Received June 27, 2019, accepted July 6, 2019, date of publication July 10, 2019, date of current version August 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927866

Multi-Task Cost-Sensitive-Convolutional Neural Network for Car Detection

XIAOMING XI[1,2], ZHILOU YU[2,3], ZHAOLEI ZHAN[2], YILONG YIN[4], AND CUIHUAN TIAN[5,6]

[1]School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China
[2]Chaoyue Digital Electronics Co., Ltd., Jinan 250013, China
[3]School of Information Science and Technology, Southeast University, Nanjing 211189, China
[4]School of Software Engineering, Shandong University, Jinan 250100, China
[5]School of Medicine, Shandong University, Jinan 250100, China
[6]Health Management Center, Qilu Hospital of Shandong University, Jinan 250000, China

Corresponding authors: Zhilou Yu (yuzhl@inspur.com) and Cuihuan Tian (shandamla@163.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61701280, in part by the Natural Science Foundation of Shandong Province under Grant ZR2016FQ18 and Grant ZR2018BF012, in part by the Natural Science Foundation of China under Grant 61801263, Grant 61671274, Grant 61703235, and Grant 61701281, in part by the National Science Fund for Outstanding Young Scholars under Grant 61622114, in part by the Shandong Provincial Key Research and Development Plan under Grant 2017CXGC1504, and in part by the Foundation of Distinguished Associate Professor in Shandong Jianzhu University.

ABSTRACT This paper proposes a novel smart parking scheme for the parking lot. Automatic car detection is the core technology of the proposed scheme. However, new challenges arise in car detection in aerial views, such as a large number of tiny objects and complex backgrounds. In order to solve these issues, this paper proposes a car detection method based on multi-task cost-sensitive-convolutional neural network (MTCS-CNN). In the proposed network framework, multi-task partition layer which is composed of some sub-task selection units is first developed. The sub-task selection unit is constructed by introducing local mask and non-zero pooling, which can divide the complex detection task into many simple sub-tasks. To tackle the obtained sub-tasks, cost-sensitive sub-network is proposed based on faster R-CNN framework with the introduction of cost-sensitive loss. In the proposed Multi-task partition layer, the sub-task selection unit is used to capture the local map of the original aerial view image. In each local map, the scale and the number of objects are enlarged and decreased, respectively. Therefore, multi-task partition layer can divide a complex tiny objects detection task into many simple enlarged objects detection sub-tasks, which is helpful for performance improvement. In addition, the proposed cost-sensitive loss can effectively discount the effect of easy examples and focus attention on the hard examples, which may improve the detection performance on hard examples. Therefore, the model with incorporation of proposed cost-sensitive loss is robust to the complex background, further improving the performance. On our dataset, the proposed method obtained an mAP of 85.3%, outperformed state-of-the-art method.

INDEX TERMS Smart parking, car detection, multi-task cost-sensitive-convolutional neural network.

I. INTRODUCTION

With the increasing number of cars in metropolises, finding an available parking spot has become a conflicting and frustrating problem for drivers [1], [2]. In order to address this problem, smart parking is proposed. In smart parking scheme, car detection in park spot is crucial.

Traditional approaches utilize various of sensors such as inductive loop detectors, radar detectors, laser detectors to detect cars [3]. In these approaches, a sensor should cover a park spot. However, the cost of these schemes will be increased for large parking lot because large number of parking spots will result in large number of sensors.

In order to reduce the cost, this paper develops a novel smart parking framework, in which automatic car detection algorithm is developed to substitute sensors. In the proposed framework (as shown in Fig.1), few cameras are firstly placed at the top of building to capture the image of parking lot. After that, the automatic car detection technology is used for automatic car detection. Based on the detection result, the available parking lot can be obtained. Finally, the parking lot analysis result is given to the users. To be noticed, automatic car detection algorithm is the core of proposed smart parking system. Therefore, this paper mainly focuses on the development of automatic car detection algorithm.

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

FIGURE 1. Flowchart of smart parking in parking lot.

FIGURE 2. Image example of the parking lot.

Car detection is the fundamental step of many applications, such as intelligent driving, Intelligent Transport Systems and so on. The existing methods mainly include motion-based methods, shallow learning -based methods and deep learning -based methods. Among these methods, deep learning has ability to learn more effective representation of data [4].Therefore, deep learning -based methods outperformed traditional methods and made remarkable progress in the development of car detection.

However, compared with traditional car detection in ground view images, in our task, the images are captured by the cameras at the top of build, which is in aerial view. Thus, new challenges are raised (as shown in Fig.2): (1) the objects have smaller scale. Moreover, there are large number of objects in each image. In deep learning framework, the feature maps which contain effective features of objects are learned by employing convolution operation. However, small scale and large number of objects may result in important information loss of some objects in convolution operation procedure, degrading performance. (2) The background is complex. As shown in Fig.2, the background contains different objects, such as car, tree, sky, building and so on. Different objects may be affected with each other. For example, occlusion and illumination variations appeared in the image. Some cars are occluded by trees or their neighboring cars. Therefore, the complex background may result in the performance degradation.

In order to address these issues, this paper proposes a car detection method based on Multi-Task Cost-sensitive - Convolutional Neural Network (MTCS-CNN). In the proposed network, two novel units are introduced: (1) Multi-task partition layer. This layer is composed of some sub-task selection units. The sub-task selection unit is constructed by introducing local mask and non-zero pooling to adaptively generate different local maps. After processing of multi-task partition layer, the complex task that car detection in original image is formulated into the some sub-tasks that car detection in local maps. In each sub-task, the scale and the number of objects in local map are larger and smaller than original task, reducing the complexity of tasks. Therefore, multi-task partition layer is able to divide the single complex detection task into several simple sub-tasks, which may avoid performance degradation. (2) Cost-sensitive sub-network. In order to tackle the sub-tasks, cost-sensitive sub-network is proposed. The cost-sensitive loss is developed to reduce the influence of complex background such as occluded cars. The proposed sub-network is constructed based on Faster R-CNN framework with incorporation of cost-sensitive loss, further improving the detection performance. We conduct the experiments on our database and the experimental results demonstrate the effectiveness of the proposed method.

We summarize the contribution of this paper as follows:

(1)A novel smart parking scheme which is dependent on automatic car detection is proposed to reduce the cost.

(2)Multi-task cost-sensitive convolutional neural network is proposed for automatic car detection. In the proposed network, multi-task partition layer and cost-sensitive sub-network is introduced to address the problem of large number of tiny objects and complex background.

(3)The proposed method outperforms state-of-the-art method of deep learning framework in our application scene.

II. RELATED WORK

Car detection has been a topic of great interest to researchers over the past decade [5], [6]. A variety of methods have been proposed for this task. The existing methods can be divided into 3 categories: (1) motion-based approach (2) shallow learning -based approach (3) deep learning -based approach.

Motion-based approach: These approaches aim to detect moving cars. The idea is to separate the moving object from the static background by mining the valuable motion information. Cucchiara.et.al has proposed Sakbot, a moving car detection method which can address the problem of cast shadows and ghosts [7]. In order to achieve reliable detection of overtaking vehicles, Zhu.et.al has proposed a robust car detection method by integrating variable bandwidth density fusion and multiscale mean shift, which can obtain reliable motion estimation and improve the detection performance [8]. Shen.et.al has proposed a dynamic visual model which can be used to detect critical motions of nearby vehicles while driving on a highway [9]. Considering that optical flow [10]–[12] is an useful tool for moving car detection, Arrospide.et.al has proposed a robust car detection method by combining optical flow and symmetry tracking [13].

Shallow learning -based approach: These methods aim to formalize the detection task into the classification task. In this framework, the cars are regarded as the positive instances while other objects are regarded as the negative instances. Effective handcrafted features and shallow classifiers have been applied for positive instances classification. Based on the idea, typical features such as Haar-like features [14],HoG [15],SIFT [16] and classifiers such as adaboost [17], artificial neural network [18], [19] have been used for car detection. Chen et.al has presented a multiorder feature descriptor and a fast sparse representation classification method for car detection [20]. In order to address insufficient samples problem of car detection in satellite images, Cao et.al has proposed a weakly supervised car detection with multi-instance discriminative learning [21]. In addition, the parts-based model [22], [23] is also a useful framework for car detection. Niknejad et.al has proposed a novel multicar detection method based on deformable parts-based model [24]. Satzoda et.al has presented a multipart-based car detection method with active learning and symmetry-based regression model [25].

Deep learning -based approach: In the recent years, deep convolutional neural networks (CNN) has achieved breakthrough in car detection area [26]–[34]. As an end to end framework, deep learning eliminates the handcrafted feature engineering of shallow learning-based method, instead of a network which contains feature learning and detection simultaneously. As a typical detection method, R-CNN is proposed by introducing Convolutional Neural Network(CNN) for feature extraction and achieves amazing detection performance [29]. In order to solve the problem of fixed size input requirement, SPP-net was proposed by introducing pyramid pooling , which can generate a fixed-length representation regardless of image size [28]. Inspired by SPP-net, ROI pooling has been proposed to construct Fast Region-based Convolutional Network (Fast R-CNN) [30] which outperforms R-CNN and SPP-net on speed and detection accuracy. Moreover, Faster R-CNN [32] has been proposed by introducing a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, reducing the cost of region proposals, further improving the detection performance and speed. Considering different scale of objects, SSD combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes, which can improve detection accuracy and efficiency [31]. YOLO has been proposed for real time object detection, the whole detection pipeline is in a single network [33]. Various improvements to the YOLO model is developed for constructing YOLO 9000(YOLO2) [34] which can achieve better detection performance and faster speed [34]. In order to reduce the redundancy of the current networks, Kim.et.al developed PVAnet, a thin and light network which is capable of complex vision tasks [35].

III. MULTI-TASK COST-SENSITIVE-CONVOLUTIONAL NEURAL NETWORK

In this section, we describe the proposed multi-task cost-sensitive-convolutional neural network. The architecture of proposed network is shown in Fig.3. Firstly, the multi-task partition layer is constructed by using several sub-task selection units which can generate some sub-tasks. In each sub-task, the scale and number of objects are enlarged and decreased, reducing the complexity of task. And then, the cost-sensitive loss is introduced into Faster R-CNN to construct cost-sensitive sub-network which can tackle the complex background such as occlusion effectively. Finally, the final detection result can be obtained by combining detection result of cost-sensitive sub-networks.

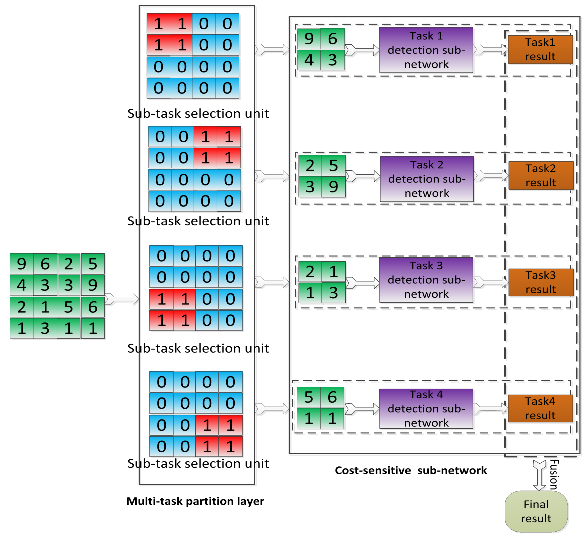

FIGURE 3. Multi-task cost-sensitive-convolutional neural network architecture.

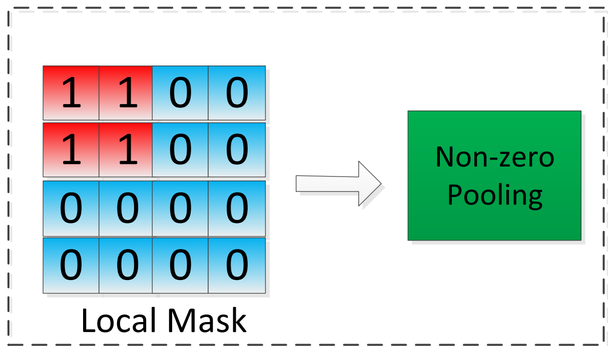

FIGURE 4. Sub-task selection unit.

A. MULTI-TASK PARTITION LAYER

The proposed multi-task partition layer is composed of sub-task selection units. The unit consists of a local mask and the non-zero pooling (shown in Fig.4). The local mask aims to focus on local information of interesting. For example, in Fig.4, the left top region (red region) of the images can be concerned via the local mask. After that, non-zero pooling is used to extract local map from original image.

Direct car detection in the original image is difficult because there is large number of tiny object. However, the proposed multi-task partition layer has ability to transfer the original task into many sub-tasks. The sub-task is to detect cars in local maps whose objects has larger scale and less number. Therefore, multi-task partition layer can divide the complex detection task of large number of tiny object into many simple detection sub-tasks of enlarged objects, which is helpful for improving performance.

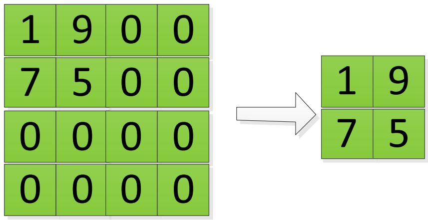

FIGURE 5. An example of non-zero pooling.

In sub-task selection unit, the image should be processed firstly by local mask which is the fundamental of the unit. In this procedure, a mask matrix with the same size of original image is given to generate interesting information, according to equation (1) as follows:

$$\mathrm{TR} = \mathrm{I} \bullet \mathrm{M} \tag{1}$$

In above equation, variable I denotes original image intensity matrix, M denotes mask matrix with the same size of original image, $\bullet$ denote dot product operation, TR can be used to represent local interesting information. In this paper, in order for convenience of computation, the value of elements in matrix M is 0 or 1. After calculation, matrix TR is composed of the elements whose value is non-zero or zero. Therefore, local maps can be concerned according to the non-zero elements in matrix TR.

After generating the interesting information by using local mask, Non-zero pooling is developed to extract local maps. Pooling layer is the basic layer in CNN [36], [37] to further extract the feature maps. Different from most of the pooling layer, there are no any parameters in non-zero pooling. According to Eq.(1), matrix TR contains local interesting information. The non-zero elements in matrix TR are used to represent interesting information. In the proposed system, TR is used as the input of Non-zero pooling. Non-zero pooling can obtain the local maps by extracting the elements whose value is not zero. Fig.5 gives an example of Non-zero pooling. The input is a 4×4 matrix which is obtained by Eq.(1). Non-zero pooling can extract the local map with size of 2×2 upper-right region whose values are not zero.

Fig.6 gives an example of original image while Fig.7 gives an example of local map obtained by using a sub-task selection unit. As shown in these two figures, compared with the cars in original image, we can see that the scale and the number of cars in the obtained local map are enlarged and decreased respectively, reducing the complexity of original task.

B. COST-SENSITIVE LOSS

In order to tackle complex background such as occlusion, the cost-sensitive loss function is introduced into detection subnetwork. In this paper, we classify the training samples as the hard sample and easy sample. On one hand, the hard sample is the car which may be failed to detected. For example, the car which is occluded by the trees. On the other hand, the easy sample is the car which can be detected easily. In the constructed loss function, a cost is assigned to each training samples. Intuitively, the proposed cost can automatically down-weight the contribution of easy examples during training and rapidly focus the model on hard examples. The cost of the training samples is calculated as shown in Eq.(2)

FIGURE 6. An example of original image.

FIGURE 7. An example of local map obtained by sub-task selection unit.

$$\mathrm{C}(x) = \begin{cases} \dfrac{p}{N} \times acc & x \in \Omega_h \\ \dfrac{q}{N} \times (1 - acc) & x \in \Omega_y \end{cases} \tag{2}$$

In above equation, variable $\mathrm{C}(x)$ denotes the cost of sample, Ω_h denotes hard sample set while Ω_y denotes easy sample set. Variable p denotes the number of easy samples while q denotes the number of hard samples. acc denotes the detection accuracy of Faster-RCNN in training set. As shown in Eq.(2), the cost is calculated according to the number and the detection accuracy of the samples. Generally speaking, the number of easy samples is larger than hard samples. Moreover, the detection accuracy is larger than 50%. According to Eq.(2), the cost of the hard samples is higher than easy samples.

In this paper, Faster R-CNN [32] is used as the basic network, the cost of each sample is introduced into the class prediction term in Faster R-CNN loss function [32], shown in Eq.(3)

$$L_{cls}(x, p, u) = -C(x) \log p_u \tag{3}$$

In above equation, p_u denotes the probability that sample x belongs to the class u. $C(x)$ is the cost of sample x that give an

FIGURE 8. Image examples in our dataset.

erroneous detection result. The proposed cost-sensitive loss can effectively discount the effect of easy examples and focus attentions on the hard examples, which may improve the detection performance on hard examples. Therefore, the proposed cost-sensitive loss function is ability to reduce the influence of the complex background, which may further improve the detection performance.

IV. EXPERIMENT

In order to demonstrate the effectiveness of the proposed method, we conduct 3 experiments on our database. In this section, we refer the proposed Multi-task Cost-sensitive-Convolutional Neural Network as MTCS-CNN. Faster R-CNN is used as the basic network of MTCNN, we compare MTCNN with Faster R-CNN in the first experiment. In the second experiment, we compare MTCS-CNN with the state-of-the-art method, such as SSD [31], PVAnet [35] and YOLO2 [34]. Considering the importance of the number of sub-tasks, we also give the parameter analysis of multi-task partition layer in the last experiment.

A. DATABASE AND EXPERIMENTAL SETUP

We construct a parking lot database which consists of 3000 images. The resolution of each image is 5456 × 3632. The image covers the parking lot of Inspur Group Co., Ltd, which are captured in aerial view(shown in Fig.8). There are hundreds of cars in each image. In addition, the image has complex background. For example, in the image, some cars are occluded by trees or their neighboring cars.

In the experiment, the proposed method is implemented by using TensorFlow [38] and runs on a workstation configured with an NVIDIA 1080 GPU card. In order to train the network, 2000 images are selected for training while the remain 1000 images are used for testing. In the experiment, the mean Average Precision (mAP) is reported to evaluate the detection accuracy.

B. COMPARISON WITH BASIC NETWORK

In this experiment, we compare MTCS-CNN with the basic network Faster R-CNN to demonstrate the effectiveness of the multi-task partition layer and cost-sensitive loss function. Fig.9 gives several detection results example. Fig.10 gives mAP of MTCS-CNN and Faster R-CNN. From these two figures, we can infer that MTCS-CNN outperformed Faster R-CNN in our task due to proposed multi-task partition layer and cost-sensitive loss.

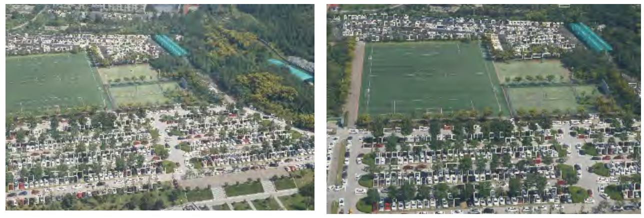

FIGURE 9. Two examples result of MTCS-CNN.

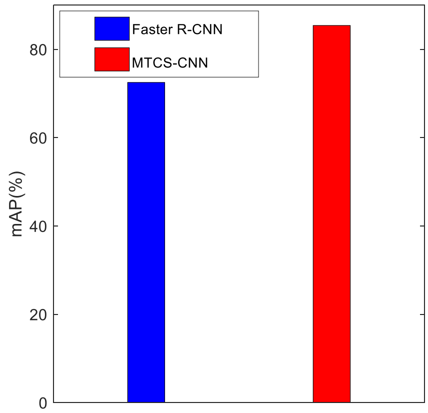

FIGURE 10. mAP of MTCS-CNN and Faster-R-CNN.

In our task, there is large number of cars in each image and the resolution of the image is high. For traditional Faster R-CNN, the image should be resized to a fixed size which is much smaller than the size of original image, resulting in the smaller scale of object. However, effective information loss of small objects may be appeared in feature map learning procedure, resulting in detection performance degradation. On the contrast, the proposed multi-task partition layer can divide the large number of tiny car detection task in the high resolution image into many simple sub-tasks. In each sub-task, the number of cars is decreased while the scale of cars is enlarged relatively, which can avoid the effective information loss in the feature map learning procedure. In addition, cost-sensitive loss is introduced to make the proposed method more robust to complex background, further improving detection performance. As shown in Fig.9, some cars which are occluded by trees can also be detected accurately.

C. COMPARISON WITH STATE-OF-THE-ART METHOD

In this experiment, we compare MTCS-CNN with state-of-the-art method, such as SSD [31], PVANET [35] and YOLO2 [34]. Fig.11and Fig.12 gives the performance of these methods, MTCS-CNN achieve the best detection performance.

In our task, the scale of the cars is small while the number of cars is large. In addition, complex background may affect the detection performance. In order to tackle these problems, MTCS-CNN is constructed by introducing multi-task partition layer and cost-sensitive loss. In the proposed method,

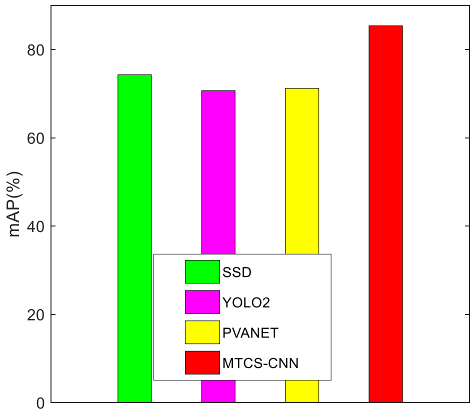

FIGURE 11. mAP of MTCS-CNN and state-of-the-art method methods.

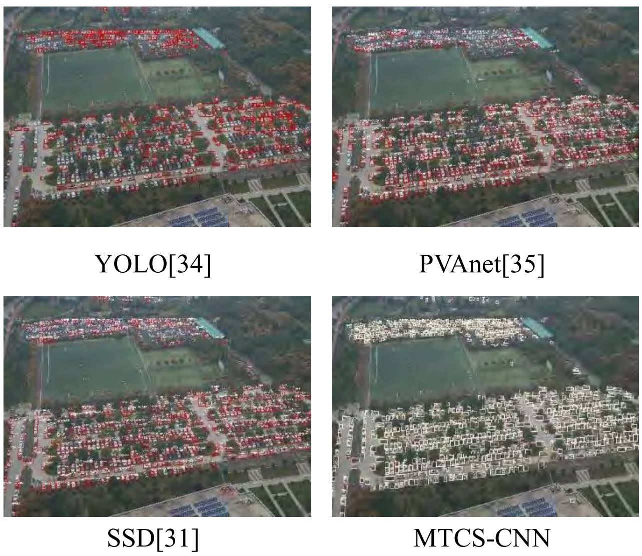

FIGURE 12. Detection examples of different methods.

multi-task partition layer can avoid that effective information loss of small object while cost-sensitive loss can make the model more robust to the complex background. Therefore, compared with state-of-the-art method, MTCS-CNN can achieve the best performance in our task.

D. PARAMETER ANALYSIS OF MULTI-TASK PARTITION LAYER

In this experiment, we analyze the influence of number of the sub-task units on detection performance. As shown in Fig.13, mAP is improved with the increasing of number of the sub-task units. However, when the number is larger than 64, mAP is decreased slowly.

The reason is that this parameter can control the number and the scale of cars indirectly. When this parameter is enlarged, the number of sub-task is increased. Accordingly, the number and the scale of cars are decreased and enlarged in each sub-task, which is helpful for improving detection performance. As shown in Fig.13, the best performance is achieved when the number of sub-task unit is 64.

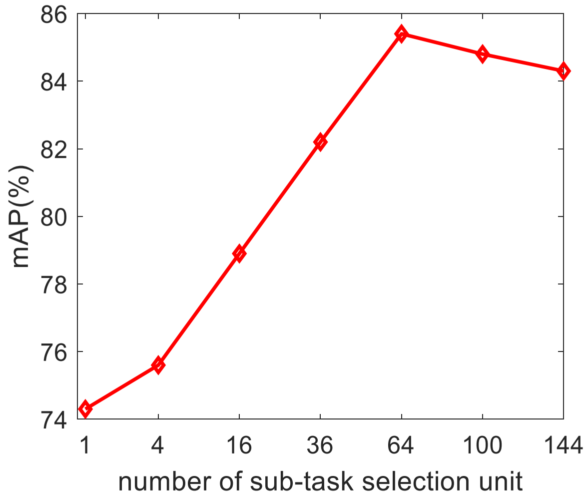

FIGURE 13. mAP of MSCS-CNN with different number of sub-task selection unit.

V. CONCLUSION

In order to address the problems of large number of tiny objects and complex background in car detection task in the aerial view image. This paper proposed a car detection method based on Multi-task Cost-sensitive-Convolutional Neural Network (MTCS-CNN).

Compared with traditional CNN architectures-based detection methods, the novelty of the proposed MTCS-CNN is as follows:

(1) Multi-task partition layer which is composed of some sub-task selection units is firstly developed. The sub-task selection unit is constructed by introducing local mask and non-zero pooling.

(2)The novel cost-sensitive loss is introduced in each sub-network.

Multi-task partition layer is used to tackle the challenge of large number of tiny objects. It is composed of some sub-task selection units. In each sub-task selection units, the local mask aims to focus on local interesting information of images while non-zero pooling is used to extract local map with local interesting information. Considering that the input size of the network is fixed, the size of the extracted local map should be enlarged as same as the original image. We can infer that more details of the object in the local map is ability to be learned due to the enlarged scale, which is helpful for performance improvement. Therefore, multi-task partition layer can divide a complex detection task of large number of tiny objects into many simple detection sub-tasks of few enlarged objects, result in the accurate detection results.

The proposed model has ability to be robust to complex background by introducing the novel cost-sensitive loss. In this paper, hard examples are failed to be detected due to influence of the complex background, result in the performance degradation. Cost-sensitive learning is a typical learning framework in machine learning that takes the misclassification costs into consideration [39]. It aims to guarantee the trained model has ability to learn enough knowledge to classify important samples by enlarging the cost to these samples. Inspired by this idea, a novel cost-sensitive loss is introduced. According to the proposed cost, the hard samples can be assigned a large cost. In order to reduce the loss, the model should focus training on a set of hard samples. We can infer that the trained model has ability to learn enough knowledge to classify these hard instances, further improving performance. Therefore, the model with incorporation of proposed cost-sensitive loss is robust to complex background, which may further improve the performance.

In this paper, the number of the sub-task is setting manually. Our future work will focus on automatic learning of best parameters.

ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of China (61701280),Natural Science Foundation of Shandong Province (ZR2016FQ18, ZR2018BF012), Natural Science Foundation of China (61801263,61573219, 61671274, 81371629, 61703235, 61701281,61703234), National Science Fund for Outstanding Young Scholars (61622114), Shandong Provincial Key Research and Development Plan (Grant.no 2017CXGC1504), Foundation of Distinguished Associate Professor in Shandong Jianzhu University. The authors would like to greatly thank the editors and the reviewers for their valuable comments and suggestions.

REFERENCES

[1] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. Shen, "Privacy-preserving smart parking navigation supporting efficient driving guidance retrieval," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6504–6517, Jul. 2018.

[2] R. Lu, X. Lin, H. Zhu, and X. Shen, "An intelligent secure and privacy-preserving parking scheme through vehicular communications," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2772–2785, Jul. 2010.

[3] T. Celik and H. Kusetogullari, "Solar-powered automated road surveillance system for speed violation detection," *IEEE Trans. Ind. Electron.*, vol. 57, no. 9, pp. 3216–3227, Sep. 2010.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, May 2015.

[5] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[6] W. Chu, Y. Liu, C. Shen, D. Cai, and X.-S. Hua, "Multi-task vehicle detection with region-of-interest voting," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 432–441, Jan. 2018.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[8] Y. Zhu, D. Comaniciu, M. Pellkofer, and T. Koehler, "Reliable detection of overtaking vehicles using robust information fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 401–414, Dec. 2006.

[9] S. Cherng, C. Y. Fang, C. P. Chen, and S. W. Chen, "Critical motion detection of nearby moving vehicles in a vision-based driver-assistance system," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 70–82, Mar. 2009.

[10] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3867–3876.

[11] D. Rüfenacht and D. Taubman, "HEVC-EPIC: Fast optical flow estimation from coded video via edge-preserving interpolation," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3100–3113, Jun. 2018.

[12] R. Ke, Z. Li, J. Tang, Z. Pan, and Y. Wang, "Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 54–64, Jan. 2019.

[13] J. Arrospide, L. Salgado, M. Nieto, and F. Jaureguizar, "On-board robust vehicle detection and tracking using adaptive quality evaluation," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2008–2011.

[14] P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," *J. Adv. Signal Process.*, vol. 1, p. 136, Jan. 2008.

[15] R. G. J. Wijnhoven and P. H. N. de With, "Unsupervised sub-categorization for object detection: Finding cars from a driving vehicle," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 2077–2083.

[16] X. Zhang, N. Zheng, Y. He, and F. Wang, "Vehicle detection using an extended Hidden Random Field model," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1555–1559.

[17] T. Liu, N. Zheng, L. Zhao, and H. Cheng, "Learning based symmetric features selection for car detection," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2005, pp. 124–129.

[18] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2019–2034, Jul. 2006.

[19] O. L. Junior and U. Nunes, "Improving the generalization properties of neural networks: An application to car detection," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 310–315.

[20] Z. Chen, C. Wang, H. Luo, H. Wang, Y. Chen, C. Wen, Y. Yu, L. Cao, and J. Li, "Vehicle detection in high-resolution aerial images based on fast sparse representation classification and multiorder feature," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2296–2309, Feb. 2016.

[21] L. Cao, F. Luo, L. Chen, Y. Sheng, H. Wang, C. Wang, and R. Ji, "Weakly supervised vehicle detection in satellite images via multi-instance discriminative learning," *Pattern Recognit.*, vol. 64, pp. 417–424, Apr. 2017.

[22] M. Ra, H. G. Jung, J. K. Suhr, and W. Y. Kim, "Part-based vehicle detection in side-rectilinear images for blind-spot detection," *Expert Syst. Appl.*, vol. 101, no. 1, pp. 116–128, Jul. 2018.

[23] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2241–2248.

[24] H. T. Niknejad, A. Takeuchi, S. Mita, and D. McAllester, "On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 748–758, Jun. 2012.

[25] R. K. Satzoda and M. M. Trivedi, "Multipart vehicle detection using symmetry-derived analysis and active learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 926–937, Apr. 2016.

[26] S.-H. Kong and S.-Y. Jun, "Cooperative positioning technique with decentralized malicious vehicle detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 826–838, Mar. 2018.

[27] H. Kuang, X. Zhang, Y. J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 927–936, Aug. 2016.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[30] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[31] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 21–37.

[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[34] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," Dec. 2016, *arXiv:1612.08242*. [Online]. Available: https://arxiv.org/abs/1612.08242

[35] K.-H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park, "PVANET: Deep but lightweight neural networks for real-time object detection," Sep. 2016, *arXiv:1608.08021*. [Online]. Available: https://arxiv.org/abs/1608.08021
[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[38] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and M. Kudlur, "TensorFlow: A system for large-scale machine learning," in *Proc. Symp. Operating Syst. Design Implement.*. May 2016, pp. 265–283.
[39] C. X. Ling and V. S. Sheng, *Cost-Sensitive Learning and the Class Imbalance Problem*. New York, NY, USA: Springer 2011.

XIAOMING XI is currently an Associate Professor with Shandong Jianzhu University. His research interests include machine learning, data mining, and medical image analysis.

ZHILOU YU is currently the Director of the Technology Center, Inspur Group Company, Ltd. His research interests include deep learning and image analysis.

ZHAOLEI ZHAN is currently a Research Fellow with the Technology Center, Inspur Group Company, Ltd. His research interests include deep learning and image processing.

YILONG YIN is currently the Director of the MLA Laboratory and a Professor with Shandong University. His research interests include machine learning, data mining, medical image analysis, and biometrics.

CUIHUAN TIAN is currently a Professor with Shandong University. Her research interests include image analysis and data mining.