The CNN and DPM based approach for multiple object detection in images

[1]Dr.S. Rabindranath , *, [2] N. Mrs. Suma Manjunath, [3] Ms. S. Kamala Harsha
[1]Associate Professor ,[23]Assistant Professor
[12 3]Department of CSE,
[123] AMC Engineering College, Banglore

*Corresponding Author
E-mail: Rabindranath.surendran@amceducation.in

***Abstract*— With the development of intelligent device and social media, the bulk of data on Internet has grown with high speed. There are so many important aspect in image processing, object detection is one of the international demanded research field. Multiple object detection is an important concept in object detection. In object detection extracting the features and handling the occlusion are two most important components. A Region-based Convolution Neural Network (R-CNN) has achieved great success in extracting the region based features which may used for extracting multiple regions from the images and Deformable Part Based Model (DPM) improve the ability for handling the occlusion. Occlusion handling is nothing but when multiple objects are near to each other that time some objects are not detected so this drawback will be handled by DPM. Existing method not performing well in the aspect of detecting multiple objects. In this paper R-CNN and DPM are to be integrated to detect multiple objects. By combining these two models we are able to notice every single object with high accuracy.**

Keywords— ***R-CNN, DPM, Multiple object detection.***

I. Introduction

Now a days the deep learning has achieved good results in number of research areas and companions by the continual improvement of convolution neural networks. Object detection is an important application in convolution neural network. The CNN has build valuable progress in object detection. In the actual application it is very challenging task to use computer technology to detect the objects. Complex background, noise disturbance, occlusion, low resolution and other factors will seriously affect the object detection performance. Feature extraction is very difficult in object detection if size of the object is very small than alternative object then this types of objects are excluded throughout the detection method. The R-CNN model is deep learning model for object detection. Here we are going to apply the Region based model for detecting multiple objects but in R-CNN Selective Search Algorithm (SSA) not performing well when there are several objects which is close to one another. Although the deformable part based model will be able to find every single object that is partially occluded. In this paper we take the benefits of R-CNN and DPM and we are going to develop a new system framework that integrates R-CNN with DPM. The DPM can almost cover the entire object with high resolution that covers smaller parts of the object. DPM can rectify the capability for handling the occlusion and R-CNN will not only used for extracting the features but also handling the incorrect bounding boxes that may generated by DPM.

The contribution of this letter is as follows. First to propose a framework which can merge the R-CNN and DPM for object detection. The proposals which are achieved by DPM are cleaned by using filter based Density Subgraph Algorithm (DSD). In this system framework for detecting each of objects and guess all the available proposals of objects initially apply the DPM. After this the proposals that are generated by SSA and proposals generated by DPM are one by one send to the CNN model for extracting features. These features are used to detect object. Using PASCAL VOC 2007 dataset we calculate the performance of our framework.

II. Literature Rview

P. Felzenszwalb [1] presents paper of an object detection system that will supported merging of multiscale DPM. In the PASCAL object detection this system represents the deep object categories and achieves the progressive decisions. Although the deformable part based models have become quite common, on the difficult benchmark their value had not been demonstrated like PASCAL datasets. The discriminative training system is depends on new methods for partially labeled information. For data-mining examples the author combines margin sensitive approach called SVM.

R. Girshick, [2] presents paper in which the PASCAL VOC dataset is used to measure performance of the object detection. In multiple low level features there are some complex methods which are best for combining the images of features with high-level context. Author also proposes one algorithm for object detection which is simple and scalable and also enhance mean average precision (mAP) beyond 30% corresponding to the earlier finest result on VOC 2012 achieving a mAP of 53.3%.

O. Barinova [3] proposes, for multiple object detection the system uses nonmaxima suppression to locate and differentiate spike in Hough images. When objects are located near to one another then the postprocessing needs standardization of many parameters. For object detection author proposes a new probabilistic framework and that is said to be Hough transform. It distributes the wide relevance and simplicity of the Hough transform however, at the constant time, permits detection of multiple objects while not invoking heuristics of Nonmaximum Suppression Algorithm (NSA) and ignore the multiple spike identification problem in Hough images.

B. Wu [4] proposes a new technique that handles the objects that are near to one another in images. The detection and bisection tasks are developed as binary classification problem. The several parts of detectors and complete object segmentor are examine by boosting local shape feature primarily based on weak classifiers.

R. Girshick [5] represents one detection algorithm that is clean and flexible which can improves mAP by also 50% corresponds to the current finest result on VOC 2012 manage a mAP of 62.4%. This System integrates two ideas: For bounded and bisected object it can apply high Capacity Convolutional Networks (CNNs) to downside-up part proposals. The performance boost significantly when labeled training information data are supervised pre-training for an auxiliary and sporadic task is followed by domain-specific fine-tuning. With CNNs the system will combine region proposals then system call the resulting model as Region-based convolutional network port vector machine.

III. IMPLEMENTATION METHEDOLOGY

In this framework the DPM and R-CNN for detecting multiple objects are integrated. After taking image as input SSA rule and DPM is applied. Our framework will be split into three parts: first part is generating the proposals using DPM and the SSA, second part is extracting the features using CNN model and the third is filtering the proposals using DSD algorithm.

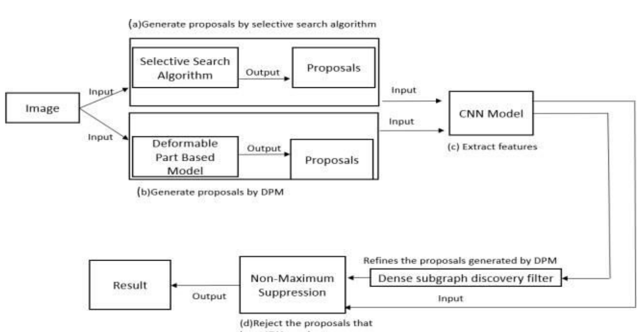

Fig1. Multiple object detection with DPM and R-CNN

Fig shows that the proposals that are generated from SSA and the proposals that are generated from DPM will be separately send to CNN model and then CNN model can extract the features from these proposals. The proposals generated by DPM will be clarify by using DSD. Between the clarified proposals and the obtained proposals by SSA rule, the ones whose count is more than the predefined threshold limit are selected. Using greedy NSA proposals which are selected are filtered which will removes proposals that have large IoU overlaps. In final result every single object is accurately detected.

A. *PROPOSAL GENERATION*

DPM is used to extract the features from whole object even the object is near to another object. In deformable part based model we tend to save every detected window as generated proposals while not reduce their expected bounding boxes.

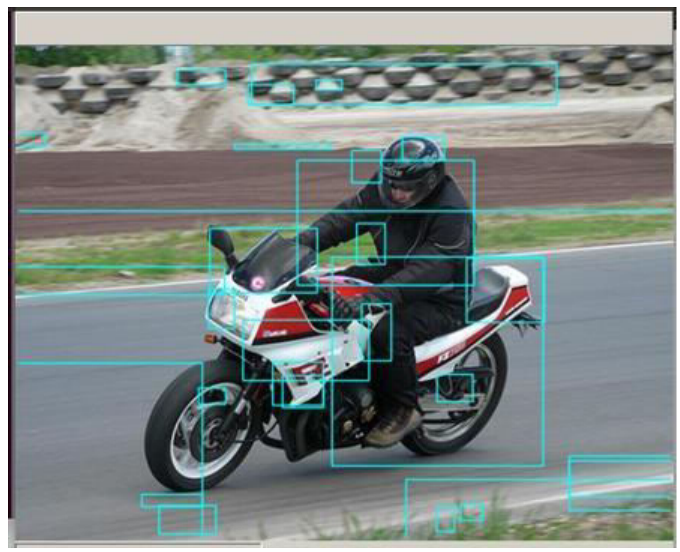

Fig2. Proposals generated by SSA

B. *FEATURE EXTRACTION*

In general the feature extraction is used for reducing the amount of resources which is required for large dataset. In our framework the CNN model will extract the features from the different proposals generated by SSA rule and deformable part based model then CNN outputs the feature and provides the score for every within the SVM.

C. *DENSITY SUBGRAPH ALGORITHM FILTER*

The generated proposals from DPM contain the limited things of an object or the full object. There are various types of cases in DPM. If the proposals are filtered by only NSA then in the DPM there are only few cases that the two bounding boxes have the identical object. It is because of the actual fact that NSA keeps biggest analysis one in all of proposals of identical objects. Also the IOU overlaps with biggest evaluation proposals those having beyond the typical threshold limit are get rejected. The proposals are kept who don't having IOU overlaps with the most important marking and includes only portion of the object.

IV. EXPERIMENTAL RESULTS

We evaluate our system based on PASCAL VOC 2007 dataset and in that dataset there are thousands of images of real world scene. For distinct object classes the dataset define ground truth bounding boxes. On test time the goal is to find the bounding boxes of all objects for the given image. The system will output set of bounding boxes here with corresponding scores and get a precision-recall curve across all images in the test set at different points. Threshold will be depending on metric, it is usually 50%, 75% or 95%. Precision is the ability of a model to identify only the relevant objects it is the percentage of correct positive predictions and is given by:

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{TP}{\text{all detection}}$$

Recall is the ability of a model to find all ground truth boxes. It is percentage of true positive detected among all relevant ground truths and it is given by:

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{TP}{\text{all ground truths}}$$

The performance metric will be calculated using IOU. IOU can evaluate the overlap between two bounding boxes. It requires predicted bounding box and ground truth bounding box. If overlap of ground truth bounding box is more than 50% then the predicted bounding box will be considered correct or else the bounding box will be cosidered as a false positive detection. Just a single prediction is considered correct, if the framework predicts few bounding boxes that overlap with a single ground-truth bounding box, and others are viewd false positives. One scores a system by the average precision (AP) of its precision-recall curve across a testset.

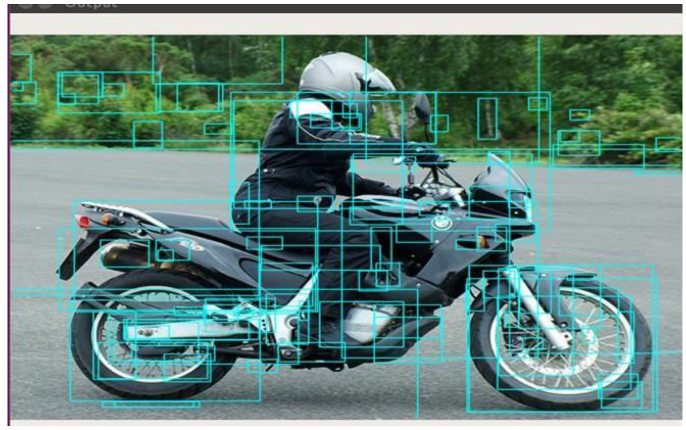

Fig3. Proposals generated through SSA and DPM

Fig3 shows the proposals generated through the SSA and DPM. Here we use the ground truth data that is available in dataset, in that ground truth dataset rectangles will be present with height and width. Here we compare the co-ordinates of the boxes with ground truth data; if co-ordinates will be matched then box will be generated.

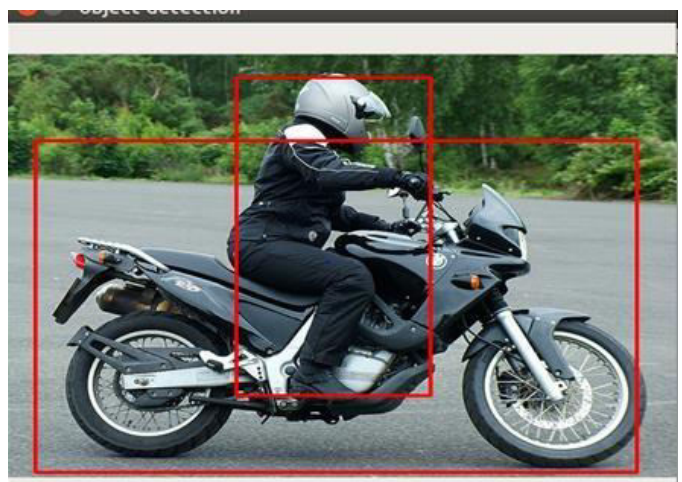

Fig4. Multiple object detection using SSA and DPM

The fig4 shows the final output in which we get the objects. Here we get the two objects by combining two methods which are SSA and DPM.

Image	IOU
2007_000027.jpg	0.9076
2007_000648.jpg	0
2007_002768.jpg	0.5336
2011_005768.jpg	0.8058
2007_00053.jpg	0.76

Table 1: Result of test images by combining DPM and R-CNN

From the above table however the obtained results are precision is 0.83 and whereas recall and average are 0.83 and 0.65 respectively.

CONCLUSION

In above study we conclude that whenever we integrate DPM and R-CNN techniques we get better result for multiple object detection. If we are using the DPM and R-CNN separately then it will not give more accuracy. In this framework we can use the DPM which can be used to generate the proposal that contains the part of the object for handling the occlusion and R-CNN is used for region based feature extraction. The proposals that are achieved from DPM and the proposals that are achieved from the SSA will be one by one send to the CNN model for extracting the features. When the CNN model is done with their process the DSD filter will clean the proposals generated from DPM. The proposals that are detected for every single object are outputted as an individual object.

REFERENCES

P. Felzenszwalb, R. B. Grishick, D.McAllister, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, Sep. 2010

R. Girshick, J. Donahue, T. Darrell, and J.Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEEConf. Comput. Vision Pattern Recognit., 2014, pp. 580–587.

O. Barinova, V. Lempitsky, and P. Kholi, “On detection of multiple object instances using hough transforms,” IEEE Trans. Pattern Anal. Mach. Intell., 34, no. 9, pp. 1773–1784, Sep. 2012.

B. Wu and R. Nevatia, “Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses,” Int. J. Compute. Vision, vol. 82, no. 2, pp. 185–204, 2009

R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-based convolutional networks for accurate object detection and segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142–158, Jan. 2016

W. Ouyang, X. Zeng, and X. Wang, “Partial occlusion handling in pedestrian detection with a deep model,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 11, pp. 2123–2137, Nov. 2016.

J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, “Selective search for object recognition,” *Int. J. Comput. Vision*, vol. 104, no. 2, pp. 154–171, 2013

B. Leibe, A. Leonardis, and B. Schiele, “Robust object detection with interleaved categorization and segmentation,” *Int. J. Comput. Vision*, vol. 77, nos. 1–3, pp. 259–289, 2008

Y. Jia *et al.*, “Caffe: Convolutional architecture for fast feature embedding,” in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–555, 2015.

A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.