[small object detection]

Report submitted in fulfillment of the requirements
for the Exploratory Project of

Second Year IDD.

by

[sahil manikchand chuahari],[manish regar]

Under the guidance of

[Dr.Tanima Dutta]

Department of Computer Science and Engineering
INDIAN INSTITUTE OF TECHNOLOGY (BHU) VARANASI
Varanasi 221005, India
May 2017

Declaration

I certify that

1. The work contained in this report is original and has been done by myself and the general supervision of my supervisor.
2. The work has not been submitted for any project.
3. Whenever I have used materials (data, theoretical analysis, results) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.
4. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Place: IIT (BHU) Varanasi
Date:

[**sahil manikchand chuahari**],[**manish regar**]
B.Tech. or IDD Student
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Varanasi, INDIA 221005.

Certificate

*This is to certify that the work contained in this report entitled "**[small object detection]**" being submitted by **[sahil manikchand chuahari],[manish regar]** (**Roll No. [18074014][18074009]**), carried out in the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, is a bona fide work of our supervision.*

	[**Dr. Tanima Dutta**]
Place: IIT (BHU) Varanasi	Department of Computer Science and Engineering,
Date:	Indian Institute of Technology (BHU) Varanasi,
	Varanasi, INDIA 221005.

Acknowledgments

we would like to express our sincere gratitude to parents and teachers.

Place: IIT (BHU) Varanasi

Date: **[sahil manikchand chuahari],[manish regar]**

Abstract

SMALL OBJECT DETECTION

As we know object detection is well studied topic , and many state of the art method such as fast rcnn [1] , faster rcnn [2] etc have given excellent accuracy over object detection and classification. But the accuracy of the model decreases as the object size with respect to image decreases.

So we implement two research paper based on small object detection. They are

1) CNN-based small object detection and visualization with feature activation mapping [3]

2) improved faster RCNN based on faster RCNN. [4]

We use spider dataset for the implementation of both papers and obtain accuracy comparable to the accuracy mention in papers. We implement both papers on keras.

Contents

List of Figures

Chapter 1

CNN-based small object detection and visualization with feature activation mapping:[2]

This classification model is based on VGG16[5], a CNN network.

An existing VGG16 model trained on 'imagenet' dataset[6] is fined tuned on spider dataset.

A new feature activation technique is used to generate the heat maps by using both high level and low-level feature maps.

In feature activation technique all feature maps of each pooling layer are summed up and upsampled to the resolution of input image, then all upsampled heat maps are summed to get overall heat map.

$$H = \sum_{l=1}^{L} H^l * R^l = \sum_{l=1}^{L} \left(\sum_{k=1}^{K} H^l_k \right) * R^l$$

[3]

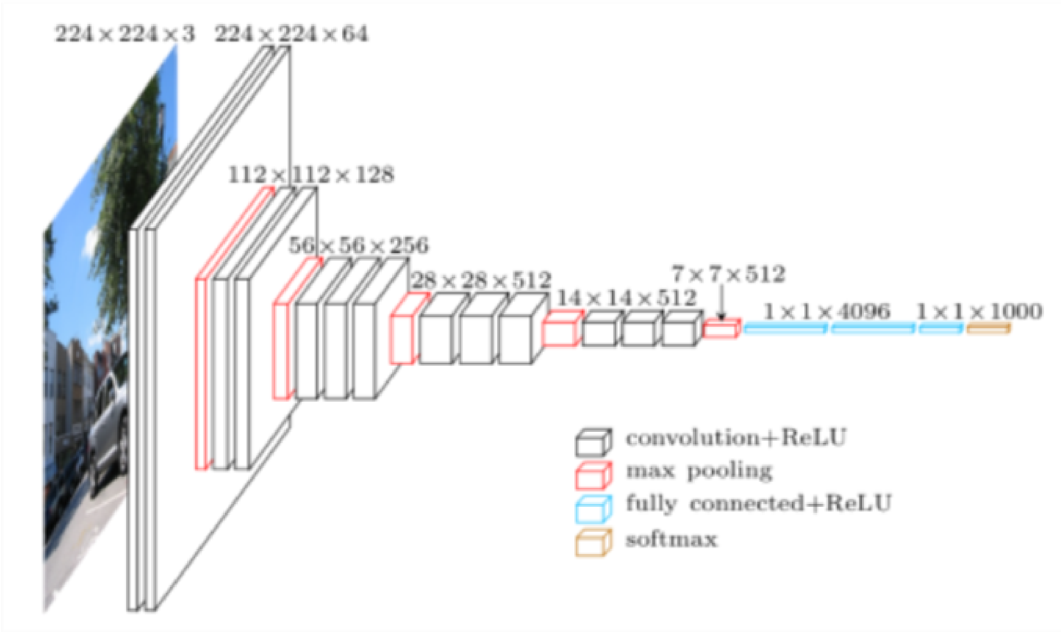
where

Hlk represents feature map of each pooling layer

Rl represents resolution factor

L represents layer

VGG16 model :



Architecture of VGG16

Figure 1.1 VGG16 model
[3],[5]

Modified VGG16 model with feature activation map

In above figure up1, up2, up3, up4, up5 gives heat maps of same resolution as input image

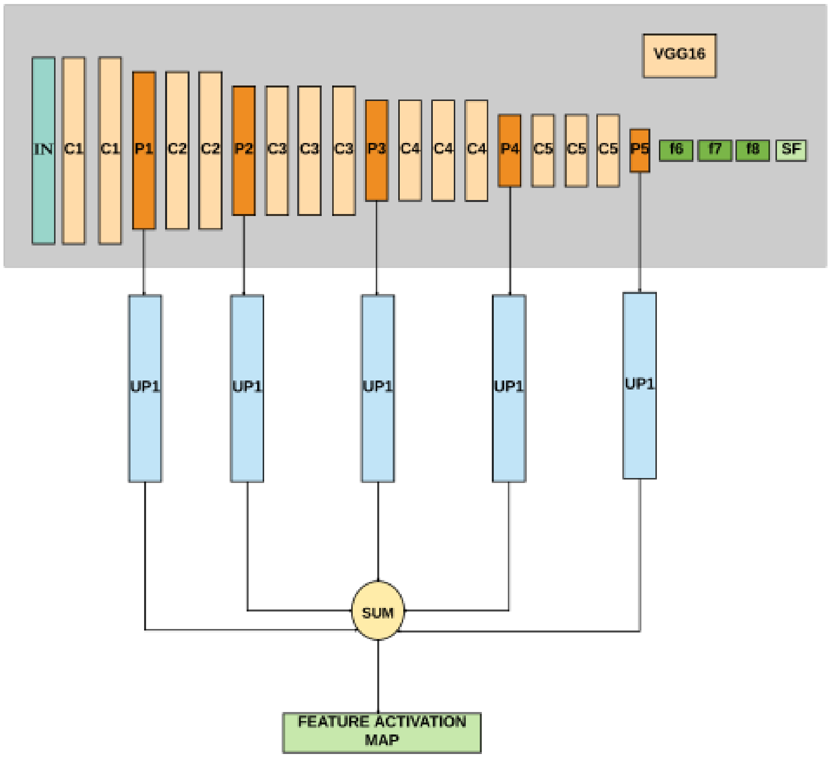


Figure 1.2 Feature activation map model [3]

As we the numbers of feature in convo layers decreases as we go down the VGG16 model. So we combine each pooling layer which is upsampled to input image size. The starting convolutional layer contain feature like boundary, edges of the object which helps in localization of the object in feature activation map.

TRANING:

We fine-tuned our model in two steps. First we fine-tuned our model with images of spider which include image with small scale spider and large scale spider. Second we fine-tuned with image of small scale spider only with image augmentation.

We used ADAM optimizer with sparse_categorical_crossentropy loss function

We trained our model in online platform Kaggle with learning rate of 1e-5

RESULTS:

Our loss decreased and accuracy increased from loss 0.3500 accuracy 0.8844 to loss 0.0193 accuracy 0.9956 in our first fine-tuning , in our small scale image accuracy approaches zero and accuracy to 1.

We evaluate our model and got loss 0.0348 and accuracy 0.9911.

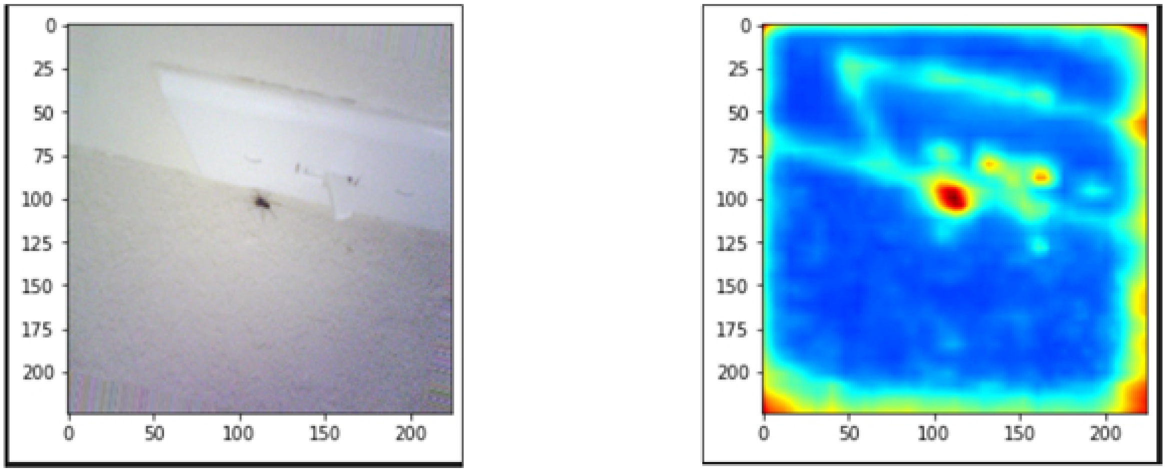


Fig3

ORIGINAL IMAGE **FEATURE ACTIVATION MAP**

Figure 1.3 Input and output images

Chapter 2

Improved Faster RCNN for small object detection

The following modification are made in the two stages of faster rcnn:

We take faster rcnn implemented code in keras and modify it given modification to make it improved faster rcnn

Improved loss function for bounding box regrsssion :

1. Improved IoU:

   Intersection over Union (IoU) is an important indicator in the system of object detection. The formula for Iou is

$$IoU = \frac{S_{\text{DetectionResult}} \cap S_{\text{GroundTruth}}}{S_{\text{DetectionResult}} \cup S_{\text{GroundTruth}}}$$

[4]

   improved the IoU, denoted as IIoU (Improved IoU):

here C is minimum rectangle containing A and B

$$\mathrm{IIoU} = \mathrm{IoU} - \frac{C - (A \cup B)}{C}$$

[4]

2. Loss function:

The regression loss function of bounding box is :

$$L_{IIoU} = 1 - IIoU$$

[4]

Where,

$$IIoU = IoU - \frac{S^C - \left(S + S^* - S^I\right)}{S^C}$$

[4]

$$IoU = \frac{S^I}{S + S^* - S^I}$$

[4]

Where S= area of detection box

S∗ =area of ground truth box

Sl=area overlapped by S and S∗

Sc=minimum closure area of S and S∗

LILoU is non negative lies between 0 and 2

3. Bilinear Interpolation[4]:

The normal RoI pooling layer causes the deviation of the feature image. The feature image mapped to the original image is larger so the box losses accuracy.

Bilinear Interpolation is used to solve this problem. The below fig shows how to solve it

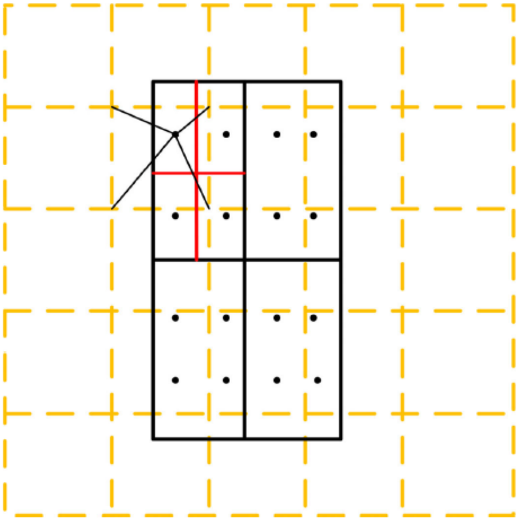


Figure 2.1 Bilinear Interpolation
[4]

The dotted line in Fig indicates the feature image, the intersection point of dotted line is the pixel point, and the black solid line indicates the region proposal. The region proposal is divided into 2× 2 bins, the sampling point of each bin is set to 4, and the bin is evenly divided into 4 small areas, as shown by the red line, the center point of each small area is the sampling point, but the coordinates of the sampling point are usually decimal, so it is needed Bilinear interpolation of the pixel values of the sampling points, as shown by the four arrows, after obtaining the pixel values of the point, and then perform the maximum pooling operation on the four sampling points of each bin[2]. We implement this using **tensorflow crop and resize**

2. In recognition stage:

1. Convolution feature fusion[4]:

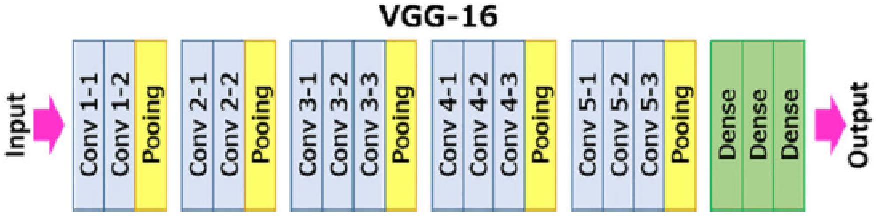


Figure 2.2 VGG16 model
[4],[5]

In faster rcnn uses only Conv5_3 layer's output as feature map for subsequent network. Therefore, they combine the features extracted by Conv3_3, Conv4_3 and Conv5_3 convolutional layers according to the addition of elements.

the size of feature map of Conv4_3 unchanged, the max pooling by subsampling is used for Conv3_3 and upsampling is used for Conv5_3 to make them consistent with Conv4_3.

the merged feature map can be obtained by adding the feature map.

Before the fusion of the three-layer convolutional feature maps, we first use local response normalization (LRN) to process each feature map so that the activated values of the feature map are the same.

b. Improved NMS[4]:

The soft-NMS algorithm is used instead of NMS algorithm.

If a bounding box overlaps with M most, the bounding box will get a low score. If the degree of overlap is low, the score is unchanged.

NMS algorithm:

$$s_i = \begin{cases} s_i & IoU(M, B_i) < p \\ 0 & IoU(M, B_i) \geq p \end{cases}$$

[4]

Soft NMS algorithm defined as:

$$s_i = \begin{cases} s_i & IoU\ (M, B_i) < p \\ s_i \times (1 - IoU\ (M, B_i)) & IoU\ (M, B_i) \geq p \end{cases}$$

[4]

Where p is threshold of IoU. then the lower scores of boxes are removed. At the same time, as the bounding box with the highest score is M, if there are bounding boxes with high scores or the IoU with M are greater than 0.9, make them recombined. The positions of the recombined bounding boxes are weighted and averaged by the corresponding score weights to the original coordinates of bounding box, and the combined scores of bounding boxes are set to the average value

The overall framework:

Overall algorithm:

A. Convolution feature fusion is done in red dotted Box 1 of above figure and then output feature map is sent to the subsequent model

B. The feature map is merged into the RPN network to generate region proposals, The LIIoU is used as the loss function for bounding box regression. This loss function

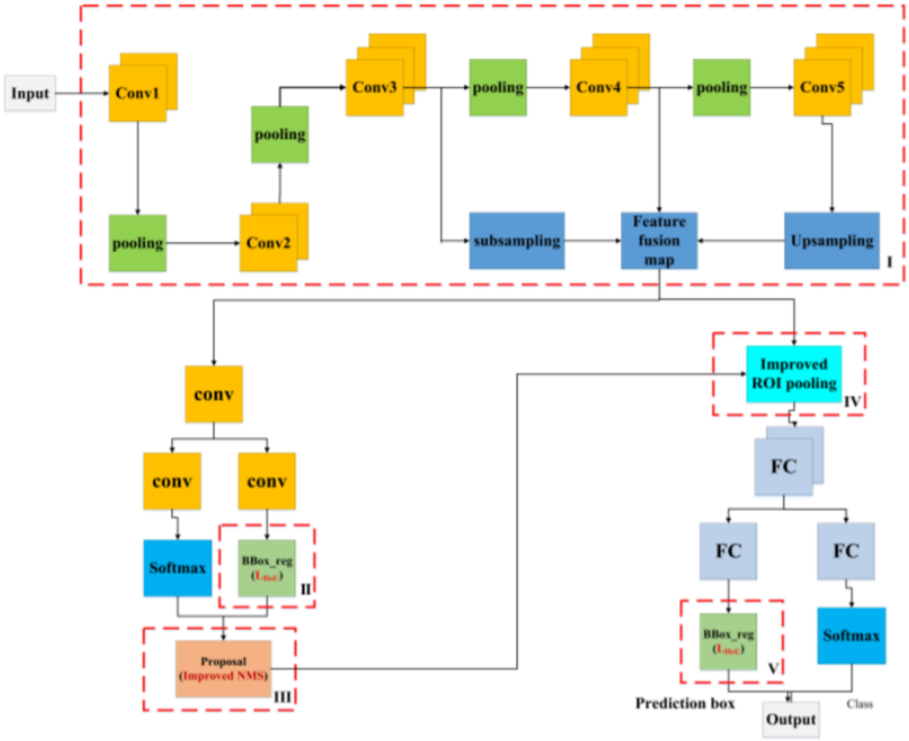


Figure 2.3 improved faster rcnn model [4]

is applied to the portion by the red dashed boxes II and V

C. The soft NMS algorithm is applied in Box III

D. Bilinear interpolation is done BOX IV. Finally, through the full connection layer into the subsequent bounding box regression, the border trimming and specific category classification by softmax are performed

RESULTS:

We train our this model on spider dataset , but as we didn't have the bounding box data we manually made bounding box using online platform LABELBOX and fine tuned with learning rate of 1e-5 on ADAM optimizer.

We got accuracy 0.85 on spider data set.

We have four losses rpn classification loss, rpn regression loss, class loss classification, class regression loss.

All four losses decreased to zeros in training, we trained our model on Kaggle CPU.

OUTPUT IMAGE

Chapter 3

Conclusions and Discussion

In this report we have implemented two papers based on small object detection This report gives us the importance of small object detection

- we have decreased our losse to nearly zero in both the model
- we got good accuracy in test
- this models will help in autonomous driving which requires detection of small traffic signs
- this may be useful in agriculture sector in detecting small insects.

Bibliography

[1] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.

[3] M. Menikdiwela, C. Nguyen, H. Li, and M. Shaw, “Cnn-based small object detection and visualization with feature activation mapping,” in *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2017, pp. 1–5.

[4] C. Cao, B. Wang, and W. Zhang, “An improved faster r-cnn for small object detection,” *IEEE Access*, vol. 7, pp. 106 838–106 846, 2019.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.