© March 2019 | IJIRT | Volume 5 Issue 10 | ISSN: 2349-6002

# Object Detection Using Deep Learning

Prof. Jogi John[1], Sagar Balpande[2], Prashul Jain[3], Anshul Chatterjee[4], Rohit Gupta[5], Samiksha Raut[6]
[1]*Professor, Dept. of Computer Technology Priyadarshini College of Engineering, Nagpur, India*
[2,3,4,5,6]*Dept. of Computer Technology Priyadarshini College of Engineering, Nagpur, India*

***Abstract*— This paper deals with the field of computer vision, mainly for the application of deep learning in object detection task. On the one hand, there is a simple summary of the datasets and deep learning algorithms commonly used in computer vision. On the other hand, a new dataset is built according to those commonly used datasets, and choose one of the networks called faster r-cnn to work on this new dataset. Through the experiment to strengthen the understanding of these networks, and through the analysis of the results learn the importance of deep learning technology, and the importance of the dataset for deep learning.**

***Index Terms*— deep learning, faster r-cnn, object detection, convnet, neural network**

## I. INTRODUCTION

In recent years, with the rapid development of deep learning, a number of research areas have achieved good results, and accompanied by the continuous improvement of convolution neural networks, computer vision has arrived at a new peak. In addition, the return of the convolution neural network also makes the application of computer vision greatly improve, such as face recognition, object detection, object tracking, semantic segmentation, and so on.

Object detection – technology in the field of computer vision for finding and identifying objects in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes and scales or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. This task is still a challenge for computer vision systems. Deep learning has formed a mainstream object recognition algorithm based on RCNN, and these algorithms is refreshing the higher accuracy in a number of famous datasets.

In this paper, we first summarize some algorithms related to deep learning for object detection, and then apply one of the algorithms to a new dataset to verify its wide applicability.

## II. DATASETS AND NEURAL NETWORK

For deep learning, dataset and neural network are two important parts. The dataset is the fuel for deep learning so that the number and quality of the dataset will affect the accuracy of the neural network output, and the choice of neural network or the network architecture will also affect the accuracy.

### *A. Dataset*

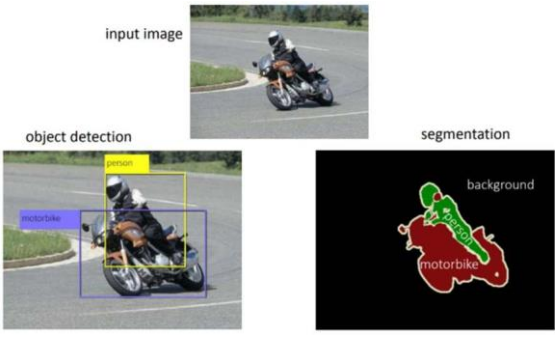
Dataset is one of the foundations of deep learning, for many researchers to get enough data to carry out the experiment just by themselves is a big problem, so we need a lot of open source dataset for everyone to use. Some commonly used datasets in computer vision is the following.

#### *1) ImageNet*

ImageNet is an image dataset organized according to the WordNet hierarchy. Each meaningful concept in WordNet, possibly described by multiple words or word phrases, is called a "synonym set" or "synset". There are more than 100,000 synsets in WordNet, majority of them are nouns (80,000+). In ImageNet, we aim to provide on average 1000 images to illustrate each synset. Images of each concept are quality-controlled and human-annotated. In its completion, we hope ImageNet will offer tens of millions of cleanly sorted images for most of the concepts in the WordNet hierarchy. It is very widely used in the field of computer vision research, and has become the "standard" dataset of the current deep learning of image domain to test algorithm performance. There is a well-known challenge called "ImageNet International Computer Vision Challenge" (ILSVRC) based on the Imagenet dataset.

#### *2) PASCAL VOC*

The PASCAL VOC (pattern analysis, statistical modelling and computational learning visual object classes) provides standardized image data sets for object class recognition and provides a common set of tools for accessing the data sets and annotations. The PASCAL VOC dataset



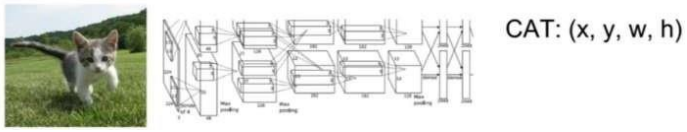
includes 20 classes and has a challenge based on this dataset. The PASCAL VOC Challenge is no longer available after 2012, but its dataset is of good quality and well- marked, and enables evaluation and comparison of different methods. And because the amount of data of the PASCAL VOC dataset is small, compared to the imagenet dataset, very suitable for researchers to test network programs.

*3) COCO*

COCO (Common Objects in Context) [10] is a new image recognition, segmentation, and captioning dataset, sponsored by Microsoft, Facebook. COCO dataset has more than 300,000 images covering 80 object categories. The open source of this dataset makes great progress in semantic segmentation in recent years, and it has become a "standard" dataset for the performance of image semantic understanding, and also COCO has its own challenge.

*B. Neural Network*

Deep learning used by the network has been constantly improving, in addition to the changes in the network structure, the more is to do some tune based on the original network or apply some trick to make the network performance to enhance. The more well-known algorithms of object detection are a series of algorithms based on R-CNN, mainly in the following.




*1) R-CNN*

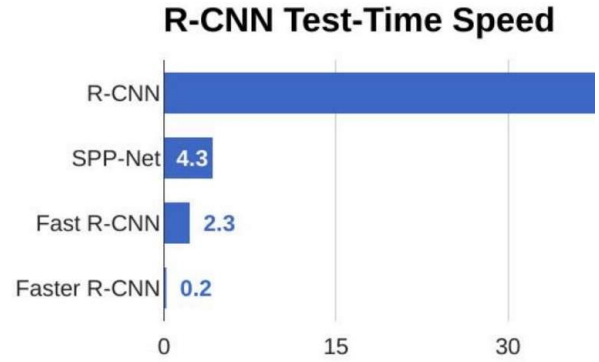
Paper which the R-CNN (Regions with Convolutional Neural Network) is in has been the state-of-art papers in field of object detection in 2014 years. The idea of this paper has changed the general idea of object detection. Later, algorithms in many literatures on deep learning of object detection basically inherited this idea which is the core algorithm for object detection with deep learning. One of the most noteworthy points of this paper is that the CNN is applied to the candidate box to extract the feature vector, and the second is to propose a way to effectively train large CNNs. It is supervised pre-training on large dataset such ILSVRC, and then do some fine-tuning training in a specific range on a small dataset such PASCAL.

*2) SPP-Net*

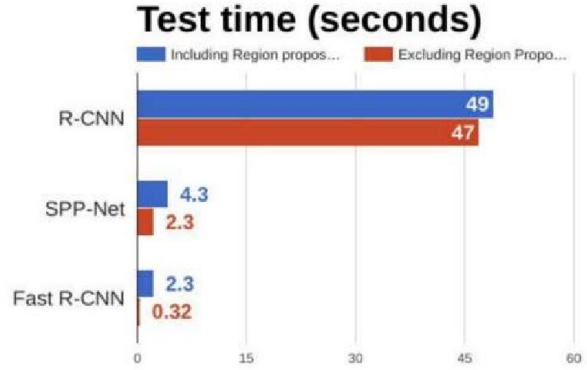
SPP-Net [11] is an improvement based on the R-CNN with faster speed. SPP-Net proposed a spatial pyramid pooling (SPP) layer that removes restrictions on network fixed size. SPP-Net only needs to run the convolution layer once (the whole image, regardless of size), and then use the SPP layer to extract features, compared to the R-CNN, to avoid repeat convolution operation the candidate area, reducing the number of convolution times. The speed for SPP-Net calculating the convolution on the Pascal VOC 2007 dataset by 30-170 times faster than the R-CNN, and the overall speed is 24-64 times faster than the R-CNN.

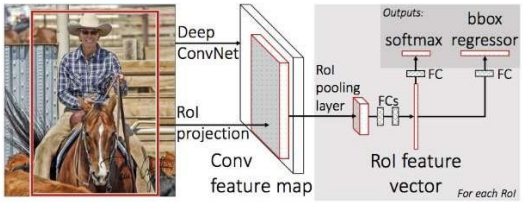
*3) Fast R-CNN*

For the shortcomings of R-CNN and SPP- Net, Fast R-CNN did the following improvements: higher detection quality (mAP) than R-CNN and SPP-Net; write the loss function of multiple tasks together to achieve single-level training process; in the training can update all the layers; do not need to store features in the disk. Fast R-CNN can improve the speed of training deeper neural networks, such as VGG16. Compared to R-CNN, the speed for Fast R- CNN training stage is 9 times faster and the speed for test is 213 times faster. The speed for Fast R- CNN training stage is 3 times faster than SPP-net and the speed for test is 10 times faster, the accuracy rate also has a certain increase.



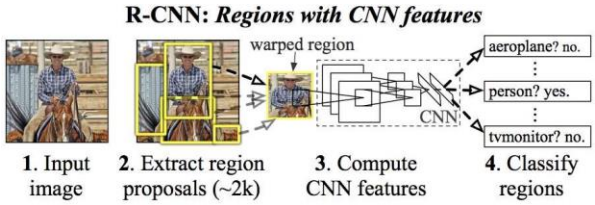







*4) YOLO -*

How YOLO works is that we take 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 outputs a class probability and offset values for the bounding box. The bounding boxes having the class probability above a threshold value is selected and used to locate the object within the image.




Comparison of different algorithms

## III. APPLICATION OF FASTER R-CNN ON NEW DATASET

In the experiment, we must need a new dataset, the dataset format as seem as VOC data set format. We have created a football game image dataset, which has the four categories of objects that is player, football, soccer goal, corner flag. As shown in Figure.

Fig 1. Marked Image

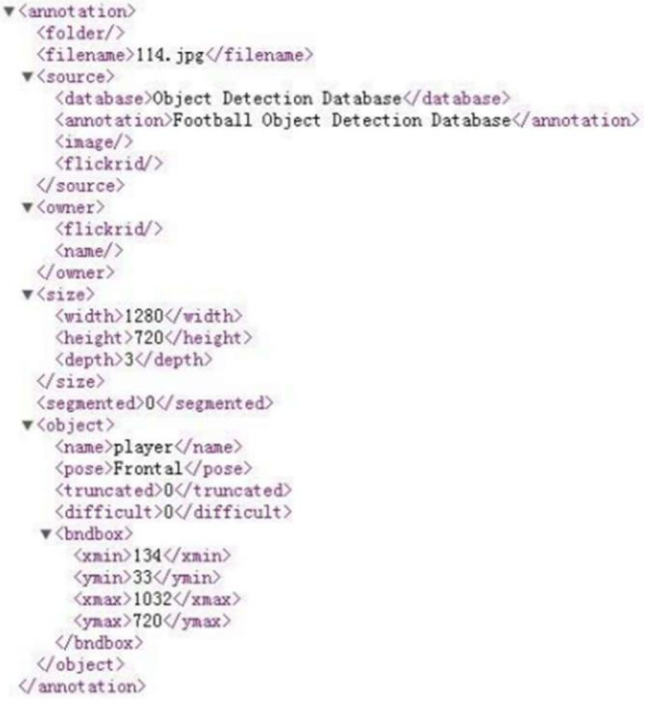
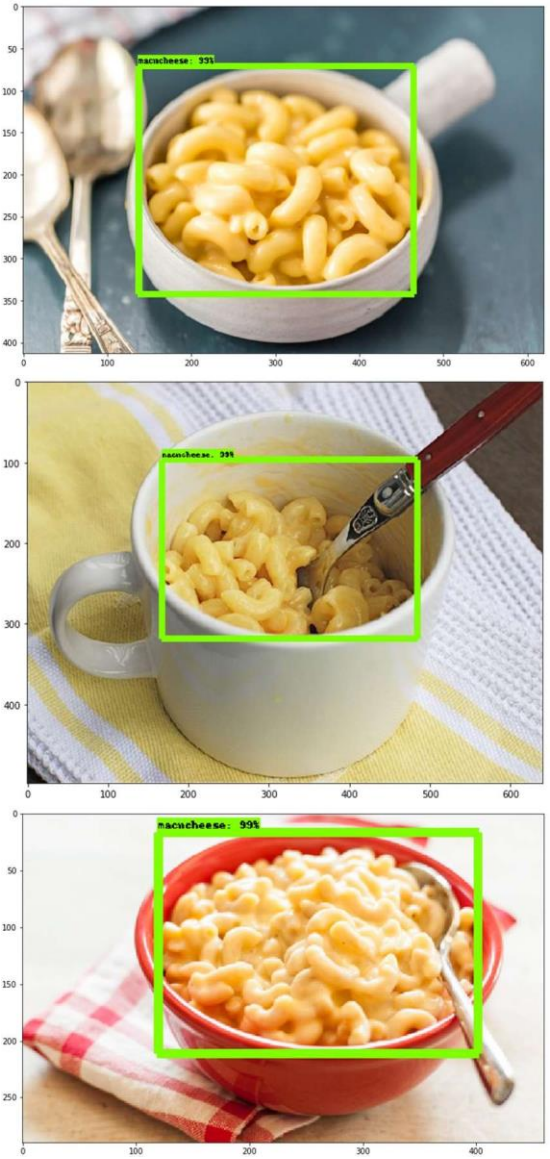



Fig 2 Labelled Image Information

From the label information as show in Figure 3 can be seen some other information, including the name of the image and the name of the dataset. The size and depth of the image are recorded under the ‘size’ tab. The information under the ‘object’ tab is the content we marked before.

## IV. RESULTS

After training, some effect on the test set is shown below




## V. CONCLUSION

This paper expresses the importance of deep learning technology applications and the impact of dataset for deep learning through the use of the faster r-cnn on new datasets. In recent years, the technology of deep learning in image classification, object detection and face identification and many other computer vision tasks have achieved great success.

Experimental data shows that the technology of deep learning is an effective tool to pass the man-made feature relying on the drive of experience to the learning relying on the drive of data. Large data is the base of the success of deep learning, large data just as fuel to the rocket for deep learning. More and more applications are continually accumulating increasingly rich application data, which is critical to the further development and application of deep learning. However, the quality of the data affects the deep learning in deed, of course, in addition to these real data, maybe we can also consider some of synthetic data to increase the amount of data in the further.

## ACKNOWLEDGMENT

We would like to thank the Priyadarshini College of Engineering (Computer Technology Department) for providing the resources and guidance.

Special thanks to Prof. Jogi John, CT dept. Priyadarshini College of Engineering, Nagpur for guiding and providing the solution and appropriate references.

## REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS,2012.

[2] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. In ECCV,

[3] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[4] SENtDEX - https://pythonprogramming.net

[5] Data Towards Sciences -https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

[5] Kaggle - https://www.kaggle.com/

[6] https://arxiv.org/abs/1505.04597

[7] M. Betke, N. Makris, "Fast object recognition in noisy images using simulated annealing", *Proceedings of the Fifth International Conference on Computer Vision*, pp. 523-520, 1995.

[8] B. Moghaddam, A. Pentland, *Probabilistic visual learning for object detection*, 1995.

*[9] "ImageNet Summary and Statistics". ImageNet. Retrieved 22 June 2016.*