Object Detection Using Convolutional Neural Networks and CAFFE Models

Manjula.B[1] Roshini.A.R[2] S.Julia Faith[3]
[1,2]UG Scholar [3]Assistant Professor
[1,2,3]Department of Information Technology
[1,2,3]S.A. Engineering College, India

Abstract— An object detection system recognises and searches the objects of the real world out of a digital image or a video, where the object can belong to any class or category. For example humans, cars, vehicles and so on. We have used Caffe model and convolutional neural network in order to complete this task of detecting an object. And this review paper discusses how deep learning technique is used to detect a live object, localise an object, categorise an object, extract features, appearance information, and many more, in images and videos using OpenCV and how caffee model and CNN is used to implement caffe model over other frameworks. To build our deep learning-based real-time object detector with OpenCV we need to access webcam in an efficient manner and to apply object detection to each frame.

Keywords: Caffe model, CNN, OpenCV

I. INTRODUCTION:

There are many applications and devices available online to assist the visually impaired person in their daily life. Some application required live connection to the server or cloud to process the objects. But none of these applications is using Machine Learning to recognize objects. Machine learning can help the blind persons recognize object. Blind people must use these senses to know about their surroundings. These senses include hearing, smelling and touching. But this is also limited sense to get the idea of their complete surroundings as they can't use their eyes. To get the idea of object in their surrounding without touching, smelling or hearing they need someone assistant. Machine learning software can help and work as an assistant for them. The concept of object detection has been there for quite some time. Methods like object detection, it's classification, the feature extraction can help us in detection object. But the major problem with these methods was that they can be only work for some specific object and required lot of supervision. Since the advancement of Convolutional neural network and Caffe model, researchers started using machine learning techniques to detect object. They used some statistical models and update their parameters using training and come up with the final model. However, the old machine learning methods were still having less accuracy until in year 2012 researchers applied deep learning technique to detect object and the accuracy of object detection greatly increased. In this paper one of the foundations deep neural network for detecting object will be discussed. Thus the neural network is used which is also known as CNN (Convolutional Neural Network) and Caffe models as well.

II. PROBLEM STATEMENT:

The problem definition of object detection is to determine where objects are located in a given image or a video or in a case of live scenario (object localization) and which category each object belongs to (object classification) nothing but extracting it's features in order to detect which particular object is that. So the pipeline of traditional object detection models can be mainly using Convolutional neural network and Caffe model.

III. EXISTING SYSTEM:

Object detection is one of the important problems in the field of machine learning. Deep learning architectures are used in many machine learning applications such as image classification and object detection. The ability to manipulate large image clusters and implement them quickly makes deep learning a popular method in classifying images. This study points out the success of the convolutional neural networks which is the architecture of deep learning, in solving image classification problems. In the study, the convolutional neural network model of the winner of ilsvrc12 competition is implemented. The method distinguishes 1.2 million images with 1000 categories in success. The application is performed with the caffe library, and the image classification process is employed. In the application that uses the speed facility provided by GPU, the test operation is performed by using the images in Caltech-101 dataset.

A. Disadvantages of existing system:

The classification accuracy is less and mismatching of object due to low level training

IV. PROPOSED SYSTEM:

The proposed system used Caffe model and convolutional neural network in order to complete this task of detecting an object in an image or a video. This review paper discusses how deep learning technique is used to detect a live object, localise an object, categorise an object, extract features, appearance information, and many more, in images and videos using OpenCV and how caffee mode and CNN is used to implement caffe model over other frameworks. To build our deep learning-based real-time object detector with OpenCV we need to access webcam in an efficient manner and to apply object detection to each frame.

A. Advantages of Proposed System:

Faster detection and recognisation using pretrained model. This allows us to protect the content from the unauthorized access since the image/audio/video file in which we have stored information remains unchanged, no one can easily identify whether the image contains any other data inside it or not

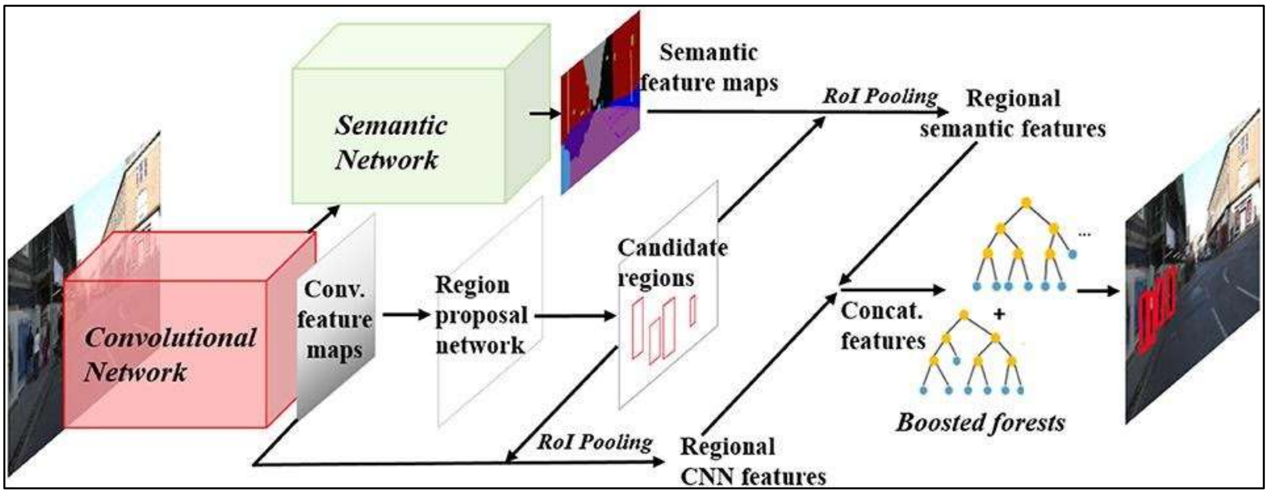

Fig. 1: Convolutional Neural Network

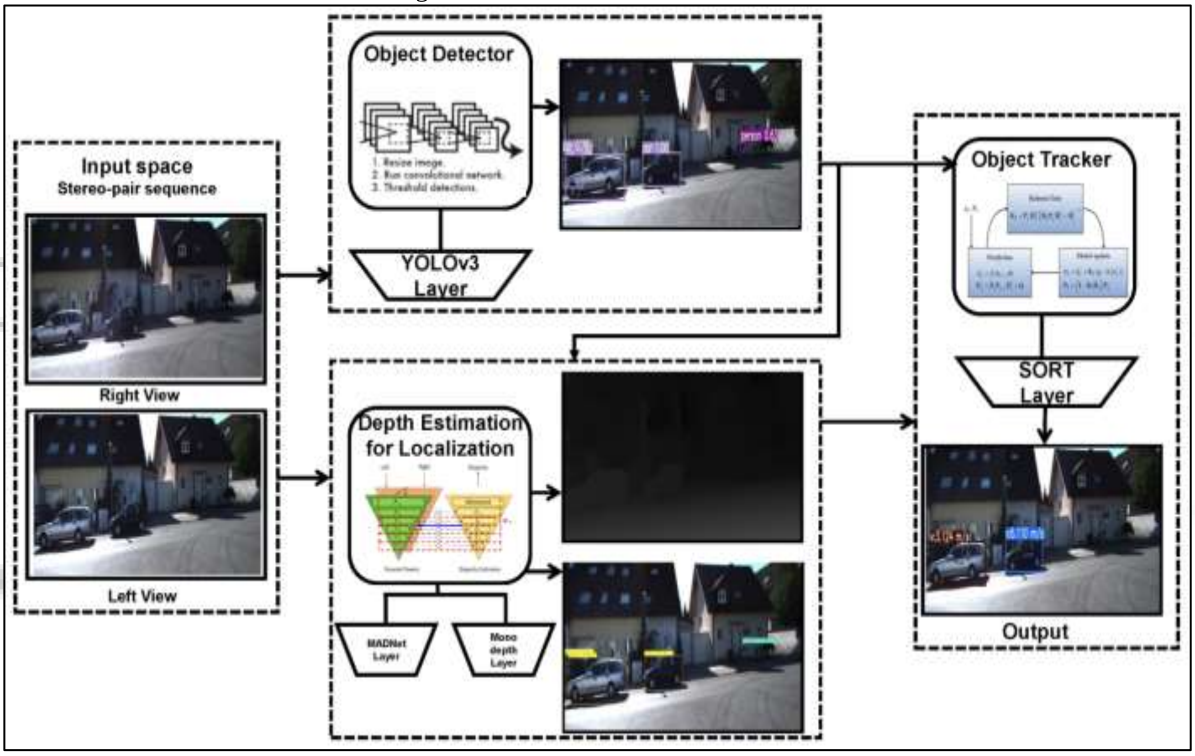

Fig. 2: Caffe Model

V. LIST OF MODULES:

1) Object detection
2) Object tracking
3) Distance measurement
4) Audio output

A. *Object detection:*

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos Well-researched domains of object detection include face detection and pedestrian detection.

B. *Object tracking:*

This object tracking algorithm is called *centroid tracking* as it relies on the Euclidean distance between (1) *existing* object centroids (i.e., objects the centroid tracker has already seen before) and (2) new object centroids between subsequent frames in a video.

C. *Distance measurement:*

In this study, a computer vision system is developed using the stereo camera system for measuring object distances. In the study, first of all, for the distance measurement, the distance of the face images obtained from the stereo camera system to the screen is calculated. In measuring the distances of the face images to screen, the disparity maps are first extracted and the face region is detected.

D. *Audio output:*

WAV files contain a sequence of bits representing the raw audio data, as well as headers with metadata in RIFF (Resource Interchange File Format) format.

For CD recordings, the industry standard is to store each audio sample (an individual audio data point relating to air pressure) as a 16-bit value, at 44100 samples per second.

To reduce file size, it may be sufficient to store some recordings (for example of human speech) at a lower sampling rate, such as 8000 samples per second, although this does mean that higher sound frequencies may not be as accurately represented. simple audio allows you to play NumPy and Python arrays and bytes objects using simple audio.play_buffer(). Make sure you have NumPy installed for the following example to work, as well as simple audio. (With pip installed, you can do this by running pip install numpy from your console.)

VI. CONCLUSION:

In conclusion, we trained Faster R-CNN to detect objects in real-time. And then we extracted features, for instance, color, HoG, SIFT descriptors, and result (the last layer of networks) given by faster R-CNN. Finally, we compared the object detected with those in our database and decided the matching one, based on the features extracted. We also recognized that the observation angle is very important for object matching. It’s essential to do the "detection" alignment.

REFERENCES:

[1] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, p. 1627, 2010.
[2] K. K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 1, pp. 39–51, 2002.
[3] C. Wojek, P. Dollar, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, p. 743, 2012.
[4] H. Kobatake and Y. Yoshinaga, “Detection of spicules on mammogram based on skeleton analysis.” IEEE Trans. Med. Imag., vol. 15, no. 3, pp. 235–245, 1996.
[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in ACM MM, 2014.
[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in NIPS, 2012.
[7] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in CVPR, 2017.
[8] Z. Yang and R. Nevatia, “A multi-scale cascade fully convolutional network face detector,” in ICPR, 2016.
[9] C. Chen, A. Seff, A. L. Kornhauser, and J. Xiao, “Deepdriving: Learning affordance for direct perception in autonomous driving,” in ICCV, 2015.
[10] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, “Multi-view 3d object detection network for autonomous driving,” in CVPR, 2017.
[11] A. Dundar, J. Jin, B. Martini, and E. Culurciello, “Embedded streaming deep neural networks accelerator with applications,” IEEE Trans. Neural Netw. & Learning Syst., vol. 28, no. 7, pp. 1572–1583, 2017.
[12] R. J. Cintra, S. Duffner, C. Garcia, and A. Leite, “Low-complexity approximate convolutional neural networks,” IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–12, 2018.
[13] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, “Cost-sensitive learning of deep feature representations from imbalanced data.” IEEE Trans. Neural Netw. & Learning Syst., vol. PP, no. 99, pp. 1–15, 2017.
[14] A. Stuhlsatz, J. Lippel, and T. Zielke, “Feature extraction with deep neural networks by a generalized discriminant analysis.” IEEE Trans. Neural Netw. & Learning Syst., vol. 23, no. 4, pp. 596–608, 2012.
[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in CVPR, 2014.