AAAI-21

Dynamic Anchor Learning for Arbitrary-Oriented Object Detection

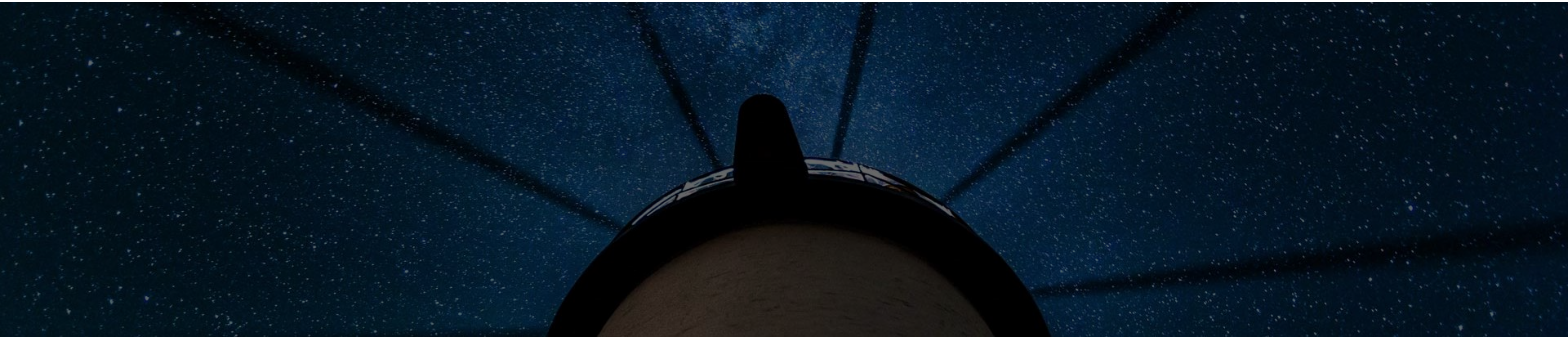
Qi Ming

Beijing Institute of Technology, China

2021.2.9

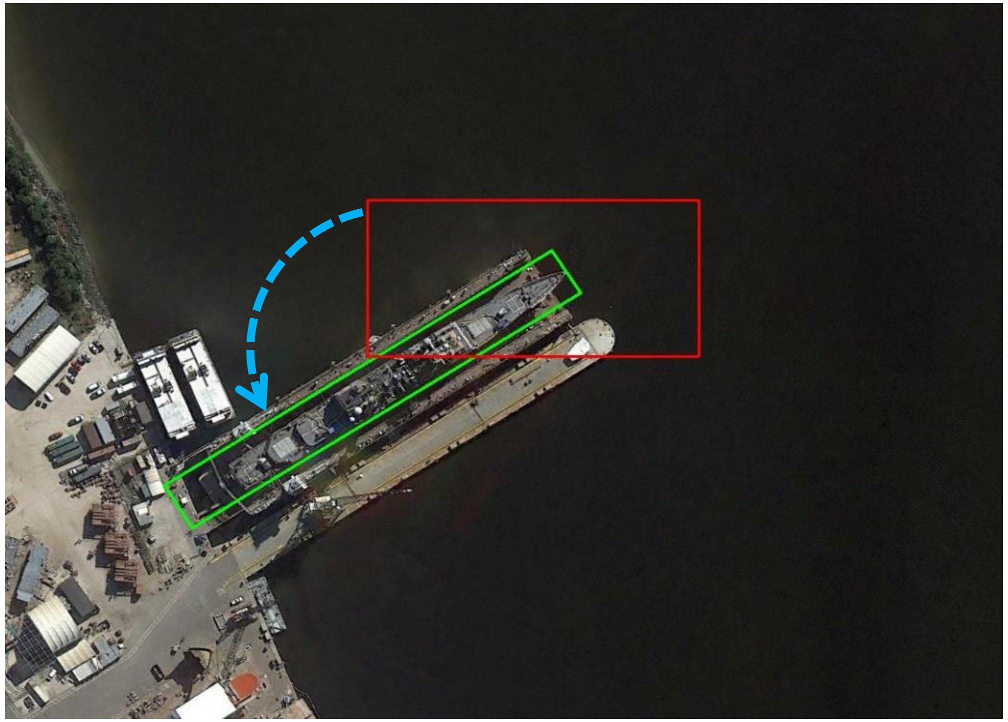
CONTENT

01 Case Study

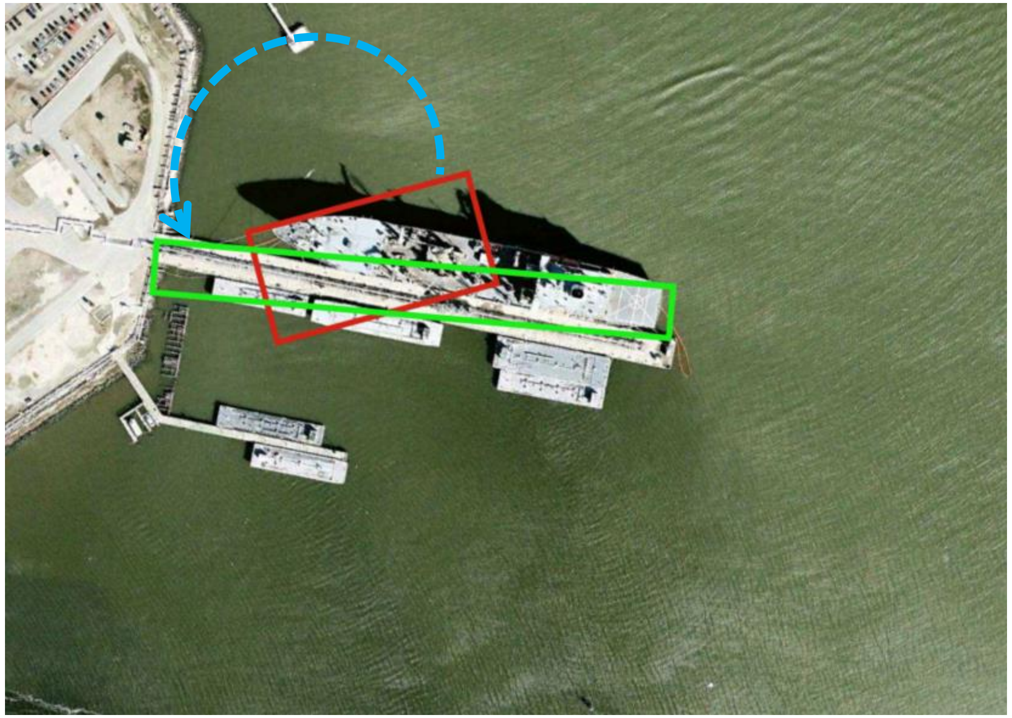


01 Case Study

False negative anchor

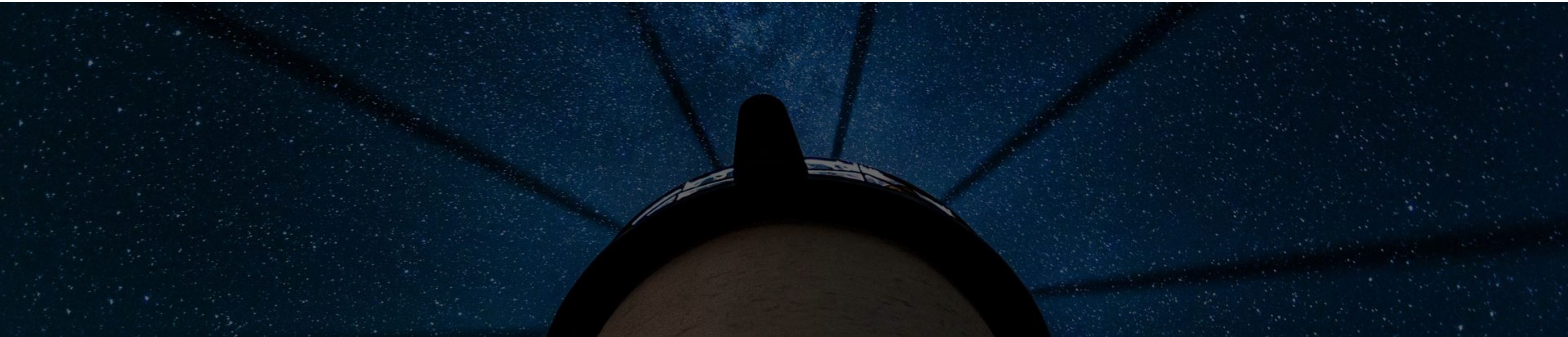


False positive anchor

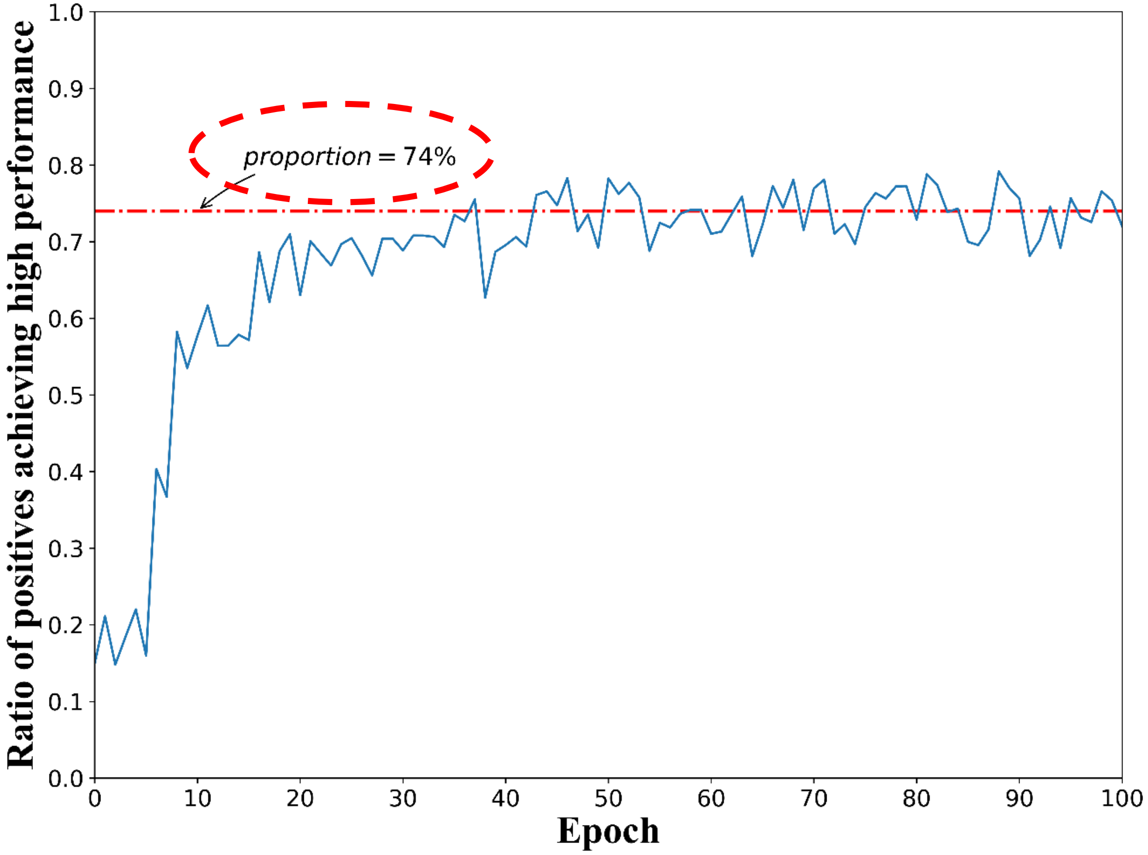


Common problems?

02 Discussion



02 Discussion

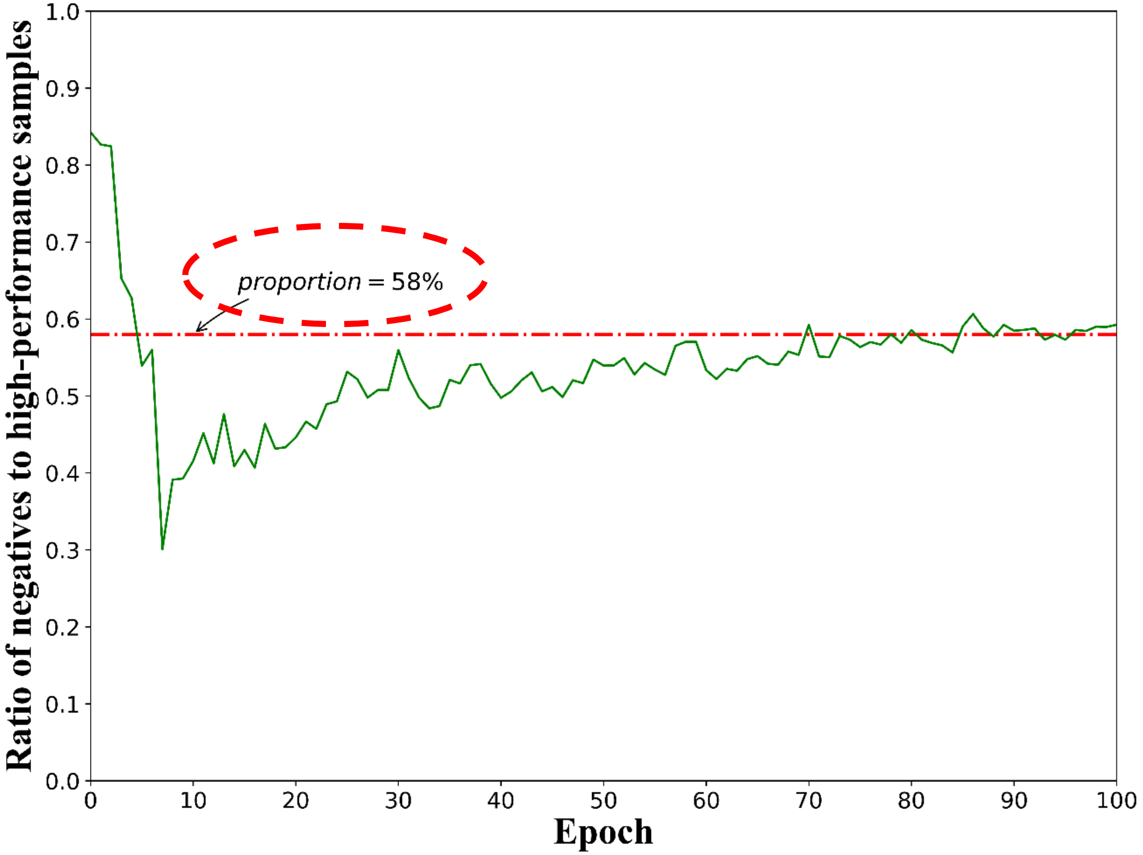


concl.1:

Positive anchors cannot always ensure accurate regression results.

(26% in this case)

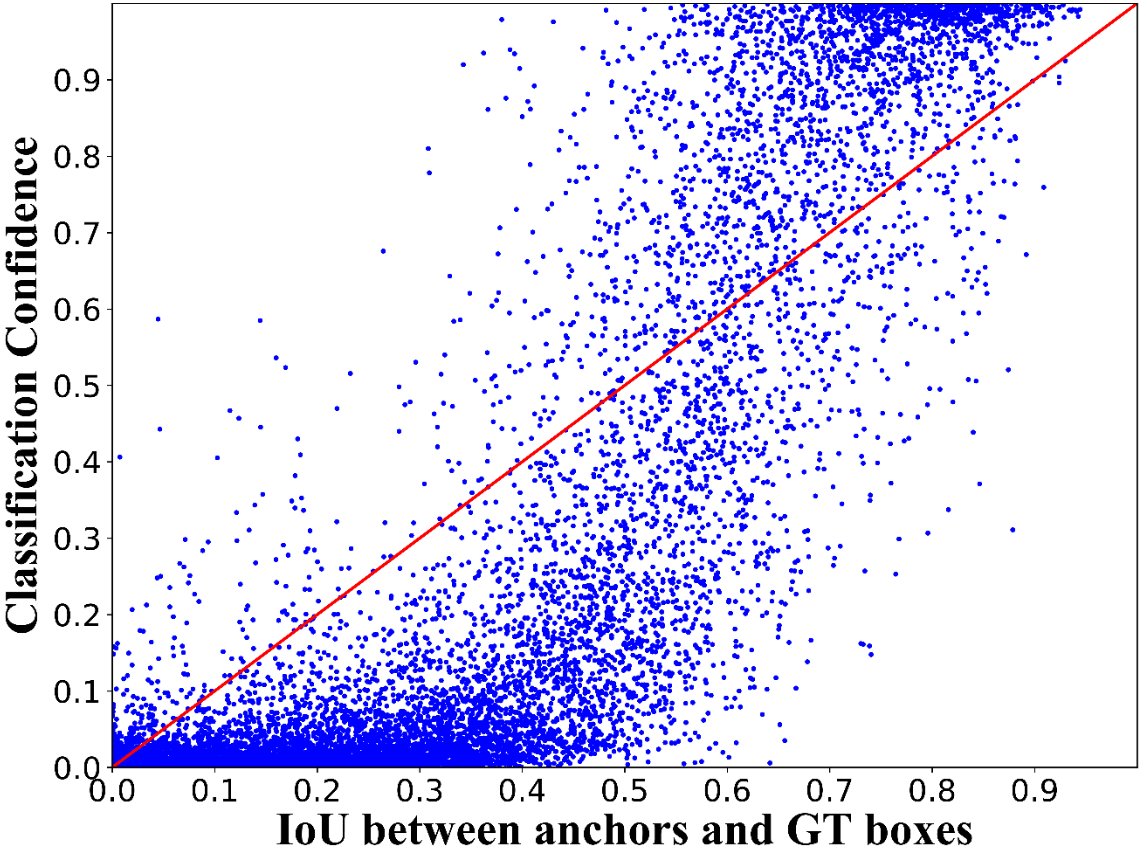
02 Discussion



concl.2:

The initial negative sample may achieve high-quality regression results, but cannot be utilized effectively.

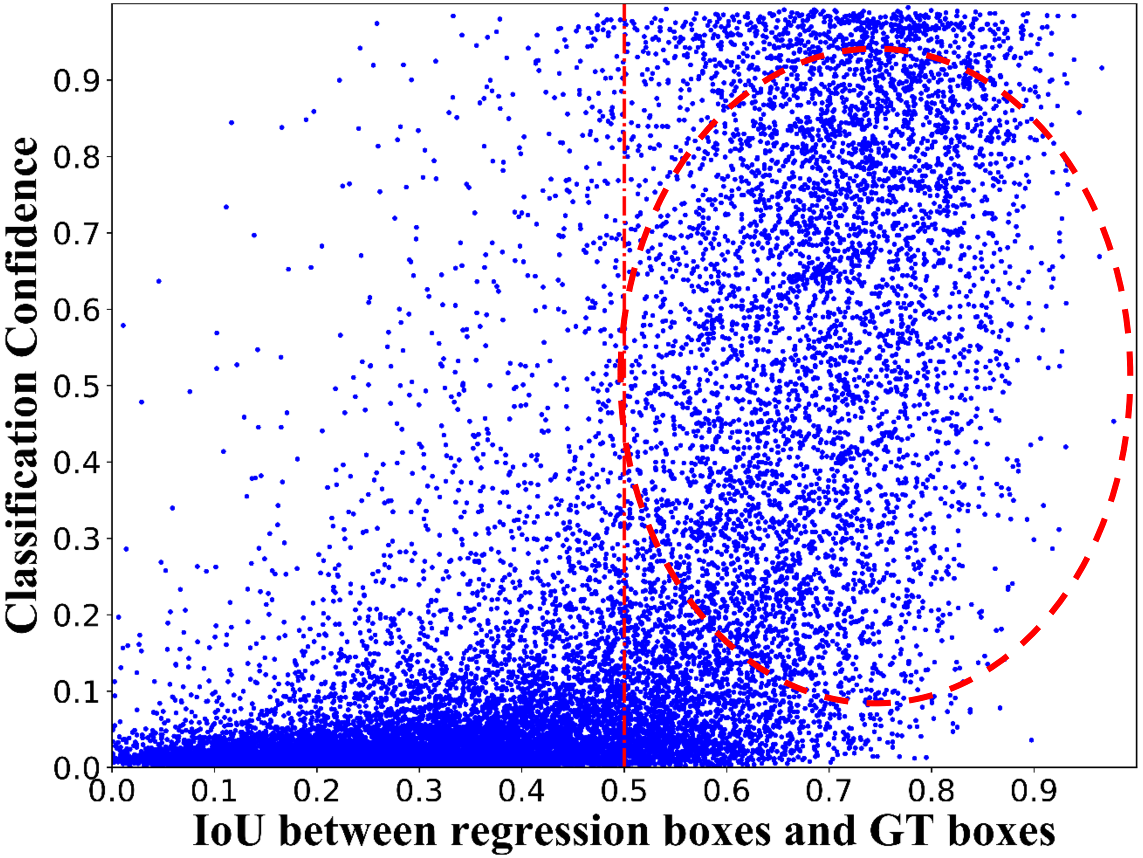
(58% in this case)



concl.3：

The IoU-based label assignment strategy will lead to a positive correlation between the classification score and the initial localization ability of the anchor.

02 Discussion

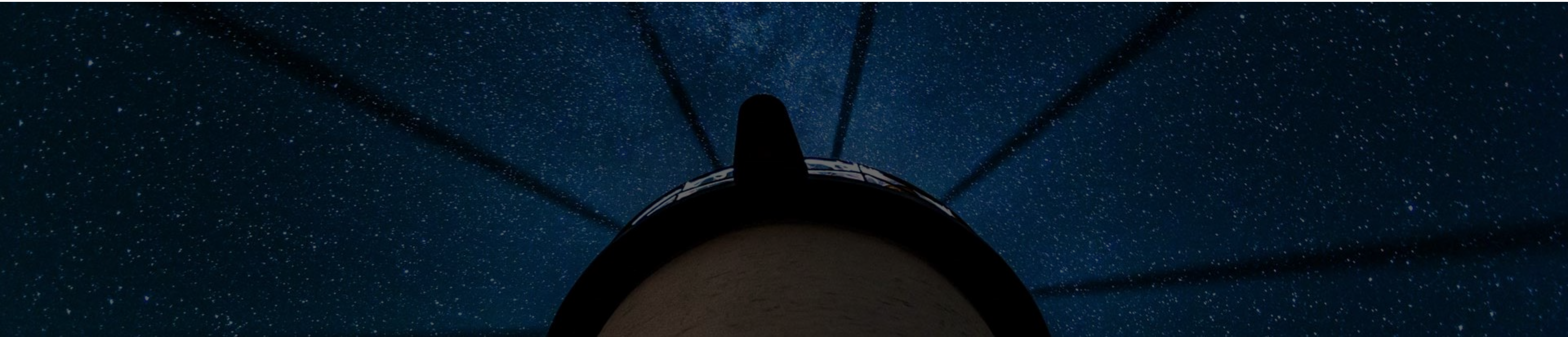


concl.4:

The weak correlation between the loc ability of the prediction and the classification score leads to inconsistencies between classification and regression.

Unreasonable label assignment

03 Methods



“

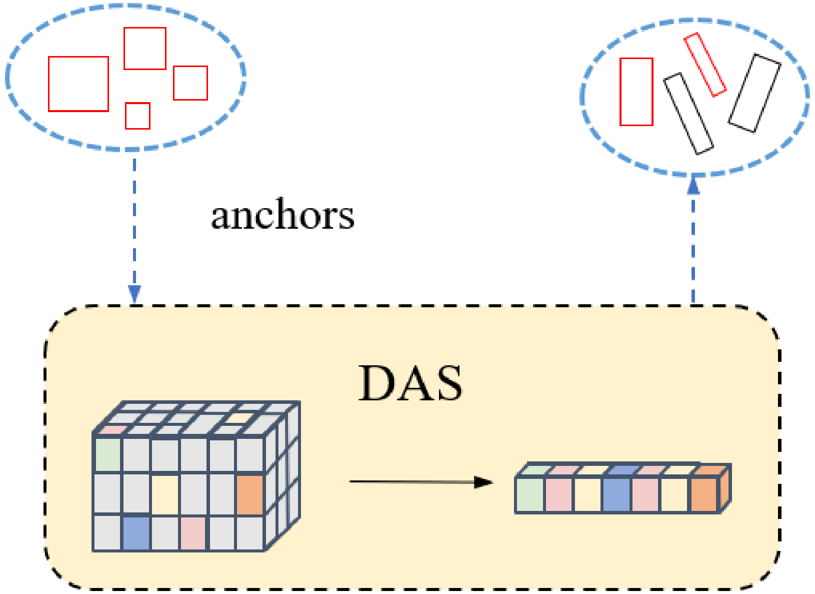
Why not directly use the output IoU as the standard for label assignemnt?

”

Does not converge!

- Output IoU judgment is unreliable at the beginning of training.
- interference in special circumstances.

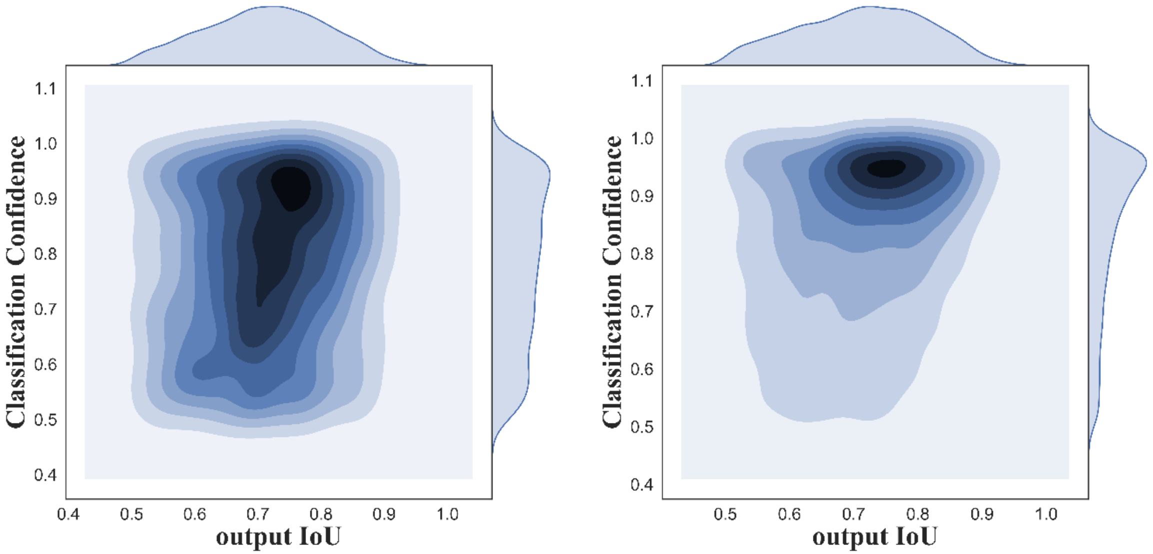
Dynamic Anchor Selection (DAS)



$$md = \alpha \cdot sa + (1 - \alpha) \cdot fa - u^{\gamma}$$

- sa denotes spatial alignment ability
- fa denotes feature alignment ability
- u denotes regression uncertainty

Matching-Sensitive Loss (MSL)



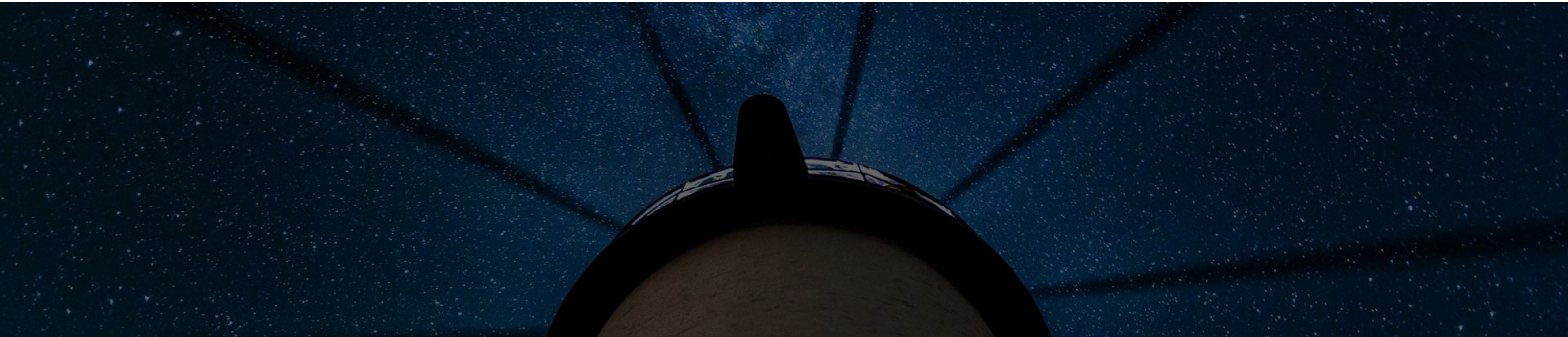
$$\Delta md = 1 - md_{\max}$$

$$\boldsymbol{w} = \boldsymbol{md}_{pos} + \Delta md$$

$$L_{reg} = \frac{1}{N_p} \sum_{j \in \psi_p} w_j \cdot L_{smooth_{L_1}} \left(\boldsymbol{t_j}, \boldsymbol{t_j^*} \right)$$

$$L_{cls} = \frac{1}{N} \sum_{i \in \psi} FL\left(p_i, p_i^*\right) + \frac{1}{N_p} \sum_{j \in \psi_p} w_j \cdot FL\left(p_j, p_j^*\right)$$

04 Experiments



Experiments

1 Ablation study

- Component-wise ablation
- Hyper-parameters
- With other label assignment methods

2 OBB datasets

- Remote Sensing: DOTA, HRSC2016, UCAS-AOD
- Scene Text: ICDAR2015

3 HBB datasets

- Single class: ICDAR2013,
- Multi classes: VOC2007, NWPU VHR -10

Component-wise

	Different Variants			
with Input IoU	✓	✓	✓	✓
with Output IoU		✓	✓	✓
Uncertainty Supression			✓	✓
Matching Sensitive Loss				✓
AP_{50}	80.8	78.9	85.9	**88.6**
AP_{75}	52.4	50.4	57.7	**67.6**

Table 1: Effects of each component in our method on HRSC2016 dataset.

Analysis

- Penalty of IoU variation is the key to effective use of feature alignment information.
- Greatly improved high-precision detection performance (described by AP75)

Component-wise

γ	α	mAP	γ	α	mAP	γ	α	mAP
3	0.2	84.1	4	0.2	88.1	5	0.2	87.3
	0.3	88.3		0.3	88.2		0.3	**88.6**
	0.5	86.2		0.5	85.5		0.5	88.4
	0.7	84.1		0.7	77.9		0.7	88.1
	0.9	70.1		0.9	75.5		0.9	83.5

Table 2: Analysis of different hyperparameters on HRSC2016 dataset.

Analysis

- alpha increases as gamma increases
- alpha = 0.3, gamma = 5

Compared with other methods

Baseline	(Yang et al. 2018)	HAMBox(Liu et al. 2020)	ATSS(Zhang et al. 2020b)	DAL(Ours)
80.8	82.2	85.4	86.1	**88.6**

Table 3: Comparisons with other label assignment strategies on HRSC2016.

Analysis

- **S3FD** (ICCV 2017) : compensate relative low-quality anchors
- **HAMBox** (CVPR 2020): directly compensate negative anchors
- **ATSS**(CVPR 2020): Dynamic selection of samples with appropriate IoU threshold

Results on DOTA

Methods	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
FR-O(Xia et al. 2018)	R-101	79.09	69.12	17.17	63.49	34.20	37.16	36.20	89.19	69.60	58.96	49.40	52.52	46.69	44.80	46.30	52.93
R-DFPN(Yang et al. 2018)	R-101	80.92	65.82	33.77	58.94	55.77	50.94	54.78	90.33	66.34	68.66	48.73	51.76	55.10	51.32	35.88	57.94
R^2CNN(Jiang et al. 2017)	R-101	80.94	65.67	35.34	67.44	59.92	50.91	55.81	90.67	66.92	72.39	55.06	52.23	55.14	53.35	48.22	60.67
RRPN(Ma et al. 2018)	R-101	88.52	71.20	31.66	59.30	51.85	56.19	57.25	90.81	72.84	67.38	56.69	52.84	53.08	51.94	53.58	61.01
ICN(Azimi et al. 2018)	R-101	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
RoI Trans.(Ding et al. 2019)	R-101	88.64	78.52	43.44	**75.92**	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
CAD-Net(Zhang, Lu, and Zhang 2019)	R-101	87.80	82.40	49.40	73.50	71.10	63.50	76.70	**90.90**	79.20	73.30	48.40	60.90	62.00	67.00	62.20	69.90
DRN(Pan et al. 2020)	H-104	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
O^2-DNet(Wei et al. 2019)	H-104	89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	61.03	71.04
SCRDet(Yang et al. 2019b)	R-101	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	86.86	65.02	66.68	66.25	68.24	65.21	72.61
R^3Det(Yang et al. 2019a)	R-152	89.49	81.17	50.53	66.10	70.92	78.66	78.21	90.81	85.26	84.23	61.81	63.77	68.16	69.83	67.17	73.74
CSL(Yang and Yan 2020)	R-152	**90.25**	**85.53**	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	**69.60**	**68.04**	73.83	71.10	**68.93**	76.17
Baseline	R-50	88.67	77.62	41.81	58.17	74.58	71.64	79.11	90.29	82.13	74.32	54.75	60.60	62.57	69.67	60.64	68.43
Baseline+DAL	R-50	88.68	76.55	45.08	66.80	67.00	76.76	79.74	90.84	79.54	78.45	57.71	62.27	69.05	**73.14**	60.11	71.44
Baseline+DAL	R-101	88.61	79.69	46.27	70.37	65.89	76.10	78.53	90.84	79.98	78.41	58.71	62.02	69.23	71.32	60.65	71.78
S^2A-Net(Han et al. 2020)	R-50	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
S^2A-Net+DAL	R-50	89.69	83.11	**55.03**	71.00	**78.30**	**81.90**	**88.46**	90.89	84.97	**87.46**	64.41	65.65	**76.86**	72.09	64.35	**76.95**

Table 4: Performance evaluation of OBB task on DOTA dataset. R-101 denotes ResNet-101(likewise for R-50), and H-104 stands for Hourglass-104.

Results on HRSC2016

- less anchors
- faster inference
- high-perfoemence

Methods	Backbone	Size	NA	mAP
Two-stage:				
R^2CNN(Jiang et al. 2017)	ResNet101	800×800	21	73.07
RC1&RC2(LB et al. 2017)	VGG16	-	-	75.70
RRPN(Ma et al. 2018)	ResNet101	800×800	54	79.08
R^2PN(Zhang et al. 2018)	VGG16	-	24	79.60
RoI Trans. (Ding et al. 2019)	ResNet101	512×800	5	86.20
Gliding Vertex(Xu et al. 2020)	ResNet101	512×800	5	88.20
Single-stage:				
RRD(Liao et al. 2018)	VGG16	384×384	13	84.30
R^3Det(Yang et al. 2019a)	ResNet101	800×800	21	89.26
R-RetinaNet(Lin et al. 2017b)	ResNet101	800×800	121	89.18
Baseline	ResNet50	416×416	3	80.81
Baseline+DAL	ResNet50	416×416	3	88.60
Baseline+DAL	ResNet101	416×416	3	88.95
Baseline+DAL	ResNet101	800×800	3	**89.77**

Table 5: Comparisons with state-of-the-art detectors on HRSC2016. NA denotes the number of preset anchor at each location of feature map.

Results on UCAS-AOD

High-quality detection can be achieved with DAL method.

eg. An improvement of **35.15% AP75** here.

Methods	car	airplane	AP_{50}	AP_{75}
FR-O(Xia et al. 2018)	86.87	89.86	88.36	47.08
RoI Transformer (Ding et al. 2019)	87.99	89.90	88.95	50.54
Baseline	84.64	**90.51**	87.57	39.15
Baseline+DAL	**89.25**	90.49	**89.87**	**74.30**

Table 6: Detection results on UCAS-AOD dataset.

Results on ICDAR2015

- Long text is often segmented into multiple detections
- It`s hard for a small number of preset anchors to adapt to excessively extreme aspect ratios of texts.

Methods	P	R	F
CTPN(Tian et al. 2016)	74.2	51.6	60.9
Seglink(Shi, Bai, and Belongie 2017)	73.1	76.8	75.0
RRPN(Ma et al. 2018)	82.2	73.2	77.4
SCRDet(Yang et al. 2019b)	81.3	78.9	80.1
RRD(Liao et al. 2018)	85.6	79.0	82.2
DB(Liao et al. 2020)	**91.8**	**83.2**	**87.3**
Baseline	77.2	77.8	77.5
Baseline+DAL	83.7	79.5	81.5
Baseline+DAL*	84.4	80.5	82.4

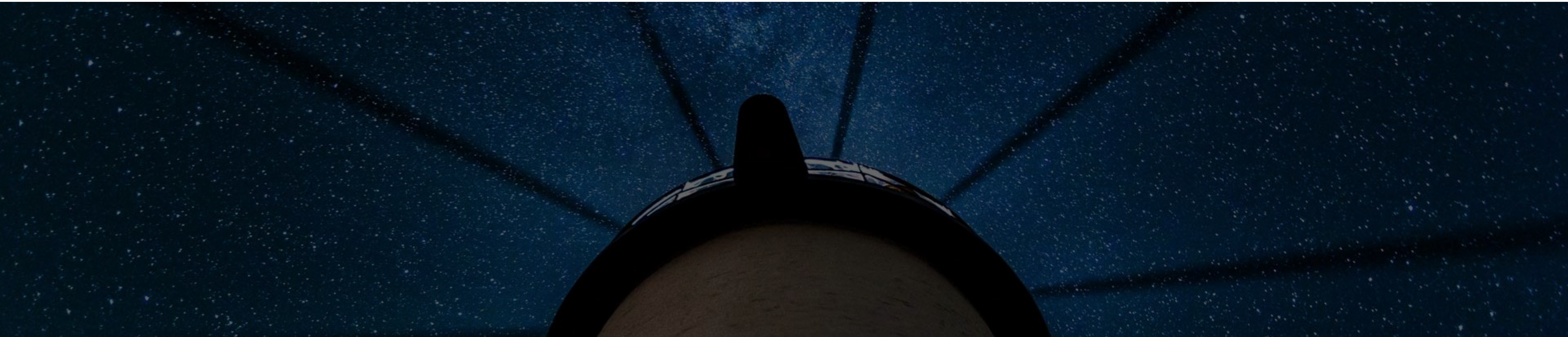
Table 7: Comparisons of different methods on the ICDAR 2015. P, R, F indicate recall, precision and F-measure respectively. * means multi-scale training and testing.

Results on HBB

Dataset	Backbone	mAP/F
ICDAR 2013	RetinaNet	77.2
	RetinaNet+DAL	**81.3**
NWPU VHR-10	RetinaNet	86.4
	RetinaNet+DAL	**88.3**
VOC 2007	RetinaNet	74.9
	RetinaNet+DAL	**76.1**

Table 8: Performance evaluation of HBB task on ICDAR 2013 and NWPU VHR-10.

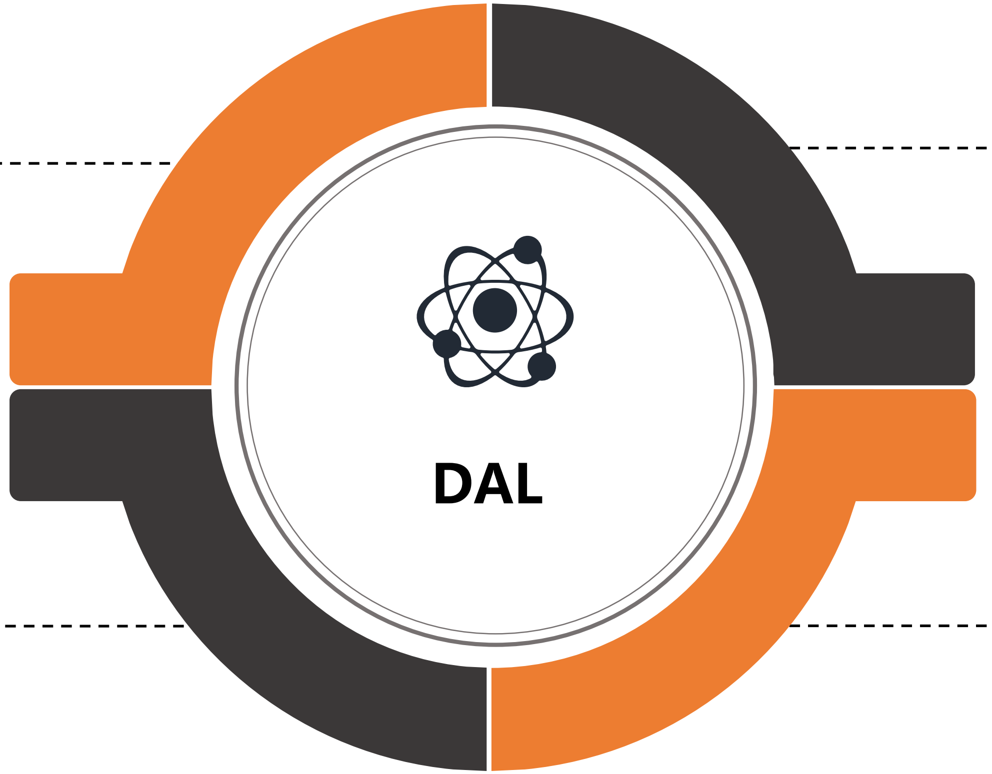
05 Conclusion



05 Conclusion

Inconsistency localization ability before and after the regression

Inconsistent classification and regression performance caused by label assignment

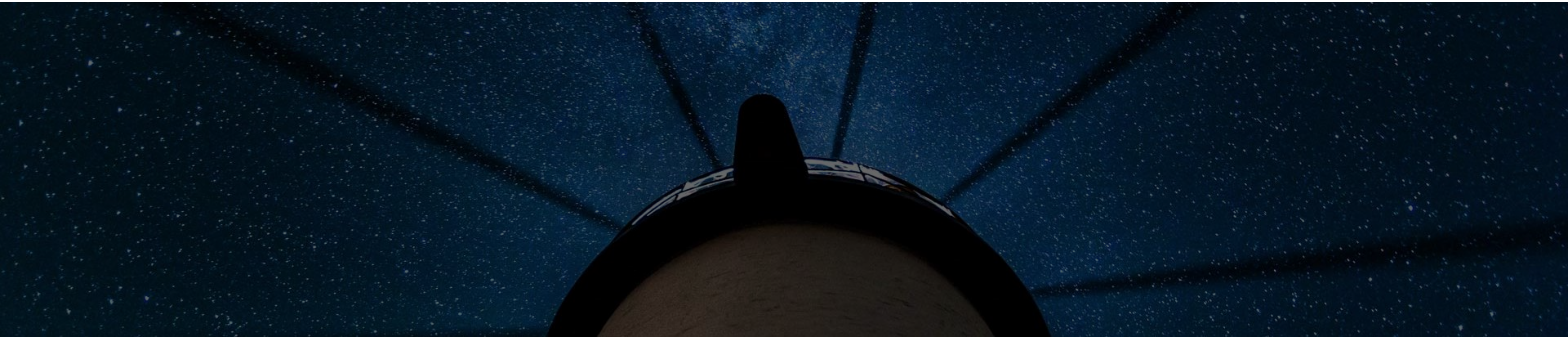


Matching degree is proposed to evaluate potential localization ability

Matching Sensitive Loss is used to improve high-precision detection performance

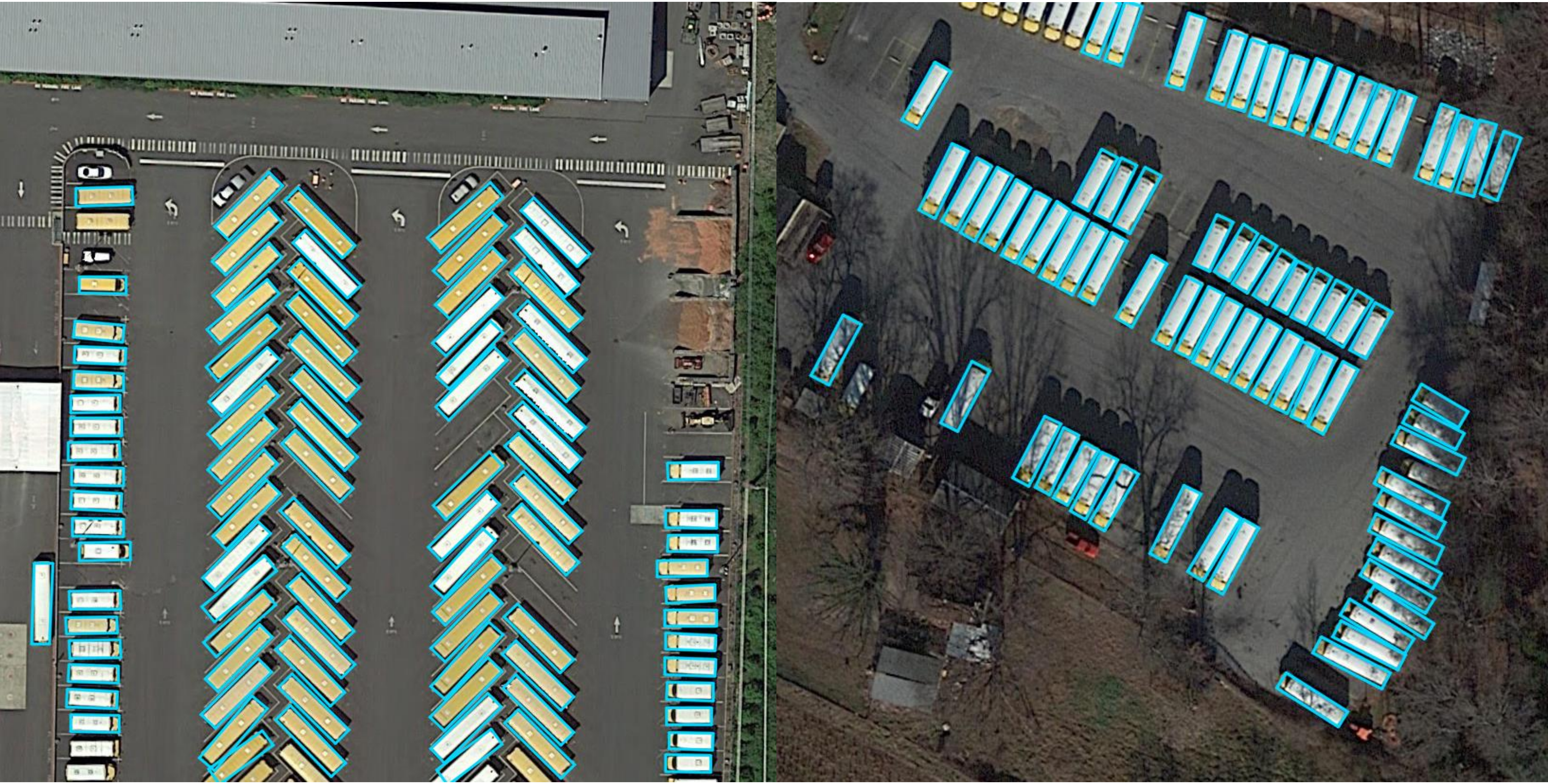
06

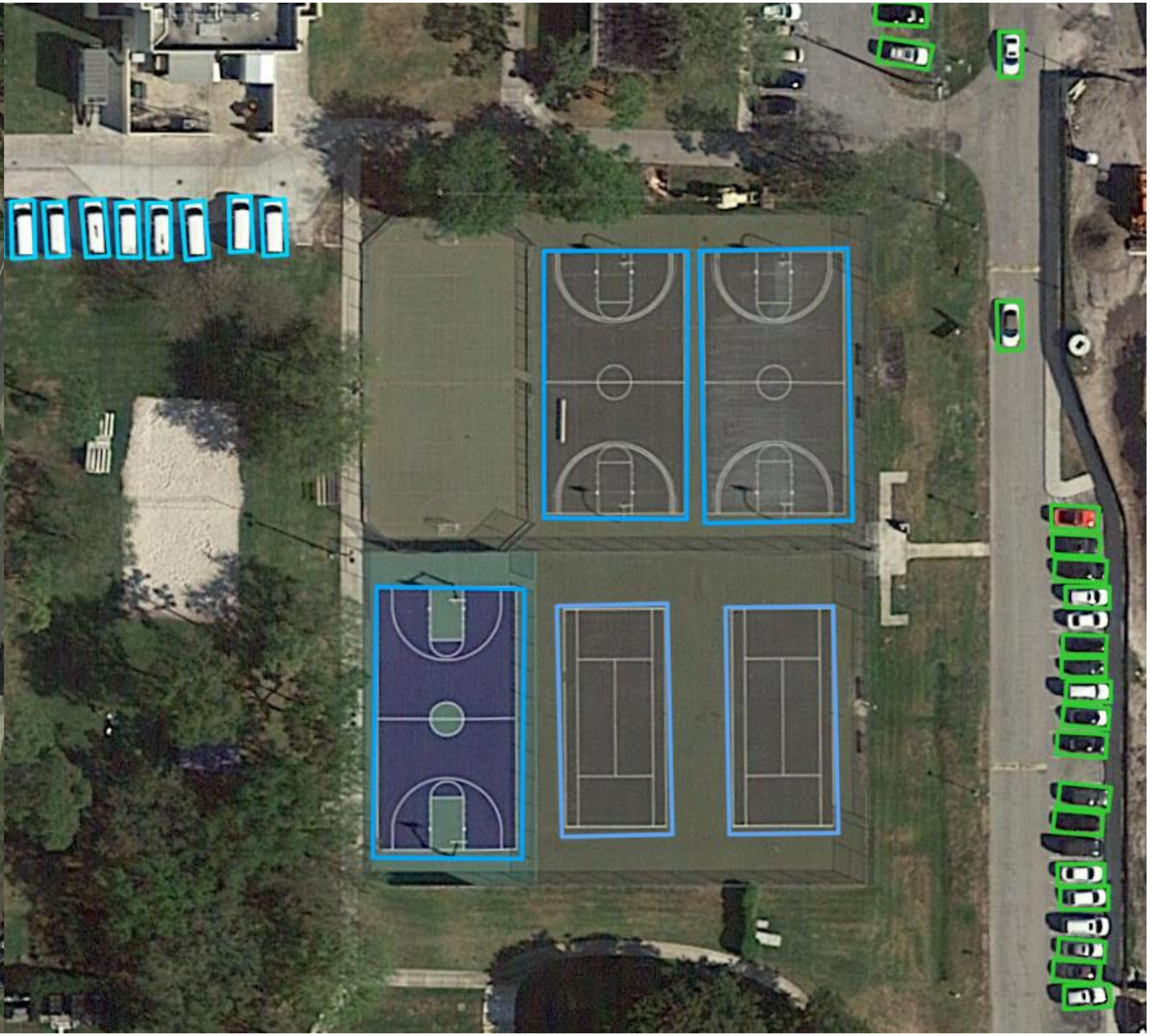
Detections

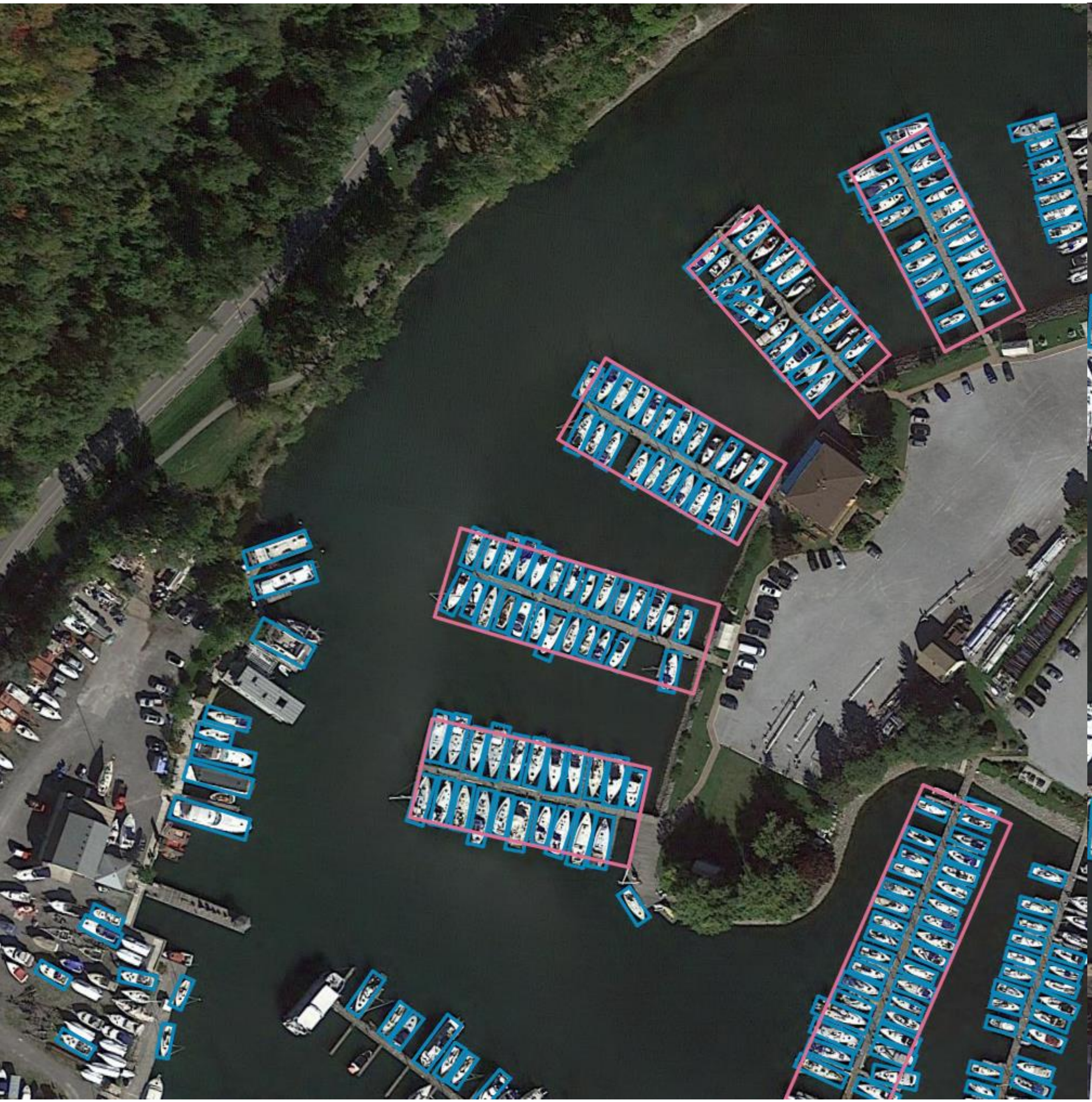


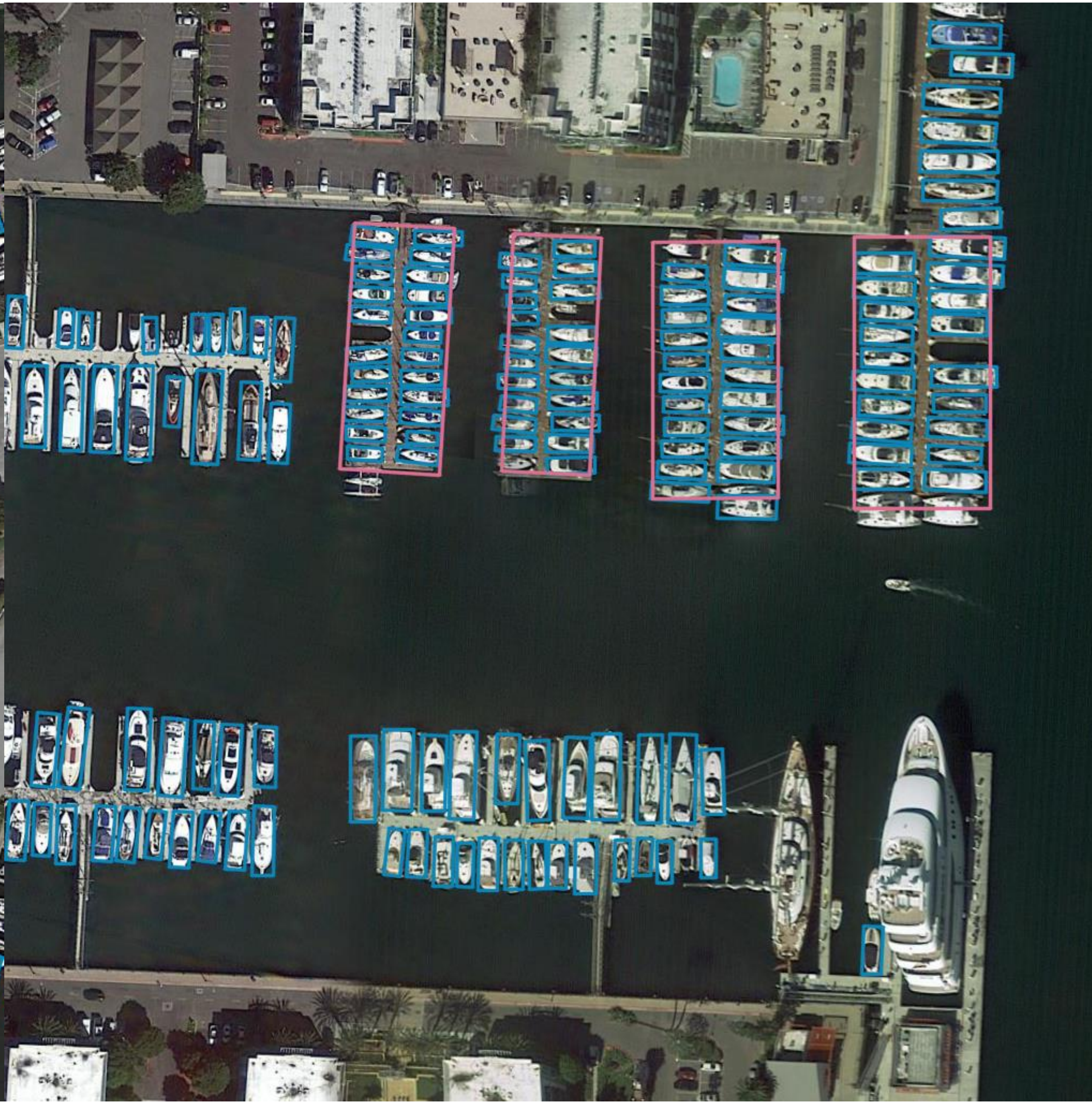
Detections with OBB

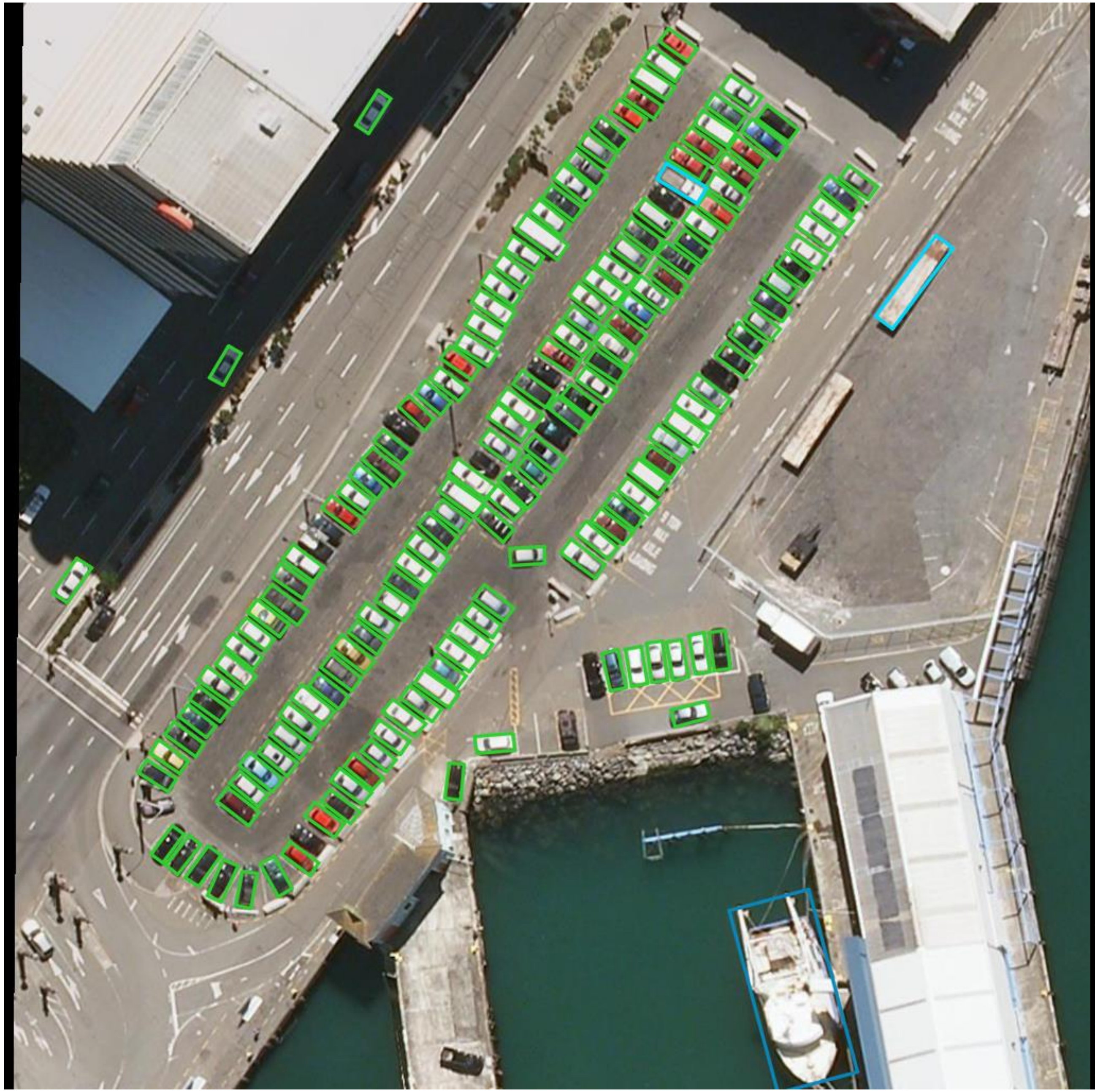
Detections on DOTA datasets

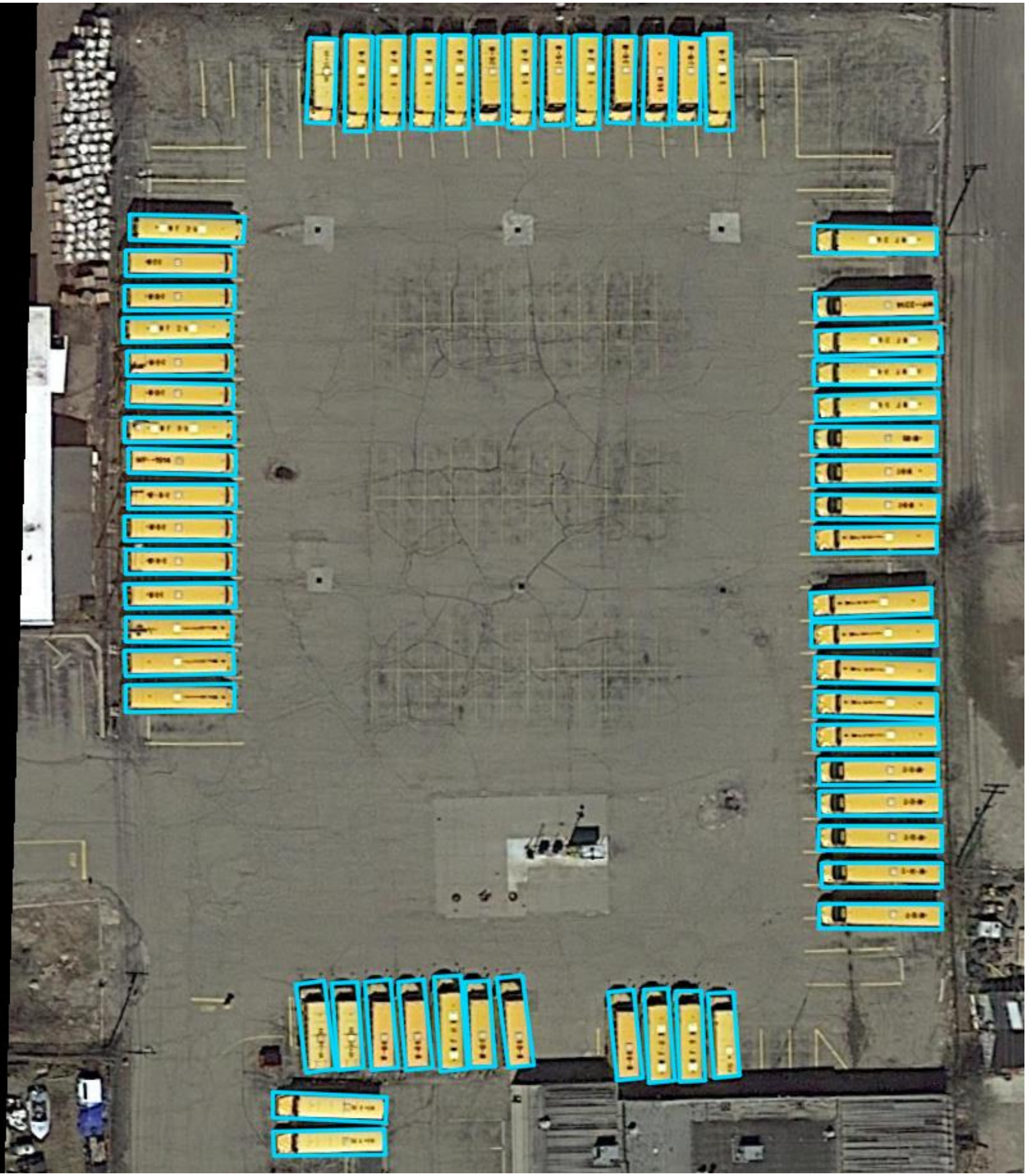








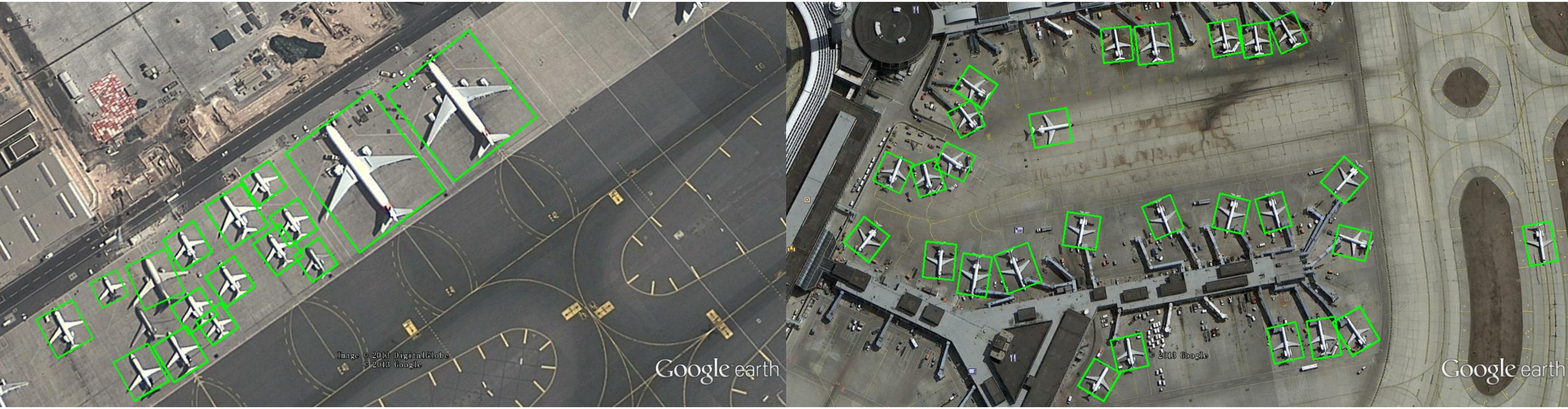


Detections on HRSC2016 datasets

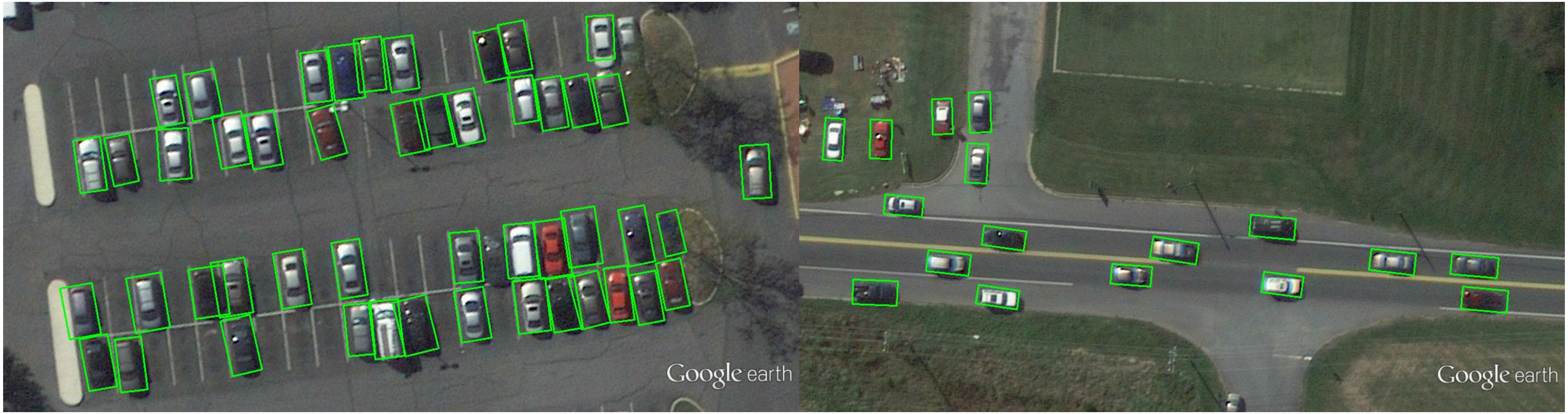



Detections on UCAS-AOD datasets


© 2013 Google
Google earth
Google earth


Google earth
Google earth


Google earth
Google earth


Google earth
Google earth

Detections on ICDAR 2015 datasets


GOLDEN VILLAGE
GOLD CLASS
HABA
LUSH
More shopping & dining options at B1, Marina Link
This way to
Esplanade Station
MARINA:SQUARE


epiCentre
VICTORIA'S SECRET
VICTORIA'S SECRET
VICTORIA'S SECRET

CATHAY
THE THEORY OF EVERYTHING
IN CINEMAS JANUARY 8
MORTDECAI


promod
More Shops Cafes
and Restaurants
Food Republic
Giant Hyperfresh
SALE SALE SALE
COTTON ON
NEW
NOW
SINCERE
exciting
privileges
Enjoy
great deals
Suntec City
ESPRIT

Detections with HBB

Detections with ICDAR 2013


TESCO
VALUE
Washing up
liquid

COSTA MED
mediterranean cuisine

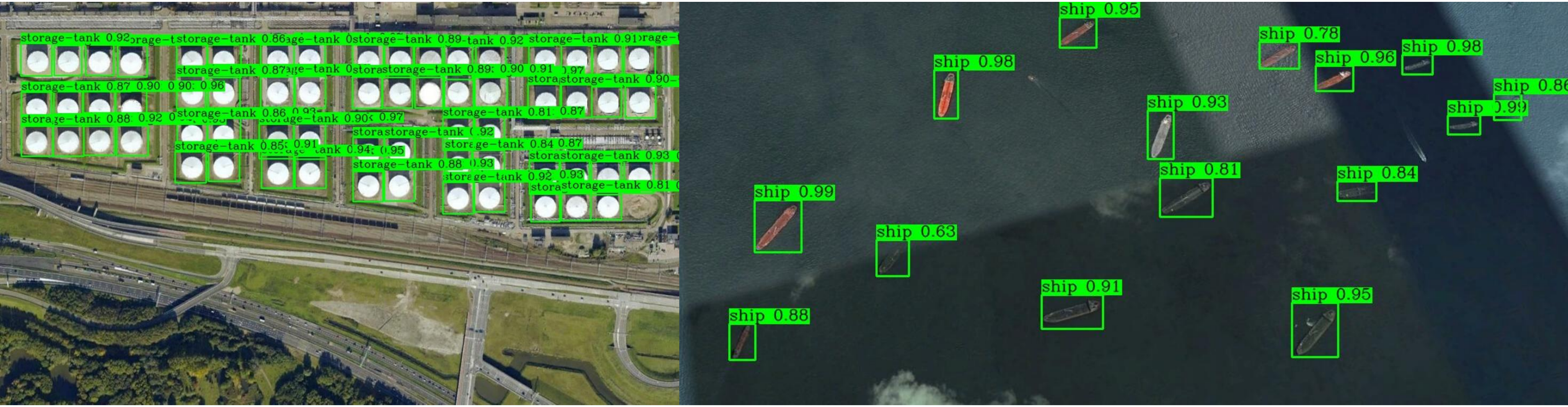

UK DANCE
PROTOTYPE &
INSPIRED RECORDS


Next
Departure
12:31
20
FIRST


University of Essex
Day Nursery
The Houses
Keynes, Rayleigh, Tawney
and William Morris Towers
Wolfson Court
Buses and Cycles Only
CONFERENCE CAR PARK
Entrance

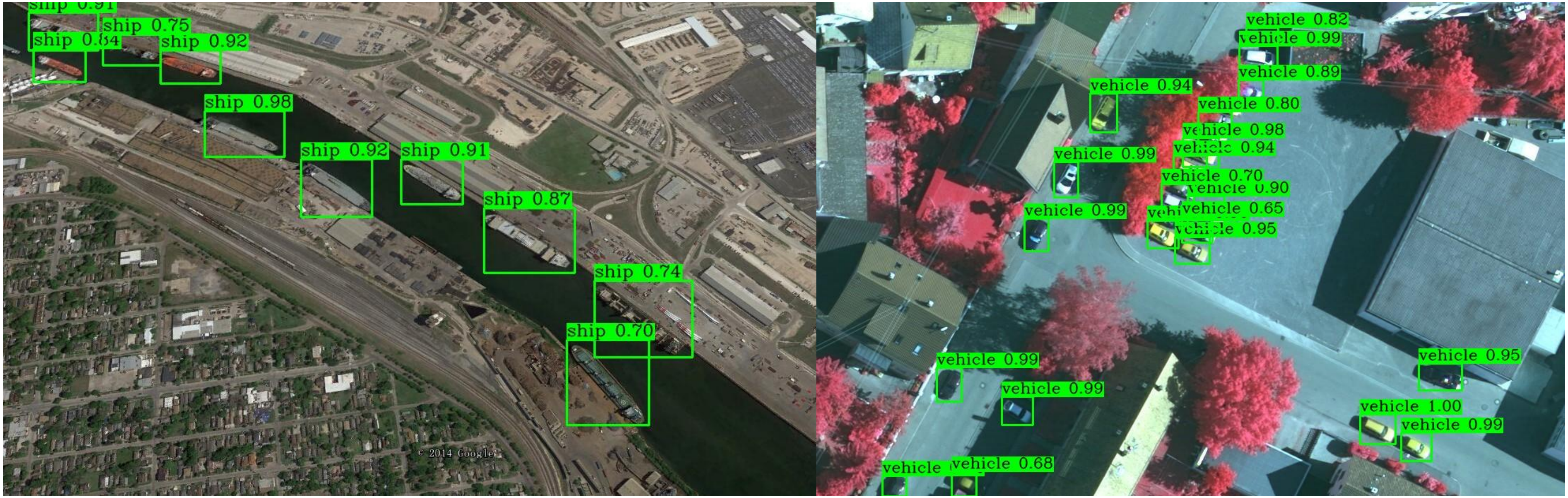

Please wait
here until
a cashier is
available

Detections with NWPU VHR-10



storage-tank 0.92
storage-tank 0.86
storage-tank 0.89
storage-tank 0.91
storage-tank 0.87
storage-tank 0.86
storage-tank 0.81
storage-tank 0.84
storage-tank 0.88
storage-tank 0.93
storage-tank 0.92
storage-tank 0.81
ship 0.95
ship 0.78
ship 0.96
ship 0.98
ship 0.98
ship 0.93
ship 0.99
ship 0.81
ship 0.84
ship 0.99
ship 0.63
ship 0.91
ship 0.95
ship 0.88

baseball-diamond 0.94
baseball-diamond 0.98
baseball-diamond 0.97
baseball-diamond 0.97
baseball-diamond 0.98
baseball-diamond 0.98
baseball-diamond 0.97
baseball-diamond 0.96
baseball-diamond 0.95
baseball-diamond 0.84
baseball-diamond 0.62


ship 0.91
ship 0.75
ship 0.84
ship 0.92
ship 0.98
ship 0.92
ship 0.91
ship 0.87
ship 0.74
ship 0.70
© 2014 Google
vehicle 0.82
vehicle 0.99
vehicle 0.89
vehicle 0.94
vehicle 0.80
vehicle 0.98
vehicle 0.94
vehicle 0.99
vehicle 0.70
vehicle 0.90
vehicle 0.65
vehicle 0.99
vehicle 0.95
vehicle 0.99
vehicle 0.99
vehicle 0.95
vehicle 1.00
vehicle 0.99
vehicle
vehicle 0.68

Comparision

with DAL

without DAL

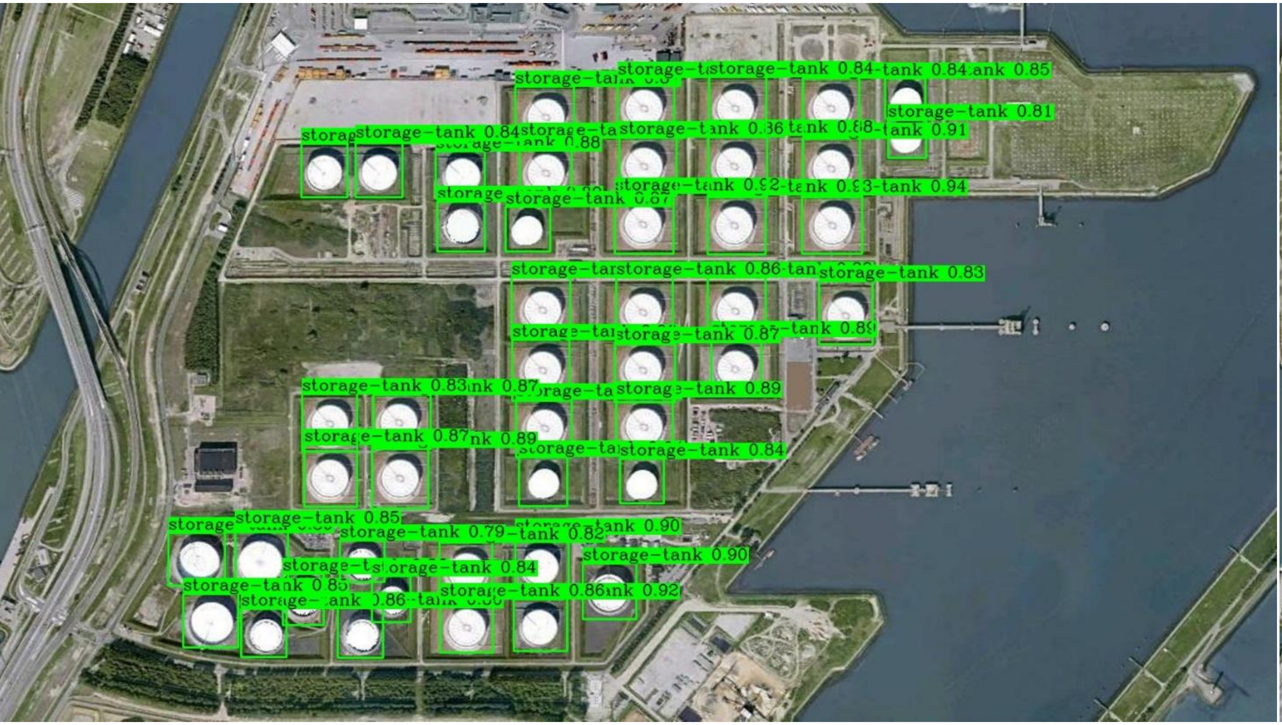
with DAL

without DAL

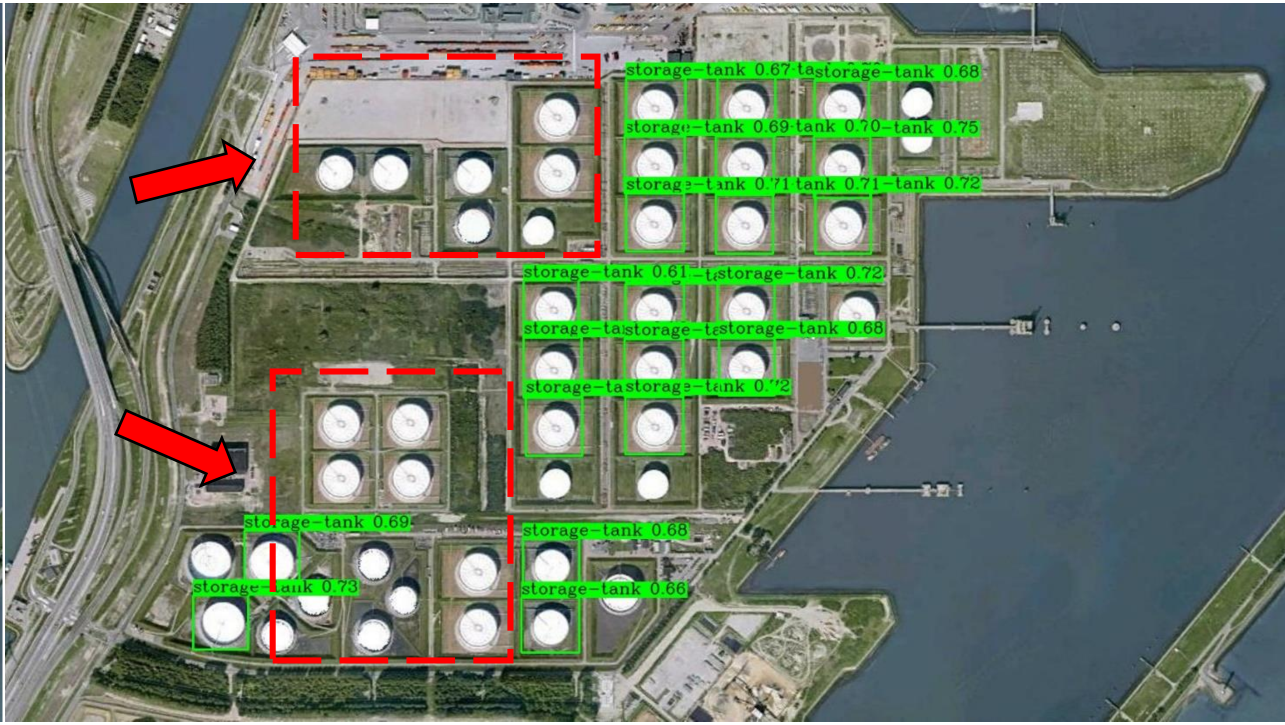
with DAL

without DAL

with DAL



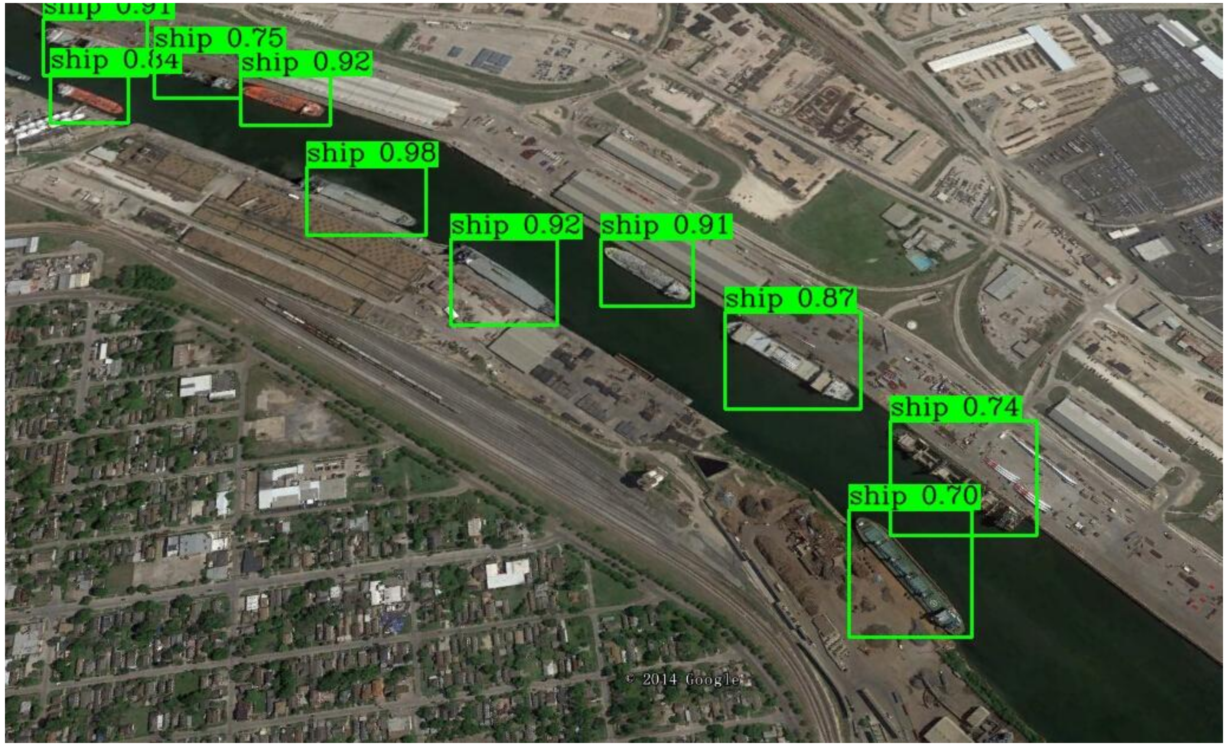
without DAL



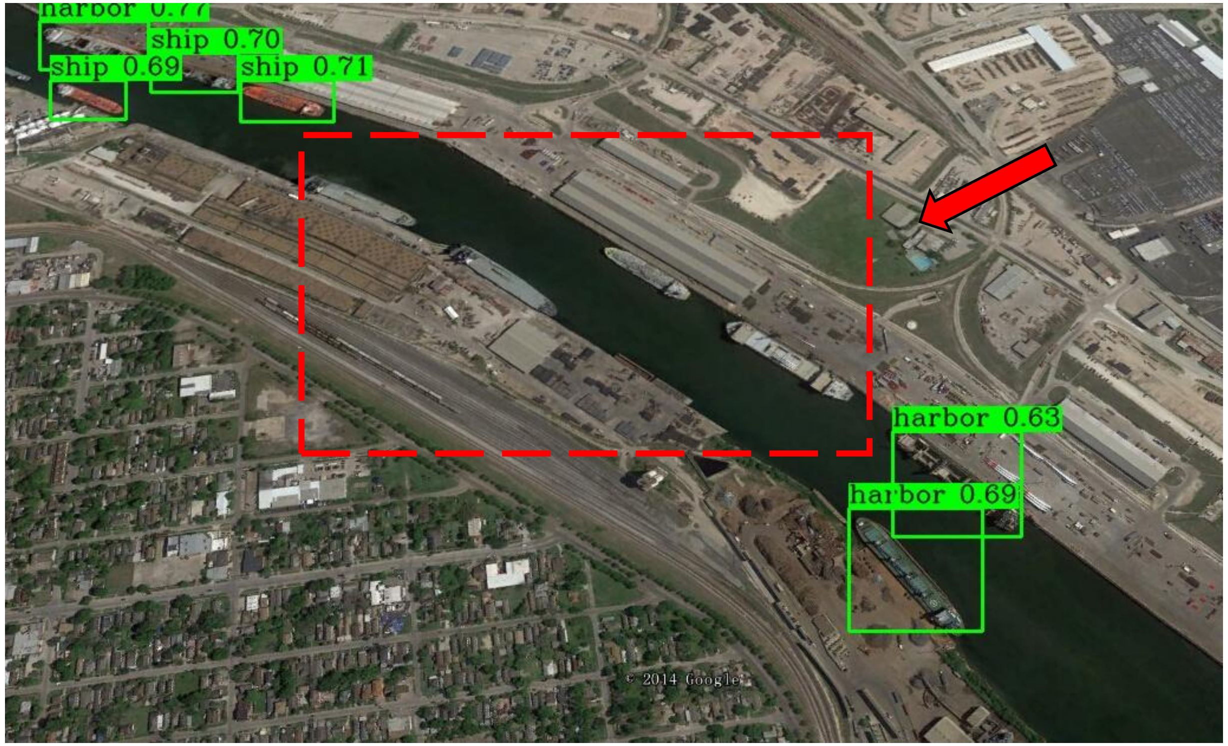
with DAL

without DAL

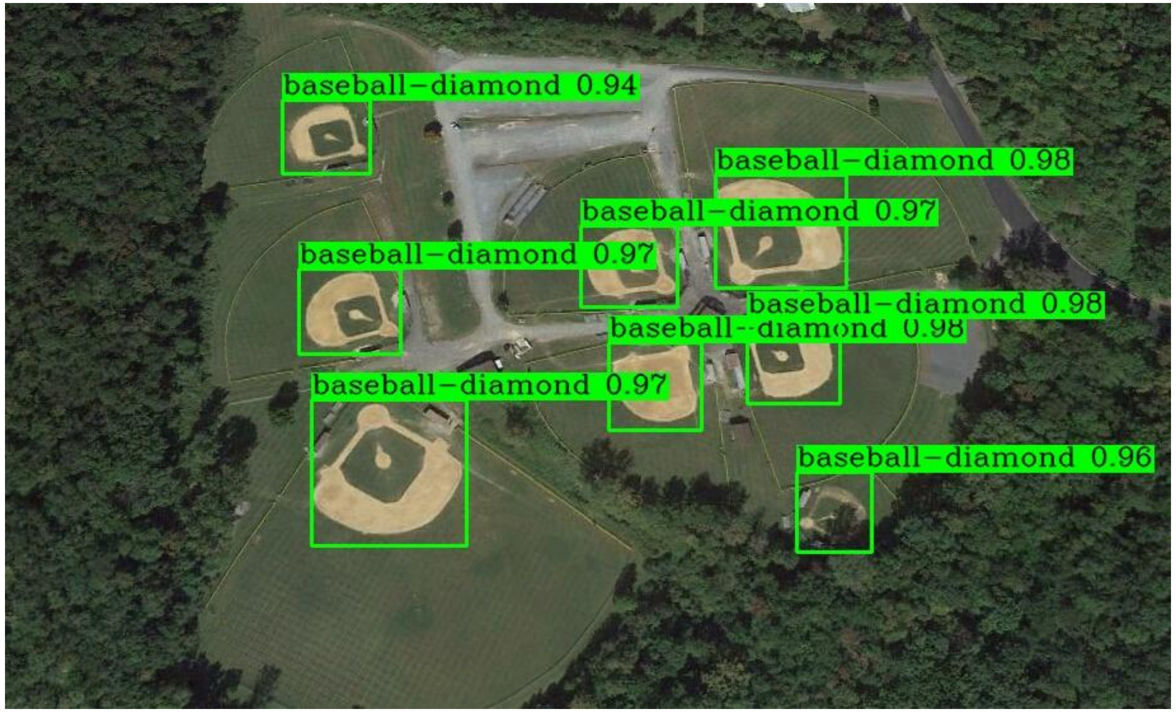
with DAL

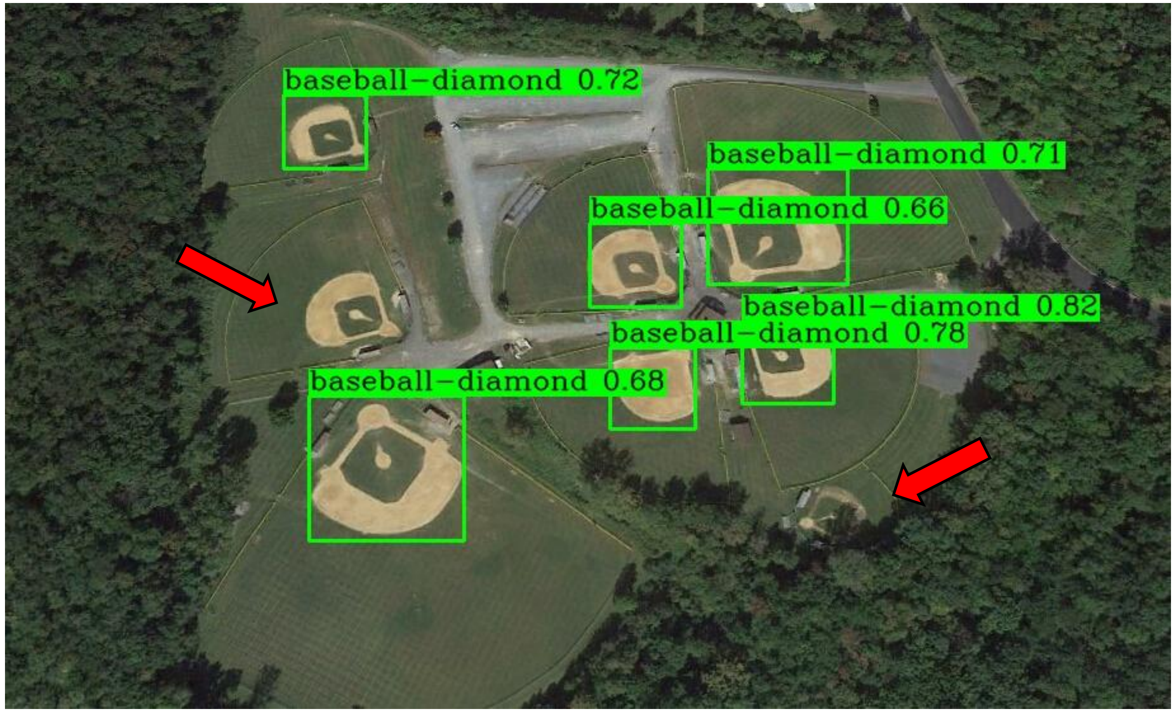


without DAL



with DAL **without DAL**





Q&A

Thank you!

You can contact me via mq_chaser@126.com.

Qi Ming

2021.2.9