AVDNet: A Small-Sized Vehicle Detection Network for Aerial Visual Data

IEEE Geoscience and Remote Sensing Letters, to be published

By: Murari Mandal

AVDNet

- AVDNet: A Small-Sized Vehicle Detection Network for Aerial Visual Data

- We proposed a single-stage vehicle detection network AVDNet to robustly detect small-sized vehicles in aerial scenes.

- Introduced ConvRes residual blocks at multiple scales to alleviate the problem of vanishing features for smaller objects caused due to the inclusion of deeper convolutional layers

AVDNet

- We proposed a recurrent-feature aware visualization (RFAV) technique to analyze the network behavior.

- Created a new airborne image dataset (ABD) by annotating 1396 new objects in 79 aerial images for our experiments.

- The effectiveness of AVDNet is validated on VEDAI, DLR-3K, DOTA and the combined (VEDAI, DLR-3K, DOTA and ABD) dataset

AVDNet: Architecture

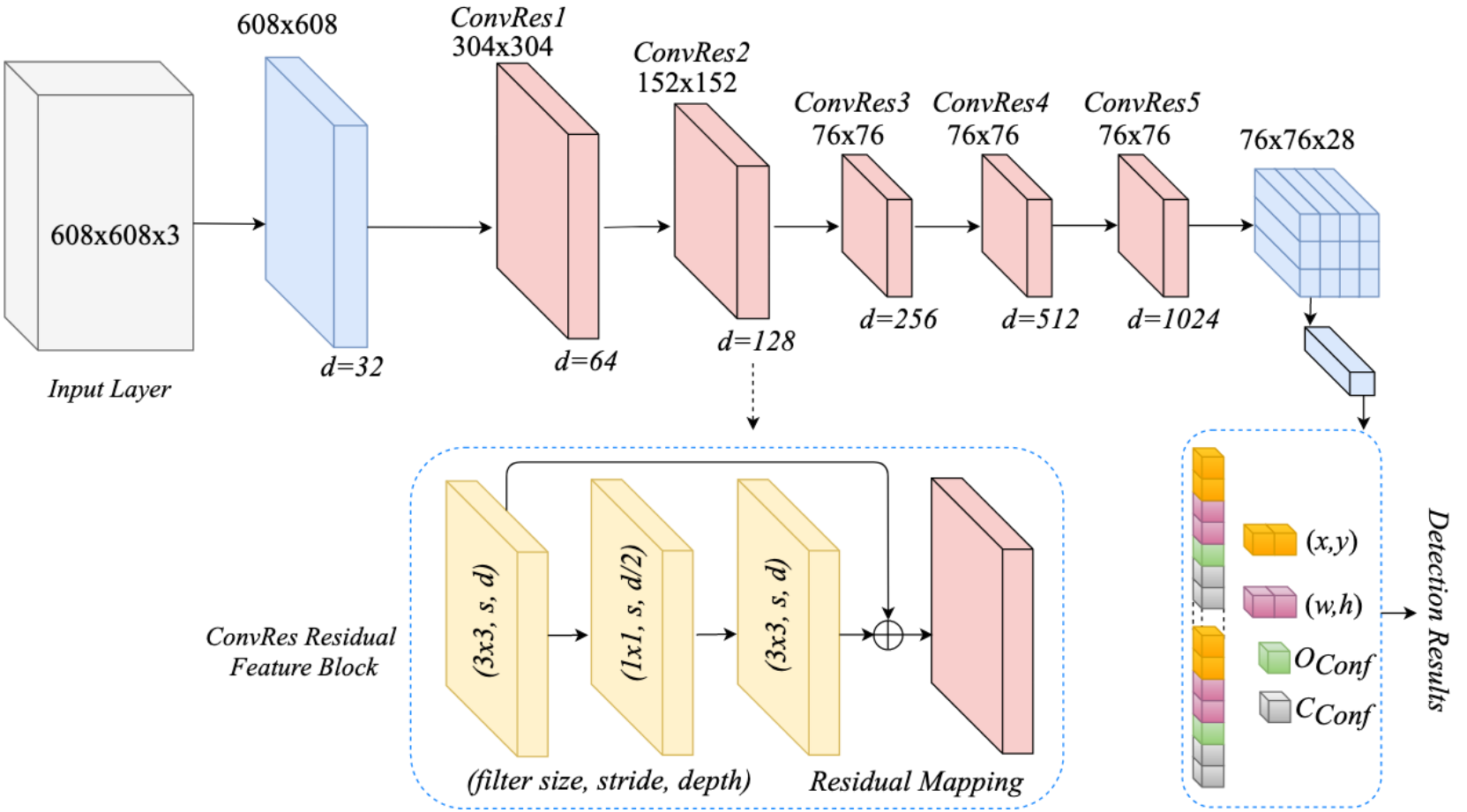


Fig. 1. The proposed AVDNet vehicle detection framework (for 2 vehicle classes). x, y, w, h: bounding box coordinates (centre location, width, height), O_{Conf}: object confidence, C_{Conf}: class confidence.

RFAV Visualization

- RFAV - Recurrent-feature aware visualization

For d feature maps in *conv* layer l, the recurrent-feature aware visual representation is computed using the equation.

$$RFAV_l(a,b) = \arg\max_z(H_l^{(a,b)}(z)); z \in [0,255]$$

where $\arg\max(\cdot)$ collects the histogram bin index of the maximum value. The temporal histogram $H_l^{(a,b)}(\cdot)$ at pixel location (a,b) of feature maps F is calculated using the equation.

$$H_l^{(a,b)}(z) = \sum_{k=1}^{d} \delta\left(F_l^k(a,b) - z\right); z \in [0,255]$$

RFAV Visualization

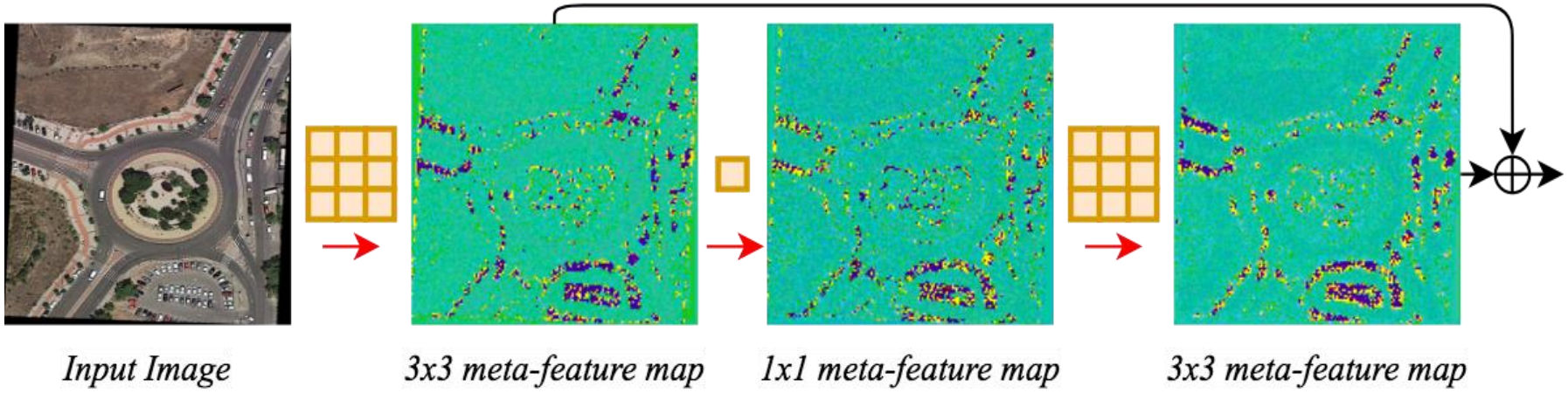


Fig. 2. RFAV visualization of the *ConvRes3* block of AVDNet.
*The composite visual representation of the multiple feature maps generated at the end of a conv3 operation is shown.

Activation Layers Visualization

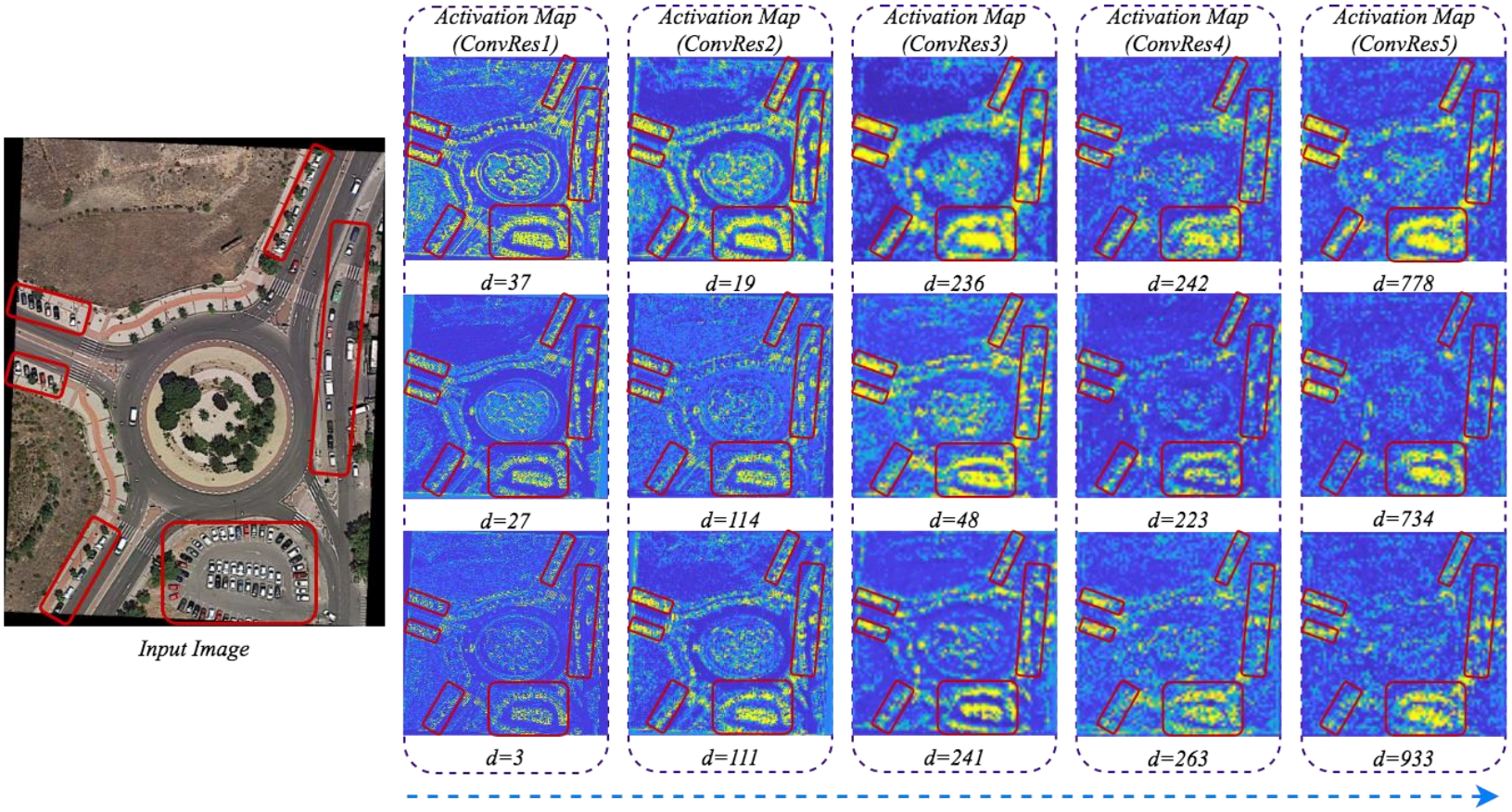
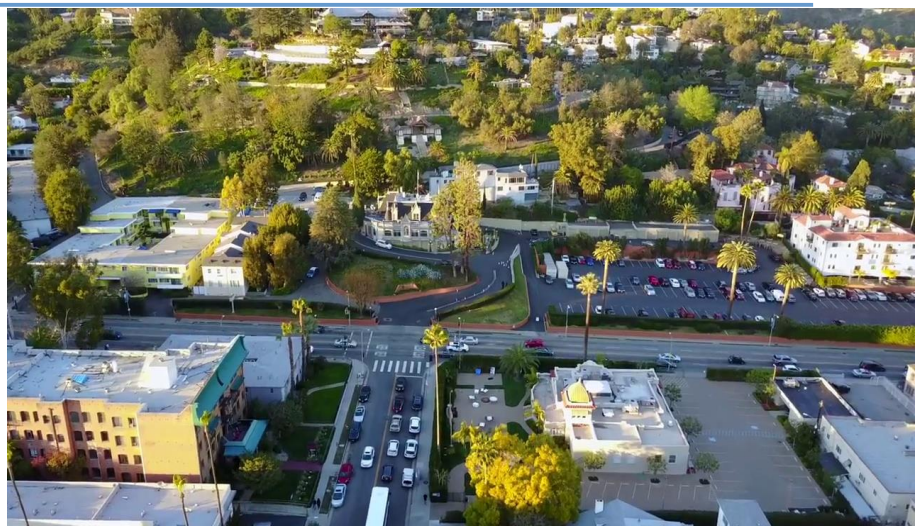


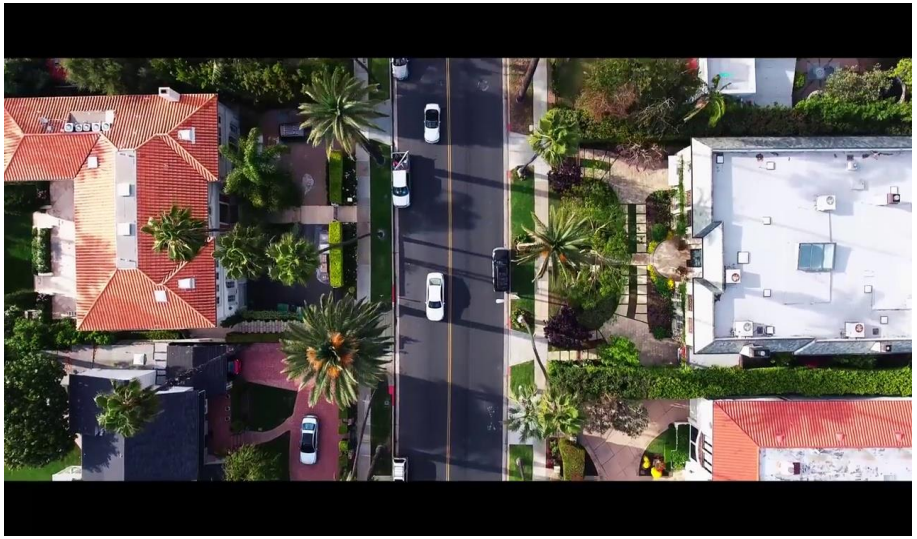
Fig. 3. Sample activation responses after each ConvRes block of AVDNet. The red boxes highlight the activations in different regions for the presence of vehicles in the input image, *d=depth of the activation map*

Datasets

- VEDAI
- DLR-3K
- DOTA
- ABD
- Complete Dataset (VEDAI + DLR-3K + DOTA + ABD)

ABD Dataset





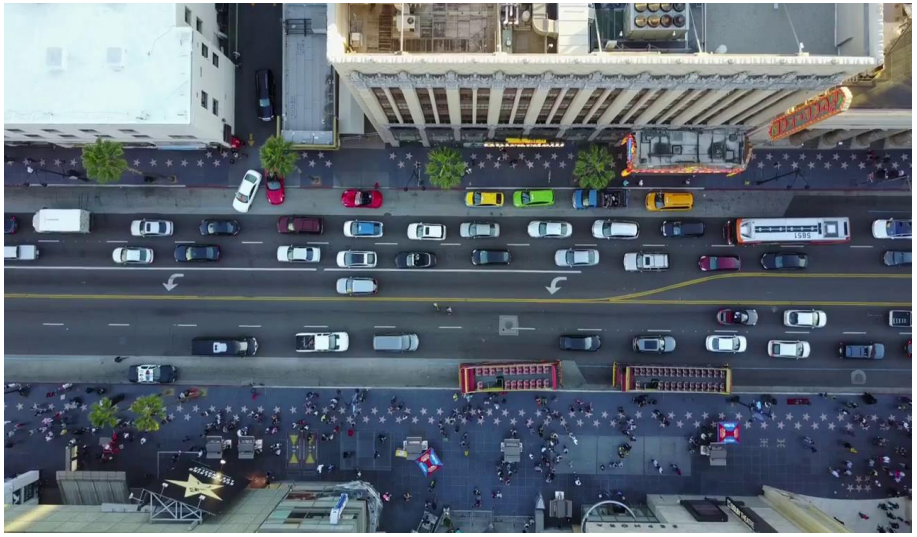


Fig. 4. Sample images of our ABD dataset

Data Description

TABLE I
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 our AVDNet paper*

Evaluation Metrics

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

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

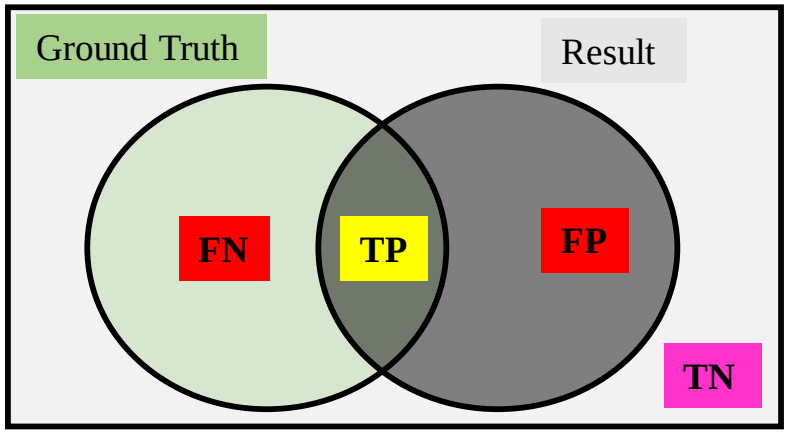


Fig. 5. Venn diagram representing true positives, true negatives, false positives and false negatives.

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 Analysis

TABLE II
COMPARATIVE DETECTION PERFORMANCE OF THE AVDNET AND EXISTING STATE-OF-THE-ART TECHNIQUES

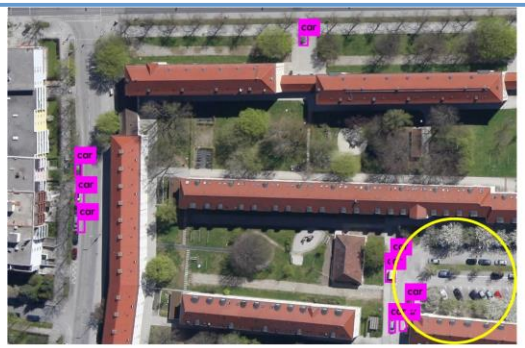
Method/mAP (%)	**VEDAI**	**DLR-3K**	**DOTA**	**Complete**
Coupled R-CNN	12.04	11.74	25.60	19.66
YOLOv2_416x416	9.08	9.61	33.36	28.86
YOLOv2_608x608	25.12	26.81	47.45	48.04
Faster R-CNN	34.82	20.04	42.29	38.02
YOLOv3_416x416	32.07	52.11	74.46	70.35
YOLOv3_608x608	38.98	54.49	76.60	75.21
RetinaNet	43.47	54.77	73.77	71.28
AVDNet	**51.95**	**56.24**	**79.65**	**80.02**

Comparative Efficiency Analysis

TABLE III
COMPUTATIONAL AND SPACE COMPLEXITY OF THE PROPOSED METHOD AND EXISTING STATE-OF-THE-ART TECHNIQUES

Method	No. of params. (in millions)	Model size
YOLOv2	67	255 MB
YOLOv3	61	235 MB
Faster R-CNN	59	253 MB
RetinaNet	36	146 MB
AVDNet	**13**	**53 MB**

Qualitative Analysis



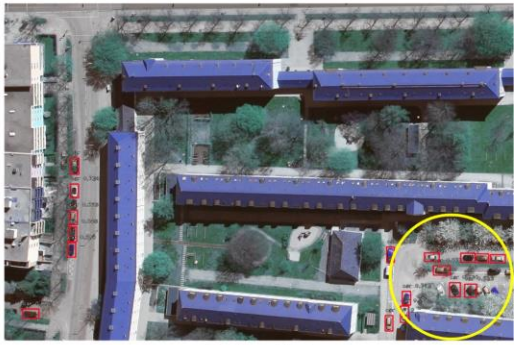
YOLOv2_416x416

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 AVDNet	24	0	0	0

YOLOv2_608x608

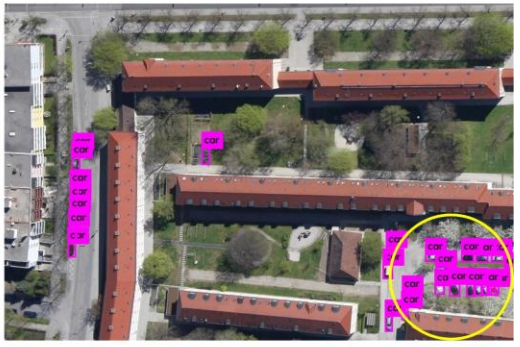
YOLOv3_416x416

Input



RetinaNet

YOLOv3_608x608



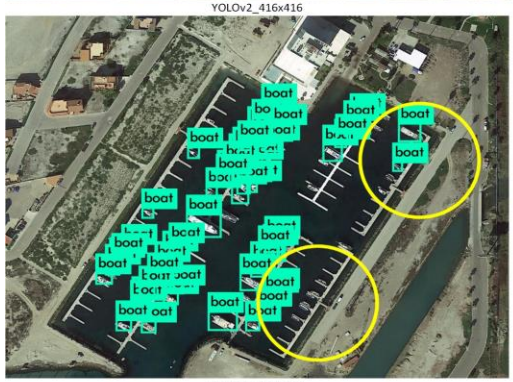
Proposed AVDNet

Qualitative Analysis

YOLOv2_416x416

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 AVDNet	4	0	0	58

YOLOv2_608x608



YOLOv3_416x416

Input

RetinaNet

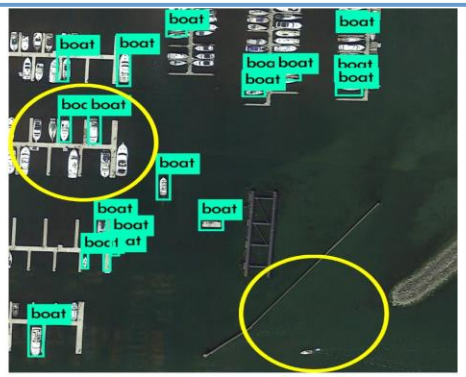
YOLOv3_608x608

Proposed AVDNet

Qualitative Analysis

YOLOv2_416x416

Method/ # Detected Objects	Car	Heavy Vehicle	Plane	Boat
YOLOv2_416	0	0	0	21
YOLOv2_608	0	0	0	17
YOLOv3_416	0	0	0	81
YOLOv3_608	0	0	0	26
RetinaNet	0	0	0	69
Proposed AVDNet	0	0	0	90



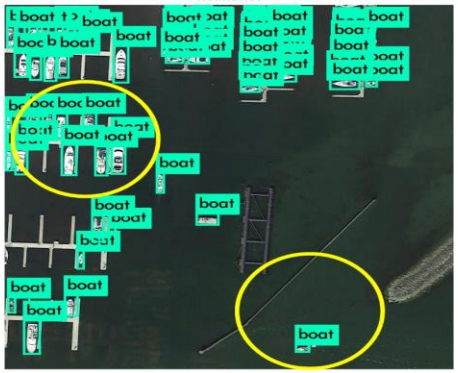
YOLOv2_608x608

YOLOv3_416x416

Input

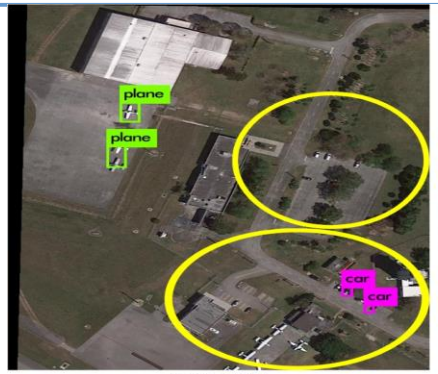
RetinaNet

YOLOv3_608x608



Proposed AVDNet

Qualitative Analysis



YOLOv2_416x416

Method/ # Detected Objects	Car	Heavy Vehicle	Plane	Boat
YOLOv2_416	2	0	2	0
YOLOv2_608	7	0	7	0
YOLOv3_416	0	0	7	0
YOLOv3_608	2	0	6	0
RetinaNet	0	0	8	0
Proposed AVDNet	9	0	8	0

YOLOv2_608x608

YOLOv3_416x416

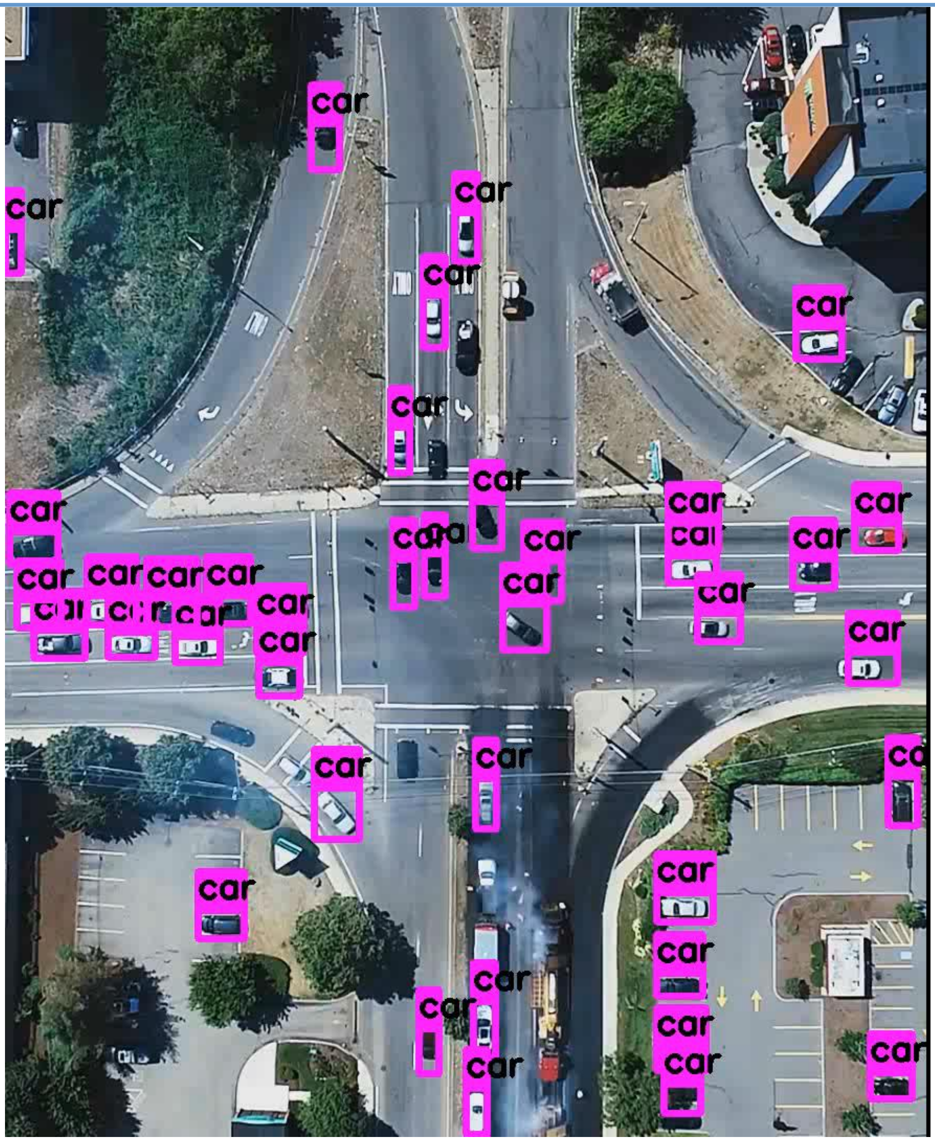

Input


RetinaNet

YOLOv3_608x608

Proposed AVDNet

Qualitative Results



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.