Faster RCNN for Concurrent Pedestrian and Cyclist Detection

Anjali S
M.Tech student
Applied Electronics and Communication System
Nehru College of Engineering and Research Centre
Thrissur, India

Nithin Joe
Asst. Professor
Electronics and communication Department
Nehru College of Engineering and Research Centre
Thrissur, India

Abstract—*Pedestrian and cyclist detection systems are increasing attention with the development of autonomous automobiles and robotics. Many researches have been done for protecting vulnerable road users particularly pedestrians and cyclists. Little effort has been made to detect the pedestrian and cyclist concurrently. Here we are using a method called UB-MPR-Upper Body Multiple Potentil Region to detect them concurrently. For the classification and localization, we are using faster RCNN network. Experimental results indicate that the faster RCNN method outperforms the already existing fast RCNN method.*

Keywords— *faster RCNN; fast RCNN; pedestrian; cyclist detection; upperbody detection*

I. INTRODUCTION

Over the past decade, many researches has been made on improving driving safety. One of the important task in intelligent transport system is pedestrian and cyclist detection. Half of the world's road traffic deaths occur among vulnerable road users (VRU)[1]. Vulnerable road users include pedestrians, cyclists, motor cyclists. Pedestrians and cyclists are more vulnerable to accidents as there is no special protection device for them. Therefore, more attention is given to cyclist and pedestrians.

Pedestrians and cyclists are involved in different situations of the real life. Some of the applications are: vehicle automation, surveillance, robotics and many others. However, some problems exist in their detection as in a real life scenario pedestrians and cyclists are involved in different activities, which will create variations in clothing, size, poses, occlusion etc., as shown in Fig 1.

Pedestrian and cyclist detection has been extensively studied as an application in Automatic Driving Assistance Systems (ADAS) during the last decade. Companies are currently trying to develop intelligent vehicles, which are able to get information of their surrounding through the use of various sensors[2], as shown in Fig. 2.

Fig. 1. Pedestrians and Cyclists

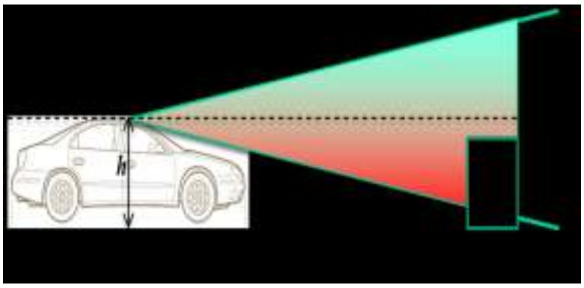

Fig. 2. The use of sensors in intelligent vehicles

Many approaches based on different such as monocular camera, stereo camera, lidar and radar are employed in vehicle environment perception systems. But for pedestrian and cyclist detection, vision sensors are the best a it can capture a high-resolution perspective view of the scene with useful color and texture information. Furthermore, vision technology is cost effective.

Traditional pedestrian or cyclist detection methods always consider pedestrians and cyclists separately [3], [4], although pedestrians and cyclists often appear in one picture. It causes confused detection results as input image has to be scanned several times. Therefore, detecting pedestrians and cyclists concurrently and differentiating them clearly is needed for the proper working of ADAS and autonomous vehicles

II. RELATED WORKS

Earlier pedestrian and cyclist detection are based on the use of feature information, commonly known as "hand-crafted feature based methods" and later deep neural networks are used.

A. *Hand-crafted feature-based methods*

Earlier pedestrian and cyclist detection includes Viola and Jones detector [6] and HOG detector [7]. In Viola and Jones detector, we first introduce an image representation called "Integral Image", which allows to quickly compute the features used by the detector. In HOG detector, we introduces the use of complex features by the application of Histogram of Oriented Gradients. Following the ideas of the methods mentioned above, other methods are proposed, like Local Binary Pattern [8], Dense Sift [9, 10], and Deformable Part Based Model [11] to gain accuracy.

B. *Deep Learning*

Recently, some advances in Computer Vision and Machine Learning have allowed to increase the accuracy of this area. In the recent years, since the concept of Deep Learning has been emerged and with this the use of large number of data to increase the accuracy. A pioneer work into this area is proposed in [12], known as AlexNet. In this method Deep Convolutional Neural Networks are applied to classify the 1.2 million high resolution images in the ImageNet contest into 1000 different classes.

C. *Region-Based Convolutional Neural Networks*

The goal of R-CNN is to take in an image, and correctly identify where the main objects (via a bounding box) in the image. R-CNN creates these bounding boxes, or region proposals, using a process called Selective Search. Selective Search looks at the image through windows of different sizes, and for each size tries to group together adjacent pixels by texture, color, or intensity to identify objects. Once the proposals are created, R-CNN warps the region to a standard square size and passes it through to a modified version of AlexNet. On the final layer of the CNN, R-CNN adds a Support Vector Machine (SVM) that simply classifies whether this is an object, and if so what object.

R-CNN works really well, but is really quite slow for a few simple reasons:

- It requires a forward pass of the CNN (AlexNet) for every single region proposal for every single image (that's around 2000 forward passes per image!).
- It has to train three different models separately to generate image features, the classifier that predicts the class, and the regression model to tighten the bounding boxes. This makes the pipeline extremely hard to train.

Fast R-CNN has different models to extract image features (CNN), classify (SVM), and tighten bounding boxes (regressor) but its descendant Fast R-CNN instead used a single network to compute all three. Fast R-CNN replace the SVM classifier with a softmax layer on top of the CNN to output a classification. It also added a linear regression layer parallel to the softmax layer to output bounding box coordinates. In this way, all the outputs needed came from one single network

III. SYSTEM MODEL

A. *Problem Formulation*

Detecting pedestrians and cyclists in urban areas is a difficult task. We may find people walking and cycling with different postures, occlusions, size. Therefore the use of just feature information at one scale is not enough to deal with all those situations. In that case the support of Deep Neural Networks provides the systems with robustness to detect objects on complex scenarios.

Not only the pedestrians and cylicts, but also the location of the pedestrians and the cyclists has to be identified.

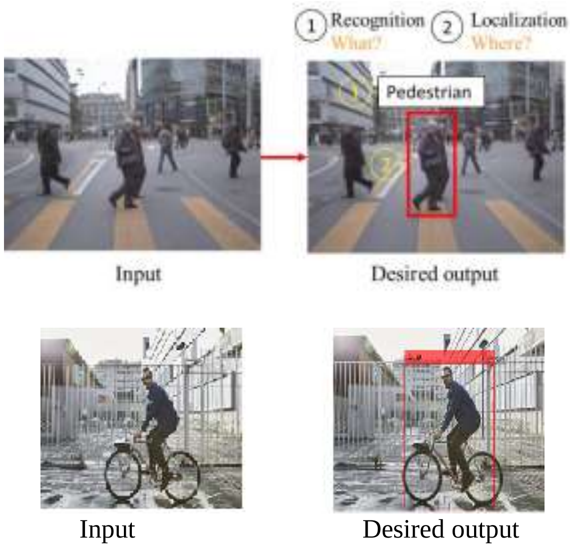

Fig. 3. Problem formulation

B. *Labels*

In order to provide the system with the estimated location of the candidates in the image, we propose to label all the all the images in the scene, assigning the candidates to either "Pedestrian" or "Cyclist".

C. *Proposed Solution*

Pedestrian and cyclist detection approach using Faster R-CNN should not only be able to accurately detect pedestrians and cyclists, but also has to detect them as quickly as possible. It is the real challenge. Recently, Deep Neural Networks allows to increase the recognition rate, but at expenses of high computational cost. To overcome the problem of computation, the use of GPU has facilitated the use of machine learning techniques. So we propose the Faster R-CNN method, to train and test our data. In Fig. 4, block diagram of our system is given.

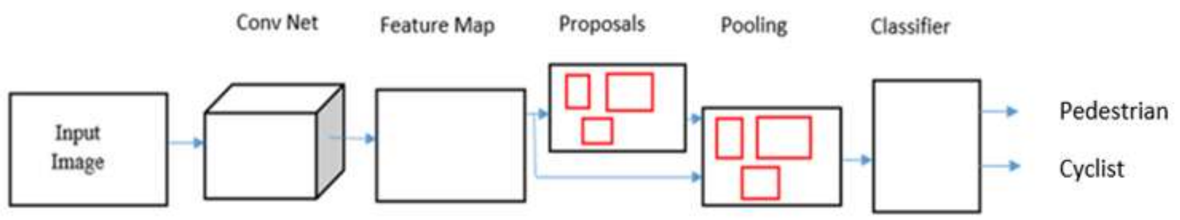

Fig 4. Block Diagram

After following the labeling procedure of the samples in each image. An image with its respective ground truth containing positive and negative samples are the inputs of the Faster R-CNN, which is able to use the same Convolutional Network to train the Region Proposals and the R-CNN. To get the Region Proposals it uses the information generated by the feature map, followed by a pooling step and a classifier to determine the class at which the candidate belongs to.

D. Classes

Here we are defining two classes: (1) Pedestrian (2) Cyclist.

IV. EXPERIMENTAL RESULT

A. Dataset

In order to train the network, we use the Tsinghua Daimler Dataset. The dataset provides a benchmark dataset for cyclist and pedestrian. Bounding box labels are provided for the classes “pedestrians” and “cyclists”.

B. Experiments

We trained and tested our system on a computer with an Intel® Core TM i3-4030U CPU@ 1.90 GHz 3 and a GPU GeForce GTX Titan X.

V. APPLICATIONS

- It can be used in driver less cars.
- It can be used in robotics.
- It can be used in military purposes.

VI. ADVANTAGES

- Less time required for detection
- More accurate when compared to already existing methods
- Huge data set

VII. CONCLUSION

A unified method for concurrent pedestrian and cyclist detection is presented, where Faster R-CNN based model is used for classification.

The proposed method outperforms the state of art method in terms of accuracy and speed. The system can be used in driving assistance systems for detecting the location of pedestrians and cyclists. Final outputs are shown in Fig 5.

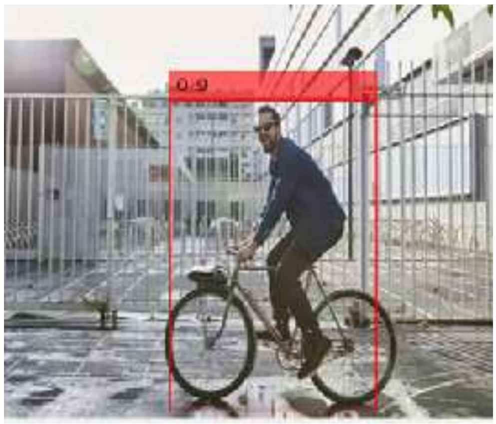

Fig 5. Final Outputs

Acknowledgment

We thank Dr. S Raj kumar , HOD, ECE Department, NCERC and Muruganantham C, Assistant Proffessor, ECE Department, NCERC for helpful discussions and encouragement.

References

[1] W. H. Organization, "WHO," 2013. [Online]. Available:http://www.who.int/mediacentre/news/notes/2013/make_walking_safe_20130502/en/.

[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Dec. 2010.

[3] C. W. B. S. P. B. P Dollar, "Pedestrian Detection: An Evaluation of the State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[4] A. L. A. S. T. G. D Geronimo, "Survey of Pedestrian Detection for Advance Driver Assistance Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, 2010.

[5] D. G. M Enzweiler, "Monocular Pedestrian Detection: Survey and Experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 12, 2009.

[6] M. J. P Viola, "Robust Real-Time Face Detection," *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.

[7] B. T. N Dala, "Histogram of Oriented Gradients for Human Detection,"

[8] H. T. S. Y. Xiayu W., "An HOG-LBP Human Detector with Partial Occlusion Handling," in *IEEE International Conference on Computer Vision*, 2009.

[9] L. D., "Object recognition from local scale-invariant features.," in *IEEE International Conference on Computer Vision*, Corfu, Greece, 1999.

[10] D. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[11] R. B. G. D. M. a. D. R. Pedro F. Felzenszwalb, "Object Detection with Discriminatively Trained," *IEEE Trans. Pattern Analysis on Machine Intelligence,* vol. 32, pp. 1627-1645, 2010.

[12] M. M. R. T. L. V. G. Rodrigo Benenson, "Pedestrian detection at 100 frames per second," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR*, Providence, RI, 2012.