ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization

Volume 8, Special Issue 1, March 2019

A Two Days National Conference on Emerging Trends in Electronic and Instrumentation Engineering (NCETEIE 19)

12th & 13th March 2k19

Organized by

Department of Electronics and Instrumentation Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

# Object Recognition from Images and Videos Using Python with CNN

Bagyalakshmi.N[1], Sakthi.S[2], Suresh Kumar.G[3]

Associate Professor, Department of EIE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India[1]

UG Student, Department of EIE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India[2, 3]

**ABSTRACT:** Object recognition is an interfiled of computer vision that is currently heavily based on machine learning. Which machine learning subtype of a neural network called a convolution neural network (CNN) is well – suited for image and video –related tasks. The network is trained to look different features, such as edges, corners and shapes. For object recognition, the system has to both estimate the location of probable objects and to classify these. It utilizes the dark scale spatial data to fabricate word reference of pixel life length to make apparition shadows and item's leftover shadows immediately mixed into the examples of the foundation. The improved calculation takes great post-handling strategy to limit dynamic commotion. In this paper, we likewise plan a strategy utilizing two classifiers to additionally tackle the issue of inability to follow vehicles with impediments and obstruction. YOLO (you only look once) uses deep learning and CNN to detect objects, and distinguishes itself from its competitor because, as its name suggests. This speed allows you to easily detect objects in real time in videos (up to 30FPS).

**KEYWARDS:** Machine learning**,** Convolutional Neural Network (CNN), You Only Look Once (YOLO), Frame per Second (FPS), Computer vision.

## I.INTRODUCTION

There is an ever-increasing amount of image datasets in the world. Legal estimates that in 2016 still cameras and mobile devices captured more than 1.1 trillion images. According to the same estimate, in 2020 the figure will increase to 1.4 trillion. Many of these images are stored in cloud services or published on the library files. In 2014, over 1.8 billion images and videos were uploaded daily to the most popular platforms. Going beyond consumer devices, there are cameras all over the world that capture images and videos for automation purposes. Cars monitor the road, and traffic cameras monitor the same cars. In unsupervised learning, the algorithm attempts to learn useful properties of the data without a human teacher telling what the correct output should be. Classical example of unsupervised learning is clustering [1].These includes the choice of data set, the choice of success measurement, the representation of the image contents such as person, animals, vehicles & etc., the selection of inference engine, and representation of the relations between objects[2]. Imaging devices are used by engineers, doctors and space explorers alike. Automated processing of image and video contents is useful for a wide and narrow variety of image-related tasks [3]. For computer systems, this means crossing the so-called semantic gap between the pixel level information stored in the image files and the human understanding of the same images. And also it is used for sequentially scanning the objects to capture the extra hopeful infoinformation [4].Objects contained in image files can be located and identified automatically. This is called object detect and is one of the basic problems of computer vision [5]. As well as this process will demonstrate, convolutional neural networks are currently the state-of-the-art solution for object aconvolutional object recognition methods have improved in the past minority years [6]. In the experimental part, we study how easily a convolutional object recognition system can be implemented in practice, test how well a recognition system trained on general image data performs in a septic task and explore, both experimentally and based on the literature, how the current systems can be improved [7]. Edge detection is one of the major role in image recognition for computer vision system. In the local images are used for accurate and this values are computationally efficient on edge detector [8].

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019

A Two Days National Conference on Emerging Trends in Electronic and Instrumentation Engineering (NCETEIE 19)

12th & 13th March 2k19

Organized by

Department of Electronics and Instrumentation Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

## II.CONVOLUTIONAL NEURAL NETWORK

The convolutional neural network (CNN) is achieved a main role on image recognition on previous state of the art in shallow representations. The CNN methods create the new benchmark on object detection on bounding box technique [9]. R-CNN, in this method used for generating the class agnostic regions of interest. Next, step is to discuss general principles of RoI generation, and have a closer look at two popular methods are Selective Search,Edge Boxes [10].The aim of region proposal generation in object recognition is to maximize recall i.e. to generate enough regions so that all true objects are detected. The generator is less concerned with precision, since it is the task of the object detector to identify correct regions from the output of the region proposal generator [11].

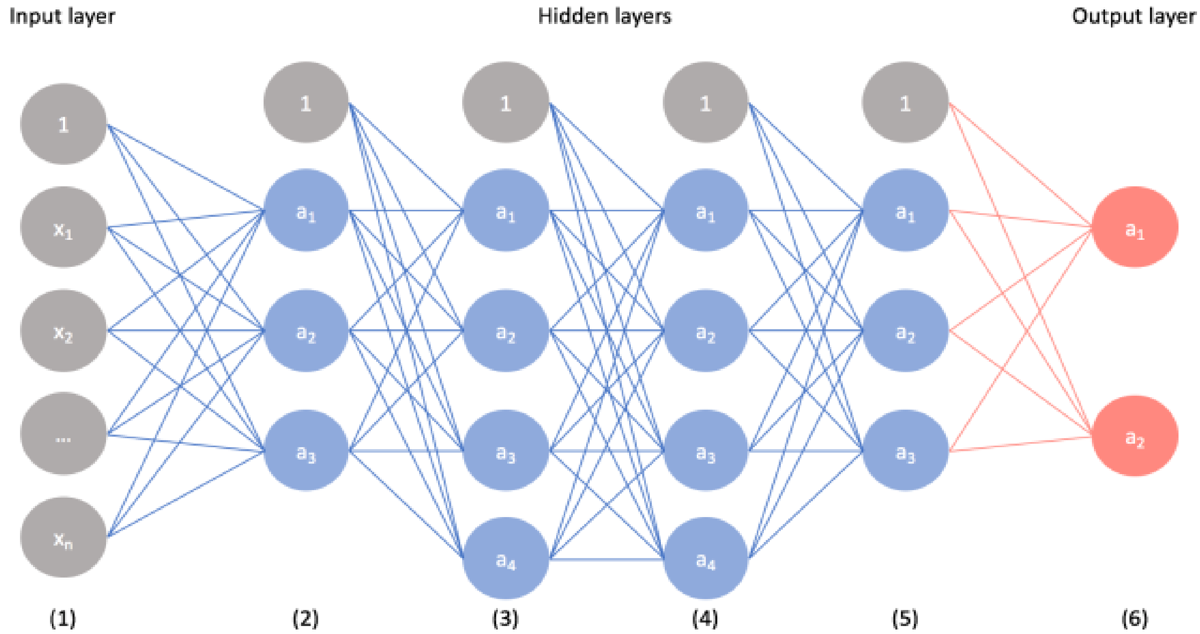



Figure.1 Convolutional Neural Network

**Selective Search:** Selective Search utilizes a hierarchical partitioning of an image and videos to create a sparse set of object locations. The main design philosophy is not to use a single strategy, but to combine the best features of bottom-up segmentation and exhaustive search [12]. The authors had three main design considerations: the search should capture all scales, be diverse i.e. not use any single strategy for grouping regions and be fast to compute.Both the region generating method and the similarity measures were selected to be fast to compute, making the method fast in general [13]. In addition to using diverse similarity measures, the search can be further diversified by using complementary color spaces (to ensure lighting invariance) and using complementary starting regions.

**Edge Boxes:** Edge Boxes is based on recognizing objects from edge maps. The main contribution of the authors of the method is the observation that the number of edge contours wholly enclosed by a bounding box is correlated with the likelihood that the box contains an object. First, the edge map is calculated using a method by the same authors called Structured Edge Detector [14]. Then, thick edge lines are thinned using non-maximum suppression.Instead of operating on the edge pixels directly, the pixels are grouped using a greedy algorithm. A unity measure is devised to calculate whether edge groups are part of the same contour. The region proposals are found by scanning the image using the traditional sliding window method and calculating an objectness score at each position, aspect ratio and scale [15].

**Deep Learning:** Deep learning is a machine learning algorithm. This algorithm are used for image recognition and visual art processing. In image recognition process is a common computational set for image classification from the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019

A Two Days National Conference on Emerging Trends in Electronic and Instrumentation Engineering (NCETEIE 19)

12th & 13th March 2k19

Organized by

Department of Electronics and Instrumentation Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

data base and the users test multiple configurations. Visual art processing is closely related to the image recognition and this process used for the visual art tasks [16].

**CONVOLUTIONAL OBJECT RECOGNITION**

Yolo v3 makes recognition across different scales, each of which deputies in detecting objects of different sizes depending upon whether they capture coarse features, fine grained features or something between you can experiment with these scales and sizes by the scales flag.Pytorch is used for implementing CNN algorithm to detect the object in videos [17].Yolo v3 makes recognitions across different scales in each object in video in each frame.Track related object recognition, such as pedestrian and vehicle recognition are popular research topics in computer vision [18]. Such objects are annotated in many publicly available collections of street view data. This provided an excellent source of data to test the generic object detector on. Cars, person, traffic light and truck are also annotated in the benchmark datasets, providing cross compatibility [19].Example esteem isn't refreshed at time t is (N-1)/N. Expecting that the time is constant, after dt time, the likelihood that the example esteem is as yet saved is appeared in the equation

P (t, t +dt) = e −lnN N−1 dt

The back propagation training algorithm, described in section is also applicable to convolutional networks. In theory, the layers closer to the input should learn to detect low-level features of the image, such as edges and corners, and the layers closer to the output should learn to combine these features to detect more meaningful shapes and sizes [20]. In an interested in studying whether convolutional networks can learn to detect complete objects.

## III.PROPOSED METHODOLOGY

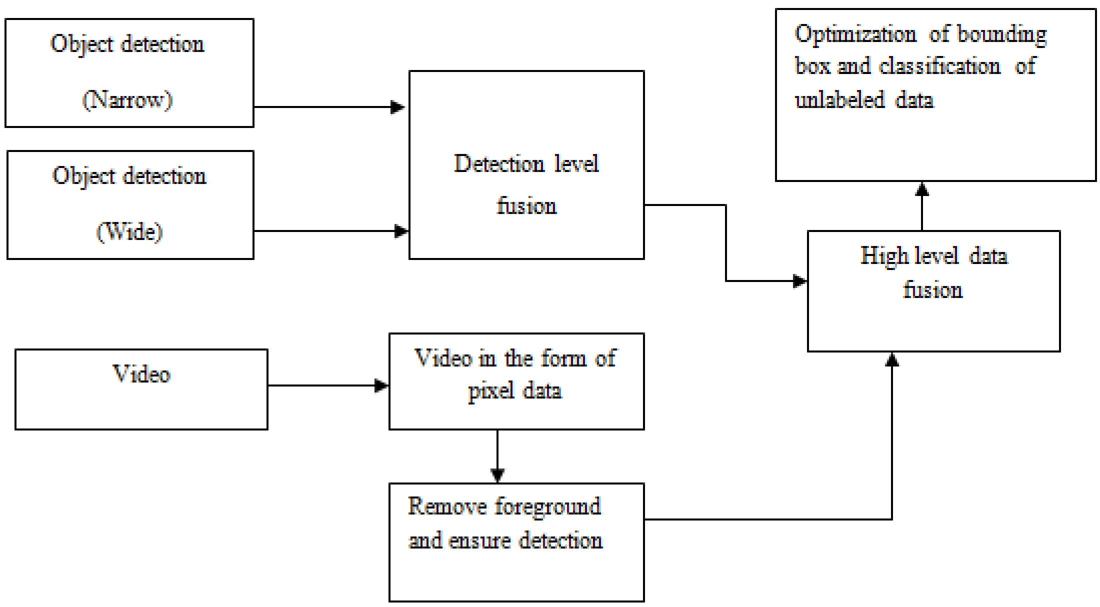



Figure.2 Proposed block diagram

Object detection for images were both (narrow and wide) input data set of image has been loaded to the program files so whenever the input image was given to the process it will compare the loaded data set to detection level fusion and high level data fusion for image identification process and output stored to the files with objects bounding box. Object recognition from videos input was measured in FPD(Form of Pixel Data) to remove the extra parts to the objects and it will compare to the loaded input data sets output stored to the files with object the bounding box. In this system was gives the more accuracy and correct labeled for the test images.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization Volume 8, Special Issue 1, March 2019

A Two Days National Conference on Emerging Trends in Electronic and Instrumentation Engineering (NCETEIE 19)

12th & 13th March 2k19

Organized by

Department of Electronics and Instrumentation Engineering, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India

## IV.RESULTS AND DISCUSSIONS

### 1. IMAGE RECOGNITION

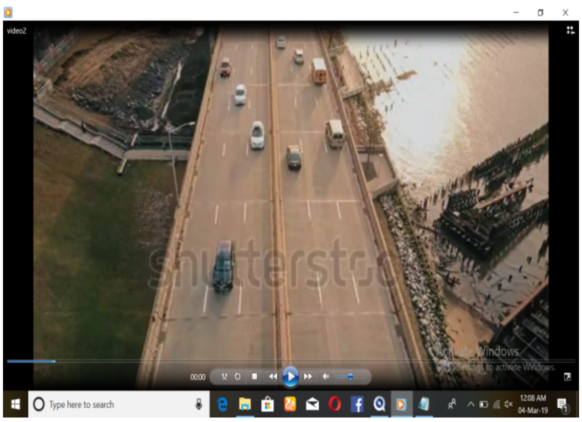
Figure 3 shows the original objects are recognized by CNN and the multi-objects are to be detected by YOLO (You Only Look Once) in frame by frame.The objects to be recognized are Cars, truck which are identified in the benchmark datasets, providing cross compatibility

Original input image


Output image recognition

Fig. 3 object recognized

### 2. VIDEO RECOGNITION



Original video input


Output video recognition

Fig. 4 object recognized from videos

In the fig .4, the different types of vehicles in the video are shown. It recognizes the shape (Selective search)andsizes (Edge Boxes) of the vehicle. The objects to be recognized are Cars, truck which are identified in the benchmark datasets, providing cross compatibility. In the multi object recognition images are detected by Single Shot Multi Detection (SSD) and the video is evaluated in frame by frame. The frame pixel speed is up to 30fps in the video.

## V.CONCLUSION

To experiment with a convolutional method in practice, created a working PYTHONimplementation of CNN. We learned that the most challenging part of implementing a deep learning system is collecting the testing data and performing the testing itself. The publicly available benchmark datasets serve as a useful starting point for both research and practical implementations. Training time can be further shortened by using a pre-trained network. Even if the final system does not feature the same objects classes as the benchmark data, visual problems are universal enough to benefit from detectors trained for a deferent problem. The optimal bottom-layers of a convolutional network are often similar regardless of the problem, just like human eye uses the same receptive fields for all visual tasks. Thus, it makes sense to initialize the layers using a pre-trained network. Additionally, this paper remains the best way of object recognition from the computer vision technique and convolutional neural networks.

## REFERENCES

[1].Bengio, Y., et al. "Deep learning of representations for unsupervised and transfer learning" ICML Unsupervised and Transfer Learning 27 (2012), 17–36.
[2].Bileschi, S. M. "Street Scenes: Towards scene understanding in still images". PhD thesis, Citeseer, 2006.
[3].Bishop, C. M. "Pattern Recognition and Machine Learning (Information Science and Statistics)". Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006
[4].Bodla, N., Singh, B., Chellappa, R., and Davis, L. S. "Improving object detection with one line of code". arXiv preprint arXiv:1704.04503 (2017).
[5].Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. "Return of the devil in the details: Delving deep into convolutional nets". arXiv preprint arXiv:1405.3531 (2014).
[6].Dalal, N., and Triggs, B. "Histograms of oriented gradients for human detection In Computer Vision and Pattern Recognition", 2005. CVPR 2005. IEEE Computer Society Conference on (2005), vol. 1, IEEE, pp. 886–893.
[7].Dollar, P., and Zitnick, C. L." Fast edge detection using structured forests". IEEE transactions on pattern analysis and machine intelligence 37, 8 (2015), 1558–1570.
[8].Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. "Scalable object detection using deep neural networks". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014), pp. 2147–2154.
[9].Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. "The pascal visual object classes (voc) challenge". International Journal of Computer Vision 88, 2 (June 2010), 303–338.
[10].Felzenszwalb, P. F., and Huttenlocher, D. P. "Enceinte graphbased image segmentation". International journal of computer vision 59, 2 (2004), 167–181.
[11].Fukushima, K. "Neocognitron: A hierarchical neural network capable of visual pattern recognition". Neural networks 1, 2 (1988), 119–130.
[12].Airsick, R. "Fast r-cnn". In Proceedings of the IEEE International Conference on Computer Vision (2015), pp. 1440–1448.
[13].Girshick, R., Donahue, J., Darrell, T., and Malik, J. "Rich feature hierarchies for accurate object detection and semantic segmentation". In Proceedings of the IEEE conference on computer vision and pattern recognition (2014), pp. 580–587.
[14].Goodfellow, I., Bengio, Y., and Courville, A. "Deep Learning". MIT Press, 2016. http://www.deeplearningbook.org.
[15].He, K., Zhang, X., Ren, S., and Sun, J. "Deep residual learning for image recognition". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016), pp. 770–778.
[16].Hoiem, D., Efros, A. A., and Hebert, M. "Automatic photo popup". ACM transactions on graphics (TOG) 24, 3 (2005), 577–584.
[17].Hoiem, D., Efros, A. A., and Hebert, M. "Geometric context from a single image. In Computer Vision", 2005. ICCV 2005. Tenth IEEE International Conference on (2005), vol. 1, IEEE, pp. 654–661.
[18].Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., and Smeulders, A. W. M. "Selective search for object recognition". International Journal of Computer Vision 104, 2 (2013), 154–171.
[19].Lowe, D. G. "Object recognition from local scale-invariant features. In Computer vision", 1999. The proceedings of the seventh IEEE international conference on (1999), vol. 2, IEEE, pp. 1150–1157.
[20].Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: "A simple way to prevent neural networks from overfitting". Journal of Machine Learning Research 15 (2014), 1929–1958.