VEHICLE CLASSIFICATION BASED ON BACKGROUND SUBTRACTION WITH DEEP LEARNING IN TRAFFIC SCENE

[1]K.Kishore Anthuvan Sahayaraj, [2]Dr. K. Venkatachalapathy, [3]A. Sabitha, [4] S. Ramya, [5]S.Mehathaj Begam
[1]Research Scholar, [2]Professor & Head,
[1,3,4,5]Department of Computer Science and Engineering,
[2]Department of Computer and Information Science,
[1,2]Annamalai University, Chidambaram, India
[3,4,5]Christ College of Engineering and Technology, Puducherry, India

Abstract : Intelligent traffic surveillance system plays a vital role in modern day traffic analysis. Numerous ways have been developed to streamline the process of analyzing traffic. The counting and classification of the vehicle during some period of time in an area require more effort even the couple of proximity sensors can calculate the track of moving the vehicle, but they are not economical. Therefore this paper proposes a framework for vision-based moving vehicle counting and classification system by combining background subtraction, proposal generation network and iterative refinement Convolution Neural Network (CNN). These approaches are applied in a cascade way to obtain the classification and counting. The experiment results show that the proposed approach has the best classification accuracy on the MOT dataset.

***IndexTerms* -** Background subtraction, faster RCNN, object detection,

I. INTRODUCTION

In real time traffic management and monitoring have increased in the last few years because of the increase in the vehicles and an increase in surveillance cameras. Previously the wireless sensor network [1] EM microwave [3] has been deployed but these are not cheap to install. So the traffic management system uses a vision-based surveillance system [4].

Video surveillance systems have become cheaper and better because of the increase in storage capabilities and computational power. The videos stored by these surveillance systems are generally evaluated by humans, which is time-consuming. To overcome this limitation, the need for more robust, automatic video-based surveillance systems has increased interest in the field of computer vision. The objectives of a traffic surveillance system are to detect, track and classify the vehicles but they can be used to do complex tasks such as driver activity recognition, lane recognition etc. The traffic surveillance systems can have applications in a range of fields such as public security, detection of anomalous behavior, accident detection, vehicle theft detection, parking areas, and personal identification. A Traffic surveillance system usually contains two parts, hardware, and software. Hardware is a static camera installed on the roadside that captures the video feed and the software part of the system is concerned with processing and analyses.

The traditional method of target detection is generally divided into three phases: First, select a few candidate regions on a given image, and then extract features from these regions and classify them by trained classifiers. Here we introduce each of these three stages separately. Area selection is to locate the location of the target. As the target may appear anywhere in the image, and the size of the target, the aspect ratio is not sure, so the first sliding window strategy to traverse the entire image, and you need to set a different scale, different aspect ratio. For feature extraction, it is not so easy to design a robust feature due to the morphological diversity of the target, the diversity of light variations, and the diversity of the background. However, the quality of extracted features has a direct impact on the accuracy of classification.

To summarize, there are two main problems in traditional target detection: one is that the region selection strategy based on sliding window is not specific, the time complexity is high, and the window is redundant; secondly, the hand-designed features are not very good for the diversity change Robustness. For now, target detection based on traditional machine learning methods has encountered bottlenecks and a more scientific approach is expected. With the rapid development of deep learning theory and practice, the goal of machine learning based detection and classification has entered a new phase. Unlike traditional feature extraction algorithms that rely on prior knowledge, deep convolutional neural networks have some degree of invariance to geometric transformations, deformations, and illumination and effectively overcome the variability of vehicle appearance and are adaptive to training data Build feature descriptions for greater flexibility and generalization. For target detection, the recognition accuracy is an indicator that researchers want to improve all the time. Speaking of recognition accuracy, we must mention the mean average pre-measurement (mAP), which measures the detection accuracy in target detection.

The contribution of the paper for vision based vehicle counting and classification system is as follows. First, the background subtraction algorithm is applied to each frame to extract the moving region of interest. Second the proposal generation network to extract the bounding box and confidence score. Finally iterative refinement convolution neural network is applied to accurately detect and classify the vehicles. This paper is organized as follows. In section 2 is an survey of related works for object detection can classification using CNN . Section 3 explains the proposed approach. In section 4, the experimental outcomes are given. In section 5 conclusions are drawn.

II. Related Work

2.1 Video surveillance

The video surveillance systems can automatically monitor specific environments using advanced cameras and CCTV infrastructure. The motions and interactions of objects are analyzed by these systems. In [1], the abnormal behavior detection is explained. The system which is shaped like an ellipse with a head position is tracked to identify any change in posture is developed in [2]. The detection of burglary and robbery through consumer camera is proposed in [3]. The conversion of 2D to 3D space is done in the research [4]. The loitering customer or an unattended cash desk is monitored by surveillance more recently which is developed in [5]. Later, the harassment in public places, unauthorized access is detected by smart surveillance in [6]. Moving objects (car, truck. etc.) are detected and then classified according to features such as size, shape, pattern, color and so on. The quality of the video is must (i.e.) that must be free from noise and then occluded mostly. These are done in [7]. The existing methods are designed for daytime surveillance, but in the night, then performance would be less. This is overcome in [8].

2.2 Background Subtraction

Background subtraction is used to detect the moving objects in video streams. The main aim is to distinguish the moving object into background and foreground is proposed in [9]. Before performing complicated detection processes the objects are detected in [10]. In [11] the three stages of background subtraction is proposed as 1) background initialization 2) foreground detection 3) background maintenance. These are various models of the algorithm such as basic modeling, statistical modeling, cluster modeling, neural network background modeling. The difference between the background image and the present frame is adjusted based on the threshold. It uses a static image as a background. These are explained by the authors in [12] for the variation with continuous images. In statistical modeling, the given color spaces are defined with Gaussian distribution which is suggested by [13]. GMM (Gaussian Mixture Models) is the most widely used technique with moving background. But it does not perform well over high lightning, shadows etc. This is solved and improved in [14]. In [15] a new framework is proposed combining the thresholding and motion. In cluster modeling, suggested in [16] performs background subtraction by color and texture of input image with the fuzzy concept. In neural network background modeling, the networks are trained how to classify each pixel into background or foreground. In [17] a SOM network (Self Organizing Map) is used to perform background subtraction.

2.3 Convolutional Neural Network

CNN is a machine learning technique that is inspired by human brain. Many CNN architectures have emerged.eg AlexNet, VGNet, GoogleNet, ZFNet. This is designed to map image data to an output variable. These find applications in image classification, image and media recognition.

2.3.1 Convolutional Layer:

Feature detection and extraction happens in convolutional layer. Each layer has its own set of kernels. This consists of 2 convolutional layers, 2 pooling layers using max pooling function and 2 fully connected layer.

2.3.2 Activation Function:

After the convolutional layer, the activation function is applied. Mostly Rectified Linear Unit (ReLU).

2.3.3 Pooling Layer:

This follows the convolutional layer. This is used to eliminate unimportant information from feature maps. This is achieved by using a pooling function such as max pooling, average pooling. This pooling function is applied to feature map regions of size n*n. Omitting the max pooling layers does not ruin the accuracy.

2.3.4 Fully Connected Layer:

This is the final layers in CNN's. The computational units in a fully connected layer are connected to every unit of previous layer. This is one dimensional.

III. Proposed Methodology

Our proposed framework vision-based surveillance system consisting of a cascade of operations, each of which focuses on a specific duty, i.e. the background subtraction method extracts moving vehicles from the video frame as a region of interest (ROI), and then the ROI is passed through several convolution neural network and max-pooling layers to produce feature maps. The produced feature maps are given as input for the proposal generation network to generate candidate bounding box. Finally, the iterative refinement of feature maps refines the score and location of the generated bounding box. In the next subdivision, we explain the cascade of operations.

3.1Feature Extraction and Annotation of Moving Objects with Background Subtraction

The background subtraction algorithm is one of the widely used methods to extract moving objects from each video frame because of fast processing time. The Gaussian Mixture Model used as a primary algorithm since its works efficiently with a simple background subtraction method. Numerous post-processing steps are conducted to extract the region of interest. After eliminating shadow and applying morphological operation we obtain the desired region of interest by the rectangle. The extracted ROI is used as the input for the CNN classification. The process is given in Fig. 1.

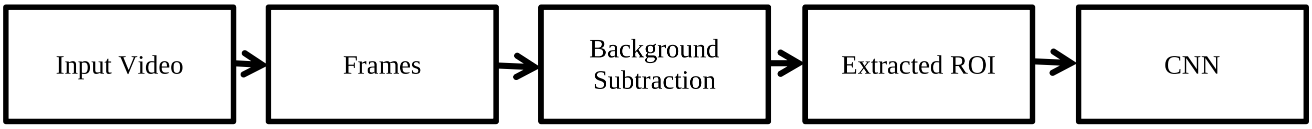

Figure1.Feature Extraction of Moving Objects with Background Subtraction

The extracted ROIs image size varies because all moving objects are the goal of extraction. The default input image size is set to 64 × 64 pixels because more than 50% of the extracted images are smaller. Table 1 shows the cumulative percentage of the images for every size. To handle the duplication problem, the subtraction outcome is saved once every 2s. Each vehicle is annotated as small, medium and large vehicle classes. The training and test set likely to have similar objects and backgrounds. We annotate as one of the three vehicle classes: small, medium and large. Sample images for each class are given separately and trained using MOT dataset. The image is cropped into a rectangle, so that unnecessary background or objects are frequently included in the image.

Table 1 Growing proportion of the extracted images of different sizes

Image size(smaller than)	32	64	96	128	160
Proportion (%)	2	41	65	77	86

3.2 Convolutional Neural Network for Object classification

The purpose of CNN is to classify the vehicles using the extracted ROI from background subtraction process. The default input images are set to 64 × 64 pixels of RGB color image of three vehicle categories: small, medium, and large. The network has five convolutional layers trained for class specific task. We drop the last classification layer and extract the convolution feature maps output, for the input for proposal generation network.The proposal generation network produces a set of object proposal, each of which has a predicted confidence score. The RPN proposed in [18] is arranged with convolution layer followed by regression layer and a bounding box classification layer. By doing this we can minimize the loss in [18] and the network can also be optimized.

3.3 Iterative Refinement

The iterative refinement in CNN relies on the VGG-16 model [19] that aims to get the input object and refine their bounding box location following the faster R-CNN detection pipeline [20]. The iterative refinement in CNN utilizes the ROI pooling layer to generate a fixed length feature descriptor of size 7×7×512 from the last pooled convolutional feature maps for every proposal provided by the proposal generation network Then, following the feature combination theme adopted in [21], we tend to concatenate every pooled feature descriptor on the channel axis and cut back the dimension with a one × one convolution to match the shape of 7× 7 × 512 required by the primary fully connected layer (fc6) of the pre-trained VGG-16 model. To match the initial amounts, every pooled feature map is L2 normalized and re-scaled make a copy by a hard and fast scale of a thousand. The generated feature is then fed into 2 fully-connected layers (fc6 and fc7) to predict the confidences over three classes, including three object categories, further because the bounding-box regression offsets. The parameters of those predictors are optimized by minimizing soft-max loss.

\

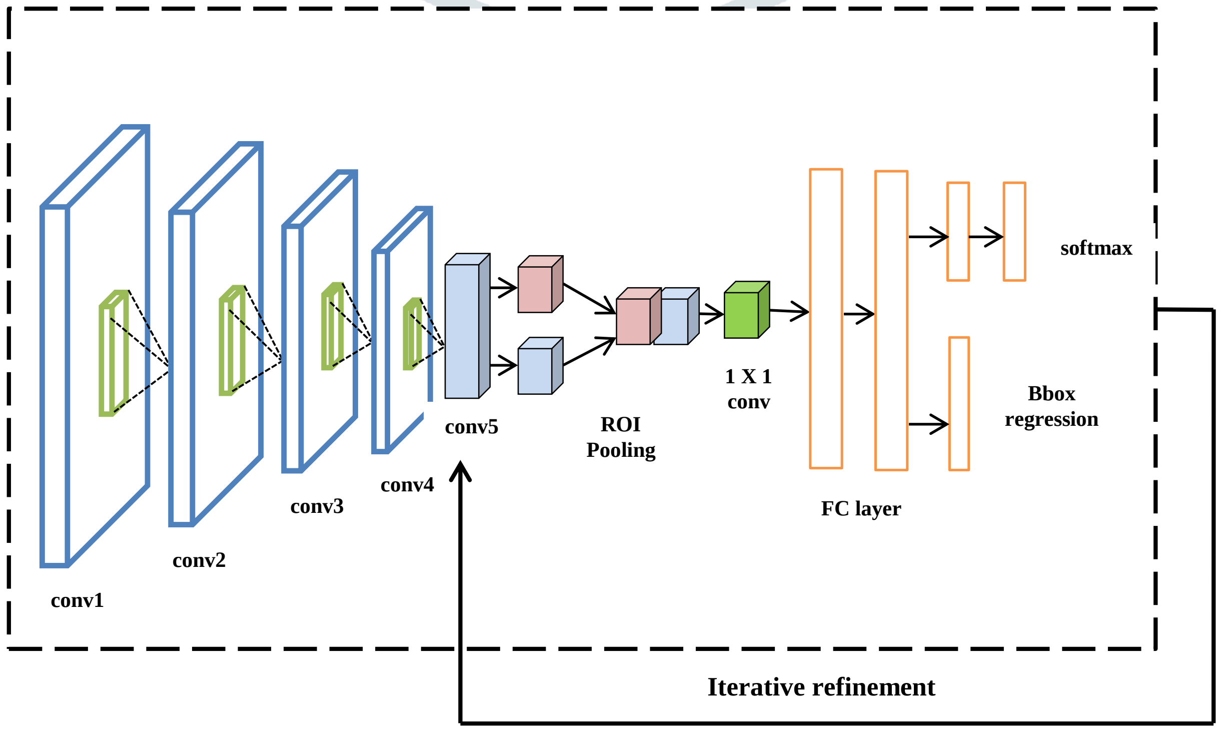

Figure 2 Iterative Refinement

IV. EXPERIMENT RESULTS AND ANALYSIS

Here the ZF model and VGG16 model of faster CNN algorithm are used to fine-tune the parameters in the model. In the paper, the original image data of training samples are from MOT dataset and the pictures of different types of vehicles on the network. The original data samples are processed to experimental requirements and then trained. We need to improve the accuracy of the target detection of the three kinds of vehicles such as cars, minibus, and truck with different networks and the different number of samples. The average accuracy (AP) (average precision) formed by the curve is commonly used evaluation indicators in the field of target detection Table 2. By increasing the convolution depth of model the number of extracted features can be increased. The experimental results using VGG16 detects more different types of vehicle targets more and more accurately shown in fig 3.

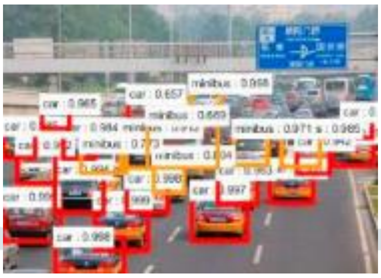

Figure 3 Classification of vehicles using VGG16

Table 2: comparison of average precision for various networks and different size samples

MODEL	5042(ZF)	5042(VGG16)
Car	84.0%	84.6%
Minibus	81.0%	84.1%
Truck	73.7%	79.1%

V. CONCLUSION

This paper proposes a framework for vision-based moving vehicle counting and classification system by combining background subtraction, proposal generation network and iterative refinement Convolution Neural Network (CNN). These approaches are applied in a cascade way to obtain the classification and counting. The experimental results show that compared with the traditional machine learning methods, the model used in this paper (i.e) Background subtraction has been improved both in average target detection accuracy and detection rate. The classification test result of this is suitable for vehicle type detection of three types of cars, minibus and truck ,in different scenarios and has achieved good results.

REFERENCES

[1] Mabrouk AB, Zagrouba E. Abnormal behavior recognition for intelligent video surveillance systems: a review. Expert Syst Appl. 2018;91:480–91.

[2] Foroughi H, Aski BS, Pourreza H. Intelligent video surveillance for monitoring fall detection of elderly in home envi- ronments. In: 11th international conference on computer and information technology, 2008. ICCIT 2008. New York: IEEE; 2008. p. 219–24.

[3] Lao W, Han J, De With PH. Automatic video-based human motion analyzer for consumer surveillance system. IEEE Trans Consum Electron. 2009;55(2):591–8.

[4] Chen DY, Huang PC. Motion-based unusual event detection in human crowds. J Vis Commun Image Represent. 2011;22(2):178–86.

[5] Arroyo R, Yebes JJ, Bergasa LM, Daza IG, Almazán J. Expert video-surveillance system for real-time detection of suspi- cious behaviors in shopping malls. Expert Syst Appl. 2015;42(21):7991–8005.

[6] Sidhu RS, Sharad M. Smart surveillance system for detecting interpersonal crime. In: 2016 International Conference on communication and signal processing (ICCSP). New York: IEEE; 2016. p. 2003–7.

[7] Valera M, Velastin SA. Intelligent distributed surveillance systems: a review. IEEE Proc Vis Image Signal Process. 2005;152(2):192–204.

[8] Huang K, Wang L, Tan T, Maybank S. A real-time object detecting and tracking system for outdoor night surveillance. Pattern Recog. 2008;41(1):432–44.

[9] Toyama K, Krumm J, Brumitt B, Meyers B. Wallfower: principles and practice of background maintenance. In: The Proceedings of the seventh IEEE international conference on computer vision, 1999, vol. 1. New York: IEEE; 1999. p. 255–61.

[10] Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. Comput Vis Image Underst. 2014;122:4–21.

[11] Bouwmans T. Background subtraction for visual surveillance: a fuzzy approach. Handb Soft Comput Video Surveill. 2012;5:103–38.

[12] Zheng J, Wang Y, Nihan N, Hallenbeck M. Extracting roadway background image: mode-based approach. Transp Res Rec J Transp ResBoard. 1944;82–88:2006.

[13] Staufer C, Grimson WEL. Adaptive background mixture models for real-time tracking. In: IEEE computer society conference on computer vision and pattern recognition, vol. 2. New York: IEEE; 1999. p. 246–52.

[14] Kaewtrakulpong P, Bowden R. An improved adaptive background mixture model for realtime tracking with shadow detection. In: Proceedings of 2nd European workshop on advanced video based surveillance systems. Dordrecht: Brunel University; 2001.

[15] Yeh C-H, Lin C-Y, Muchtar K, Lai H-E, Sun M-T. Three-pronged compensation and hysteresis thresholding for moving object detection in real-time video surveillance. IEEE Trans Ind Electron. 2017;64:4945–55.

[16] Zhang H, Xu D. Fusing color and texture features for background model. In: Proceedings 3 of the third international conference fuzzy systems and knowledge discovery, FSKD 2006, Xi'an, China, September 24–28, 2006. Berlin: Springer; 2006. p. 887–93

[17] Maddalena L, Petrosino A. A self-organizing approach to background subtraction for visual surveillance .IEEETrans Image Process. 2008;17(7):1168–77

[18] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, pages 91–99, 2015.

[19] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[20] Ross Girshick. Fast r-cnn. In CVPR, pages 1440–1448, 2015

[21] Sean Bell, C Lawrence Zitnick, Kavita Bala, and Ross Girshick. Insideoutside net: Detecting objects in context with skip pooling and recurrent neural networks. arXiv preprint arXiv:1512.04143, 2015