Object Recognition and Classification for Security Surveillance System using Single Board Computer

by

MUHAMMAD FARIEZ FIKHI BIN FAUZI
(1932322905)

A dissertation submitted in partial fulfillment of the requirements for the degree of
Master of Science in Embedded System Design Engineering

School of Computer and Communication Engineering
UNIVERSITI MALAYSIA PERLIS

2020

ACKNOWLEDGMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along the completion of my project. All that I have done is only because of supervision and assistance and I would not forget to thank them. I am extremely thankful to Dr Shuhaizar Daud for providing such a nice support and guidance, although he had busy schedule class teaching. I also owe my deep gratitude to dissertation committee, who have been responsibility on this project work and guided all along till the completion of our project work by providing all the necessary information for developing a good system.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PENGESANAN OBJEK DAN PENGELASAN UNTUK SISTEM PENGAWASAN KESELAMATAN MENGGUNAKAN KOMPUTER PAPAN TUNGGAL

ABSTRAK

Projek ini adalah untuk membangunkan pengiktirafan objek dan pengelasan untuk sistem pengawasan keselamatan menggunakan komputer papan tunggal, dan juga projek ini dapat mengenali jenis objek. Pengunaan system televisyen litar tertutup (CCTV) boleh beroperasi secara berterusan atau hanya diperlukan untuk memantau sesuatu peristiwa. Sekiranya, ada sesuatu kejadian berlaku, orang yang bertanggungjawab perlu memeriksa video yang dirakam untuk melihat keadaan yang telah terjadi dan juga CCTV tidak dapat mengesan dan mengenali sebarang objek. Projek ini adalah untuk melaksanakan kaedah pengenalan objek dengan menggunakan pembelajaran mendalam ataupun "Deep learning" untuk mengesan dan mengklasifikasikan sesuatu objek.Laporan projek ini adalah untuk melaksanakan system pengawasan untuk mengesan objek secara siaran langsung bagi mengurangkan aktiviti pengawasan oleh manusia untuk mengklasifikasikan sesuai objek. Dengan menggunakan kaedah pengiktirafan, system ini dapat mengenali dan mengklasifikasikan objek yang datang melalui pandangan kamera dan ciri-ciri untuk klasifikasi akan mengeluarkan objek dari pengiktirafan untuk mengklasifikasikan jenis objek yang datang melalui kamera. Projek ini dibangunkan menggunakan kaedah rangkaian neural yang mendalam dan SSD Mobilenet untuk pengesanan. MobileNets adalah rangkaian saraf yang merupakan rangkaian asas bagi menyediakan ciri-ciri tahap tinggi untuk klasifikasi dan penggunaan SSD adalah lapisan perlingkaran terakhir pada rangkaian asas untuk tugas pengesanan. Dengan menggunakan set data MS-COCO yang mempunyai 80 objek kelas dapat membantu projek ini untuk mengesan objek asas iaitu manusia, haiwan dan kenderaan. Projek ini menggunakan komputer riba dan Raspberry Pi untuk perbandingan ketepatan. Untuk analisis data, ia mengambil data ketepatan dalam masa dan persekitaran yang berbeza untuk menguji peratusan ketepatan dan klasifikasi. Untuk setiap ujian 6 sampel diambil untuk tujuan analisis. Selepas data telah diambil, projek ini akan membandingkan prestasi komputer riba dan Raspberry Pi melalui peratusan ketepatan dan klasifikasi. Hasil menunjukkan bahawa purata ketepatan komputer riba lebih baik iaitu 94.90% dan 83.22% untuk Raspberry Pi. Kesimpulannya, bahawa dengan prestasi Raspberry Pi mampu untuk menggunakan didalam sistem pengawasan kerana factor saiznya yang kecil dan mengunnakan kuasa proses yang rendah.

Object Recognition and Classification for Security Surveillance System using Single Board Computer

ABSTRACT

This project presents a development of object recognition and classification for security surveillance system using single board computer, the system is able to recognize object type. Then, close circuit television (CCTV) systems may operate continuously or only as required to monitor a particular event. If something happen, person in charge will needs to access recorded video to see the situation, then CCTV also cannot detect and recognize any object. The project is to implement an object recognition method with Deep learning to detect object and classify. The report also describes the project to implement surveillance system for real time object detection to reduce monitoring activities by human to classify an object. By using recognition method, the system can recognize and classify any object that come through camera view and classification features will extract the object from recognition to classify type of object that comes through the camera. The project was developed using deep neural network method and SSD Mobilenet for detection. MobileNets is a neural network that is base network provide high level features for classification and SSD use of last convolutional layer on base networks for detection task. By using MS-COCO dataset that has 80 class of object helps this project to detect basic object which are human, animals and vehicle. This project uses a laptop and Raspberry Pi for comparison of accuracy. The data analysis takes accuracy data in different time and environment to test accuracy percentage and classification. For each test, 6 sample collected for analysis purposes. Afterwards, performance comparison between laptop and Raspberry Pi are made to compare the difference between in accuracy and classification performance. The result show that laptop accuracy is better which averaged of at 94.90% for laptop and 83.22% for Raspberry Pi. Raspberry Pi performance is found to be acceptable for surveillance system implementation as it smalls and requires less power.

CHAPTER 1 : INTRODUCTION

1.1 Research Background

The traditional close circuit television (CCTV) is to monitor of any area for security purposes, from the CCTV it will record any scene that happen through the camera that place in a wall. Mostly, CCTV is use for residential or warehouse, security of area is an important consideration in keeping the area from undesirable event or accident, by using this device it will monitor situation of the area. Current surveillance system is common in many areas around the world, and the camera has been introduced as a new form of surveillance, often used in law enforcement, with cameras located at insecurity area. CCTV systems may operate continuously or only as required to monitor a particular event. If something happen, person in charge will need to access recorded video to see the situation, because CCTV features is not capable to detect and recognize any object. Some problem of CCTV implementation is streaming video continuously to capture any of the events that occur in the environment also even when there is no any suspicious object or activity happen. Therefore, it is difficult to trace back to the event because rewinding need to be done manually to the past event location.

The aim of this project is to develop a security surveillance system that is implemented with object recognition and classification features. By using recognition method, the system can detect any object that come through camera view and classification features will extract the object from recognition to classify type of object that coming through the camera. Vision system on a Single Board Computer (SBC) with camera, the system will be able to recognize and classify basic object that comes into view. This will help to improve current CCTV to recognize any object automatically for surveillance monitoring without needs of man power to focus on the CCTV footage to classify type of object.

Then, this study implement deep neural network (DNN) that learn from many layers between input to provide probability from data set that have thousand trained image for recognition and using MobileNet with single shot detector (SSD) which is base network that provide high level features for classification to classify type of object in dataset. This project will improve current security surveillance system by using computer vision that allow to process image and video. By processing this will provide recognition for the object and classify type of object which cannot be recognize in current CCTV features. This will reduce monitoring time to visualize type of object in video footage.

1.2 Problem Statement

Current implementation with security systems has some problems regarding the system capabilities because of security system currently not able to recognize and classify object type. To recognize and classify object that come through CCTV, it will need manual monitoring by person in charges. This will be lead to any suspicious object can be overlooked because of human errors that can create unsafe situation.

For person in charges to check the footage from CCTV will increase time consuming because their need to check entire recorded minutes to find the object. The project is for recognize and classify any object from camera for security surveillance, using camera module that attach to SBC and will process all the video to recognize and classify basic object. CCTV only use simple background subtraction technique to record activity, and not able to differentiate objects in view causing any detected motion to be recorded.

From the CCTV video footage, it need to be clear to seen the object to classify type of object in the footage. If the object image is not clear it will lead to less accurate type of object because the object image is not clear enough to classify. This project will provide high accuracy for classification to classify type of object.

1.3 Objectives

The objectives of the projects are :

I. To implement object recognition using Deep Neural Network to detect recognize objects which is human class, animal class and vehicle class.

II. To develop a surveillance system capable of differentiating objects that come through the Raspberry Pi camera.

III. To provide high accuracy percentage of object type classification in live stream video .

1.4 Research Question

- How to recognize object using Deep Neural Network ?

- What effect the accuracy percentage ?

1.5 Research Scopes

The project is focus to identify object that come into camera view, and classify type of object that show in the camera. Using camera module that attach to Single Board Computer (SBC) which is Raspberry Pi , the system will process all the image processing will use feature from OpenCV and TensorFlow to classify type of object from the image processing. MS-COCO (common object in context) dataset will be used which have 80 object categories with more than 200k pre trained image for recognition and classification. COCO dataset will store into media storage as a pre-train model, this to able the system to recognize and classify the object that comes into view based on the model.

Besides that, this project also has its own limitation because its only uses Raspberry Pi that has limited performance to process all the data. Moreover, only certain type of object will be detects for this project which is Human (person),Vehicle (car) and animals (Cat) by the system because low of SBC processing performance. Then, camera module use for this project cannot detect any object in a dark places because doesn’t has infra-red features.

1.6 Expected Outcome

- The project is able to recognize object in the view from the camera, and also by using OpenCV and tensor flow features for image and video processing it should able to classify detected type of object.

- This system also will show an object in the real-time streaming video from camera as type of object such as person, cat and car. Also, this project

- Expected more than 50% accuracy percentage on recognition of object type from video.

1.7 Report Outline

Chapter 1:The first chapter give a brief overview on the project introductions, problem statement, objectives and also the project scope.

Chapter 2:The second chapter explains about previous related works regarding the project and a thorough research on projects implementation.

Chapter 3:This chapter discusses on research methodology which consist of proposed design for the project, technical aspects used and workflow of the project.

Chapter 4:This chapter briefly discusses the result obtained through this system with graph and images.

Chapter 5:This chapter concludes about the whole systems and also includes future recommendations.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

In this chapter, the research explains about related work that can be refer for the project, From related work data, system can be use output for critical analysis to compare the project with previous related project and collect what needs for the project. This chapter also improve understanding about object recognition, classification and detection, then explains about the idea of the project.

2.2 Single Shot Multi-box Detector

Single shot multibox detector (SSD) is method for detecting objects in images using deep neural network. SSD discretize the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location (Liu et al., 2016). At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustment to the box to better match the object shape. Additionally, the network combines prediction from multiple feature map with different resolutions to naturally handle object of various size.

SSD is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. This make SSD easy to train and straight forward to integrate into systems that require a detection component. SSD has competitive accuracy to methods that utilize an additional object proposal step and much faster.

SSD a single shot detector for multiple categories that is faster than You Only Look Once (YOLO) method and significantly more accurate. The core of SSD is predicting category scores and box offsets for a fixed set of default bounding boxes using small convolutional filters applied to feature maps. To achieve high detection accuracy in SSD method, it is by using separate prediction by aspect ratio. These design features lead to simple end to end training and high accuracy, even on low resolution input images.

2.3 Single Shot Detector Model

SSD only needs an input image and ground truth boxes for each object during training. In a convolutional method, evaluation using small set of default boxes of different aspect ratios at each location in several feature maps with different scales in figure 2.1 SSD Feature map (8 x 8 and 4 x 4 in (b) and (c)). For each default box, prediction of both shape offsets and the confidence for all object categories on the figure 2.1 (c). On the training, first match of default boxes to the ground truth boxes. For example, matched two default boxes with the cat and one with the dog, which are treated as positives and the rest as negatives.

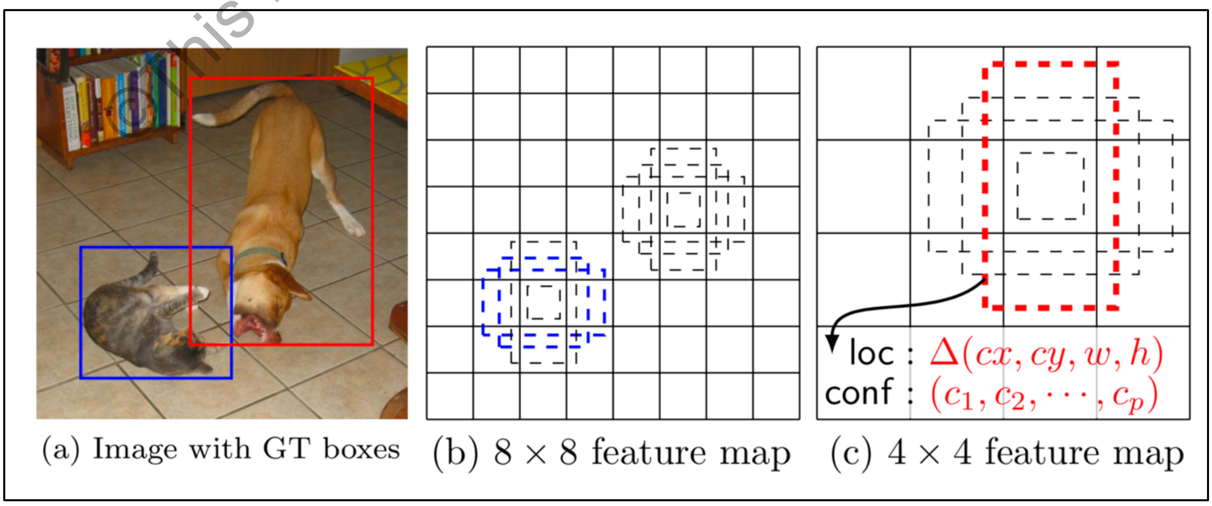

Figure 2.1 SSD Feature Map

SSD approach is based on a feed forward convolutional network that produces a fixed size collection of bounding boxes and scores for the presence of object class instances in those boxes, followed by a non-maximum suppression step to produce the final detection (Liu et al., 2016). The early network layers are based on a standard architecture used for image classification which is the base network.

Multiscale feature maps for detection added convolutional feature layers to the end of truncated base network. These layers decrease in size progressively and allow prediction of detection at multiple scales. The convolutional model for predicting detection is different for each feature layer. Convolutional predictors for detection with each added feature can produce a fixed set of detection prediction using set of convolutional filters. These are indicated on top of the SSD network architecture in Figure 2.2. For a feature layer of size m x n with p channels, the basic element for predicting parameters of potential detection is a 3x3xp small kernel that produces either a score for a category or a shape offset relative to the default box coordinates. At each of the m x n location where the kernel is applied, it produces an output value.

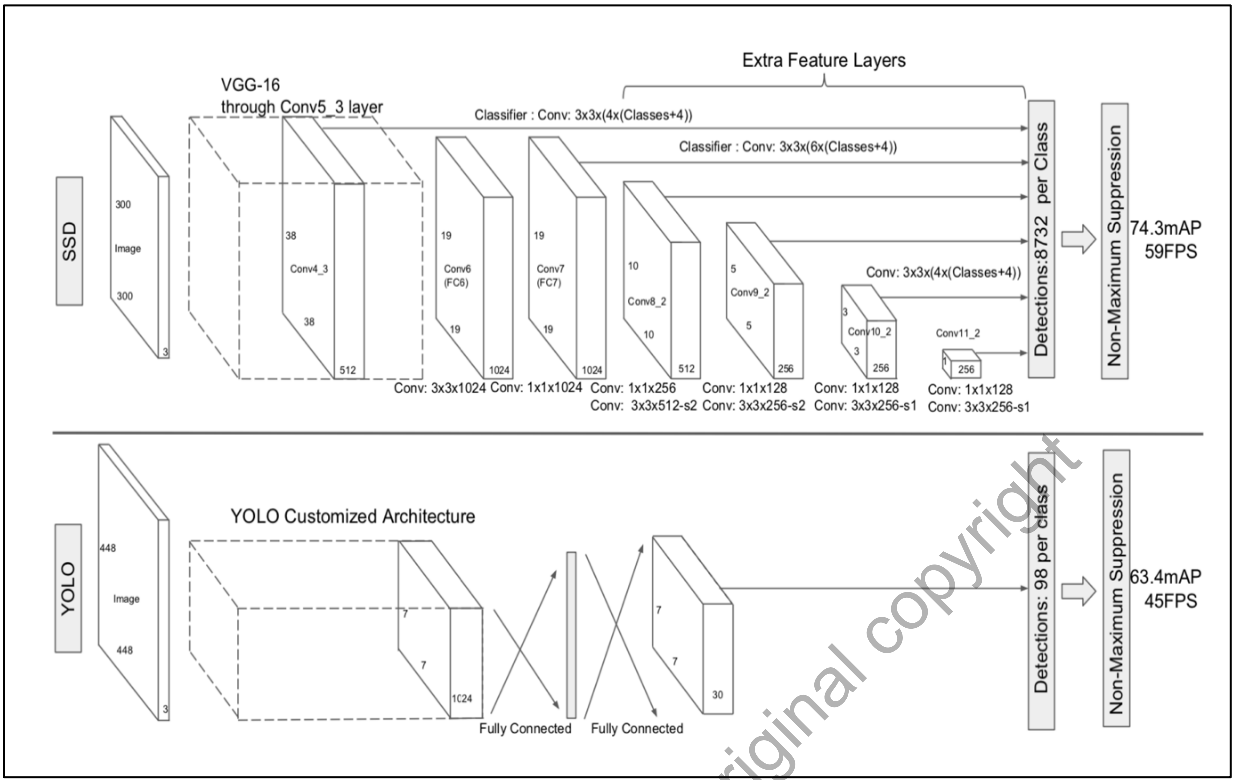

Figure 2.2 SSD Network Architecture

2.4 MobileNets

MobileNets is a family of mobile first computer vision models for TensorFlow, designed to effectively maximize accuracy while being mindful of the restricted resources for an on device or embedded application. MobileNets are small, low-latency, low-power models parameterized to meet the resource constraints of a variety of use cases. It can be built upon for classification, detection, embeddings and segmentation. Figure 2.3 shows example use cases included detection, classification, attribute and location.

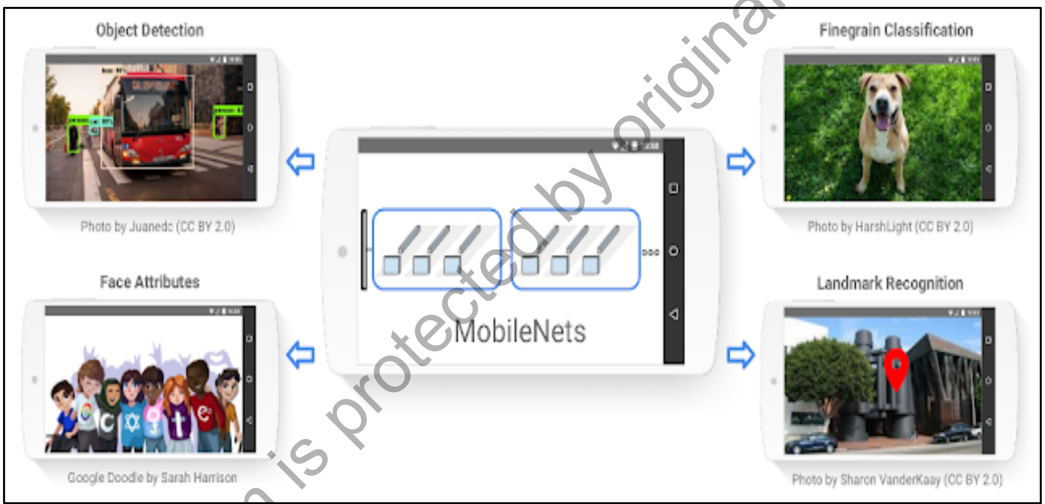

Figure 2.3 Use Cases Example

MobileNets is a Convolutional Neural Network (CNN) architecture model for Image Classification and Mobile Vision. There are other models as well but what makes MobileNets special that it very less computation power to run or apply. This makes it a perfect fit for Mobile devices, embedded systems and computers without GPU or low computational efficiency with compromising with the accuracy of the results. It is also best suited for web browsers as browsers have limitation over computation, graphic processing and storage.

MobileNets architecture for mobile and embedded vision applications is proposed, which are based on a streamlined architecture that uses depthwise separable convolutions to build light weight deep neural networks. There are two simple global hyper-parameters that efficiently trade-off between latency and accuracy are introduced. The core layer of MobileNet is depthwise separable filters, named as Depthwise Separable Convolution. The network structure is another factor to boost the performance. Finally, the width and resolution can be tuned to trade-off between latency and accuracy.

Depthwise separable convolutions which is a form of factorized convolutions which factorize a standard convolution into a depthwise convolution and a 1×11×1 convolution called a pointwise convolution. In MobileNet, the depthwise convolution applies a single filter to each input channel. The pointwise convolution then applies a 1×11×1 convolution to combine the outputs the depthwise convolution. The following figure 2.4, illustrates the difference between standard convolution and depthwise separable convolution. Which the standard convolutional filters in (a) are replaced by two layers, depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter.

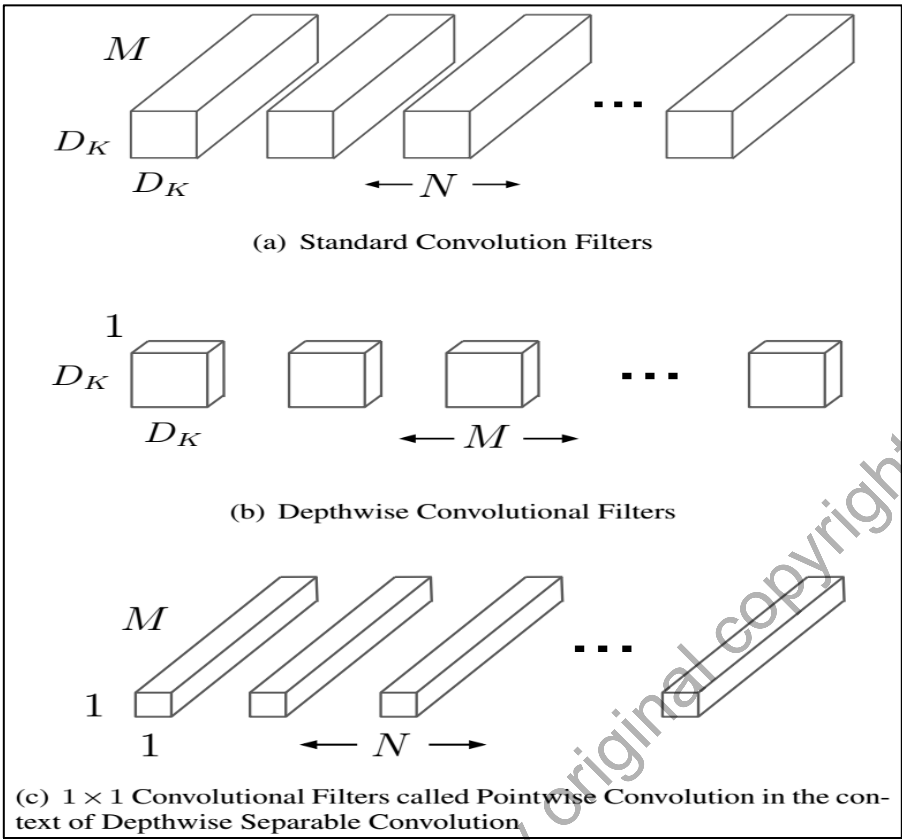

Figure 2.4 Standard and Depthwise Convolution Filters