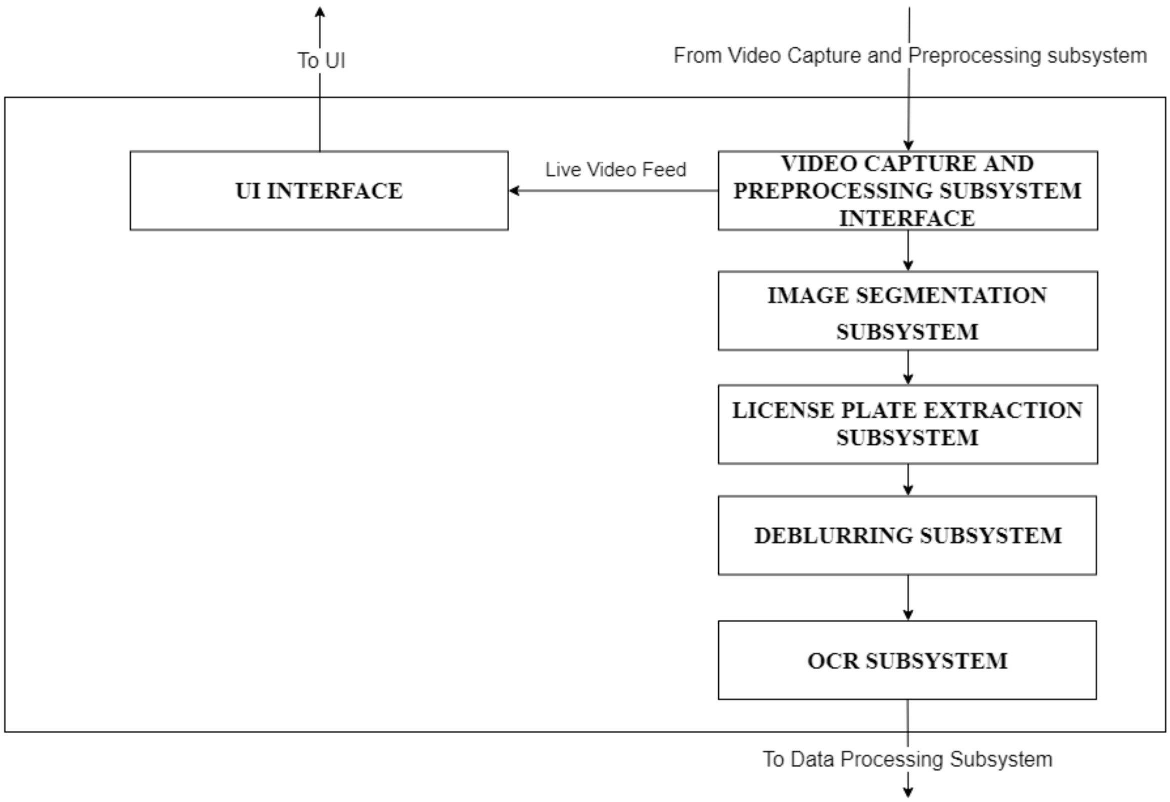


Fig 3.3

The image processing subsystem consists of the following subsystems in it:

- **Video capture and preprocessing subsystem interface**

  It is the interface module that recieves data from the video capture and preprocessing subsystem and feeds it to the lower layers of the image processing subsystem. It also feeds the video feed to the user interface.

- **Image segmentation subsystem**

  This subsystem takes the traffic image as an input and creates segments of vehicles according to the algorithms described in section 2.1.

- **License plate extraction subsystem**

  This subsystem extracts license plate images from the segmented vehicle images for further processing.

- **Deblurring subsystem**

  This subsystem deals with removing the blur and noise from the license plate images by using the variational method described in section 2.2.

- **OCR subsystem**

  This subsystem deals with recognizing the license plate numbers by using optical character recognition technology as described in section 2.3. This subsystem also provides the recognized vehicle data to the data processing subsystem.

3.2.5. DATA PROCESSING SUBSYSTEM

The data processing subsystem takes the input from the image processing subsystem, processes the license plate information, stores it in the database, checks the database to see if any vehicles identified are reported as stolen and provides information to the user interface subsystem.

4. EXPECTED OUTCOME

The final outcome shall be a web application that incorporates the complete functionality of extracting frames from a traffic video feed, then performing object detection, license plate extraction, and finally Optical Character Recognition. The information thus extracted shall be fed to a database and queried to see if any stolen vehicles are found. Some images of expected outcome of different subsystems are depicted below.

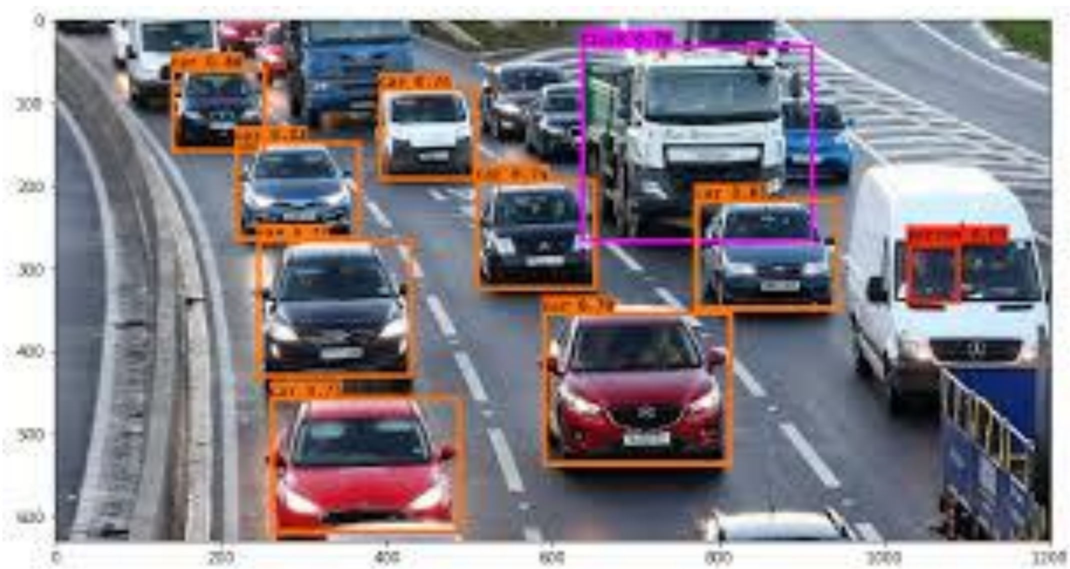


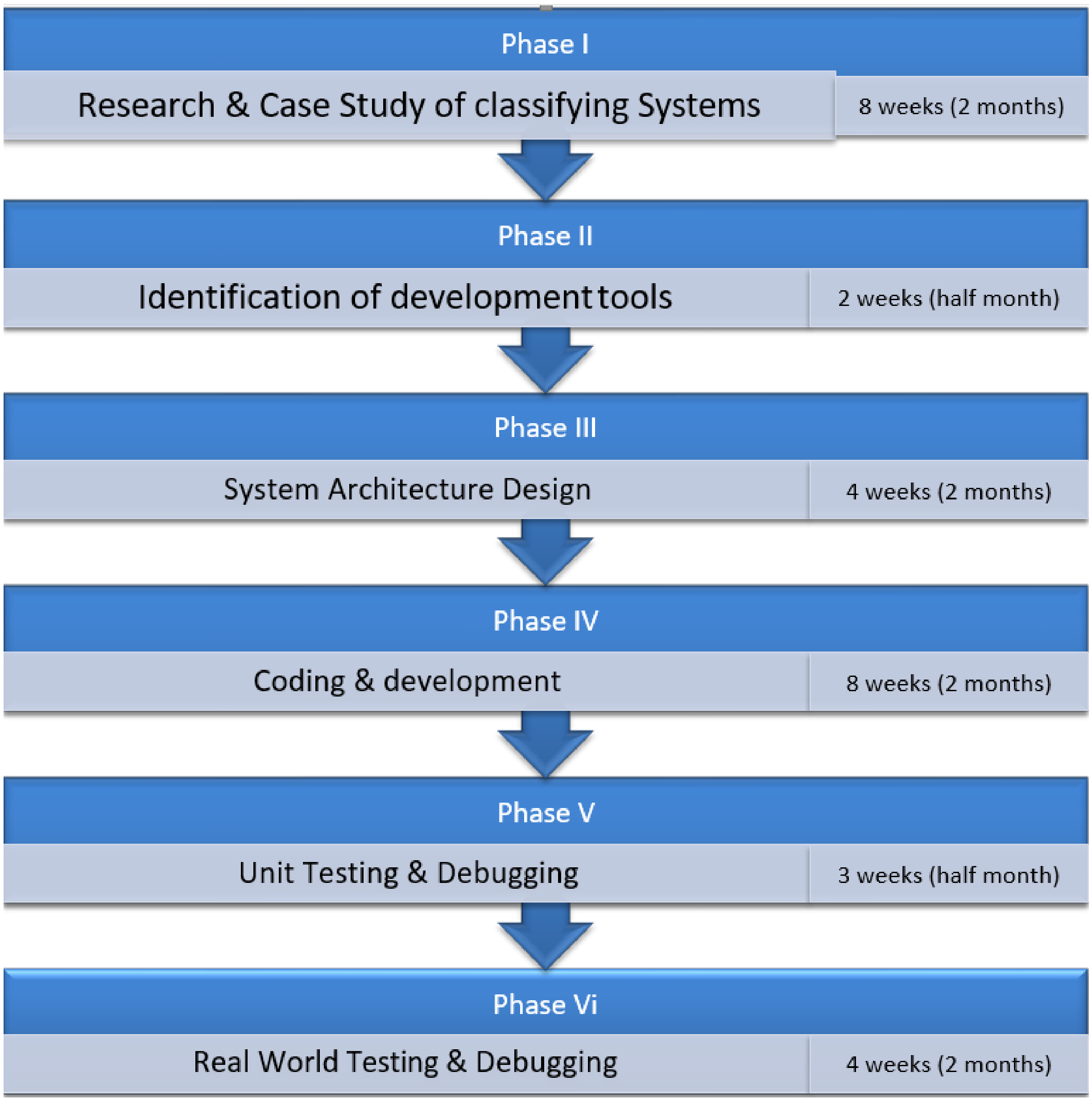
Fig. 4.1: Object Detection

Fig. 4.2: License Plate Extraction

Fig. 4.3. Deblurring with the variational method

Fig. 4.4: Character segmentation and recognization

5. PROJECT PLAN



6. REQUIREMENTS

1. Software requirements:
   1.1. Programming Languages: Python, JavaScript, Java
   1.2. Libraries / Frameworks:
      - Computation Model: Tensorflow, Pytorch, Numpy, Matplotlib, Scipy, Scikit-learn, OpenCV
      - Web App: NodeJs, React framework
2. Resource Requirements:
   2.1. GPU powered Laptops
   2.2. Google CoLab
   2.3. High speed Internet services
   2.4. Past reports on research in ML
   2.5. Past reports on Digital Signal Processing
3. Hardware Requirements
   3.1. A high-resolution digital video camera with a IEEE 802.11 interface

7. BIBLIOGRAPHY

[1]. Aittala M., Durand F.: Image Deblurring Using Convolutional Neural Network, Reshma Vijay V.J., Deepa P.L (2016)

[2]. Burst Image Deblurring Using Permutation Invariant Convolutional Neural Networks (2018)

[3]. Subashini, P., Krishnaveni A.: Image Deblurring Using Back Propagation Neural Network,(2011)

[4]. Neetin, K., Shrivastava, M., Rahim N,: Image Deblurring Using a Neural Network Approach (2012)

[5]. Gilbert, A., Messingher, S., Patel A.: Non-Blind Image Deblurring using Neural Networks

[6]. Chakrabarti, A.: A neural approach to blind motion deblurring. In: ECCV (2016)

[7]. Deblurring Images: Matrices, Spectra, and Filtering. Society for Industrial & Applied Mathematics, U.S., 2007.

[8]. Aubert, A., Vese L.: A variational method in image recovery. Society for Industrial and Applied Mathematics, U.S. In: Siam J. Numer Analytical. Vol. 34, No. 5, pp. 1948-1979. (1997).

[9]. Cash, G.L., Hatamain, M. Optical character recognition by the method of moments. In: Computer vision, graphics and image processing 39, pp. 291-310. (1987).

[10]. Redmond, J., Divvala, S. et. al. You Only Look Once: Unified, Real-Time Object Detection. Extracted from: https://arxiv.org/abs/1506.0260v5. (2016).