Advances in Engineering Research (AER), volume 124
2nd International Symposium on Advances in Electrical, Electronics and Computer Engineering
(ISAEECE 2017)

Context-Aware Object Region Proposals for Efficient Vehicle Detection from Traffic Surveillance Videos Using Deep Neural Networks

Jianhe Yuan[1, a], Wenming Cao[1], Fangfang Lv[1*]

[1]College of Information Engineering, Shenzhen University, Shenzhen, China

[a]2141130410@email.szu.edu.cn

*corresponding author E-mail: 1517799809@qq.com

Keywords: Region Propose, Vehicle Detection, Image Segmentation, Traffic Surveillance, Deep Convolutional Neural Network (DCNN)

Abstract. Recently, many methods based on deep neural networks have been developed for object recognition, which dominate various performance competitions on public datasets such as ImageNet and Pascal VOC. Existing methods suffer from high computational complexity and/or insufficient recognition accuracy for practical use. In this paper, we demonstrate that, in specific application domains, such as traffic video surveillance, the priori knowledge or environmental context information can be utilized to dramatically reduce the computational complexity and improve the object detection performance. Specifically, our method models the traffic scene background, using the model as a context to guide the generation of a much smaller number of high quality object region proposals that maintain 100% coverage. We then train a deep convolutional neural network (DCNN) to classify these proposal regions and have achieved 99% accuracy on a large test dataset, which outperforms existing methods DCNN-based methods, such as YOLO.

Introduction

Automated vehicle detection and tracking from aerial surveillance platforms, such as drones, has many important applications, such as traffic flow control, traffic jam prediction and transportation statistics analysis [3]. Recently, many methods based on deep neural networks, such as RCNN and YOLO (You Look Only Once) [1, 2], have been developed for object recognition, which dominate various performance competitions on public datasets such as ImageNet and Pascal VOC. In this paper, we demonstrate that, in specific application domains, such as traffic video surveillance, the priori knowledge or environmental context information can be utilized to dramatically reduce the computational complexity and improve the object detection performance.

This work is also related to object detection. Recently, selective Search [2] proposes a set of bounding boxes and scores those patches using a convolutional neural network. However, many bounding boxes are generated to ensure sufficient object coverage. The Region-based Convolutional Network method (R-CNN) applies high-capacity convolutional networks (CNNs) to bottom-up region proposals to localize and segment objects [5]. Szegedy et al. [4] propose a saliency-inspired convolutional neural network model to predict regions of interest.

Proposed Vehicle Detection Framework

We observe that existing methods suffer from high computational complexity and the recognition accuracy are still inadequate for successful use in practice. In this section, we explain our proposed framework for context-aware vehicle detection using DCNN. As illustrated in Fig. 1, the proposed approach has the following four major components: segmentation, merging procedure, patches grouping, and DCNN classification.

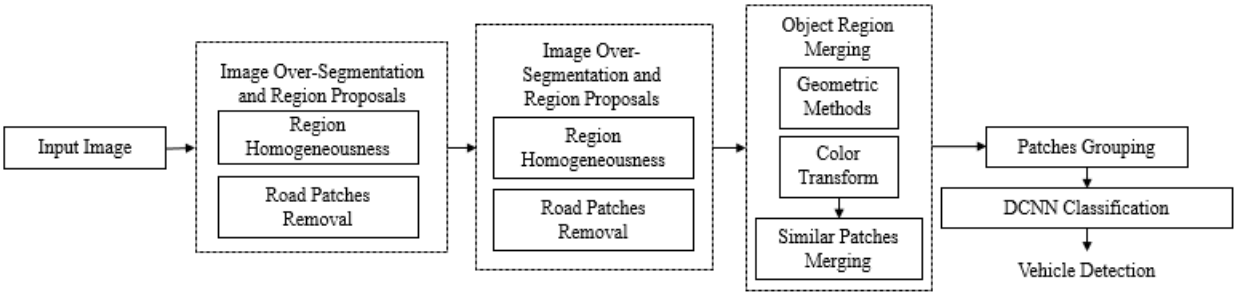

Fig. 1. The framework of proposed approach.

Homogeneous Regions and Over Segmentation. We perform homogeneousness analysis of local image regions to achieve over segmentation of the input image. The goal of region homogeneousness analysis is to construct a set of texture regions for vehicle object proposals. To this end, we propose an adaptive connected component approach. A pixel X is connected to pixel Y if their distance metric $\delta(X,Y) < \Delta(X,Y)$. Here,

$$\Delta(X,Y) = \lambda \cdot \sigma(X,Y) \quad (1)$$

is an adaptive threshold related to the local noise level of the image at location (X,Y). One approach to estimate the local noise level is to use the energy of high-frequency components after spatial transform, such as discrete cosine transform (DCT). We use 8-connectivity for adaptive connected components analysis in the YUV color space, which will produce a set of small image regions from the input image.

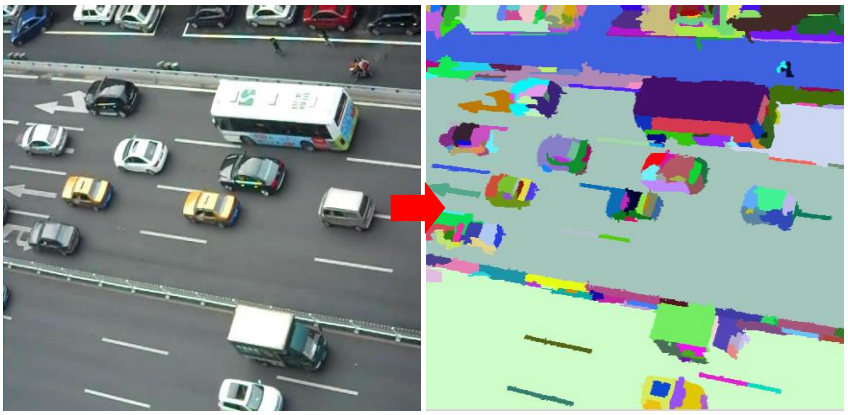

Fig. 2. The result of adaptive connected component analysis.

Fig. 2 shows one example of the output. Certainly, the parameter λ will control the granularity of the image regions. Removing road regions that covers most parts of the entire image, we can not only concentrate on the rest of regions but also prevent the regions grouping between vehicles and the road.

Object Region Grouping. The goal of object region grouping is to combine a much larger vehicle patches so that generate a much smaller number of bounding boxes. Our merging procedure works as follows. First all the patches size is calculated. For the small size regions, the process of searching the large size patch from neighboring regions is repeated until the patch is larger than threshold. Then we merge these two regions together. Several connected regions may form a vehicle patch due to similarity in color or texture. Therefore, we perform the merging procedure using Chi-Square histogram distance comparison. Given the neighboring patch pairs (r_n, p_m), the similarity d_{nm} can be computed.

Vehicle Object Proposals Generation. After the segmentation and merging procedure, vehicle and other objects can divide into several patches. Furthermore, some vehicles may have no clear boundaries or even overlap with neighboring vehicles. Therefore, we group neighboring patches together using the following algorithm in Table 1. For each neighboring patch pair S_k (r_i, p_j)

between patch r_i and p_j, we want to group them respectively and record the new grouped patch. Furthermore, the neighbor of new grouped patch should be based on the initial patches P, which means that the new grouped patch r_i needs to be grouped with initial neighboring patch p_j.

Table 1. Vehicle object proposals generation

1. Obtain initial patches $P = \{p_1, \ldots, p_n\}$using the region homogeneousness analysis
2. Set the grouping patches $R_0 = P$
3. Initialize neighboring patches set S_0 $S_0 = \{s_0(r_1, p_1), \ldots, s_0(r_n, p_m)\}$
4. Set the number of grouping patches C
While $C \neq 0$ do
For each neighboring patch pair S_k (r_i, p_j)
Group the neighboring patch $R_t = r_i \cup p_j$
Record the new neighboring patches set S_i
$S_k = \{s_k(r_1, p_1), s_k(r_1, p_2), \ldots, s_k(r_n, p_m)\}$
Remove the same sets of neighboring patches
Decrease C progressively
Create bounding boxes per patch R_t

Experimental Results

Dataset. In this section, we evaluate the performance of our approach and present the experimental results. To analyze the performance of the proposed approach, different scenes from video sequences are used. We compare our approach with YOLO (You Only Look Once). We establish a dataset for vehicle object detection from aerial video surveillance with 200 test images and manually labelled ground truth for vehicle objects.

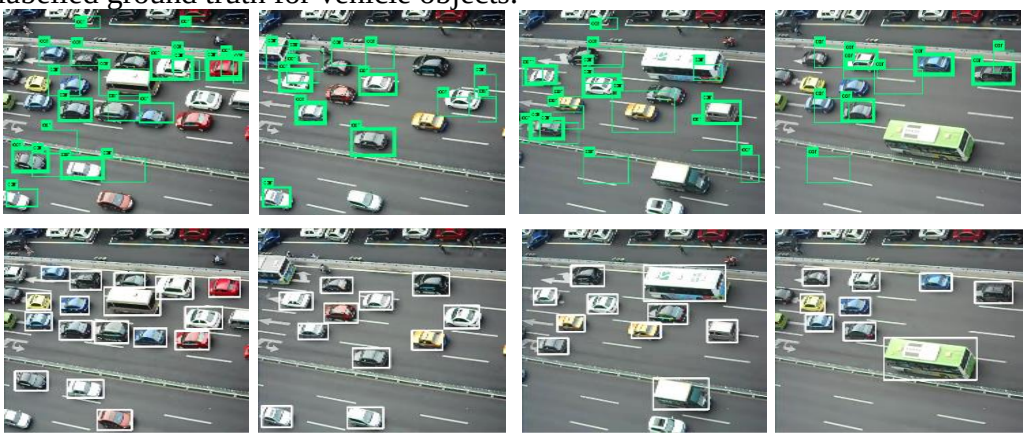

Fig. 3. The results of YOLO and our approach.

Comparisons with YOLO. We compare the performance of our algorithm with YOLO [1], which represents the state-of-the-art object detection method based on DCNN. For evaluating YOLO on our dataset, we fine-tuned the YOLO model. Fig. 3 shows some examples of object detection results produced by our algorithm (bottom row in white) and YOLO (top row in green). As we know, YOLO presents a good performance on classifying many different classes in a same image. But, for the specific domain application, such as the traffic surveillance, the accuracy can be further significantly improved. In addition, YOLO suffers from performance degradation for small objects that appear in groups, such as many vehicles in the same image. We see that the YOLO performs

not well in locating vehicles. Fig. 4 shows the precision-recall curves of our method and YOLO. We can see that our algorithm significantly outperforms YOLO. For the specific traffic, aerial images, YOLO is having a hard time to detect vehicles of different sizes. Therefore, the precision of YOLO considerably dropped with the recall rate. Although the number of bounding box is more than YOLO, our approach achieves a sufficient coverage, even the recall rate reaches to 0.9.

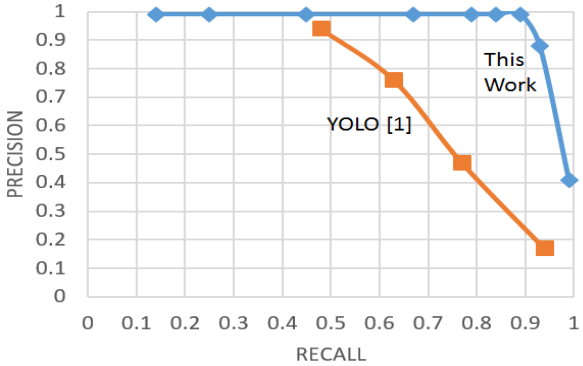

Fig. 4. Precision-recall curve for vehicle object detection.

Comparison to Selective Search. In this section, we compare our algorithm against Selective Search [2] in terms of the number of vehicle object proposals. Selective search uses region proposals to find objects in images and it performs hierarchical grouping of local image regions. However, to capture the objects at different scales, selective search often produces a very large number of bounding boxes per image.

Our approach shares some similarities with selective search. For specific domain application, we use much effective region proposals to detect vehicle. Our approach maintains a much smaller number of bounding boxes, only 827 per image compared to about 2147 by the Fast-Selective Search. Furthermore, our approach covers all vehicle objects with much fewer bounding boxes.

Conclusion

This paper presents an efficient approach for vehicle detection in traffic image. We model the traffic surveillance scene using image region homogeneousness analysis and use this context information to guide the generation of object region proposals. We demonstrate that the priori knowledge or environmental context information can be utilized to dramatically reduce the computational complexity and improve the object detection performance. The proposed method outperforms existing methods, such as YOLO and Selective Search in terms detection accuracy and complexity.

Acknowledgements

This work was supported by National Natural Science Foundation of China under (NSFC) Grant No. 61375015.

References

[1] J Redmon, S Divvala, R Girshick, A Farhadi: You Only Look Once: Unified, Real-Time Object Detection. Computer Vision and Pattern Recognition, Vol. 4, p. 1-10, 2015.

[2] J. R. Uijings, K. E. van de Sande, T. Gevers, and A. W. Smeulders: Selective Search for Object Recognition, International Journal of Computer Version, Vol. 104, p. 154-171, 2013.

[3] H. -Y. Cheng, C. -C. Weng, and Y. Chen: Vehicle detection in aerial surveillance using dynamic Bayesian networks, IEEE Trans. Image Process, Vol. 21, p. 2152-2159, 2012

[4] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov: Scalable Object Detection Using Deep Neural Networks, Computer Vision and Pattern Recognition, p. 2155-2162, 2014

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Computer Vision and Pattern Recognition, p. 580-587, 2014.