Εθνικό Μετσόβιο Πολυτεχνείο
Επιστήμη Δεδομένων και Μηχανική Μάθηση (ΔΠΜΣ)

Deep Object Detectors for Remote Sensing Data

Κωνσταντίνος Γιαννακέλος
A.M.: 03400047

Introduction

Object Detection in Aerial Imagery

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Horizontal Object Detection

Rotated Object Detection

Object Recognition in Aerial Imagery

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Aerial Image

Object Detection

Instance Segmentation

Thesis Goals

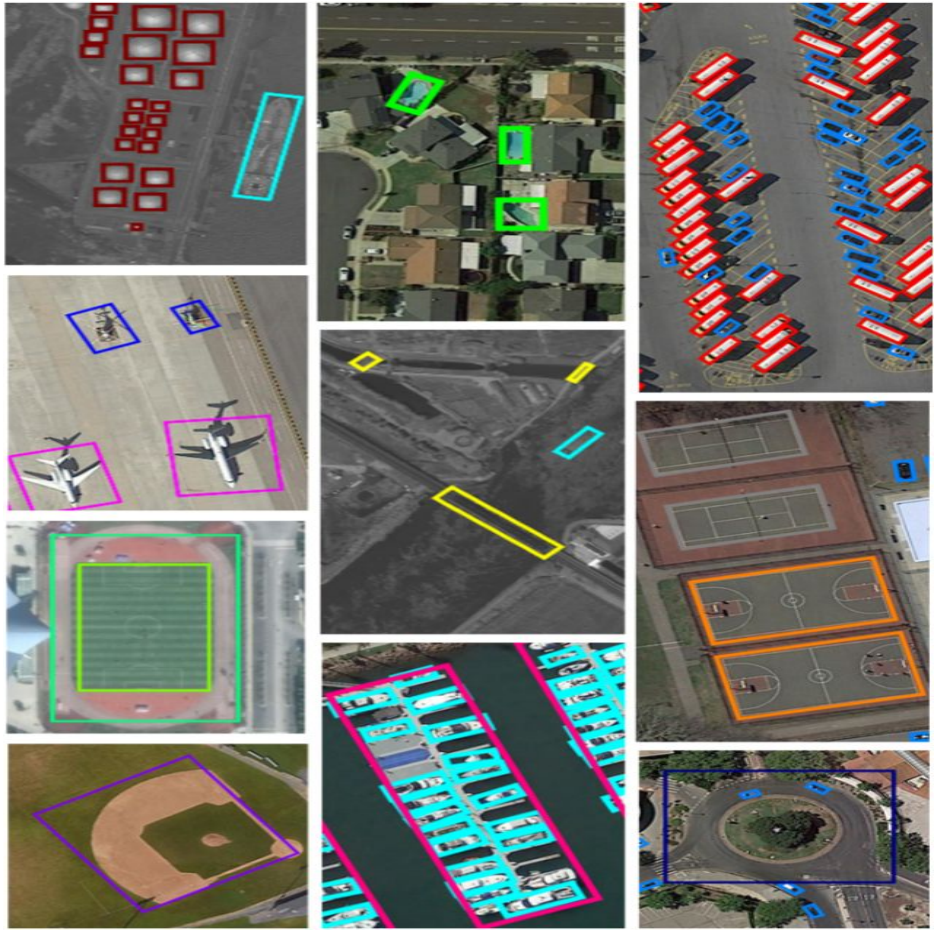
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Study current emerging challenges Object Detection faces in Aerial Imagery
- Investigate the state-of-the-art methods currently used and how they approach these challenges
- Evaluate their performance on a large scale Aerial Imagery dataset
- Propose a method to perform Instance Segmentation

Challenges - Aerial Imagery

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

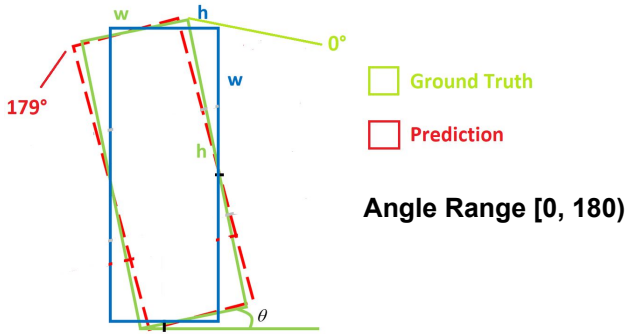
- Large Image Sizes
- Object Size Variations
- Crowded Objects
- Large Aspect Ratio
- Imbalanced Frequency



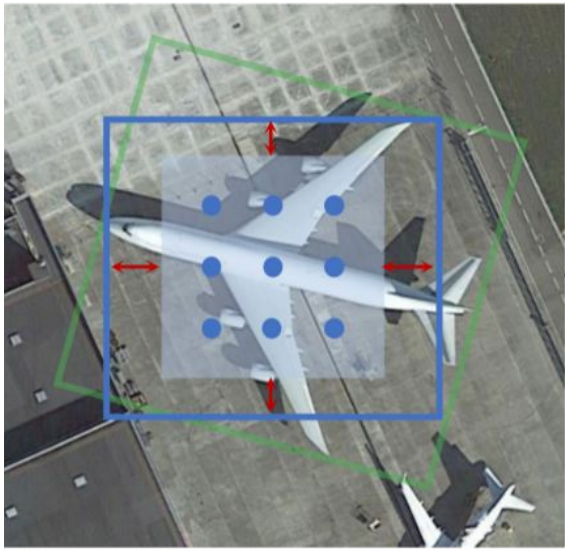
Challenges - Rotated Object Detection

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- **Periodicity of Angle:** the periodic nature causes problems in the values of the loss function in the case of boundary discontinuity



- **Feature Misalignment:** Anchors and bounding boxes are misaligned with the features extracted by the backbone which are x-axis aligned with fixed receptive field

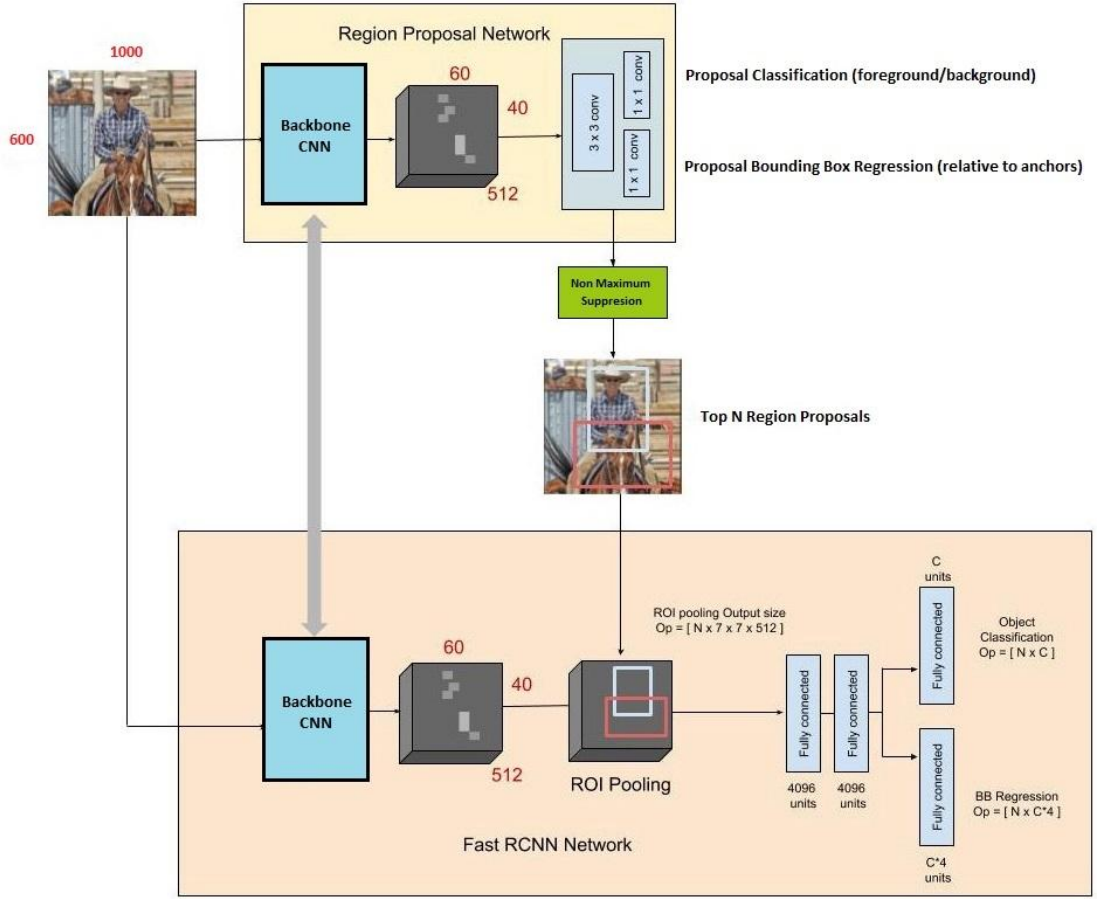


Literature Review

Faster R-CNN

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

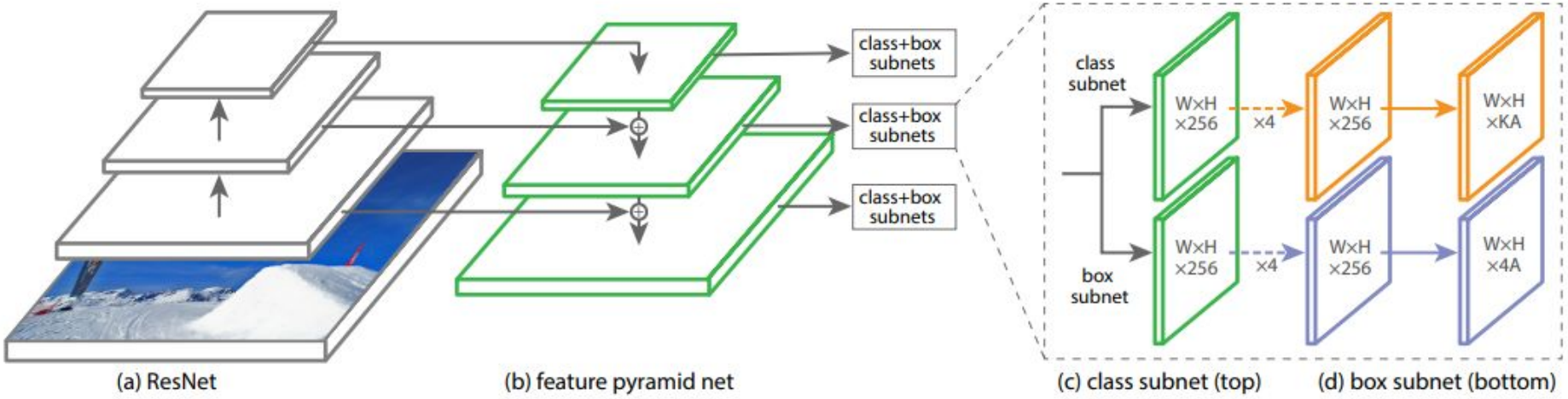
- Two Stage Detector
- First stage uses Region Proposal Network to crop the regions which may contain an object from the image
- Second stage classifies their categories and tightens the bounding boxes



RetinaNet

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

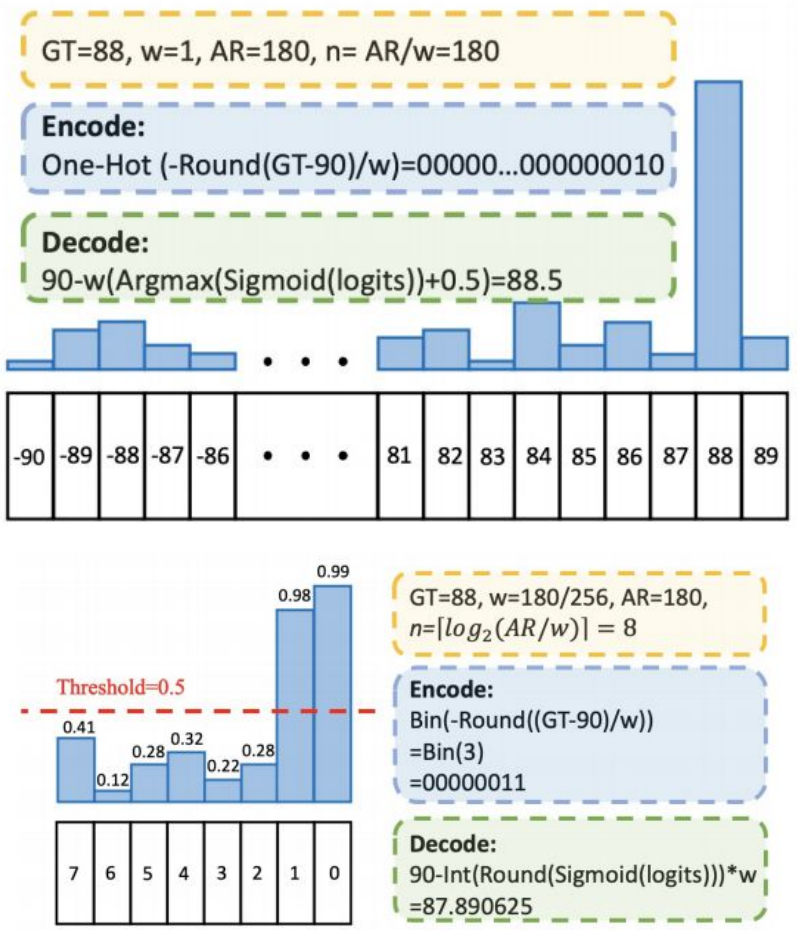
- One Stage Detector
- Uses Feature Pyramid Network to extract multiscale features
- Use Focal Loss to deal with class imbalance



Dense Coded Label - DCL RetinaNet

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

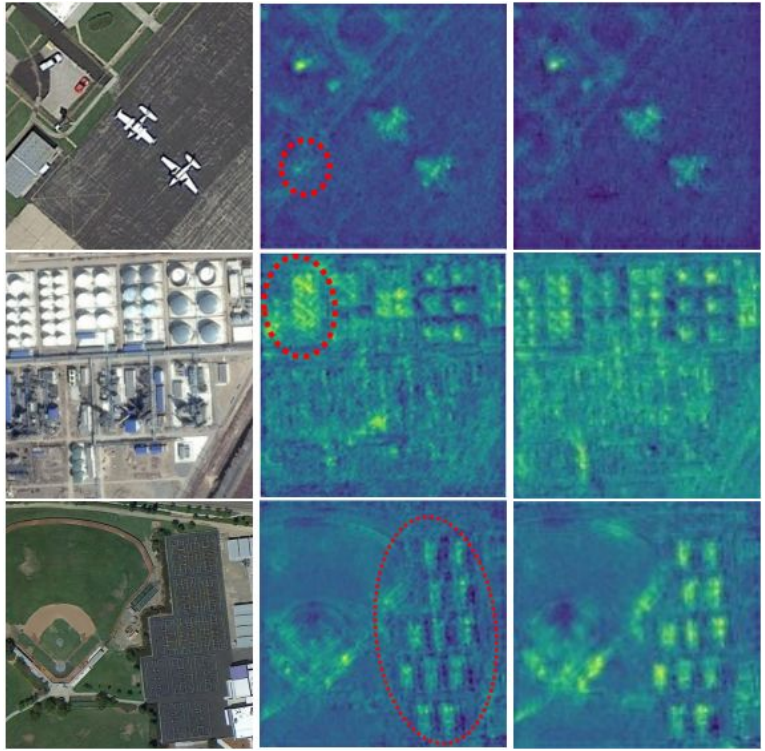
- Predicts angle by solving a classification task
- [0, 180) angle range is divided into finite number of categories
- 2^8 = 256 categories, where each division interval represents $\boldsymbol{\omega}$ = 180/256 = 0.703125^0
- Uses Binary and Gray coded labels to reduce the complexity of the classification task



Small, Cluttered, Rotated - SCRDET++

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Uses **Instance-level Denoising Module** for suppressing instance noise after feature extraction
- Aims to solve inter-class feature coupling and confusion between intra-class and background
- Solves a Coarse Semantic Segmentation task to create denoised feature maps
- The denoised feature maps are ordered and contain feature information per category.
- Uses IOU Smooth L1 Loss to eliminate the large loss values produced by the periodicity of the angle



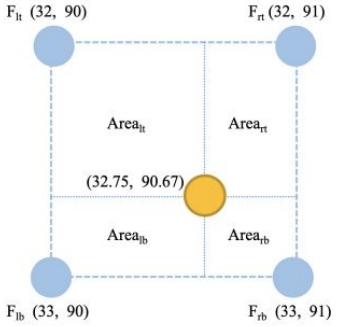
Original Image | Before Denoising | After Denoising

Refined Rotation RetinaNet - R^3Det

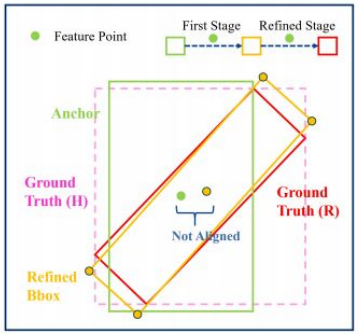
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Tries to solve the misalignment problem
- Uses Feature Refinement Module to reconstruct feature maps pixel-by-pixel based on their most confident detection
- Uses SkewIoU loss to solve the angle periodicity problem (similar to IoULoss)

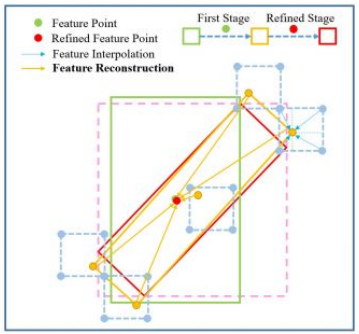
(a) Original image



(b) Feature interpolation



(c) Refine box with misaligned feature due to bounding box location changes.

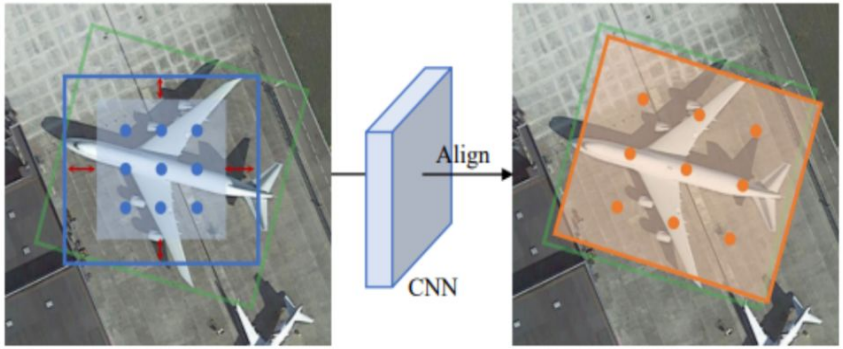


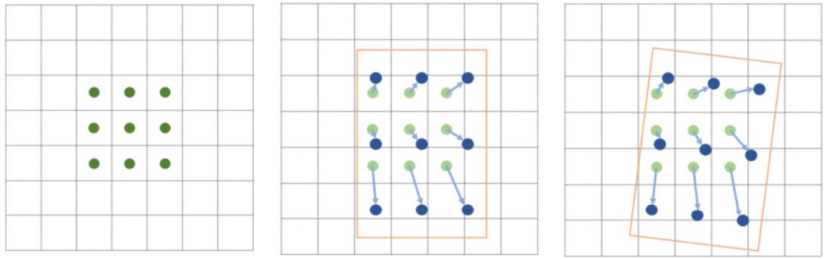
(d) Refine box with aligned features by reconstructing the feature map.

Single-Shot Alignment Network - S^2ANet

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Tries to solve the misalignment problem
- Converts horizontal anchors to rotated
- Uses Feature Alignment Module to reconstruct feature maps by applying Alignment Convolution on the most confident detection
- Alignment Convolution samples parts of object compared to 2D Convolution which samples fixed parts of an image





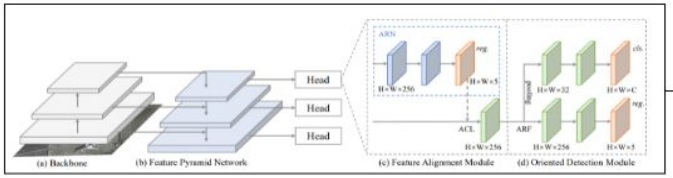
Proposed Method

Our contribution - Instance Segmentation - S^{2}AU-Net

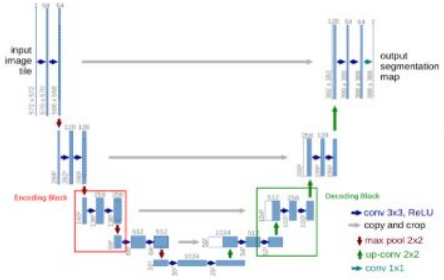
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

S^{2}AU-Net

S^{2}ANet

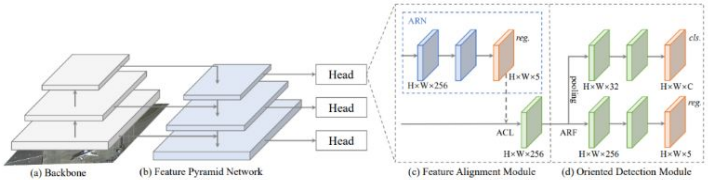


U-Net



Instance Segmentation

S^{2}ANet

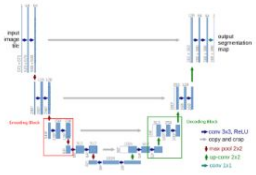


Output BBoxes/Mask Proposals

Assign top-100 predicted bboxes to ground truths based on MAX IOU (at least 50%) + first 2000 ground truth masks

Resize to 32 x 32

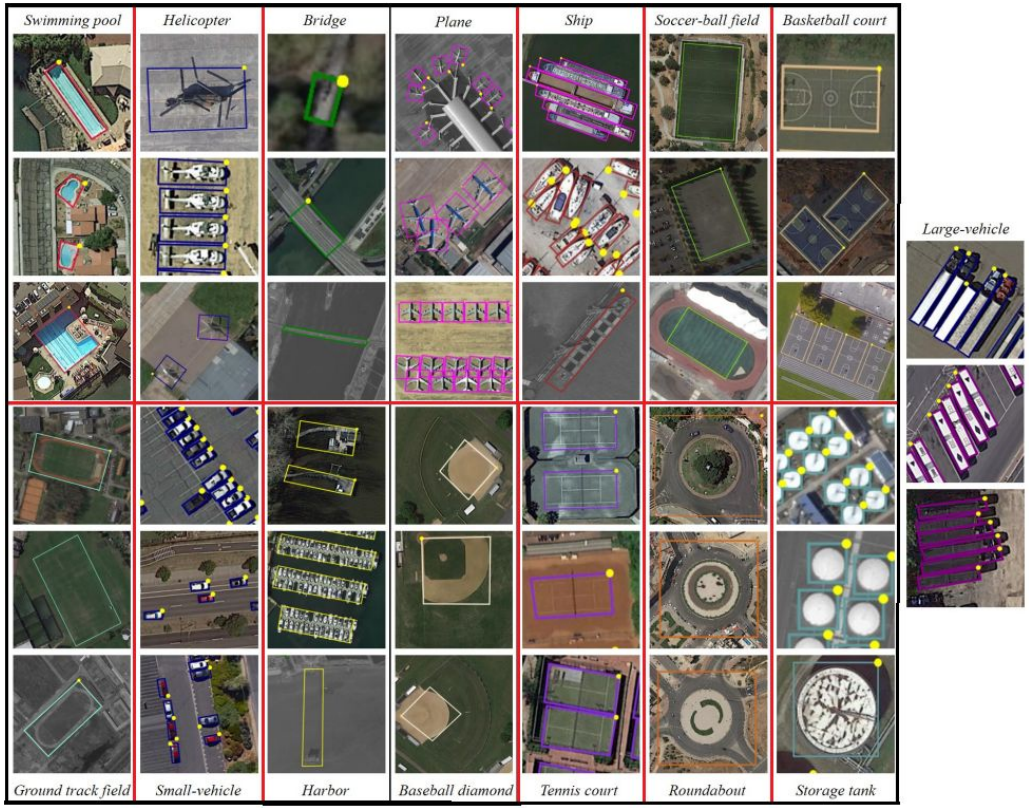
U-Net Head



DOTA: A large scale Dataset for Object DeTection in Aerial Images

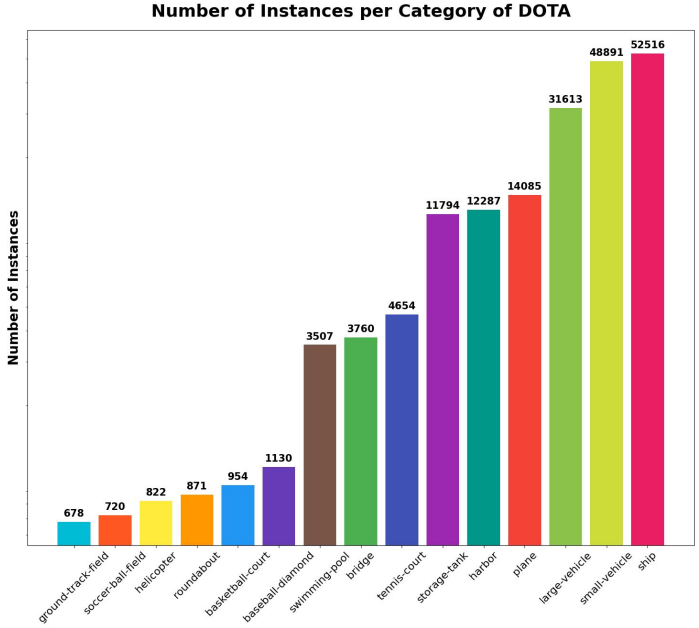
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

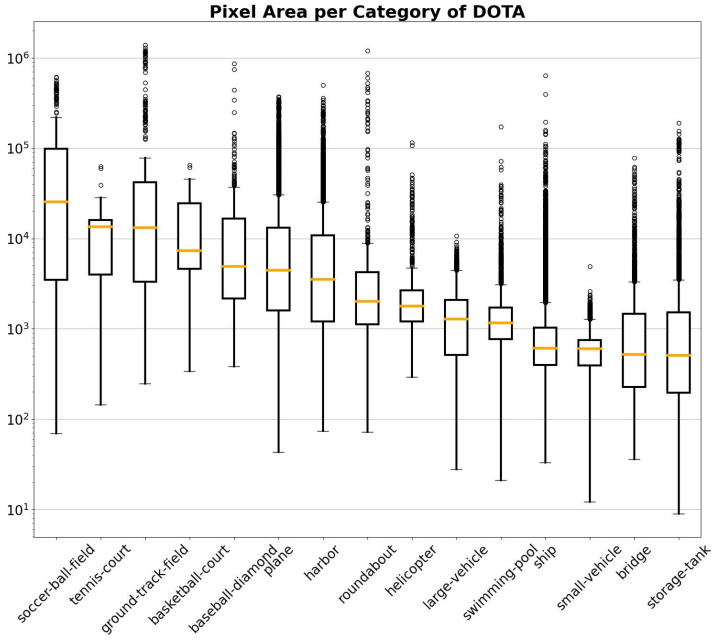
- Contains **2806** aerial images collected from multiple sensors and platforms (e.g. Google Earth) with multiple resolutions.
- Images from **800 × 800** to **4000 × 4000** pixels
- **15** common object categories
- **188.282** total objects to detect

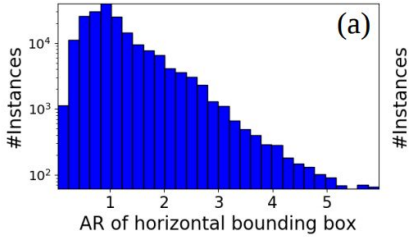


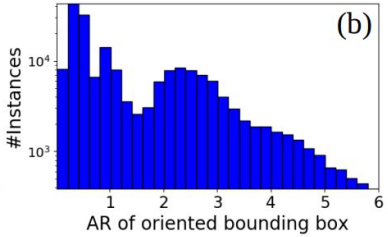
DOTA: Object Statistics

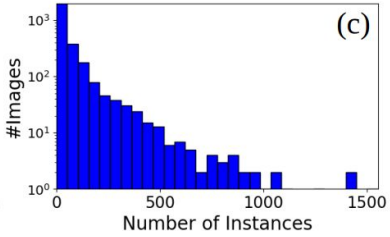
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση







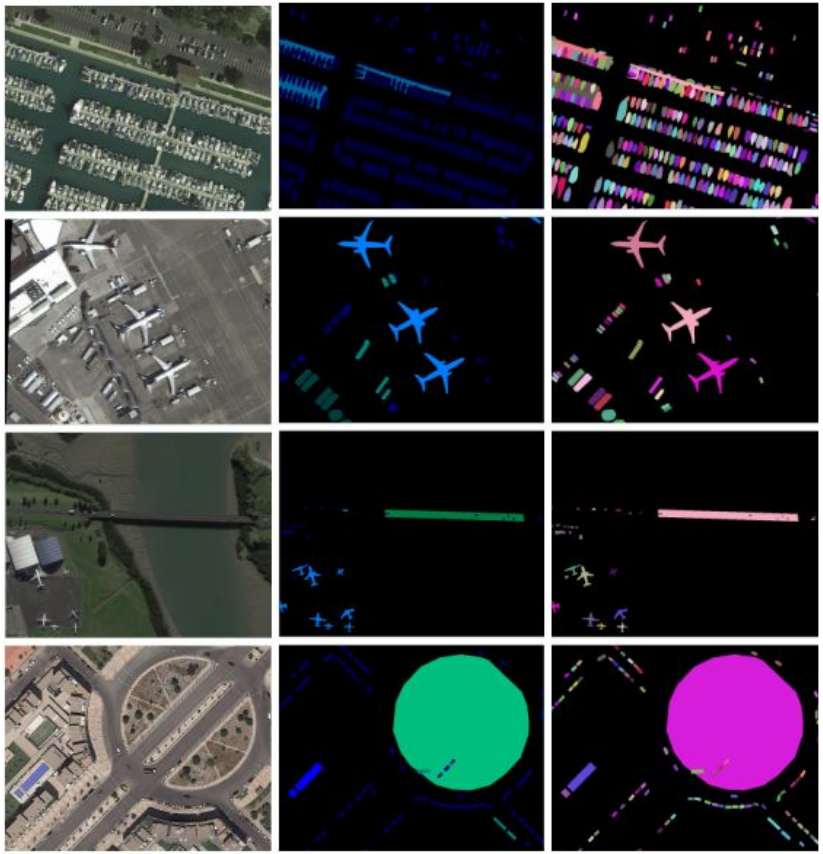




iSAID: A large scale Instance Segmentation in Aerial Images Dataset

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Is a complement dataset to DOTA
- Contains RGB masks used for Semantic/Instance Segmentation
- Annotated from scratch, contains 655,451 instances, 250% more than DOTA

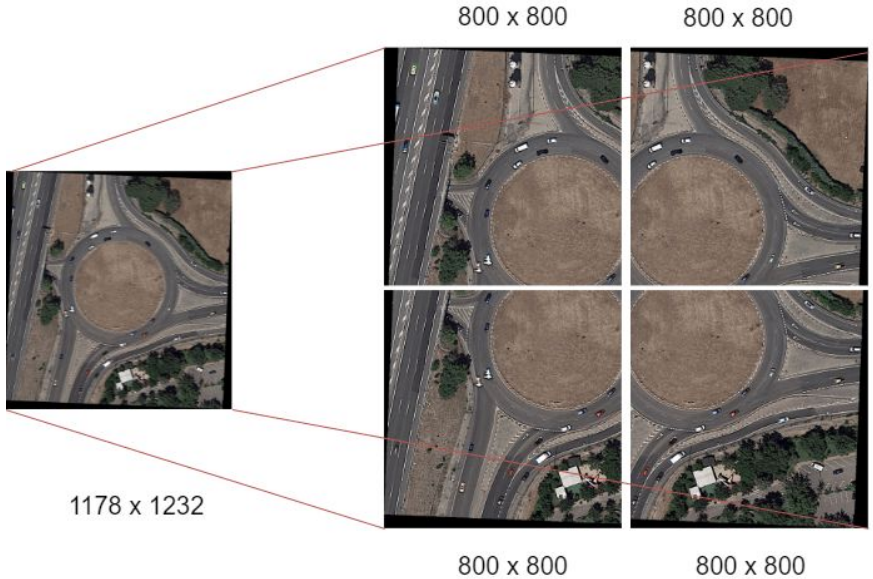


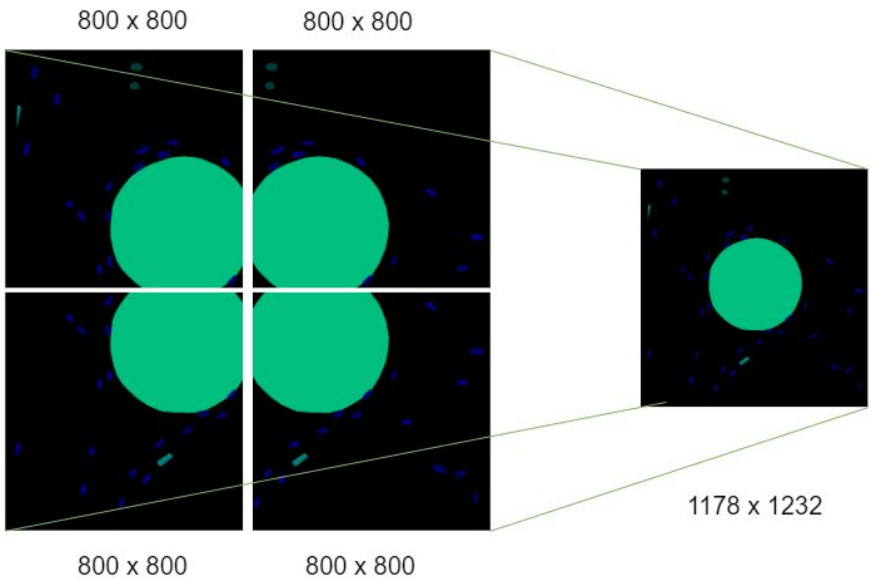
(a) Original DOTA Image (b) Semantic Segmentation Mask (c) Instance Segmentation Mask

Data Preprocessing

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Images are split into smaller patches with 200 pixels overlap between them
- 1024 x 1024 for DOTA
- 800 x 800 for iSAID





Experiments & Results

Object Detection Experiments

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Trained from scratch a Faster R-CNN, RetinaNet and S^2ANet
- Acquired pretrained versions of DCL RetinaNet, R^3Det, SCRDET++
- All methods were trained on DOTA train + val sets and evaluated on test set
- Used mean Average Precision (mAP) as the main metric at 50% IoU

Object Detection Quantitative Results (1/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.1: Evaluation results on DOTA test set. * indicates multiscale training.

Method	Backbone	Train	Input Size	Speed (FPS)	$mAP_{50\%IoU}$
RetinaNet	ResNet 50	Ours	1024×1024	6.4	**65.87** %
Faster R-CNN	ResNet 50	Ours	1024×1024	4.2	**68.93** %
S^2ANet [33]	ResNet 50	Ours	1024×1024	5.9	**72.43** %
DCL RetinaNet* [32]	ResNet 152	Pretrained	800×800	< 1	**72.58** %
R^3DET [30]	ResNet 152	Pretrained	800×800	< 1	**74.74** %
SCRDET++* [29]	ResNet 152	Pretrained	800×800	< 1	**77.73** %

Object Detection Quantitative Results (2/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.1: Evaluation results on DOTA test set. * indicates multiscale training.

Method	Backbone	Train	Input Size	Speed (FPS)	$mAP_{50\%IoU}$
RetinaNet	ResNet 50	Ours	1024×1024	6.4	**65.87** %
Faster R-CNN	ResNet 50	Ours	1024×1024	4.2	**68.93** %
S^2ANet [33]	ResNet 50	Ours	1024×1024	5.9	**72.43** %
DCL RetinaNet* [32]	ResNet 152	Pretrained	800×800	< 1	**72.58** %
R^3DET [30]	ResNet 152	Pretrained	800×800	< 1	**74.74** %
SCRDET++* [29]	ResNet 152	Pretrained	800×800	< 1	**77.73** %

Lowest mAP
Highest Inference Speed

Object Detection Quantitative Results (3/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.1: Evaluation results on DOTA test set. * indicates multiscale training.

Method	Backbone	Train	Input Size	Speed (FPS)	$mAP_{50\%IoU}$
RetinaNet	ResNet 50	Ours	1024×1024	6.4	**65.87** %
Faster R-CNN	ResNet 50	Ours	1024×1024	4.2	**68.93** %
S^2ANet [33]	ResNet 50	Ours	1024×1024	5.9	**72.43** %
DCL RetinaNet* [32]	ResNet 152	Pretrained	800×800	< 1	**72.58** %
R^3DET [30]	ResNet 152	Pretrained	800×800	< 1	**74.74** %
SCRDET++* [29]	ResNet 152	Pretrained	800×800	< 1	**77.73** %

Better than RetinaNet
at the cost of 2.2 FPS

Object Detection Quantitative Results (4/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.1: Evaluation results on DOTA test set. * indicates multiscale training.

Method	Backbone	Train	Input Size	Speed (FPS)	$mAP_{50\%IoU}$
RetinaNet	ResNet 50	Ours	1024×1024	6.4	**65.87** %
Faster R-CNN	ResNet 50	Ours	1024×1024	4.2	**68.93** %
S^2ANet [33]	ResNet 50	Ours	1024×1024	5.9	**72.43** %
DCL RetinaNet* [32]	ResNet 152	Pretrained	800×800	< 1	**72.58** %
R^3DET [30]	ResNet 152	Pretrained	800×800	< 1	**74.74** %
SCRDET++* [29]	ResNet 152	Pretrained	800×800	< 1	**77.73** %

Better than Faster R-CNN in both mAP and speed

Good trade-off

Object Detection Quantitative Results (5/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.1: Evaluation results on DOTA test set. * indicates multiscale training.

Method	Backbone	Train	Input Size	Speed (FPS)	$mAP_{50\%IoU}$
RetinaNet	ResNet 50	Ours	1024×1024	6.4	**65.87** %
Faster R-CNN	ResNet 50	Ours	1024×1024	4.2	**68.93** %
S^2ANet [33]	ResNet 50	Ours	1024×1024	5.9	**72.43** %
DCL RetinaNet* [32]	ResNet 152	Pretrained	800×800	< 1	**72.58** %
R^3DET [30]	ResNet 152	Pretrained	800×800	< 1	**74.74** %
SCRDET++* [29]	ResNet 152	Pretrained	800×800	< 1	**77.73** %

Better mAP
multi-scale training
poor inference times

Object Detection Quantitative Results (6/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.2: Evaluation results on DOTA test set per category. The short names for categories are defined as: PL-Plane, BD–Baseball Diamond, BR-Bridge, GTF–Ground Track-Field, SV–Small Vehicle, LV–Large Vehicle, SH-Ship, TC–Tennis Court, BC–Basketball Court, ST–Storage tank, SBF–Soccer-Ball Field, RA–Roundabout, SP–Swimming Pool, HB-Harbor and HC–Helicopter.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HB	SP	HC
RetinaNet	89.00	78.50	31.20	66.51	75.49	60.97	75.81	90.74	81.18	75.56	54.67	42.97	51.96	63.21	50.33
Faster RCNN	88.38	74.93	44.65	57.76	73.72	71.77	77.67	90.67	82.74	82.74	47.36	59.40	63.31	66.17	52.70
S^2ANet [33]	88.63	81.57	44.99	69.73	76.46	79.29	87.08	**90.82**	84.64	84.53	59.01	42.28	66.27	65.25	65.68
DCL RetinaNet* [32]	89.09	**84.18**	47.17	69.31	71.34	58.36	72.92	90.79	86.65	85.67	64.44	**64.78**	64.50	73.08	66.47
R^3DET [30]	88.88	82.74	51.87	65.50	76.13	**81.49**	**87.39**	90.79	85.07	84.48	59.60	61.49	68.08	73.31	64.20
SCRDET++* [29]	**89.82**	84.10	**56.71**	**71.42**	**79.29**	73.35	78.56	90.72	**87.86**	**87.18**	**68.47**	64.55	**76.47**	**82.02**	**75.36**

Object Detection Quantitative Results (7/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.2: Evaluation results on DOTA test set per category. The short names for categories are defined as: PL-Plane, BD–Baseball Diamond, BR-Bridge, GTF–Ground Track-Field, SV–Small Vehicle, LV–Large Vehicle, SH-Ship, TC–Tennis Court, BC–Basketball Court, ST–Storage tank, SBF–Soccer-Ball Field, RA–Roundabout, SP–Swimming Pool, HB-Harbor and HC–Helicopter.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HB	SP	HC
RetinaNet	89.00	78.50	31.20	66.51	75.49	60.97	75.81	90.74	81.18	75.56	54.67	42.97	51.96	63.21	50.33
Faster-RCNN	88.38	74.93	44.65	57.76	73.72	71.77	77.67	90.67	82.74	82.74	47.36	59.40	63.31	66.17	52.70
S^2ANet [33]	88.63	81.57	44.99	69.73	76.46	79.29	87.08	**90.82**	84.64	84.53	59.01	42.28	66.27	65.25	65.68
DCL-RetinaNet* [32]	89.09	**84.18**	47.17	69.31	71.34	58.36	72.92	90.79	86.65	85.67	64.44	**64.78**	64.50	73.08	66.47
R^3DET [30]	88.88	82.74	51.87	65.50	76.13	**81.49**	**87.39**	90.79	85.07	84.48	59.60	61.49	68.08	73.31	64.20
SCRDET++* [29]	**89.82**	84.10	**56.71**	**71.42**	**79.29**	73.35	78.56	90.72	**87.86**	**87.18**	**68.47**	64.55	**76.47**	**82.02**	**75.36**

Object Detection Quantitative Results (8/8)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

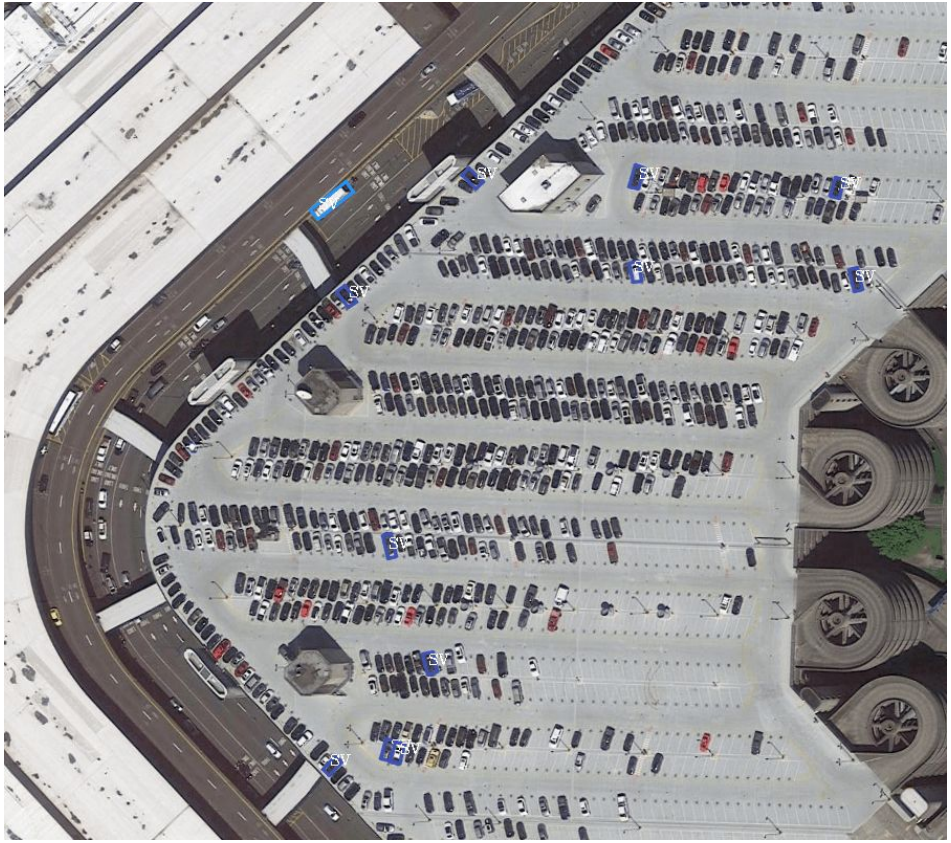
Table 4.2: Evaluation results on DOTA test set per category. The short names for categories are defined as: PL-Plane, BD–Baseball Diamond, BR-Bridge, GTF–Ground Track-Field, SV–Small Vehicle, LV–Large Vehicle, SH-Ship, TC–Tennis Court, BC–Basketball Court, ST–Storage tank, SBF–Soccer-Ball Field, RA–Roundabout, SP–Swimming Pool, HB-Harbor and HC–Helicopter.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HB	SP	HC
RetinaNet	89.00	78.50	31.20	66.51	75.49	60.97	75.81	90.74	81.18	75.56	54.67	42.97	51.96	63.21	50.33
Faster RCNN	88.38	74.93	44.65	57.76	73.72	71.77	77.67	90.67	82.74	82.74	47.36	59.40	63.31	66.17	52.70
S^2ANet [33]	88.63	81.57	44.99	69.73	76.46	79.29	87.08	**90.82**	84.64	84.53	59.01	42.28	66.27	65.25	65.68
DCL RetinaNet* [32]	89.09	**84.18**	47.17	69.31	71.34	58.36	72.92	90.79	86.65	85.67	64.44	**64.78**	64.50	73.08	66.47
R^3DET [30]	88.88	82.74	51.87	65.50	76.13	**81.49**	**87.39**	90.79	85.07	84.48	59.60	61.49	68.08	73.31	64.20
SCRDET++* [29]	**89.82**	84.10	**56.71**	**71.42**	**79.29**	73.35	78.56	90.72	**87.86**	**87.18**	**68.47**	64.55	**76.47**	**82.02**	**75.36**

Object Detection Qualitative Results - Small and Crowded Objects (1/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

DCL RetinaNet

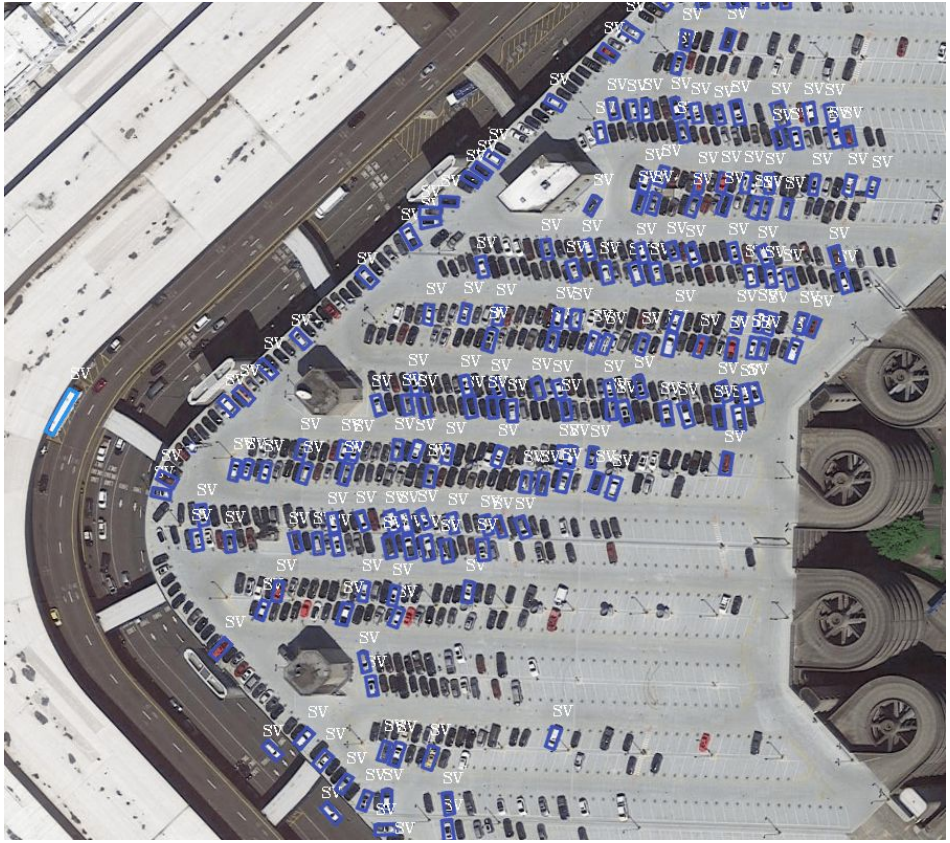


R^3Det

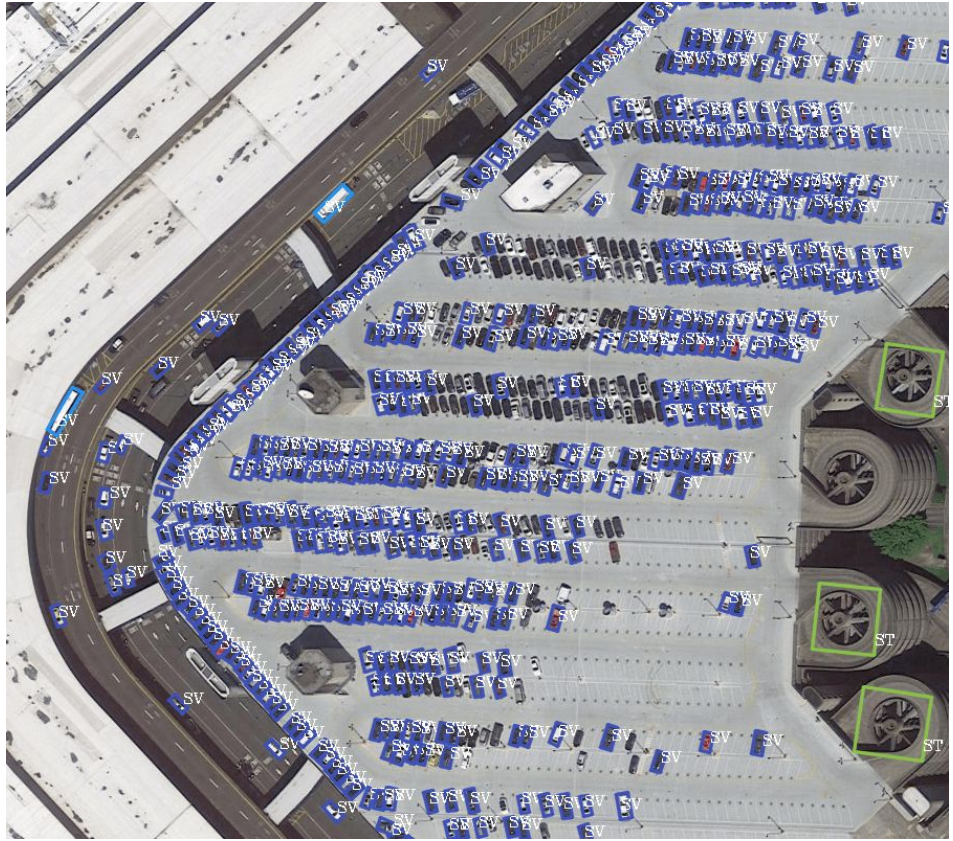
Object Detection Qualitative Results - Small and Crowded Objects (2/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

S^2ANet



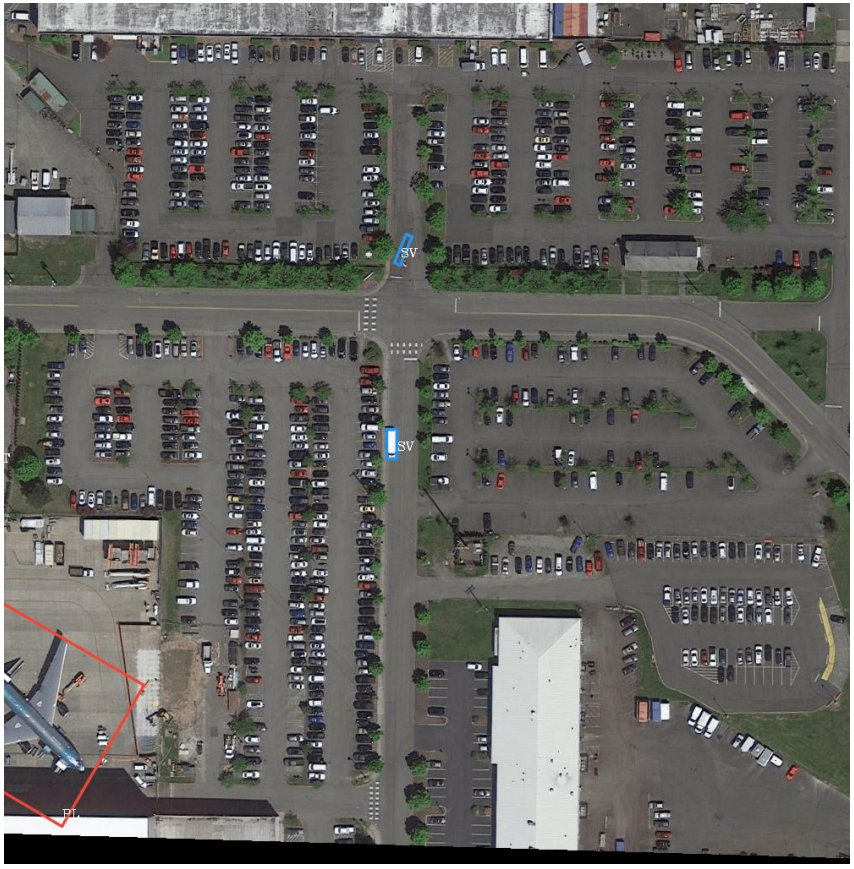
SCRDET++



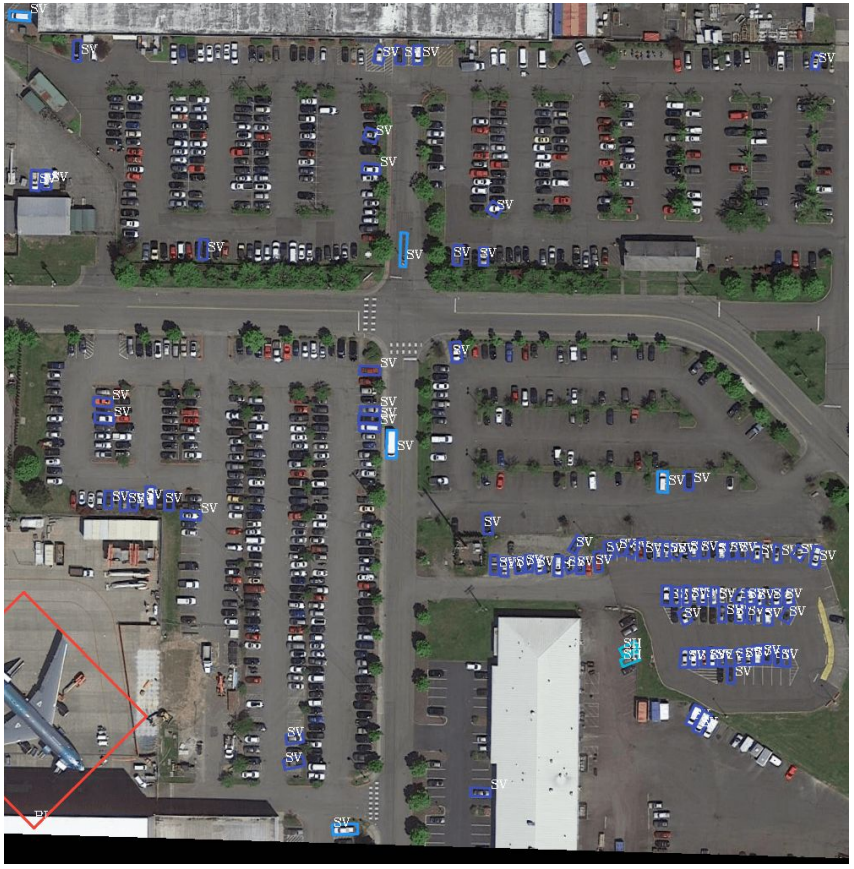
Object Detection Qualitative Results - Small and Crowded Objects (3/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

DCL RetinaNet



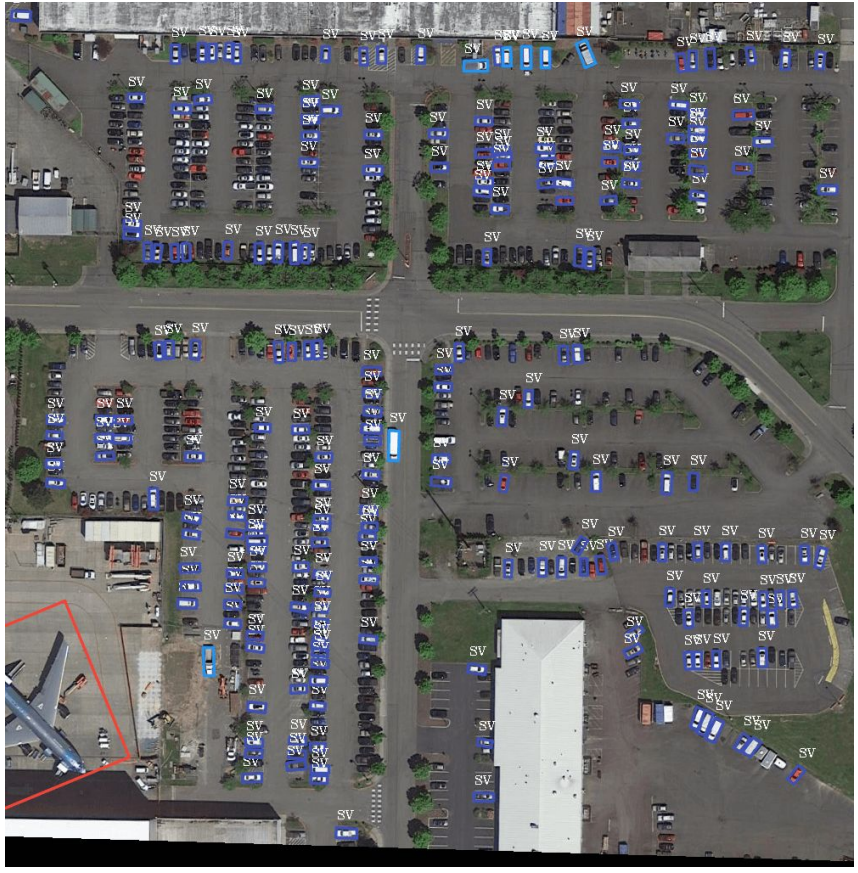
R^3Det



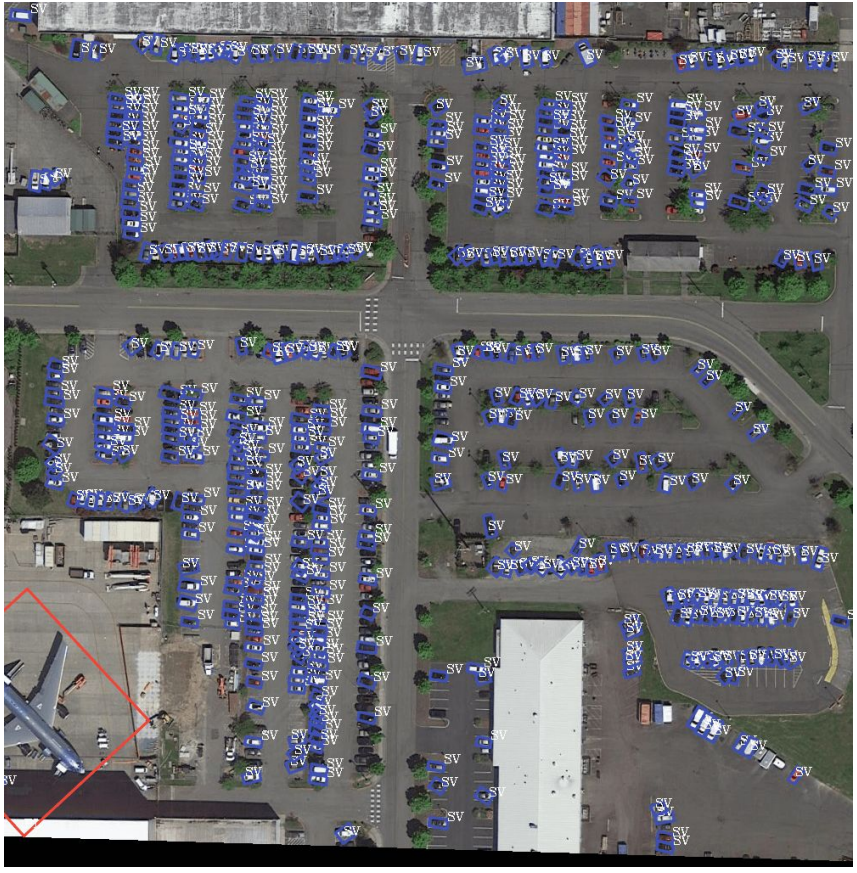
Object Detection Qualitative Results - Small and Crowded Objects (4/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

S^2ANet



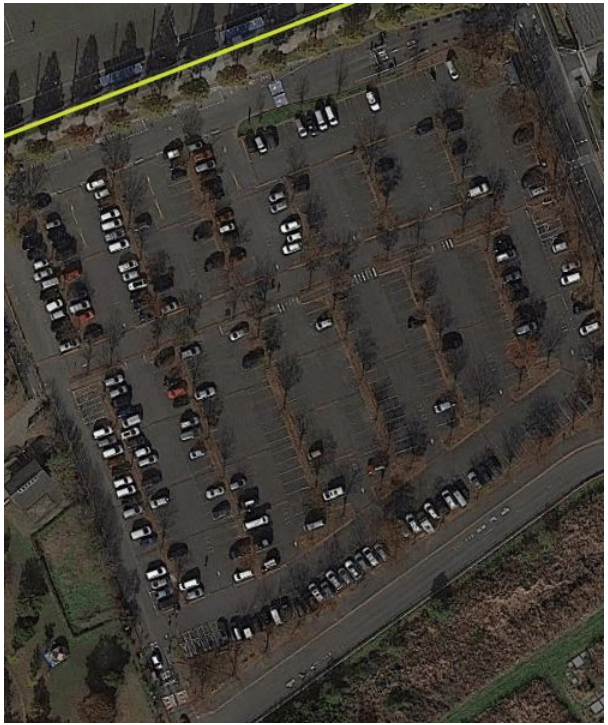
SCRDET++



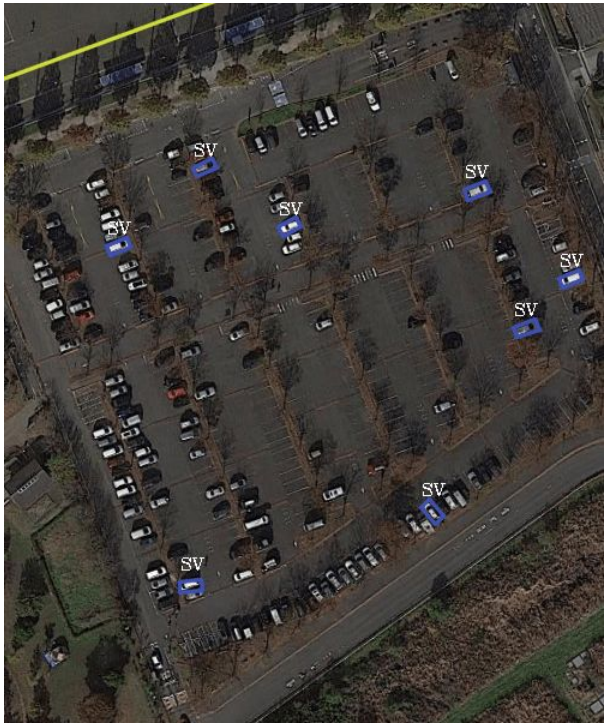
Object Detection Qualitative Results - Small and Crowded Objects (5/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

DCL RetinaNet & R^3Det



S^2ANet



SCRDET++

Object Detection Qualitative Results - Orientation (1/3)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

DCL RetinaNet

R^3Det

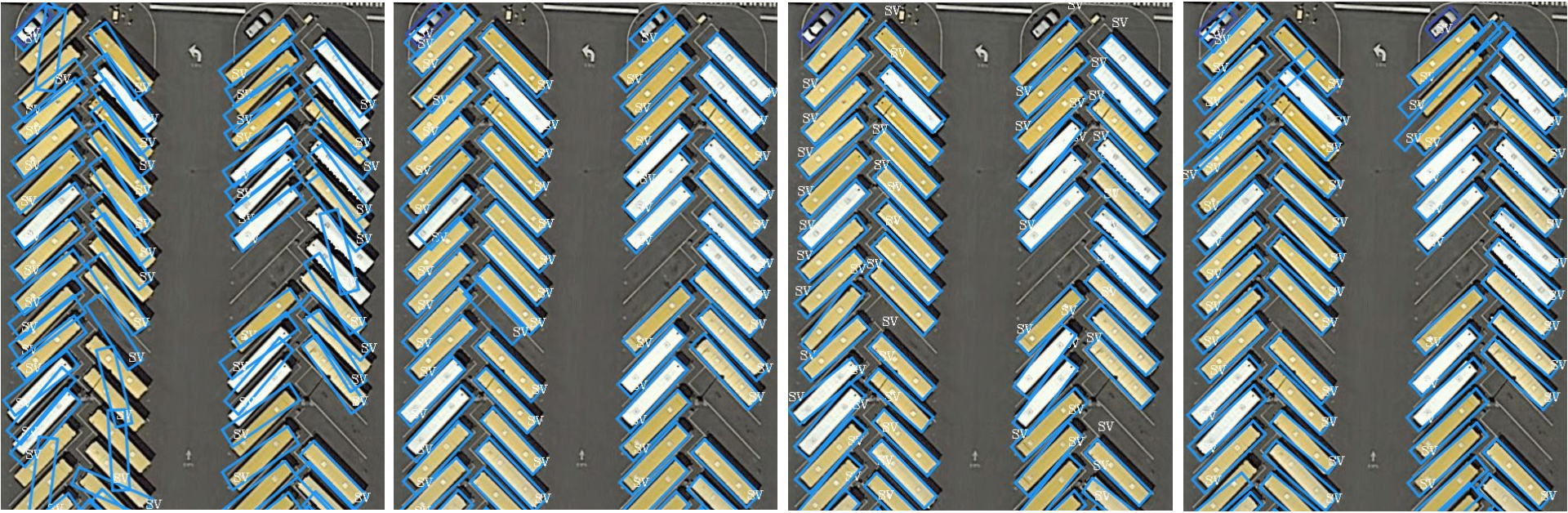
S^2ANet

SCRDET++

Object Detection Qualitative Results - Orientation (2/3)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

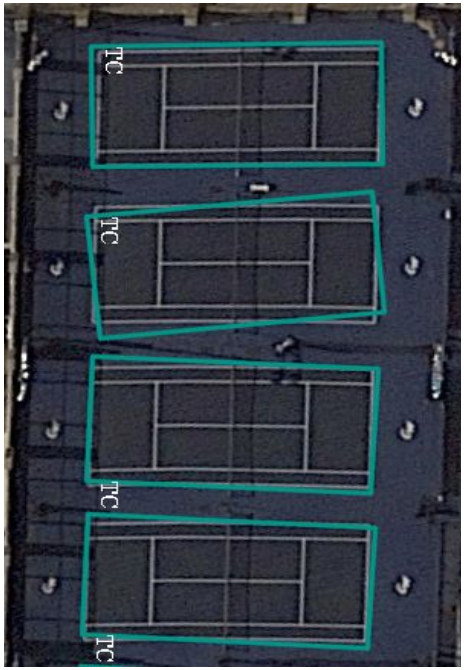
DCL RetinaNet | **R^3Det** | **S^2ANet** | **SCRDET++**



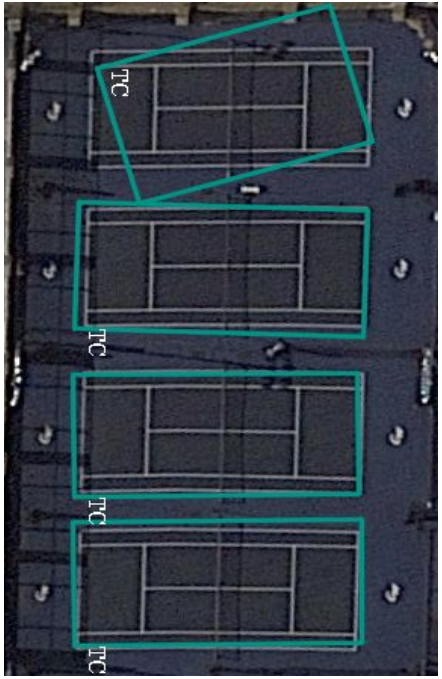
Object Detection Qualitative Results - Orientation (3/3)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

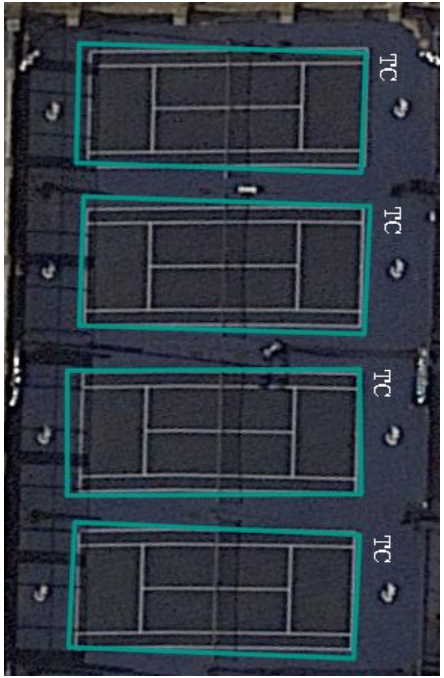
DCL RetinaNet



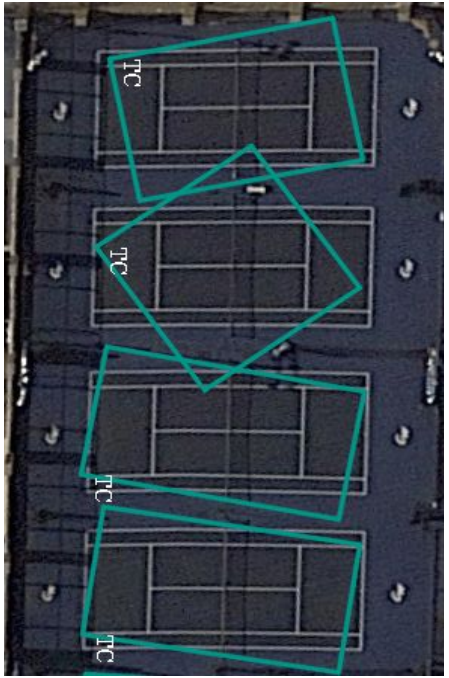
R^3Det



S^2ANet



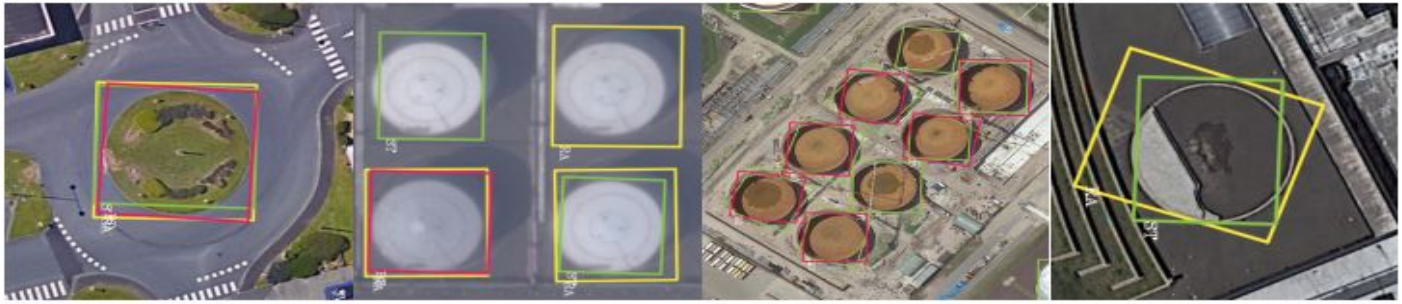
SCRDET++



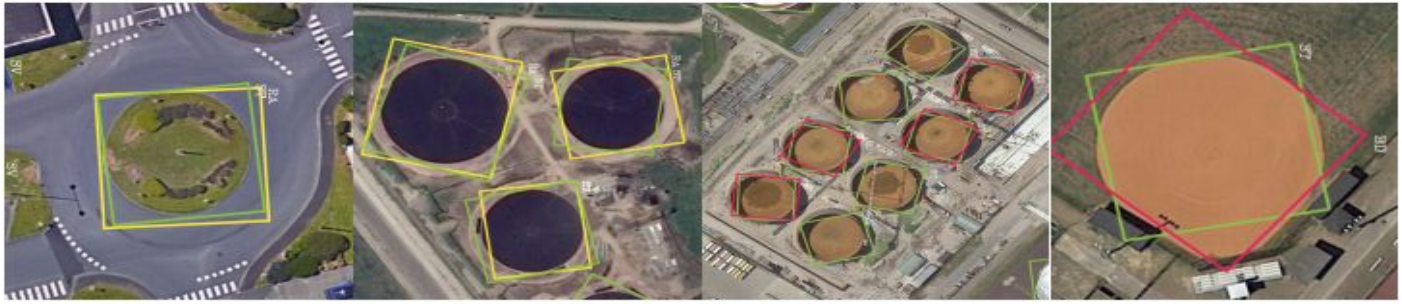
Object Detection Qualitative Results - Circular Objects

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

R^3Det



S^2ANet

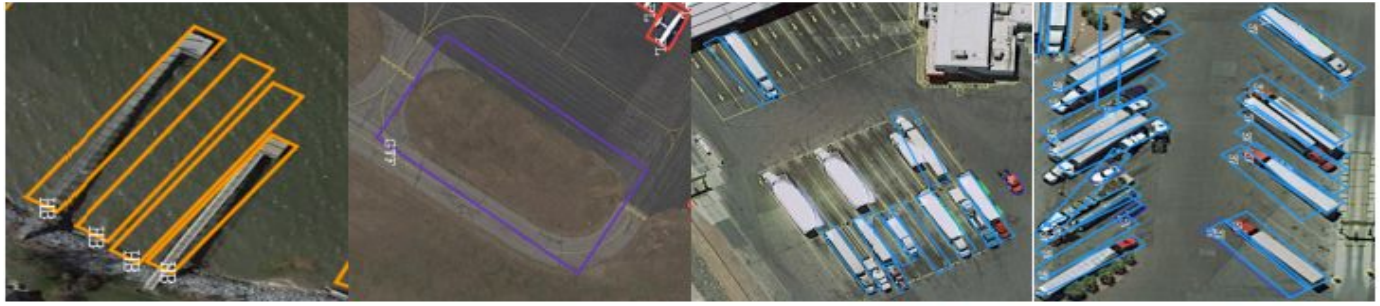


SCRDET++

Object Detection Qualitative Results - Non existing Objects

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

R^3Det



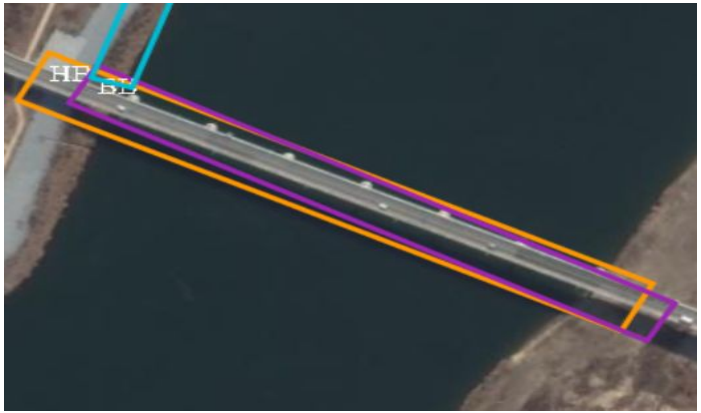
S^2ANet

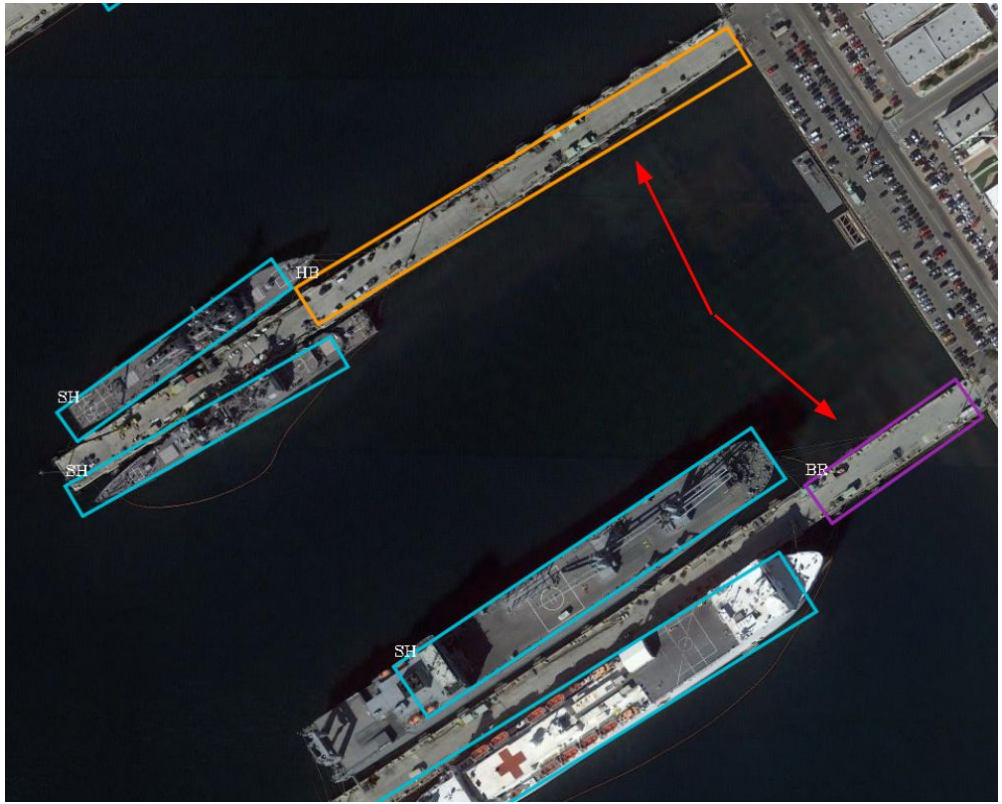
SCRDET++

Object Detection Qualitative Results - Large Aspect Ratio (Confusion)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

All methods

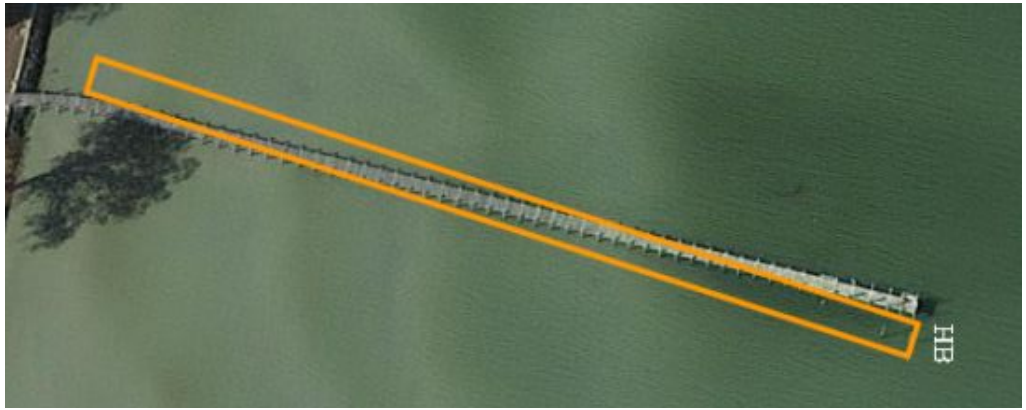


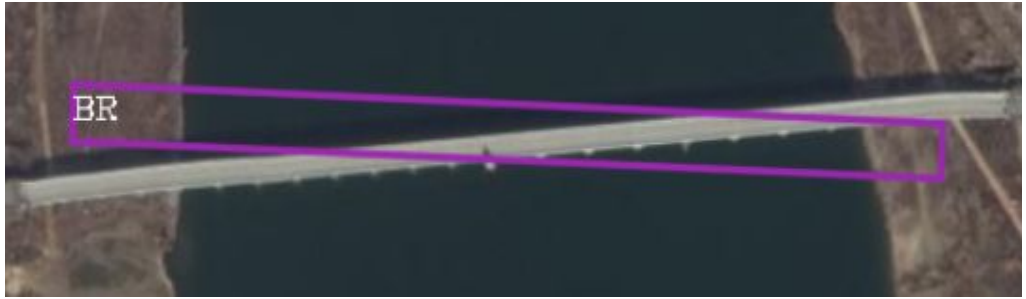


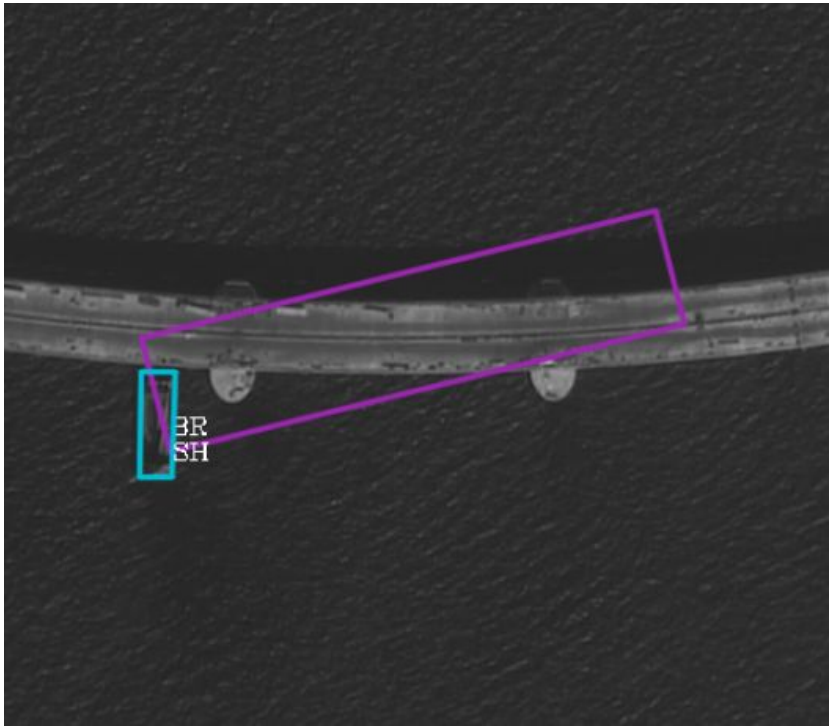
Object Detection Qualitative Results - Large Aspect Ratio (Rotation)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

All methods



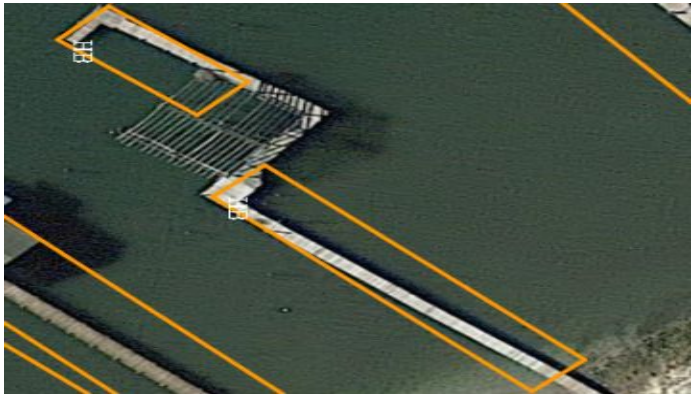


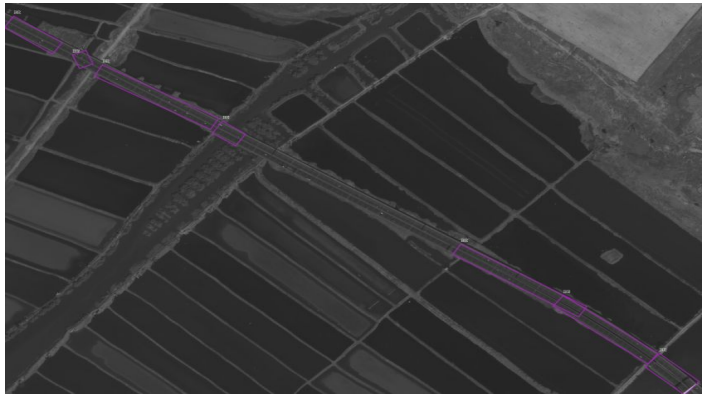


Object Detection Qualitative Results - Large Aspect Ratio (Split)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

All methods





Proposed Method Experiments

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Trained from scratch our proposed method
- Trained from scratch D2Det, a two-stage horizontal instance segmentation algorithm
- All methods were trained on iSAID train and evaluated on val set

Proposed Method Quantitative Results

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

Table 4.3: Evaluation results of proposed method on iSAID Validation set.

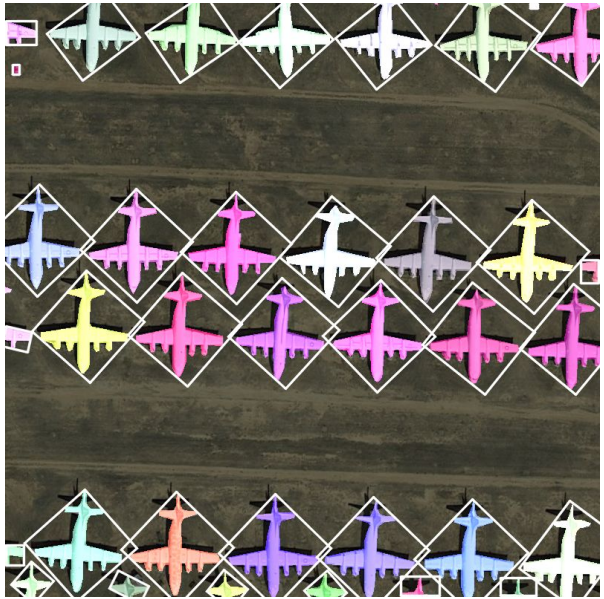
Method	Backbone	Train	Input Size	$mAP_{50\%IoU}$
D2Det	ResNet 101	Ours	800×800	**57.80** %
S^2AU-Net	ResNet 50	Ours	800×800	**51.50** %

Proposed Method Qualitative Results (1/5)

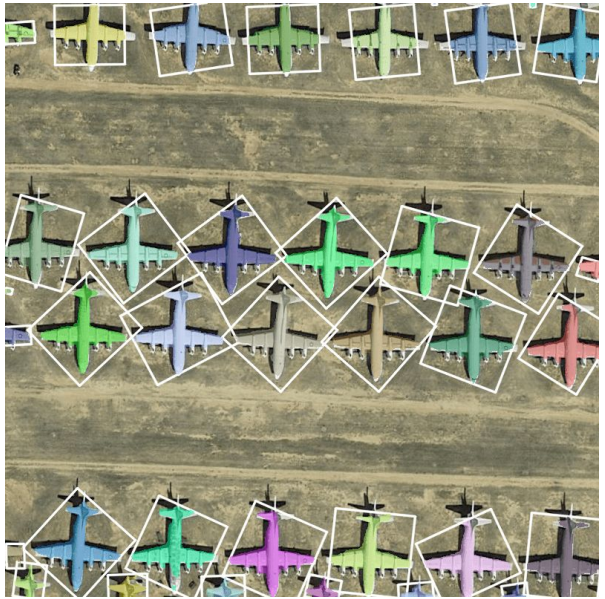
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

(a) Original Image

(b) Ground Truth



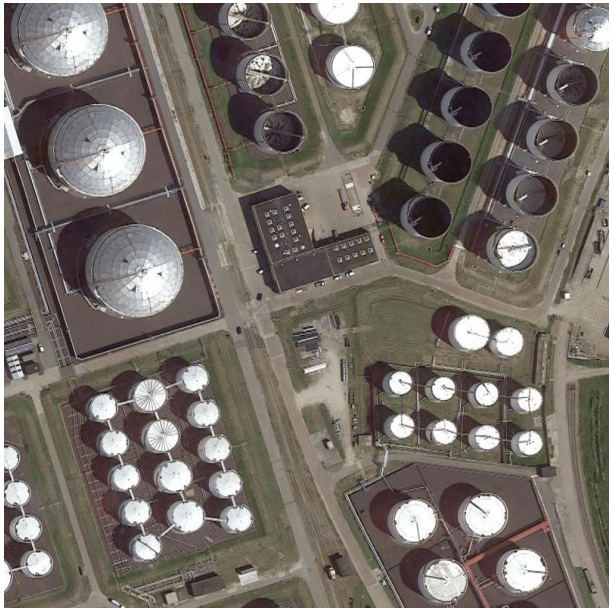
(c) S^{2}AU-Net



Proposed Method Qualitative Results (2/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

(a) Original Image



(b) Ground Truth

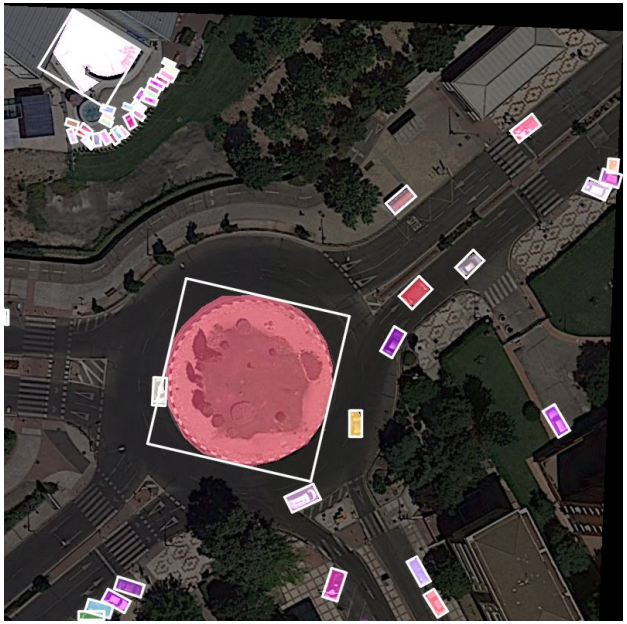
(c) S^2AU-Net

Proposed Method Qualitative Results (3/5)

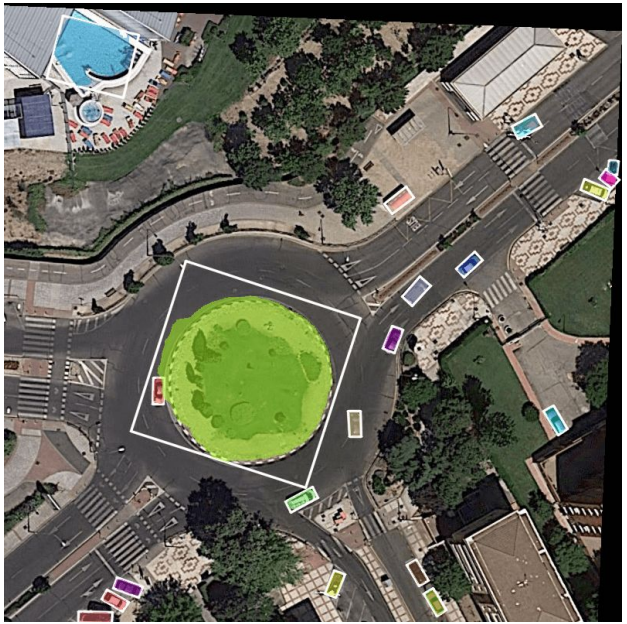
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

(a) Original Image

(b) Ground Truth



(c) S^{2}AU-Net

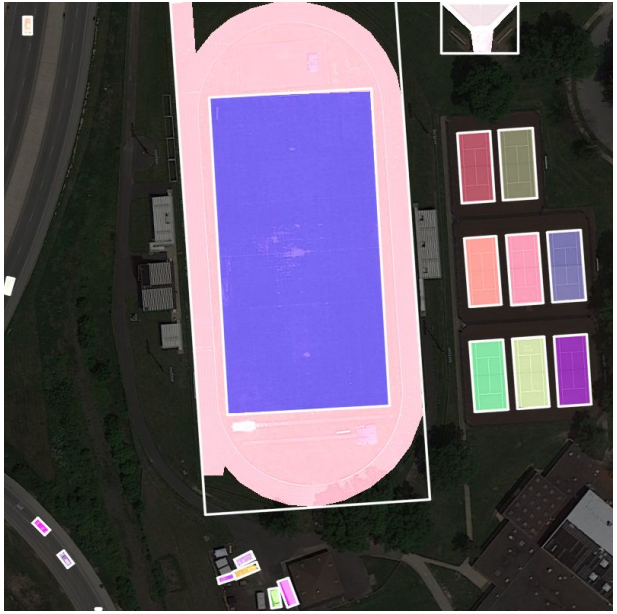


Proposed Method Qualitative Results (4/5)

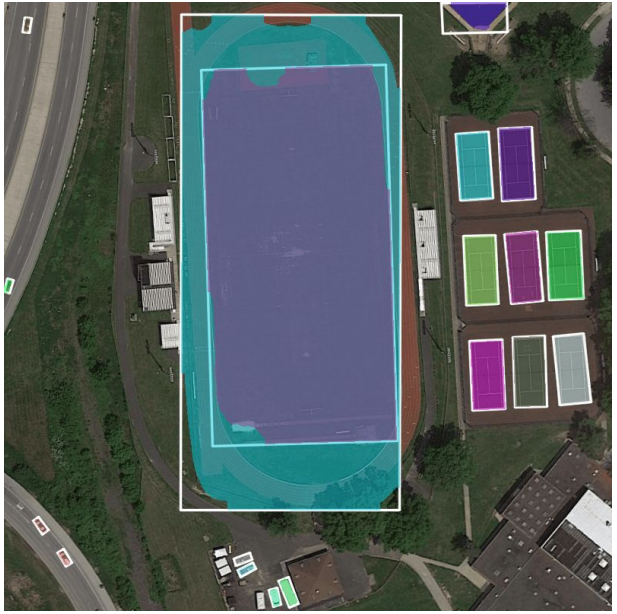
Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

(a) Original Image

(b) Ground Truth



(c) S^2AU-Net



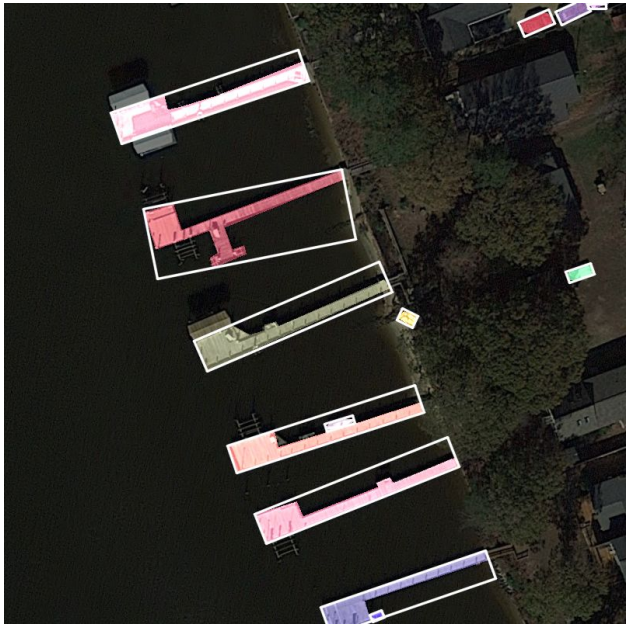
Proposed Method Qualitative Results (5/5)

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

(a) Original Image

(b) Ground Truth

(c) S^2AU-Net



Conclusions & Future Directions

Conclusions - Object Detection

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Most SOTA methods add a computational performance overhead to achieve greater results. **S^2ANet** achieves the most **optimal** trade-off between accuracy and speed

- Most SOTA methods are **resource intensive** and require multiple GPUs to train.

- **SCRDET++**'s Instance Level Denoising is the most effective way to detect **small and crowded** objects

- **R^3Det's and S^2ANet's** feature alignment/refinement modules are the most effective ways to detect arbitrary **oriented** objects

Conclusions - Proposed Method

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- S^2AU-Net offers an effective solution with a concise architecture, producing satisfactory results
- To reach state of the art results, a different approach may be more appropriate (ROI Align and Mask Head similar to Mask R-CNN)
- Horizontal bounding boxes are more preferred than rotated bounding boxes when performing Instance Segmentation.

Future Directions

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Επιστήμη Δεδομένων και Μηχανική Μάθηση

- Use horizontal bounding boxes instead rotated on S^2AU-Net
- Integrate ideas of SCRDET++ (or/and DCL RetinaNet) into S^2AU-Net
- Multi-scale Training
- Multi-GPU Training
- Investigate/Evaluate CenterMap and Gliding Vertex object detectors

Questions

Thank you for your time!