

## AN EFFECTIVE TIME OBJECT DETECT AND ANALYSING OF DEEP LEARNING

Ms. Suvarna Ashok Amle<sup>1</sup>, Dr. Pankaj Dashore<sup>2</sup>, Dr. Mohd Muqem<sup>3</sup>,  
Dr. Pawan R. Bhaladhare<sup>4</sup>

<sup>1</sup>Student of M.Tech , Sandip University, Nashik, Maharashtra, India.

<sup>2,3,4</sup>Professor, Sandip University, Nashik, Maharashtra, India.

### ABSTRACT

Actual time 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 Object Detection. This detects the semantic objects of a class in digital images and videos. The applications of real time object detection include tracking objects, video surveillance, pedestrian detection, people counting, self-driving cars, face detection, ball tracking in sports and many more. Convolution Neural Networks is a representative tool of Deep learning to detect objects using OpenCV(Opensource Computer Vision), which is a library of programming functions mainly aimed at Realtime computer vision.

**Keywords:** Computer vision, Deep Learning, Convolution Neural Networks.

### 1. INTRODUCTION

#### The Importance Of Object Detection

##### Project Objective:

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 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.

The searching or recognition process in real time scenario is very difficult. So far, no effective solution has been found for this problem. Despite a lot of research in this area, the methods developed so far are not efficient, require long training time, are not suitable for real time application, and are not scalable to large number of classes. Object detection is relatively simpler if the machine is looking for detecting one particular object. However, recognizing all the objects inherently requires the skill to differentiate one object from the other, though they may be of same type. Such problem is very difficult for machines, if they do not know about the various possibilities of objects.

##### 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 REAL TIME OBJECT DETECTION System. We are interested in this project after we went through few papers in this area. As a result we are highly motivated to develop a system that recognizes objects in the real time environment

Object Detection is the process of finding and recognizing real-world object instances such as car, bike, TV, flowers, and humans out of an images or videos. An object detection technique lets you understand the details of an image or a video as it allows for the recognition, localization, and detection of multiple objects within an image.

It is usually utilized in applications like image retrieval, security, surveillance, and advanced driver assistance systems (ADAS).Object Detection is done through many ways:

- Feature Based Object Detection
- Viola Jones Object Detection

- SVM Classifications with HOG Features
- Deep Learning Object Detection

Object detection from a video in video surveillance applications is the major task these days. Object detection technique is used to identify required objects in video sequences and to cluster pixels of these objects.

The detection of an object in video sequence plays a major role in several applications specifically as video surveillance applications.

Object detection in a video stream can be done by processes like pre-processing, segmentation, foreground and background extraction, feature extraction.

Humans can easily detect and identify objects present in an image. The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects with little conscious thought. With the availability of large amounts of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy.

### DIGITAL IMAGE PROCESSING

Computerized picture preparing is a range portrayed by the requirement for broad test work to build up the practicality of proposed answers for a given issue. A critical trademark hidden the plan of picture preparing frameworks is the huge level of testing and experimentation that

Typically is required before touching base at a satisfactory arrangement. This trademark informs that the capacity to plan approaches and rapidly model hopeful arrangements by and large assumes a noteworthy part in diminishing the cost and time required to land at a suitable framework execution.

### WHAT IS AN IMAGE?

A picture is spoken to as a two dimensional capacity  $f(x, y)$  where  $x$  and  $y$  are spatial co-ordinates and the adequacy of "T" at any match of directions  $(x, y)$  is known as the power of the picture by then.

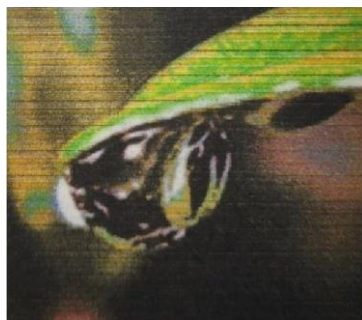


Fig. 1.1 digital image

### Processing on image:

Processing on image can be of three types They are low-level, mid-level, high level.

#### Low-level Processing:

- Preprocessing to remove noise.
- Contrast enhancement.
- Image sharpening.

### APPLICATION OF OBJECT DETECTION

The major applications of Object Detection are:

#### FACIAL RECOGNITION

“Deep Face” is a deep learning facial recognition system developed to identify human faces in a digital image. Designed and developed by a group of researchers in Facebook. Google also has its own facial recognition system in Google Photos, which automatically separates all the photos according to the person in the image.

## 2. LITERATURE SURVEY

### The Model

Deep learning is a popular technique used in computer vision. We chose Convolutional Neural Network (CNN) layers as building blocks to create our model architecture. CNNs are known to imitate how the human brain works when analyzing visuals.

A typical architecture of a convolutional neural network contain an input layer, some convolutional layers, some dense layers (aka. fully-connected layers), and an output layer. These are linearly stacked layers ordered in sequence

### Input Layer

The input layer has pre-determined, fixed dimensions, so the image must be pre-processed before it can be fed into the layer. We used OpenCV, a computer vision library, for object detection in the video.

The OpenCV contains pre-trained filters and uses Adaboost to quickly find and crop the object. The cropped object is then converted into gray scale using `cv2.cvtColor` and resized to 48-by-48 pixels with `cv2.resize`. This step greatly reduces the dimensions compared to the original RGB format with three colour dimensions (3, 48, 48). The pipeline ensures every image can be fed into the input layer as a (1, 48, 48) numpy array.

### Convolutional Layers

The numpy array gets passed into the Convolution2D layer where we specify the number of filters as one of the hyper parameters. The set of filters are unique with randomly generated weights. Each filter, (3, 3) receptive field, slides across the original image with shared weights to create a feature map.

Convolution generates feature maps that represent how pixel values are enhanced, for example, edge and pattern detection. A feature map is created by applying filter 1 across the entire image. Other filters are applied one after another creating a set of feature maps.

Pooling is a dimension reduction technique usually applied after one or several convolutional layers. It is an important step when building CNNs as adding more convolutional layers can greatly affect computational time. We used a popular pooling method called MaxPooling2D that uses (2, 2) windows across the feature map only keeping the maximum pixel value. The pooled pixels form an image with dimension's reduced by 4.

## 3. METHODOLOGY

### LEARNING FEATURE HIERARCHY:

Learn hierarchy all the way from pixels classifier One layer extracts features from output of previous layer, train all layers jointly

Zero-One Loss- The models given in these deep learning tutorials are largely used for classification. The major aim of training a classifier is to reduce the amount of errors (zero-one loss) on unseen examples Negative Log- Likelihood Loss Optimizing it for large models (thousands or millions of parameters) is prohibitively expensive (computationally) because the zero-one loss isn't differentiable. In order to achieve this maximization of the loglikelihood is done on the classifier given all the labels in a training set [14]. The likelihood of the correct class and number of right predictions is not the equal, but they are pretty similar from the point of view of a randomly initialized classifier. As the likelihood and zero-one loss are different objectives but we should always see that they are co-related on the validation set but sometimes one will rise while the other falls, or vice-versa.

Stochastic Gradient Descent- Ordinary gradient descent is an easy rule within which we repeatedly create tiny steps downward on an error surface defined by a loss function of some parameters. For the aim of normal gradient descent we take into account that the training data is rolled into the loss function. Then the pseudo code of this algorithm can be represented as Stochastic gradient descent (SGD) works according to similar principles as random gradient descent (SGD) operates on the basis of similar principles as normal gradient descent. It quickly proceeds by estimating the gradient from simply a few examples at a time instead of complete training set. In its purest kind, we use simply one example at a time to estimate the gradient. Caffe is a deep learning framework or else we can say a library it's made with expression speed and modularity in mind they will put by Berkeley artificial intelligence research and created by young King Gia there are many deep learning or machine learning frameworks for computer vision like tensorflow, Tiana, Charis and SVM[2]. But why exactly we implement edition cafe there as on is its expressive architecture we can easily switch between CPU and GPU while training on GPU machine modules and optimization for Our problem is defined by configuration without hard coding. It supports extensible code since cafes are open source library. It is four foot by over twenty thous and developers and github since its birth it offers coding platform in extensible languages like Python and C++. The next reason is speed for training the neural networks speed is the primary constraint. Caffe can process over million images in a single day with the standard media GPU that is milliseconds per image. Whereas the same dataset of million images can take weeks for Tiana and Kara's Caffe is the fastest convolution neural network present community as mentioned earlier since its open source library huge number of research are powered by cafe and every single day something new is coming out of it.

#### 4. RESULT

Here, in this project we've considered around 15 to 20 objects to be detected during the training. Some of those include 'person', 'car', 'train', 'bird', 'sofa', 'dog', 'plant', 'aero plane', 'bicycle', 'bus', 'motorbike', etc. The output of this project displays the objects detected with a rectangular box around the object with a label indicating its name and therefore the exactness with which the object has been detected on the top of it. It can dig out any number of objects existing during a single image with certainty.

#### 5. CONCLUSION

Deep-learning based object detection has been a search hotspot in recent years. This project starts on generic object detection pipelines which give base architectures for other related tasks. With the assistance of this the 3 other common tasks, namely object detection, face detection and pedestrian detection, are often accomplished. Authors accomplished this by combining 2 things: Object detection with deep learning and OpenCV and Efficient, threaded video streams with OpenCV. The camera sensor noise and lightening condition can change the result because it can create problem in recognizing the objects. generally, this whole process requires GPU's rather than CPU's. But we've done using CPU's and executes in much less time, making it efficient. Object Detection algorithms act as a mixture of both image classification and object localization. It takes the given image as input and produces the output having the bounding boxes adequate to the amount of objects present within the image with the category label attached to every bounding box at the highest. It projects the scenario of the bounding box up the shape of position, height and width.

#### 6. REFERENCES

- [1] Bruckner, Daniel. ML-o-scope: a diagnostic visualization system for deep machine learning pipelines.
- [2] No. UCB/EECS-2014-99.CALIFORNIA UNIV BERKELEY DEPT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES, 2014.
- [3] K Saleh, Imad, Mehdi Ammi, and Samuel Szonieczky, eds. Challenges of the Internet of Things: Technique, Use, Ethics. John Wiley & Sons, 2018.
- [4] Petrov, Yordan. Improving object detection by exploiting semantic relations between objects.MS thesis.Universitat Politècnica de Catalunya, 2017.
- [5] Nikouei, SeyedYahya, et al. "Intelligent Surveillance as an Edge Network Service: from Harr-Cascade, SVM to a Lightweight CNN." arXiv preprint arXiv:1805.00331 (2018).
- [6] Thakar, Kartikey, et al. "Implementation and analysis of template matching for image registration on DevKit-8500D." OptikInternational Journal for Light and Electron Optics 130 (2017): 935-944.
- [7] Bradski, Gary, and Adrian Kaehler.Learning OpenCV: Computer vision with the OpenCV library." O'Reilly Media, Inc.", 2008.
- [8] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).
- [9] Kong, Tao, et al. "Ron: Reverse connection with objectness prior networks for object detection." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).IEEE, 2017.
- [10] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision.Springer, Cham, 2016. Veiga, Francisco José Lopes. "Image Processing for Detection of Vehicles In Motion." (2018).
- [11] Huaizheng Zhang, Han Hu, GuanyuGao, Yonggang Wen, Kyle Guan, "Deepqoe: A Unified Framework for Learning to Predict Video QoE", Multimedia and Expo (ICME) 2018 IEEE International Conference on, pp. 6, 2018. Shijian Tang and Ye Yuan, "Object Detection based on Conventional Neural Network".
- [12] R. P. S. Manikandan, A. M. Kalpana, "A study on feature selection in big data", Computer Communication and Informatics (ICCCI) 2017 International Conference on, pp. 1-5, 2017 Warde-Farley, David. "Feedforward deep architectures for classification and synthesis." (2018).
- [13] Shilpisingh et al" An Analytic approach for 3D Shape descriptor for face recognition", International Journal of Electrical, Electronics, Computer Science & Engineering (IJECSSE), Special Issue - ICSCAAIT-2018| E-ISSN: 2348-2273 | P-ISSN: 2454-1222,pp-138-140.