

Fast Enhanced Unidentifiable Object Detection using Deep Learning Algorithm

+Maheswari M, Monisha V, Ramya S, Ramya S

Associate Professor⁺, Department of Computer Science Engineering, ⁺Panimalar Engineering College, Chennai, Tamilnadu, India

Students⁺, ⁺Department of Computer Science Engineering, ⁺Panimalar Engineering College, Chennai, Tamilnadu, India

Abstract— Over the years the number of sensors, cameras, and cognitive pieces of equipment placed in the wilderness has been growing exponentially. However, the resources (human) to leverage these data into something meaningful are not improving at the same rate. Our framework detects all the objects based on training set provided to it. Main view of this paper is to increase the accuracy rate of the detection. It recognizes the object even in blur stage or under less brightness. In recent days the detection of objects is done with RCNN. Our framework takes full advantage of extracting from FCN providing more advanced representation at each layer, this property is used for segment detection. It has advantages in terms of efficiency i.e., 0.08 second per image, effectiveness and simplicity over existing algorithms. It can be analyzed on the role of training data on performance, the experimental results provide a more reasonable and powerful training set for future research. Over 60-65k images can be trained for detection. This framework can be constructed using deep learning, which is a subset of machine learning in AI that has network capability of learning unsupervised from data that is unstructured or unable.

Keywords— Object Detection, RCNN Algorithm, YOLO Algorithm, Deep Learning

1. INTRODUCTION

A few years ago the development of software and hardware vision systems was mainly limited to the development of the user interface, in which most of the programmers in each company participated. The developers solved the problems of image processing themselves.

Object recognition describes a collection of Computer vision tasks that include activities such as identifying objects in video. Image classification includes activities such as predicting the class of one or more objects in an image. Object position refers to identifying the location of one or more objects in an image and drawing a bounding box around its extent. Detection does the job of combining these two tasks and finding and classifying one or more objects in an image. Image classification also assigns a class name to an image, while when objects are found, a bounding box is drawn around one or more objects in an image.

Object recognition is getting harder and harder and you combine these two tasks and draw a bounding box around each object of interest in the image and assign it a class name. Collectively, all of these problems are referred to as objects. Object recognition refers to a collection of related tasks used to identify objects in video.

Region-Based Convolutional Neural Networks (R-CNNs) is a family of techniques for addressing object location and recognition tasks designed for model performance. You Only Look Once or YOLO is known as the second family of object recognition techniques designed for speed.

Our framework recognizes all objects based on the provided training set. The main view of this paper is to increase the recognition accuracy rate. It recognizes the object even in the unsharp stage or with less brightness and 60-65,000 scanned images are trained for recognition.

II. RELATED WORKS

The methodology proposed uses multilayer convolutional neural networks to develop a model of a multilayer system in order to classify a given object into one of the defined categories. The solution then uses multiple images and identifies objects from these images by assigning them appropriate class labels. In order to improve computing power, the proposed algorithm is used in combination with a multilayer convolutional neural network, which uses a large number of predefined blocks and leads to more accurate recognition. Use various parameters to check the accuracy of object detection, such as frame drop function, frames per second (FPS), average accuracy (mAP) and aspect ratio. The experimental results prove that the method we proposed has higher accuracy.

Here, we use two different algorithms for detecting an object from video/image:

1. FRCNN (Fast Regional Convolutional Neural Network)
2. YOLO (You Only Look Once)

The combination of the results of algorithm will help us to achieve better gain.

region and the bounding box is generated in each recognition stage. Generate a region of interest. These are the regions that the network believes could contain an object. The large number of bounding boxes, each of which has an objectivity rating and get into the shifts. First, the zero padding fills the input volume around the edge with zero. If we want to apply the same convolution layer but keep the output volume the same, we will use this. Then the convolution layer extracts the features from the input images. is carried out between the images and the filter. The visual properties are extracted for each of the images. They are evaluated based on the visual properties and determined whether and which object is present. Then ReLU (Rectified Linear Unit). This level only changes all negative triggers to 0. Apply the function $f(x) = \max(0, x)$ to all values in the input volume. This layer increases the nonlinear nature of the model. Then the maximum pooling selects the highest intensity pixels in the image, which are useful when the image background is dark and we are only interested in the lightest pixel in the image. On the other hand, the convolution layer occurs in every single stage for feature extraction. Then the fully connected layer is usually placed in front of the output layer. In this input image, the previous level is flattened and fed to the output level fully connected layer, at this stage the classification process begins, then the maximum suppression, in which several regions are proposed for the same object, is not applied. After the NMS, only the box that best fits the object remains, the rest is ignored. Layers are connected to each other and there are hidden layers between them too. At last the objects are detected using the combination of these algorithms and the accuracy rate is increased.

OBJECT DATA COLLECTION

Real-time data from the Coco dataset. Data collection is one of the most important and important tasks of a machine learning project. Because the input we feed into the algorithms is data. Therefore, the efficiency and precision of the algorithm depends on the accuracy and quality of the data collected. So the same data is output. Common Objects in Context (COCO): COCO is the recognition, segmentation and labeling of objects on a large scale. It contains around 330,000 images, 200,000 of which are marked for 80 different object categories.

DATA PREPROCESSING

The data collected from various means is in a messy format, and there may be many null values, invalid data values, and unwanted data. Clean all this data and replace it with appropriate or similar data. Delete empty dates and missing data and replace them with some dates. The fixed substitute value is the most important preprocessing step of the data. All these cases need to be checked and replaced with substitute values so that the data is meaningful and useful for further processing. Data should be stored in an organized format.

EXPERIMENTAL RESULTS

Table1. Comparison of accuracy(%) between frcnn,yolo and combination of both in COCO dataset.

Number	Category	FRCNN	YOLO	Combined
1	cat	86.3	81.4	89.4
2	bird	70.8	57.7	73.5
3	person	72.0	63.5	74.7
4	car	71.6	55.9	73.1
-	mAP	75.17	64.62	77.67

The above table shows that performance is better when we combine both the algorithms for detecting an object and we get a gain of 2.5.

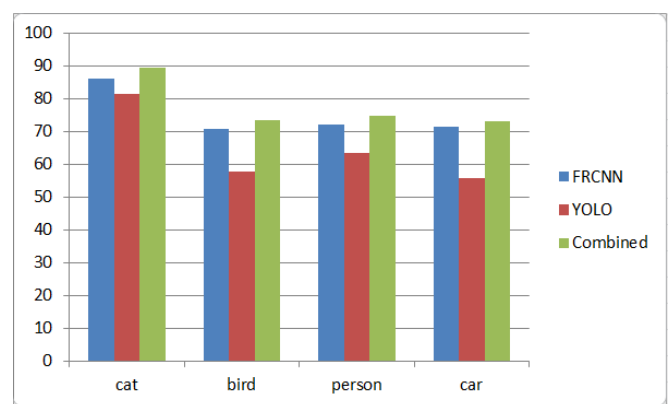


Figure 4. Comparison of the performance among Fast RCNN, YOLO and Fast RCNN+YOLO in COCO dataset.

CONCLUSION

With the help of this paper and based on the experimental results, we can recognize objects more precisely and identify the objects individually with the exact position of an object in the image on the x, y axis. This paper also contains experimental results on various methods for the detection and identification of objects and compares each method based on its efficiency.

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. Dunder, 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.