Data Driven 3D Voxel Patterns for Object Category Recognition

Yu Xiang[1,2], Wongun Choi[3], Yuanqing Lin[3], and Silvio Savarese[1]

[1]Stanford University, [2]University of Michigan at Ann Arbor

[3]NEC Laboratories America, Inc.

CVPR 2015

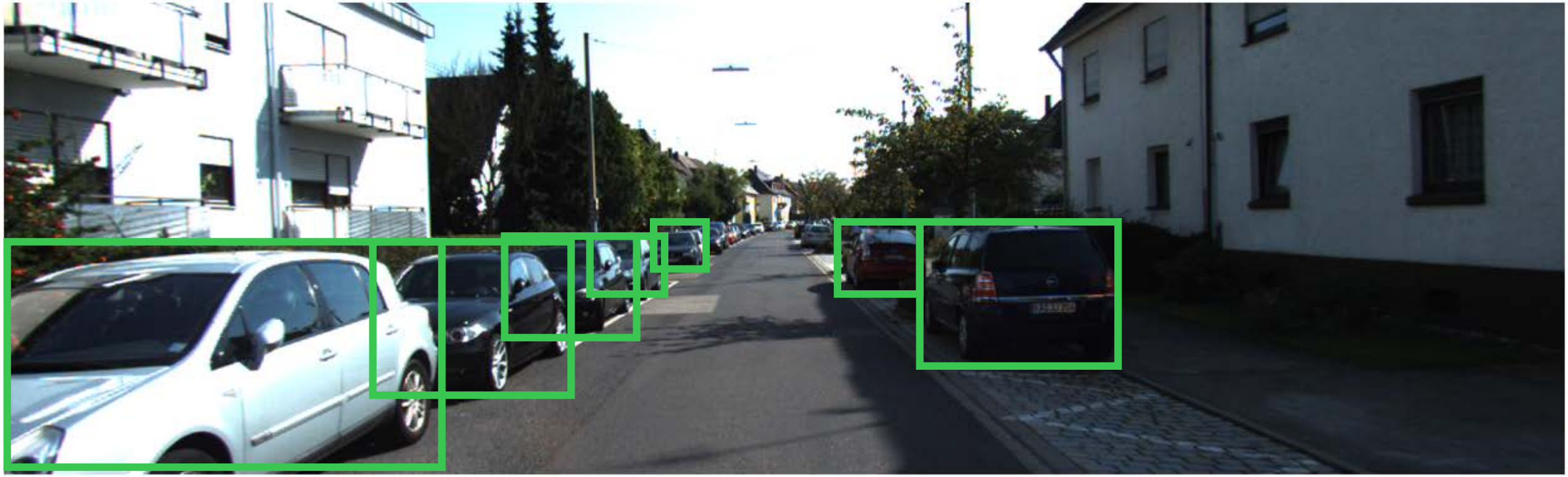
CV GL Computational Vision & Geometry Lab

NEC Laboratories America

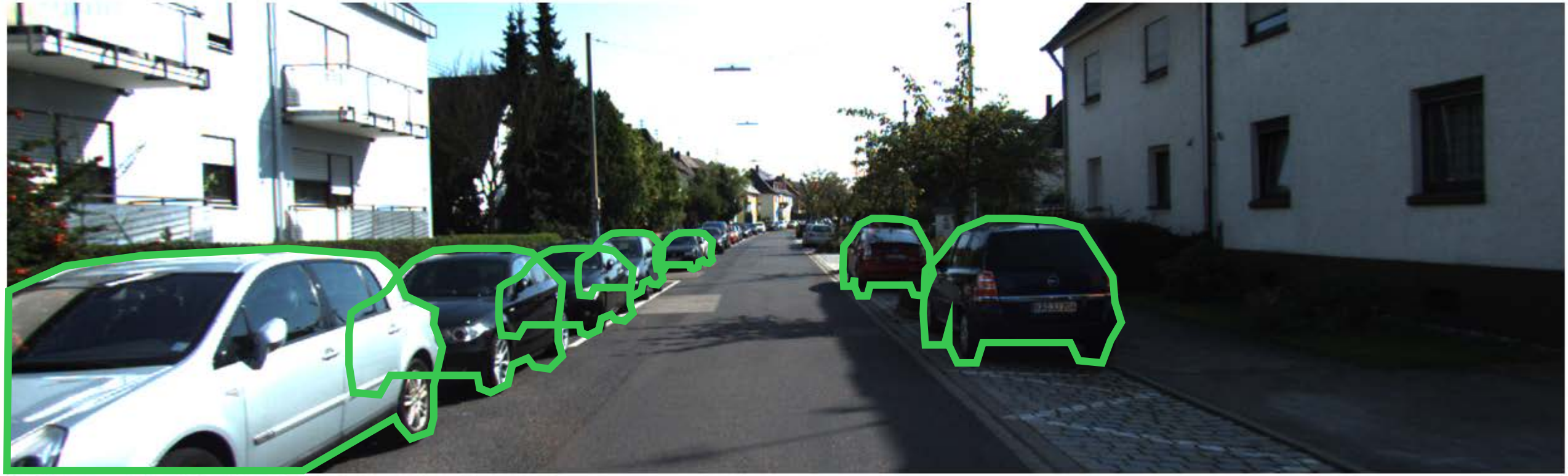
Relentless passion for innovation

The image is from the KITTI detection benchmark (Geiger et al. CVPR’12)

2D Object Detection

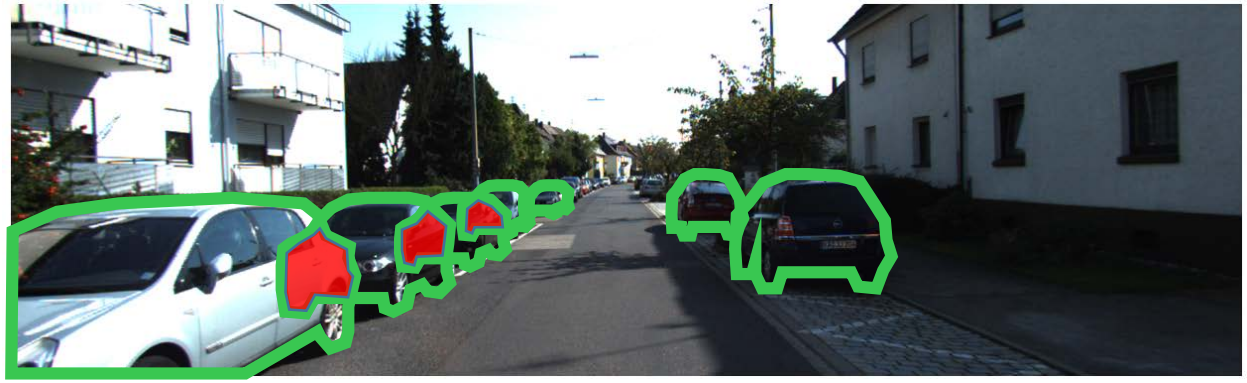


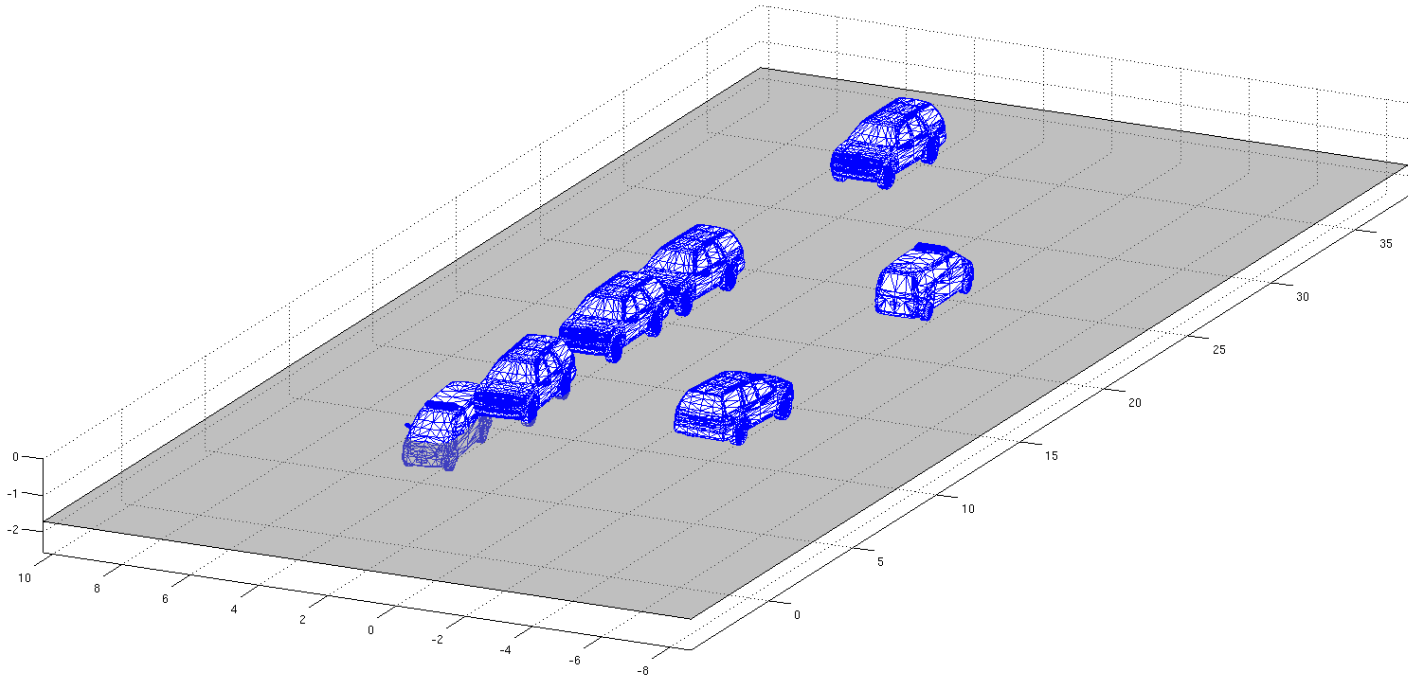
2D Object Segmentation



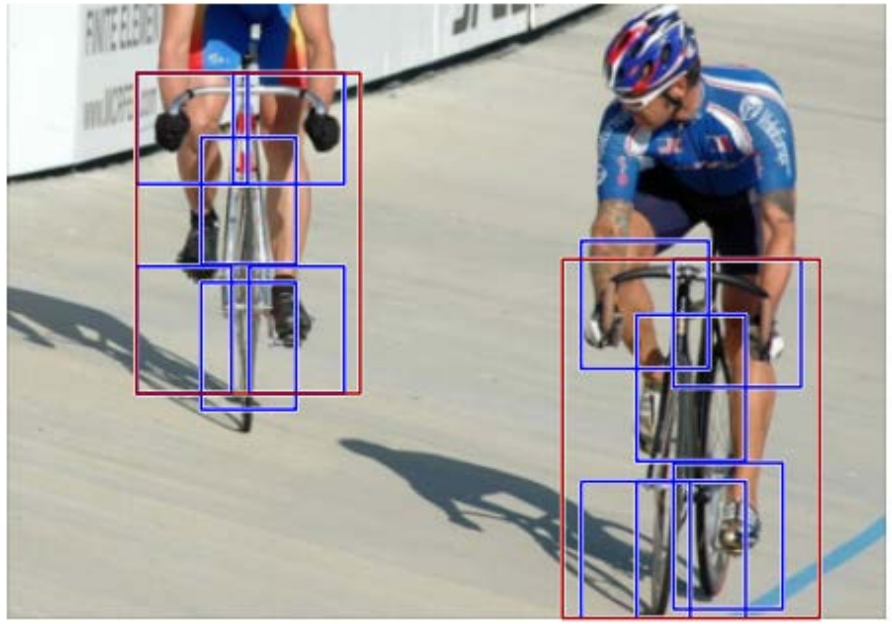
Occlusion Reasoning

3D Localization





Related Work: 2D Object Detection



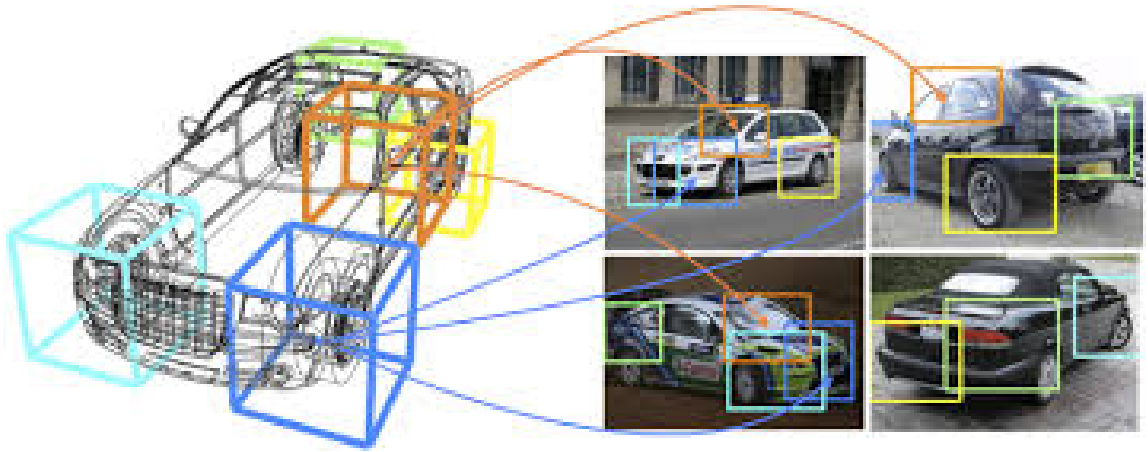
Deformable part model
Felzenszwalb et al., TPAMI'10

- ✓ 2D detection
- ✘ 3D pose
- ✘ Occlusion
- ✘ 3D location

- Viola & Jones, IJCV'01
- Fergus et al. , CVPR'03
- Leibe et al., ECCVW'04
- Hoiem et al., CVPR'06
- Vedaldi et al., ICCV'09
- Maji & Malik, CVPR'09
- Felzenszwalb et al., TPAMI'10
- Malisiewicz et al., ICCV'11
- Divvala et al., ECCVW'12
- Dolla´r et al., TPAMI'14

Etc.

Related Work: 3D Pose Estimation



3DDPM
Pepik et al., CVPR'12

- ✓ 2D detection
- ✓ 3D pose
- ✗ Occlusion
- ✗ 3D location

- Thomas et al., CVPR'06
- Savarese & Fei-Fei ICCV'07
- Yan et al., ICCV'07
- Hoiem et al., CVPR'07

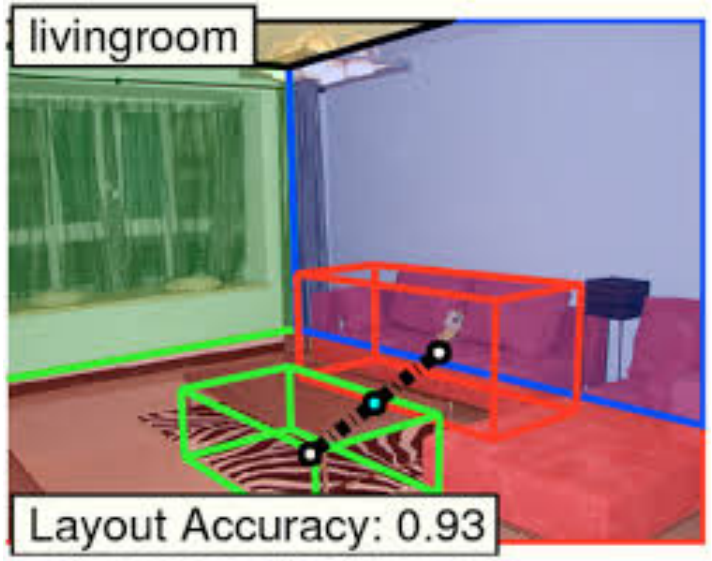
- Kushal et al., CVPR'07
- Su et al., ICCV'09
- Sun et al., CVPR'10
- Liebelt et al., CVPR'08, 10

- Glasner et al. ICCV'11
- Pepik et al., CVPR'12
- Xiang & Savarese, CVPR'12
- Hejrati & Ramanan, NIPS'12

- Fidler et al., NIPS'12

Etc.

Related Work: Model Object Relationships



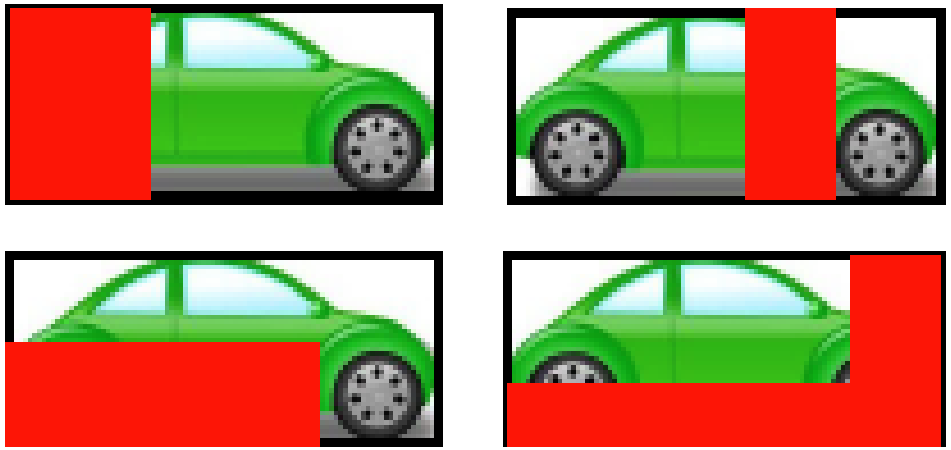
Geometric Phrases
Choi et al., CVPR'13

- ✓ 2D detection
- ✓ 3D pose
- ✗ Occlusion
- ✗ 3D location

- Desai et al., ICCV'09
- Yang et al., CVPR'10
- Gupta et al., ECCV'10
- Sadeghi & Farhadi, CVPR'11
- Li et al., CVPR'12
- Choi et al., CVPR'13

Etc.

Related Work: Handle Occlusion



Occlusion masks
Zia et al., CVPR'13

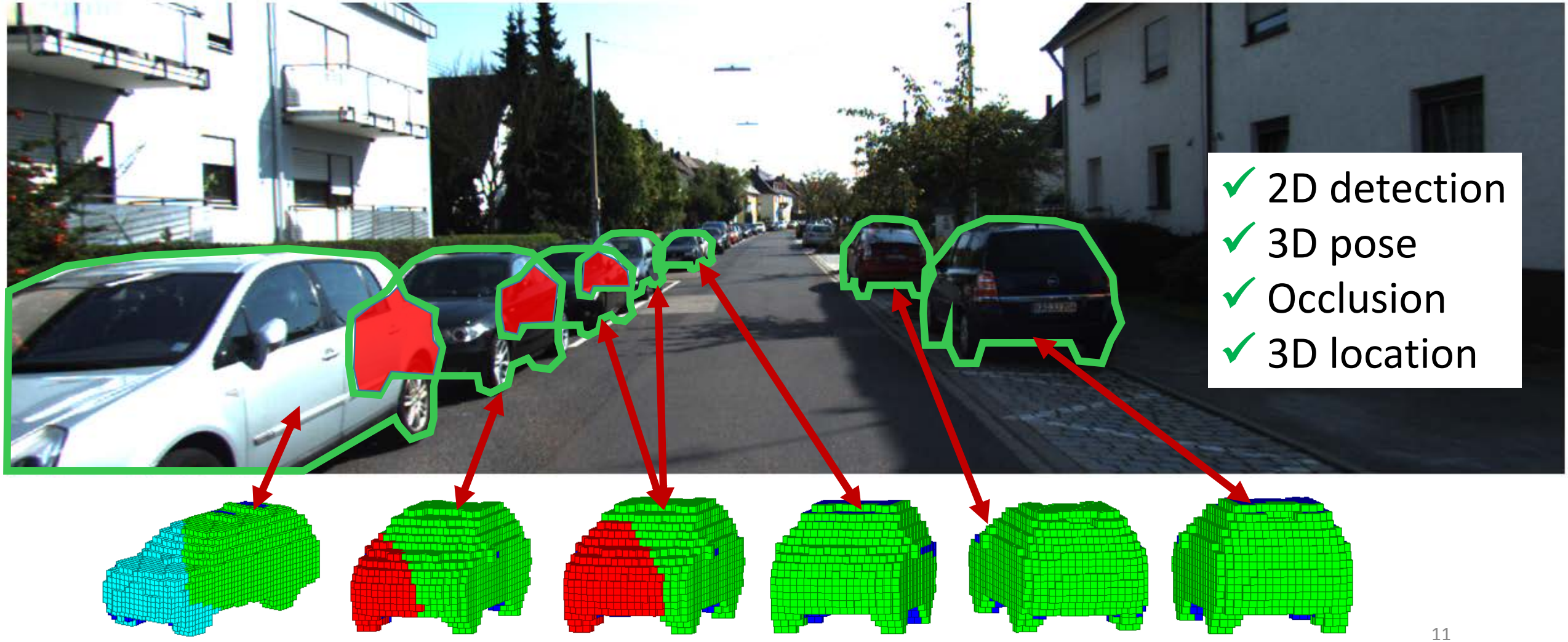
- ✓ 2D detection
- ✗ 3D pose
- ✓ Occlusion
- ✗ 3D location

- Wu and Nevatia, ICCV'05
- Wang et al., ICCV'09
- Gao et al., CVPR'11
- Meger et al., BMVC'11

- Wojek et al., CVPR'11
- Pepik et al., CVPR'13
- Xiang & Savarese, ICCVW'13
- Zia et al., CVPR'13, 14

Etc.

Our Contribution: Data-Driven 3D Voxel Patterns



Outline

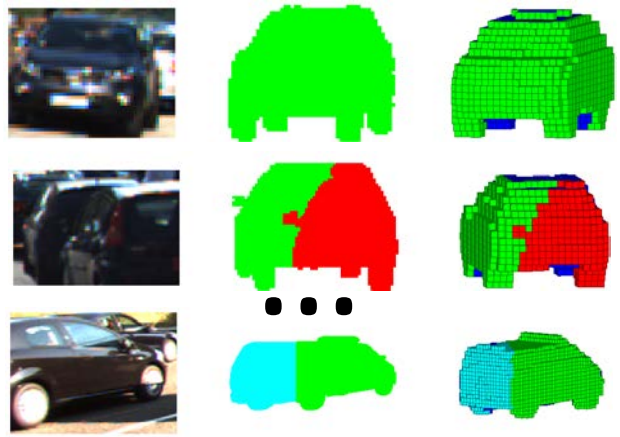
- Training Pipeline
- Testing Pipeline
- Experiments
- Conclusion

Outline

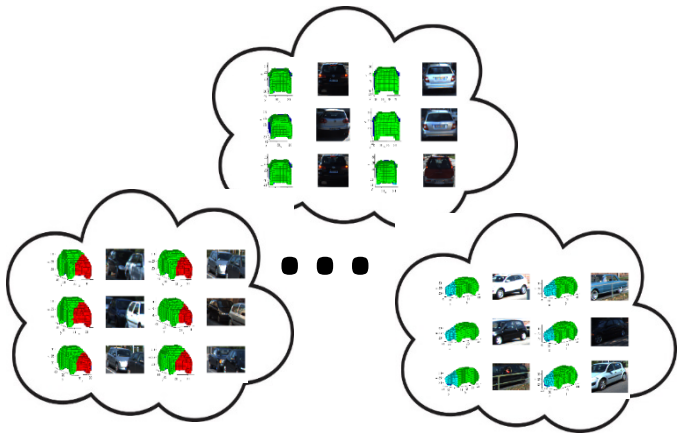
- Training Pipeline
- Testing Pipeline
- Experiments
- Conclusion

Training Pipeline Overview

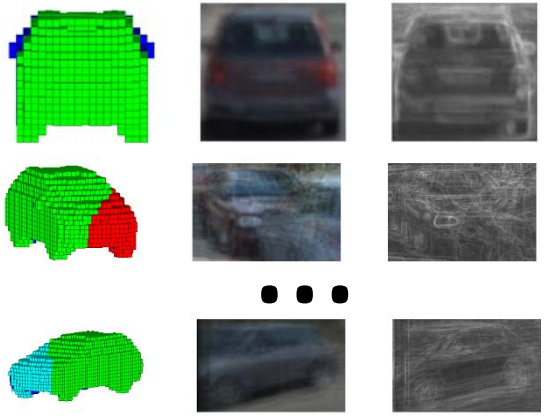
1. Align 2D images with 3D CAD models



2. 3D voxel exemplars

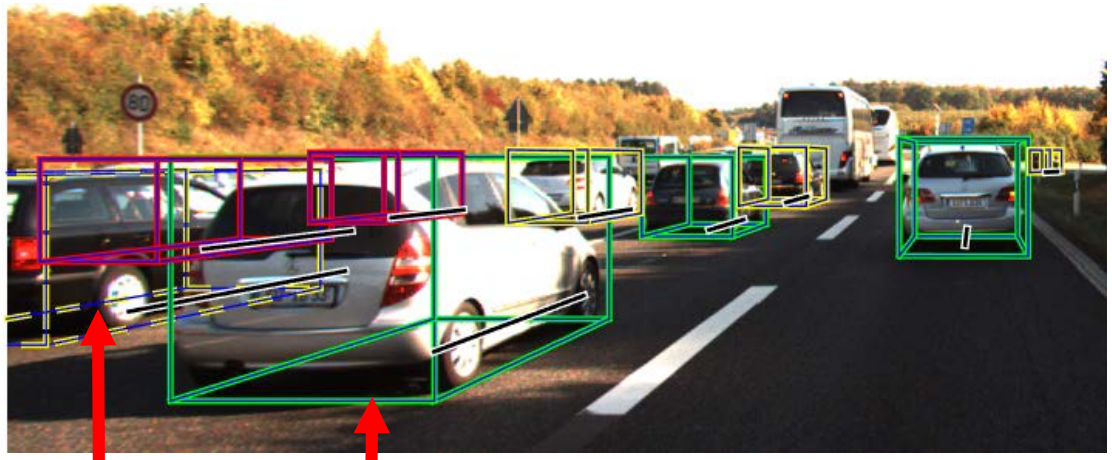


3. 3D voxel patterns

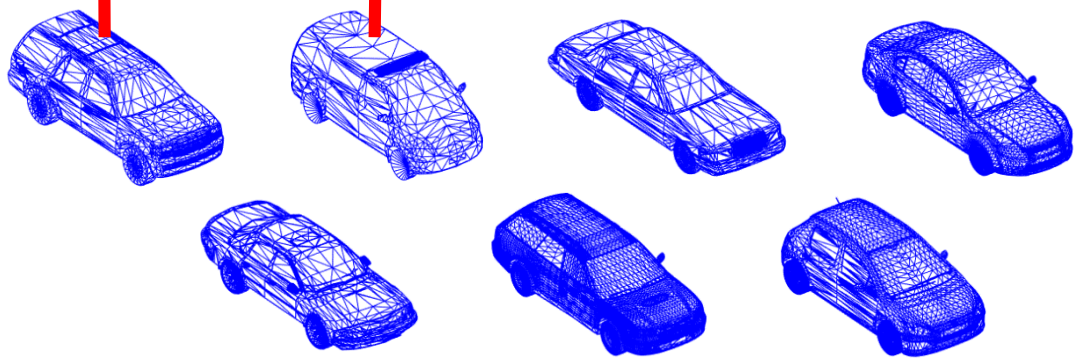


4. Training 3D voxel pattern detectors

1. Align 2D Images with 3D CAD Models



3D annotations • • •



3D CAD models

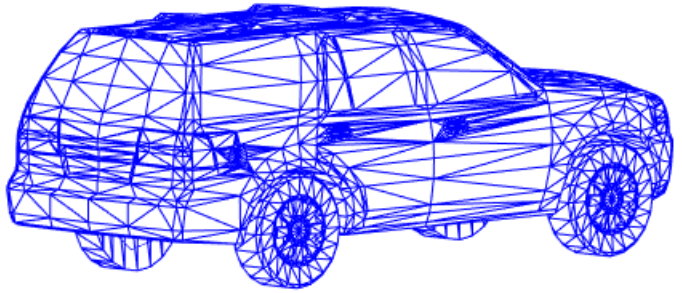
Project of 3D CAD models

Depth ordering

A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012

2. Building 3D Voxel Exemplars

Depth ordering



3D CAD model

Voxelization

truncated

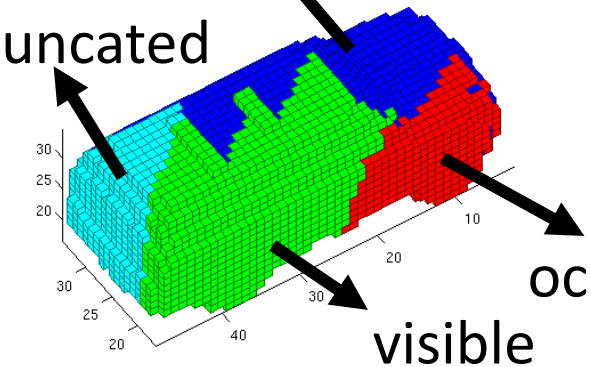
visible

occluded

2D mask labeling

self-occluded

truncated



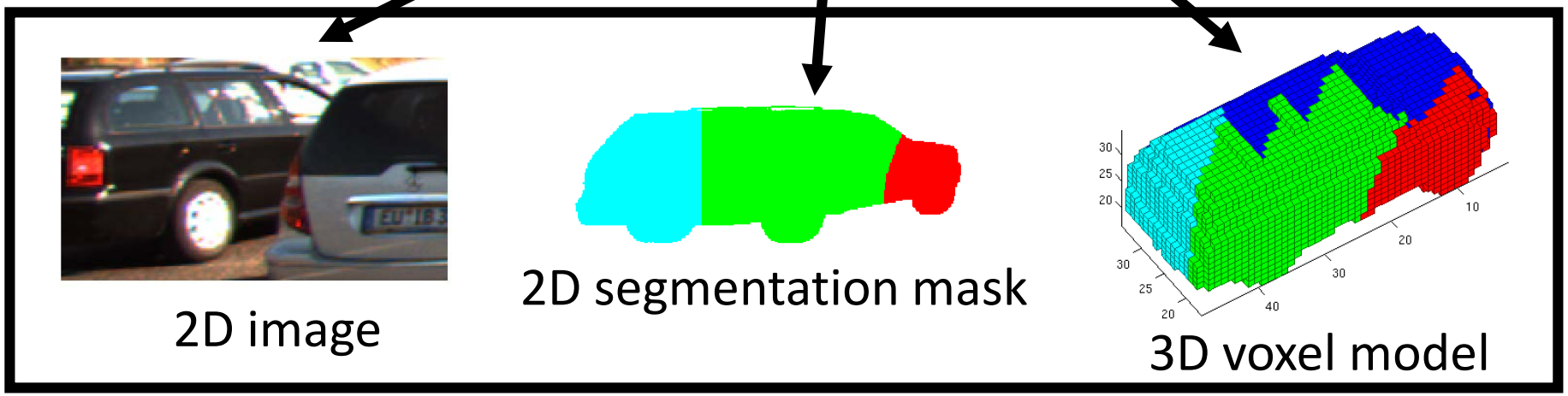
occluded

visible

3D voxel model

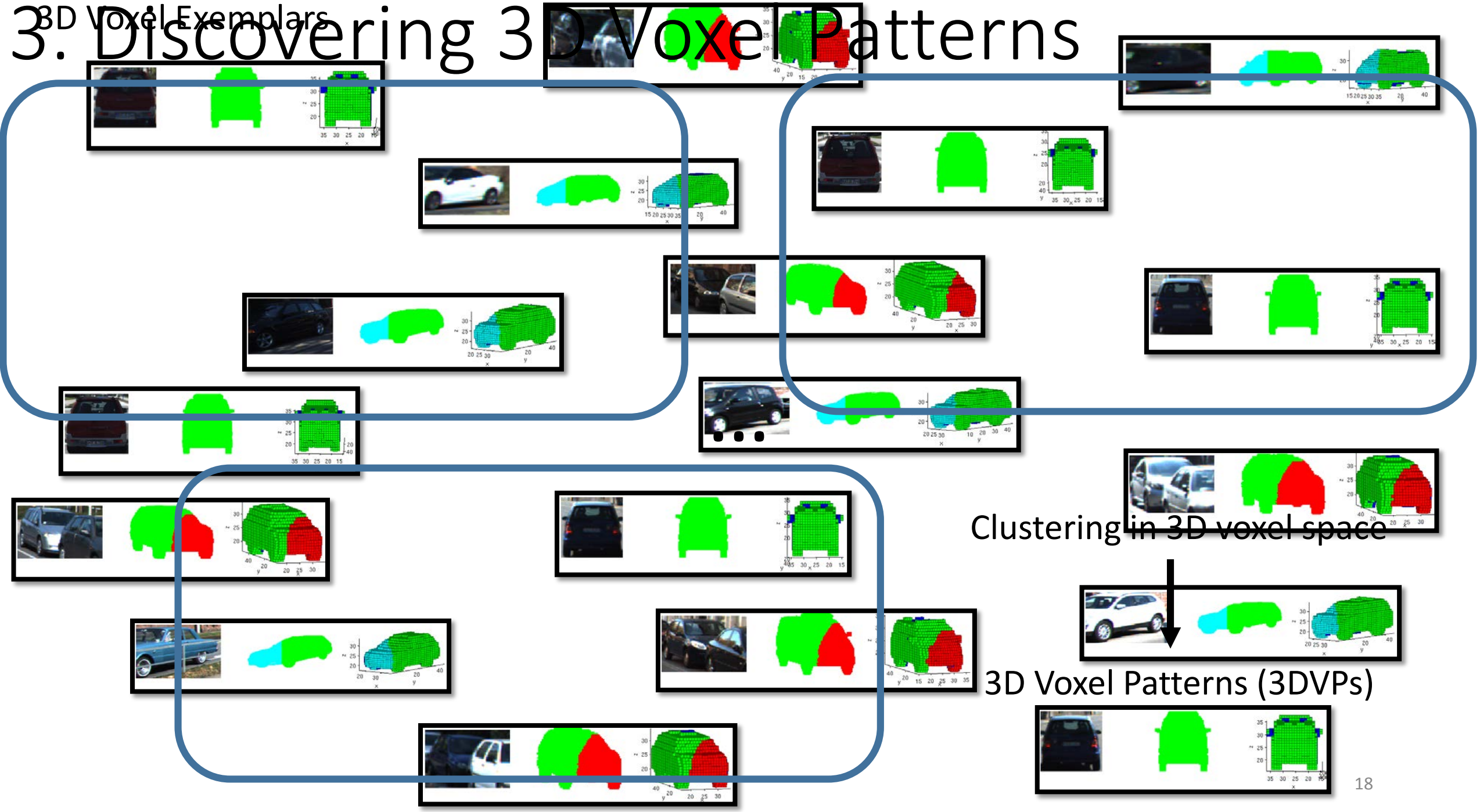
2. Building 3D Voxel Exemplars

A 3D voxel exemplar $E_i = (I_i, M_i, V_i)$



3D Voxel Exemplars

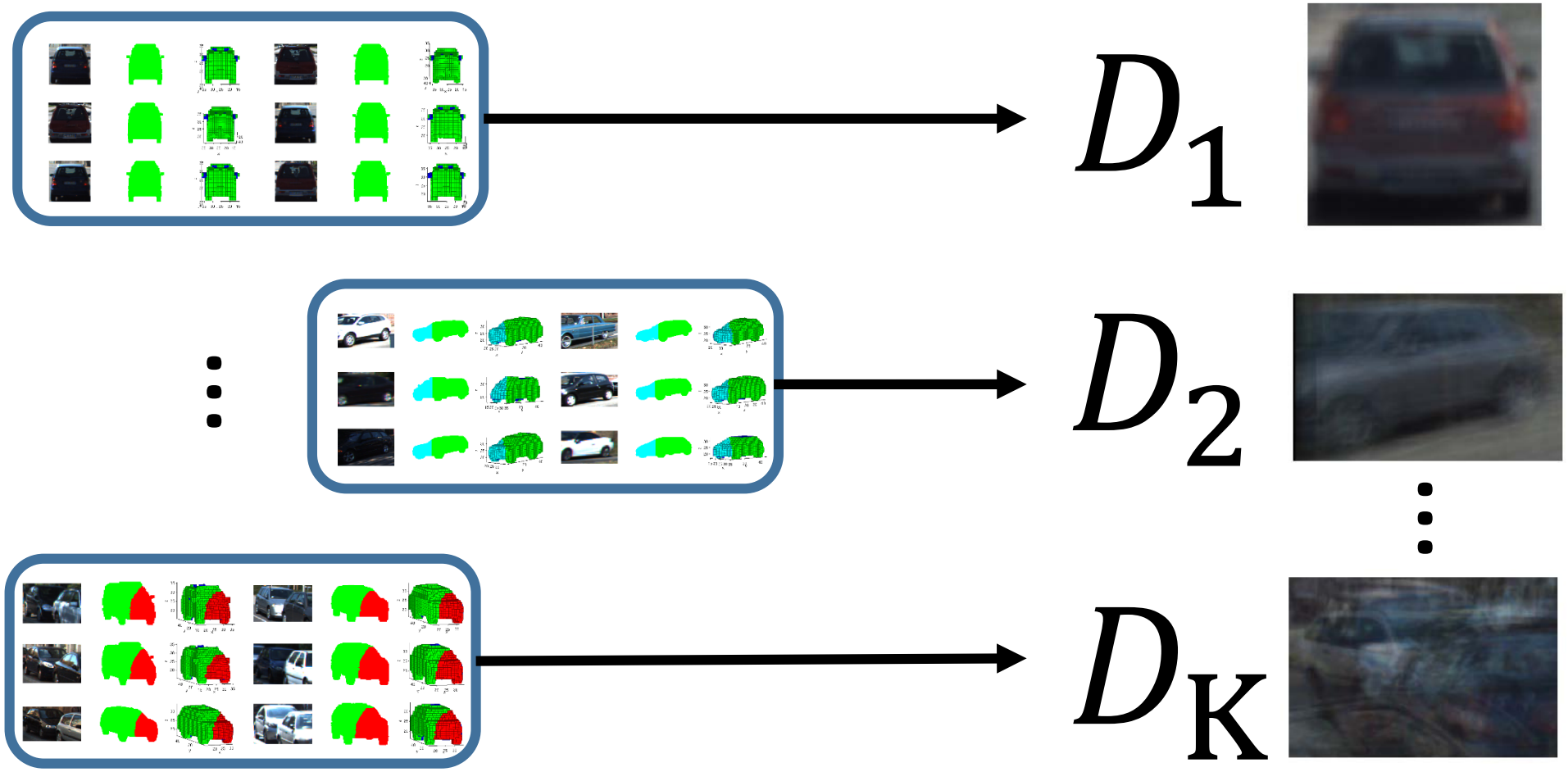
3. Discovering 3D Voxel Patterns



Clustering in 3D voxel space

3D Voxel Patterns (3DVPs)

4. Training 3D Voxel Pattern detectors



- Train a ACF detector for each 3DVP.

P. Dolla´r, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. TPAMI, 2014.

Outline

- Training Pipeline
- Testing Pipeline
- Experiments
- Conclusion

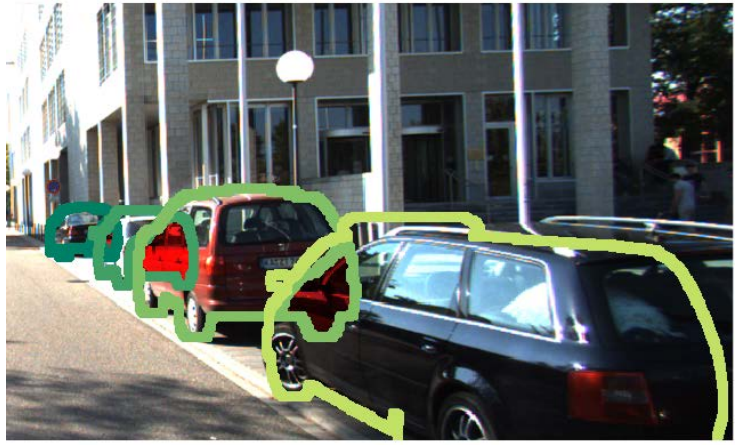
Testing Pipeline Overview

Input 2D image

1. Apply 3DVP detectors

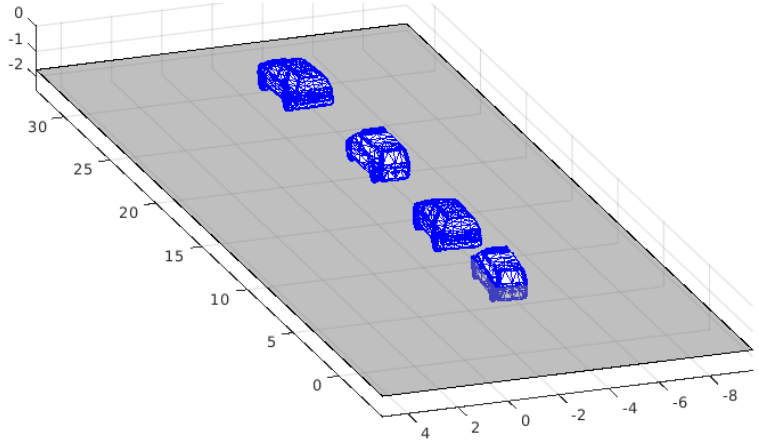
2D detection

2. Transfer meta-data
3. Occlusion reasoning



2D segmentation

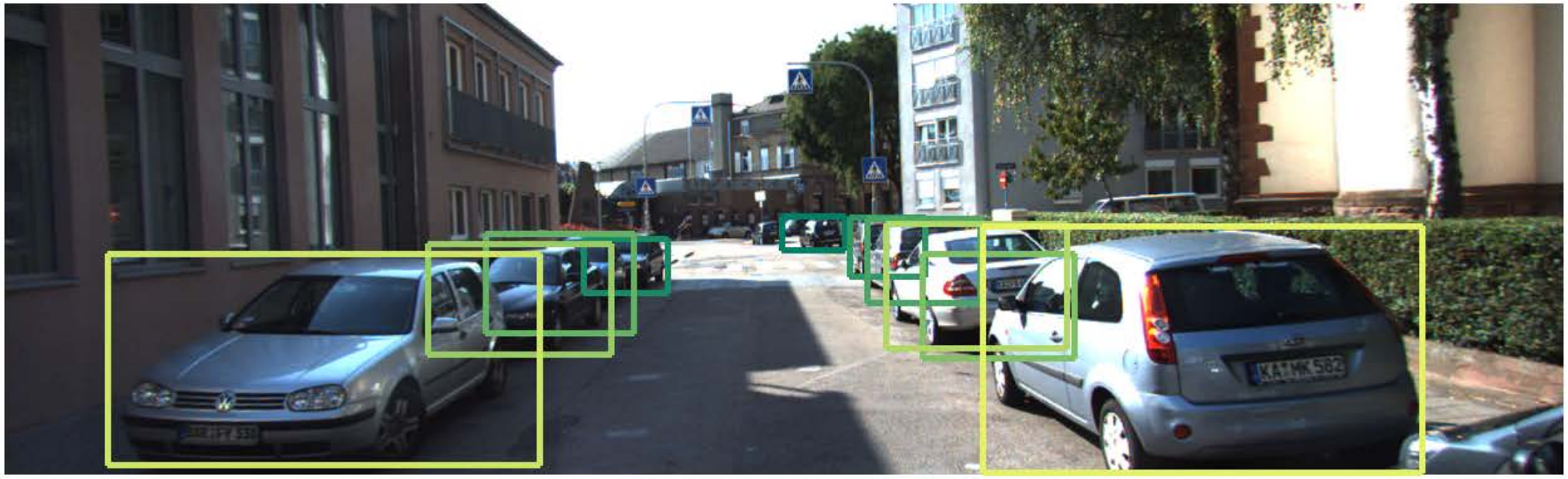
4. Backproject to 3D



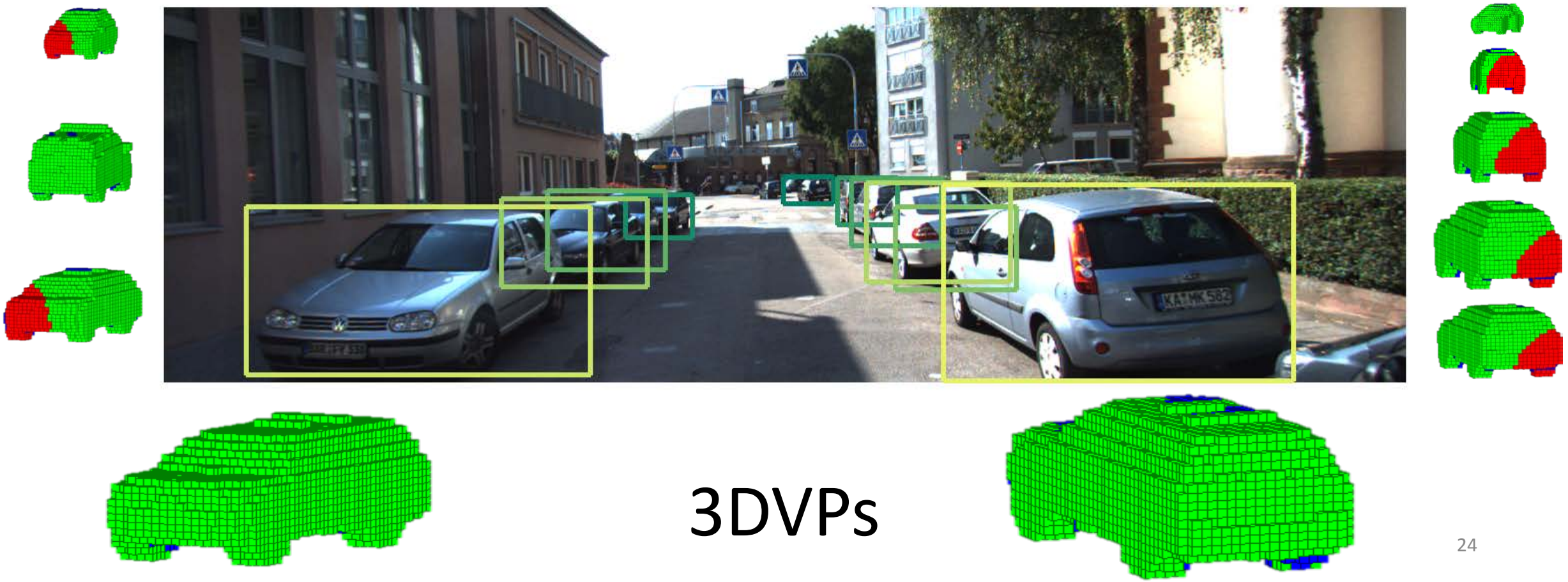
3D localization

1. Apply 3DVP Detectors

1. Apply 3DVP Detectors



2. Transfer Meta-Data



3DVPs

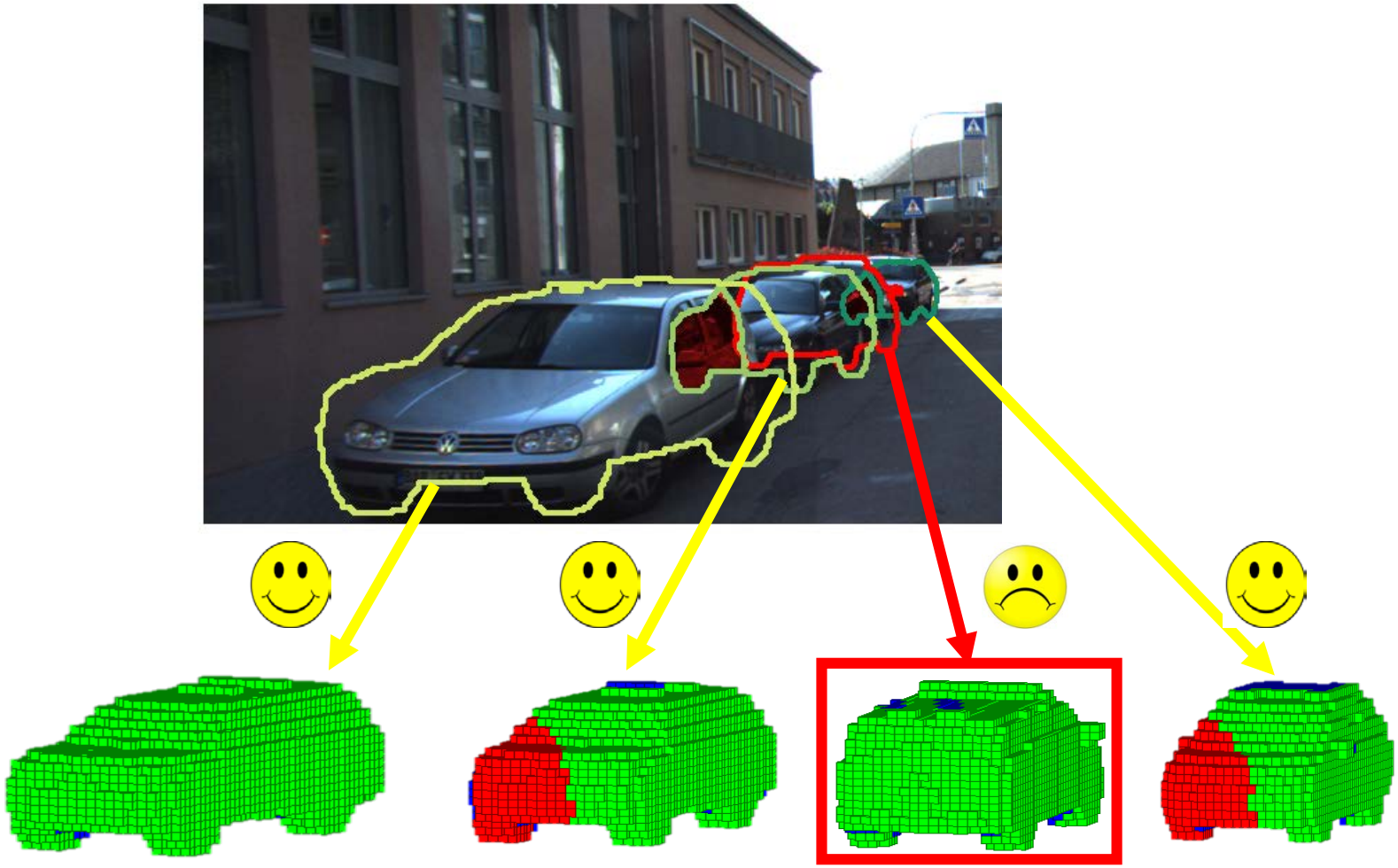
2. Transfer Meta-Data

3. Occlusion Reasoning

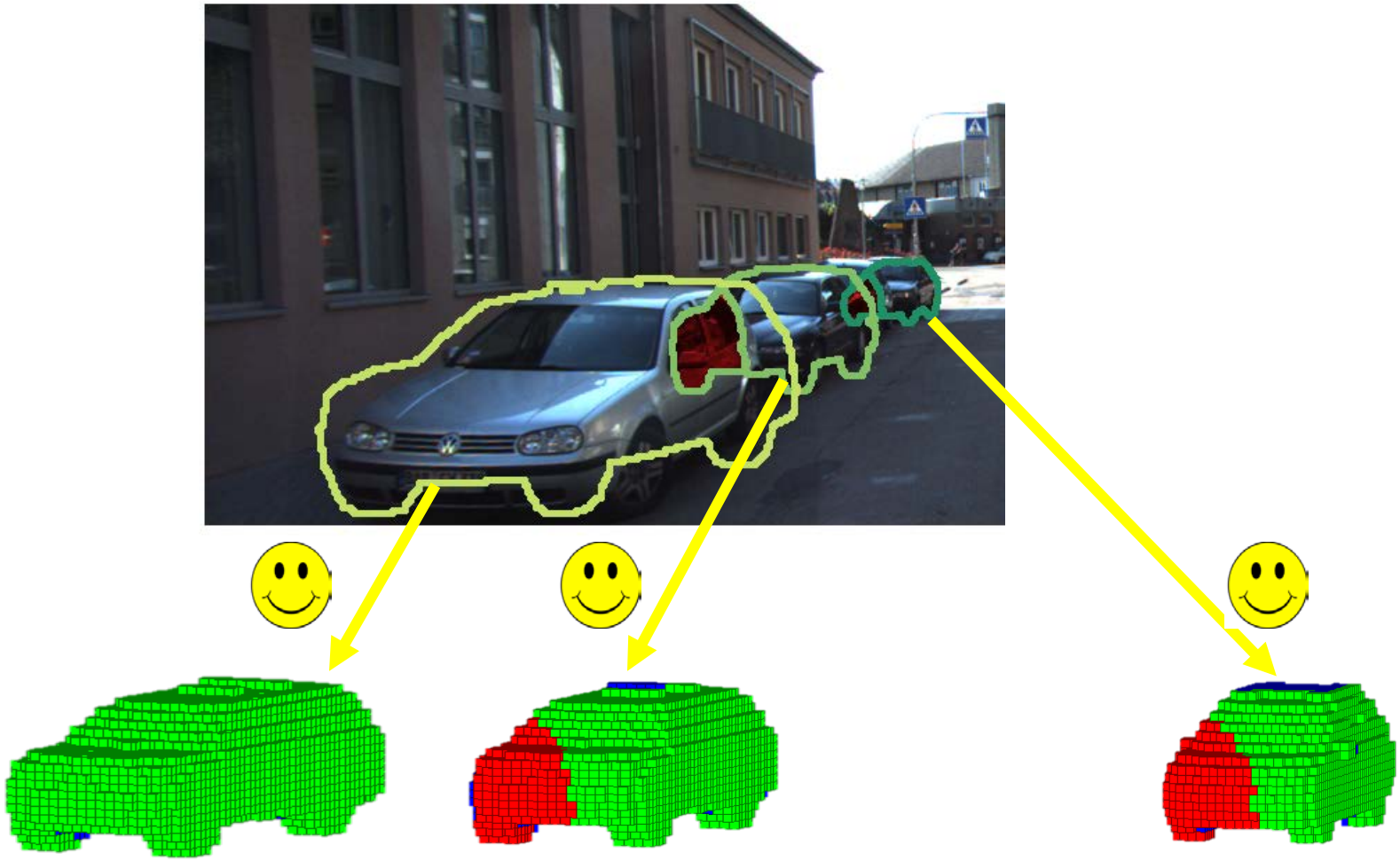
Occlusion reasoning: find a set of visibility-compatible detections

$$E = \sum_i (\psi_{\text{detection\_score}} + \psi_{\text{truncation}}) + \sum_{ij} \psi_{\text{occlusion}}$$

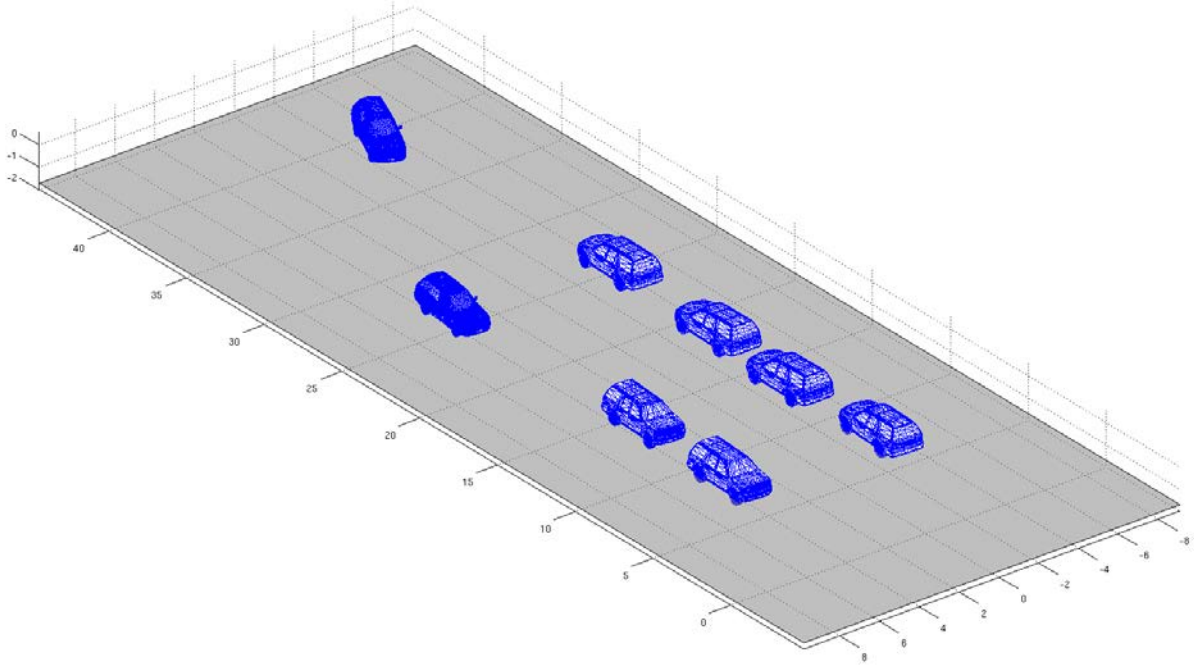
3. Occlusion Reasoning



3. Occlusion Reasoning



4. 3D Localization



Backprojection

Outline

- Training Pipeline
- Testing Pipeline
- Experiments
- Conclusion

Experiments: Datasets

- KITTI detection benchmark [1]
  - Autonomous driving scene
  - Test: 7,481 images for training (28,612 cars), 7,618 images for testing
  - Validation: 3,628 images for training, 3,799 images for testing

- Outdoor-scene dataset in [2]
  - Various scenarios: street, parking plot, free way, garage, etc.
  - 200 images for testing only
  - 659 cars with 235 occluded cars and 135 truncated cars

[1] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[2] Y. Xiang and S. Savarese. Object detection by 3d aspectlets and occlusion reasoning. In ICCVW, 2013.

Car Detection and Orientation Estimation on KITTI

	Object Detection (AP)			Object Detection and Orientation estimation (AOS)		
Method	**Easy**	**Moderate**	**Hard**	**Easy**	**Moderate**	**Hard**
ACF [1]	55.89	54.77	42.98	N/A	N/A	N/A
DPM [2]	71.19	62.16	48.43	67.27	55.77	43.59
DPM-VOC+VP [3]	74.95	64.71	48.76	72.28	61.84	46.54
OC-DPM [4]	74.94	65.95	53.86	73.50	64.42	52.40
SubCat [5]	81.94	66.32	51.10	80.92	64.94	50.03
AOG [6]	84.36	71.88	59.27	43.81	38.21	31.53
SubCat [7]	84.14	75.46	59.71	83.41	74.42	58.83
Regionlets [8]	84.75	**76.45**	59.70	N/A	N/A	N/A
Ours NMS	84.81	73.02	63.22	84.31	71.99	62.11
Ours Occlusion	**87.46**	75.77	**65.38**	**86.92**	**74.59**	**64.11**

[1] P. Dolla´r, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. TPAMI, 2014.
[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[3] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Multi-view and 3d deformable part models. TPAMI, 2015.
[4] B. Pepikj, M. Stark, P. Gehler, and B. Schiele. Occlusion patterns for object class detection. In CVPR, 2013.
[5] E. Ohn-Bar and M. M. Trivedi. Fast and robust object detection using visual subcategories. In CVPRW, 2014.
[6] B. Li, T. Wu, and S.-C. Zhu. Integrating context and occlusion for car detection by hierarchical and-or model. In ECCV, 2014.
[7] E. Ohn-Bar and M. M. Trivedi. Learning to detect vehicles by clustering appearance patterns. T-ITS, 2015.
[8] X.Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013.

Joint Car Detection and Segmentation on KITTI

Method	Easy	Moderate	Hard
DPM [1] + box	38.09	29.42	22.65
Ours NMS + box	57.52	47.84	40.01
Ours Occlusion + box	59.21	49.74	41.71
Ours NMS + 3DVP	63.88	52.57	43.82
Ours Occlusion + 3DVP	**65.73**	**54.60**	**45.62**

Evaluation on validation set

Metric: Average Segmentation Accuracy (ASA)

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.

Joint Car Detection and 3D Localization on KITTI

Method	Easy	Moderate	Hard
DPM [1] < 2m	40.21	29.02	22.36
Ours NMS < 2m	64.85	49.97	41.14
Ours Occlusion < 2m	**66.56**	**51.52**	**42.39**
DPM [1] < 1m	24.44	18.04	14.13
Ours NMS < 1m	44.47	33.25	26.93
Ours Occlusion < 1m	**45.61**	**34.28**	**27.72**

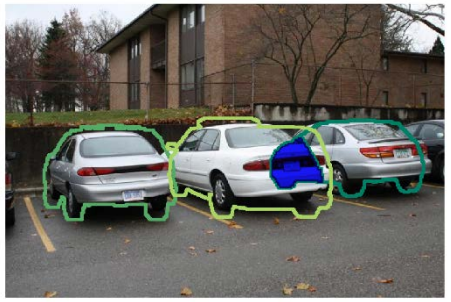
Evaluation on validation set

Metric: Average Localization Precision (ALP)

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.

Car Detection on the Outdoor-Scene Dataset

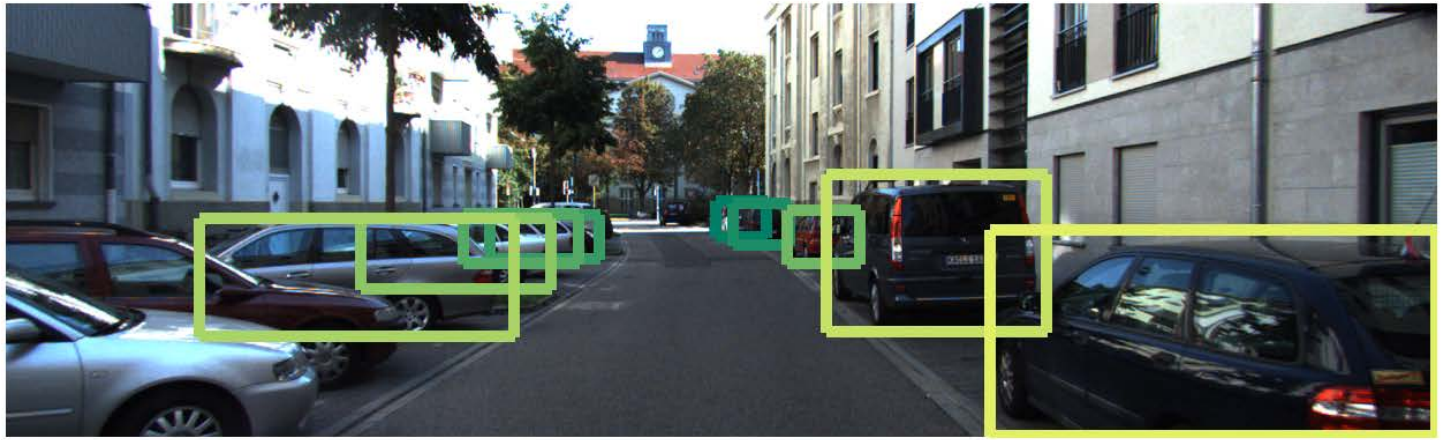
% occlusion	< 0.3	0.3 – 0.6	> 0.6
#images	66	68	66
ALM [1]	72.3	42.9	35.5
DPM [2]	75.9	58.6	44.6
SLM [3]	80.2	63.3	52.9
Ours NMS	89.7	76.3	55.9
Ours Occlusion	**90.0**	**76.5**	**62.1**

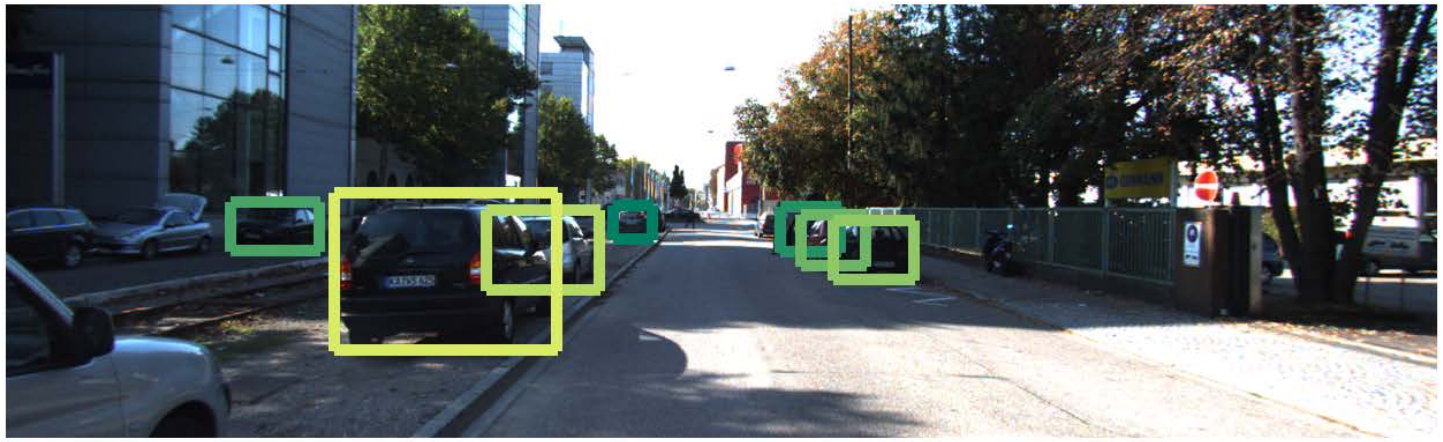



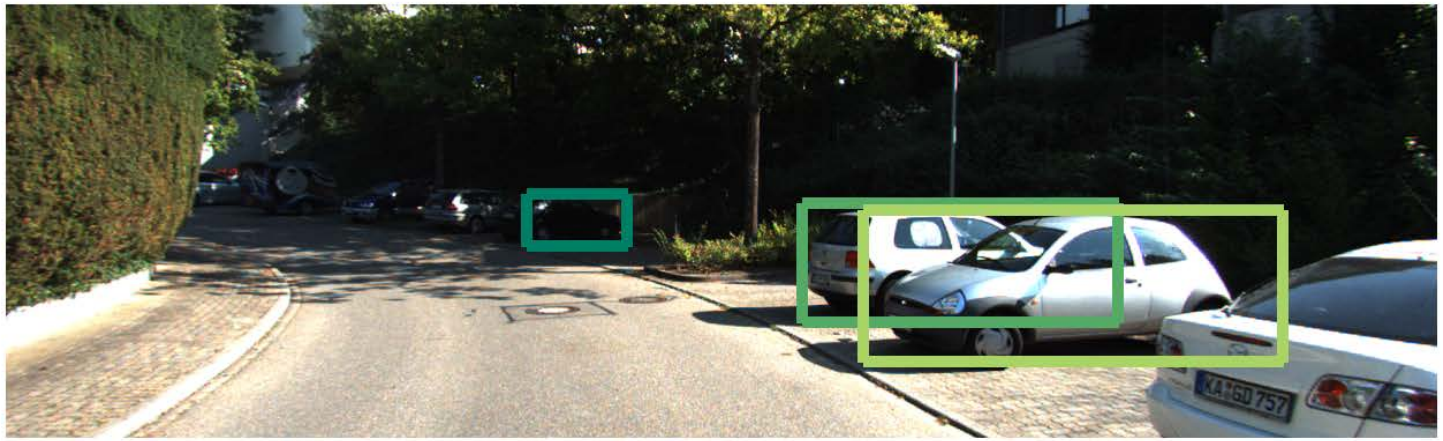

[1] Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.
[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[3] Y. Xiang and S. Savarese. Object detection by 3d aspectlets and occlusion reasoning. In ICCVW, 2013.

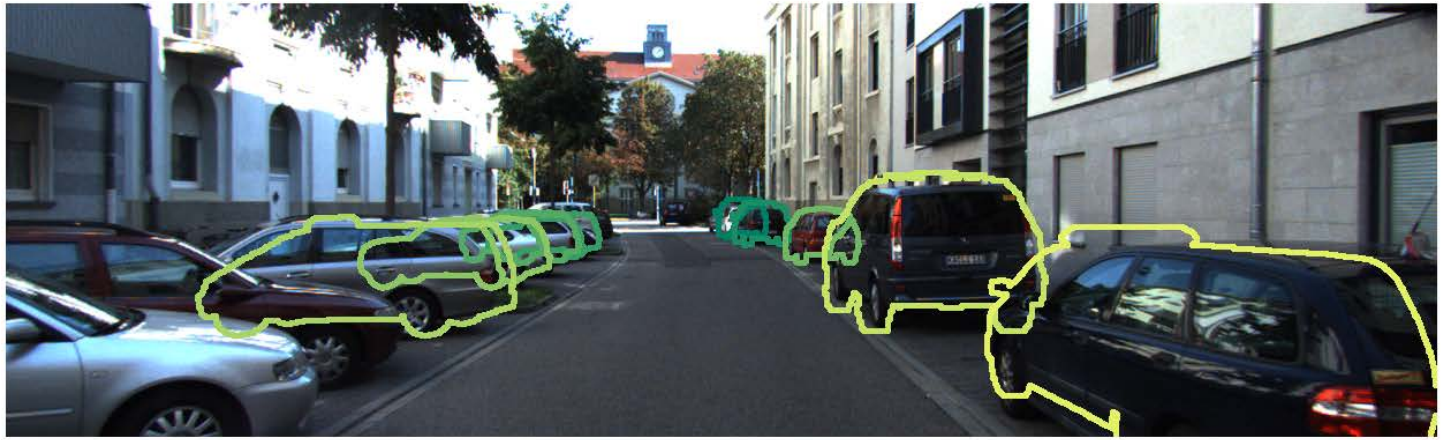
Anecdotal Results on KITTI

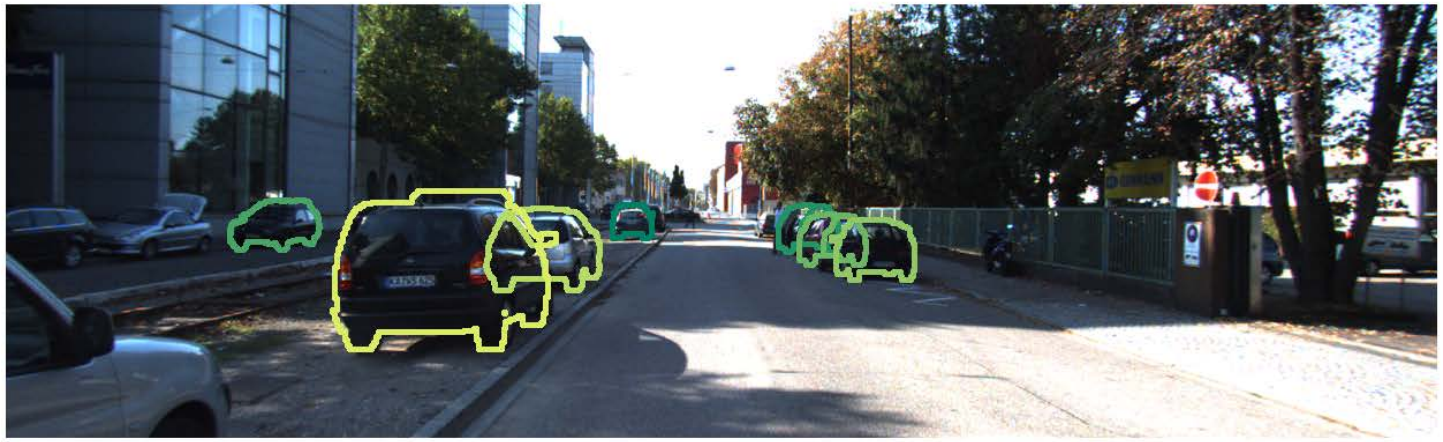

KA 160 757

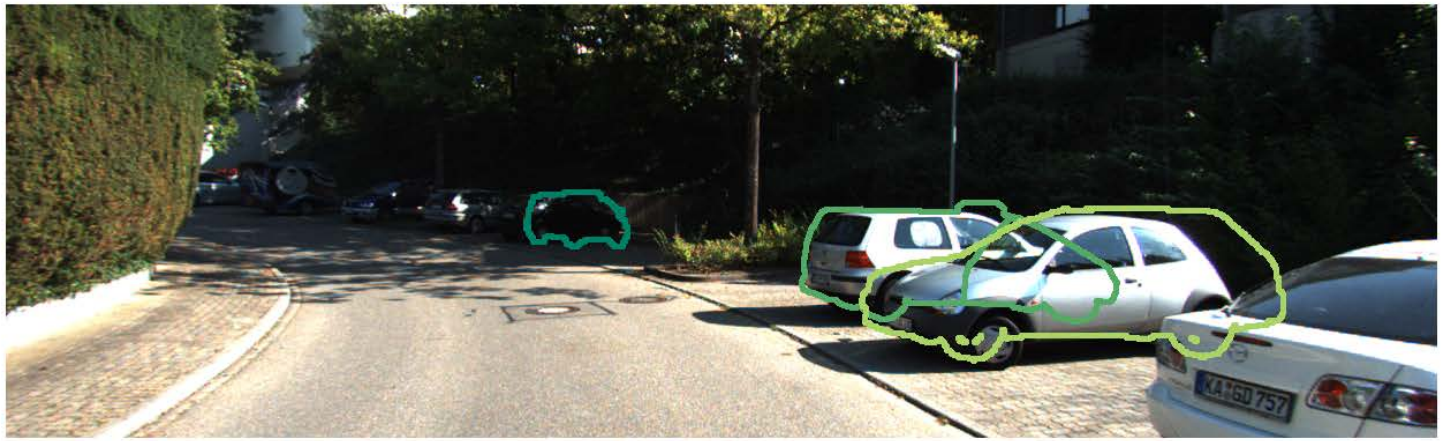


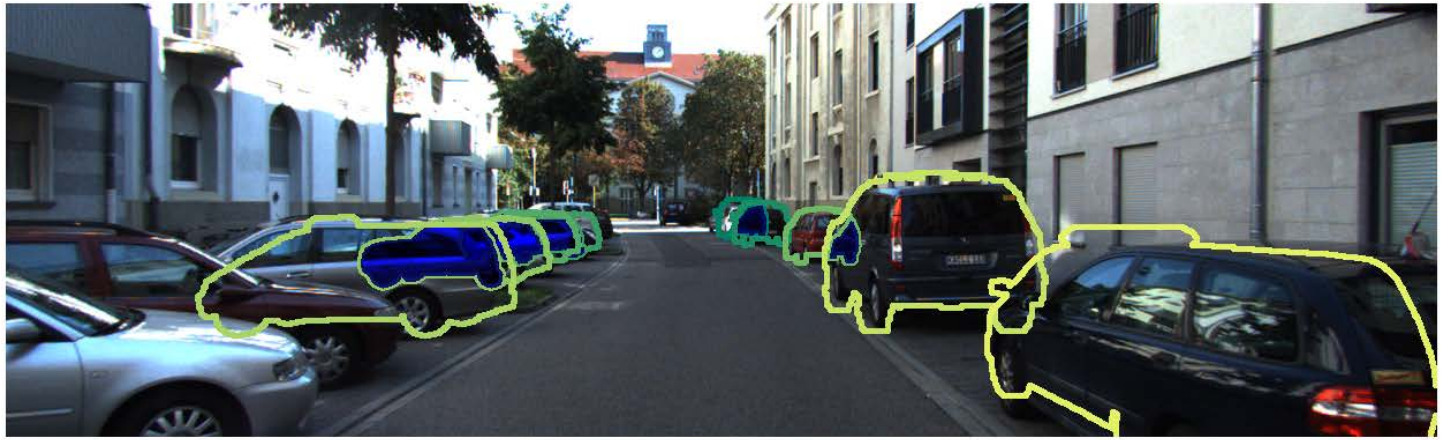


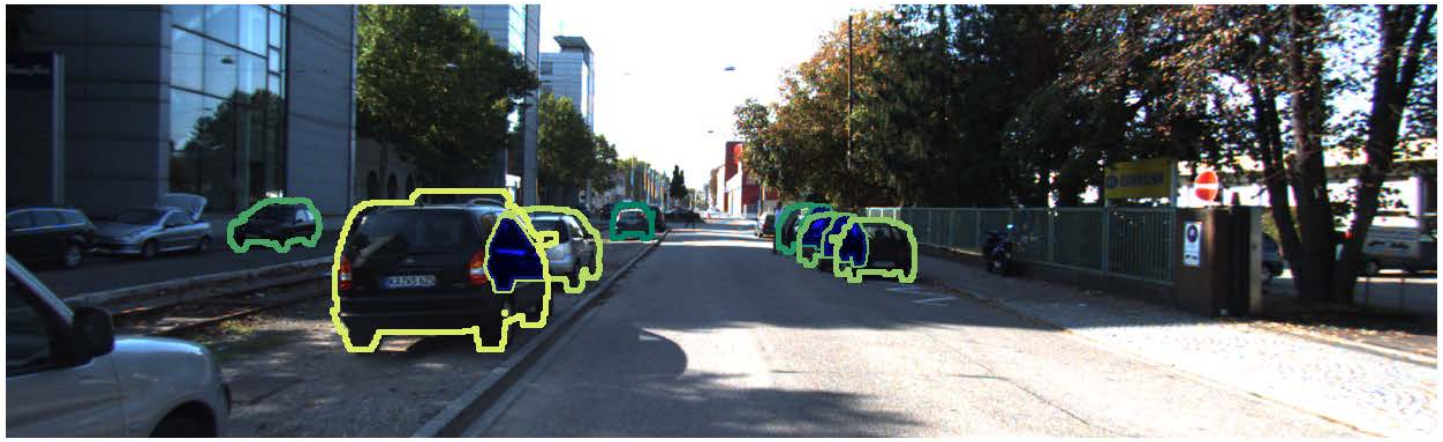


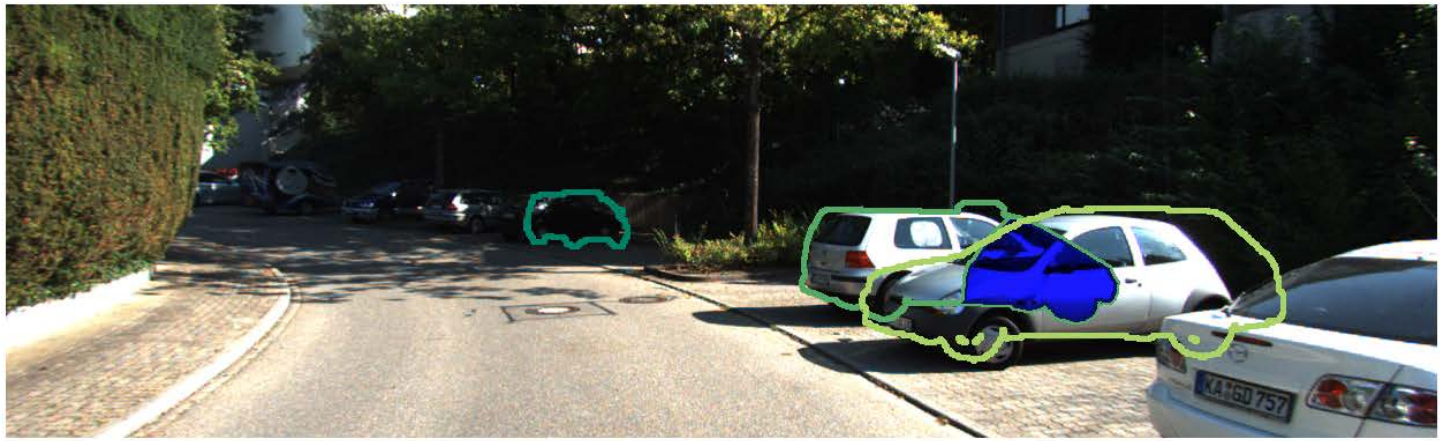


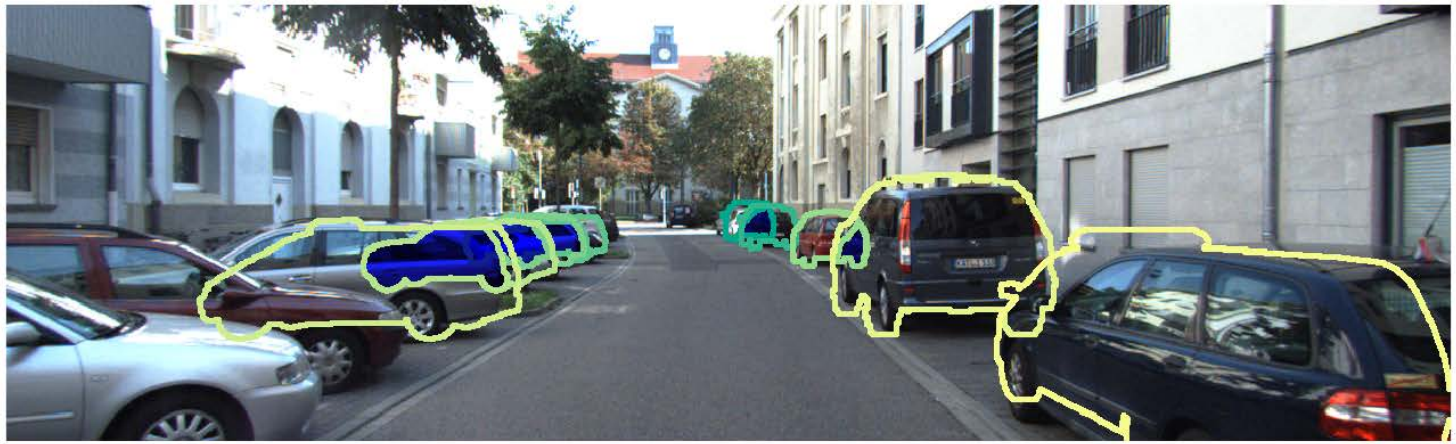


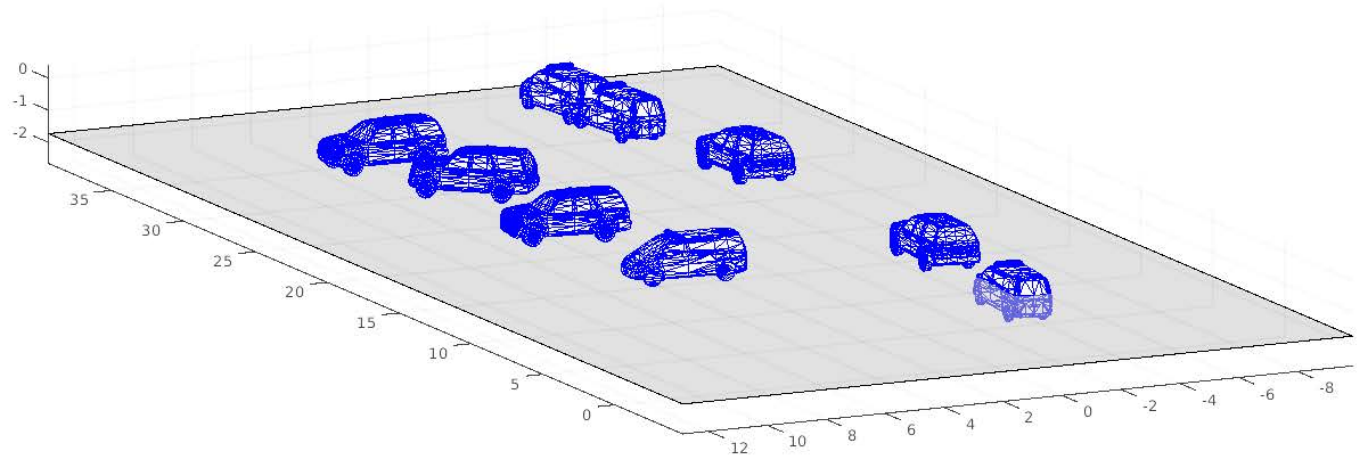


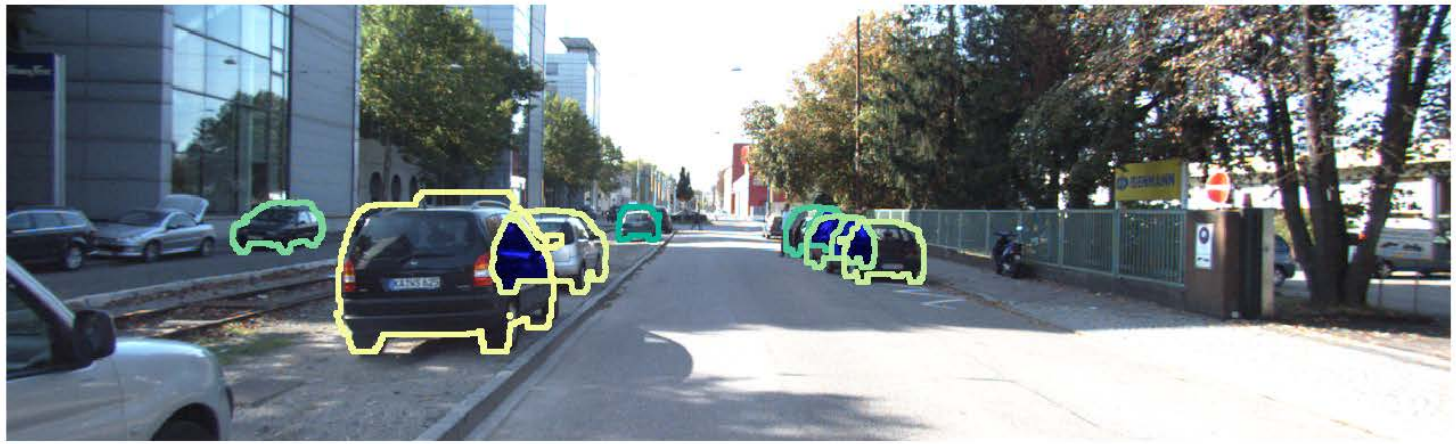


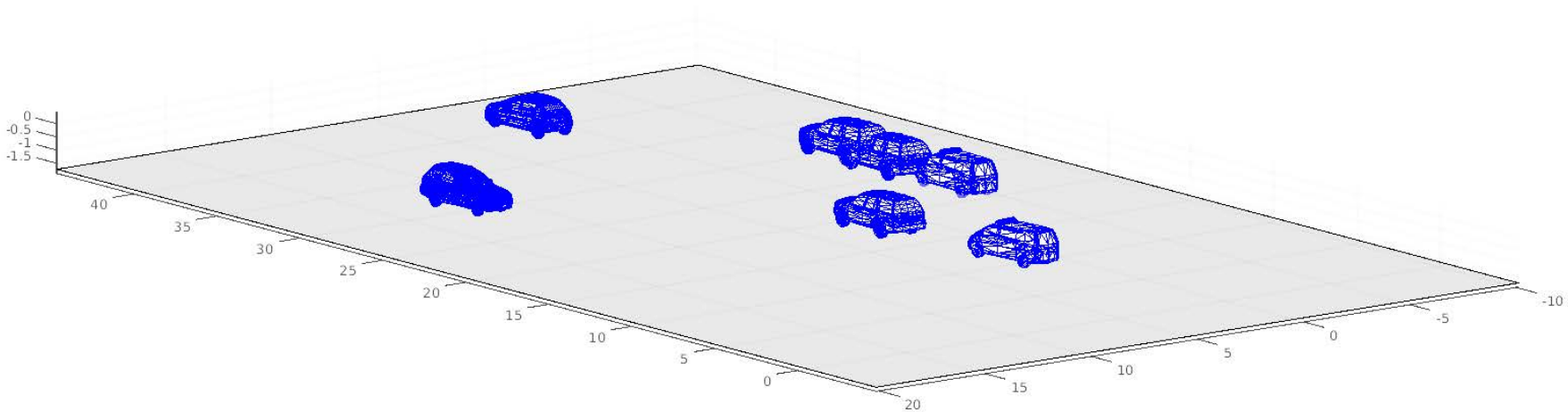


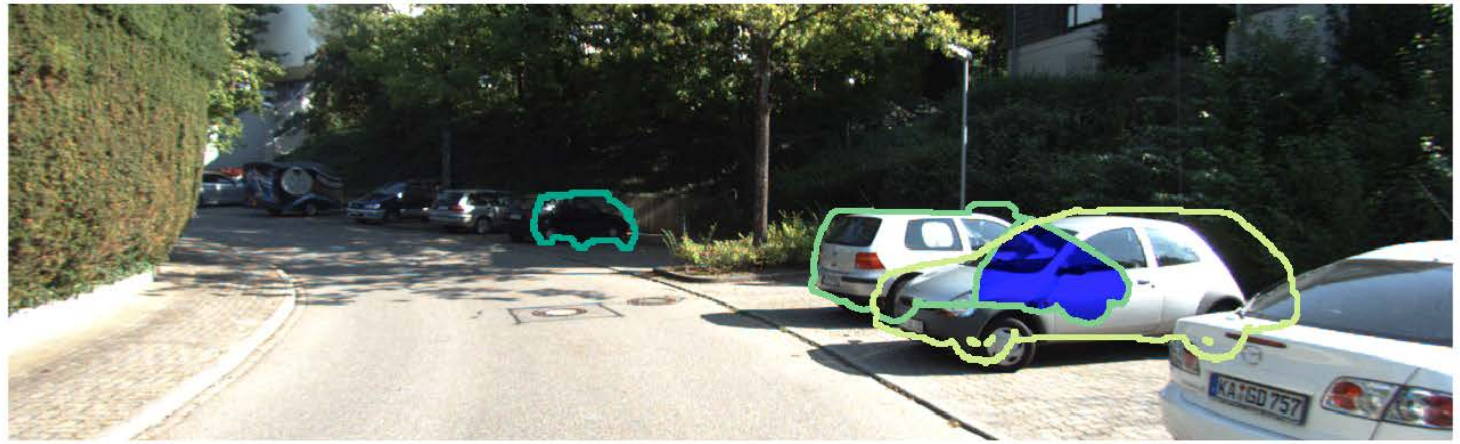


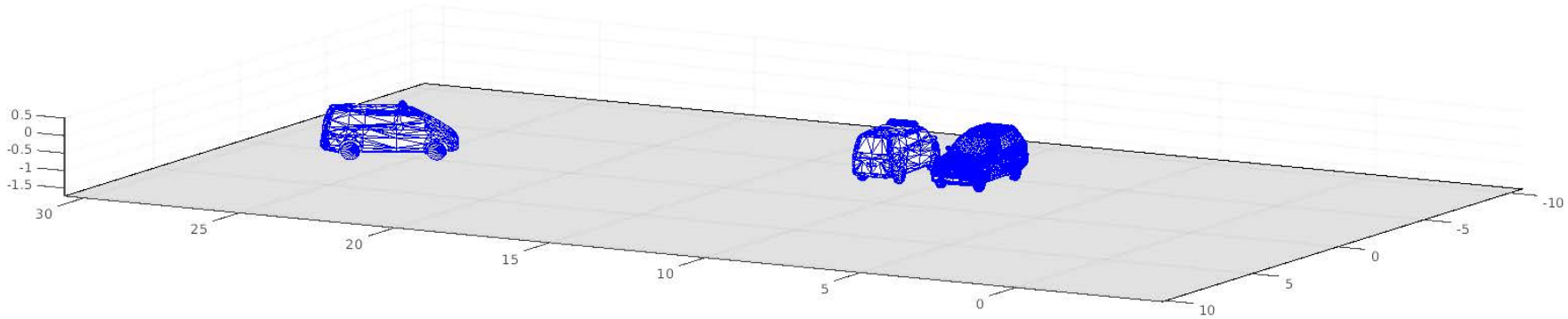



0
-1
-2
35
30
25
20
15
10
5
0
12
10
8
6
4
2
0
-2
-4
-6
-8




0
-0.5
-1
-1.5
40
35
30
25
20
15
10
5
0
20
15
10
5
0
-5
-10



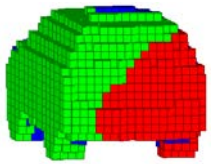

0.5
0
-0.5
-1
-1.5
30
25
20
15
10
5
0
10
5
0
-5
-10

Outline

- Training Pipeline
- Testing Pipeline
- Experiments
- Conclusion

Conclusion

- A novel 3D object representation: 3D Voxel Pattern (3DVP)



- 3DVP handles 3D pose, occlusion and truncation jointly
- A contextual model to reason about occlusions between objects
- The idea of 3DVP is applicable to generic rigid object categories

Acknowledgements

Thank you!