Preprocessed Faster RCNN for Vehicle Detection

Mduduzi Manana, Chunling Tu and Pius Adewale Owolawi
Department of Computer Systems Engineering
Tshwane University of Technology
Pretoria, South Africa
mdwens@gmail.com, tclchunling@gmail.com, owolawipa@tut.ac.za

***Abstract*—This paper presents a pre-processed faster region convolution neural network (faster RCNN) for the purpose of on-road vehicle detection. The system introduces a preprocessing pipeline on faster RCNN. The preprocessing method is for the improvement on training and detection speed of Faster RCNN. A preprocessing lane detection pipeline based on the Sobel edge operator and Hough Transform is used to detect lanes. A Rectangular region is then extracted from lane coordinates which is a reduced region of interest (ROI). Results show that the proposed method improves the training speed of faster RCNN when compared to faster RCNN without preprocessing.**

Keywords—Convolution Neural Network, RPN, ROI, Hough Transform, Sobel edge detection

I. INTRODUCTION

Object detection methods have been used in many applications including that of public safety [1]. Vehicle detection methods which are based on visual object detection can help to increase driving security and decrease road crime such as on-road hijacking [2]. In this context a monocular camera is used as a passive sensor for data input and visual object detection methods are applied to detect vehicles from the traffic scene. In recent times, most state-of-the-art visual object detection methods are based on convolution neural networks (CNN) .

Region convolution neural network (RCNN) has achieved impressive detection results compared to other methods of the same class. The RCNN [3] operates by running region proposals, generated by proposal methods, on a CNN network. The drawback of RCNN is its high computational cost because each region is processed separately. Fast region convolution neural network (fast RCNN) [4] improves on the downside of RCNN by using region proposals as attention detectors for a shared feature map. The shared feature map eliminates the need of processing each region separately, this improves the computation cost of RCNN. The proposal generating method is a drawback for fast RCNN, which is slow compared to the CNN network. As a result, it has a negative impact on the overall speed of fast RCNN.

To overcome this drawback, faster RCNN [5] shares CNN computation between the proposals and the detection network. Faster RCNN creates an almost cost free region proposal generating mechanism. The Region Proposal Network (RPN) of faster RCNN generates proposals that are used as attention directors for a shared feature map. The RPN replaces an external region proposal generating process such as selective search , and this results in improved computational cost.

Faster RCNN performs well for general object detection but performs unimpressively when applied to vehicle detection. This can be improved through parameter tuning and algorithmic modification [6] .

In this paper we focus on the algorithmic modification of Faster RCNN. A preprocessing method is integrated to faster RCNN to improve training and detection speed.

The preprocessing pipeline reduces the region of interest which results in a reduced number of pixels to be processes by faster RCNN. The preprocessing pipeline is based on the Sobel edge detection [7] and Hough transform [8] with an extracted rectangular region as the end result. This improves the processing speed of the Faster RCNN. Since we are focusing on vehicles around the ego vehicle and their behavior, the tracking procedure in lane detection is ignored. This provides a simplified preprocessing method for the initial stage with computation advantages.

The preprocessed faster RCNN can then be used for vehicle detection through a vehicle mounted monocular camera which captures the traffic scene. This is for the purpose of detecting suspicious vehicles around the ego vehicle.

The rest of this paper is arranged in the following manner: Section II introduced related work, Section III presents the proposed approach, Section IV validates the proposed method by using experiments, Section V concludes the paper.

II. RELATED WORK

Vehicle detection using a camera mounted on a vehicle has been implemented in [1][9][10][11] and object detection methods were applied to detect vehicles. The improvement of a CNN based algorithm was done in [1] for vehicle detection with image pre-processing integrated. This was shown to improve computational cost by reducing the region of interest (ROI). The reduction of the ROI which is based on lane detection and tracking has been shown to be reliable on roads [10].

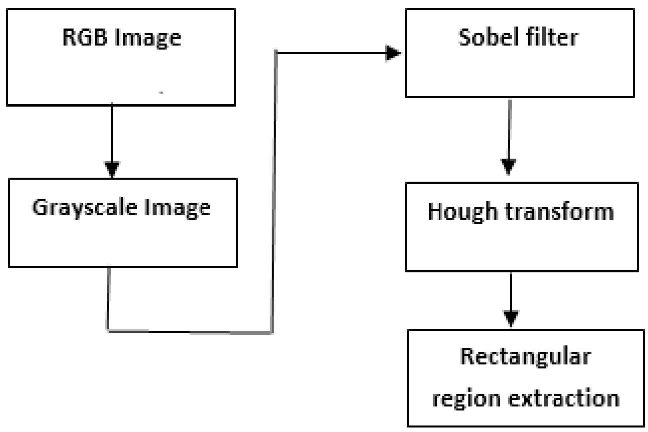

Fig. 1 Pipeline for the lane detection and rectangular region extraction

A. Lane detection

Lane detection has been used in many areas, such as lane departure warning systems. In [12] a camera based lane detection algorithm was developed to detect lane markers on the road.

Lane detection and tracking consists of the following steps: (1) detecting lane markings on the road with respect to the ego vehicle; (2) fitting those lane markings onto a model to detect the trajectory of the markings.

Edged detection methods are used to detect road markings based on pixel changes in an image. The gradient based edge detection is commonly used, where an edge is detected if the gradient is above a certain threshold. Sobel, Roberts cross operator, Prewitt operator and Canny edge [13] all comes from the gradient edge detection family. The Laplacian based edge detection is based on locating zeros at the second derivative of an image, which detects regions of rapid intensity changes indicating edges. Laplacian of Gaussian [13] is a representative of this branch. Lane marking localization also include steerable filters [14] which rotates to different angles and then synthesize those filters of arbitrary orientation from linear combinations of basis filters [10] for lane tracking and adaptive thresholds [15]. This method takes a greyscale image or color image as input then outputs a binary image representing the segmentation. For each pixel in the image a threshold is calculated.

Hough transform is used to detect straight lines in an image by connecting points that lie in a straight line. It has been used after edge detection to detect lane markers on the road [1], then other a tracking based method such as curve fitting or a lane position tracker can used as a final stage [10].

B. Vehicle detection

Object detection methods have been used in [1][9][10] for vehicle detection with region of interest reduction before vehicle detection and tracking. Conventional object detection methods were based on sliding windows to produce features that determine whether they contain objects or not. Deformable part models (DMP) [16] achieved the best results among conventional object detectors. Many vehicle detection systems get improved from DPM.

Deep learning techniques such as CNN dominate the state-of-the-art object detection methods, of which RCNN has achieved impressive detection results [3]. RCNN computes 2000 bottom-up proposals from an input image, the proposals features are extracted using CNN and each region is then classified using sector vector machine (SVM) [3]. RCNN uses selective search for proposal generation, then from each region proposal a 4096-feature vector is produced. The image is then warped into 227 x 227 pixels for computation of features. The warped image is then processed through five convolution neural networks and two fully connected layers [17]. A trained sector vector machine is then used to score the computed features.

The Fast RCNN method computes a convolutional feature map for the entire input image and then classifies each object proposal using a feature vector extracted from the shared feature map. This improves the processing speed [4]. The bottle neck of Fast RCNN is the proposal generating method which is slow compared to the detection network.

Faster RCNN addresses this bottleneck by introducing a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, and this results in nearly cost-free region proposals. RPN serves as an attention director telling the unified network where to look. A RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high quality region proposals, which are used by Fast R-CNN for detection. RPN and Fast R-CNN are then merged into a single network by sharing their convolutional features. Faster RCNN has been implemented in [6] for vehicle detection. In [6] faster RCNN is examined specifically for application in vehicle detection.

III. PROPOSED APPROACH

In this work an algorithm based on lane detection, a rectangular region of interest and faster RCNN object detection method is introduced for the purpose of vehicle detection. It begins with a lane detection pipeline for vehicles that are on the road because the lane markings are the most prevalent and visible features on the road. The lane markings are used as an estimate of the road because this is where vehicles are mostly found. This estimate enables the extraction of an accurate rectangular region of interest. The pipeline reduces the region of interest as a pre-processing step to improve the computation cost of faster RCNN. This pre-processing pipeline is similar to [1] [10] but our pipeline is quicker because there is no lane tracking involved. This provides our pipeline with a computational advantage. Method [1] also uses deep multibox which is slower compared to faster RCNN. Our lane detection pipeline reduces the pixels to be processed by faster RCNN and therefore improves on the computation cost of faster RCNN. A rectangular region of interest that is based on lane coordinates is then used as input to the Faster RCNN.

A. Lane detection

As shown in Fig. 1, the lane detection is composed of the following parts:

1) The RGB road image is converted into grayscale image.
2) The Sobel filter extracts edges.
3) The Hough transform detects straight lines.
4) Finally, a rectangular region of interest is cropped using the lines obtained from Hough transform.

The proposed pipeline uses the Sobel operator to detect edges which slightly compares to the pipeline of [1] that uses a canny edge detector which is computationally expensive. The Sobel edge detector is a simple filter with low computational cost and is ideal for the lane detection method proposed in this paper. It detects edges in a grayscale image which are then used as candidates of lines for the subsequent stage of Hough transform. The Sobel operator is considered in this situation because there is no in-depth edge detection required. Figure 2 shows the Sobel operator on an image.

The Sobel operator use two 3x3 kernels one rotated 90 degrees of the other, as shown in (1) and (2).

(a) (b)

Fig. 2 (a) greyscale image and (b) detected edges

$$Gx = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \quad (1)$$

$$Gy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (2)$$

The Sobel operator can be broken down as a product of an averaging kernel and a differentiating kernel (combining gradient with smoothing). *Gx* and *Gy* can be broken down into

$$Gx = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}, \quad (3)$$

$$Gy = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \quad (4)$$

At each point in the image the resulting gradient can be combined to give the magnitude in (5) and direction of the edge (6)

$$G = \sqrt{Gx^2 + Gy^2} \quad (5)$$

$$\theta = \mathrm{atan}(Gy / Gx) \quad (6)$$

This reduces the computational cost as there are fewer calculations. The direction (θ) can be set to 0 for a vertical edge detection only. This filter only needs 8 image points around a point to compute a result and then integer arithmetic is used for gradient computation. Areas of high gradient are represented as white lines on a black background which are the edges shown in Figure 2.

When the edge detection is complete, the Hough transform is applied to the detected edges to produce candidates of straight lines to represent lane markings on the road. The Hough transform is robust for eliminating noisy data on an edge image and produce accurate lines. It has an added advantage of handling incomplete data very well as a result of occlusion and shadows on the road [8]. Method on [1] uses Hough transform which it is applied to a canny edge detector.

Eq. (7) is used as a description of a line for Hough transform.

$$r = x\cos\theta + y\sin\theta \quad (7)$$

In (7) (x, y) represents a point (pixel) which has to be computed as to whether it belongs to a line or not. r represents the distance from the line to the origin. The parameter θ represents the angle between the origin of the line and the x axis. Each point in this context is represented by (r, θ). Every straight line that passes a single point corresponds to a sinusoidal curve in the (r, θ) plane which is unique to that point. Two or more points are considered a straight line if sinusoids are produced that intersect at (r, θ) which is a candidate of a straight line for those points. In this paper Hough transform is applied to the edging image obtained from the Sobel edge detector.

To detect a lane, a minimum length of edges is set. The coordinates of lines representing lane markings are then used to crop a rectangular region of interest as shown in Figure 3 (which shows the regular image and the cropped image). This rectangular region shown in Figure 3(b) is used as input of faster RCNN.

B. Vehicle detection

The extracted rectangular region is fed into Faster RCNN for vehicle detection. The proposed algorithm as shown in Figure 4 shows a preprocessed image being input into the convolution layer of Faster RCNN.

The purpose of the pre-processing step in Figure 4 serves to reduce the region of interest in order to speed up processing of image pixels. This will reduce training time of the detector.

The advantage of Faster RCNN is the proposal generation that is almost cost free. The region generates 300 proposals. RPN takes an image and produces rectangular proposals with an objectness score. This is done by sliding a small network over the shared feature map. At each sliding window region proposals are predicted. The regression layer outputs the coordinates of the predicted proposals and the classification layer produces a score as to the possibility of there being an object for each predicted proposal. A single-scale detection is used in this paper, because it has been shown that, for the similar performance, multi-scale detection is more expensive computationally than single scale detection [6].

(a) (b)

Fig. 3 (a) RGB image and (b) the cropped rectangular image based on points of a line detect by Hough transform

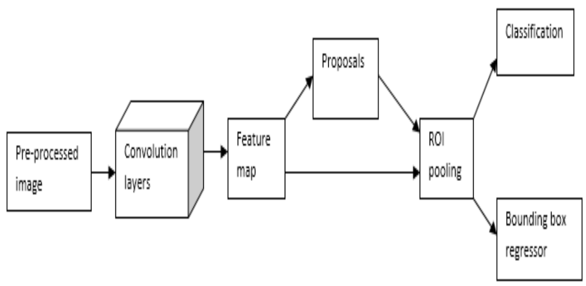

Fig. 4 Combined method with preprocessed image fig 3b based on fig. 1 and faster RCNN [5]

IV. EXPERIMENT

A similar training process for both faster RCNN and the proposed method are used to compare their training speed. Same computation resources and data are used for training. The training process is done on 102 images of vehicles on both methods. The training follows the process laid out in faster RCNN pipeline, with the following strategies:

- The regional proposal (RPN) network is first trained
- The RPN is then used to train fast RCNN
- RPN is retrained sharing weights with Fast RCNN
- Fast RCNN is retrained using the updated RPN

Training is done on a single Intel Celeron CPU, with 2GB of memory. The learning rate of the detector is 0.000001. The same data is used for the proposed method and for faster RCNN. The results of the above procedure are compared between faster RCNN and the proposed method. The results of training time, number of iterations, and mini batch accuracy are shown in Table I for the first round training, and Table II for the re-training process. The results of the final training shows that our method trains quicker than Faster RCNN with a slightly lower mini batch accuracy and more iterations.

V. CONCLUSION

A vehicle detection method is proposed that reduces a region of interest by preprocessing an image before it is processed by faster RCNN. The results show that preprocessing an image into a reduced rectangular portion of the original image improves training speed. The image is processed faster as a result of the rectangular cropped image based on lane detection. This is because with the cropped region there are a fewer pixels to process and this results in a shorter training time.

TABLE I. COMPARISON OF THE FIRST ROUND TRAINING FOR EXISTING METHODS AND PROPOSED METHOD

Method	***Iteration***	***Time elapsed(s)***	***Mini batch accuracy***
Faster RCNN (Training RPN)	72	218.13	50%
Proposed (training RPN)	67	93.33	100%
Faster RCNN (fast RCNN training using RPN)	45	81.20	81.25%
Proposed (fast RCNN training using RPN)	49	40.03	56.25%

TABLE II. RETRAINING RESULTS EXISING METHODS AND OUR PROPOSED METHOD

Method	***Iteration***	***Time elapsed (s)***	***Mini batch accuracy***
Faster RCNN (retrain RPN)	72	193.01	100%
Proposed (retrain RPN)	67	88.23	100%
Faster RCNN (fast RCNN training using updated RPN)	45	60.59	71.43%
Proposed	49	36.31	69.8%

ACKNOWLEDGMENT

This work is based on the research supported by the National Research Foundation of South Africa (Grant Number 99383).

REFERENCES

[1] H. Guan, W. Xingang, W. Wenqi, Z. Hanand W. Yuanyuan, “Real-Time Lane-Vehicle Detection and Tracking System”, Beijing: Beijing Information Science & Technology University, 2016.

[2] “Crime statistics”, 2017. [Online]. Available: https://africacheck.org/factsheets/factsheet-south-africas-201516-crime-statistics . [Accessed: 2 June 2017].

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Richfeature hierarchies for accurate object detection andsemantic segmentation," in Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014, pp,

[4] R. Girshick, "Fast R-CNN," arXiv preprint arXiv:1504.08083, 2015.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," arXiv preprint arXiv:1506.01497, 2015.

[6] Q. Fan, L. Brown, J. Smith, “A closer look at vehicle detection for Faster RCNN”, 2016 IEEE Intelligent Vehicles Symposium (IV) Gothenburg, Sweden, June 19-22, 2016.

[7] M. Wen, C. Zhong, “Application of Sobel algorithm in edge detection of images”, China High-tech Enterprise, pp.57-62.

[8] J. Illingworth, J. Kittler, “A survey of Hough Transform”, April 1987.

[9] Q. Shiru. L. Xu, “Research on multi-feature front vehicle detection algorithm based on video image”, 28th Chinese Control and Decision Conference (CCDC), 2016.

[10] S. Sivaraman, M. Manubhai, “Integrated lane and vehicle detection localization and tracking: a synergistic approach”, IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 2, June 2013.

[11] C. L. Tu and S. Z. Du, “Moving vehicle detection in dynamic traffic contexts”, in Electronics, Communications and Networks V, Singapore: Springer, 2016, pp 263-269.

[12] C. Kreucher, S. Lakshmanan, “Lana: A lane extraction algorithm that uses frequency. Domain features”, IEEE Transactions on Robotics and Automation, Vol. 15, No. 2, April 1999.

[13] R. Maini, H. Aggarwal. “Study and comparison of various edge detection techniques”.

[14] T. Freedman, E. Adelson, “Design and user of steerable filters”, IEEE Transaction Patern Analysis and machine intelligence, vol. 13, no. 9, September 1991.

[15] M. Meuter, S. Muller-Schneiders, A. Mika, S. Hold, C. Nunn, andA. Kummert, “A novel approach to lane detection and tracking,” in Proc. 12th Int. IEEE ITSC, Oct. 2009, pp. 1–6.

[16] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively Machine Intelligence”, IEEE Transactions on, vol. 32, pp. 1627-1645, 2010.

[17] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS*, 2012.