

# Object Detection from Images Using Deep Learning

Saaod M .Rasheed<sup>1</sup>, Ahmet cinar<sup>2</sup>

Student, Dept. of Computer Engineering, Firat University, Elazig, Turkey<sup>1</sup>

Associate Professor, Dr, Dept. of Computer Engineering, Firat University, Elazig, Turkey<sup>2</sup>

**ABSTRACT:** Object detection is the process of classifying and locating variant objects in images. Object detection is a challenging problem that computer vision and machine learning involved. by applying various deep learning methods, computer vision and machine learning achieved the most impressive results, Based on new examination of discoveries on new researches, the technique that is fit for our works is Faster Region Convolution Neural Network (Faster R-CNN) that we tend to utilized it to classify numerous varieties of dogs in image. Faster R-CNN for object detection and classifying beats several different ways particularly for the detection purpose. We chose faster R-CNN method because it's accuracy and speeds. We present a Faster R-CNN trained to classify and detect Dogs from image.

**KEYWORDS:** Object detection, Deep learning, Computer Vision, machine learning, Faster R-CNN.

## I. INTRODUCTION

Nowadays the world become smart and smarter and the technique of image processing is delivered from all over the world such as camera surveillance, image texture, classification detection and classification.

Object detection is that the most crucial and particular part in image processing, the benefits of this technique are that we can use it for many purposes to do our works, so that it become easier and faster than human. object detection [1][2] is one of the most difficult task in computer vision and has attracted plenty of attentions all the time. In this paper we only care about detecting objects [2][3]in imagespecially in deep learning technique and used it for our study to detect dogs in an image because there are10 classes of dogs for that we need a method which an achieved a high result in object detection and deep learning is so suitable to do our work.

Deep learning methods are widely utilized in numerous issue areas, including object detection. Deep learning models achieved better results when trained on more data. It has ability to localize and detect objects in images. Deep learning techniques in computer vision have reached good results (Krizhevsky et al., 2012[2]) in speech recognition. The most important field used very widely in deep learning is convolution neural network [5][6] for image detection, recognition, localization and segmentation. Convolutional networks have a great role in the history of deep learning [4]. Convolutional networks were additionally a portion of the first neural networks to solve the most crucial applications in commercial area and stay at the highest of commercial area applications of deep learning. In 2013, Girshick et al. published a new why [2] to achieve good results in object detection. This why is called R-CNN ("CNN with region proposals"). Extract features are used from each region proposal in convolutional network. R-CNN is a crucial method that provided practical solution for using object detection, but there were some problems for training, this method need multiple stages and also training for SVM and region proposal is expensive, because it need a lot of locations in disc, also another problem for detecting objects was that about a minute was required.

in 2015 Faster R-CNN [8] published by Ren et al. is an integrated method. The thought is to utilize shared convolutional layers for region proposal creation and for classifying. The creators found that feature maps created by detection networks can also be utilized to create the region proposals. Explained in IV also it is need to shorten the time

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

spent during the test time on region proposal step. so the Faster R-CNN authors of came with the idea that can be combined region proposal step with Fast R-CNN[7], so that it can take advantage of GPUs as well as sharing computations. two modules come out by Faster R-CNN. The first one is Region Proposal Network (RPN), which share convolutional layers with the second module, Fast R-CNN. So we find out the faster R-CNN its better than other for dog detection.

## II. RELATED WORK

Object detection is an active point form of research and development for over a decade. A major drawback of traditional computer vision approaches was that the features for constructing object classifiers and detectors were hand-designed. some of the new work in object detection and recognition revolves around the need to speed up detection phase of recognition pipeline A different closer work to ours, depends on thought that objects must be localized without having to find their category a particular of these methodology generate on bottom-up classless segmentation. The task of object localization and detection has been largely addressed by using CNNs and formulating the detection problem as that of classification of a smaller region, thus obtaining a reduced spatial map of classifications. One of the most advancement in this area was made by introducing R-CNN type of networks, in three iterations R-CNN [1], Fast R-CNN [2][7] and Faster R-CNN [3]. We can say ultra-modern techniques for classifying and detecting items of general categories are mainly dependent on deep CNNs. Regions with Convolutional Neural Networks (R-CNN) presented Girshick et al. [3] in multi-stage for the classification of region proposals to find objects explains the detection issue into a few steps such as CNN pre-training, bounding-box proposal, SVM training, CNN fine-tuning, and bounding box regression. It accomplishes great results however the pipeline is less effect since features of each region proposal necessary be computed over and again. In SPP-net [9], this issue has been self-addressed by presenting a pooling method to compute the feature map just once and create features in arbitrary regions. Faster R-CNN [8] joins a CNN classifier the region proposal and within a single network by presenting a Region Proposal Network .it is one of the most crucial methods and widely utilized methods for object detection, which have achieved excellent results in object detection like Microsoft COCO [10] and PASCAL VOC [11]. It consists of 2 modules: a detector module and a Region Proposal Network (RPN). The RPN in component in Faster R-CNN is a fully convolutional neural network which generates object proposals. Which is that predicting bounding boxes with respect to reference boxes of multiple sizes. Another crucial thing we used convolutional network frameworks Caffe. While Caffe is most suitable choice mostly because original Faster R-CNN was implemented using Caffe, in our paper. For object classification There are several case of work on the general issue. For instance, the Pascal Visual Objects Challenge (VOC) that use the VOC dataset it consists 20 different classes also faster R-CNN [3] trained on it. we explain some object detection methods which is based to CNN.

## III. OBJECT DETECTION

Computer vision involved in many problems object detection is one of that problem and also known as one of the difficult tasks. Both of locating and classifying regions of an image makes a problem to detect objects. The object detection objective is to detect all instances of objects in a known class, for example cars or faces in an image. Typically, there are just a small number of object instances in the image, but there are a very large number of possible locations and scales where they can occur and need to be explored in some way. To detect an object, we need to know some ideas about where the object might be or how the image is segmented. The location of an object is needed to recognize its shape, and the shape of an object is needed to recognize its location.

## IV. FASTER R-CNN

Faster R-CNN [8] attempts to struggle the complex training that both R-CNN and Fast R-CNN revealed and it is an integrated method [4]. There are two modules in Faster R-CNN. fully convolutional network is first one that proposes regions, and Fast R-CNN detector [7] is the second one, which uses the proposed regions. The whole network is a single, unified network for object detection. The goad is using the final convolutional layer to infer region proposal

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

network (RPN) for detection. This network can simply take focus at the final convolutional feature map and region proposals generate from that. It can show the object's exact location so that is a good knowledge for the computer. Today for object detection Faster R-CNN has become the state-of-art. The authors discovered by object detection networks that feature maps generated, also it can be used to create the region proposals. feature proposals generated by fully convolutional part of the Faster R-CNN network also is called a region proposal network (RPN) [8] Region proposal network that shares full convolutional features with the detection network and it is trained to produce best quality region proposals, which are utilized by Fast R-CNN architecture for detection network. region proposals can be generating by convolutional feature map when slide a small network over the result by the final shared convolutional layer. A Faster R-CNN network is trained by alternating between training for detection and RoI generation. two separate networks first trained. After that, these networks are joined and fine-tuned. During fine-tuning, certain layers are trained in turn and kept fixed. Only as input single image receive by trained network. feature maps will generate by The shared fully convolutional from the image. when feature maps are generated it used to input to the RPN. The RPN results region proposals, which are input, together with the said feature maps, to the last detection layers. These layers consist output of final classifications and a RoI pooling layer. By using shared convolutional layers, region proposals are computationally nearly cost less. Processing the region proposals on a CNN has the additional useful of being feasible on a GPU.

The below steps are typically followed in a Faster RCNN approach:

- 1- First taking an image for the input and pass it to the ConvNet that returns that image's feature map.
- 2- These feature maps are applied to the region proposal network. This returns the proposals for objects together with their score for objects.
- 3- On these proposals, a RoI pooling layer is applied to bring all the proposals down to the same size.
- 4- Finally, the proposals are passed on to a fully connected layer with a softmax layer at the top and a linear regression layer to classify and output the object bounding boxes.

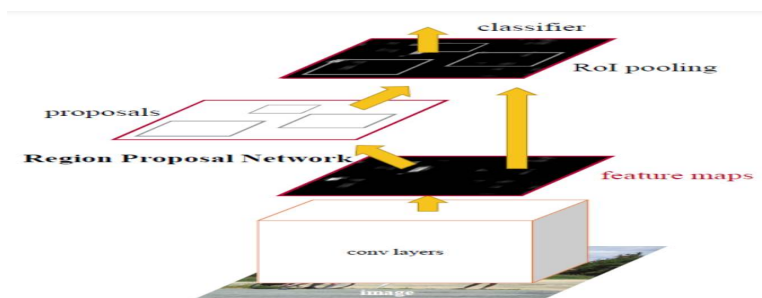


Fig: faster R-CNN

## V. TRAINING OF FASTER R-CNN

For training faster R-CNN depends on the arrangement in the network itself and that make four-step training proposal process. During the process sequence followed.

- 1- The first step to do training should be Region Proposal network(RPN). The RPN network is done by a pre-trained ImageNet and fine-tuned by generating region proposal
- 2- The generated proposal is used second stage for training is Fast R-CNN detector network which it created by the first training in RPN.
- 3- For the third stage, step 2 results used to utilized the RPN where only unique RPNs layers are fine-tuned.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

4- The lastly in fourth stage only fine-tuned last layers of faster R-CNN.

## VI. EXPERIMENTAL RESULTS

In the experiment, we chosen ~4k images form 10 difference classes and we chose ~2.5k images for training and another ~1.5k for testing to classifying dogs in image using Faster R-CNN methods, the categories is of Scottish deerhound, afghan hound, basset, Saluki, Samoyed, beagle, Staffordshire bullterrier, Yorkshire terrier, scotch terrier and Cocker spaniel. all images we used is obtained from Stanford dog dataset [12] and oxford animals pet dataset[13] which all images are resized 500 pixel or smaller than 500 pixels and those are power full dataset to our study. we labelled most of images as shown in the Fig.3. we use ZF fast version for the network ImageNet pre-trained which consistthree fully-connected layers and five convolutional layers, The value of the mean average precision across whole categories predictions is utilized as metric measure of execution. Furthermore, number of iterations and run time per image are noted for the objective of confirming the technique's execution. The mean Average scores in shown in table 1.

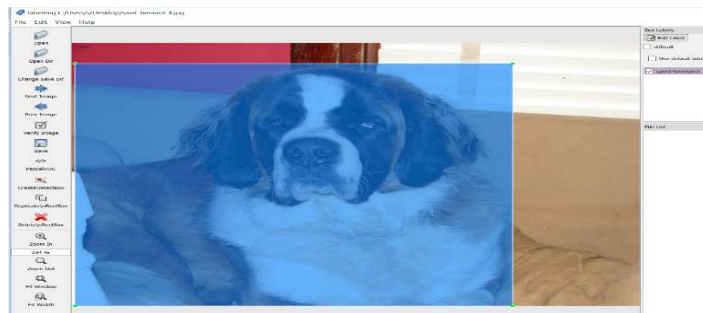


Fig 3: labelling image

We must labelling all images like above write name and selection its position we nearly labelled ~2k of images.

Table 1: Average precision and mean AP scores for each of 10 classes

Class	AP
Scottish Deerhound	<b>0.876</b>
Afghan Hound	<b>0.892</b>
Saluki	<b>0.837</b>
Samoyed	<b>0.728</b>
Basset	<b>0.799</b>
Beagle	<b>0.844</b>
Staffordshire Bullterrier	<b>0.816</b>
Yorkshire Terrier	<b>0.823</b>
Scotch Terrier	<b>0.523</b>
Cocker Spaniel	<b>0.893</b>
Mean AP	<b>0.803</b>

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

In here we show dog examples in our tests



Fig 4: Examples of Scottish deerhound, afghan hound, basset, Saluki, Samoyed, beagle, Staffordshire bullterrier, Yorkshire terrier, scotch terrier and Cocker spaniel

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

The mean AP is 0.803 its good achievement the highest one is Afghan Hound AP 0.892 it mean the quality of Afghan Hound's image is very good but in other hand the lowest one is Scotch Terrier AP 0.523 because of its bad quality images. In generally our test it achieved good result to detect all 10 categories dog it shown in fig 4 it consist all of 10 categories.

## VII. CONCLUSION

We explained briefly object detection and Faster R-CNN. so in our paper main in our paper main objective was successfully implemented to present the application of Faster R-CNN technique for the detection of 10 categories of difference dogs for the purpose of detecting dogs in an image and showing their class names. We also had a perfect result in the outputs average mean precision value can have achieved high result 0.803 in dog's detection during training and testing process.

## REFERENCES

- [1] Liu, Lei, Zongxu Pan, and Bin Lei. "Learning a Rotation Invariant Detector with Rotatable Bounding Box." arXiv preprint arXiv:1711.09405 (2017).
- [2] Szegedy, Christian, Alexander Toshev, and Dumitru Erhan. "Deep neural networks for object detection." In Advances in neural information processing systems, pp. 2553-2561. 2013.
- [3] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.
- [4] Stenroos, Olavi. "Object detection from images using convolutional neural networks." (2017).
- [5] Kang, Kai, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang. "Object detection from video tubelets with convolutional neural networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 817-825. 2016.
- [6] Kloss, Alina. "Object Detection Using Deep Learning-Learning where to search using visual attention." PhD diss., Universität Tübingen Tübingen, Germany, 2015.
- [7] Girshick, Ross. "Fast r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.
- [8] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems, pp. 91-99. 2015.
- [9] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In ECCV. 2014.
- [10] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740-755. Springer, Cham, 2014.
- [11] Everingham, Mark, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. "The pascal visual object classes (voc) challenge." International journal of computer vision 88, no. 2 (2010): 303-338. Muthukumar S, Dr. Krishnan .N, Pasupathi.P, Deepa. S, "Analysis of Image inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods", International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010
- [12] Khosla, Aditya, et al. "Novel dataset for fine-grained image categorization: Stanford dogs." Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC). Vol. 2. No. 1. 2011.
- [13] Parkhi, Omkar M., et al. "Cats and dogs." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.