SSSDET: Simple Short and Shallow Network for Resource Efficient Vehicle Detection in Aerial Scenes

26th IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019.

By: Murari Mandal

Vision Intelligence Lab

Introduction: Aerial Vehicle Detection

- Advances in unmanned aerial vehicle (UAV) technology has unlocked a new frontier of computer vision which requires analysis and interpretation of aerial images and videos.

- Detection of small-sized targets in airborne images from urban scenes, industrial sites and agricultural landscapes has numerous applications in surveillance, remote sensing and other commercial purposes

Vehicle Detection in Aerial View: Applications

- Aerial surveillance: search and rescue
- Maritime surveillance
- Urban and rural scene understanding.
- Industrial site inspection
- Agricultural land monitoring and analysis
- Event recognition
- Other commercial purposes.

Vehicle Detection in Aerial View: Challenges

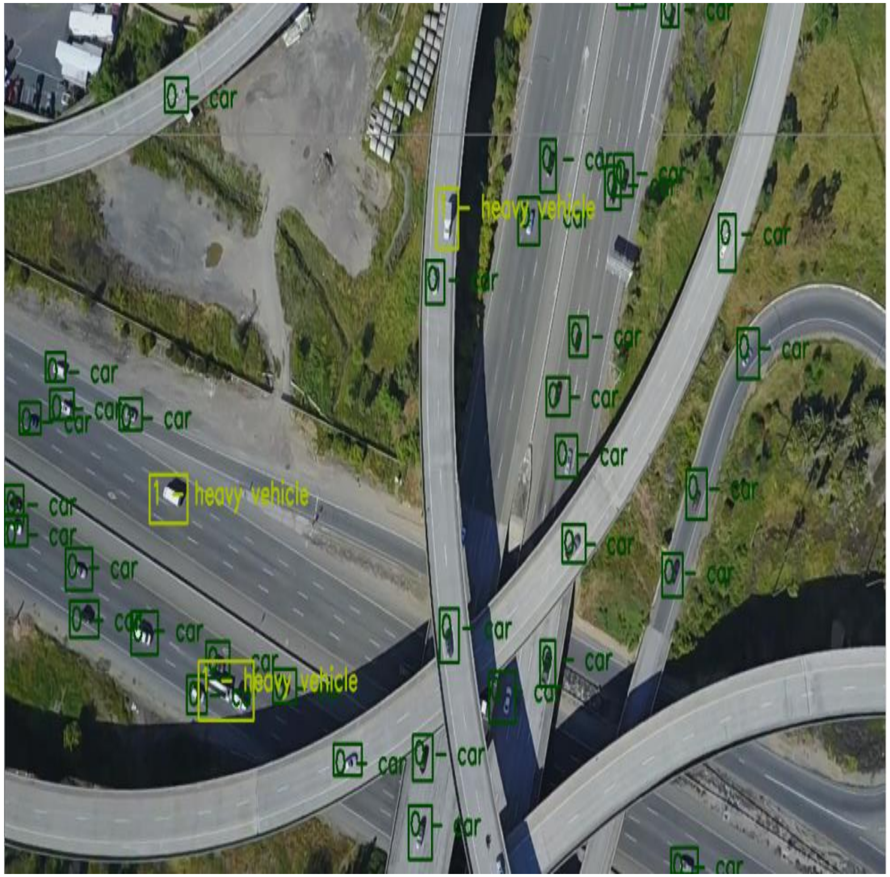
- Vehicle detection in aerial images is a challenging task due to the variable sizes of the vehicles (small, medium and large).

- High/low density of vehicles and complex background in the cameras field of view.

- Moreover, the aerial scenes in urban setup usually comprises of a varieties of object types leading to excessive interclass object similarities.

Regular Vs Aerial View

Regular View



Aerial View

Fig. 1. Difference between regular and aerial view

Expected Features for UAV based Applications

- Resource Efficient Model.
- Memory – The model must take very less memory space.
- Compute – The model must operate even with minimal computational support.
- Accuracy – The model must offer reasonable accurate results.
- Real-time – The model must offer scope for real-time inference.

SSSDet

- SSSDet: Simple Short and Shallow Network for Resource Efficient Vehicle Detection in Aerial Scenes

- Proposed a lightweight network for vehicle detection in aerial scenes.

- The SSSDet is a simple short and shallow convolutional network optimized for fast inference and high accuracy.

- We proposed deep learning based solution to develop a robust and resource efficient vehicle detector for aerial images.

SSDet

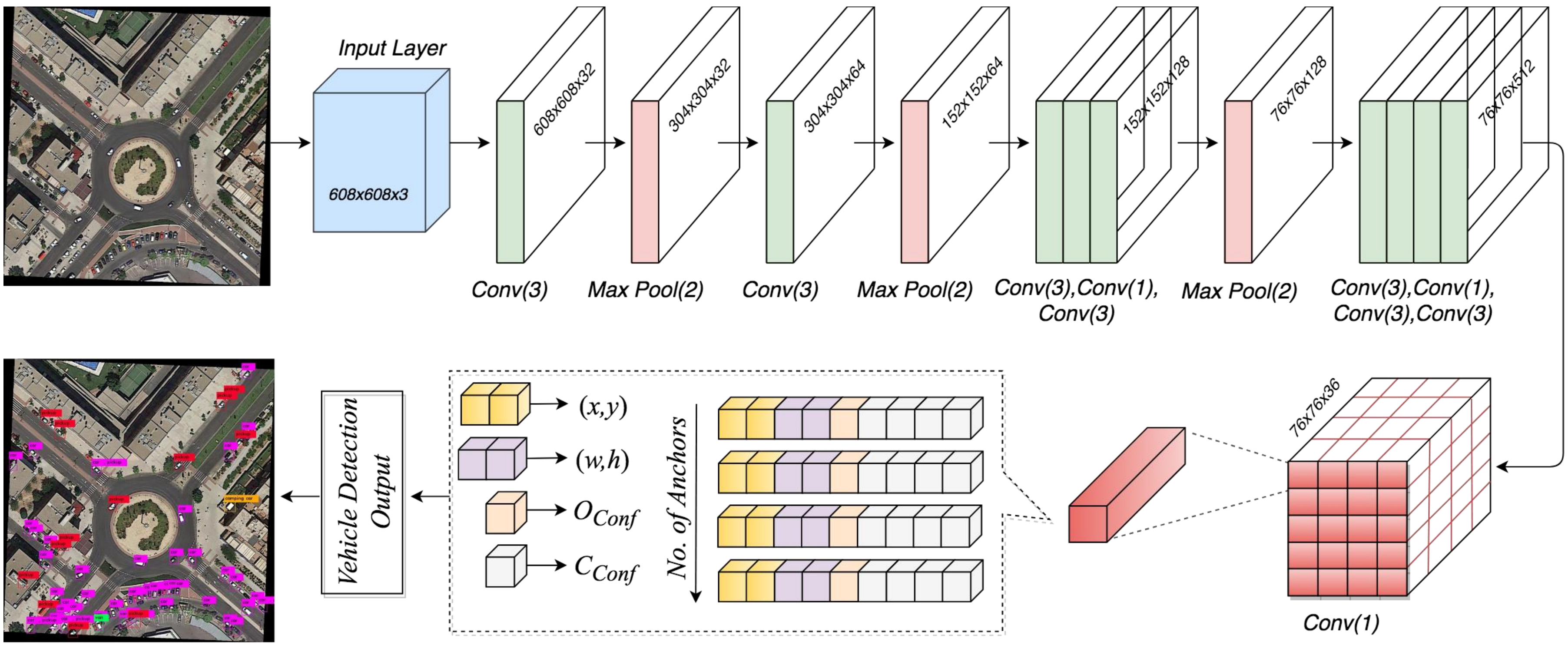


Fig. 2. The proposed SSSDet architecture is shown for object detection and classification in 4-class DOTA dataset. The final layer features are composed of 5776 tensors of size 1 ×1×36. Each tensor contains the bounding box coordinates (x, y, w, h), object confidence (OConf) and class confidence (CConf) for every anchor box

Vision Intelligence Lab

SSSDet: Layer Visualization

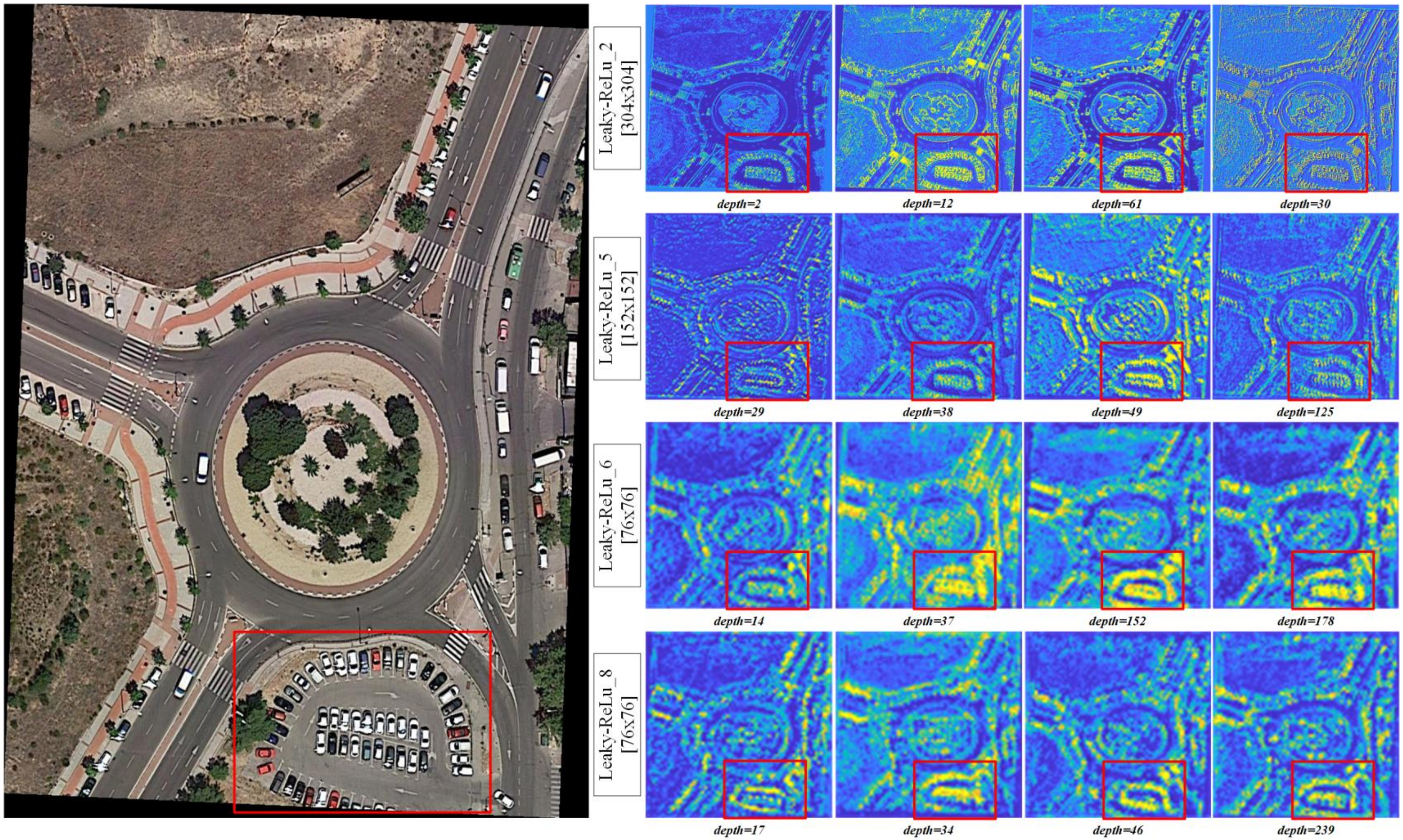


Fig. 3. The heat map visualization of the various activation layers of SSSDet. The yellow color indicates the activation of relevant regions in the original image.

Qualitative Comparison

Method/ # Detected Objects	Car	Heavy Vehicle	Plane	Boat
YOLOv2_416	0	0	0	12
YOLOv2_608	0	0	0	33
YOLOv3_416	0	0	0	48
YOLOv3_608	0	0	0	21
RetinaNet	0	0	0	45
Proposed SSSDet	2	0	0	53

Fig. 4. Qualitative results of the proposed SSSDet, YOLOv2, YOLOv3, RetinaNet

Qualitative Comparison

YOLOv2_416x416

Method\ # Detected Objects	Car	Heavy Vehicle	Plane	Boat
YOLOv2_416	7	0	0	0
YOLOv2_608	15	0	0	0
YOLOv3_416	19	0	0	0
YOLOv3_608	11	0	0	0
RetinaNet	17	0	0	0
Proposed SSSDet	24	0	0	0

YOLOv2_608x608

YOLOv3_416x416

Input

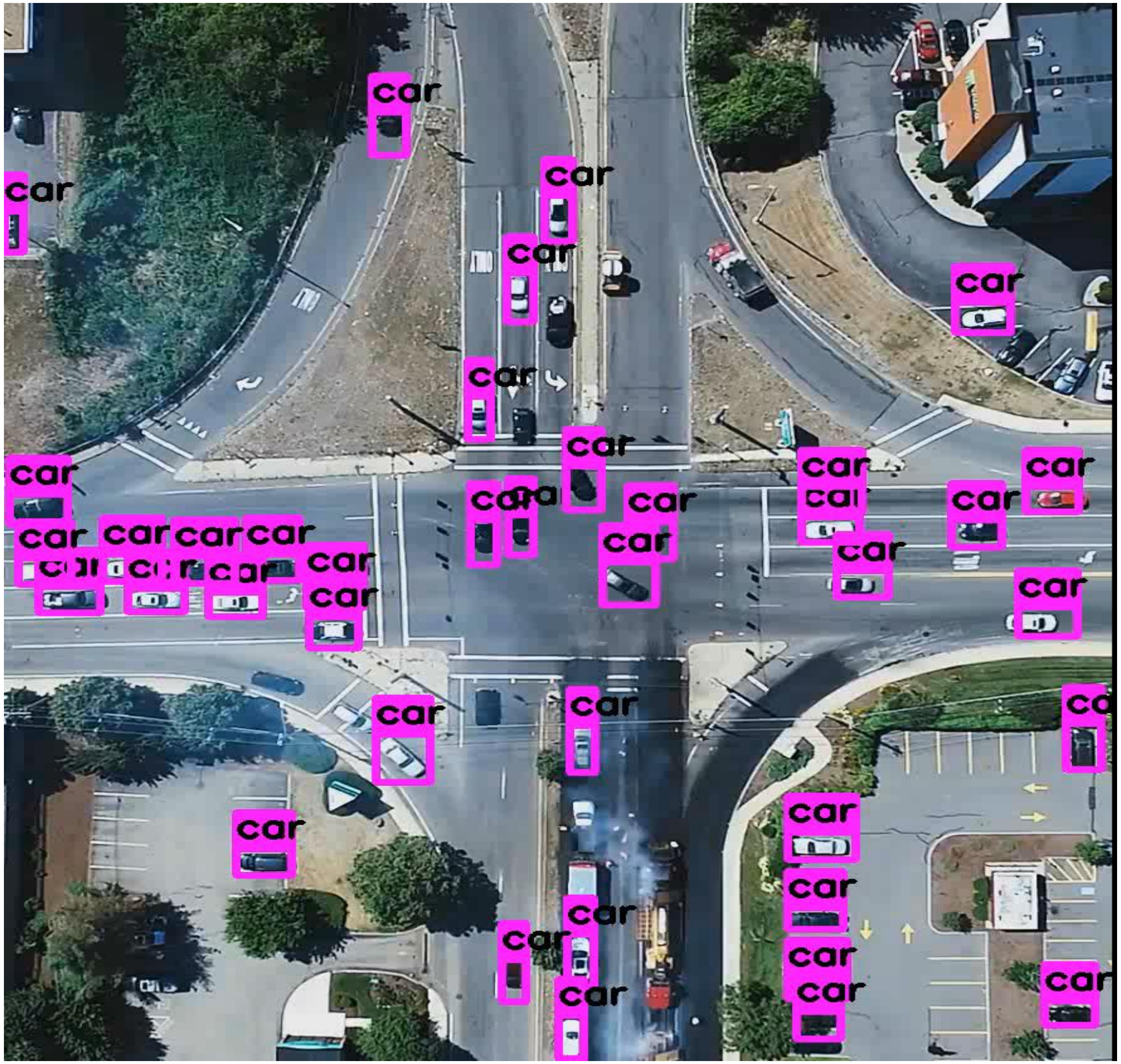
RetinaNet

YOLOv3_608x608

Proposed SSSDet

Fig. 5. Qualitative results of the proposed SSSDet, YOLOv2, YOLOv3, RetinaNet

Qualitative Results



Data Description

Table 1. Summarization of the evaluated datasets

Dataset	#Images	#Objects	#Object per class
VEDAI	1248	3773	car: 1393, truck: 307, pickup: 955, tct: 190, cc: 397, bt: 171, mc: 4, bus: 3, van: 101, other: 204, large: 48
DLR-3K	262	8401	car: 8210, hv: 191
DOTA	1558	55235	car: 24516, hv: 11307, pln: 4733, bt: 14679
ABD	79	1396	car: 1353, hv: 11, bt: 32
Complete	3099	68579	car: 36510, hv: 12406, pln: 4781, bt: 14882

**tct: tractor, cc: camping car, mc: motorcycle, hv: heavy vehicle, pln: plane, bt:boat*

**This table is from SSSDet paper*

ABD Dataset

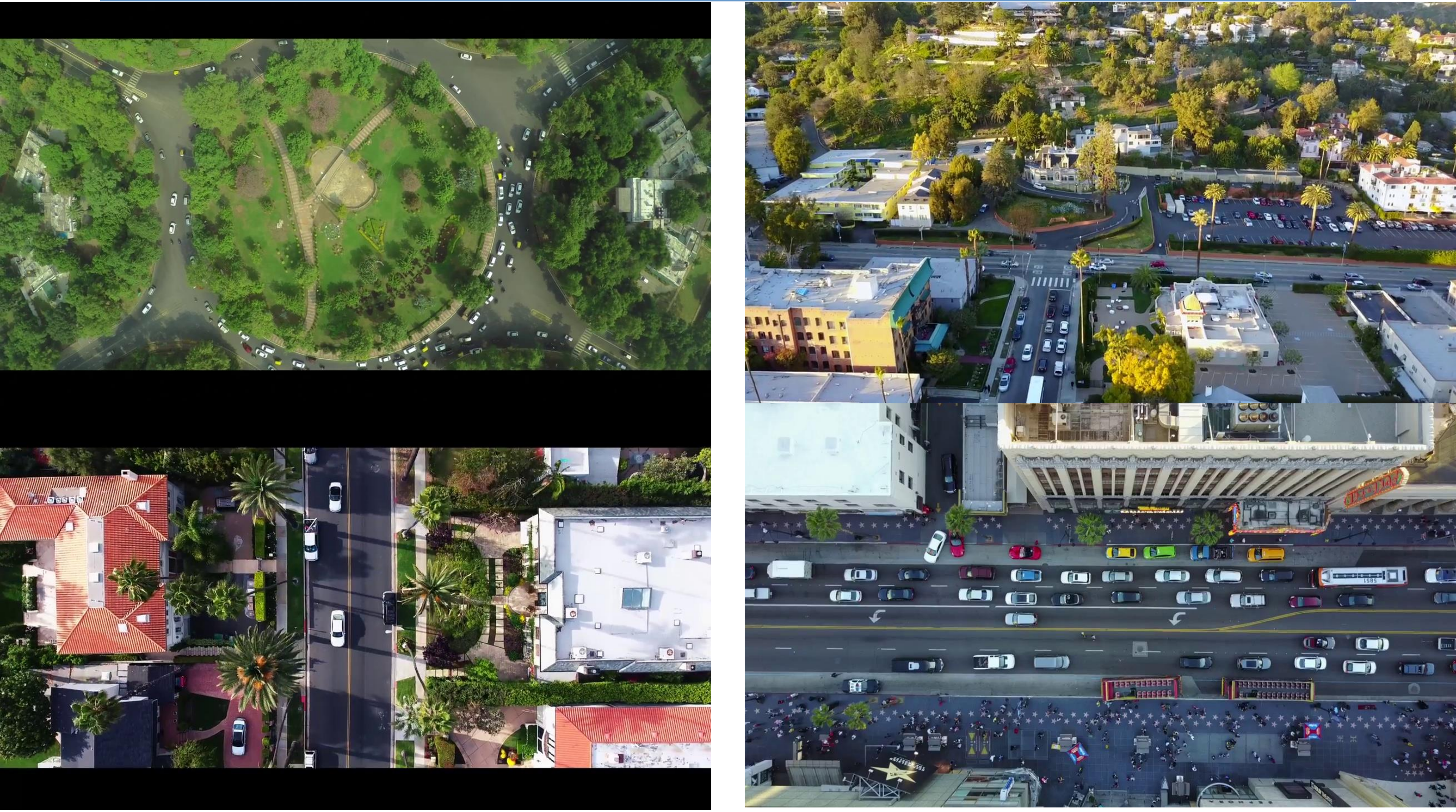


Fig. 6. Sample images of our ABD dataset

Evaluation Metrics

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

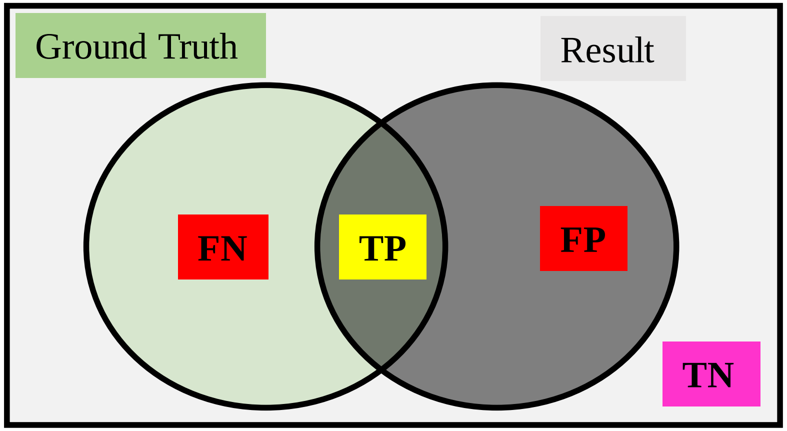
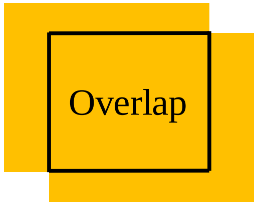
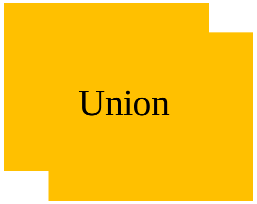


Fig. 7. Venn diagram representing true positives, true negatives, false positives and false negatives.

Intersection Over Union

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union}$$





Evaluation Metrics

- Mean Average Precision (**mAP**)

Mean of Average Precisions: AP@ [IOU 0.5 : 0.95] corresponds to the average AP for IoU from 0.5 to 0.95 with a step size of 0.05.

Compute AP for each class and take the average of all the APs - **mAP**

Quantitative Comparison

Table 2. Comparative detection performance in terms of mean average precision (mAP) of the proposed SSSDet and existing state-of-the-art approaches

Method	VEDAI	DLR	DOTA	Complete
YOLOv2_416	9.08	9.61	33.36	28.86
YOLOv2_608	25.12	26.81	47.45	48.04
Faster R-CNN	34.82	20.04	42.29	38.02
YOLOv3_416	32.07	52.11	74.46	70.35
YOLOv3_608	38.98	54.49	76.60	75.21
RetinaNet	43.47	54.77	73.77	71.28
YOLOv3-tiny_416	11.10	26.42	47.88	46.73
YOLOv3-tiny_608	31.73	39.74	65.89	59.17
SSSDet	**45.97**	**58.25**	**79.52**	**77.22**

**This table is from SSSDet paper*

Efficiency Analysis

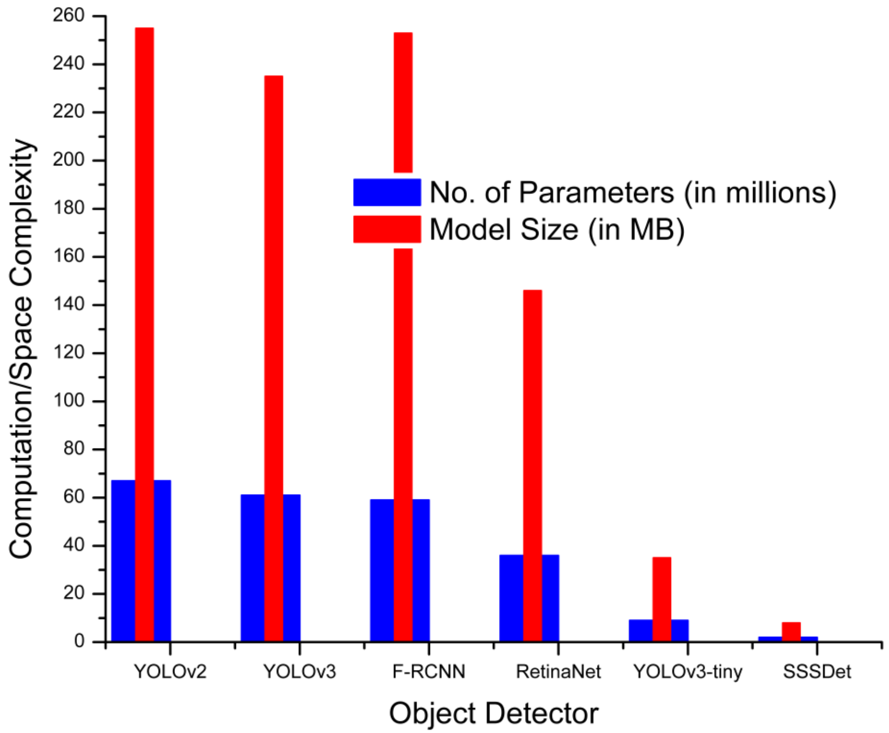


Fig. 8. Computation and space complexity comparison of the proposed SSSDet with the existing state-of-the-art object detectors.

References

[1] K. Liu and G. Mattyus, “Fast multiclass vehicle detection on aerial images,” IEEE Geosci. Remote Sens. Lett., vol. 12, no. 9, pp. 1938-1942, 2015.

[2] M. ElMikaty and T. Stathaki, “Detection of Cars in HighResolution Aerial Images of Complex Urban Environments,” IEEE Trans. Geosci. Remote Sens. , vol. 55, no. 10, pp. 5913-5924, 2017.

[3] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell. , vol. 39, no. 6, pp. 1137-1149, 2017.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. CVPR, 2016.

[5] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in Proc. CVPR, 2017.

[6] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.

[7]] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, “Focal loss for dense object detection,” in Proc. ICCV, 2017

[8] S. Razakarivony and F. Jurie, “Vehicle detection in aerial imagery: a small target detection benchmark,” J. Visual Communicat. Image Representation, vol. 34, pp. 187-203, 2016.

[9] G. S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo and L. Zhang, “DOTA: A Large-scale Dataset for Object Detection in Aerial Images,” in Proc. CVPR, 2018.