2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. June 4-5, 2020

Vehicle Detection and Classification using Improved Faster Region Based Convolution Neural Network

Usha Mittal
Department of Computer Science and Engineering
Lovely Professional University
Phagwara, India
usha.20339@lpu.co.in,

Rahul Potnuru
Department of Computer Science and Engineering
Lovely Professional University,
Phagwara, India
rahulpotnuru1998@gmail.com

Priyanka Chawla
Department of Computer Science and Engineering, Lovely Professional University,
Phagwara, India
priyanka.22046@lpu.co.in

Abstract-- **Over the last few years, object detection has arisen as a powerful tool for developing reliable systems like detection of vehicles using computer vision. Convolutional neural network can be used to employ feature synthesis methods to concatenate low level and high level features as well as to detect vehicles of different size and scale. In this paper, an improved faster R-CNN method has been proposed. The proposed method has been evaluated using FLIR_ADAS dataset for both thermal and RGB images. Experiments have been performed on thermal images and results demonstrated significant increase in accuracy as well as non-ambiguous detections as compared to conventional Faster R-CNN method.**

Keywords-- Convolutional Neural Network (CNN), Vehicle detection, classification, transfer learning, faster R-CNN

I. INTRODUCTION

By looking at an image, people immediately notice the different objects contained in the image, their location, and find out the correlation between objects. When image processing techniques could be precise as well as quick enough, machines would be able to detect and locate vehicles without sophisticated sensors, and additional technologies would be able to deliver real-time data to users. Similarly, if these procedures could perform deep learning tasks with a highly effective and superior standard in the same manner mankind do, it would be true Artificial Intelligent. The main functions of image processing are therefore a sequence of recognition, classification, and localization and object detection. The main issues of these techniques are speed, reliability, accuracy, computational cost, and overall complexity.

A more generalized (multi-class) application of image processing can be used in vehicle detection where a various type of vehicles needs to be detected. It also has a key role to play in surveillance systems. Still it is an important challenge in computer vision. Earlier Gaussian Mixture Model [11] has given acceptable results but it is not ideal because of background clutter, illumination changes, occlusion etc. A machine learning algorithm generally takes a fixed size of input and output to be trained. Another significant challenge to the widespread adoption of object detection systems is the need for real-time (30fps) while being precise in detection. Because of these reasons, these methods provide less accuracy as different vehicles have different aspect ratio and scale. So, a complex system may be able to produce acceptable results. The time required for inference is more in complex models. The trade-off between accuracy and performance must be selected as per the requirement.

Faster Regional Convolutional Neural Network

The key idea behind the use of CNN is to make reliable, effective, firm and efficient object detection system. For detection and localization, various methods such as the Region with CNN (R-CNN), the Fast Region with CNN (Fast R-CNN), and the Faster Region with CNN (Faster R-CNN) were proposed by different researchers. R-CNN uses selective search to produce 2000 bounding boxes with fixed size. Its efficiency is very low. Fast R-CNN [6] produces bounding boxes on the feature map of last layer and accepts ROI pooling. Faster R-CNN [12] introduces the Region Proposal network (RPN) and then classification layer. Time and space complexity of this method is high at the training and testing phase. The RPN layer selects candidate boxes efficiently.

1. It uses the aspect ratio of [0.5, 1, and 2] but vehicles have low aspect ratio.
2. Candidate regions are extracted on high-level feature map, as it is having more semantic information but it is not able to locate the objects properly.

II. RELATED WORK

In this section, introduction of vehicle detection and classification techniques have been discussed. Vision based object detection and classification systems can be categorized into three classes' i.e. Hand crafted feature based, motion based and CNN based methods.

Motion based methods comprise optical flow, frame subtraction and background subtraction. In frame subtraction, two or more consecutive frame sequences are subtracted from each other to analyze the motion object. Motion vector of each pixel is calculated in optical flow for tracking the objects, but the complexity of the model is very high. In background subtraction, background and foreground objects are differentiated by comparing the background image with input image but, these methods cannot detect and classify static vehicles.

Hand crafted or Machine learning feature based methods comprise Harr-like, Histogram of Oriented Gradients (HOG) [13] and, SIFT [14]. Feature representation ability is very low in these methods.

CNN based methods are very good in feature representation and got favorable results. R-CNN makes use of object proposal produced by selective search [15] to train CNN for object detection.

SPP-Net and Fast R-CNN have high computational speed as compared to R-CNN as these methods generate candidate region on feature map. Faster R-CNN uses the Region proposal network (RPN) rather than selective search; also it provides high accuracy as compared to other methods. YOLO is a single stage detector which uses a single feed-forward CNN for the localization and prediction of object class. This method is computationally very fast; though accuracy is less. SSD is also a single stage detector which filters anchors of various aspect ratios and scales on multiple feature maps.

Dataset

Datasets are very important when using machine learning methods to solve detection problems. Here, dataset is collected from open source library which contains thermal and RGB images. Vehicles present in the images are classified in six classes and annotations are performed using labelimg tool. To increase the dataset, augmentation is also performed.

III. Limitations of Faster R-CNN

Although faster R-CNN provides promising results but it has some limitations. The major limitation of Faster R-CNN is duplicate detections i.e. sometimes it detects multiple labels for a single object. Figure 1 show that a car is detected as light vehicle as well as heavy vehicle with confidence score of 98% and 81% respectively. So, number of vehicles present in the image is two rather than one.

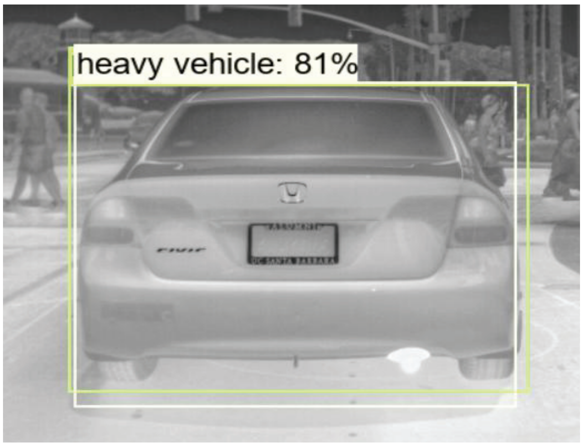

Fig. 1. Duplicate Detections

IV. Proposed Method

Here, a method is proposed to fix the limitations of Faster R-CNN. First input image is given to Faster R- CNN mode and detections of the model are examined for duplicate detections. If in a particular region, multiple labels are detected then confidence scores of multiple labels are compared, and the label with the highest confidence score is retained and others are discarded. Final Output image will have the unique detections only. Figure 2 shows the flow chart of the proposed methodology.

TABLE I. Literature Survey

Year	Authors	Approach	Result and Discussion
2011	Y. M. Chan et al	The imaging working principle of the proposed sensor device is considered with the process of the signals acquisition and fusion.	Algorithm achieved an accuracy of 92.84% under poor lighting conditions.
2013	Sebastian Tuermer et al	Road database was used for training. Disparity map was used to exclude elevated objects and for the classification of remaining objects Histogram of Gradient is image.	The algorithm showed a top accuracy up to 70%.
2013	Yi-ling Chen et al	Computer vision and cloud computing procedures based on image processing have been used.	Higher detection accuracy and higher positioning accuracy.
2014	Ross Girshick, Jitendra Malik, Jeff Donahue, Trevor Darrell.	Regional Convolutional Neural Network (RCNN)	This algorithm achieved a mAP of 62% on PASCAL VOC 2012
2015	Ross Girshick	Fast Regional Convolutional Neural Network (Fast RCNN)	This algorithm achieved a mAP of 66% on PASCAL VOC 2012
2015	Yongbin Gao et al.	Frame distinction algorithm has been used to detect the travelling vehicles. Symmetry filter was used to extract the front view of the vehicle. Three layers of restricted Boltzmann machines (RBM) were used to classify the car model.	The proposed algorithm achieved 100% accuracy, hence it is reliable model.
2016	Jifeng Dai, Yi Li, Kaiming He, Jian Sun	R-FCN	This algorithm achieved a mAP of 83.6% on PASCAL VOC 2007
2016	Kaiming He, Shaoqing Ren, Ross Girshick, Jian Sun	Faster R-CNN	This algorithm achieved a mAP of 75.9% on PASCAL VOC 2012
2019	Aparna et al	Pre-trained CNN model based on residual networks was used. Augmentation techniques were used on the data for the pothole's detection.	Proposed method provided accuracy of 97.08%

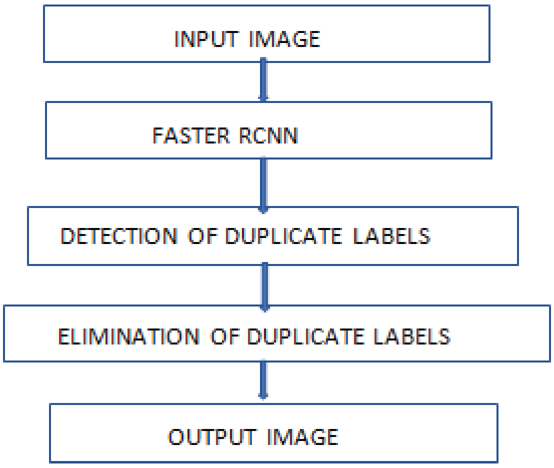

Fig. 2. Proposed Methodology

V. RESULTS AND DISCUSSION

The model is implemented using Tensorflow API and trained on NVIDIA GPU with 4 GB capacity. Figure 3 to Figure 8 shows the output of the proposed methodology. Figure 3 shows total six objects are detected out of which two vehicles are having duplicate labels. One vehicle is detected as heavy vehicle with confidence score of 93% and light vehicle with confidence score of 85%. Another vehicle is detected as light vehicle and heavy vehicle with 82% and 80% confidence score respectively. Where Figure 4 shows detections made by proposed method in which duplicate labels are removed and label with highest confidence score i.e. one vehicle with confidence score of 93% and another vehicle with 82% score is retained. Figure 5 shows vehicle is detected as light vehicle as well as heavy vehicle with confidence score of 84% and 89% respectively.

Where figure 6 shows vehicle is detected as heavy vehicle with confidence score of 89%. Figure 7 shows vehicle is detected as light vehicle and heavy vehicle with confidence score of 98% and 81% respectively. Where Figure 8 shows vehicle is detected as light vehicle with 98% confidence score.

Table 3, 4 and 5 shows the confusion matrix of Faster R-CNN, Improved Faster R-CNN and F1 score of both methods respectively for the images given in experimental results. From the confusion matrix, it can be analyzed that number of false negatives are more in case of Faster R-CNN as compared to improved Faster R-CNN method. Thus, F1-Score of improved Faster R-CNN is better than conventional Faster R-CNN method and it shows improved accuracy.

TABLE II. CONFUSION MATRIX OF FASTER R-CNN

		Actual	
		Light Vehicle	**Heavy Vehicle**
Predicted	**Light Vehicle**	3	1
	Heavy Vehicle	3	3

TABLE III. CONFUSION MATRIX OF IMPROVED FASTER R-CNN

		Actual	
		Light Vehicle	**Heavy Vehicle**
Predicted	**Light Vehicle**	2	0
	Heavy Vehicle	1	3

Faster R-CNN

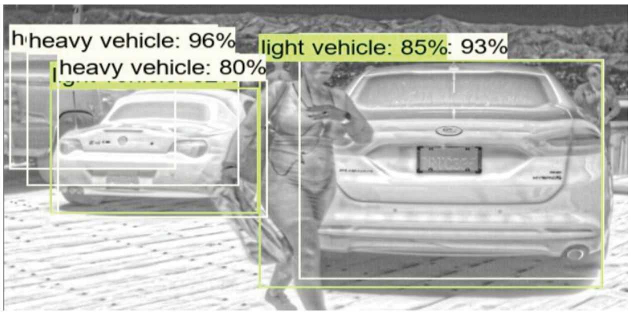

Fig. 3. Detections of Faster R-CNN

Proposed Method

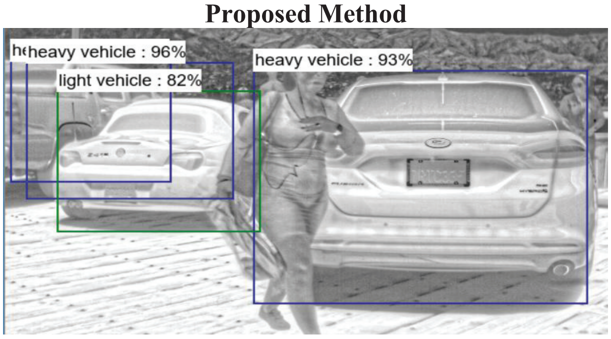

Fig. 4. Detections of Proposed Method

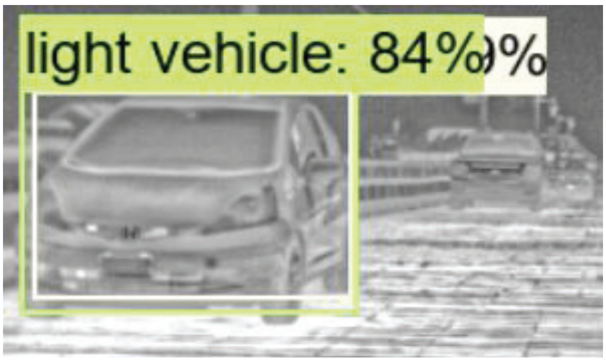

Fig. 5. Detection of Faster R-CNN

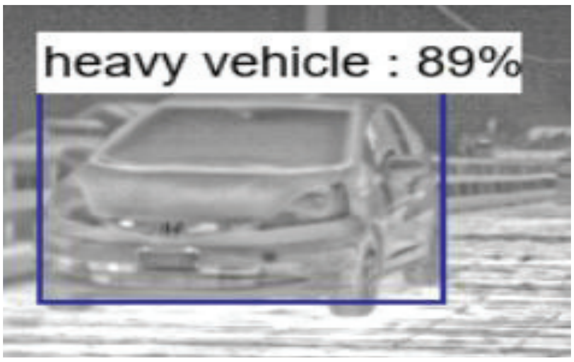

Fig. 6. Detections of Proposed Method

Fig. 7. Detections of Faster R-CNN

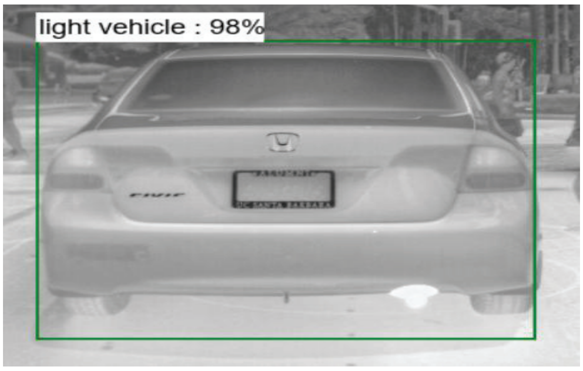

Fig. 8. Detections of Proposed Method

TABLE IV. F1 SCORE OF FASTER R-CNN AND IMPROVED FASTER R-CNN

	Faster R-CNN	Improved Faster R-CNN
Light Vehicle	60%	75%
Heavy Vehicle	60%	85.7%

VI. CONCLUSION

In this paper, vehicle detection and classification has been done using Faster R-CNN method and improved Faster R-CNN method. Although Faster R-CNN is very efficient model for object detection as it can detect objects at different size and scale. But this model has some limitations too. During the experiment, one of major limitation which is analyzed is that it generates duplicate labels for the same object. Duplicate detection produces incorrect results in terms of number of objects counted in image. In this paper, to solve the issue of duplicate detections, Faster R-CNN method is modified which provides better results as compared to conventional Faster R-CNN method.

REFERENCES

[1] Xu, J. (2017). Deep Learning for Object Detection: A Comprehensive Review. " https://towardsdatascience.com/deep-learning-for-object-detection-a-comprehensive-review-73930816d8d9" last accessed on 25-11-2019

[2] Shaoqing Ren, Kaiming. He, Ross Girshick, Jian Sun. (2016). Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Network

[3] Chan Y.M. Huang, S. S., Fu, L. C., Hsiao, P. Y., & Lo, M. F. (2012). Vehicle detection and tracking under various lighting. IET Intelligent Transport Systems

[4] Tuermer, S., Kurz, F., Reinartz, P., & Stilla, U. (2013). Airborne Vehicle Detection in Dense Urban Areas Using HoG Features and Disparity Maps. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

[5] Chen. Y, L. Chen, T. S. Huang, T. W. Yin, L. C., Wang, S.-Y., & Chiueh, T.-C. (2013). Intelligent Urban Video Surveillance System for Automatic Vehicle Detection and Tracking in Clouds. IEEE 27th International Conference on Advanced Information Networking and Applications (AINA) (pp. 814-821). Barcelona, Spain: IEEE

[6] Ross. Girshick. (2015). ICCV 2015. IEEE International Conference on Computer Vision. Fast R-CNN. DOI: 10.1109/ICCV.2015.169

[7] Ross. Girshick, Jeff. Donahue, Trevor. Darrell, Jitendra. Malik. (2014). The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Rich feature hierarchies for accurate object detection and semantic segmentation. DOI: 10.1109/CVPR.2014.81

[8] Jifeng. Dai, Yi. Li, Kaiming. He, Jian. Sun. (2016). Advances in Neural Information Processing Systems 29 (NIPS 2016)

[9] Gao. Y, Lee. H. J. (2015). Moving Car Detection and Model Recognition based on Deep Learning. Advanced Science and Technology Letters

[10] Bhatia. Y, Rai. R, Gupta. V, Aggarwal. N, & Aparna, A. "Convolutional neural networks-based potholes detection using thermal imaging," Journal of King Saud University Computer and Information Sciences, 2019.

[11] C. Stauffer and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on, Fort Collins, 1999, p. 252 Vol. 2

[12] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (Nips, Montréal, 2015.

[13] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, 2001.

[14] H. Harzallah, F. Jurie, C. Schmid, Combining efficient object localization and image classification, in: ICCV, 2009.

[15] P. Viola, M. J. Jones, Robust real-time face detection, in: IJCV, 2004.