Motion-Aligned and Hardness-Aware Dynamic Update Network for Weakly-Supervised Vehicle Detection in Satellite Videos

Quanpeng Jiang, Jie Feng, Yuping Liang, Ziyu Zhou, Xiangrong Zhang, Licheng Jiao

Motion-Aligned and Hardness-Aware Dynamic Update Network for Weakly-Supervised Vehicle Detection in Satellite Videos

Quanpeng Jiang, Jie Feng*, Yuping Liang, Ziyu Zhou, Xiangrong Zhang, Licheng Jiao

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi Province 710071, China

Abstract. Though the deep learning methods have achieved effective moving vehicle detection in satellite videos, there is a non-negligible premise that these methods require lots of object-level annotations for hundreds of small and blurry vehicles in the vast observation scene. These annotations can be quite labor-intensive and time-consuming. To address this problem, this paper is committed to realizing the vehicle detection based on point-level annotations, and a motion-aligned and hardness-aware dynamic update network is proposed, which consists of the basic detector, motion-aligned initialization method and online pseudo label update scheme. Specifically, the high-quality pseudo bounding boxes are initialized by revising the Gaussian mixture model to fully exploit the motion information in the video sequence and the location information from the point annotations. Then, the pseudo bounding boxes are utilized as the supervision for the basic detector. During the training phase, an online label refinement scheme is designed to refine the pseudo bounding box continuously, and the confidence-aware loss function is defined to adjust the example weight dynamically according to its learning hardness. Extensive experiments on the Jilin-1 and SkySat satellite video datasets show that our method achieves the comparative performance compared with fully-supervised learning methods.

Keywords: Weakly-supervised Vehicle Detection, Satellite Video, Point-level Annotations.

1 Introduction

With the continuous improvement of remote sensing technology and satellite imaging technology, high quality remote sensing images and videos with analytical values are becoming easier to obtain. Remote sensing videos obtained by optical sensors of vid-

This work was supported in part by the National Natural Science Foundation of China under Grant 61871306, Grant 61836009, Grant 62172600, Grant 62077038, by the Innovation Capability Support Program of Shaanxi (Program No. 2021KJXX-08), by the Natural Science Basic Research Program of Shaanxi under Grant No. 2022JC-45 and 2022GY-065, and by the Fundamental Research Funds for the Central Universities under Grant JB211901.

eo satellites staring at a specific area contain richer temporal information and a larger range of observation than natural images. They have been widely used in dynamic traffic detection, agriculture, forestry, water conservancy, mining, land management, ocean observation, atmospheric observation and other fields [1].

In recent years, the vehicle detection has become a research hotspot of satellite video processing and analysis. Specifically, computer vision combining with satellite remote sensing technology has broad application prospects in the field of intelligent transportation. Compared with the traditional traffic target monitoring equipment, the vehicle detection methods based on satellite remote sensing videos have many advantages, such as one-time investment and lasting application, no damage to road surface, no impact on ground traffic, large coverage area, rich traffic information acquisition, and so on, which provides new data and method source for traffic management and traffic flow dynamic monitoring [2,22,23].

To handle vehicle detection tasks, traditional methods consider it as a segmentation problem of background and foreground. They are divided into three categories: optical flow, background subtraction and frame difference. The Optical flow methods [4] find the corresponding relationship between the last frame and the current frame by using the changes of pixels in the time domain and the correlation between adjacent frames in the image sequence, so as to calculate the motion information of objects between adjacent frames [5].The basic principle of background subtraction [6] is to subtract the current frame and the background image determined by the model, and calculate regions whose pixels difference with the background exceeds a certain threshold as the moving region, so as to acquire the position, contour and size of the moving object. The frame difference [9] methods detect moving objects by looking for pixels that differ in adjacent frames. Unfortunately, most background subtraction are susceptible to moving backgrounds and changes in brightness and contrast, which are commonly present in satellite videos.

With the rise of deep learning and its powerful feature representation ability, object detection based on deep learning becomes a better choice. In the previous work, the classical object detectors can be divided into two categories: anchor-based methods and anchor-free methods. Based on anchor-based methods [10,12], these detectors usually design a large number of anchor boxes with pre-defined size and aspect ratio, and then classify and regression them to get the bounding box. Compared to the anchor-based approach, anchor-free detectors [11,14,15] no longer need to preset anchor boxes, but they can still achieve comparable performance with the former method.

However, both anchor-based and anchor-free methods have a premise that annotation boundary box is needed, which brings a strong demand for data annotation. Collecting bounding box-level annotations [16] is very expensive and laborious, especially for remote sensing images containing hundreds of objects. Specifically, in the vehicle detection task, there are usually 150-200 vehicles in a remote sensing image, and the edge of each vehicle is very fuzzy, even if manual annotation, it is difficult to accurately annotate a bounding box. Compared to bounding box-level annotations, point-level annotations have been widely used in object detection or segmentation, greatly reduces annotation time. Specifically, it takes about 12 seconds to annotate a bounding box-level instance and only 4 seconds to annotate a point-level instance. It

means we can save twice as much time on the same remote sensing image. For example, if we need to annotate 100 remote sensing images with an average number of 150 vehicles, the bounding box-level annotation takes 50 hours, while the same point-level annotation takes less than 17 hours. In order to solve the problem of object detection under fully-supervised network, a large number of bounding box annotations are labeled on the image firstly, which requires a lot of manpower and time, especially for satellite images containing hundreds of vehicles [18,20]. While, weakly-supervised object detection network based on point-level annotations needs to operate a small amount of weakly labeled training data to learn the model, which reduces a large amount of human labor in labeling training samples [17].

In this paper, a motion-aligned and hardness-aware dynamic update network is proposed for moving vehicle detection in the satellite images based on CenterNet. First, combining unsupervised object detection algorithm GMM, some low-quality pseudo label boxes are generated. The motion-aligned initialization method which bases on the multi corresponding relation-ship between pseudo-label boxes and points generates higher quality pseudo label boxes. And then, the size of the pseudo-label boxes with higher confidence is constantly updated during the training process through the online pseudo label update mechanism designed. Finally, a newly-designed confidence-aware loss function is proposed to assist fully-supervised networks to better mine hard training samples for learning.

According to what have been argued above, the main contributions of this paper can be summarized as follows:

1. A motion-aligned and hardness-aware dynamic update network (MHDUN) based on only point annotations is proposed to reduce the manual labeling time and achieve performance comparable to the fully-supervised method.
2. An accurate motion-aligned initialization method is designed to initialize the pseudo-label box precisely, taking full advantage of the point annotations to mine the size information and combining with the motion information in the videos.
3. An online pseudo label update scheme is proposed, which contains a novel confidence-aware loss function by adjusting the training example weight to further improve the quality of the size of the pseudo label box during training and ensure the stability of the entire training process.

The rest of this paper is organized as follows. In section 2, the overall structure and details of MHDUN are described. In section 3, the detailed experimental results and analysis are discussed to verify the effectiveness of the proposed network. Finally, the conclusions and some suggestions for future work are given in section 4.

2 Proposed Method

2.1 Overview

In this section, the proposed method is introduced in detail. Fig. 1 shows the overall structure of MHDUN, which is capable of training an object detector only with point-level annotations for vehicle detection. Specifically, the network is based on an anchor-free object detector, CenterNet. Firstly, the motion-aligned initialization (MAI)

method is proposed to generate the relatively accurate initial pseudo size for every vehicle. Furthermore, an online pseudo label update scheme (OPLU Scheme) is proposed to refine the pseudo sizes in every training epoch. Besides, a novel confidence-aware loss function contained in the OPLU Scheme is designed to pay more attention to hard training samples by adding a bigger weight in vehicle's size regression.

2.2 Motion-aligned initialization

In order to be able to train a fully supervised network later, we need to first initialize the pseudo label bounding box from the point annotations. The closer a natural image is taken, the larger the object is, and vice versa [20]. However, the size of objects in remote sensing images is not affected by this, and the size difference of object is not obvious. Therefore, based on this discovery, we propose a novel motion-aligned initialization method combined with the GMM to initialize the size of the object.

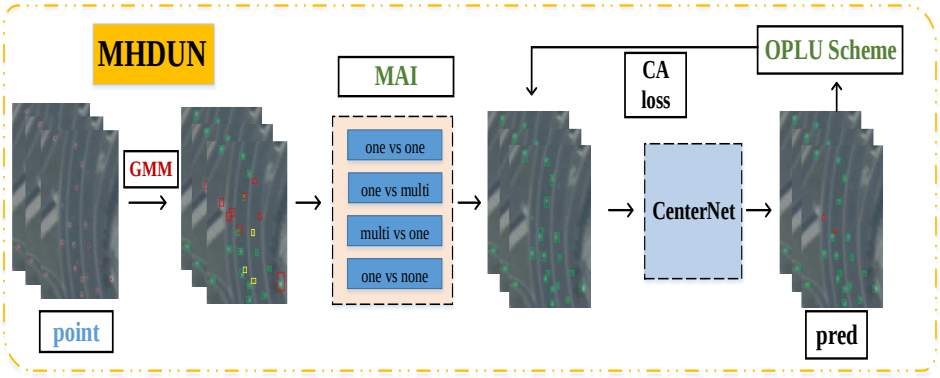


Fig. 1. Overview of motion-aligned and hardness-aware dynamic update network (MHDUN) for moving vehicle detection in the satellite videos. It consists of the basic detector, motion-aligned initialization method and online pseudo label update scheme.

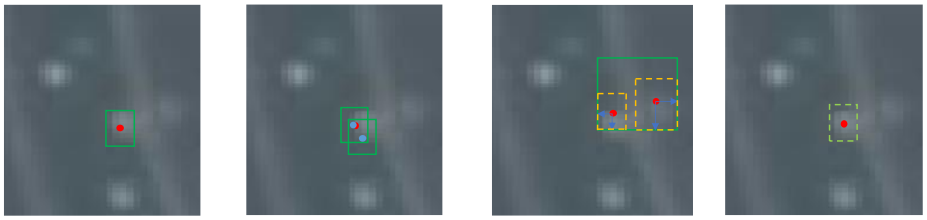
2.3 Motion-aligned initialization

In order to be able to train a fully supervised network later, we need to first initialize the pseudo label bounding box from the point annotations. The closer a natural image is taken, the larger the object is, and vice versa [20]. However, the size of objects in remote sensing images is not affected by this, and the size difference of object is not obvious. Therefore, based on this discovery, we propose a novel motion-aligned initialization method combined with the GMM to initialize the size of the object.

First, as for how to generate a pseudo label bounding box, our strategy is to use the GMM, which can obtain a relatively accurate initial pseudo label bounding box for us.

Then, the different position inclusion relationship of the points and boxes need to be judged, there are four corresponding relationships between the pseudo label bounding box generated by GMM and the real point annotation as can be seen from Fig.2.

Specifically, 1) one vs. one (one point corresponds to one box): the box predicted by GMM is used as the pseudo label bounding box of the real point. 2) one vs. multi (one point corresponds to multi boxes): It means that there is only one real object here in the image, but GMM predicts multiple boxes. Therefore, in order to maintain the uniqueness and accuracy of the initial pseudo label box, the distance between the center point of each prediction box and the real point mark is calculated, and then the prediction box corresponding to the smallest distance is taken as the corresponding pseudo label bounding box of the point finally. 3) multi vs. one (multi points corresponds to one box): The situation shows that GMM prediction is not accurate enough. Therefore, in order to avoid vehicle missing detection in subsequent fully-supervised training as much as possible, a fairly reasonable generation method of multiple pseudo-label boxes is proposed. First of all, the vertical distance from each point to predict box of four sides are calculated. Then, the minimum distance between each point and both sides of the predict box in the horizontal and vertical directions is calculated. Finally, we consider the real point as the center point, and take the minimum distance to the vertical boundary and horizontal boundary as its height and width to generate the corresponding pseudo label box of each point. 4) one vs. none (one point corresponds to no box): In order to make up for the deficiency of the traditional algorithm, a pseudo label box is generated for each point without corresponding box. To be specific, the mean and variance of the size of the pseudo label box generated in the first three cases are calculated. According to this, a three-sigma rule is adopted to random generate the size of the pseudo label box.



(a) one vs. one (b) one vs. multi (c) multi vs. one (d) one vs. none

Fig. 2. Motion-aligned initialization. (a) is the GT point denoted in red point and pseudo label box predicted by the GMM denoted in green solid box, (b) is the center point of the pseudo label box denoted in blue point, (c) is new pseudo label boxes denoted in yellow dashed boxes, (d) is the pseudo label box generated by three-sigma rule denoted in light green dashed box.

2.4 Online pseudo label update scheme

Due to the existence of inaccurate initial pseudo-label boxes, inputting such boxes will lead to the instability of target detection network training or convergence to local optimum. In order to train a reliable and stable object detection network, we propose an online pseudo label updating mechanism iteratively updating and refining pseudo label boxes in every epoch. We set an initial confidence value for each initial pseudo-label box. In the training process, the size of the pseudo-label box will be updated

only when the predicted corresponding box in next epoch is greater than the initial confidence. At the same time, the initial confidence level will be replaced with a higher confidence level, which ensures that the fully supervised detection network trains with increasingly confident training samples.

Besides, a new-designed confidence-aware loss function is designed to focus on hard training samples by adding a bigger weight in vehicle's size regression to perfect and supplement online updating refinement scheme. Specifically, we can get the score of each prediction box and instinctively believe that the positive box with higher score is an easy sample while the positive box with lower score is a difficult one. Thus, we define the weights for each of these prediction bounding boxes:

$$w_i = 1/\,sigmoid(s_i) \tag{1}$$

where w_i denotes the weight of the i-th bounding box and s_i denotes the score of the i-th bounding box. Then, we apply to multiplied with the heatmap loss to get confidence-aware loss (CA Loss):

$$CA\ Loss = \frac{-1}{N}\sum_{xyc}\begin{cases}(1-Y^{\wedge}_{xyc})^{\alpha}\log(Y^{\wedge}_{xyc})w_i & if\ \ Y^{\wedge}_{xyc}=1\\ (1-Y_{xyc})^{\beta}(Y^{\wedge}_{xyc})^{\alpha}\log(1-Y^{\wedge}_{xyc})(1-w_i) & otherwise\end{cases} \tag{2}$$

which replace the original heatmap loss, where α and β are hyper-parameters of the focal loss and N is the number of keypoints in the image just same to CenterNet. The keypoint heatmap predicted by the network is $Y^{\wedge}_{xyc}$ and Y_{xyc} represents the ground truth keypoint heatmap, x and y represent the position on the heatmap and c represents the category. α = 2 and β = 4 are used in our experiments.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Datasets

In this paper, Dubai and San Diego datasets captured by Jilin-1 satellite and the Las Vagas dataset captured by SkySat satellite are used to validate the effectiveness of the proposed network, as shown in Fig. 3. Dubai dataset is captured over Dubai, UAE, on November 9, 2018. AOIs 1-3 (areas of interests) come from the dataset and the size of the three frames are 1000 × 1000 pixels. AOIs 1, 2 are used for training, AOIs 3 is used for testing. San Diego dataset is captured over San Diego, USA, on May 23, 2017. AOIs 4–6 come from the dataset and the size of AOIs 5,6 are 1000 × 1000 pixels, while the size of AOI 4 is 1500 × 700 pixels. AOIs 4, 5 are used for training, AOIs 6 is used for testing. Las Vagas dataset is captured over Las Vagas, USA on March 25, 2014. Video 001, 002 come from the dataset and the size of Video 001 is 400 × 400 pixels, while the size of Video 002 is 600 × 400 pixels. Video 002 is used for training and Video 001 is used for testing.

3.2 Experimental Setups

We implemented MHDUN in PyTorch 1.10.0 without pretrained parameters, and initialized by PyTorch default setting. The networks are trained on the Windows server 2019 system with a single RTX 3090 GPU and AMD Ryzen 9 5950X CPU. The benchmark network used in the experiment is CenterNet, and the backbone is Hourglass-104 [21]. Specifically, the network is trained with a batch size of 4 and learning rate 0.32e-4 with the Adam optimizer for 20 epochs. We use random cropping, and color jittering as data augmentation for the training data, and no any data augmentation for the testing data.

3.3 Detection Results of Different Methods

To verify the effectiveness of motion-aligned initialization and hardness-aware based dynamic update network, some representative traditional methods, FastMCD [13], ViBe [3], GMM [7], and deep learning methods, Faster R-CNN [10], YOLO v4 [16], CenterNet [11], CornerNet [14], and CentripetalNet [15], are selected for comparison. For a fair comparison, Faster R-CNN, YOLO v4, Cor-nerNet, CenterNet and CentripetalNet take three stacked frames as the input to exploit the temporal information but these models are based on bounding box annotations, not point annotations. The common evaluation of Precision, Recall and F1 score are used for all experiments. GMM algorithm has a relatively good performance in the traditional algorithm and thus it is selected to generate initial pseudo-label boxes in our network. It is not difficult to find that deep learning methods obtain better detection perfor-mance compared with traditional methods. At the same time, the most important thing is that motion-aligned initialization and hardness-aware based dynamic update network is based on point annotations, obtaining similar results with deep learning methods based on box annotations, which further illustrates the effec-tiveness of the network.

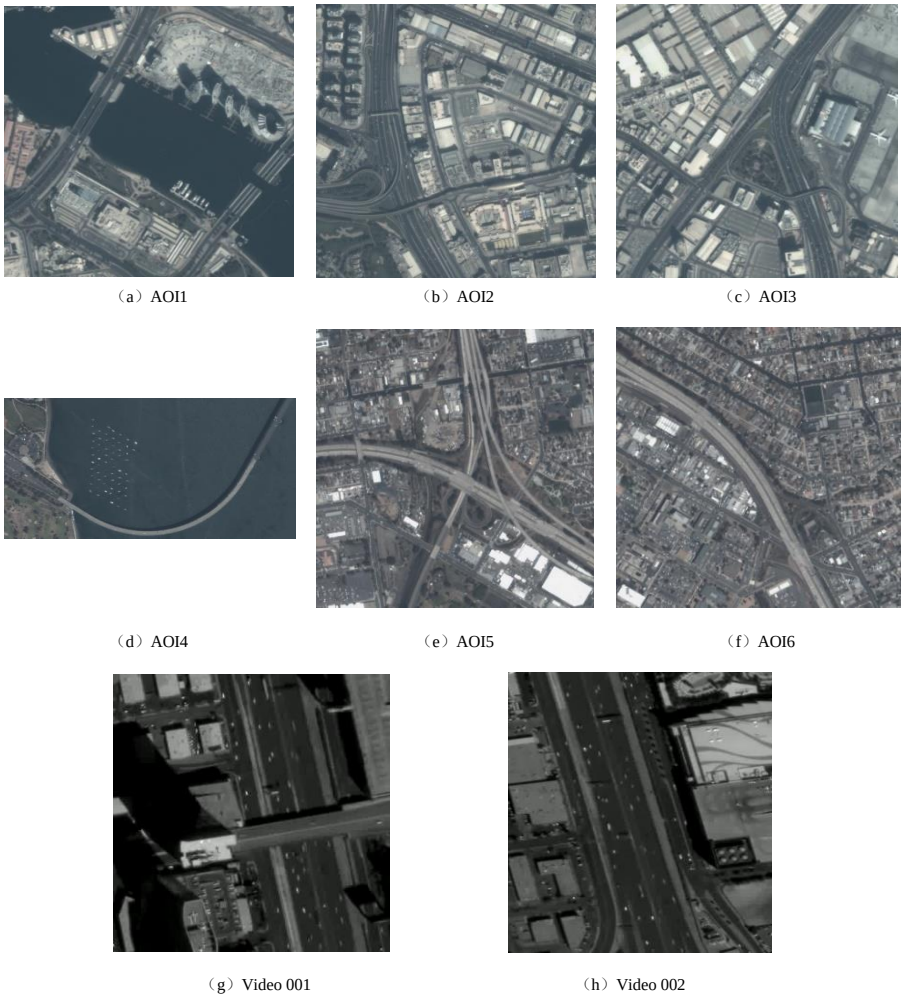


Fig. 3. Examples of satellite video datasets. (a), (b) and (c) belong to Dubai dataset from the Jilin-1 satellite, (d), (e) and (f) are from San Diego dataset from the Jilin-1 satellite, (g) and (h) are from Las Vegas dataset from the SkySat satellite.

As shown in Fig.4, most of the vehicles are detected rightly, and there is little difference in size. Besides, there is only some false detection for blurry objects and a little missing detection. In all, the overall detection effect is relatively good and the network can achieve similar performance to fully-supervision network.

Table 1. Detection results of different methods in satellite videos.

	Las Vegas (SkySat)			Dubai (JiLin-1)			San Diego (JiLin-1)		
	Prec↑	Rec↑	F1↑	Prec↑	Rec↑	F1↑	Prec↑	Rec↑	F1↑
Traditional methods	(%)								
FastMCD	72.46	71.74	72.10	90.01	63.21	74.27	87.33	49.96	63.56
ViBe	53.24	78.43	63.43	47.00	66.85	55.20	61.17	61.07	61.12
GMM	55.63	72.56	62.98	82.92	71.48	76.78	88.46	62.38	73.16
Deep learning methods	(%)								
Faster R-CNN	86.13	87.27	86.70	88.47	84.28	86.32	85.14	80.81	82.92
YOLOv4	88.98	87.48	88.22	88.93	86.97	87.94	87.71	82.47	85.01
CenterNet	85.48	72.55	78.49	84.25	75.63	79.71	86.57	65.92	74.85
CornerNet	87.54	84.48	85.98	87.08	83.48	85.24	86.93	58.99	68.19
CentripetalNet	87.69	88.13	87.91	87.87	88.69	88.28	82.58	85.32	83.93
MHDUN(Our)	82.87	78.31	80.53	88.24	85.63	86.39	83.24	79.26	81.20

(a) Las Vegas (SkySat) (b) Dubai (JiLin-1) (c) San Diego (JiLin-1)

Fig. 4. Detection results in the Las Vegas, Dubai and San Diego satellite video datasets. The detection box is represented in blue, the ground truth box is represented in red.

3.4 Ablation Experiments

The ablation experiments are implemented to investigate the effectiveness of MAI, OPLU Scheme and CA Loss on the San Diego Datasets. From the comparison between GMM and MAI, we can find that Prec., Rec. and F1 increase by 6.27%, 14.45% and 10.73% respectively, which indicates that OPLU Scheme has a strong ability to correct the pseudo-label box generated by GMM algorithm. From the comparison between GMM and OPLU Scheme, we can observe that Prec., Rec. and F1 increase by 2.92%, 4.18% and 4.31% respectively, which indicates that OPLU Scheme has the ability to update the pseudo-label box generated by GMM algorithm during the training. Comparing the results of the first three lines in the table above, we

can draw a conclusion that MAI is more effective than OPLU Scheme in correcting the size of the pseudo-label box. Also, CA Loss can further improve the results which achieves a respectable performance compared to fully-supervised methods of 83.24% Prec., 79.26% Rec. and 81.20% F1.

Table 2. Ablation experiments on the San Diego Datasets.

	Prec (%) ↑	Rec (%) ↑	F1(%) ↑
CenterNet w/ GMM	73.46	63.38	68.04
CenterNet w/ MAI	79.73	77.83	78.77
CenterNet w/ GMM + OPLU Scheme	76.38	67.56	72.35
CenterNet w/ MAI + OPLU Scheme	81.57	78.92	80.22
CenterNet w/ MAI + OPLU Scheme + CA Loss	**83.24**	**79.26**	**81.20**

4 CONCLUSION

In this paper, a novel motion-aligned initialization and hardness-aware based dynamic update network is proposed for moving vehicle detection in satellite videos. Motion-aligned initialization method can accurately initialize the pseudo-label box of each object by combining GMM algorithm with the different correspondence between boxes and points. Online pseudo label update scheme can iteratively update and refine the size of pseudo label boxes every epoch to ensure stability of training. Besides, CA Loss even can further mine difficult training samples to guide the pseudo label box regression. These innovations presented in this paper are likely to be universal under the weakly supervised learning framework. In the future, it is not difficult to find that there are some defects in using the traditional background difference algorithm GMM to assist the generation of pseudo label boxes. Therefore, in the future, there is an attempt to adopt other different traditional algorithms, such as Vibe algorithm, or integrate multiple traditional models to achieve more accurate initialization of pseudo-label boxes.

References

1. Ao, Wei, Yanwei Fu, Xiyue Hou and Feng Xu. “Needles in a Haystack: Tracking City-Scale Moving Vehicles From Continuously Moving Satellite.” IEEE Transactions on Image Processing 29 (2020): 1944-1957.
2. Ahmadi, Seyed Ali, Arsalan Ghorbanian and Ali Mohammadzadeh. “Moving vehicle detection, tracking and traffic parameter estimation from a satellite video: a perspective on a smarter city.” International Journal of Remote Sensing 40 (2019): 8379 - 8394.
3. Barnich, Olivier and Marc Van Droogenbroeck. “ViBe: A Universal Background Subtraction Algorithm for Video Sequences.” IEEE Transactions on Image Processing 20 (2011): 1709-1724.

4. Roy, Sourav Dey and Mrinal Kanti Bhowmik. “A Comprehensive Survey on Computer Vision Based Approaches for Moving Object Detection.” 2020 IEEE Region 10 Symposium (TENSYMP) (2020): 1531-1534.
5. Ranjan, Anurag and Michael J. Black. “Optical Flow Estimation Using a Spatial Pyramid Network.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017): 2720-2729.
6. Dale, Matthew A., Melissa K Suh, Shijia Zhao, Trevor Meisinger, Linxia Gu, Vicki J. Swier, Devendra K. Agrawal, Timothy C Greiner, Jeffrey S. Carson, Bernard Timothy Baxter and Wanfen Xiong. “Background differences in baseline and stimulated MMP levels influence abdominal aortic aneurysm susceptibility.” Atherosclerosis 243 2 (2015): 621-9 .
7. Stauffer, C. and W. Eric L. Grimson. “Adaptive background mixture models for real-time tracking.” Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) 2 (1999): 246-252 Vol. 2.
8. Sheather, Simon J. and M. Chris Jones. “A reliable data-based bandwidth selection method for kernel density estimation.” Journal of the royal statistical society series b-methodological 53 (1991): 683-690.
9. Chaohui, Zhang, Du Xiaohui, Xu Shuoyu, Song Zheng and Luo Min. “An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection.” Fourth International Conference on Image and Graphics (ICIG 2007) (2007): 519-523.
10. Ren, Shaoqing, Kaiming He, Ross B. Girshick and Jian Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015): 1137-1149.
11. Zhou, Xingyi, Dequan Wang and Philipp Krähenbühl. “Objects as Points.” ArXiv abs/1904.07850 (2019): n. pag.
12. Bochkovskiy, Alexey, Chien-Yao Wang and Hong-Yuan Mark Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection.” ArXiv abs/2004.10934 (2020): n. pag.
13. Rousseeuw, Peter J. and Katrien van Driessen. “A Fast Algorithm for the Minimum Covariance Determinant Estimator.” Technometrics 41 (1999): 212-223.
14. Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." Proceedings of the European conference on computer vision (ECCV). 2018.
15. Dong, Zhiwei, Guoxuan Li, Yue Liao, Fei Wang, Pengju Ren and Chen Qian. “CentripetalNet: Pursuing High-Quality Keypoint Pairs for Object Detection.” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020): 10516-10525.
16. Lin, Tsung-Yi, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár and C. Lawrence Zitnick. “Microsoft COCO: Common Objects in Context.” ECCV (2014).
17. Branson, Steve, Pietro Perona and Serge J. Belongie. “Strong supervision from weak annotation: Interactive training of deformable part models.” 2011 International Conference on Computer Vision (2011): 1832-1839.
18. Yao, Xiwen, Xiaoxu Feng, Junwei Han, Gong Cheng and Lei Guo. “Automatic Weakly Supervised Object Detection From High Spatial Resolution Remote Sensing Images via Dynamic Curriculum Learning.” IEEE Trans. Geosci. Remote. Sens. 59 (2021): 675-685.
19. Liu, Yuting, Miaojing Shi, Qijun Zhao and Xiaofang Wang. “Point in, Box Out: Beyond Counting Persons in Crowds.” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019): 6462-6471.
20. Li, Youyou, Binbin He, Farid Melgani and Teng Long. “Point-Based Weakly Supervised Learning for Object Detection in High Spatial Resolution Remote Sensing Images.” IEEE

Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021): 5361-5371.
21. Newell, Alejandro, Kaiyu Yang and Jia Deng. “Stacked Hourglass Networks for Human Pose Estimation.” ECCV (2016).
22. Casagli, Nicola, William Frodella, Stefano Morelli, Veronica Tofani, Andrea Ciampalini, Emanuele Intrieri, Federico Raspini, Guglielmo Rossi, Luca Tanteri and Ping Lu. “Space-borne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning.” Geoenvironmental Disasters 4 (2017): 1-23.
23. Yang, Tao, Xiwen Wang, Bowei Yao, Jing Li, Yanning Zhang, Zhannan He and Wencheng Duan. “Small Moving Vehicle Detection in a Satellite Video of an Urban Area.” Sensors (Basel, Switzerland) 16 (2016): n. pag.