Object Detection using Deep Learning

[1]Mrs. G.PRATHIBHA PRIYADARSHINI[2]K.ARCHANA ,

[3] K.RUQSAR AHMED, [4]B.RUSHMITHA, [5]G.GIRIJA RANI

[1] GUIDE [2,3,4,5]U.G SCHOLAR

[1,2,3,4,5]DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERNG

[1,2,3,4,5]RAVINDRA COLLEGE OF ENGINEERING FOR WOMEN

ABSTRACT

. Object detection is a vast, vibrant and complex area of computer vision. If there is a single object to be detected in an image, it is known as Image localization and if there are multiple objects in an image, then it is known as Object Detection. This detects the semantic objects of a class in digital images and videos. The applications of object detection include tracking objects, video surveillance, pedestrian detection, people counting, self driving cars, face detection, ball tracking in sports and many more. Convolutional Neural Networks is a representative tool of Deep Learning detects objects using OpenCV (Open-Source Computer Vision), which is a library of programming functions mainly aimed at real time computer vision.

Mobile networks and binary neural networks are the most commonly used techniques for modern deep learning models to perform a variety of tasks on embedded systems. In this project, we develop a technique to identify an object considering the deep learning pre-trained model MobileNet for Single Shot Multi-Box Detector (SSD). This algorithm is used for real

time detection, and for webcam feed to detect the purpose webcam which detects the object in a video stream. Therefore, we use an object detection module that can detect what is in the video stream. In order to implement the module, we combine the MobileNet and the SSD framework for a fast and efficient deep learning-based method of object detection. The main purpose of our research is to elaborate the accuracy of an object detection method SSD and the importance of pre-trained deep learning model MobileNet. The experimental results show that the Average Precision (AP) of the algorithm to detect different classes as car, person and chair is 99.76%, 97.76% and 71.07%, respectively. This improves the accuracy of behavior detection at a processing speed which is required for the real-time detection and the requirements of daily monitoring indoor and outdoor.

I. INTRODUCTION

1.1 INTRODUCTION

Object detection is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization is referring to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term "object

recognition ", they often mean "object detection ". It may be challenging for beginners to distinguish between different related computer vision tasks.

Image classification also involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. Object detection is always more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label. Together, all these problems are referred to as object recognition.

1.2 MOTIVATION

Blind people do lead a normal life with their own style of doing things. But, they definitely face troubles due to inaccessible infrastructure and social challenges. The biggest challenge for a blind person, especially the one with the complete loss of vision, is to navigate around places. Obviously, blind people roam easily around their house without any help because they know the position of everything in the house. Blind people have a tough time finding objects around them. So, we decided to make a OBJECT DETECTION System. We are interested in this project after we went through few papers in this area. The main intention of the project is when we view an image the objects present in it are recognized by our brain instantaneously.

1.3 PROBLEM DEFINITION

Object detection is the problem of finding and classifying a variable number of objects in an image. The important difference is the "variable" part. In contrast, with the problems like classification, the output of object detection is variable in length, since the number of objects detected may change from image to image. The main purpose of object detection is to identify and locate one or more effective targets from still image or video data.

1.4 OBJECTIVE OF THE PROJECT

The motive of object detection is to recognize and locate all known objects in a scene. Preferably in 3D space, recovering pose of objects in 3D is very important for robotic control systems. Imparting intelligence to machines and making robots more and more autonomous and independent has been a sustaining technological dream for the mankind. It is our dream to let the robots take on tedious, boring, or dangerous work so that we can commit our time to

DEPARTMENTOF CSE,RCEW, KURNOOL Page 1

INTRODUCTION CHAPTER 1

more creative tasks. Unfortunately, the intelligent part seems to be still lagging behind. In real life, to achieve this goal, besides hardware development, we need the software that can enable robot the intelligence to do the work and act independently. One of the crucial components regarding this is vision, apart from other types of intelligences such as learning and cognitive thinking. A robot cannot be too intelligent if it cannot see and adapt to a dynamic environment.

1.5 LIMITATIONS

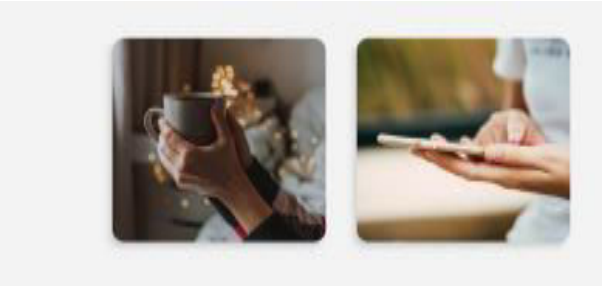
· Viewpoint Variation :

One of the biggest difficulties of object detection is that an object viewed from different angles may look completely different. For example, the images of the cakes that you can see below differ from each other because they show the object from different sides.

· Deformation :

The subject of computer vision analysis is not only a solid object but also bodies that can be deformed and change their shapes, which provides additional complexity for object detection. Look at the images of football players in different poses

· Occlusion :



Sometimes objects can be obscured by other things, which makes it difficult to read the signs and identify these objects. For example, in the first below image, a cup is covered by the hand of the person holding this cup.

II. SPECIFICATIONS

SYSTEM SPECIFICATIONS CHAPTER 2 **2. SYSTEM SPECIFICATIONS**

2.1 SOFTWARE SPECIFICATIONS

Install python on your computer system.

1. Install Image AI and its dependencies like tensor flow, NumPy, OpenCV, etc. 2. Download the Object Detection

STEPS TO BE FOLLOWED: -

1. Download and install Python version 3 from official Python Language website https://python.org

2. Install the following dependencies via pip

i. TensorFlow:

TensorFlow is an open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning application such as neural networks, etc... It is used for both research and production by Google.

TensorFlow computations are expressed as stateful dataflow graphs. The name TensorFlow derives from operations that such neural networks perform on multidimensional data arrays, which are referred to as tensors.

pip install tensor flow –command

ii. NumPy:

NumPy is library of Python programming language, adding support for large, multi dimensional array and matrices, along with large collection of high-level mathematical function to operate over these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several developers. In 2005 Travis Olphant created NumPy by incorporating features of computing Numarray into Numeric, with extension

modifications. NumPy is open-source software and has many contributors. pip install numPy

iii. SciPy:

SciPy contain modules for many optimizations, linear algebra, integration, interpolation, special function, FFT, signal and image processing, ODE solvers and other tasks common in engineering. SciPy abstracts majorly on NumPy array object, and is the part of the NumPy stack which include tools like Matplotlib, pandas and SymPy, etc., and an

expanding set of scientific computing libraries. This NumPy stack has similar uses to other applications such as MATLAB, Octave, and Scilab. The NumPy stack is also sometimes referred as the SciPy stack. The SciPy library is currently distributed under BSD license, and its development is sponsored and supported by an open communities of developers. It is also supported by NumFOCUS, community foundation for supporting reproducible and accessible science. pip install scipy – command

iv. OpenCV:

OpenCV is an library of programming functions mainly aimed on real time computer vision. originally developed by Intel, it is later supported by Willow Garage then Itseez. pip install opencv-python-command

DEPARTMENTOF CSE,RCEW, KURNOOL Page 3

SYSTEM SPECIFICATIONS CHAPTER 2

v. Pillow:

Python Imaging Library is a free Python programming language library that provides support to open, edit and save several different formats of image files. Windows, Mac OS X and Linux are available for this.

pip install pillow-command

vi. Matplotlib:

Matplotlib is a Python programming language plotting library and its NumPy numerical math extension. It provides an object-oriented API to use general-purpose GUI toolkits such as Tkinter, wxPython, Qt, or GTK+ to embed plots into applications.

pip install matplotlib – command

vii. H5py:

The software h5py includes a high-level and low-level interface for Python's HDF5 library. The low interface expected to be complete wrapping of the HDF5 API, while the high-level component uses established Python and NumPy concepts to support access to HDF5 files, datasets and groups..

pip install h5py– command

viii. Keras :

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

pip install keras– command

ix. ImageAI:

Image AI provides API to recognize 1000 different objects in a picture using pre-trained models that were trained on the ImageNet-1000 dataset. The model implementations provided are SqueezeNet, ResNet, InceptionV3 and DenseNet.

pip3 install imageai—upgrade

3.Download the RetinaNet model file that will be used for object detection using following link

https://github.com/OlafenwaMoses/ImageAI/releases/download/1.0/resnet50_coco_best_v2.0. 1.h5

2.2 HARDWARE SPECIFICATIONS

The algorithms that are shown throughout this paper are designed for a specific target, NVIDIA Jetson TX1. This system-on-a-chip contains a CPU module and a GPU module [9]. The CPU module is a quad-core ARM Cortex-A57 CPU. The GPU module is a Maxwell GPU that has 256 cores, evenly distributed onto two Stream Multiprocessors.

It was selected because of its low power consumption which is estimated lower than 20W when fully utilized, yet its high computational capacity per watt. Moreover this system-on-a-chip supports standard and widely used frameworks such as Compute Unified Device Architecture (CUDA). In addition it supports 16-bit floating point operations which may enable to reach a better performance if precision is not the most important aspect.

III. LITERATURE SURVEY

3.2 EXISTING SYSTEM

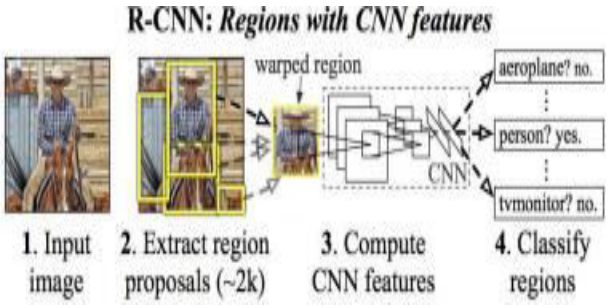
ResNet :

To train the network model in a more effective manner, we herein adopt the same strategy as that used for DSSD (the performance of the residual network is better than that of the VGG network). The goal is to improve accuracy. However, the first implemented for the modification was the replacement of the VGG network which is used in the original SSD with ResNet. R-CNN :

Therefore, instead of trying to classify the huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated by using the selective search algorithm which is written below.

Selective Search:

1. Generate the initial sub-segmentation, we generate many candidate regions. 2. Use the greedy algorithm to recursively combine similar regions into larger ones. 3. Use generated regions to produce the final candidate region proposals



These 2000 candidate regions which are proposals are warped into a square and fed into a convolutional neural network that produces a 4096-dimensional feature vector as output. The CNN plays a role of feature extractor and the output dense layer consists of the features extracted from the image and the extracted features are fed into an SVM for the classify the presence of the object within that candidate region proposal In addition to predicting the presence of an object within the region proposals. For example, given the region proposal, the algorithm might have predicted the presence of a person but the face of that person within that region proposal could have been cut in half.

DEPARTMENTOF CSE,RCEW, KURNOOL Page 5

LITERATURE SURVEY CHAPTER 3

FAST R-CNN:

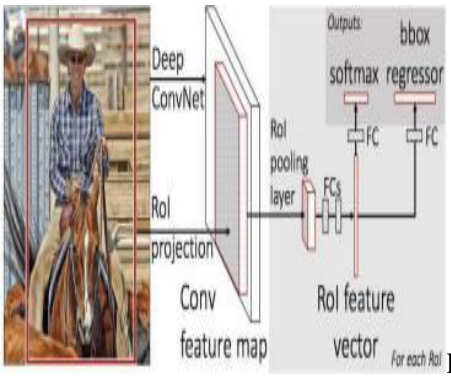


Figure : Fast R-CNN

The same author solved some of the drawbacks of R-CNN to build a faster object detection algorithm it was called Fast R-CNN. The approach is similar to the R-CNN algorithm.

But, instead of feeding the region proposals to the CNN, we feed the input image to the CNN to generate a convolutional feature map. From the convolutional feature map, we can identify the region of the proposals and warp them into the squares and by using an RoI pooling layer we reshape them into the fixed size so that it can be fed into a fully connected layer.

The reason "Fast R-CNN" is faster than R-CNN is because you don't have to feed 2000 region proposals to the convolutional neural network every time. Instead, the convolution operation is always done only once per image and a feature map is generated from it.

Figure : Comparison of object detection algorithms

From the above graphs, you can infer that Fast R-CNN is significantly faster in training and testing sessions over R-CNN. When you look at the performance of Fast R-CNN during testing time, including region proposals slows down the algorithm significantly when compared to not using region proposals. Therefore, the region which is proposals become bottlenecks in Fast R-CNN algorithm affecting its performance.

FASTER R-CNN:

Both of the above algorithms (R-CNN & Fast R-CNN) uses selective search to find out the region proposals. Selective search is the slow and time-consuming process which affect the performance of the network.

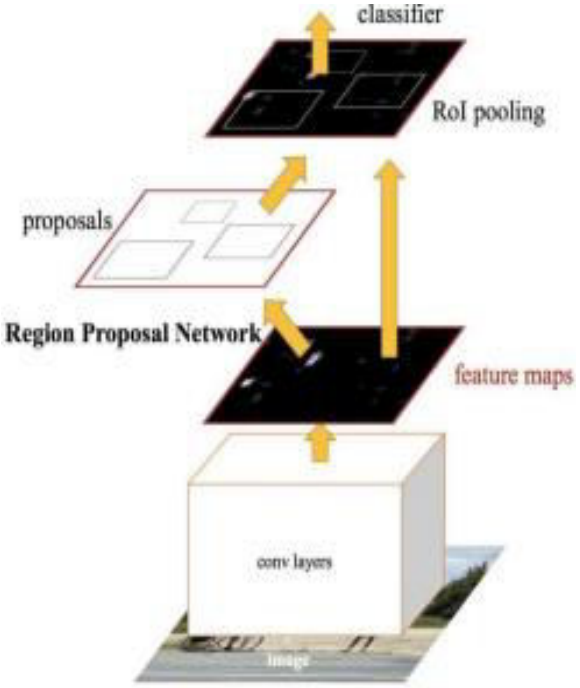


Figure : Faster R-CNN

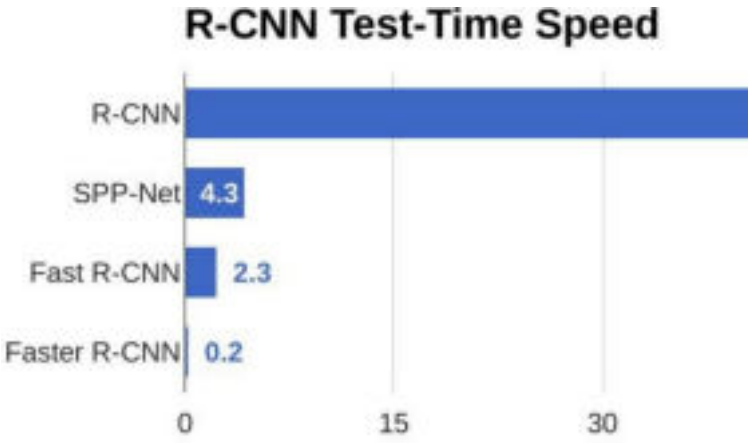


Figure : Comparison of test-time speed of object detection algorithm

From the above graph, you can see that Faster R-CNN is much faster than its predecessors. Therefore, it can even be used for real-time object detection.

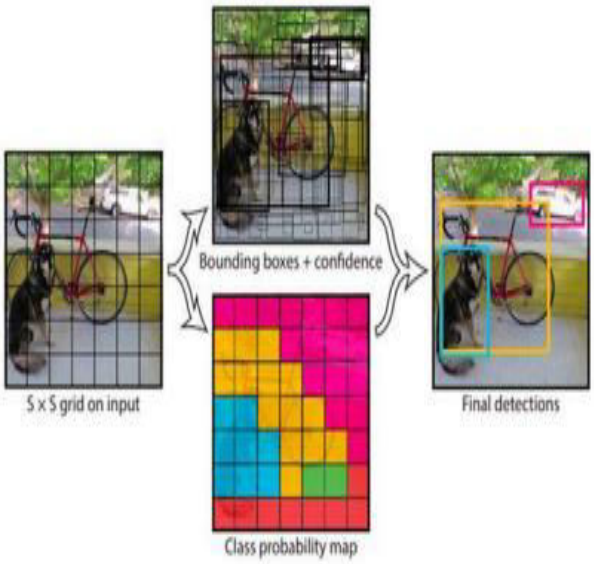
YOLO--You Only Look Once:

All the previous object detection algorithms have used regions to localize the object within the image. The network does not look at the complete image. Instead, parts of the image which has high probabilities of containing the

object. YOLO or You Only Look Once is an object detection algorithm much is different from the region-based algorithms which seen above. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes.

DEPARTMENTOF CSE,RCEW, KURNOOL Page 7

LITERATURE SURVEY CHAPTER 3



YOLO works by taking an image and split it into an SxS grid, within each of the grid we take m bounding boxes. For each of the bounding box, the network gives an output a class probability and offset values for the bounding box. The bounding boxes have the class probability above a threshold value is selected and used to locate the object within the image.

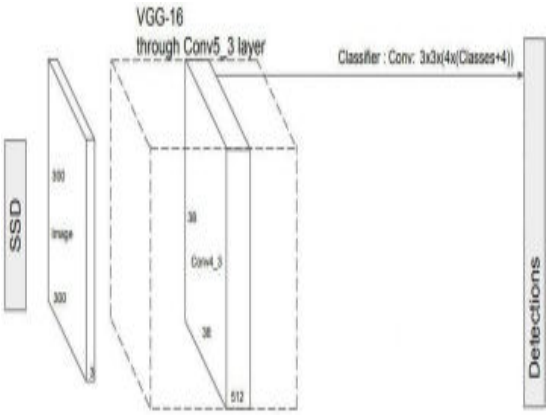
3.3. PROPOSED SYSTEM:

SSD:

The SSD object detection composes of 2 parts:

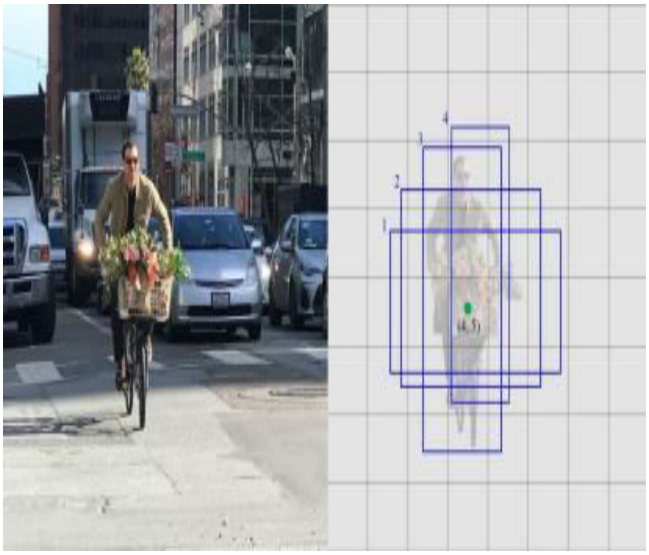
1. Extract feature maps, and

2. Apply convolution filters to detect objects.

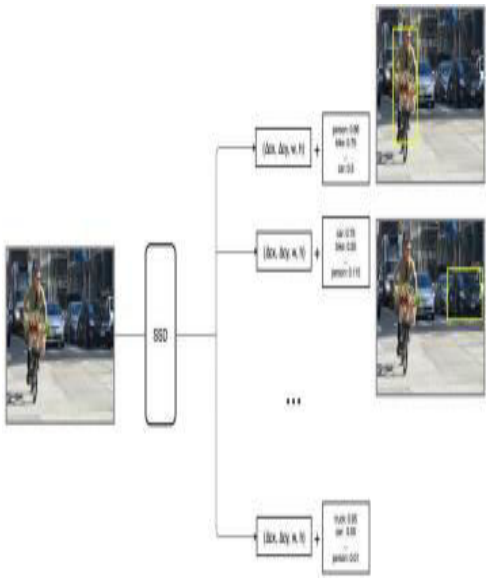


SSD uses VGG16 to extract feature maps. Then it detects objects using the Conv4_3 layer. For illustration, we draw the Conv4_3 to be 8 × 8 spatially (it should be 38 × 38). For each cell in the image (also called location), it makes 4 object predictions.

LITERATURE SURVEY CHAPTER 3



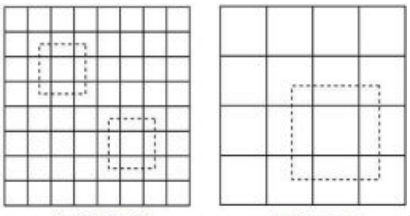
Each prediction composes of a boundary box and 21 scores for each class (one extra class for no object), and we pick the highest score as the class for the bounded object. Conv4_3 makes total of 38 × 38 × 4 predictions: four predictions per cell regardless of the depth of feature maps. As expected, many predictions contain no object. SSD reserves a class “0” to indicate



It computes both the location and class scores using small convolution filters. After extraction the feature maps, SSD applies 3 × 3 convolution filters for each cell to make predictions. (These filters compute the results just like the regular CNN filters.) Each filter gives outputs as 25 channels: 21 scores for each class plus one boundary box.

Beginning, we describe the SSD detects objects from a single layer. Actually, it uses multiple layers (multi-scale feature maps) for the detecting objects independently. As CNN reduces the spatial dimension gradually,. For example, the 4× 4 feature maps are used for the larger scale object.

LITERATURE SURVEY CHAPTER 3



. SSD adds 6 more auxiliary convolution layers to image after VGG16. Five of these layers will be added for object detection. In which three of those layers, we make 6 predictions instead of 4. In total, SSD makes 8732 predictions using 6 convolution layers

Multi-scale feature maps enhance accuracy. The accuracy with different number of feature map layers is used for object.

MANet:

Target detection is fundamental challenging problem for long time and has been a hotspot in the area of computer vision for many years. The purpose and objective of target detection is, to determine if any instances of a specified category of objects exist in an image. If there is an object to be detected in a specific image, target detection returns the spatial positions and the spatial extent of the instances of the objects (based on the use a bounding box, for exampleTarget detection is also widely used in areas such as artificial intelligence and information technology, including machine vision, automatic driving vehicles, and human– computer interaction. In recent times, the method automatic learning of represented features from data based on deep learning has effectively improved performance of target detection. Neural networks are foundation of deep learning. Therefore, design of better neural

networks has become an key issue toward improvement of target detection algorithms and performance. Recently developed object detectors that has been based on convolutional neural networks (CNN) has been classified in two types: The first is two-stage detector type, such as Region-Based CNN (R–CNN), Region-Based Full Convolutional Networks (R–FCN), and Feature Pyramid Network (FPN), and the other is single-stage detector, such as the You Only Look Once (YOLO), Single-shot detector (SSD), and the Retina Net. The former type generates an series of candidate frames as samples of data, and then classifies the samples based on a CNN; the latter type do not generate candidate frames but directly converts the object frame positioning problem into a regression processing problem.

3.4. DISADVANTAGES OF EXISTING SYSTEM:

Problems mainly occur with R-CNN.

• It takes huge amount to train n the network as you would have to classify 2000 region proposals per image.

• It cannot be implemented real time as it takes around 47 seconds for each test image.

• The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage. This could lead to the generation of bad candidate region proposals.

3.5 CONCLUSION

Due to its powerful learning ability and advantages in dealing with occlusion, scale transformation and background switches, deep learning-based object detection has been a research hotspot in recent years. This paper provides a detailed review on deep learning based object detection frameworks which handle different subproblems, such as occlusion, clutter and low resolution, with different degrees of modifications on R-CNN. The review starts on generic object detection pipelines which provide base architectures for other related tasks. Then, three other common tasks, namely salient object detection, face detection and pedestrian detection, are also briefly reviewed. Finally, we propose several promising future directions to gain a thorough understanding of the object detection landscape.

METHODOLOGY

4.1 SqueezeNet

SqueezeNet is name of a DNN for computer vision. SqueezeNet is developed by researchers at Deep Scale, University of California, Berkeley, and Stanford University together. In SqueezeNet design, the authors goal is to create a smaller neural network with few parameters that can more easily fit into memory of computer and can more easily be transmitted over a computer network.SqueezeNet is originally released in 2016. This original version of SqueezeNet was implemented on top of the Caffe deep learning software framework. The open-source research community ported SqueezeNet to a number of other deep learning frameworks. And is released in additions, in 2016, Eddie Bell released a part of SqueezeNet for the Chainer deep learning framework. in 2016, Guo Haria released a part of SqueezeNet for the Apache MXNet framework. 2016, Tammy Yang released a port of SqueezeNet for the Keras framework. In 2017, companies including Baidu, Xilnx, Imagination Technologies, and Synopsys demonstrated SqueezedNet running on low-power processing platforms such as smartphones, FPGAs, and custom processors. SqueezeNet ships as part of the source code of a number of deep learning frameworks such as PyTorch, Apache MXNet, and Apple CoreML.In addition, 3rd party developers have created implementation of SqueezeNet that are compatible with frameworks such as TensorFlow. Below is summary of frameworks that support SqueezeNet

4.2 InceptionV3

Inception v3 is widely used as image recognition model that has showed to obtain accuracy of greater than 78.1% on the ImageNet dataset. The model is the culmination of many ideas developed by researchers over years. It is based on “Rethinking the Inception Architecture Computer Vision” by Szegedy.

The model is made of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, concats, dropouts, and fully connected layers. Batchnorm is used more throughout the model and applied to activation inputs. Loss is

computed via Softmax. A high-level diagram of the model is shown below

IV. .DESIGN

5.1 CAFFE MODEL

Caffe is a framework of Deep Learning and it was made used for the implementation and to access the following things in an object detection system.

· Expression: Models and optimizations are defined as plaintext schemas in the caffe model unlike others which use codes for this purpose.

· Speed: For research and industry alike speed is crucial for state-of-the-art models and massive data.

· Modularity: Flexibility and extension is majorly required for the new tasks and different settings.

· Openness: Common code, reference models, and reproducibility are the basic requirements of scientific and applied progress.

TYPES OF CAFFE MODELS

1. Open Pose: The first real-time multi-person system is portrayed by OpenPose which can collectively sight human body, hand, and facial key points (in total 130 key points) on single pictures. Cnn-vis: Cnn-vis is an open-source tool that lets you use convolutional neural networks to generate images. It has taken inspiration from the Google's recent Inceptionism blog post.

2. Speech Recognition: Speech Recognition with Caffe deep learning frame work. 3. DeconvNet: Learning Deconvolutional Network for Semantic Segmentation.

5.2 OPEN CV

OpenCV stands for Open supply pc Vision Library is associate open supply pc vision and machine learning software system library. The purpose of creation of OpenCV was to produce a standard infrastructure for computer vision applications and to accelerate the utilization of machine perception within the business product [6]. It becomes very easy for businesses to utilize and modify the code with OpenCV as it is a BSD-licensed product. It is a rich wholesome library as it contains 2500 optimized algorithms, which also includes a comprehensive set of both classic and progressive computer vision and machine learning algorithms. These algorithms is used for various functions such as discover and acknowledging faces. Identify objects classify human actions. In videos, track camera movements, track moving objects. Extract 3D models of objects, manufacture 3D purpose clouds from stereo cameras, sew pictures along to provide a high-resolution image of a complete scene, find similar pictures from a picture information, remove red eyes from images that are clicked with the flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality.

APPLICATIONS OF OPENCV

--Object Detection

--Egomotion estimation

DEPARTMENTOF CSE,RCEW, KURNOOL Page 14

DESIGN CHAPTER 5

--Facial recognition system

--2D and 3D feature toolkits

--Segmentation and recognition

--Motion tracking

--Augmented Reality

--Structure from Motion (SFM)

--Motion understanding

--Stereopsis stereo vision: depth perception from 2 cameras

--Mobile robotics

--Human-computer interaction

LIBRARIES IN OPENCV

NumPy:

NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open source extension module for Python Furthermore, NumPy enriches the programming language Python with powerful data structures for efficient computation of multi-dimensional arrays and matrices. The implementation is even aiming at huge matrices and arrays.

Besides that the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays. It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

· A powerful N-dimensional array object Sophisticated (broadcasting) functions · Useful linear algebra, Fourier Transform, and random number capabilities.

Rectangular Haar-like features:

A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called 2-rectangle feature. Viola and Jones also defined 3-rectangle features and 4- rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region.

Fast Computation of Haar-like features:

One of the contributions of Viola and Jones was to use summed-area tables, which they called integral images. Integral images can be defined as two-dimensional lookup tables in the form of a matrix with the same size of the original image. Each element of the integral image contains the sum of all pixels located on the up-left region of the original image

The above figure indicates the rectangle feature.

Sum=I(C)+I(A)-I(B)-I(D)

OpenCV has a modular structure, which means that the package includes several shared or static libraries. The following modules are available:

• Core functionality (core) - a compact module defining basic data structures, including the dense multidimensional array Mat and basic functions used by all other modules.

• Image Processing (imgproc) - an image processing module that includes linear and non linear image filtering, geometrical image table-based remapping), color specific transformations (resize, affine and perspective conversion, histograms, and so on.

• Video Analysis (video) - a video analysis module that includes motion esti subtraction, and object tracking algorithms.

• Camera Calibration and 3D Reconstruction (calib3d) - basic multiple-view geometry algorithms, single and stereo camera calibration, object pose estimation, stereo correspondence algorithms, and elements of 3D reconstruction.

• 2D Features Framework (features2d) - salient feature detectors, descriptors, and descriptor matchers.

5.3 MOBILENET SSD:

Just like classification, here also, we will leverage the pre-trained models. These models have been trained on the MS COCO dataset, the current benchmark dataset for deep learning based object detection models.MS COCO has almost 80 classes of objects, starting from a person, to a car, to a toothbrush. The dataset contains 80 classes of everyday objects. We will also use a text file to load all the labels present in the MS COCO dataset for object

detection. We will use MobileNet SSD (Single Shot Detector), which has been trained on the MS COCO dataset using the TensorFlow deep learning framework. SSD models are generally faster when compared to other object detection models. Moreover, the MobileNet backbone also makes them less compute-intensive. So, it is a good model to start learning about object detection with OpenCV DNN.MobileNet is a lightweight deep neural network architecture designed for mobiles and embedded vision applications. The 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.SSD provides localization while mobilenet provides classification.

V. **RESULT ANALYSIS**

VI. **2 DESCRIPTION OF KEY PARAMETERS AND FUNCTIONS:** ImageAI provides many more features useful for customization and production capable deployments for object detection tasks. Some of the features supported are: -

VII. · tkinter: In Python, Tkinter is a standard GUI (graphical user interface) package. Tkinter is Python's default GUI module and also the most common way that is used for GUI programming in Python. Note that Tkinter is a set of wrappers that implement the Tk widgets as Python classes.

VIII. · Message box: Message Box Widget is used to display the message boxes in pyhton applications. This module is used to display a message using provides as a number of functions.

IX. · Detection Speeds: You can reduce the time it takes to detect an image by setting the speed of detection speed to “fast”, “faster” and “fastest”.

X. · Simple dialog: The SimpleDialog module is used to create dialog boxes to take input from the user in a variety of ways .Simple Dialog allows us to take input of varying data types from the user, such as float, string and integer.

XI. · file dialog: Python Tkinter (and TK) offer a set of dialogs that you can use when working with the files.File dialogs help you open, save files or directories.

XII. · readNetFromCaffee():readNetFromCaffe() function for reading a network model stored in Caffe framework with args for “prototxt ”and “model” file paths.

XIII. · BLOB: BLOB stands for Binary Large ObjectBLOB is a large complex collection of binary data which is stored in Database. Basically BLOB is used to store media files like images, video and audio files.

XIV. · Confidence:The confidence score reflects how likely the box contains an object (objectness) and how accurate is the bounding box. If no object exists in that cell, the confidence score should be zero..

XV. · destroyAllWindow():Python OpenCV destroyAllWindow() function allows user to destroy all windows at any time.

XVI. · askopenfilename:We use the askopenfilename() function to display an open file dialog that allows users to select one file.

XVII. · imutils:A series of convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV and both Python 2.7 and Python 3.

XVIII. · cv2:OpenCV-Python is a library of Python bindings designed to solve computer vision problems. cv2. imread() method loads an image from the specified file. If the image cannot be read then this method returns an empty matrix.

XIX. · NumPy:NumPy provides standard trigonometric functions, functions for arithmetic operations, handling complex numbers, etc. NumPy has standard trigonometric functions which return trigonometric ratios for a given angle in radians.

TESTING AND VALIDATION

7.1.1 Testing Strategies

· Functional Testing:Once the system is completed developed and integrated it is checked and evaluated for its functionality as whole for specific demands and requirements. This type of testing falls under the category of BlackBox testing and does not require the knowledge of in depth working and protocol off in the system.

· Structural Testing: In contrary to Functional testing Structural testing checks for the functionality of the different modules of the whole system and how well they are in link with another module. This type of testing requires full knowledge of the behaviour, protocol and working of the system as a whole and module wise. The system-based coding and programming knowledge is also a requirement to perform this testing. The tester chooses inputs to exercise paths through the code and determine the appropriate outputs.

Testing the model : To test the model we first select a model checkpoint (usually the latest) and export into a frozen inference graph checkpoints is created when we train our model with the help of checkpoint we are testing our model. We divide our data and used 70% images for training and 30% for testing purpose so we split our images

in test and train folder. We store 100 of images per object to train the model of every angle of the object.

By using this thesis and based on experimental results we are able to detect objects more precisely and identify the objects individually with exact location of an objects in the picture in x,y axis. This paper also provide experimental results on different methods for object detection and identification and compares each method for their efficiencies.

The object recognition system can be applied in the area of surveillance system, face recognition, fault detection, character recognition etc. The objective of this thesis is to develop an object recognition system to recognize the 2D and 3D objects in the image. The performance of the object recognition system depends on the features used and the classifier employed for recognition. This research work attempts to propose a novel feature extraction method for extracting global features and obtaining local features from the region of interest. Also the research work attempts to hybrid the traditional

classifiers to recognize the object. The object recognition system developed in this research was tested with the benchmark datasets like COIL100, Caltech 101, ETH80 and MNIST.

It is important to mention the difficulties observed during the experimentation of the object recognition system due to several features present in the image. The research work suggests that the image is to be preprocessed and reduced to a size of 128 x 128. The proposed feature extraction method helps to select the important feature. To improve the efficiency of the

classifier, the number of features should be less in number. Specifically, the contributions towards this research work are as follows,

• An object recognition system is developed, that recognizes the two-dimensional and three dimensional objects.

• The feature extracted is sufficient for recognizing the object and marking the location of the object. x The proposed classifier is able to recognize the object in less computational cost. • The proposed global feature extraction requires less time, compared to the traditional feature extraction method.

• The performance of the SSD is greater and promising when compared with the BPN and SVM.

• Local feature PCA-SIFT is computed from the blobs detected by the Hessian-Laplace detector.

• Along with the local features, the width and height of the object computed through projection method is used.

The methods presented for feature extraction and recognition are common and can be applied to any application that is relevant to object recognition.

The proposed object recognition method combines the state-of-art classifier SSD to recognize the objects in the image. The feature extraction method proposed in this research work is efficient.

VII CONCLUSION AND FUTURE ENHANCEMENTS

provides unique information for the classifier. The image is segmented into 16 parts, from each part the Hu's Moment invariant is computed and it is converted into Eigen component. The local feature of the image is obtained by using the Hessian-Laplace detector. This helps to obtain the objects feature easily and mark the object location without much difficulty. As a scope for future enhancement

• Features either the local or global used for recognition can be increased, to increase the efficiency of the object recognition system.

• Geometric properties of the image can be included in the feature vector for recognition. 150 • Using unsupervised classifier instead of a supervised classifier for recognition of the object. • The proposed object recognition system uses grey-scale image and discards the color information. The colour information in the image can be used for recognition of the object. Colour based object recognition plays vital role in Robotics Although the visual tracking algorithm proposed here is robust in many of the conditions, it can be made more robust by eliminating some of the limitations as listed below:

• In the Single Visual tracking, the size of the template remains fixed for tracking. If the size of the object reduces with the time, the background becomes more dominant than the object being tracked. In this case the object may not be tracked.

• Fully occluded object cannot be tracked and considered as a new object in the next frame. • Foreground object extraction depends on the binary segmentation which is carried out by applying threshold techniques. So blob

extraction and tracking depends on the threshold value. • Splitting and merging cannot be handled very well in all conditions using the single camera due to the loss of information of a 3D object projection in 2D images. 39 • For Night time visual tracking, night vision mode should be available as an inbuilt feature in the CCTV camera. To make the system fully automatic and also to overcome the above limitations, in future, multi- view tracking can be implemented using multiple cameras. Multi view tracking has the obvious advantage over single view tracking because of wide coverage range with different viewing angles for the objects to be tracked.

I. VIII REFERENCES

[1] Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490. doi:10.1109/TPAMI.2004.108

[2] Alexe, B., Deselaers, T., and Ferrari, V. (2010). "What is an object?," in ComputerVision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco,CA: IEEE), 73– 80. doi:10.1109/CVPR.2010.5540226

[3] Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. Int. J.Comput. Vis. 1, 333–356. doi:10.1007/BF00133571

[4] Andreopoulos, A., and Tsotsos, J. K. (2013). 50 years of object recognition: direc-tions forward. Comput. Vis. Image Underst. 117, 827–891. doi:10.1016/j.cviu.2013.04.005

[5] Azizpour, H., and Laptev, I. (2012). "Object detection using strongly-superviseddeformable part models," in Computer Vision-ECCV 2012 (Florence: Springer),836–849.

[6] Azzopardi, G., and Petkov, N. (2013). Trainable cosfire filters for keypointdetectionand pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 490–503.doi:10.1109/TPAMI.2012.106

[7] Azzopardi, G., and Petkov, N. (2014). Ventral-stream-like shape representation:from pixel intensity values to trainable object-selective cosfire models. Front.Comput. Neurosci. 8:80. doi:10.3389/fncom.2014.00080

[8] Benbouzid, D., Busa-Fekete, R., and Kegl, B. (2012). "Fast classification using sparsedecisiondags," in Proceedings of the 29th International Conference on MachineLearning (ICML-12), ICML '12, eds J. Langford and J. Pineau (New York, NY:Omnipress), 951–958

[9] Bengio, Y. (2012). "Deep learning of representations for unsupervised and transferlearning," in ICML Unsupervised and Transfer Learning, Volume 27 of JMLRProceedings, eds I. Guyon, G. Dror, V. Lemaire, G. W. Taylor, and D. L. Silver(Bellevue: JMLR.Org), 17–36.

[10] Bourdev, L. D., Maji, S., Brox, T., and Malik, J. (2010). "Detecting peopleusing mutually consistent poselet activations," in Computer Vision – ECCV2010 – 11th European Conference on Computer

Vision, Heraklion, Crete, Greece,September 5-11, 2010, Proceedings 11. Bourdev, L. D., and Malik, J. (2009). “Poselets: body part detectors trained using 3dhuman pose annotations,” in IEEE 12th International Conference on ComputerVision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE),1365–1372