Efficient Object Detection using Transfer Learning

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

Submitted By
Harshdeep Singh
(801732019)

Under the supervision of:
Dr. R. K. Sharma
Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

July 2019

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*Efficient Object Detection using Transfer Learning*", in partial fulfilment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. R. K. Sharma.*

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Harshdeep Singh

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Dr. R. K. Sharma 31.7.2019
Professor
Computer Science & Engineering
TIET, Patiala

i

ACKNOWLEDGMENT

No volume of words is enough to express my gratitude towards my guide ***Dr. R. K. Sharma***, Department of Computer Science & Engineering, Thapar University, Patiala, They have been very concerned and has aided for all the materials essentials for the preparation of this thesis report. They have helped me to explore this vast topic in an organized manner and provided me all the ideas on how to work towards a research-oriented venture.

I am also thankful to **Dr. S.S. Bhatia**, Dean of Academic Affairs, **Dr. Maninder Singh**, Head of Computer Science & Engineering Department and **Dr. Ashutosh Mishra**, P.G. Coordinator, for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues who were always there at the need of the hour and provided with all the help and facilities, which I required, for the completion of my thesis work.

Most importantly, I would like to thank my parents and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Harshdeep Singh

(801732019)

ii

ABSTRACT

Object detection in automated surveillance images or video is an extremely monotonous process for monitoring for crowded scenes and a variety of sensible objects can be restricted in surveillance images or video. An appropriate machine learning technique can help to train the object detection system in identifying random activities during surveillance. To this end, we present Efficient Object Detection using Transfer Learning that can be used as a tool for object detection in surveillance images or videos using the concept of artificial intelligence. The main intention of the proposed object detection system is to improve the detection time and accuracy by using the concept of Convolutional Neural Network (CNN) as artificial intelligence technique. In this paper we present CNN based VGG-16 model for object detection, which is the combination of multiple layer of hidden unit with the optimized feature by using transfer learning. Here CNN is used for classifying the random activity into objects from the surveillance images or videos based on the transfer learning which is used for the selection of optimal feature sets. Further, Self adaptive transfer learning is adopted to efficiently solve optimization problems in the continuous search domain to select the best possible feature to segregate the pattern of object. The main contribution of this research is validation of proposed system for the large scale data and we introduce a new large-scale dataset of 50 class images. Dataset consists of total 5000 long and untrimmed real-world surveillance indoor images. The experimental results of the proposed system show that our designed for object detect achieve significant improvement on detection system performance as compared to the state-of-the-art approaches. In this paper, to validate the proposed model we provide the comparison of existing results of several recent deep learning baselines on object detection. The real-time object detection in surveillance images or videos sequences using transfer learning based on CNN with feature extraction technique is implemented using anaconda3 python Software.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Context

Human brain is the most intelligent organism, which leverages cognitive reasoning and processes visual data for high level semantic interpretation and gaining selective situational awareness. Over the previous few years, the computer visualization researchers have attempted similar capabilities to video analysis systems. With the accessibility of cheaper visual sensors, the need for understanding large quantity of video data is also increased. A significant application of video analysis is intelligent surveillance system, which is used for object detection/tracking/prediction. The object detection system example is shown below in Figure 1.1 (Source: Ke Xu, Xinghao Jiang *et al.* 2018).

Figure 1.1: Irrelevant Object Detection in Image

In the Figure 1.1, red colour boundary is denoting the object which is different from the actual event. Video processing is a technique used for examining video contents in order to get an idea of the scene that the video represents. Video analytic methods are inspired by the requirement of generating machine algorithms which imitate the abilities of humans and other living organism visual systems. Video analysis is an important part of various technologies such as robotics, video surveillance, and multimedia. There are a number of research works for video analysis carried out in various fields like computer science,

statistics, signal and image processing, and system theory. The Video processing method has created changes in multimedia applications with products including Digital Versatile Disk (DVD), High Definition Television (HDTV), video cameras, and Digital Satellite Systems (DSS). There are four fields in video processing. They are video compression, video indexing, video segmentation, and video tracking.

1.2 Object Detection

New objects are detected using a histogram thresholding on the likelihood of pixels classified as background. The physics of the objects determines the minimum size of an object, providing a criterion whether or not a blob can be an object. Too small objects are re-labelled as the background. Object detection is a computer vision procedure for finding occurrences of objects in pictures or recordings. Object detection calculations ordinarily influence AI or profound figuring out how to create significant outcomes. At the point when people take a gander at pictures or video, we can perceive and find objects of enthusiasm inside a matter of minutes. The objective of object detection is to recreate this knowledge utilizing a computer.

1.2.1 Why Object Detection Matters

Object detection is a key innovation behind cutting edge driver help frameworks (ADAS) that empower vehicles to distinguish driving paths or perform walker detection to improve street wellbeing. Object detection is likewise valuable in applications, for example, video observation or picture recovery frameworks.

1.2.2 How it Works

1.2.2.1 Object Detection Using Deep Learning

You can utilize an assortment of strategies to perform object detection. Well known profound learning–based methodologies utilizing convolutional neural systems (CNNs, for example, R-CNN and YOLO v2, consequently figure out how to identify objects inside pictures. You can browse two key ways to deal with begin with object detection utilizing profound learning:

Make and train a custom object identifier: To prepare a custom object identifier starting with no outside help, you have to plan a system engineering to get familiar with the highlights for the objects of intrigue. You additionally need to incorporate an enormous arrangement of named information to prepare the CNN. The consequences of a custom object locator can be wonderful. So, you have to physically set up the layers and loads in the CNN, which requires a ton of time and preparing information.

Utilize a pre-trained object locator: Many object detection work processes utilizing profound learning influence move learning, a methodology that empowers you to begin with a pre-trained system and after that tweak it for your application. This strategy can give quicker outcomes in light of the fact that the object locators have just been prepared on thousands, or even millions, of pictures.

1.2.2.2 Object Detection Using Machine Learning

AI strategies are additionally generally utilized for object detection, and they offer unexpected methodologies in comparison to profound learning. Regular AI systems include:

- Total channel highlights (ACF)
- SVM order utilizing histograms of arranged slope (HOG) highlights
- The Viola-Jones calculation for human face or chest area detection

1.3 Other Object Detection Methods

- Notwithstanding profound learning–and AI based object detection, there are a few other normal strategies that might be adequate relying upon your application, for example,
- Picture division and mass examination, which uses basic object properties, for example, size, shape, or shading
- Highlight based object detection, which uses include extraction, coordinating, and RANSAC to assess the area of an object

1.3.1 Object Representation

Each object is spoken to by its shape and appearance. Object portrayal dependent on shape is depicted in the accompanying:

Points: Here, the objects are spoken to by a point which means centroid or a gathering of points. Point portrayal is utilized for object following in a picture where the object possesses little locales.

Crude geometric shapes: Here, the object is spoken to by crude geometric shapes, for example, an oval, square shape, etc. Relative, interpretation, and projective (homographic) change are utilized to demonstrate the movement of the object in this kind of portrayal. Crude geometric shapes are valuable for both inflexible and non-unbending objects.

Object silhouette and form: Contour portrayal is utilized to characterize an object's limit. The object silhouette speaks to the area which is inside the form. The object silhouette and shape portrayal are utilized for following non-unbending objects.

Articulated shape models: Articulated objects will be objects, in which parts of the body are connected together by joints. A case of a verbalized object is the human body where the various parts, for example, legs, hands, middle, feet, and head are associated by joints. The pieces of the body are connected by kinematic movement models, for example, a joint edge. A verbalized object is spoken to by chambers or ovals.

Skeletal models: The average hub change is connected to the object silhouette from which the object skeletal is extricated. Skeletal models are utilized to speak to the state of the objects for recognizing those. Skeletal models are commendable for speaking to unbending and explained objects. Appearance based object portrayals are depicted in the accompanying segments:

Likelihood densities of object appearance: The assessment of the likelihood densities of the object appearance is either parametric or non-parametric. Here, highlights, for example, shading and surface are determined from the locales of the picture characterized by the shape models.

Templates: Templates are utilized to encode the presence of the object delivered by a solitary view and it is commendable for following objects when there is no adjustment in the situation of the object during the season of following. Object silhouettes or crude geometric shapes are utilized to characterize the formats. Layouts consider both the appearance and spatial data of the objects, which is a noteworthy advantage of the format.

Dynamic appearance models: The shape and presence of the objects are demonstrated to produce the dynamic appearance models. The state of the objects is portrayed by a gathering of tourist spots. The milestone lives on either the limit of the objects or inside the object area. The shading, inclination extent or surface is put away as the appearance vector for each milestone. In the preparation period of the dynamic appearance model, the shape and presence of the object is developed from the arrangement of tests.

Multi-view appearance models: In Multi-see appearance models, the subspaces are created from the given perspective on the object to speak to the object in various perspectives. Subspace techniques are utilized to speak to the shape and presence of the object. Instances of subspace strategies are the Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

1.3.2 Types of Object Detection

There are two approaches in object detection. The first one is recognition-based object detection and the second one is motion-based object detection.

1.3.2.1 Recognition-Based Object Detection

In recognition-based Detection, the recognition of the object is modified. The object recognition is performed in consecutive images and the position of the image is identified. The major advantages of recognition-based object detection are, it evaluates the rotation and translation of the object and the detection is performed in three dimensions. The disadvantage of this type of object detection is that it only detects the objects that are recognizable, because the object recognition needs a high high-level operation. Therefore, recognition-based object detection systems have limited performance.

1.3.2.2 Motion-Based Object Detection

The Motion-based detection system detects the moving object by finding the motion of the object. The advantage of motion-based object detection is to track objects without considering the size or shape of the object. The two methods in motion-based detection are optical flow tracking and motion-energy tracking.

1.3.2.3 Optic Flow Object Detection

Optic flow object detection, detects the motion of the objects from videos taken by a fully active camera. Optic flow object detection can efficiently detect the object's motion in an unconstrained domain. The limitation of this object detection system is to qualitatively evaluate the motion of the object and it is more unfair than the quantitative method. The Determination of the entire optic-flow field requires high cost. Hence, the alternative approach which identifies some features of the objects in the video and detects the motions of that objects. The limitation of this approach is that the features of the object in each frame must be identical to the features of the objects in the previous frame. This problem will be intensified in the videos taken by the active camera. Since the image sequence observed is dynamic, some features are moved far away from the field of vision and the new features will be moved into the field of vision which increases the strain for finding the matching features. Velocity field withdrawal is another problem in optic flow object detection and hence, optic flow object detection is not suitable for real time applications.

1.3.2.4 Motion Energy Object Detection

The alternative approach for object detection from the motion of an object is motion energy object detection. In this approach, the noise in the image sequence is filtered and the image is segmented into movement regions and inactivity regions, by finding the temporal derivative of an image and thresholding at an appropriate level. Generally, the temporal derivative can be calculated by straightforward image subtraction, which creates noise and produces vague values. The performance of the image subtraction can be improved by using spatial edge information for finding the moving edges. The most successful and widely used method for motion detection combines image subtraction and spatial information. Motion energy object detection is computationally very simple and it is

worthy for pipeline architectures. The limitation of motion energy object detection is that it detects the motion of the pixel but does not measure the quantity. Hence, additional information like the focus of expansion cannot be determined. This type of object detection is not applicable for videos from an active camera. The active camera systems persuade the obvious motion of objects in the sequence of images, so that the indemnity for the obvious motion is made prior to the motion energy object detection. The general object detection techniques use rough shapes like ellipses and rectangles for representing the objects, while the active contour-based detection presents the more comprehensive representation of the object. Active contour-based detection is harder than the general object detection techniques for detecting the same object, because contour-based object detection tries to find the detailed representation of the objects like object boundary and object boundary detection which is harmed by the effects from the background disruption. In videos captured by stationary cameras, the object movement regions are determined by the background subtraction process and the edges of the movement regions are discovered to produce the object contours. On the other hand, in videos captured by the non-stationary cameras, background subtraction is not used for finding the object movement regions, which makes the active contour-based object detection harder for videos captured by the non-stationary cameras. However, active contour-based tracking is a more powerful approach for object detection in recent years.

Object contours are represented by two types: The first is the explicit representation and the second is the implicit representation. The explicit representation describes the features of the parameterized curves like snakes and the implicit representation describes the contour by a signed distance map like level sets. Compared to the explicit representation, the implicit representation is favourable because it produces a strong numerical solution and it performs well even if there are topological changes. Active contour-based detection is divided into three types of methods. They are,

1) Edge-based,

2) Region-based, and

3) Shape prior based methods.

1) Edge-Based Methods: The Edge-based method examines the local information like the grey level gradient throughout contours for representing the object contours. An example for the edge-based method is the snake model. The edge-based methods are simple and effective for identifying the object contours. The drawbacks of the edge-based methods are described in the following: The first contour must be described close to the objects and the object contours described in the analogous regions need not be optimized.

2) Region-Based Methods: Depending on the principle of statistic quantities like variance and mean, the region-based methods separate the objects and the framework from the image. In region-based methods, the past knowledge about the object's colour and texture are combined for evaluating the contour. The object appearance models are used to describe the previous knowledge about the object colour. Examples of object appearance models are colour histograms, and Gaussian Mixture Models (GMMs). The major advantages of region-based methods are robustness and correctness. The disadvantage of the region-based method is that it assumes that the pixel values are independent of each other for estimating the posterior probability function so that the evaluated contour is easily affected by the disruptions of texture or colour similarities across the background and object.

3) Shape Prior-Based Methods: Shape prior-based methods are used to model the shape of the objects according to statistics and recover the object contours which are affected by occlusion, and blurring. Adaptive shape-based methods twist the contour sections that are undisturbed and determine them by the colour features; on the other hand, they determine the disturbed contour sections globally. The active shape prior-based methods need continuous updating to adopt, according to the changes in the object shape. The recent method, which updates the shape prior-based methods, does not process the multiple modern shapes. A simple data fitting process for regular movements of non-rigid objects described and it is the dynamic shape model which has no knowledge about changes in shape. This model imagines that the hidden movement of the object is close to the periodic motion of the object.

There are two types of approaches in visual object detection. The first one is the bottom-up approach and the second one is the top-down approach. Bottom-up approaches are used

to restore the target by evaluating the contents of the image. An Example of bottom-up detection is the reconstruction of a parametric shape through the use of curve fitting. On the other hand, top-down approaches produce and assess state hypotheses depending on the target states and track the target by assessing and checking these hypotheses on observations of the image. The robustness of the bottom-up approaches depends on the image analysis, because activities such as grouping and tracing are deluged by the noise and clutters in the image, and they are very efficient. In contrast, the top-down approaches are slightly based on image analysis because the image can be analysed based on the target hypotheses and the performance is calculated by hypotheses generation and verification. The top-down approach requires a huge number of hypotheses in order to attain robustness. Hence, it requires additional computation for hypotheses evaluation. In some situations, the top-down and bottom-up approaches are mingled to increase the robustness. At the same time, the computation can be reduced. An adaptive object detection method improves the tracker's performance by adaptively choosing the features of the object that differentiate the object from its background. One way to increase the performance of the tracker is to combine the different cues such as texture, shape and colour.

1.4 Transfer Learning

Transfer learning is a technique of machine learning in which a model created for a task is reused as the starting point for a second task model. It is a popular approach in deep learning where pre-trained models are used as the starting point for computer vision and natural language processing tasks due to the vast computational and time resources needed to develop neural network models on these issues and the huge skill jumps they provide on related issues.

1.4.1 What is Transfer Learning?

Transfer learning is a technique of machine learning that re-purposes a model trained on one task on a second related task. There are multiple definitions of learning transfer:

1. Transfer learning and domain adaptation refer to the situation where what has been learned in one setting is exploited to improve generalization in another setting.

2. Transfer learning is an optimization that allows rapid progress or improved performance when modelling the second task.
3. Transfer learning is improving learning in a new task by transferring knowledge from a task already learned.
4. Transfer learning is linked to issues like multi-task learning and concept drift and is not just a research area for deep learning.

Transfer learning, however, is popular in deep learning given the enormous resources needed to train deep learning models or the large and challenging datasets that train deep learning models. Only if the model characteristics learned from the first task are general, transfer learning works in deep learning. We first train a base network on a base dataset and task in transferring learning, and then repurpose or transfer the learned features to a second target network to be trained on a target dataset and task. This method will tend to operate if the characteristics are general, which means that they are appropriate for both the base and target functions, rather than the base task specific.

How Transferable are Features in Deep Neural Networks?

This type of learning transfer used in deep learning is referred to as inductive transfer. This is where the scope of possible models (model bias) is reduced by using a model fit for a different but related task in a beneficial way. The following Figure 1.2 (Source: Transfer learning, Lisa Torrey and Jude Shavlik) shows this process visually.

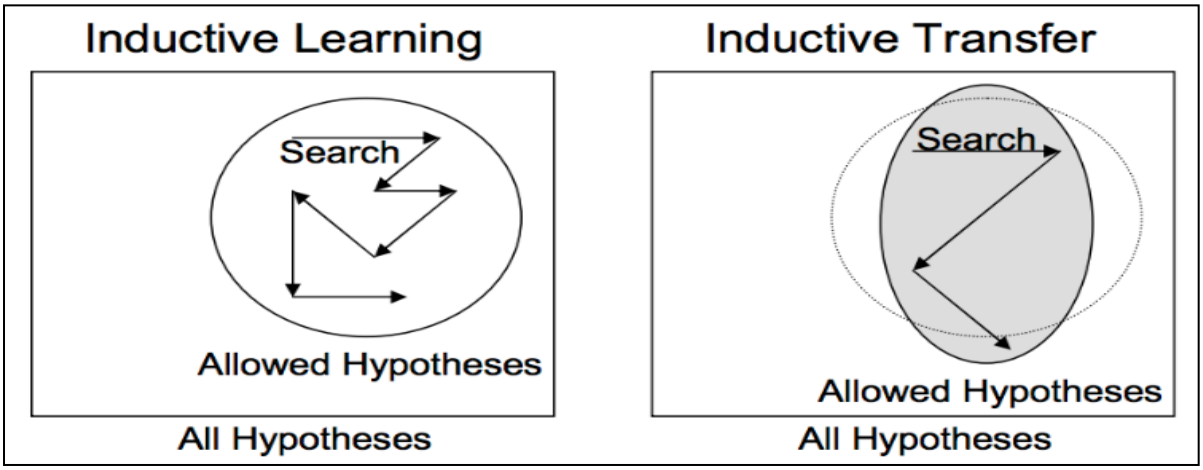

Figure 1.2: Approach of Transfer Learning

1.4.2 How to use Transfer Learning?

On your own predictive modeling issues, you can use transfer learning. Two common approaches are as follows:

1) Develop Model Approach
2) Pre-trained Model Approach

1.4.2.1 Develop Model Approach

Select source task: You need to pick an associated predictive modeling issue with an abundance of information where input information, output information, and/or ideas learned from input to output information are linked.

Develop source model: Next, for this first task, you need to create a skilful model. To guarantee that some learning features have been carried out, the model must be better than a naive model.

Reuse model: It is then possible to use the model fit on the source task as the starting point for a model on the second interest task. Depending on the modeling method used, this may require using all or sections of the model.

Tune model: The model optionally needs to be adjusted or modified for the task of interest on the input-output pair data available.

1.4.2.2 Pre-Trained Model Approach

Select source model: From accessible models, a pre-trained source model is selected. Many study organizations are releasing models on big and difficult datasets that can be included in the pool of model candidates to choose from.

Reuse model: It is then possible to use the pre-trained model as the starting point for a model on the second interest task. Depending on the modeling method used, this may require using all or sections of the model.

Tune model: The model may optionally need to be adjusted or modified for the task of interest on the input-output pair information available.

1.4.3 When to use Transfer Learning?

Transfer learning is an optimization, a time-saving shortcut or performance improvement. Generally speaking, it is not obvious that the use of transfer learning in the domain will be beneficial until after the model has been developed and assessed. There are three possible benefits to look for when using transfer learning:

Higher start: The initial skill of the source model (before refining the model) is higher than it would otherwise be.

Higher slope: During source model instruction, the level of skill improvement is steeper than it would otherwise be.

Higher asymptote: The model's converged ability is better than it would otherwise be.

1.4.4 Pre-Trained Models in Deep Learning

- VGG-16
- Inception V3
- VGG 19
- Inception V3
- ResNet-50
- XCeption

In the proposed model we used VGG-16 in Deep Learning and their description is written as: The VGG-16 model is a 16-layer (convolution and fully linked) network based on ImageNet's database, designed for image recognition and classification issues. In their paper ' Very Deep Convolutional Image Recognition Networks, ' this model was built by Karen Simonyan and Andrew Zisserman. The architecture of the VGG-16 model is depicted in the following Figure 1.3 (Source: https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning).

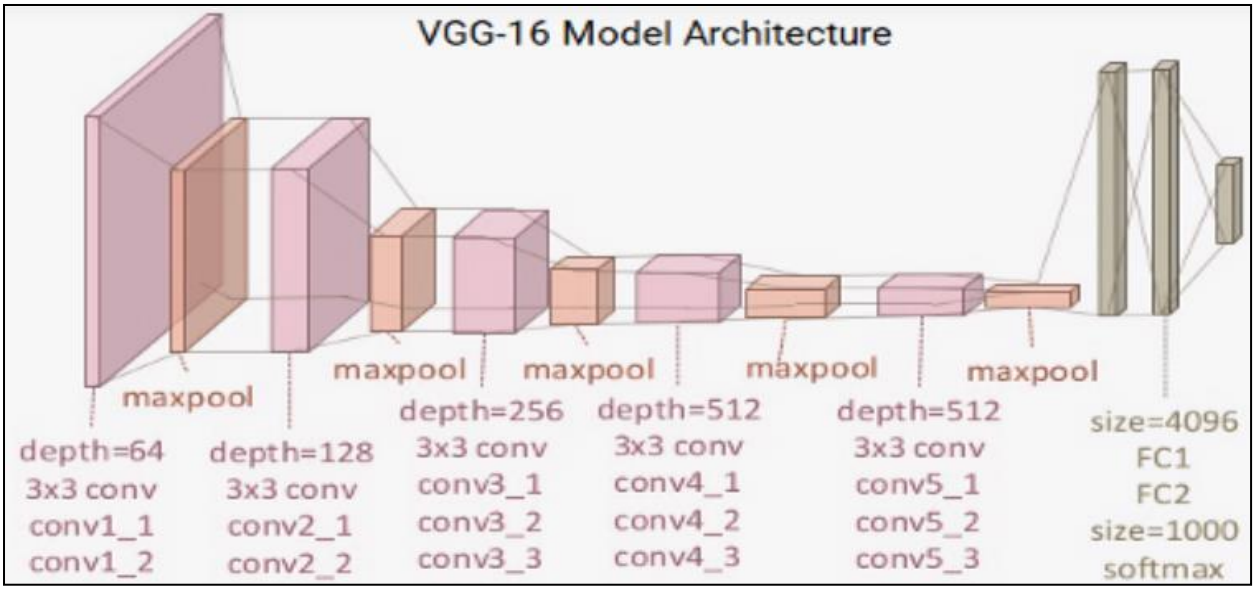

Figure 1.3: CNN Based VGG-16 Model Architecture

One can see that we have a total of 13 number of convolution layers using 3 x 3 convolution filters along with max pooling layers for down sampling and a total of two fully connected dense hidden layers of 4096 units in each layer followed by a dense layer of 1000 units, where each unit represents one of the image categories in the ImageNet database. We do not need the last three layers as we will use our own fully connected layers to predict if pictures are going to be a dog or a cat. We are more concerned with the first five blocks so we can use the VGG model as an effective extractor of features. For one of the models, by freezing all five convolution blocks, we will use it as a easy feature extractor to ensure that their weights are not updated after each epoch. For the last model, we will apply the approach of fine-tuning to the VGG model, where we unfreeze the last two blocks (Block 4 and Block 5) so that their weights get updated in each epoch (per batch of data) as we train our own model. We depict the previous architecture along with the two versions that we will use in the following block diagram (basic feature extractor and fine tuning) so that you can get a better visual perspective. Thus, we are mostly concerned with leveraging the convolution blocks of the VGG-16 model and then flattening the final output (from the feature maps) so that we can feed it into our own dense layers for our classifier.

1.4.4.1 Pre-Trained CNN Model as a Feature Extractor

Let's leverage Keras, load up the VGG-16 model, and freeze the convolution blocks so that we can use it as just an image feature extractor. This is the most important step of working with image data. During image pre-processing, we simultaneously prepare the images for our network and apply data augmentation to the training set. Each model will have different input requirements, but if we read through what ImageNet requires, we figure out that our images need to be 224x224 and normalized to a range. To process an image in anaconda (jupyter notebook), we use transforms, simple operations applied to arrays. The validation (and testing) transforms are as follows:

- Resize
- Center crop to 224 × 224
- Convert to a tensor
- Normalize with mean and standard deviation

The end result of passing through these transforms is tensors that can go into our network. The training transformations are similar but with the addition of random augmentations.

Arguments

- include_top: Whether to include at the top of the network the 3 fully connected layers.
- Weights: one of None (random initialization) or 'ImageNet' (pre-training on ImageNet).
- input_tensor: Optional Keras tensor (i.e. layer output. Input ()) to be used as the model's picture input.
- input_shape: optional shape tuple, only to be specified if include_top is False (otherwise the input shape has to be (224, 224, 3) (with 'channels last' data format) or (3, 224, 224) (with 'channels_first' data format). It should have precisely 3 channels of inputs, and not less than 32 should be the width and height. It would be one valid value, for example (200, 200, 3).
- Pooling: Optional mode for extraction of features when include top is False.

1. None implies the model's output will be the last convolutional layer's 4D tensor output.
2. ' Avg ' implies that the output of the last convolution layer will be applied to the global average pooling and therefore the output of the model will be a 2D tensor.

- 'max' In other words, global max pooling will apply.
- Classes: Optional number of classes in which pictures can be classified, to be specified only if include top is True and no weights argument is indicated.
- **Returns**

A Keras Model instance for an efficient object detection using transfer learning

1.5 Convolution Neural Network

CNN (Convolution Neural Network) is similar to ANN (artificial neural network), Which is primarily used for image classification. This algorithm can be used in different areas such as object detection, faces, tumour, heart rate and plant leaf disease.

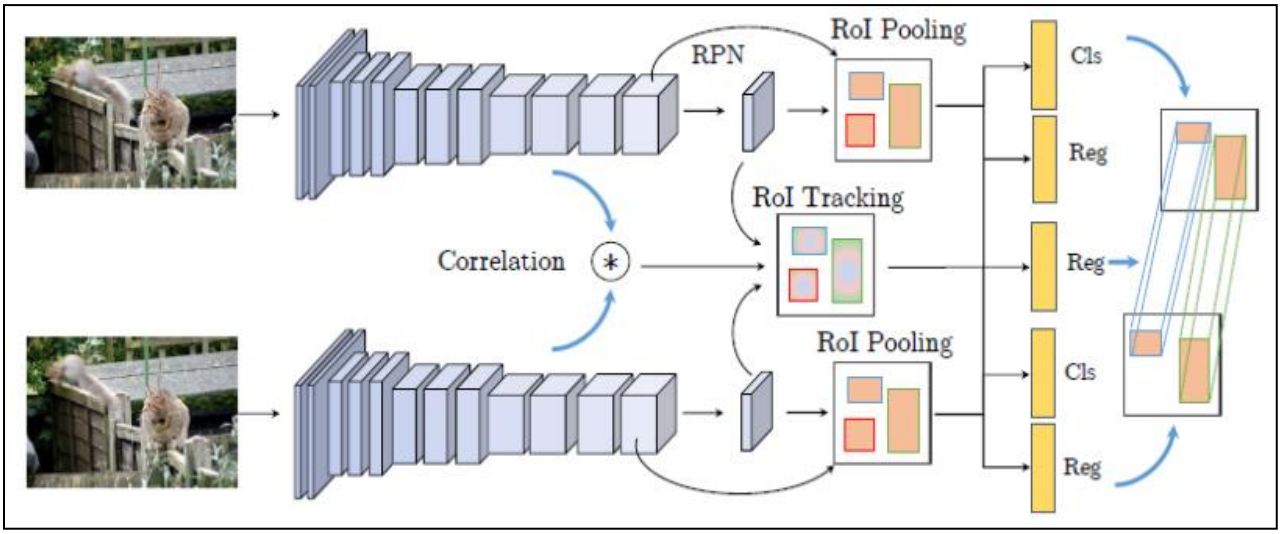

Figure 1.4: CNN Architecture

Above Figure 1.4 (Source: https://www.robots.ox.ac.uk/~vgg/research/detect-track/) represents the CNN model for proposed efficient object detection using transfer learning with different input, output and hidden layers. The layers of CNN are given as:

1.5.1 Convolution Map

In CNN, the convolution layer is a basic element and the objective of convolution is to extract features from the input image using filters. It consists of a group of learnable square filters which helps to find out the appropriate feature sets. Each filter is applied to the raw values of the input image data.

1.5.2 Max-Pooling Map

In the CNN architecture of proposed model, convolution layers are followed by sub-sampling layers and act as a unique feature extraction approach. A layer of max-pooling is an alternative of feature selection but in this work, we used feature of Region of Interest (ROI) to increase the chances of feature uniqueness by using the transfer learning as an optimization technique. The output of max-pooling layer is passes to the classification model which is used as a maximum activation value and creates a structure of model.

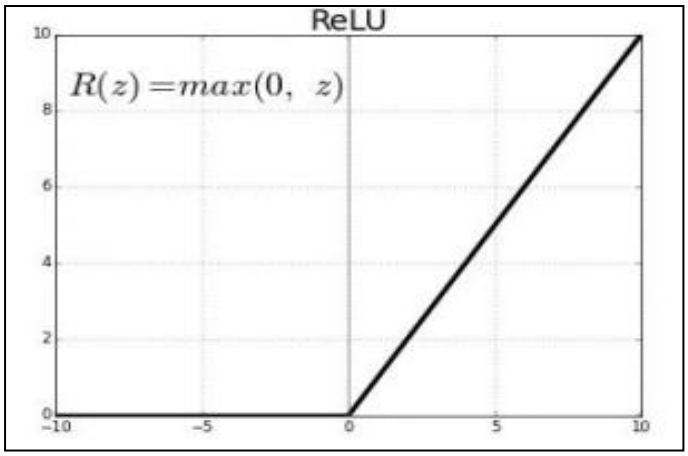

Figure 1.5: Activation Function

ReLu function helps in faster training and addressed the vanishing gradient problem. ReLu activation is shown in Figure 1.5. Image downloaded from http://csci431.artifice.cc/notes/deep-learning .The example of max-pooling is depicted in Figure 1.6 and average pooling is given in Figure 1.7. Images available at https://www.researchgate.net/figure/Illustration-of-Max-Pooling-and-Average-Pooling .

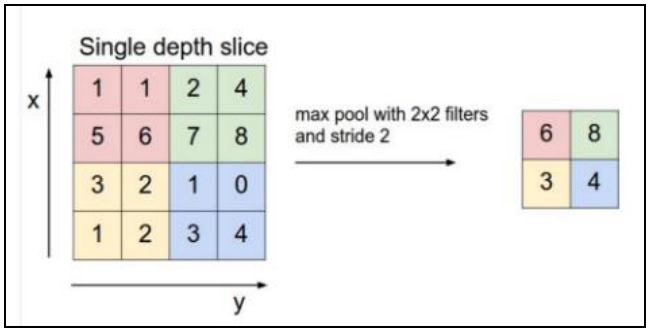

Figure 1.6: Example of Max Pool Layer

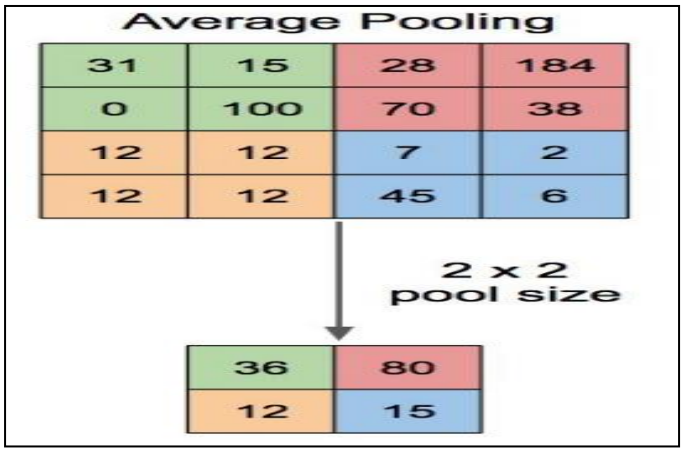

Figure 1.7: Example of Average Pooling Layer

1.5.3 Classification Model

In the CNN, we use fully dense layers in the classification phase where each neuron offers a complete link to all learned optimized function maps from the past CNN layer. To calculate the results of the class, these dense layers are based on the activation function. The classifier input is a vector of optimized features arising from the fitness function optimization method and the output is a likelihood that an picture belongs to a specified class. CNN contains multiple neurons arranged in different layers. The neurons in the adjacent layers are connected to each other and help to carried out the data from one layer to another layer. These neurons learn how to convert inputs into corresponding output and provide the details of object which make easier to detect an object.

1.6 Framework of Object Detection

Our model comprises of three stages on a wider dimension: model construction, training and testing respectively. First of all, our CNN consist of input layer of size 224 × 224 × 3. The input image has a fix size of 224 × 224 with 3 number of channels. The size of other convolution layers is 224 × 224 × 64, 112 × 112 × 128, 56 × 56 × 256, 28 × 28 × 512, 14 × 14 × 512 and 7 × 7 × 512 respectively. Also, the model contains max pooling layers and fully connected layers with rectified linear unit layers. Last fully connected dense layer is connected with SoftMax layer which consist of 50 neurons corresponding to the 50 number of classes.

A CNN model is taught in the training stage using 5000 pictures, 224×224×3 dimensions corresponding to 50 classes of objects with 100 pictures per class. Pre-Processing of the dataset is performed according to the TensorFlow library which is used on the backend. In the training phase transfer learning approach is used for the extraction of features by freezing the layers of lower part of model which is also known as convolution base. Then fine tune the model by unfreeze lower part and replace the upper head by our classifier. By capturing the images on our webcam, this trained model is now being tested. Output of the model is in the form of performance evaluation and object is detected.

Figure 1.8 shows the methodology and framework of our proposed object detection. Our object detection model is basically consisting of five phases:

- CNN Implementation
- Training
- Pre-Processing Phase
- Testing
- Output Phase

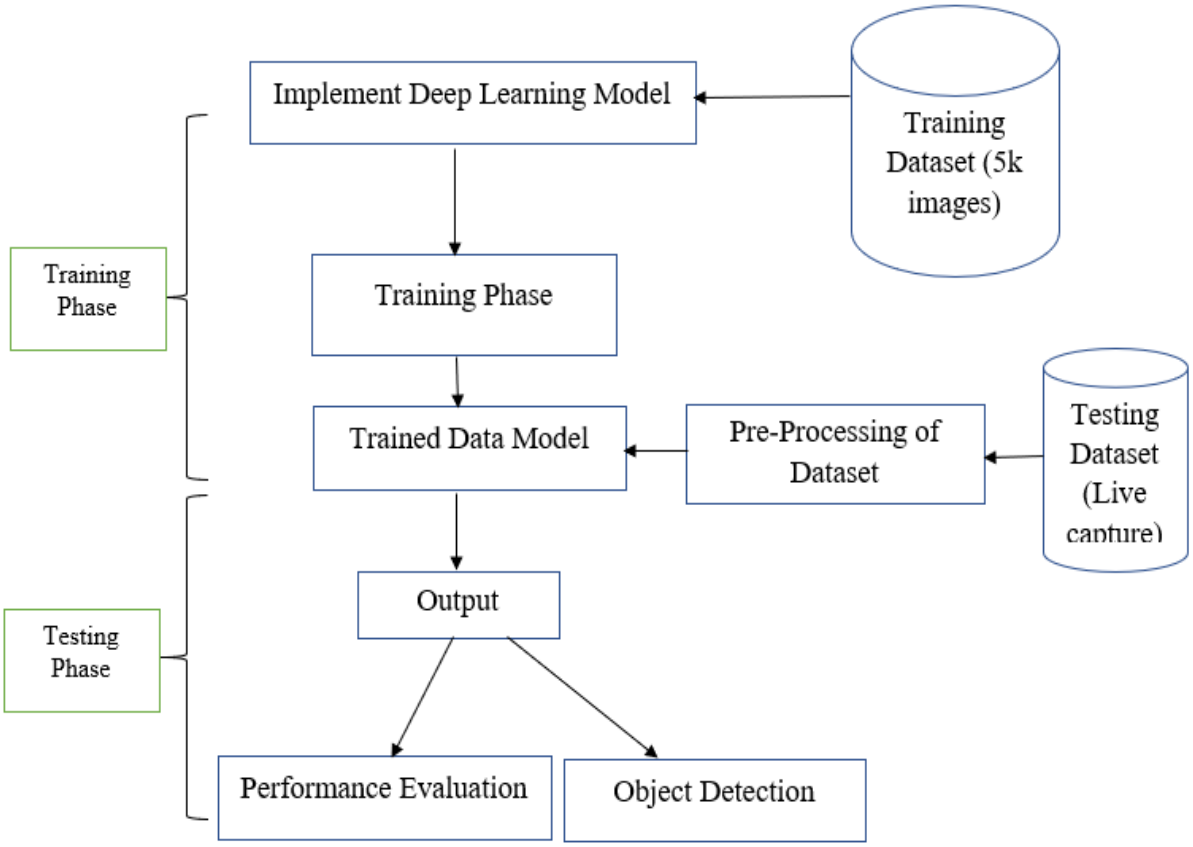

Figure 1.8: Framework of Object Detection

CHAPTER 2
LITERATURE SURVEY

This chapter illustrates the state-of-art of the current and interrelated work of object detection system. This highlights the survey, mechanism, working, benefits and limitations of the related work in the field of real-time object detection using their feature analysis from the surveillance images or videos.

Work Related with Object Detection

Bhattacharya *et al*. (2011) arranged moving article recognition and following in forward looking infra-red flying symbolism. The issues of mechanized assessment and examination for infra-red visual information are accomplished with flying stages. The issues accomplished hugeness with lightweight and solid imaging gadgets. The distinguishing proof and article following expanded the consideration in PC vision. The assignments are provided the aerial sequences of infra-red modality. A method is used for identifying the moving objects from ego-motion compensated input sequence.

Salti *et al*. (2012) structured a versatile appearance demonstrating for video following a study and assessment. Another estimation technique is planned that permits prompt examination of following exactness and following achievement without the prerequisite of use subordinate limits settings. Another estimation strategy is intended for the assessment of trackers that execute appearance model adjustment.

Bhajibhakare *et al.* (2012) distinguish and track moving article for reconnaissance framework. The planned procedure in home and business reconnaissance framework is utilized to distinguish and follow the moving articles. It is basic for the video observation framework to distinguish and follow the moving item unequivocally against unsettling influences flying creatures, trees and ecological varieties like many climate conditions. The shading foundation displaying with affectability parameter disposes of the clamours, recognize and track the moving items. Haar like component extraction strategy is utilized in article acknowledgment. Mass naming is used for bunching of moving articles.

An all-encompassing MCMC procedure is structured by **Sakaino (2013)** for following and expanded HMM technique for getting the hang of/perceiving many moving articles in recordings with jittering airs. A GUI with improved convenience is planned. MCMC and HMM based systems are recognized to encounter the disabilities, following an acknowledgment exactness and higher calculation costs. A cost decrease strategy is gotten ready for MCMC approach through taking moves when many moving items participate.

Shen and Jhang *et al.* (2013) taken the overview of appearance models in visual article following. The issues of appearance demonstrating are separated into two preparing stages, in particular visual portrayal and measurable displaying. Numerous 2D appearance models are ordered with respect to the organization modules. The key target of the model is four-overlap. The visual portrayals predictable with their element development instruments are tended to. The measurable displaying strategies for following by-location are assessed reliable with model-development techniques like generative, discriminative and half and half generative-discriminative. At that point, the demonstrating techniques are investigated.

The three parameter-autonomous measures are structured by **Nawaz *et al.* (2014)** for figuring the multitarget video following. The measures with objective size varieties join precision and cardinality mistakes that figure the long haul following exactness at numerous precision levels and register the ID varieties contrasted and term of track where they happens.

Deori and Thounaojam (2014) arranged an overview on moving item following in video. The new following strategies are ordered into numerous classes and perceive the following procedures. The following advances use numerous approaches like Mean-move, Kalman filter, Particle channel. The aftereffects of following strategies change dependent on foundation data.

Jadhav *et al.* (2014) clarified moving item location and following for video reconnaissance. Static camera is utilized for video and edge of video is taken as Reference Background Frame. This casing is taken out from the current casing to recognize the moving article and allot limit T esteem. At the point when pixel distinction is higher than

the limit esteem T, it limits the pixels from moving item or foundation pixels. The dynamic improvement limit strategy is wanted to accomplish whole moving items.

Sukanyathara *et al*. (2014) arranged a streamlined structure for recognition and following of video objects. A structure is wanted to perceive the moving items and track paying little mind to impediment and from the earlier learning of articles in scene. The division step utilizes the limit choice calculation with multi-foundation model. The video item following tracks numerous articles alongside directions dependent on Continuous Energy Minimization. A productive technique is planned as multi-target following minimization of nonstop vitality and multi foundation enrolment.

A decentralized helpful system called Pulse Counting is arranged by **Wenzhong *et al*. (2015)** for DTN restriction and probabilistic following technique called Protracting. Heartbeat Counting computes the client strolling steps and development directions through accelerometer and electronic compass in PDAs. It likewise ascertains the client area through gathering the strolling fragments and expands the estimation exactness through the experiences of portable hubs. The area estimation includes the variety of direction relying upon the reference focuses and common refinement of area estimation for experience hubs relying upon higher plausibility.

Wang *et al*. (2015) structured a system to examine the progressive qualities for visual article following. The disconnected highlights are utilized for movement designs from video arrangement. The various levelled highlights are considered with two-layer convolutional neural system. The worldly gradualness imperative in stacked auxiliary plan makes the educated highlights strong for visual item following. An objective video grouping is given adjustment module to online pre-learned highlights steady with the specific objective item.

Zhang *et al*. (2015) planned new following technique through multi-see learning structure by numerous help vector machines (SVM). The multi-see SVMs following technique is structured relying upon numerous perspectives on highlights and plans. For perceiving the exhibit, three kinds of highlights are picked, specifically dim scale esteem, histogram of situated angles (HOG), and neighbourhood parallel example (LBP) for reasonable SVMs.

The highlights indicate the article from perspectives of clarification and acknowledgment. For blend of SVMs in multi-see learning structure, another communitarian plan is planned with entropy standard by certainty circulation of applicant tests.

A Tracking-Learning-Data system is planned by **Ding *et al*. (2016)** to move nonexclusive item tracker to obscure invariant article tracker without the de-obscuring picture successions. A huge arrangement of unlabelled pictures is utilized to discover article's visual earlier information that reassigned to appearance model of specific objective. In item following procedure, internet preparing tests are assembled from following exhibitions and the setting data. Obscure pieces with preparing tests improve the power of model to obscure in molecule channel system.

Wojke *et al.* (2017) plan a basic on the web and constant following a profound affiliation metric. In this paper, we coordinate appearance data to improve the exhibition of SORT. Because of this augmentation we can track questions through longer times of impediments, viably decreasing the quantity of personality switches. In soul of the first structure we place a significant part of the computational intricacy into a disconnected pre-preparing stage where we gain proficiency with a profound affiliation metric on an enormous scale individual re-recognizable proof dataset. During on the web application, we build up estimation to-follow affiliations utilizing closest neighbour inquiries in visual appearance space. Exploratory assessment demonstrates that our expansions lessen the quantity of personality switches by 45%, accomplishing generally speaking focused execution at high edge rates

A spatially directed intermittent convolutional neural system for visual item following is structured by **Ning *et al.* (2017**) In this paper, they build up another methodology of spatially directed intermittent convolutional neural systems for visual item following. Our repetitive convolutional system misuses the historical backdrop of areas just as the particular visual highlights learned by the profound neural systems. Enlivened by late jumping box relapse strategies for article discovery, we consider the relapse capacity of Long Short-Term Memory (LSTM) in the transient space, and propose to link abnormal state visual highlights delivered by convolutional systems with district data. Rather than existing profound learning-based trackers that utilization parallel arrangement for district

up-and-comers, they use relapse for direct expectation of the following areas both at the convolutional layer and at the repetitive unit. Our test results on testing benchmark video following datasets demonstrate that our tracker is focused with best in class draws near while keeping up low computational expense.

Chandan *et al*. (2018) had led an examination on constant article recognition and following utilizing profound learning and OpenCV. Things are recognized using SSD estimation logically circumstances. Likewise, SSD has shown results with broad sureness level. Basic target of SSD count is to perceive various things constantly video gathering and track them dynamically. This model demonstrated the exceptional area and following results on the article arranged and can furthermore be utilized in express circumstances to recognize, track and respond to the particular concentrated on things in the video surveillance. This steady examination of the earth can yield mind blowing results by enabling security, solicitation and utility for any endeavour. Further extending the proposed work to recognize ammunition and weapons, to trigger alarm, if there ought to be an event of dread monger ambushes. The model can be sent in CCTVs, drifts and other surveillance contraptions to recognize strikes on various spots like schools, government work environments and facilities where arms are completely constrained.

Work Related with Transfer Learning

The concept of deep transfer learning was used by **George *et al*. (2017)** for classification and unsupervised clustering of LIGO data. They show that our profound transfer learning technique allows the ideal use of very profound convolution neural networks for glitch classification due to tiny and unbalanced training information sets, considerably decreases training time and achieves state-of - the-art precision above 98.8%, reducing the past error rate by more than 60%. More importantly, once trained on the known classes through transfer learning, they show that neural networks can be truncated and used as feature extractors for unsupervised clustering to automatically group new unknown classes of glitches and anomalous signals together. This new capacity is of paramount significance in identifying and removing new kinds of glitches that will happen as the LIGO / Virgo detectors gradually become sensitive to design.

Nie *et al*. (2017) proposed a model for ship detection using transfer learned single shot multi box detector. In this research, authors introduced a transfer learned Single Shot Multi Box Detector (SSD) for ship detection. To this end of development. To detect ships with restricted satellite images, a state-of - the-art object detection model pre-trained from a big number of natural pictures was taught. This could be one of the first studies to introduce SSD into ship detection on satellite images to the best of our knowledge. Experiments showed that using NVIDIA TITAN X, the method could achieve 87.9% AP at 47 FPS. Compared to Faster R-CNN, an increase in the AP could be accomplished by 6.7 percent. It was observed that the precision of detection decreased significantly due to the declining resolution induced primarily by the missing tiny ships.

Transfer learning by ranking for weakly supervised object annotation was designed by **Shi *et al*. (2017).** Object annotation based on techniques of transfer learning and detection is a difficult issue because each picture can contain many places of candidate objects that partly overlap the object of concern. Existing methods concentrate on how best to use binary labels for annotation of object place. Authors proposed a model in this study to fix this issue from a very distinct view by casting it as a issue of transfer learning. In particular, they formulate a novel transfer learning based on learning to rank, which efficiently transfers a model for automatic object location annotation from an auxiliary dataset to a target dataset with totally unrelated classifications of objects. They show that our approach outperforms existing, weakly supervised, state-of - the-art approach to object annotation in the challenging VOC dataset.

Transfer learning based End-to-End Airplane Detection in Remote Sensing Images was developed by **Zhong *et al*. (2018)**. They used the idea of object detection technology in this research work based on transfer learning strategy in remote sensing pictures and concentrate on aircraft detection job. Besides using some features of remote sensing images, some new techniques for data augmentation have been proposed. They also use deep learning to implement end-to-end trainable aircraft detection and adopt a single deep convolution neural network and limited training samples. Classification and positioning are no longer divided into multi-stage tasks; end-to-end detection attempts to combine them for optimisation, ensuring an optimal final stage solution. They used remote sensing images

of airports collected from Google Earth in experimental results and the results show that the algorithm designed for remote sensing object detection is highly accurate and meaningful and could be used in different types of object detection system.

Kolar *et al*. (2018) designed a model for safety guardrail detection in 2D images using the concept of Transfer learning and deep convolutional neural networks. A safety guardrail detection model based on convolutional neural network (CNN) was developed by the authors in this research work. An augmented data set is generated and used as a training set by adding a background image to guardrail 3D models. Transfer learning is used and the 16-layer architecture of the Visual Geometry Group (VGG-16) is adopted to build the basic features extraction for the neural network. In the implementation of the CNN, 4000 augmented images were used to train the proposed model, while another 2000 images were used to validate the proposed model, collected from real construction jobs and 2000 images from Google. The proposed CNN-based guardrail detection model obtained a high accuracy of 96.5%. Furthermore, this study shows that the synthetic pictures produced by augmentation technology can be used to produce a big training dataset, and the CNN-based picture detection algorithm is a promising approach to security surveillance for building jobs.

Deep neural networks and transfer learning applied to multimedia web mining by **Sánchez *et al*. (2018).** They suggested a deep learning model in this study to fix the internet categorization issue. Authors use a method known as transfer or inductive learning to drastically decrease the training phase's computational costs. Finally, they report experimental outcomes using distinct classification techniques and features from distinct depths of the profound model on the efficacy of the suggested technique.

Zhang *et al*. (2019) discussed an improved technique, weakly supervised detection which used for new categories object refer as mixed supervised detection. This work different from exist MSD scheme in which per-trained object are directly transfer from existing to new categories. In this work proposed a robust method for object transfer. Proposed MSD scheme is implemented on both intra-dataset detection and cross-dataset detection task. Two matrices are used for evaluation name as mAP and CorLoc. For testing mean average precision is used and for training correct localization is used in this proposed work. This

work achieves excellent result from existing technique. Achieve well result on the challenging ILSVRC2013 detection dataset and the PASCAL VOC datasets.

CHAPTER 3

PROBLEM STATEMENT

3.1 Research Gap Analysis

- One of the emerging fields for improving the precision of detection technologies is also the pre-processing of pictures. Relatively small number of studies using deep learning techniques had been proposed for pre-processing or enhancement of image quality.
- In existing work, clarification changes along with handling of dynamic background is a widely not covered and it is still a research challenge in the real time capture images.
- Another direction for effective detection of objects is to develop methods that can handle background clutter, occlusions, dynamic background or noisy images etc which is not appropriately covered in exiting researches. There had been some attempts with advent of deep learning based convolutional neural networks but it is still an open issue for researchers.
- In the existing popular ROI selection algorithm based on segmentation approaches is taken randomly using the blind segmentation algorithm. This will increase time to get desired solution and the detection and tacking accuracy affected.
- Currently individual objects in image are detected but in single frame, multiple objects are appearing then designed model not able to detect a particular object due to use of existing binary classifiers.

3.2 Motivation for Research Work

The purpose of image content analysis is to find consequential structures and samples from real time data. The assignment of object detection is to overpass the gap among the numerical pixel level data and a high-level abstract activity account. Intelligent ways to deal with object detection and following is well examined in the writing. More upgraded adaptations of object detection and tracking will demonstrate to be time productive and

activity proficient. In this research we endeavour to propose answers for image reconnaissance applications utilizing still cameras where human detection is utilized as an assistive innovation for the framework administrator. The stationary and real time moving object detection following can be principally made in view of foundation estimation. The camera vibration and adjustment are a real subject in detection and following of the moving object from any image succession. To try the above issues, we have accepted a skilful object following framework with the assistance of streamlining strategies. From these types of challenging task, we present a transfer learning-based CNN based real-time object detection in surveillance stationary and real time image. In simple words, this research makes the following contributions.

- To detect objects from stationary and real time capture images, CNN technique is used with segmented stationary and real time images.
- We designed a novel transfer learning approach for segmented ROI of images to optimize object features.
- For the validation of proposed object detection module, we evaluate performance parameters of proposed work and compare with state-of-the-art approaches.

3.3 Problem Description

Classification and detection of the objects from the stationary and real time images is the key to provide better secure, safe and automatic monitoring in control system. A lot of work has already been done in the field of object detection for surveillances stationary and real time data but the object detection accuracy is not satisfactory due to the proper segmentation of events data. In general, surveillances stationary and real time capture mages sequences contain time interval based spatial events which exist over a sequence of continuous frames. From the study and literature survey, it has concluded that there is a still hope that can develop a well-organized new technique to detect objects to improve the accuracy rate. The object detection accuracy is dependent on the segmentation technique; if the segmented region is proper as per the normal and abnormal events then the detection accuracy will be high. So, the detection accuracy of anomaly system is directly proportional to the segmented region and segmented region is dependent on the region values. In the proposed work, automated object detection from surveillances stationary and real time

images using the segmentation technique with CNN (Convolution Neural Network) and Transfer Learning has been presented for the surveillances stationary and real time image to the improvement in detection results based on their ranking. At the last of simulation, to verify the effectiveness of proposed work the performance parameters will have to calculate in terms of error rate and accuracy rate.

3.4 Objectives of Thesis

In this work, efficient object detection using transfer learning with deep convolutional neural network based on the audible results is proposed. The objectives of this work have been identified as follows

1. To study the previous existing objects detection methods for image data of surveillance stationary and real time capture images.
2. To develop a novel transfer learning approach for optimization of object feature.
3. To detect and track objects, CNN (Convolution Neural Network) technique is used.
4. To evaluate performance parameters of proposed work like error and accuracy.

CHAPTER 4

PROPOSED SOLUTION

4.1 Description of Algorithms

There are certain steps that are required to be followed in order to create a model for objects detection from frame of surveillance stationary and real time images using CNN (Convolution Neural Network) and transfer leaning technique. These steps are defined as follows:

Step 1: Upload stationary and real time capture images for simulation of model. After the design of simulation structure, we proceed to the next steps. Weights of secondary dataset are used in the approach of transfer learning.

Step 2: Apply pre-processing and feature extraction from the stationary and real time images for the simulation of proposed object detection system. Firstly, images are resized to the pixel value of 224 × 224. Function img_to_array () is used to add the 3 number of channels for RGB images.

Then pre-processing function of expand_dims () to add the number of pictures to array. Rearrange the array according to the TensorFlow requirements. The function preprocess_input () is used to tune our dataset to the array.

Step 3: Apply transfer learning on extracted feature to optimize the feature sets and create a set of unique features for object. For the best and optimal feature selection from the extracted feature set, transfer learning is used and it is a better option. We use VGG16 model as feature extractor. VGG16 model have 16 layers. There are 13 convolution layers in the VGG16 pre-trained model. Then we use these features and send them to dense layers trained in accordance with our dataset. The output layer is also substituted by our fresh layer of SoftMax appropriate to our issue. The last layer in the VGG16 model is a SoftMax activation layer with 1000 categories. We are removing this layer and replacing it with a SoftMax layer of 50 categories. In our custom VGG16 model we change the neurons of dense layer with 128. We train the weights of these layers.

Step 4: After that initialize CNN (Convolution Neural Network) and to train the system using following steps:

1. Optimized feature set to provide an input of CNN for training and testing data.
2. Compute the total categories which are generated by the training of data using CNN classifier to classify the normal and abnormal behaviour of events.

Step 5: Initialize CNN for classification purpose using two phases, namely, training and testing. After the training of system, we save the trained structure which is use in the classification section to classify the object from surveillance stationary and real time images. In the testing phase, the test image is uploaded and repeats the all steps. In the classification section, test image's extracted feature is matched with trained CNN structure and return results with object boundary region. CNN has inbuilt feature extraction method which is act as an input to CNN and the used CNN algorithm is written in above section.

Step 6: After that in the classification section, classification of events occurs according to the trained CNN structure.

Step 7: At last of the module, the performance parameters of object detection system like Error rate and Accuracy rate will be calculated.

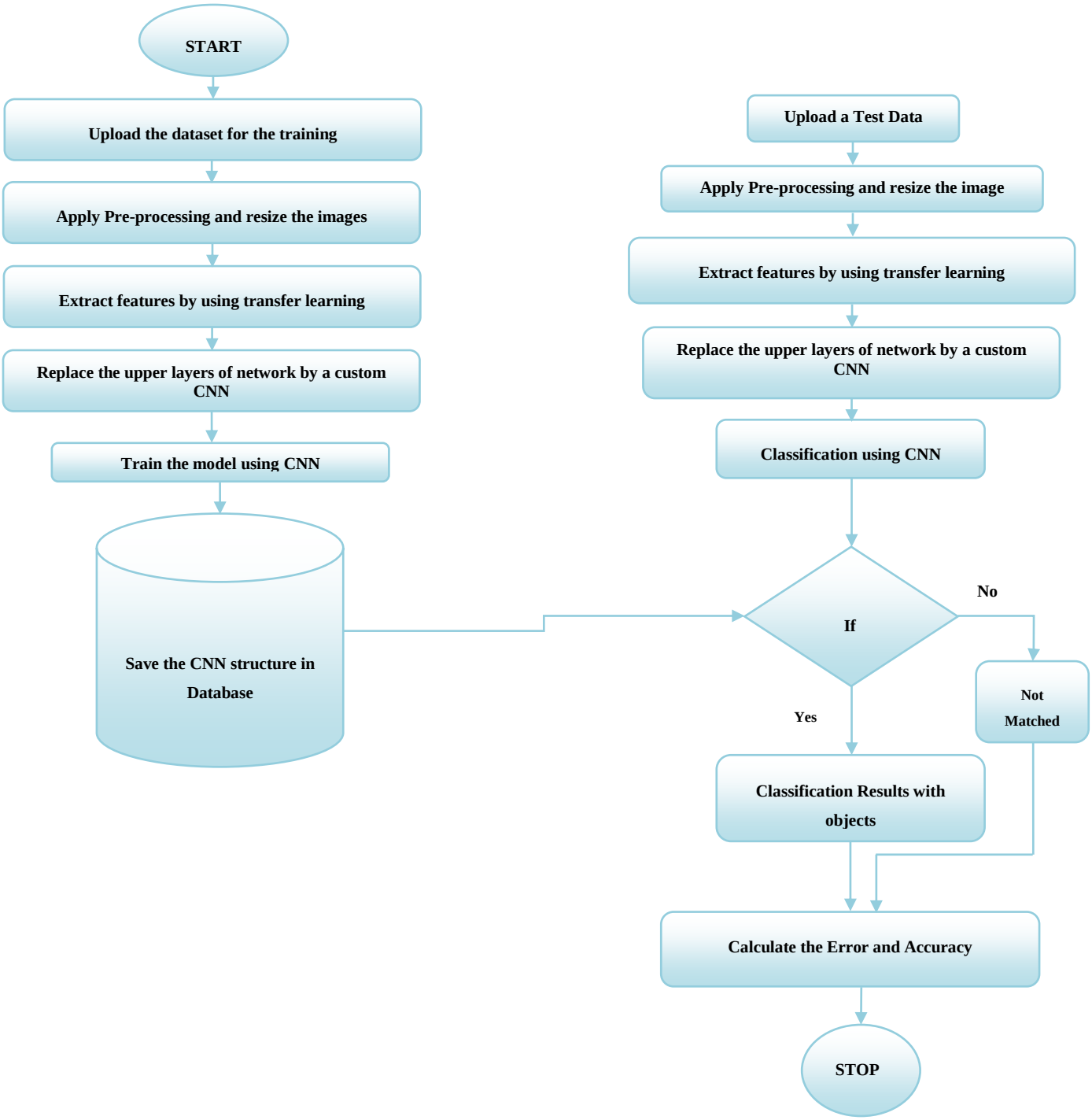

Figure 4.1: Flowchart of The Proposed Work

CHAPTER 5

DATA COLLECTION AND PRE-PROCESSING

5.1 Data Collection

Dataset 1

For the simulation of proposed object detection model, we collect images or surveillance videos frames manually. Used dataset is real-time dataset and each image is free from any type of compression or noise. We collect total 50 indoor classes of object data from the simulation purposes in which 80% data is sued to train the system and rest of the data is used for the cross validation of proposed model. In each class to 100 samples, so the total size of dataset is 5000 with image dimension of 4032 × 3024 and the modulation of images is 72 × 72 for horizontal as well as vertical. All images are captured by smart mobile phone of Apple with model iPhone-X. Some sample of collected data with results are given in the below Figure 5.1.

Dataset 2

This is the secondary dataset we used for our proposed work. We have used the weights of ImageNet dataset in the transfer learning approach. ImageNet dataset contains 1000 number of classes. The ImageNet models are trained on 1.28 million training images and evaluated on the testing images of 50 thousand.

Image Description	Samples of Dataset			
Mug				
Pen				
Backpack				
Watch				
Hammer				
Dustbin				

Screwdriver				
Sunglasses				
Remote				
Keypad mobile				
Iron				
Keyboard				
Mouse				

Screw				
Modem				
Sport shoes				
T-shirt				
Book				
Water bottle				
Fan				

Ruler				
Hairdryer				
Sharpener				
Eraser				
Carton box				
Safety pin				
Wooden chair				

Bucket				
School kit				
Belt buckle				
Helmet				
Switch				
Woolensweater				
Jeans				

Towel				
Curtain				
Desktop screen				
Handkerchief				
Candle				
File folder				
Perfume				

Pillow				
Wallet				
Nail				
Paintbrush				
Sweatshirt				
Washbasin				
Sofa				

Weighing scale				
Tie				

Figure 5.1: Samples of Dataset 1

5.2 Pre-processing of Dataset 1 Images

Apply pre-processing on the uploaded test images to increase the efficiency of proposed object detection system. So, in this work, firstly we apply resize technique to convert image from pixels 4032 × 3024 into 224 × 224. Then we add number of channels by img_to_array () function i.e., (224, 224, 3) for RGB images. Add the number of images to array by expand_dims () function. i.e., (1, 224, 224,3). To adequate our image to the format our model requires the preprocess_input function is used. The mean RGB channels of previous dataset are subtracted. Subtract the calculated mean per channel for all the images.

CHAPTER 6
IMPLEMENTATION AND EXPERIMENTAL RESULTS

6.1 Results and Discussion

In this section, the results attained after as simulating the code in python enviornment has been defined. In the proposed research work, real-time multi view object detection using deep convolutional neural network with audible results based on some pre-processing phases is designed. Objects are detected from the distance of around 25 cm in the real time capture images. Here simulation result is divided in two parts 1st is results based on the stationry images from dataset and second is based on the real time capture images.

6.2 Results for Stationry Dataset 1 Images

In this case proposed model is simulated based on the images of dataset for the object detection. The simulation results with dataset are given below for some samples.

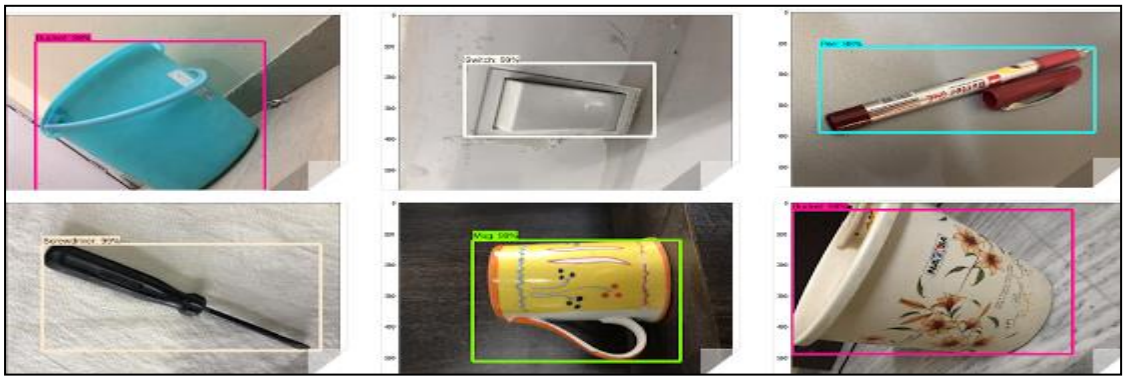

Figure 6.1: Results with Dataset 1 Images

6.3 Results for Real Time Capture Images

In this case proposed model is simulated based on the images of real time scenario for the object detection. The simulation results with real time images are given below for some samples in Figure 6.2.

Figure 6.2: Results with Real Time Capture Images

Above results represent the simulation results of proposed model where CNN is used to train the model. After feature extraction, we are applying the CNN as a classification algorithm, which detect the object by comparing the test image with the features stored into the database. In this section, the simulation result of proposed object detection is discussed and the efficiency of proposed work is compared with existing work (Chandan *et al*.). The training and testing of the proposed mechanism is evaluated by manually create dataset. By adapting the established proposed algorithms, below outcomes are computed with quality based parameters, such as Error and Accuracy. A comparison is drawn with the existing work (Chandan *et al*.) to shown the effectiveness of the proposed work with respect to the different object based on the three sample data from each categories and for the graphical representation we calculate their average value like Average Error Rate, and Average Accuracy of system.

Table 6.1: Test Results of Proposed Method

Dataset Types	Test Sample	Accuracy
Stationary	Test Sample 1	99.82
	Test Sample 2	99.62
	Test Sample 3	99.91
Real-Time	Test Sample 1	99.86
	Test Sample 2	99.74
	Test Sample 3	99.92

Table 6.2: Test Results of Existing Method (Chandan *et al.*)

Dataset Types	Test Sample	AUC
Stationary	Test Sample 1	95.09
	Test Sample 2	98.01
	Test Sample 3	99.99
Real-Time	Test Sample 1	93.10
	Test Sample 2	98.68
	Test Sample 3	97.77

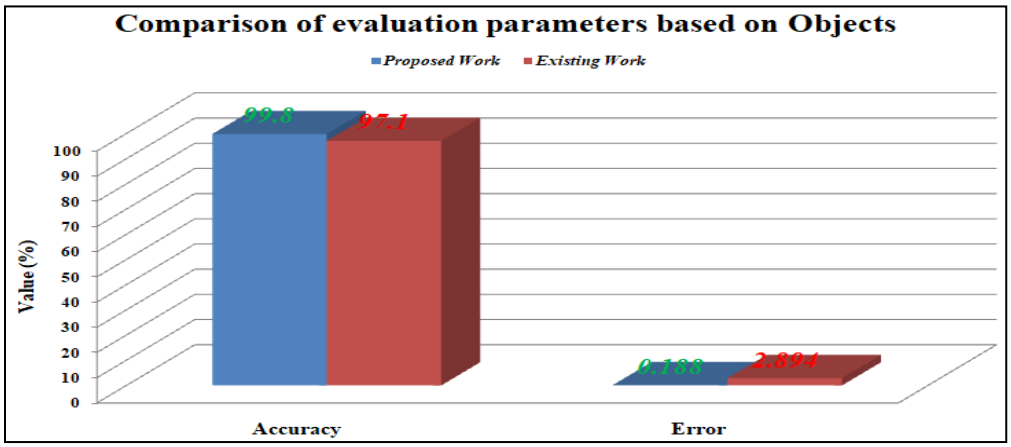

Figure 6.3: Comparison of Evaluation Parameters Based on Objects

The comparison of evaluation parameters for proposed and existing work is depicted in Figure 6.3. For the proposed work on the basis of normal object data, average error rate is 0.188% and average accuracy (AUC) 99.8%. By using the concept of transfer learning with CNN, the accuracy of proposed work is improved because the average error rate is 2.894% and average accuracy (AUC) 97.1% for the existing work. The comparison of proposed work with some other existing work, which is considered in survey of proposed work, is described in below table.

Table 6.3: Comparison of Accuracy of Proposed Work with Existing Works

Authors	**Accuracy (%)**
Chandan	97.01
Wojke	93.1
Ning	91.3
Proposed work	99.8

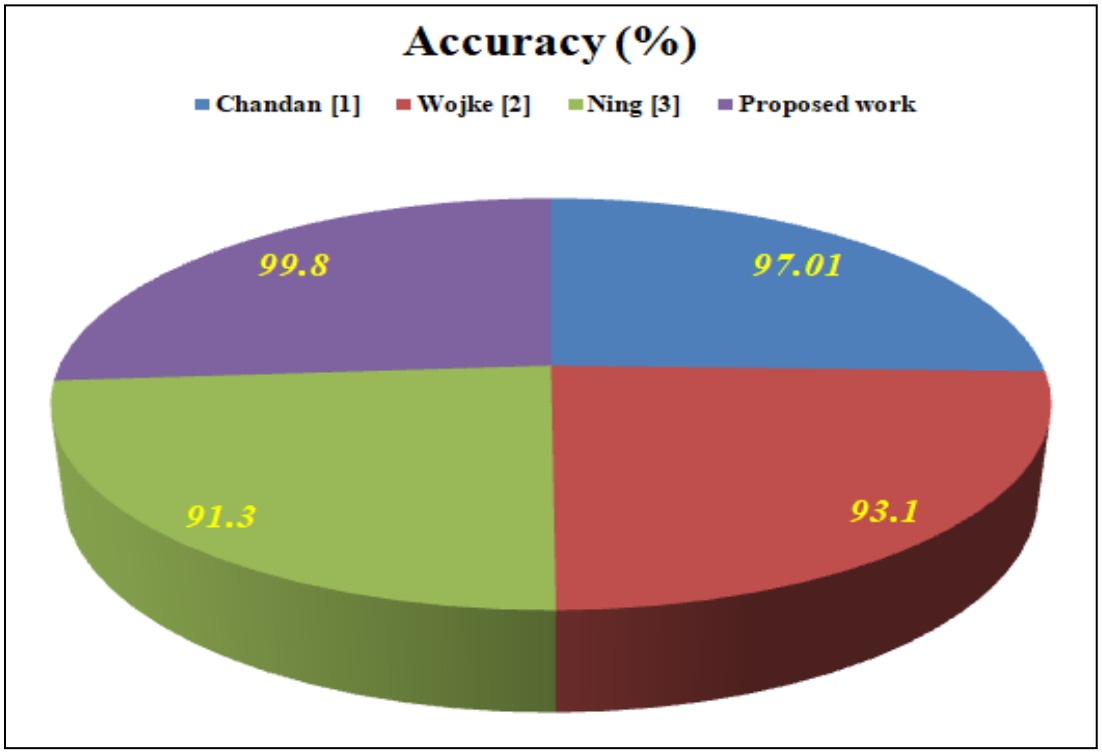

Figure 6.4: Comparison of Accuracy of Proposed Work with Existing Works

Figure 6.4 represents the comparative analysis of existing work based on the classification accuracy. From the figure we observe that the accuracy achieve by proposed work is better than other author by using the hybridization of transfer learning with CNN technique.

Table 6.4: Performance Parameters of Proposed Work

Number of Test Data	Precision	Recall	F-measure	E-Time	Accuracy
1	0.955	0.4953	0.954	0.057	99.62
2	0.958	0.3959	0.959	0.041	99.19
3	0.960	0.962	0.961	0.038	99.34
4	0.957	0.7960	0.958	0.036	99.45
5	0.957	0.8955	0.956	0.04	98.90
6	0.958	0.9608	0.959	0.037	99.81
7	0.963	0.958	0.959	0.025	99.77
8	0.943	0.8737	0.907	0.083	98.28
9	0.937	0.2934	0.446	0.074	99.73
10	0.983	0.194	0.324	0.089	97.37
Average	**0.957**	**0.682**	**0.838**	**0.052**	**99.14**

Figure 6.5: Precision, Recall and F-measure (PRF)

The above Figure 6.5 represents the comparison graph obtained for (i) Precision (ii) Recall and (iii) F-measure for proposed efficient object detection using transfer learning. X-axis shows the number of test sample of stationary as well as webcam images that are uploaded for object detection purpose. Y-axis depicts the value of different computation parameters and from the above graph it is clear that the code is simulated ten times and the average value of precision, recall and F-measure are 0.957, 0.682 and 0.838 respectively.

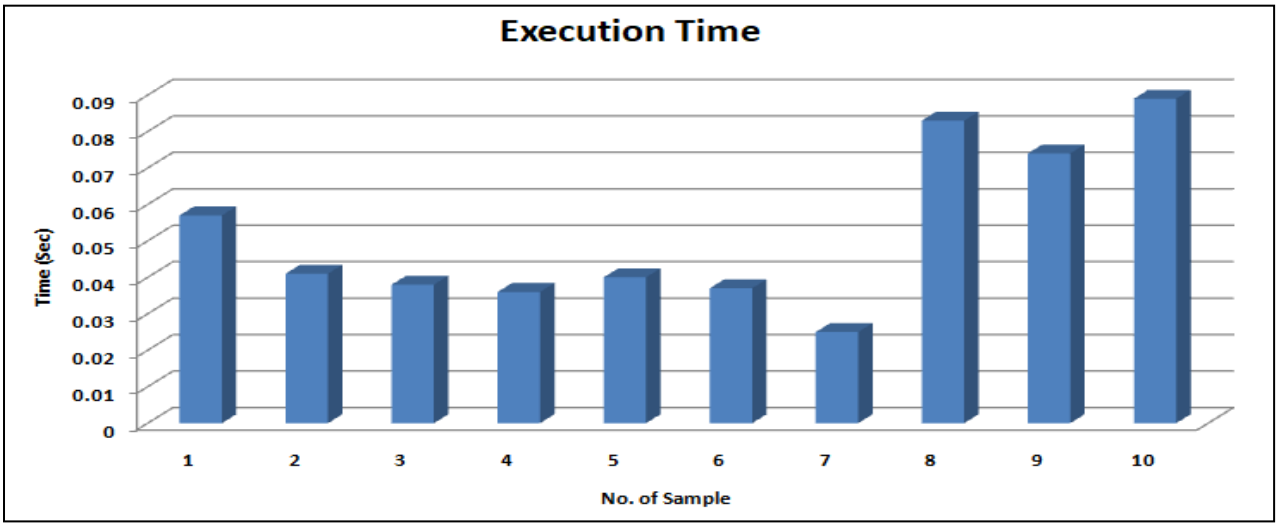

Figure 6.6: Execution Time

The execution time which is required to detect the sentiments types for 10 iterations is shown in figure above. X-axis and Y-axis signifies the number of test samples and

execution time respectively. From the graph shown in Figure 6.6, it is understandable that average execution time required for the proposed work is 0.052 seconds.

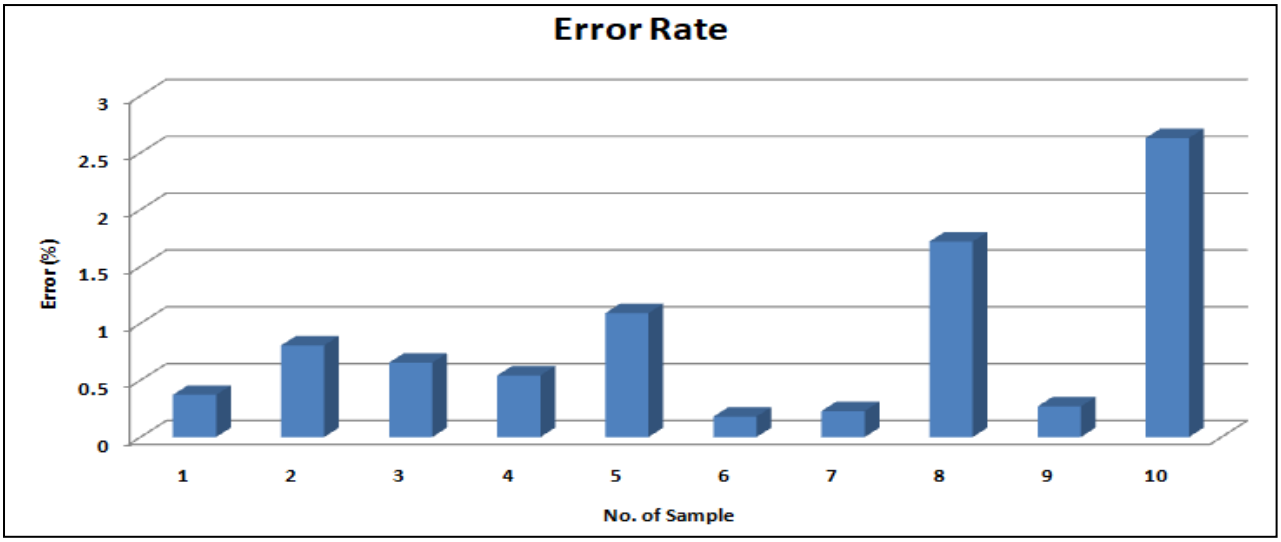

Figure 6.7: Error Rate

The error obtained while detecting the leaf disease is shown in Figure 6.7. The average error measured for 10 test text data is 0.86 %.

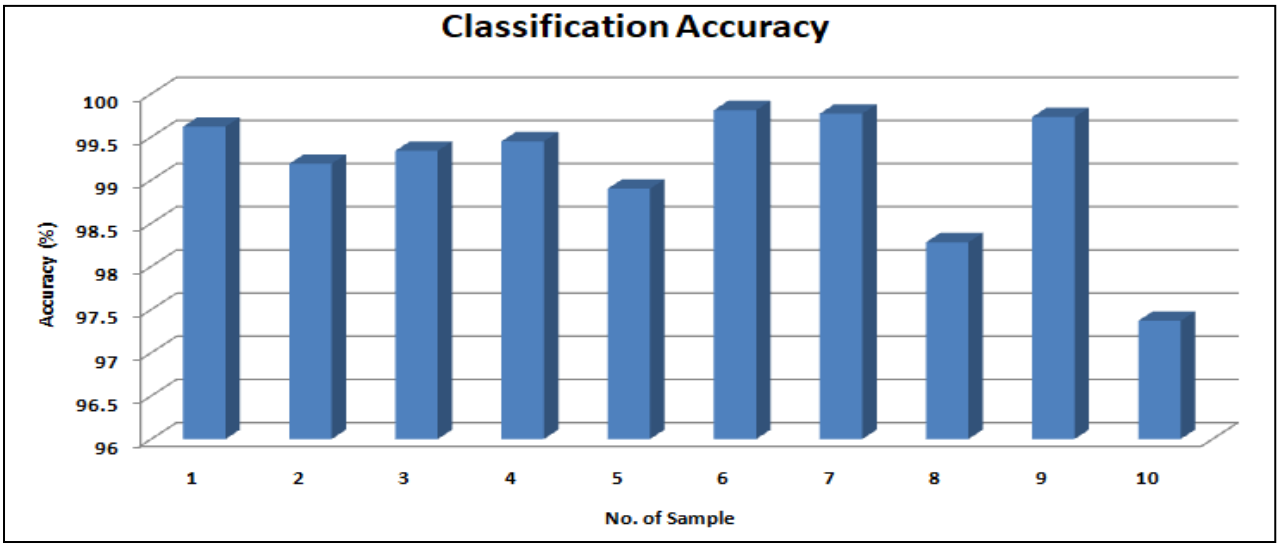

Figure 6.8: Classification Accuracy of Proposed Work

The above Figure 6.8 shows the value of classification accuracy obtained for ten test text data from different sentiments. The average value of accuracy obtained for the proposed work is 99.14 % which represents that the text sentiments detected with high accuracy. Some more simulation results with 50 different images is given in the below table.

Table 6.5: Simulation Results for Dataset 1

IMAGE NAME	PRECISION	RECALL	F- MEASURE
Mug	0.95	1.00	0.97
Pen	1.00	1.00	1.00
Backpack	1.00	1.00	1.00
Watch	1.00	0.93	0.97
Hammer	0.93	0.96	0.95
Dustbin	0.88	0.78	0.82
Screwdriver	1.00	0.96	0.98
Sunglasses	1.00	0.90	0.95
Remote	0.87	0.87	0.87
Keypadmobile	0.93	1.00	0.97
Iron	1.00	1.00	1.00
Keyboard	1.00	0.92	0.96
Mouse	1.00	0.80	0.89
Screwdriver	1.00	0.80	0.89
Modem	0.91	0.87	0.89
Sportshoes	1.00	0.94	0.97
Tshirt	1.00	1.00	1.00
Book	0.95	0.91	0.93
Waterbottle	0.90	0.95	0.92
Fan	1.00	0.86	0.93
Ruler	0.91	0.95	0.93
Hairdryer	0.95	0.87	0.91
Sharpner	1.00	0.91	0.95
Eraser	0.89	0.94	0.91
Cartonbox	0.90	0.95	0.93
Safetypin	0.94	0.94	0.94
Woodenchair	0.89	1.00	0.94
Bucket	1.00	0.77	0.87
Schoolkit	1.00	1.00	1.00
Beltbuckle	0.95	1.00	0.97
Helmet	0.86	1.00	0.93
Switch	1.00	0.94	0.97
Woolensweater	0.87	1.00	0.93

Jeans	1.00	1.00	1.00
Towel	1.00	0.88	0.94
Curtain	0.95	1.00	0.98
Desktopscreen	0.87	1.00	0.93
Handkerchief	1.00	0.79	0.88
Candle	0.90	0.90	0.90
Filefolder	0.93	0.88	0.90
Perfume	1.00	0.94	0.97
Pillow	0.86	0.96	0.91
Wallet	1.00	1.00	1.00
Nail	0.78	0.96	0.86
Paintbrush	0.89	1.00	0.94
Sweatshirt	1.00	1.00	1.00
Washbasin	0.96	1.00	0.98
Sofa	0.86	0.96	0.91
Weighingscale	0.71	0.85	0.77
Tie	1.00	0.95	0.98

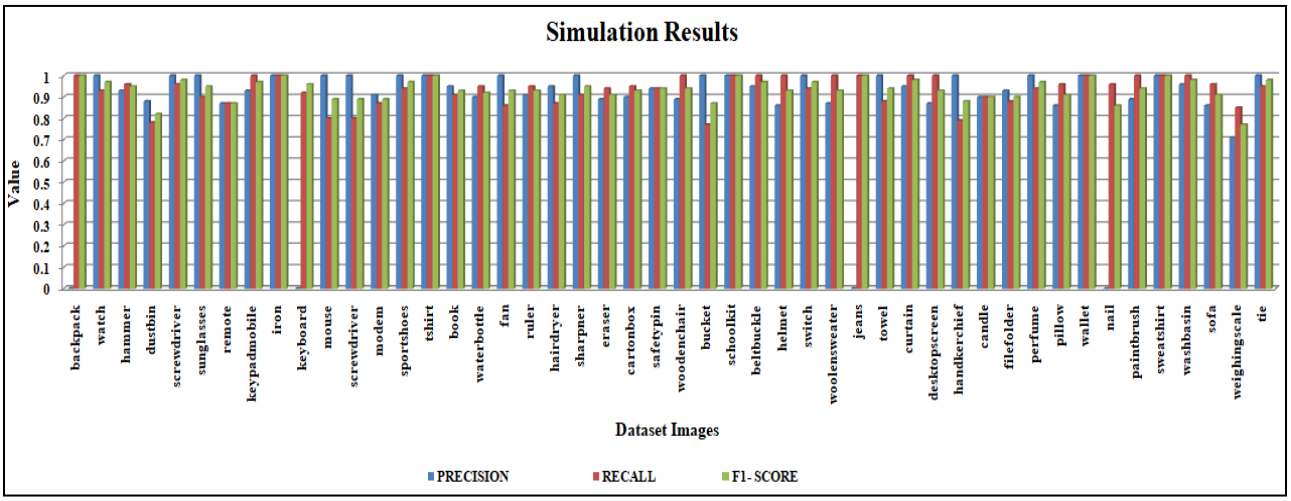

Figure 6.9: Precision, Recall and F-measure (PRF)

From the Figure 6.9, it is clear the simulation results for proposed efficient object detection using transfer learning is better for stationary and real time capture images.

CHAPTER 7

CONCLUSION & FUTURE SCOPE.

7.1 Conclusion

In this work, transfer learning based CNN is used for real time object detection in images and surveillance videos is proposed. It provides a detailed view of the different applications and potential challenges of object detection. The proposed transfer learning algorithm is experimented with several image or surveillances video frames for detecting and locating objects in the image/video according to the frames. So, new concept based on the transfer learning can be used to improve the correctness of proposed algorithms with the CNN. From the result analysis of proposed and existing work, it seems to be the system accuracy is more than 99.8% to detect and track the objects from the images or surveillance video. To validate the proposed object detection work, a new dataset 1 consisting of a variety of real-world images are introduced which is known as Image Dataset. The results of the experiment on this dataset indicate that our proposed object detection approach performs significantly better than baseline methods and the accuracy is improved by 2.7% as compare to the various existing work.

7.2 Future Scope

In the future work, hybridization artificial intelligence techniques can be used to enhance the performance of object detection from images or surveillances video. Feature extraction algorithms can also be used to classify the objects with more accurately so that the efficiency of object detection system can be improved. In addition, the framework can be extended with the objective of object detection, traffic information analysis, traffic prediction, route suggestion, smart park management, etc.

REFERENCES

[1] Alsaqre, Falah E., and Yuan Baozong. "Moving object segmentation from video sequences: an edge approach." In *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE Cat. No. 03EX667)*, vol. 1, pp. 193-199. IEEE, 2003.

[2] Chandan, G., Ayush Jain, and Harsh Jain. "Real time object detection and tracking using Deep Learning and OpenCV." In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1305-1308. IEEE, 2018.

[3] Chen, Fan, Christophe De Vleeschouwer, and Andrea Cavallaro. "Resource allocation for personalized video summarization." *IEEE Transactions on Multimedia* 16, no. 2 (2013): 455-469.

[4] Chuang, Chi-Han, Shyi-Chyi Cheng, Chin-Chun Chang, and Yi-Ping Phoebe Chen. "Model-based approach to spatial–temporal sampling of video clips for video object detection by classification." *Journal of Visual Communication and Image Representation* 25, no. 5 (2014): 1018-1030.

[5] Chunxian, Gao, Zeng Zhe, and Liu Hui. "Hybrid video stabilization for mobile vehicle detection on SURF in aerial surveillance." *Discrete Dynamics in Nature and Society* 2015 (2015).

[6] Culibrk, Dubravko, Milan Mirkovic, Vladimir Zlokolica, Maja Pokric, Vladimir Crnojevic, and Dragan Kukolj. "Salient motion features for video quality assessment." *IEEE Transactions on Image Processing* 20, no. 4 (2010): 948-958.

[7] Deori, Barga, and Dalton Meitei Thounaojam. "A survey on moving object tracking in video." *International Journal on Information Theory (IJIT)* 3, no. 3 (2014): 31-46.

[8] Ercan, Ali O., Abbas El Gamal, and Leonidas J. Guibas. "Object tracking in the presence of occlusions using multiple cameras: A sensor network

approach." *ACM Transactions on Sensor Networks (TOSN)* 9, no. 2 (2013): 16.

[9] Farin, Dirk, Peter HN de With, and W. A. Effelsberg. "Video-object segmentation using multi-sprite background subtraction." In *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, vol. 1, pp. 343-346. IEEE, 2004.

[10] Gambhir, Deepak, and Meenu Manchanda. "Adaptive threshold based segmentation for video object tracking." In *2014 IEEE International Advance Computing Conference (IACC)*, pp. 1127-1132. IEEE, 2014.

[11] Garcia-Dopico, Antonio, José Luis Pedraza, Manuel Nieto, Antonio Pérez, Santiago Rodríguez, and Luis Osendi. "Locating moving objects in car-driving sequences." *EURASIP Journal on Image and Video Processing* 2014, no. 1 (2014): 24.

[12] Giordano, Daniela, Francesca Murabito, Simone Palazzo, and Concetto Spampinato. "Superpixel-based video object segmentation using perceptual organization and location prior." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4814-4822. 2015.

[13] Izadinia, Hamid, Imran Saleemi, and Mubarak Shah. "Multimodal analysis for identification and segmentation of moving-sounding objects." *IEEE Transactions on Multimedia*15, no. 2 (2012): 378-390.

[14] Kang, Bin, Wei-Ping Zhu, and Jun Yan. "Object detection oriented video reconstruction using compressed sensing." *EURASIP Journal on Advances in Signal Processing* 2015, no. 1 (2015): 15.

[15] Kangin, Dmitry, Denis Kolev, and Garik Markarian. "Multiple video object tracking using variational inference." In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1-6. IEEE, 2015.

[16] Kermani, Elham, and Davud Asemani. "A robust adaptive algorithm of moving object detection for video surveillance." *Eurasip Journal on Image and video processing* 2014, no. 1 (2014): 27.

[17] Lemaitre, Cédric, Michal Perdoch, Abdul Rahmoune, Jiri Matas, and Johel Miteran. "Detection and matching of curvilinear structures." *Pattern recognition* 44, no. 7 (2011): 1514-1527.

[18] Ling, Chih-Hung, Chia-Wen Lin, Chih-Wen Su, Yong-Sheng Chen, and Hong-Yuan Mark Liao. "Virtual contour guided video object inpainting using posture mapping and retrieval." *IEEE Transactions on multimedia* 13, no. 2 (2010): 292-302.

[19] Liu, Dingming, and Jieyu Zhao. "Spatio-temporal video object segmentation using moving detection and graph cut methods." In *2011 Seventh International Conference on Natural Computation*, vol. 4, pp. 1859-1862. IEEE, 2011.

[20] McFee, Brian, Carolina Galleguillos, and Gert Lanckriet. "Contextual object localization with multiple kernel nearest neighbor." *IEEE Transactions on Image Processing* 20, no. 2 (2010): 570-585.

[21] Min, Alexander W., and Kang G. Shin. "Robust tracking of small-scale mobile primary user in cognitive radio networks." *IEEE Transactions on Parallel and Distributed Systems* 24, no. 4 (2012): 778-788.

[22] Nakhmani, Arie, and Allen Tannenbaum. "Self-crossing detection and location for parametric active contours." *IEEE Transactions on Image Processing* 21, no. 7 (2012): 3150-3156.

[23] Ning, Guanghan, Zhi Zhang, Chen Huang, Xiaobo Ren, Haohong Wang, Canhui Cai, and Zhihai He. "Spatially supervised recurrent convolutional neural networks for visual object tracking." In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-4. IEEE, 2017.

[24] Panagiotakis, Costas, Nikos Pelekis, Ioannis Kopanakis, Emmanuel Ramasso, and Yannis Theodoridis. "Segmentation and sampling of moving object

trajectories based on representativeness." *IEEE Transactions on Knowledge and Data Engineering* 24, no. 7 (2011): 1328-1343.

[25] Papazoglou, Anestis, and Vittorio Ferrari. "Fast object segmentation in unconstrained video." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1777-1784. 2013.

[26] Porikli, Fatih Murat. "Video Object Segmentation by Volume Growing Using Feature-Based Motion Estimator." In *proceedings of 16th International Symposium on Computer and Information Sciences (ISCIS)*. 2001.

[27] Shen, Chunhua, Sakrapee Paisitkriangkrai, and Jian Zhang. "Efficiently learning a detection cascade with sparse eigenvectors." *IEEE Transactions on Image Processing* 20, no. 1 (2010): 22-35.

[28] Varas, David, and Ferran Marques. "A region-based particle filter for generic object tracking and segmentation." In *2012 19th IEEE International Conference on Image Processing*, pp. 1333-1336. IEEE, 2012.

[29] Walha, Ahlem, Ali Wali, and Adel M. Alimi. "Moving object detection system in aerial video surveillance." In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 310-320. Springer, Cham, 2013.

[30] Walha, Ahlem, Ali Wali, and Adel M. Alimi. "Video stabilization with moving object detecting and tracking for aerial video surveillance." *Multimedia Tools and Applications* 74, no. 17 (2015): 6745-6767.

[31] Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. "Simple online and realtime tracking with a deep association metric." In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3645-3649. IEEE, 2017.

[32] Xu, Fuyuan, Guohua Gu, Kan Ren, and Weixian Qian. "Motion segmentation by new three-view constraint from a moving camera." *Mathematical Problems in Engineering* 2015 (2015).

[33] Yadav, Dileep Kumar, and Karan Singh. "A combined approach of Kullback–Leibler divergence and background subtraction for moving object detection in thermal video." *Infrared Physics & Technology* 76 (2016): 21-31.

[34] Zhang, Chenguang, and Haizhou Ai. "Video Object Segmentation by Hierarchical Localized Classification of Regions." In *The First Asian Conference on Pattern Recognition*, pp. 244-248. IEEE, 2011.

[35] Zhang, Dong, Omar Javed, and Mubarak Shah. "Video object segmentation through spatially accurate and temporally dense extraction of primary object regions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 628-635. 2013.

[36] Zhang, Fanlong, Jian Yang, Ying Tai, and Jinhui Tang. "Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling." *IEEE Transactions on Image Processing* 24, no. 6 (2015): 1956-1966.

[37] Zhang, Junge, Kaiqi Huang, and Jianguo Zhang. "Mixed supervised object detection with robust objectness transfer." *IEEE transactions on pattern analysis and machine intelligence* 41, no. 3 (2018): 639-653.