BUAFAI 2019

AI未来之光

第一届北京高校人工智能学术论坛

2019 Beijing Universities Academic Forum of Artificial Intelligence

「 主办单位 」 清华大学自动化系研究生会 中国科学院自动化研究所研究生会

北京大学信息科学技术学院研究生会 北京大学软件与微电子学院研究生会 北京交通大学电子信息工程学院研究生会 北京航空航天大学自动化科学与电气工程学院研究生会

北京理工大学自动化学院研究生会 北京邮电大学自动化学院研究生会 中国科学院计算技术研究所研究生会 中国科学院计算机网络信息中心研究生会

Deep Fitting Degree Scoring Network for Monocular 3D Object Detection

Lijie Liu, Jiwen Lu, Chunjing Xu, Qi Tian, Jie Zhou





CVPR LONG BEACH CALIFORNIA June 16-20, 2019
清華大學 Tsinghua University
i-VisionGroup
HUAWEI | NOAH'S ARK LAB

目录 / CONTENTS

BUAFAI 2019

BUAFAI 2019

01 Problem

Monocular 3D Object Detection

"Perception is our best guess as to what is in the world given our current sensory input and our prior experience"

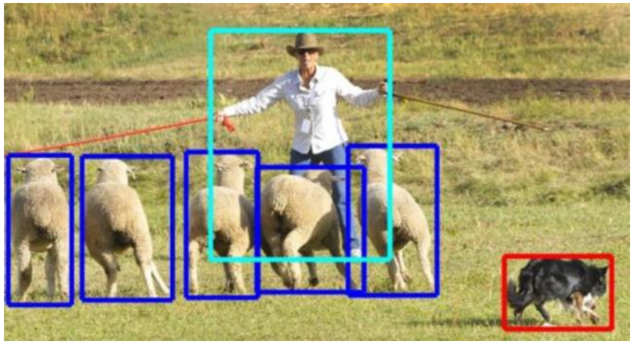
- Helmholtz (1866)

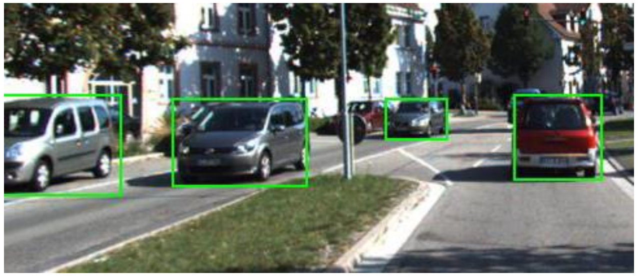
Problem Setting: Monocular 3D Object Detection

- Input: A single RGB image & camera intrinsic
- Solve: 9 DoF, including orientation, dimension, location
- Cue: Appearance (sensory input) & Projection Law (prior experience)

Why 3D perception

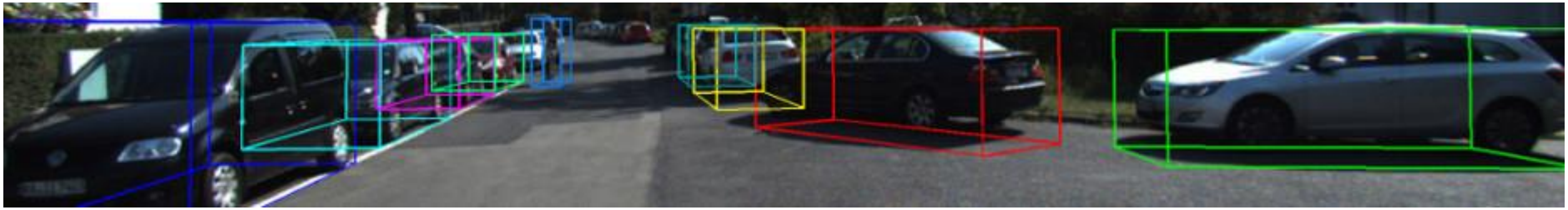
- 3D perception is the key to human intelligence
- Autonomous driving & Robotic grasping





Challenges in Location Estimation

- Dimension and orientation estimation are easier than location estimation
- Ambiguities arising from 2D-3D mapping
- Real 3D information unavailable
- Occlusions, Truncation, Scale variation……



BUAFAI 2019

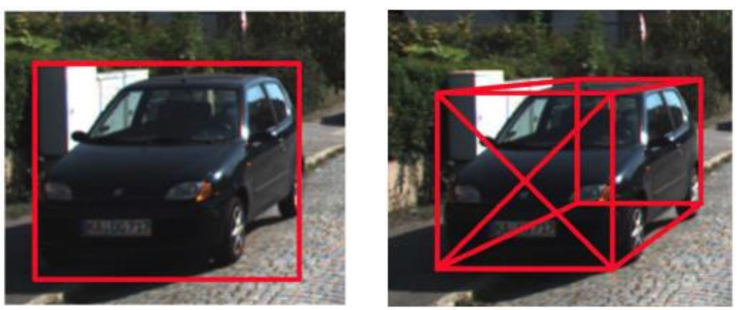
02

Motivation

How we come up with our idea

Previous Methods

- Tight Constraints （CVPR17）
- Solves the location by placing the 3D proposal in the 2D bounding box compactly.



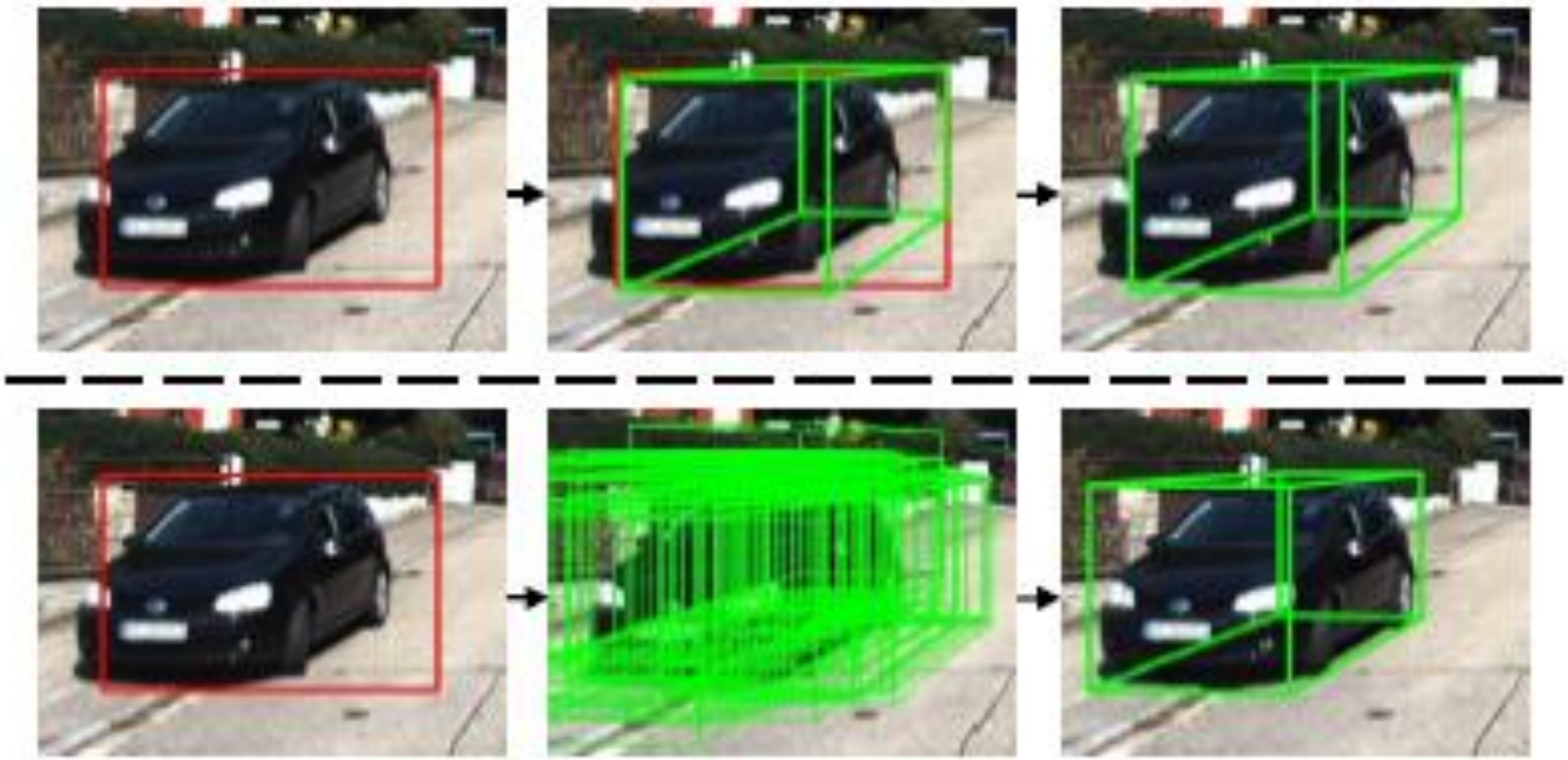
Drawbacks:

1) Image appearance clue is not used

2) Performance highly depends on the 2D detection accuracy

How do human do?

- Hypothesize-and-Verify



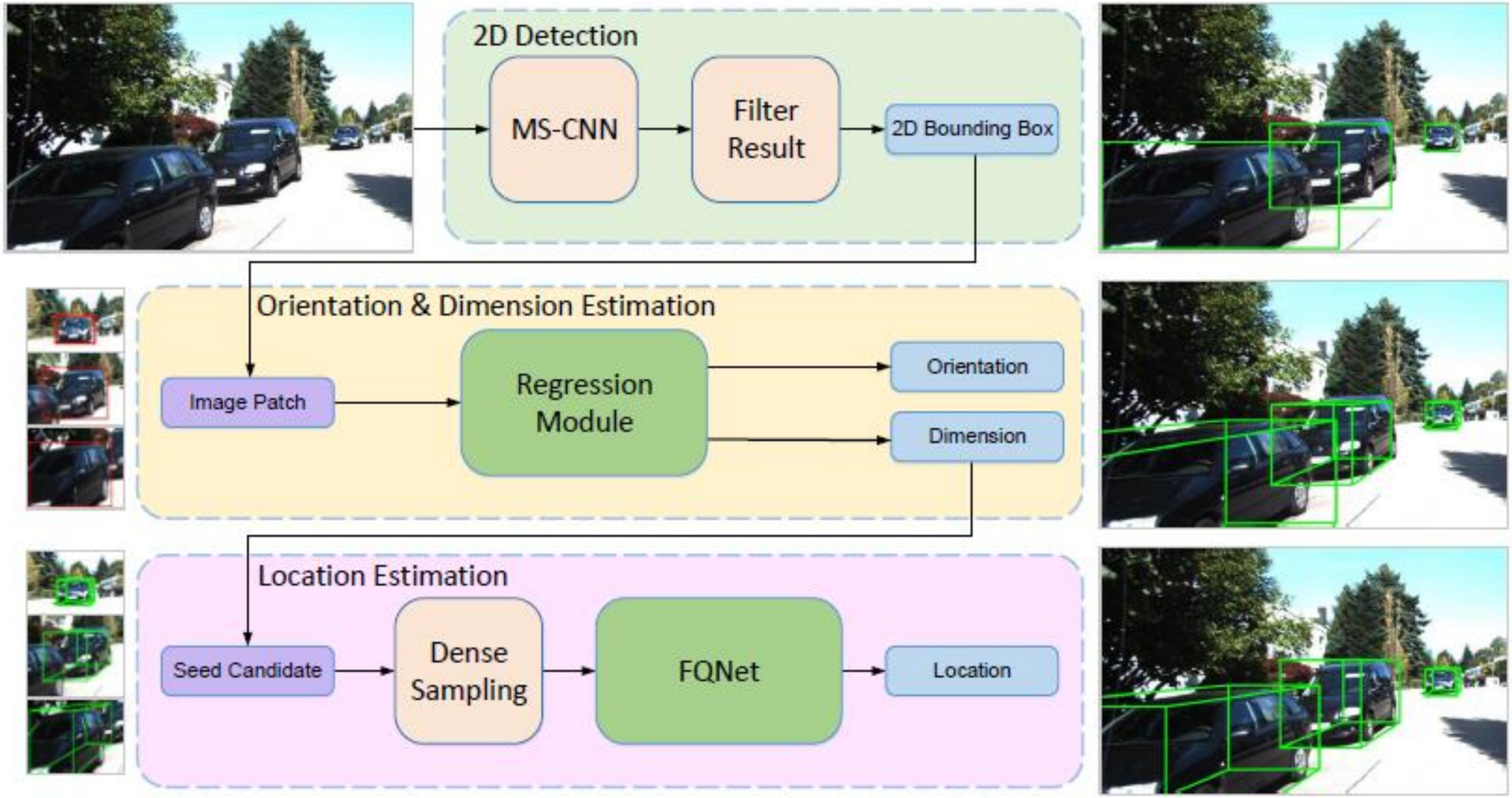
BUAFAI 2019

03 Approach

Some details

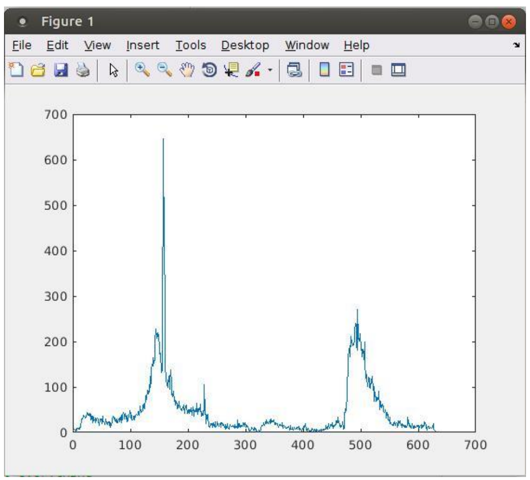
Overall framework

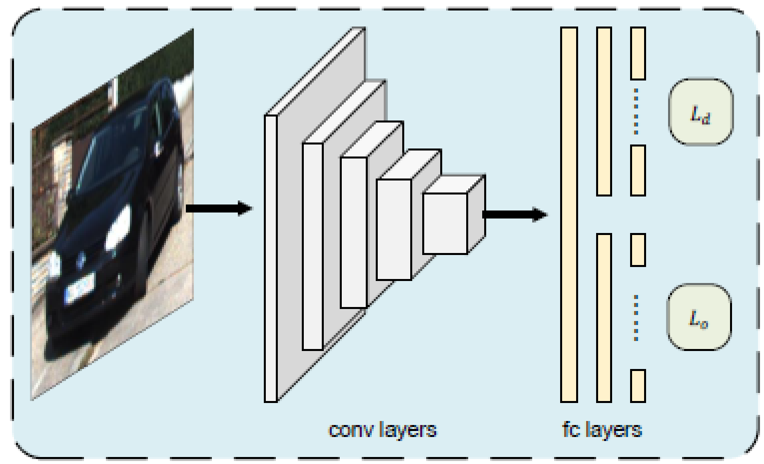
- Three-stage



Regression Module

- Anchor cuboid & Anchor angle



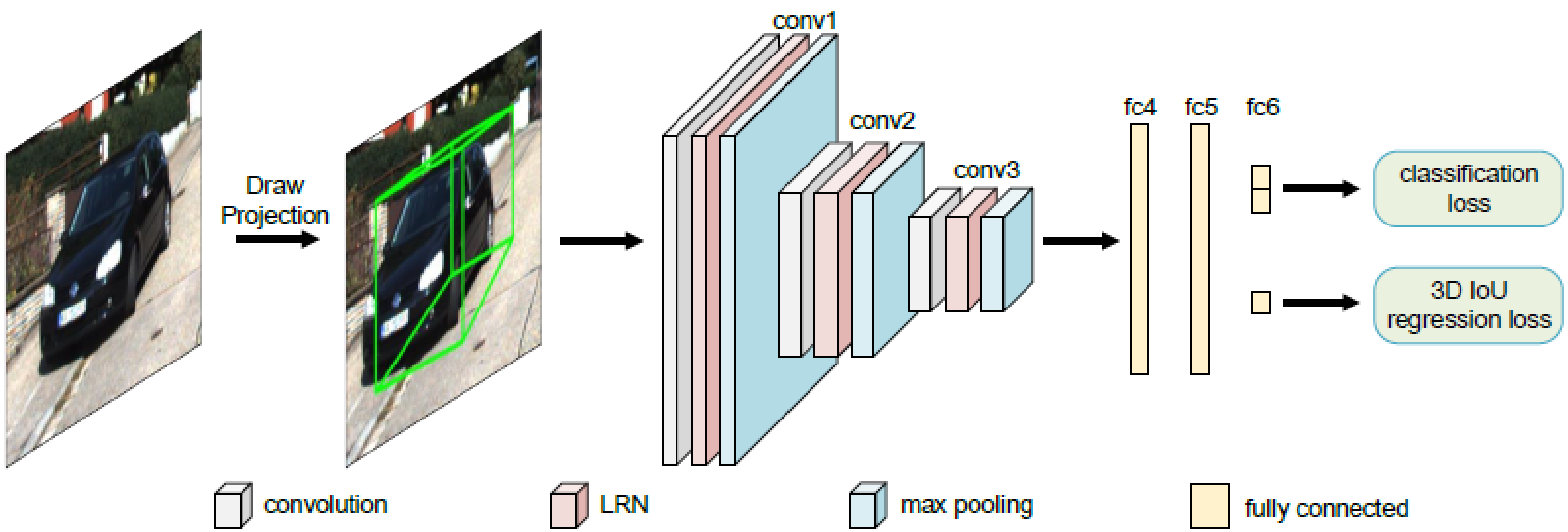


$$L_d = -\log\sigma(c_{i^\star}) + [1 - IoU(\boldsymbol{A}_{i^\star} + [\Delta w_{i^\star}, \Delta h_{i^\star}, \Delta l_{i^\star}], \boldsymbol{G})]$$

$$L_o = -\log\sigma(c_{i^\star}) + [1 - \cos(\Theta_{i^\star} + \Delta\theta_{i^\star} - \theta_G)]$$

Location Estimation

- Dense sampling
- FQNet



$$\Theta^{\star} = \arg\min_{\Theta} ||\mathcal{F}(\boldsymbol{I}, \boldsymbol{S}_i|\Theta) - IoU(\boldsymbol{I}, \boldsymbol{S}_i)||$$

BUAFAI 2019

04 Experiments

Some demonstration

Experimental Results on KITTI dataset

➢ Orientation & Dimension

Table 1. Comparisons of the Average Orientation Similarity (AOS) with the state-of-the-art methods on the KITTI dataset.

Method	Easy			Moderate			Hard		
	train/val 1	train/val 2	test	train/val 1	train/val 2	test	train/val 1	train/val 2	test
3DOP [9]	91.58	-	91.44	85.80	-	86.10	76.80	-	76.52
Mono3D [8]	91.90	-	91.01	86.28	-	86.62	77.09	-	76.84
3DVP [42]	-	78.99	86.92	-	65.73	74.59	-	54.67	64.11
SubCNN [43]	-	94.55	90.67	-	85.03	88.62	-	72.21	78.68
Deep3DBox [31]	-	97.50	**92.90**	-	96.30	88.75	-	80.40	76.76
3D-RCNN [23]	90.70	**97.70**	89.98	89.10	96.50	**89.25**	**79.50**	**80.70**	**80.07**
Our Method	**97.28**	97.57	92.58	**93.70**	**96.70**	88.72	79.25	80.45	76.85

Method	train/val 1	train/val 2
3DOP [9]	0.3527	-
Mono3D [8]	0.4251	-
Deep3DBox [31]	-	0.1934
Our Method	**0.1698**	**0.1465**

Experimental Results on KITTI dataset

➢ Location

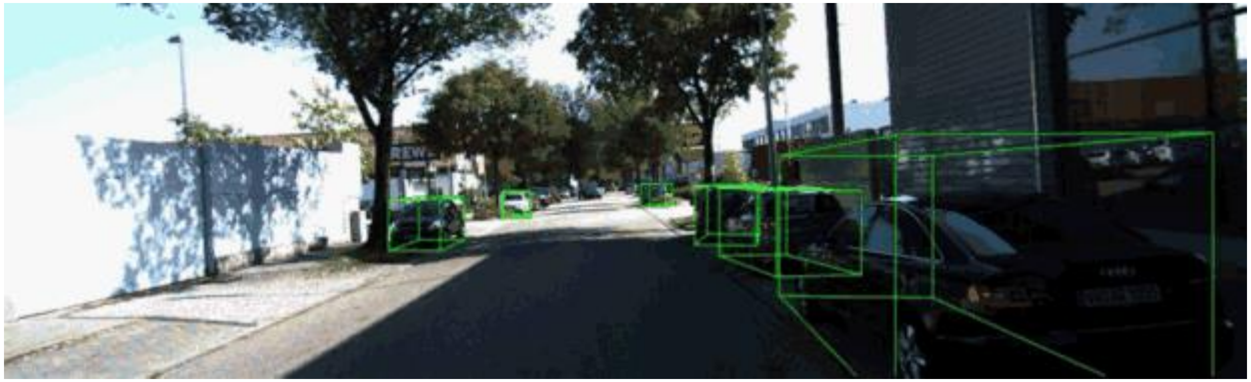
Table 2. Comparisons of the 2D AP with the state-of-the-art methods on the KITTI Birds Eyed View validation dataset.

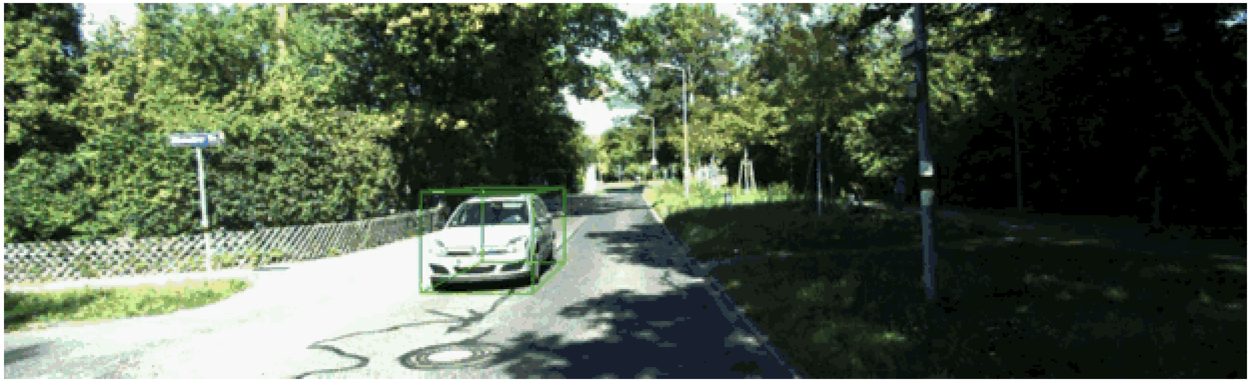
Method	IoU = 0.5						IoU = 0.7					
	Easy		Moderate		Hard		Easy		Moderate		Hard	
	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2
3DOP [9]	55.04	-	41.25	-	34.55	-	12.63	-	9.49	-	7.59	-
Mono3D [8]	30.50	-	22.39	-	19.16	-	5.22	-	5.19	-	4.13	-
Deep3DBox [31]	-	30.02	-	23.77	-	18.83	-	9.99	-	7.71	-	5.30
Our Method	32.57	33.37	24.60	26.29	21.25	21.57	9.50	10.45	8.02	8.59	7.71	7.43

Table 4. Comparisons of the 3D AP with the state-of-the-art methods on the KITTI 3D Object validation dataset.

Method	IoU = 0.5						IoU = 0.7					
	Easy		Moderate		Hard		Easy		Moderate		Hard	
	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2	t/v 1	t/v 2
3DOP [9]	46.04	-	34.63	-	30.09	-	6.55	-	5.07	-	4.10	-
Mono3D [8]	25.19	-	18.20	-	15.52	-	2.53	-	2.31	-	2.31	-
Deep3DBox [31]	-	27.04	-	20.55	-	15.88	-	5.85	-	4.10	-	3.84
Our Method	28.16	28.98	21.02	20.71	19.91	18.59	5.98	5.45	5.50	5.11	4.75	4.45

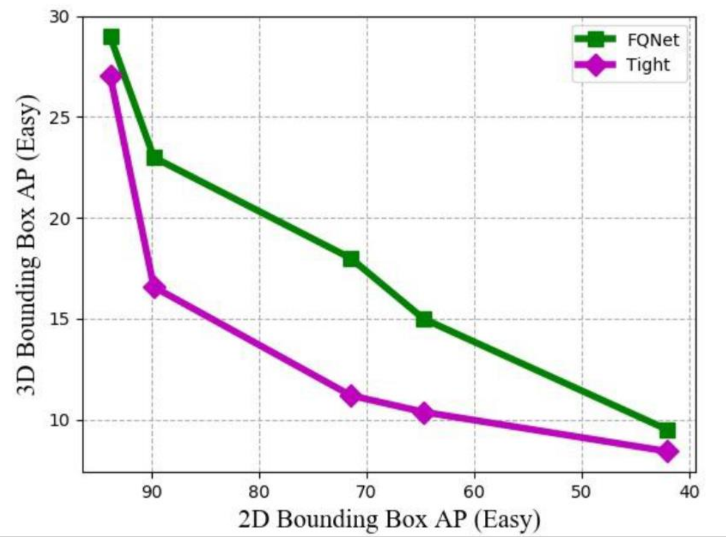
Qualitative Results

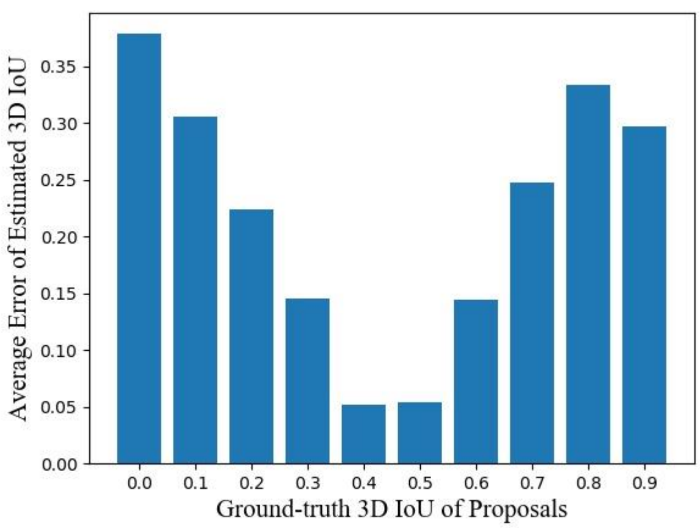




Effectiveness

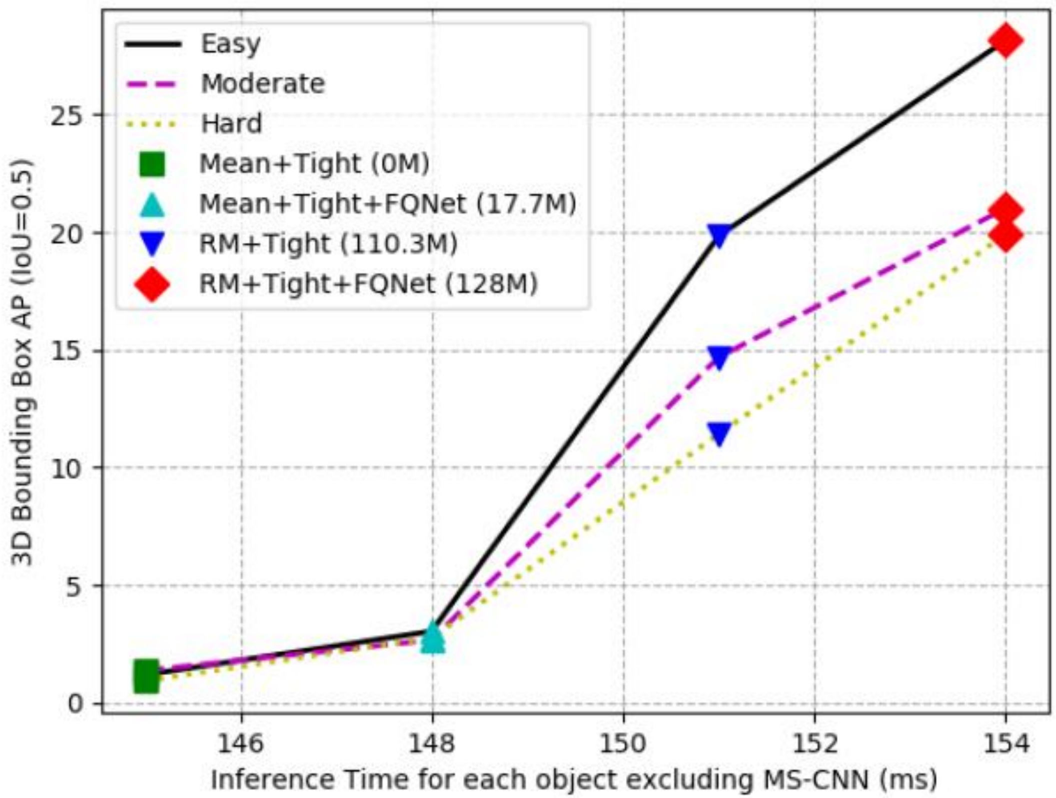
- Not sensitive to 2D detection precision
- 3D IoU regression





Accuracy vs Speed

- Ablation study
- Efficiency



THANK YOU

BUAFAI 2019