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.

yuxiang@umich.edu, {wongun, ylin}@nec-labs.com, ssilvio@stanford.edu

CV GL Computational Vision & Geometry Lab

NEC Laboratories America
Relentless passion for innovation

Object Recognition with 3D Voxel Patterns

❑ **Input**: a single 2D image
❑ **Output**:
(1) 2D location and 2D segmentation of objects
(2) 3D location, 3D pose and 3D shape of objects
(3) Occlusion relationship between objects

2D recognition

3D localization

3D voxel patterns
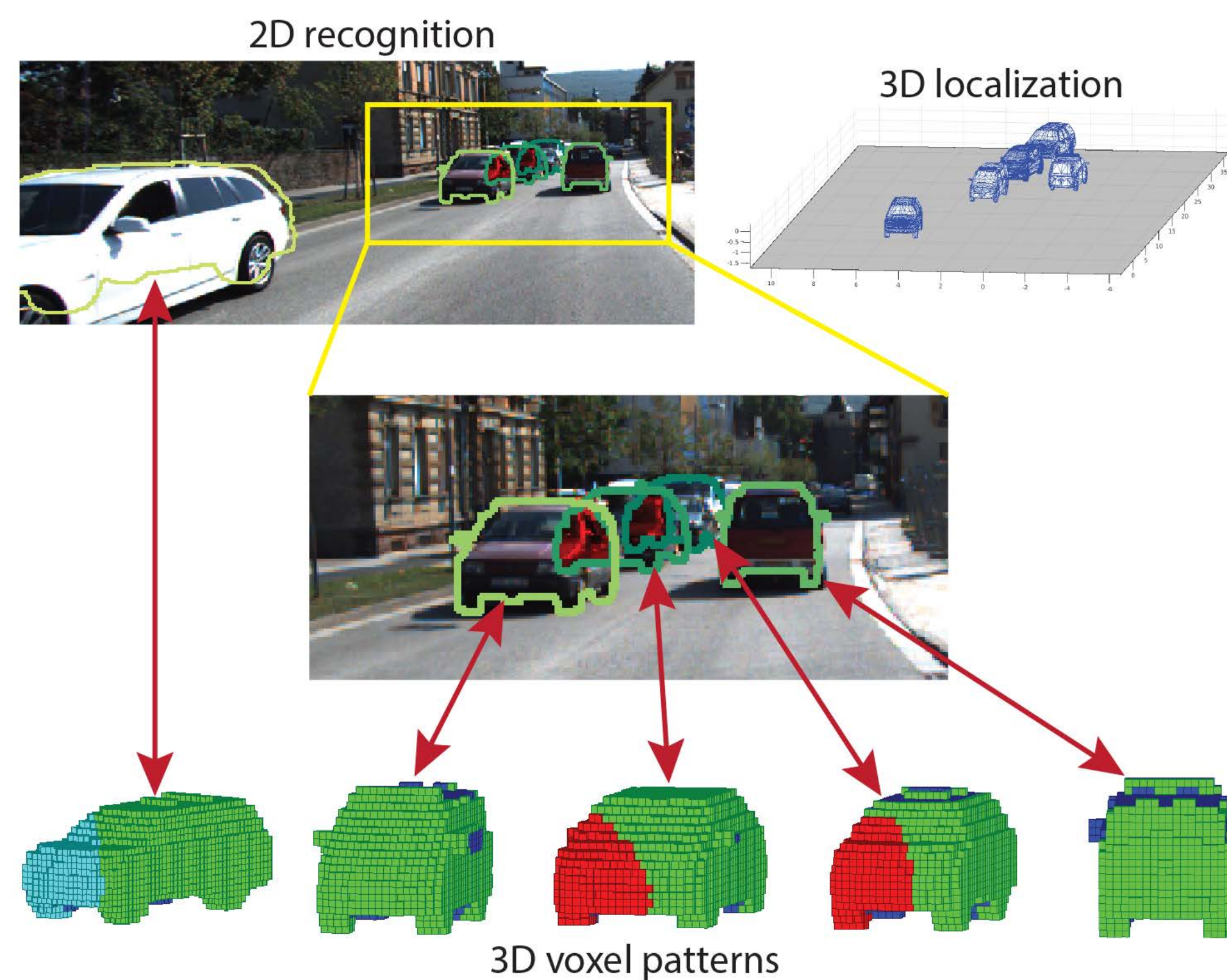
❑ **Applications**: autonomous driving, robotics, augmented reality, etc.

Challenges in Object Recognition

Part-based model
Felzenszwalb et al., TPAMI'10

Subcategory
Divvala et al., ECCVW'12

Nearest neighbor
Malisiewicz et al., ICCV'11

Structured regression
Vedaldi & Zisserman, NIPS'09

Shape variation

Truncation

2.5D representation
Su et al., ICCV'09

3D representation
Xiang et al., CVPR'12

2D occlusion patterns
Pepik et al. CVPR'13

2D Occlusion masks
Zia et al., CVPR'14

Viewpoint variation

Occlusion

Our Object Recognition Framework

❑ Overview

Align 2D images with 3D CAD models — 3D voxel exemplars — 3D voxel patterns — Train 3D voxel pattern detectors

(a) Training

Input 2D image — Apply 3DVP detectors — 2D detection — Transfer meta-data and occlusion reasoning — 2D segmentation — Backproject to 3D — 3D localization

(b) Testing
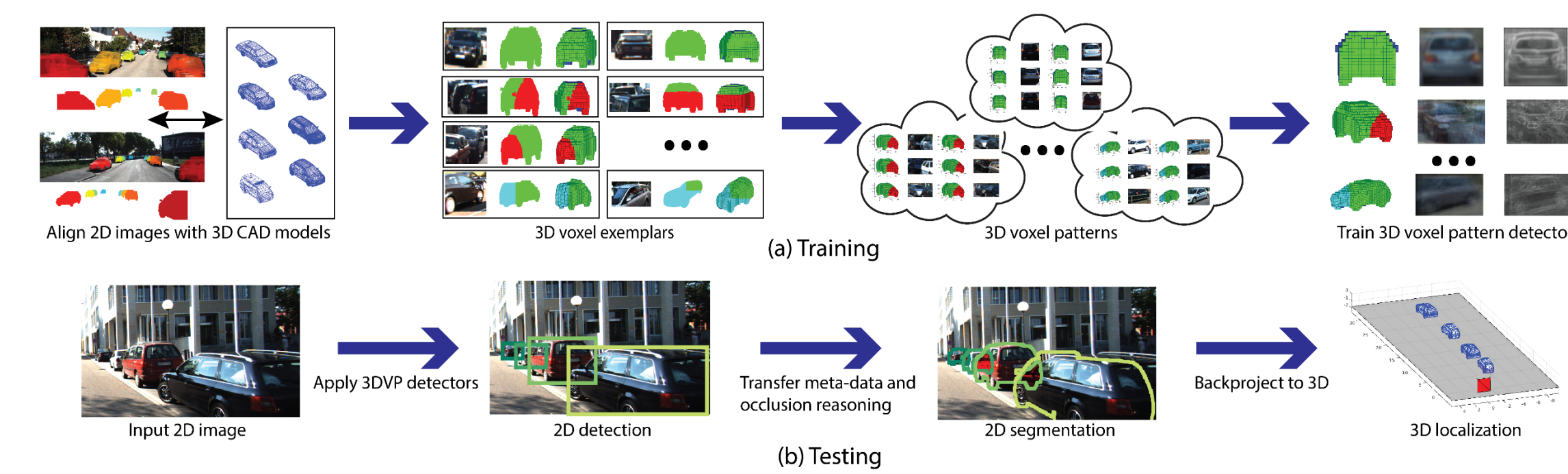
❑ 3D Voxel Exemplars from Data

3D annotations from KITTI [1]

(a) 3D CAD model association and registration   (b) project 3D CAD models to the image   (c) Label 2D segmentation mask and 3D voxel model   (d) 3D voxel exemplar
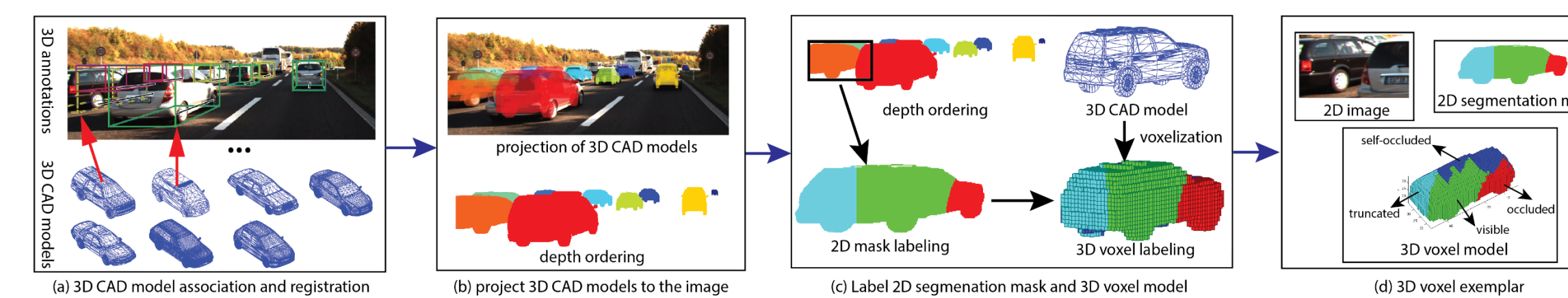
❑ Discovering 3D Voxel Patterns

Clustering in 3D voxel space

$$\mathcal{S} = \{0,1,2,3,4\}$$
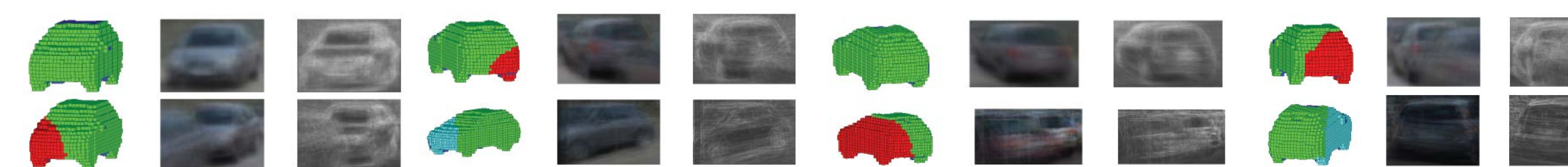$$s(\mathbf{x_1},\mathbf{x_2}) = \frac{|\mathcal{S}|}{N^3}\sum_{i=1}^{N^3} 1(x_1^i = x_2^i)\cdot w(x_1^i)$$
$$\text{s.t., } \sum_{i=0}^{|\mathcal{S}|-1} w(i) = 1$$
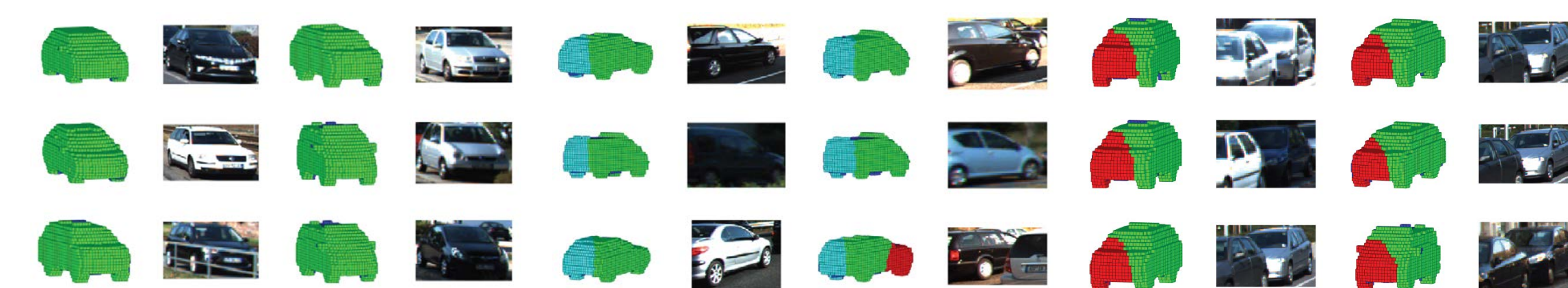
❑ Learning 3DVP Detectors

Aggregated Channel Features (ACF) [2]

❑ Occlusion Reasoning with 3DVPs

Find a set of visibility-compatible detections

$$E(\hat{\mathbb{D}}) = \sum_{i\in\hat{\mathbb{D}}}\left(\underbrace{w_d(s_i - b)}_{\text{detection score}} - w_o\underbrace{\frac{|m_o^i| + |m_t^i|}{|m_i|}}_{\text{invisibility penalty}} + w_o\underbrace{\frac{|m_t^i \not\subset I|}{|m_i|}}_{\text{truncation explained}}\right) + \sum_{i,j\in\hat{\mathbb{D}},i\neq j}\left(w_o\underbrace{\frac{|m_{far(i,j)}^o \cap m_{near(i,j)}^v|}{|m_{far(i,j)}|}}_{\text{occlusion explained}} - w_p\underbrace{\frac{\sum_{k=v,o,t}|m_i^k \cap m_j^k|}{\min(|m_i|,|m_j|)}}_{\text{overlap penalty}}\right)$$

References

[1] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[2] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. TPAMI, 2014.
[3] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[4] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Multi-view and 3d deformable part models. TPAMI, 2015.
[5] B. Pepikj, M. Stark, P. Gehler, and B. Schiele. Occlusion patterns for object class detection. In CVPR, 2013.
[6] E. Ohn-Bar and M. M. Trivedi. Fast and robust object detection using visual subcategories. In CVPRW, 2014.
[7] B. Li, T. Wu, and S.-C. Zhu. Integrating context and occlusion for car detection by hierarchical and-or model. In ECCV, 2014.
[8] E. Ohn-Bar and M. M. Trivedi. Learning to detect vehicles by clustering appearance patterns. T-ITS, 2015.
[9] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013.

Experiments

❑ Car Detection and Orientation Estimation on KITTI

Method	Object Detection (AP)			Orientation Estimation (AOS)		
	Easy	Moderate	Hard	Easy	Moderate	Hard
ACF [2]	55.89	54.77	42.98	N/A	N/A	N/A
DPM [3]	71.19	62.16	48.43	67.27	55.77	43.59
DPM-VOC+VP [4]	74.59	64.71	48.76	72.28	61.84	46.54
OC-DPM [5]	74.94	65.95	53.86	73.50	64.42	52.40
SubCat [6]	81.94	66.32	51.10	80.92	64.94	50.03
AOG [7]	84.36	71.88	59.27	43.81	38.21	31.53
SubCat [8]	84.14	75.46	59.71	83.41	74.42	58.83
Regionlets [9]	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**

Evaluation on test set
Metrics: Average Precision (AP) and Average Orientation Similarity (AOS)

❑ Joint Car Detection and Segmentation on KITTI

Method	Easy	Moderate	Hard
DPM [3] + 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)

❑ Joint Car Detection and 3D Localization on KITTI

Method	Easy	Moderate	Hard
DPM [3] < 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 [3] < 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)

2D recognition color code: true positive / false alarm / missed ground truth
3D localization color code: detection / ground truth / camera

DPM   NMS (Ours)   Occlusion Reasoning (Ours)