An Automatic Vehicle Type Classification and Counting based on Deep Learning in Traffic Environment

K. Kishore AnthuvanSahayaraj, Dr.K. Venkatachalapathy

Abstract--- A model for automatic vehicle type classification and counting based on deep learning is proposed to handle complex traffic scene. This model covers of parts, vehicle detection model and vehicle detection and classification and counting model.Faster R-CNN method is implemented in vehicle detection model to extract vehicle images from an image with disorder background which may contains numerousvehicles. In vehicle classification model, an image contains only one vehicle is fed into a CNN model to produce a feature, then a Non negative matrix factorization is used to implement the classification process. Experiments show that vehicle's detection and classification from traffic scenes can be recognized effectively by using our method. Furthermore, in order to build a large scale database easier, this paper comes up with a novel network collaborative annotation mechanism using iterative refinement in region proposal network

Keywords--- Faster R-CNN, Iterative, Non-Negative Matrix Factorization, Object Detection, Object Classification.

I. INTRODUCTION

With the promotion of Intelligent Transportation System(ITS) in intelligent town, core technologies applied in ITS develop speedily and are updated perpetually. In 1970s, solely magnetic coils were wont to observe vehicles, but now, different technologies like microwave radar, ultrasonic, infrared rays and video image area unit extremely popular in follow [1]. As a lot of and a lot of digital video surveillances are equipped to transportation roads, thus visual vehicle detection strategies became analysis problems with pc vision scientists recently [2].As associate degree application domain of object detection, vehicle detection plays associate degree necessary role in ITS, unmanned intelligent automotive and peace [3]. When detection, we will more classify them in additional detail, if applied publically security, it will helps to arrest criminals quickly.Although the criminals turn back, per automotive build, model, color and plate variety, we will begin all the cameras within the town, which may mechanically observe, acknowledge and find the automotive. During this scene, classification of auto is indispensable.

But actually, object's intra-class distinction is delicate, even generally intra-class distinction is greater than inter-class [4] [5], that the analysis subject of fine-grained classification is extremely difficult, which may advance the event of face detection [6], action recognition [7] and automatic scene description [8] and thenon. If classification of auto is applied in transportation and peace, we will acquire a lot of meta info like vehicle build, model, logo, production year, liquid ecstasy speed and acceleration and then on [9]. By feat these info dynamically, we will build an outsized intelligent transportation that may monitor the total city's road. Further, we will analyze the vehicles on the road at completely different time to seek out the discipline of people's going out, then we will schedule transportation rules consequently, these can build cities a lot of sensibleand intelligent.

II. RELATED WORK

A brief survey of the connected add intelligent traffic observation system exploitation traffic cameras is given within the section of Buch et al. [12]. Daigavane and Bajaj [13] given a background registration technique and segmentation employing a morphological operator. during this study, a system has been developed to notice and count objects dynamically on highways. The system effectively combines easy domain data concerning object categories with time domain applied math measures to spot target objects within the presence of partial occlusions and ambiguous poses. Chen et al. [14] addressed the problems relating to unsupervised image segmentation and object modeling with transmission inputs to capture the spacial and temporal behavior of the article for traffic observation. Gupte et al. [15] showed algorithms for vision-based detection and classification of vehicles in monocular image sequences of traffic scenes that area unit recorded by a stationary camera. process is finished at 3 levels: raw pictures, region level and vehicle level. Vehicles area unit modelled as rectangular patterns with bound dynamic behavior. Cheung and Kamath [16] compared the performance of an oversized set of various background models on urban traffic videos. They experimented with sequences recorded in climate like snow and fog, that a strong background model is needed. Kanhere et al. [17] applied a feature chase approach to traffic viewed from a low-angle off axis camera. Vehicle occlusions and perspective effects create an additional vital challenge for a camera placed low to the bottom. Deva et al. [18] projected a thought to mechanically track the articulations of individuals from video sequences. This can be a difficult task however contains a fashionable body of relevant literature.

Manuscript received January 25, 2019.

K. Kishore AnthuvanSahayaraj, Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram,Tamilnadu India. (e-mail:antonio.kishore@gmail.com)

Dr.K. Venkatachalapathy, Professor& Head, Department of Computer and Information Science, Annamalai University, Chidambaram,Tamilnadu, India.(e-mail:omsumeetha@rediffmail.com)

It will determine and track people and count distinct folks. Toufiq et al. [11] represented background subtraction because the widely-used paradigm for detection of moving objects in videos taken from a static camera that features a terribly big selection of applications.

The main plan behind this idea is to mechanically generate and maintain an illustration of the background, which may be later wont to classify any new observation as background or foreground. Gao et al. [20] analysis showed that a collection of Scale Invariant Feature rework (SIFT) options area unit extracted and matched within the follow-up image frames to enhance chase performance for additional correct detection. The SIFT options also are detected, tracked and clustered within the foreground blobs in Jun et al. [21]. Horizontal and vertical line options area unit extracted to make a 3D vehicle model at Leotta and Mundy's [22] work, assumptive the vehicle isn't occluded, by predicting and matching image intensity edges to suit a generic 3D vehicle model to multiple still pictures. concurrent chase can even be done throughout the form estimation in a very video. Ma and Grimson [23] projected a vehicle classification algorithmic rule that uses the feature supported edge points and changed SIFT descriptors.

Two classification tasks, cars versus minivans and sedans versus taxies, area unit tested with satisfactory performance. Buch et al. projected a 3D extended Histograms of orienting Gradients (HOG) feature for detection and classification of individual vehicles and pedestrians by combining 3D interest points and HOG. 3D vehicle models area unit pre-reconstructed by the strategies in Messelodi et al. [24]. Hsieh et al. [11] additional extracted region size and vehicle "linearity" to classify vehicles into four classes (e.g., car, minivan, truck and van truck), assumptive that individual vehicles are separated once lane and shadow detection. Alonso et al. [25] extracted image regions per high edge density areas.

Shadows, symmetry measuring and Harris corners area unit then employed in the hypothesis classification. Lou et al. [26] extracted image regions of interest supported motion detection. Then, a 3D model is fitted to the image region employing a point-to-line phase distance metric. Occlusion could be a major difficult drawback within the vehicle segmentation. several strategies are projected to influence this drawback.

Features mentioned higher than might be thought of as a collection of "parts" that area unit tracked and classified along [17]. once the 2D/3D vehicle model might be fitted into image frames, it's additionally comparatively straightforward to notice occlusions [27,28]. Liang et al. [29] extracted a "cutting region" between 2 occluded vehicles supported the motion field of consecutive image frames. Similarly, a "cutting line" is calculable in a very Traffic book [1] to separate 2 occluded vehicles supported the analysis of solid.

Image warp isn't thought of as a step or module to notice and track vehicles in Buch et al. [12]. there's some analysis mistreatment image warp as a pre-processing step to get a horizontal or vertical road phase to facilitate the detection and chase (e.g., [21,30]). Four reference points area unit elite to estimate a projective transformation in Jun et al. [13]. This transformation is applied in order that all motion vectors area unit about parallel to every different. The similar plan is applied in Salvi et al. [30] in order that lanes might be detected simply. However, image warp itself has not been applied on to notice unclassified vehicles.

This paper is organized as follows. Section 3 we come up with a unified model which can do the following process. Firstly, we use a collaborative network annotation mechanism to generate large annotated vehicles images that can build a continuously growing dataset. Secondly, we use an Iterative refinement Faster R-CNN model to detect vehicles in the existing dataset and then generate images bounding box on vehicle. At last, we use the CNN model and joint nonnegative matrix to classify and count vehicles.

III. METHODOLOGY

In this section, we come up with a unified model which can do the following process. Firstly, we use a collaborative network annotation mechanism to generate large annotated vehicles images that can build a continuously growing dataset. Secondly, we use an Iterative refinement Faster R-CNN model to detect vehicles in the existing dataset and then generate images bounding box on vehicle. At last, we use the CNN model and joint nonnegative matrix to classify and count vehicles.

The traffic scene images sometimes contain an unsure variety of vehicles. First, extract all the vehicles in the image in order to do the classification and then for these extracted images, we use a CNN model to extract features, with which we can classify vehicles easily. The overall unified model is as figure 1. This vehicle classification model takes an original image with a complex background as input, which is first fed into the vehicle detection model, then a sequence of bounding box contain only vehicle will be produced by detection model. The entire detected bounding box is then transferred to the next classification and counting model; at last, all the type and number of the vehicle can be achieved.

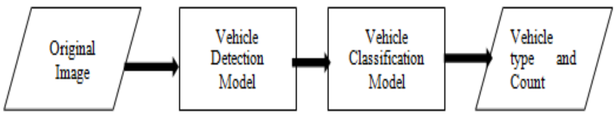

Figure 1: Overall unified model

1.1 Vehicle detection model

Vehicle detection model can detect all the vehicles in images with complex background and then extract them as bounding box. The function of this model is described as figure 2. Input image is first fed into a convolutional networkwhich uses VGG16 network arrangement as described in [31] [32]. And then feature of the Input image will be generated, on which aIterative refinement RPN network [33]isapplied to acquire region proposals.After that we use a ROI Pooling layer toobtain region proposals on the Input image accordingly, which then transferredto a vehicle LR classifier to judge whether these region proposals are vehiclesor not.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

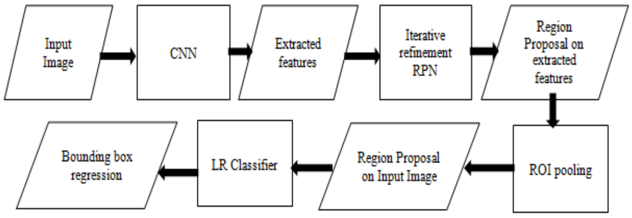

Figure 2: Vehicle detection model

1.1.1 Convolution Neural Network (CNN)

CNN is extremely effective within the field of image classification and recognition [34]. This type of Neural Network demands a large amount of memory to process the huge amount of training data. Now, we have high-end graphic cardsand hard drive and an extended volume of memory, which gives an extra advantage to run CNN algorithms powerfully. CNN has some layers of convolution with nonlinear activation functions; like ReLu, tanh, sigmoid, etc as described in Figure3.

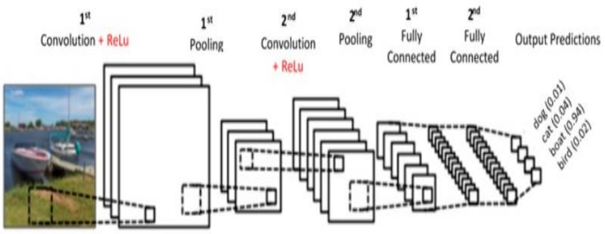

Figure 3: A simple CNN architecture

A region of input is connected to a region of output in the CNN; unlike old neural network it is fully connected. Each layer of CNN extract features based on the different filters applied. As the number of layer increases, the number of filters increases to extract more features. From direct pixels, the edges are detected at the first level. Then, the shapes are detected from the edges and further featureinformation is detected that is specific to a particular object, like the difference between trucks and vans.

There are four processes in the CNN:

a. Convolution

The Layers are the building blocks of the CNN. Features Extraction is the more important in these layers. The layers use different types of filters to extract more detailed information. The filters are responsible for extracting the features. Figure 4 describes the concept more clearly Let us assume, convolving a minor region of the image, which the filter size is 3 × 3, and, as there arefive filters, the length of the matrix after the operation will be 3 × 3 × 5

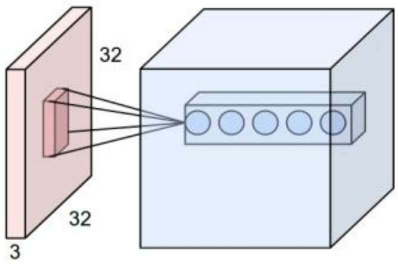

Figure 4: Convolving a minor region of the image, which the filter size is 3 × 3

b. Nonlinearity (ReLu)

The nonlinear function is applied for learning. The rectified linear unit (ReLu) is used widely among the neural network community because it solves the vanishing gradient problem. ReLu is a simple functionf(x) = max (0, x) applied after every convolution operation.

c. Pooling

The Pooling layer reduces the spatial size of the input volume, which in turn reduces the amount of parameters and computation in the network. Also Pooling helps to control the over fitting problem. Figure 5 shows the max pooling with a 2 × 2 filter and stride 2.

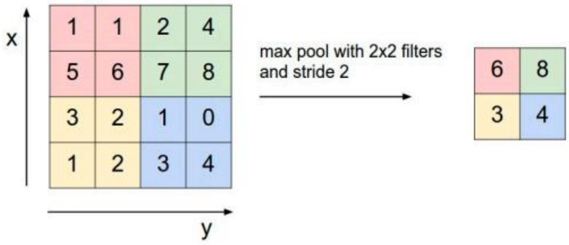

Figure 5: Max pooling over a 2 × 2 region with stride of 2

d. Fully-connected layer (FC)

FC is the decision layer and it is the final layer. Every layer is connected to the previous layer and also to the next layer, as shown in the figure 6.

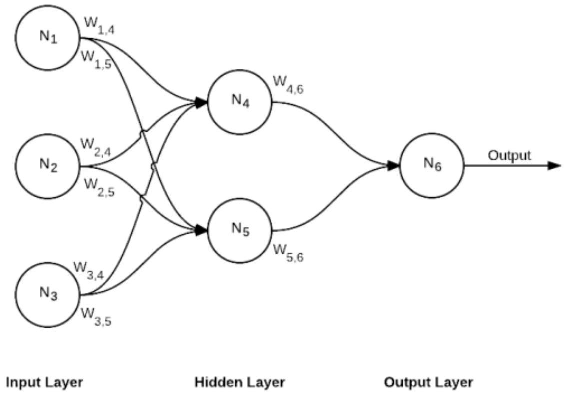

Figure 6: Fully connected Layer

1.1.2 Region Proposal Networks and Fast RCNN.

Faster RCNN proposed using Iterative refinement RPN network to extract the bounding box, compared with the Selective Search method to extract fewer candidates, more efficient. Using Fast RCNN network for network target detection and parameter training.

The RPN network takes a picture of random size as its input and outputs a batch of rectangular region boundingboxes, each region corresponding to a target's existing probability score and location information . In this study, the vehicle type detection problem is mainly to improve the accuracy in the iterative RPN layer in the Faster RCNN model.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The traditional RPN network generates 256 individual maps by 2-step convolution, in which 3 × 3 convolution the basis produces 192 feature maps, with 5 × 5 convolution kernels producing 64 feature maps. The iterative RPN network in this paper only produces 256 characteristic maps through a one-time 3 × 3 convolution kernel, which shortens the network structure and decreases the computational complexity. After the convolution operation, the last layer of the convolution feature graph is obtained, and the convolution kernel (sliding window) of 3 × 3 is used to convolve the feature graph on this feature graph. Because in this 3 × 3 region, each feature map to obtain a 1-dimensional vector, the last layer of the convolution layer 256 features, so this 3 × 3 region convolution can be a 256-dimensional feature vector, followed by classification layer cls layer and regression layer reg layer were used for classification and border (Bounding Box) regression. At the vital point of the 3 × 3 convolution kernel sliding window, there are 9 region proposals corresponding to three types of scales (eg128,256,512) and three aspect ratios (1:1,1:2,2:1) The process of this mapping is called anchor, which generates nine candidate windows. Lastly, according to region proposal score level, select the top score of 300 region proposal, as Fast RCNN input target detection.

To train RPNs, this paperallocates each anchor a binary label (whether or not it is a type of vehicle target). If thisanchor and a Ground Truth GT (Ground Truth) overlap ratio IoU (Intersection-over-Union) greater than 0.7,recorded as a positive sample; if it and any one of the calibration overlap ratio is less than 0.3, denoted as negativesamples. With these definitions, this paper uses the objective function reducing multitasking loss in Fast R-CNN.

1.2 Vehicle classification and counting model

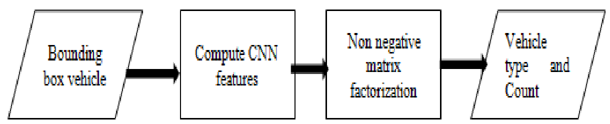

Figure 7: Vehicle classification and counting model

Vehicle type classification model use an image which contains only vehicle as an input to another classification CNN model to generate features, then with a joint Logistic regression, the contained vehicle will be correctly Classified [35]. As shown in figure 7. CNN feature and Nonnegative Matrix factorizationhave been proved to be effective in character recognition as described in [33]. We treat a vehicle as two parts, one part is the inter-difference of different vehicle models, and another part is the intra-difference of the same vehicle model, vehicle images taken from different views for example.

IV. EXPERIMENT

In this paper, the ZF model (Zeiler and Fergus) [18] and VGG16 model [19] of Faster R-CNN algorithm areused to train and fine tune the parameters in the network model. Firstly, it is initialized with the ImageNet [20]model to train an RPN network independently. Then the Fast-RCNN network is initialized with the network weight of this RPN and the proposal of the previous RPN is used as the input of the Fast-RCNN to train the Fast R- CNNnetwork; re-initialize the parameters of the RPN network using the parameters of the trained Fast R-CNN networkand fine-tune the RPN network; and then fix the convolution layer of the Fast R-CNN and use the candidate boxRPN to extract Fast R The CNN network is fine-tuned; afterwards, this process is iteratively training the RPN, FastRCNN, and eventually converging the network.

In this paper, the original image data of the training samples are from the MIT and Caltech vehicle databases and the pictures of different types of vehicles on the network. The original data samples are then processed according tothe experimental requirements and then trained. At last, we choose all kinds of scenes to detect the images of allkinds of vehicles, and get the accuracy of the target detection of the three kinds of vehicles such as cars, minibus andSUVs with different networks and different numbers of samples.

For different depth learning network models (including ZF and VGG16) and different numbers of samples, the recall curves of the experimental results of three different types of vehicles are shown in Fig. 6. Table 2 shows the results of different types of vehicles in different sample numbers and different the average accuracy of the network.Accuracy in the curve precision, recall recall, and area average accuracy AP (Average Precision) formed by thecurve are commonly used evaluation indicators in the field of target detection.. In the case of the same number of samples,network detection with more convolutional layers is better.According to the result analysis, if the number of samples is increased moderately, the recognition accuracy of all kinds of vehicles can be increased. However, if the sample size isincreased, there will be over fitting, and the detection target will be more or the position of the frame will beinaccurate. Depth networks have more convolution layers and more features to be extracted. By increasing theconvolution depth of the model (changing from ZF to VGG16), the number of extracted features can be increased.

The experimental results show that the depth network model (VGG16) detects more different types of vehicletargets more and more accurately.In this paper, the detection model under the same conditions with the Fast RCNN and Faster RCNN in the abovedata set for the detection time comparison, as shown in Table 3. It can be seen that under the same conditions, by improving the RPN network and reducing the complexity of the algorithm of the model, this method can obviouslyimprove the vehicle type detection time without reducing the effect of Faster RCNN in detection and meeting therequirements of traffic scenarios Vehicle target detection.

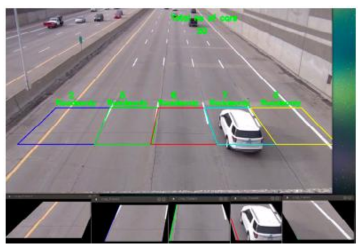

Figure 8: Unified model result

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

V. CONCLUSION

In this paper, the application of convolution neural network in vehicle target type detection and counting isimplemented. The Faster RCNN model is applied to the actual traffic environment. We focus on training ZF andVGG16 networks to detect vehicle types in traffic scenarios. The experimental results show that compared with thetraditional machine learning methods, the model used in this paper has been improved both in average targetdetection accuracy and detection rate. The classification test result of this article is also suitable for vehicle typedetection in different scenarios and has achieved good results.

REFERENCES

1. Editorial, Special issue on big data driven intelligent transportation system Neurocomputing (2016) .
2. D. Tao, Y. Rui, M. Wang, Learning to rank using user clicks and visual features for image retrieval, IEEE Transactions on Cybernetics (2015)767–779.
3. Q. Ge, T. Shao, C. Wen, R. Sun, Analysis on strong tracking filtering forlinear dynamic systems, Mathematical Problems in Engineering (2015) 1–9.
4. Z. Lu, L. Wang, J. Wen, Image classification by visual bag-of-words refinement and reduction, Neurocomputing (2016) 373–384.
5. L. Xie, J. Wang, B. Zhang, Q. Tian, Incorporating visual adjectives forimage classification, Neurocomputing (2016) 48–55
6. S. A. A. Shah, M. Bennamoun, F. Boussaid, Iterative deep learningimage set based face and object recognition, Neurocomputing (2016) 866- 874.
7. N. Nedjah, F. P. Silva, A. O. Sa, L. M.Mourelle, D. A.Bonilla, A massivelparallel pipelined reconfigurable design for m-pln based neural networks for efficient image classification, Neurocomputing (2016) 39–55.
8. Q. Ge, D. Xu, C. Wen, Cubature information filters with correlated noisesand their applications in decentralized fusion, Signal Processing (2014) 434–444.
9. Q. Ge, C. Wen, S. Duan, Fire localization based on range-range-rangemodel for limited interior space, IEEE Transactions on Instrumentation and Measurement (2014) 2223–2237
10. Traffic Scorecard. INRIX. Available online: http://inrix.com/ (accessed on 5 May 2017).
11. Hsieh, J.-W.; Yu, S.-H.; Chen, Y.-S.; Hu, W.-F. Automatic traffic surveillance system for vehicle tracking and classification. IEEE Trans. Intell. Transp. Syst. 2006, 7, 175–187.
12. Buch, N.; Velastin, S.A.; Orwell, J. A review of computer vision techniques for the analysis of urban traffic.IEEE Trans. Intell. Transp. Syst. 2011, 12, 920–939.
13. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. OverFeat: Integrated Recognition,Localization and Detection using Convolutional Networks. arXiv. 2013. Available online: https://arxiv.org/abs/1312.6229 (accessed on 6 February 2017).
14. Daigavane, P.M.; Bajaj, P.R. Real Time Vehicle Detection and Counting Method for Unsupervised TrafficVideo on Highways. Int. J. Comput. Sci. Netw. Secur. 2010, 10, 112–117.
15. Chen, S.C.; Shyu, M.L.; Zhang, C. An Intelligent Framework for Spatio-Temporal Vehicle Tracking.In Proceedings of the 4th IEEE Intelligent Transportation Systems, Oakland, CA, USA, 25–29 August 2001.
16. Gupte, S.; Masoud, O.; Martin, R.F.K.; Papanikolopoulos, N.P. Detection and Classification of Vehicles.IEEE Trans. Intell. Transp. Syst. 2002, 3, 37–47.
17. Cheung, S.; Kamath, C. Robust Techniques for Background Subtraction in Urban Traffic Video. In Proceedingsof the Visual Communications and Image Processing, San Jose, CA, USA, 18 January 2004.
18. Kanhere, N.; Pundlik, S.; Birchfield, S. Vehicle Segmentation and Tracking from a Low-Angle Off-AxisCamera. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and PatternRecognition, San Diego, CA, USA, 20–25 June 2005.
19. Deva, R.; David, A.; Forsyth, D.A.; Andrew, Z. Tracking People by Learning their Appearance. IEEE Trans.Pattern Anal. Mach. Intell. 2007, 29, 65–81.
20. Toufiq, P.; Ahmed, E.; Mittal, A. A Framework for Feature Selection for Background Subtraction.In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,New York, NY, USA, 17–22 June 2006.
21. Gao, T.; Liu, Z.; Gao, W.; Zhang, J. Moving vehicle tracking based on sift active particle choosing.In Proceedings of the International Conference on Neural Information Processing, Bangkok, Thailand,1–5 December 2009.
22. Jun, G.; Aggarwal, J.; Gokmen, M. Tracking and segmentation of highway vehicles in cluttered and crowded scenes. In Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision, Copper Mountain, CO, USA, 7–9 January 2008; pp. 1–6.
23. Leotta, M.J.; Mundy, J.L. Vehicle surveillance with a generic, adaptive, 3D vehicle model. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 33, 1457–1469.
24. Ma, X.; Grimson, W.E.L. Edge-based rich representation for vehicle classification. In Proceedings of the10th IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2,pp. 1185–1192.
25. Messelodi, S.; Modena, C.M.; Zanin, M. A computer vision system for the detection and classification ofvehicles at urban road intersections. Pattern Anal. Appl. 2005, 8, 17–31. [CrossRef]
26. Alonso, D.; Salgado, L.; Nieto, M. Robust vehicle detection through multidimensional classification for onboard video based systems. In Proceedings of the 2007 IEEE International Conference on Image Processing,San Antonio, TX, USA, 16 September–19 October 2007.
27. Lou, J.; Tan, T.; Hu, W.; Yang, H.; Maybank, S.J. 3-D modelbased vehicle tracking. IEEE Trans. Image Proc. 2005, 14, 1561–1569.
28. Gentile, C.; Camps, O.; Sznaier, M. Segmentation for robust tracking in the presence of severe occlusion.IEEE Trans. Image Proc. 2004, 13, 166–178.
29. Song, X.; Nevatia, R. A model-based vehicle segmentation method for tracking. In Proceedings of the10th IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2,pp. 1124–1131.
30. Liang, M.; Huang, X.; Chen, C.-H.; Chen, X.; Tokuta, A. Counting and Classification of Highway Vehicles by Regression Analysis. IEEE Trans. Intell. Transp. Syst. 2015, 16, 2878–2888.
31. R. S, H. K, G. R, Faster r-cnn: Towards realtime object detection withregion proposal networks, NIPS (2015) 442–451.
32. Y. Zhang, L. Zhang, P. Li, A novel bilogically inspired elm-based network for image recognition, Neurocomputing (2016) 286–298.
33. TolgaEnsari"Character Recognition Analysis with Nonnegative Matrix Factorization"International Journal of Computers, Volume 1, 2016,pp219-222.

1. J. Shotton, J. Winn, C. Rother, A. Criminisi, Textonboost: Joint appearance, shape and context modeling for multi-class object recognition andsegmentation, ECCV (2006) 1–15.
2. Rakesh N. Rajaram, EshedOhn-Bar, and Mohan M. Trivedi, RefineNet: Iterative Refinement for Accurate Object Localization016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).