3D Object Representations for Recognition

Yu Xiang
Computational Vision and Geometry Lab
Stanford University

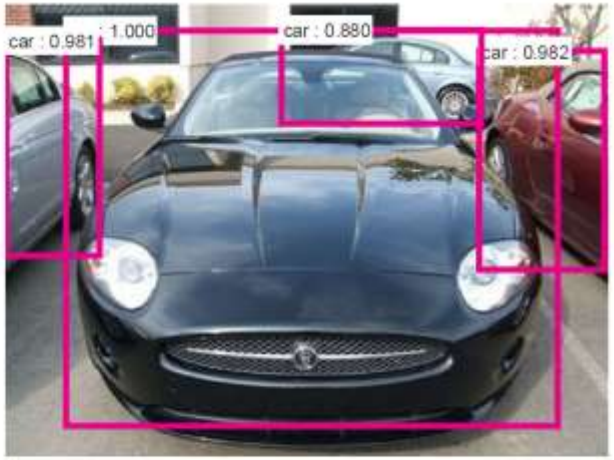
2D Object Recognition

Ordonez et al. ICCV13

horse
pasture
field
cow
fence

Image classification/tagging/annotation

Ren et al. NIPS15



Object detection

Long et al. CVPR15

Object segmentation

Karpathy et al. CVPR15

"man in black shirt is playing guitar."

Image description generation

Applications of 2D Object Recognition

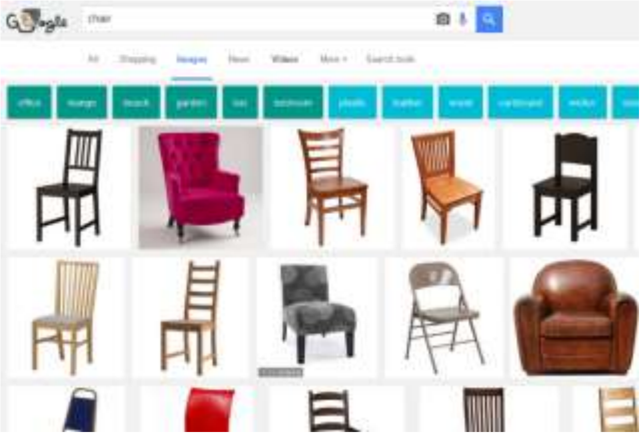


Image search/indexing

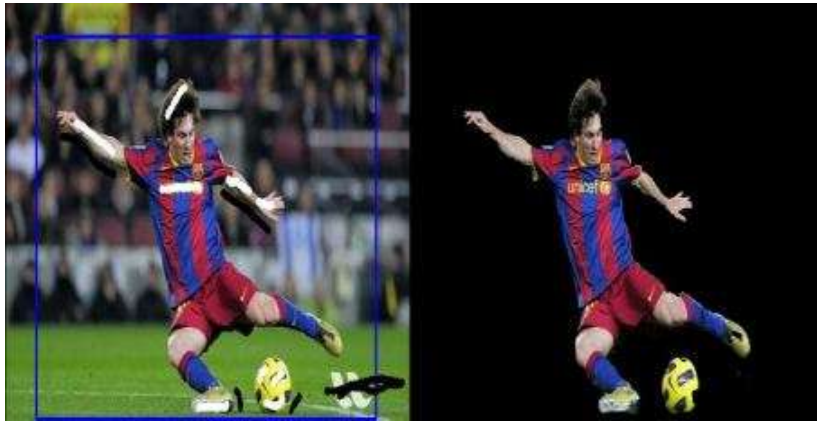


Photo editing

Visual surveillance

Biometrics authentication

These are all great, but…

2D recognition is NOT enough!

Applications that need 3D Object Recognition

Autonomous Driving

Robotics

Any application that requires interaction with the 3D world!

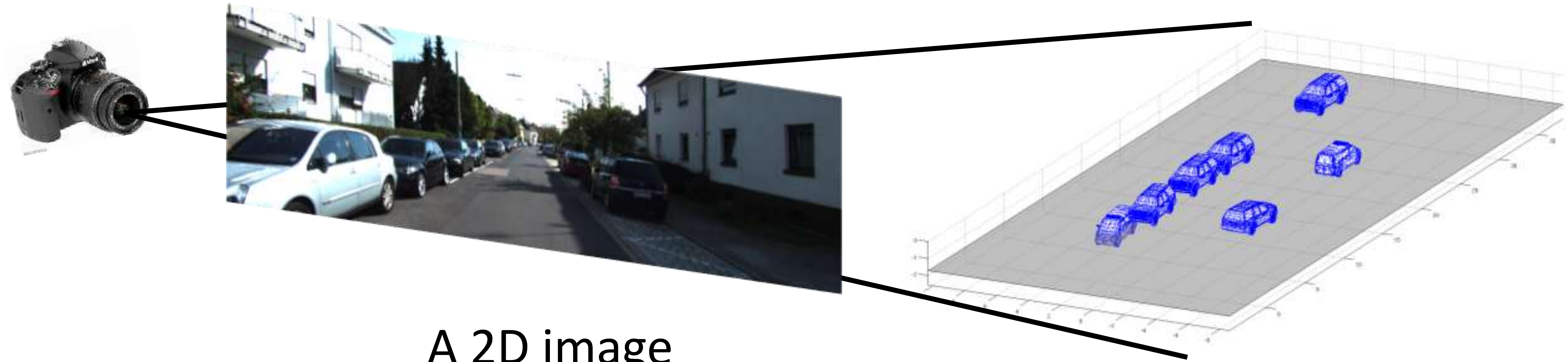
Augmented Reality

Gaming

Goal: Infer the 3D World

- Interaction
- Control
- Decision making
- Navigation

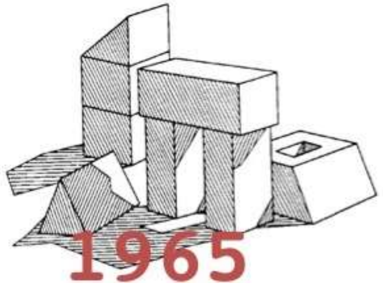
Etc.



A 2D image

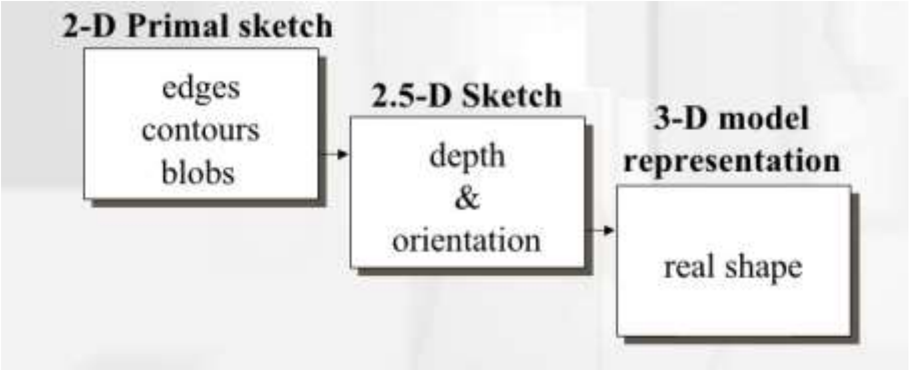
The 3D world

Goal: Infer the 3D World

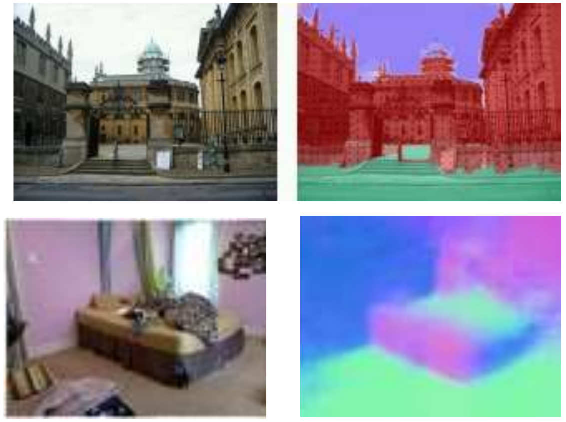


Blocks World
Larry Roberts, 1965

Blocks World Revisited
Gupta et al., ECCV'10

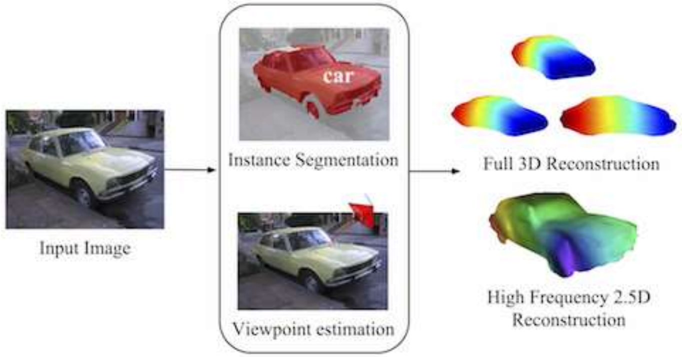


Marr's Theory
David Marr, 1978



Surface Layout/Normal Estimation
Hoiem et al., ICCV'05
Fouhey et al. ICCV'13, ECCV'14

Room Layout Estimation
Lee et al. CVPR'09
Hedau, el al., ICCV'09
Mallya & Lazebnik, ICCV'15

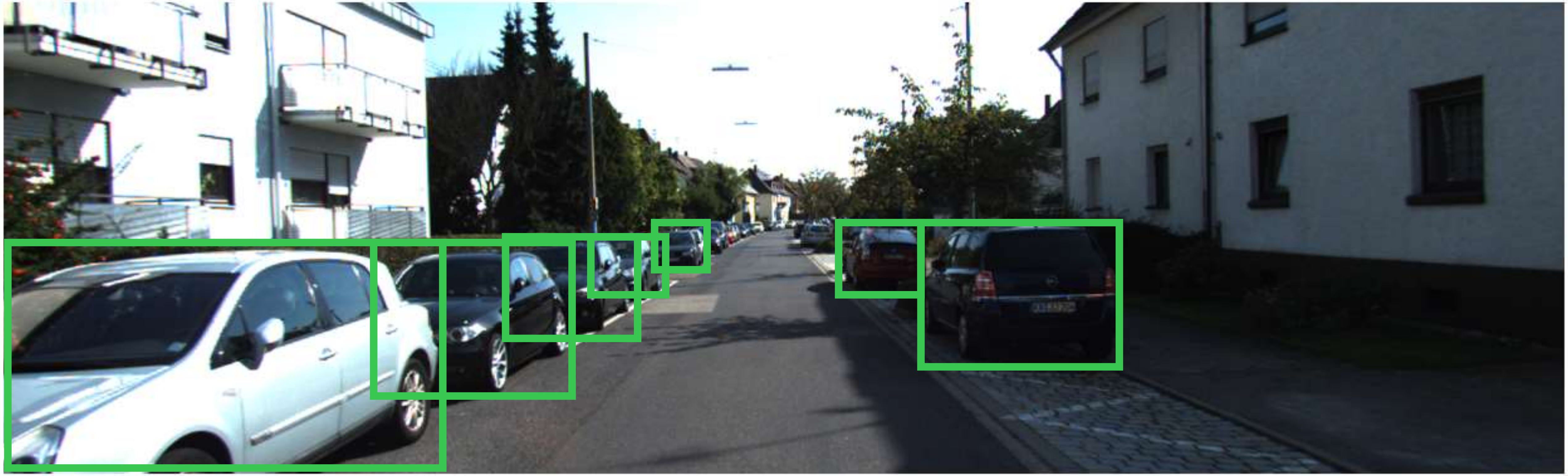


3D Object Recognition
Kar et al., ICCV'15
Tulsiani & Malik, CVPR'15

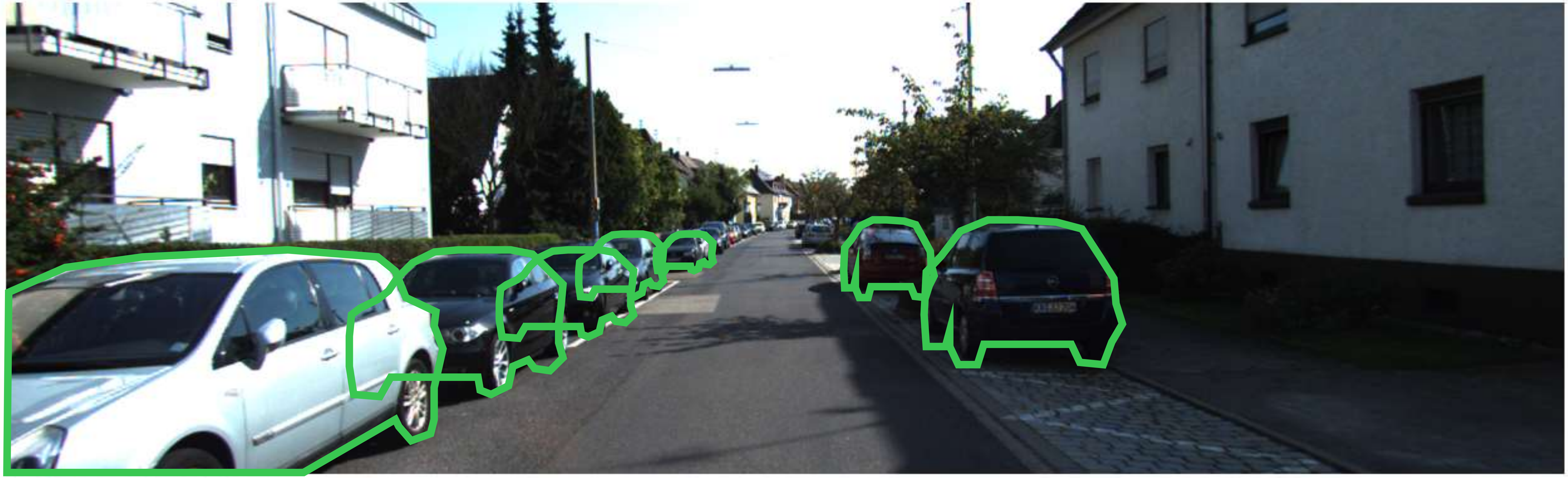
My Work: 2D Object Detection

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

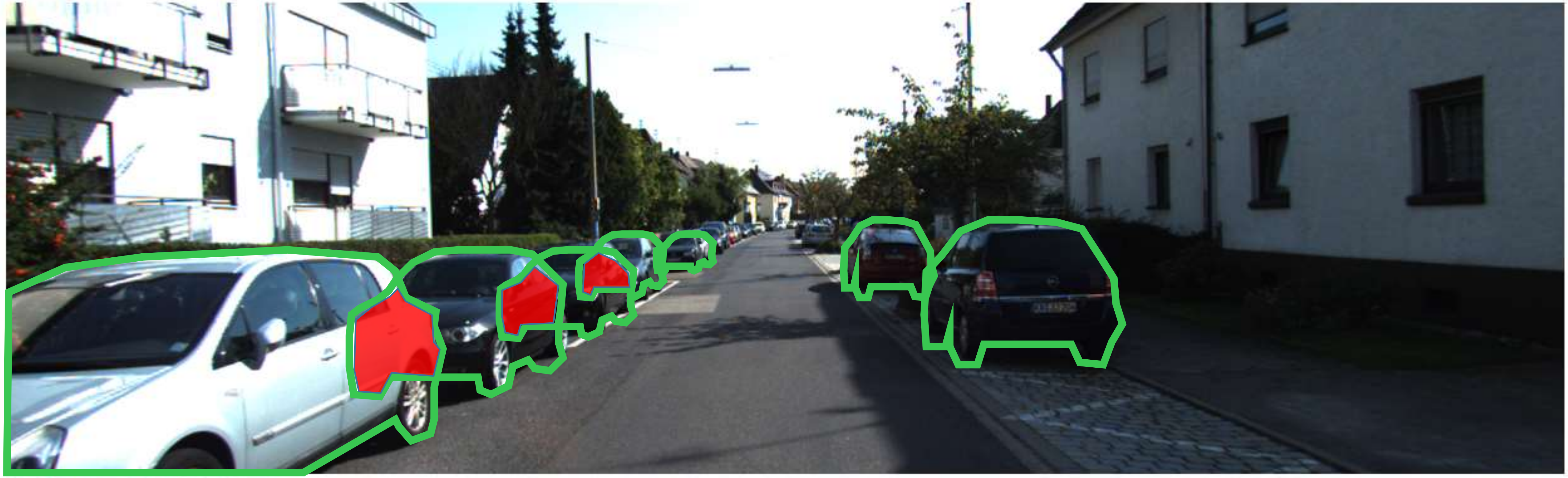
My Work: 2D Object Detection



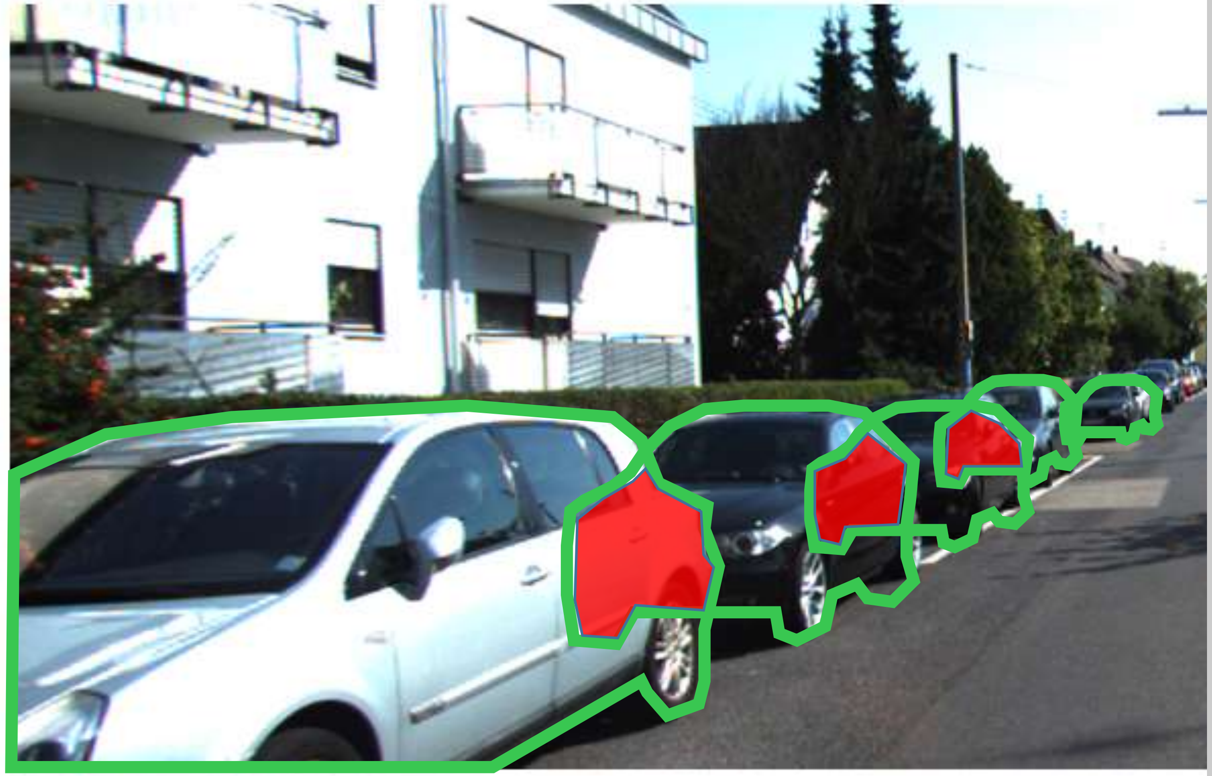
My Work: 2D Segmentation and 3D Pose Estimation

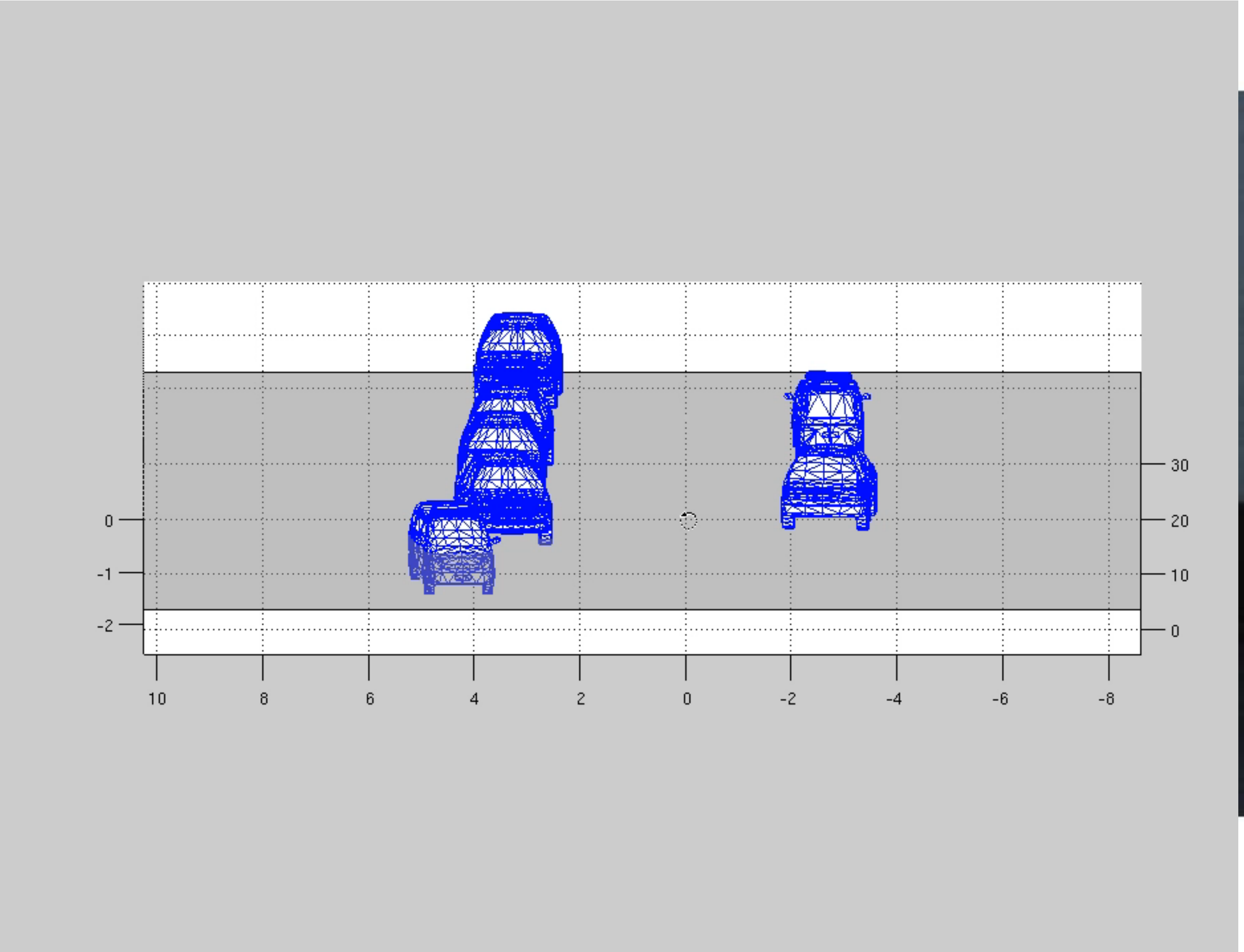


My Work: Occlusion Reasoning



My Work: 3D Localization

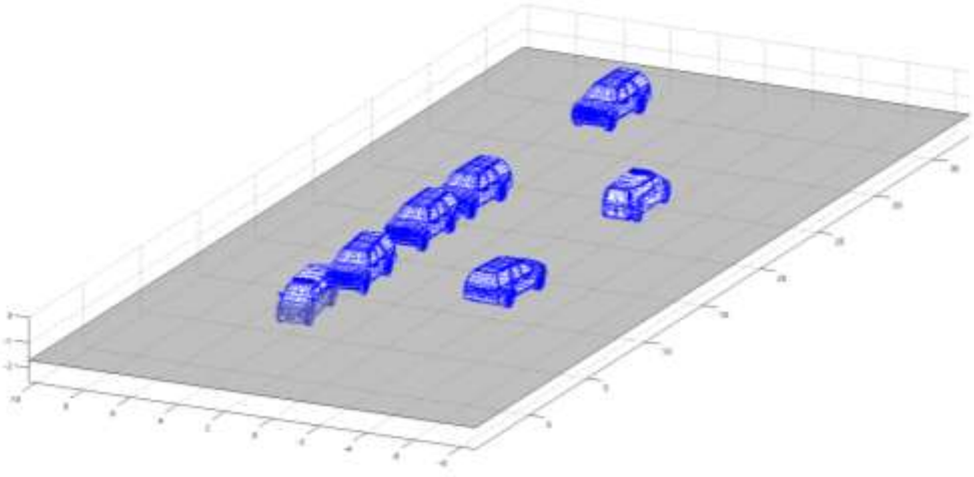




Contribution: 3D Object Representations

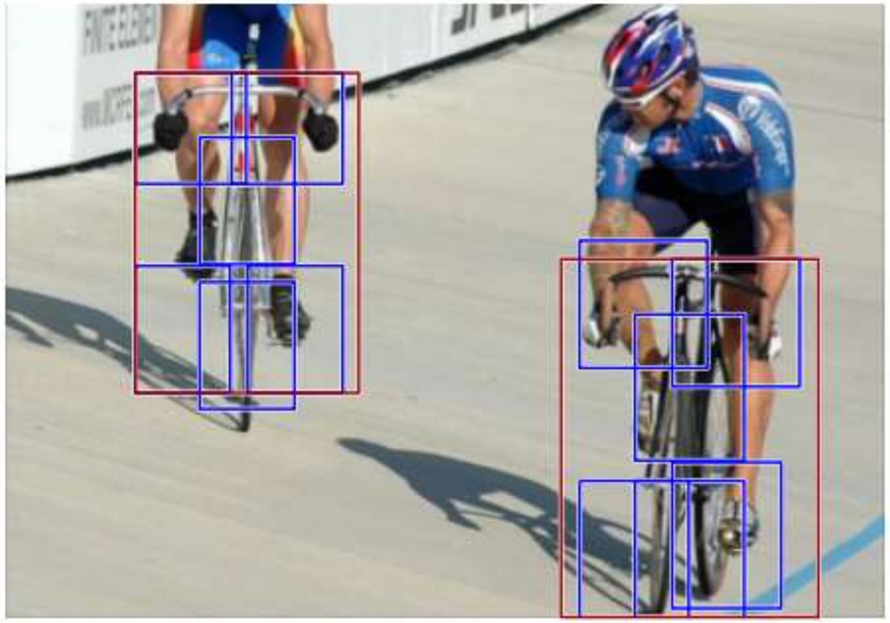
3D Object Representation

A 2D image



The 3D world

Related Work: 2D Object Representations



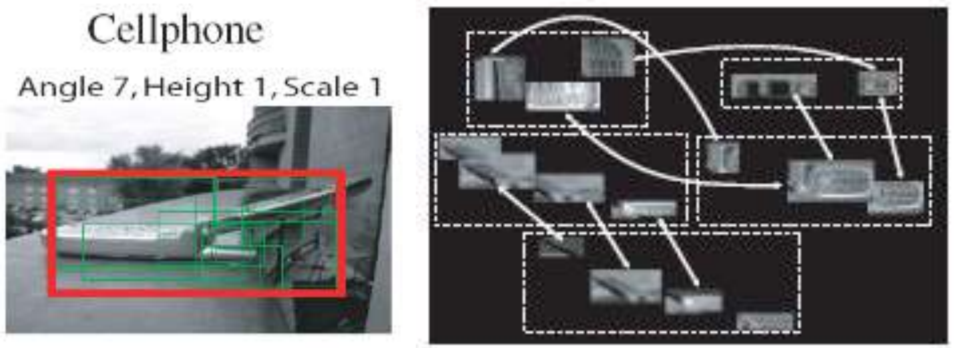
Deformable part model
Felzenszwalb et al., TPAMI'10

- ✓ 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: 2.5D Object Representations

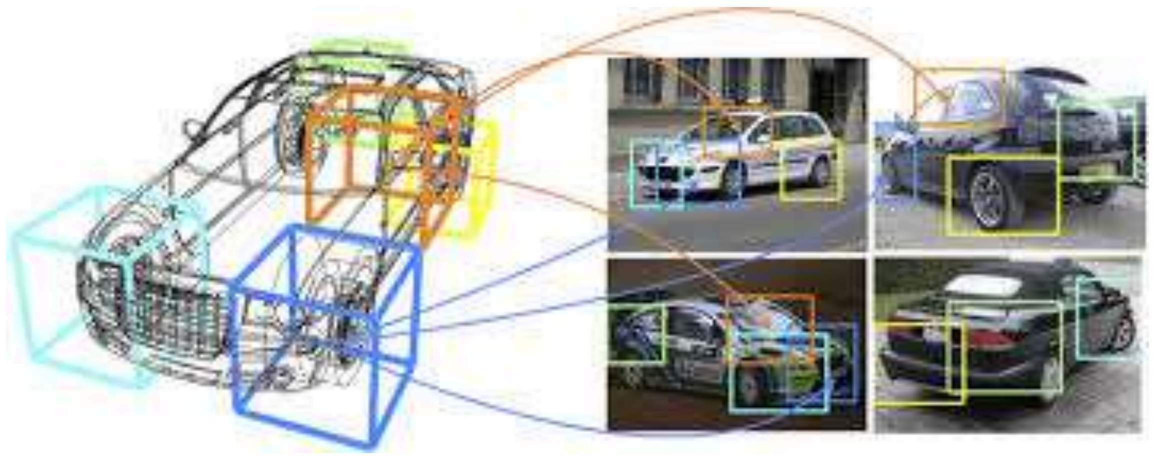


Savarese & Fei-Fei ICCV’07

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

- Thomas et al., CVPR’06
- Savarese & Fei-Fei ICCV’07
- Kushal et al., CVPR’07
- Su et al., ICCV’09
- Sun et al., CVPR’10
- Etc.

Related Work: 3D Object Representations



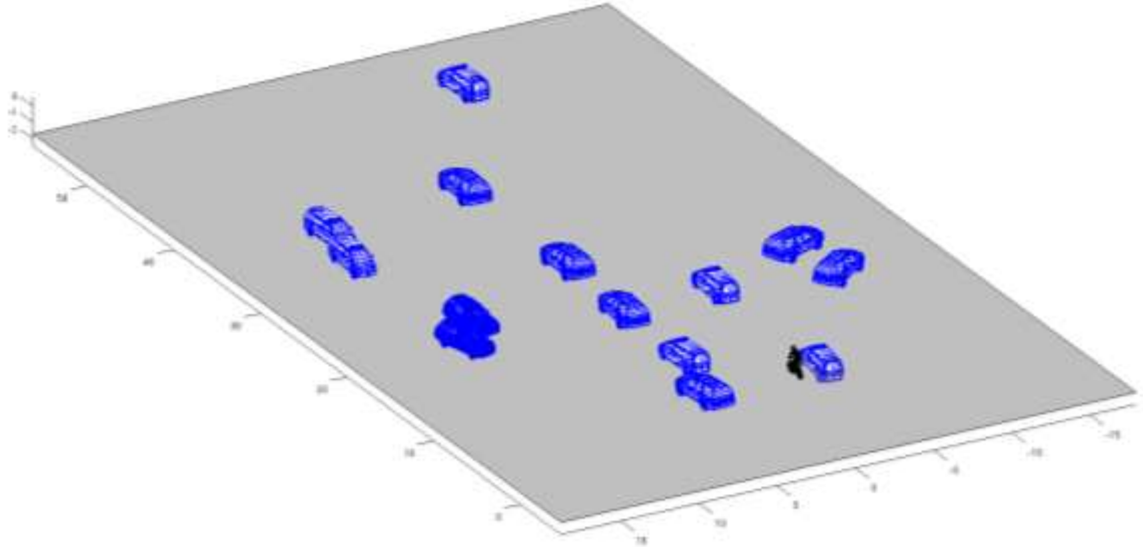
3DDPM
Pepik et al., CVPR’12

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

- Yan et al., ICCV’07
- Hoiem et al., CVPR’07
- 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

Contribution: 3D Object Representations



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

Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Conclusion and Future Work

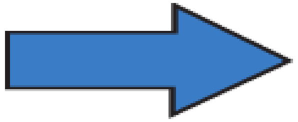
Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Conclusion and Future Work

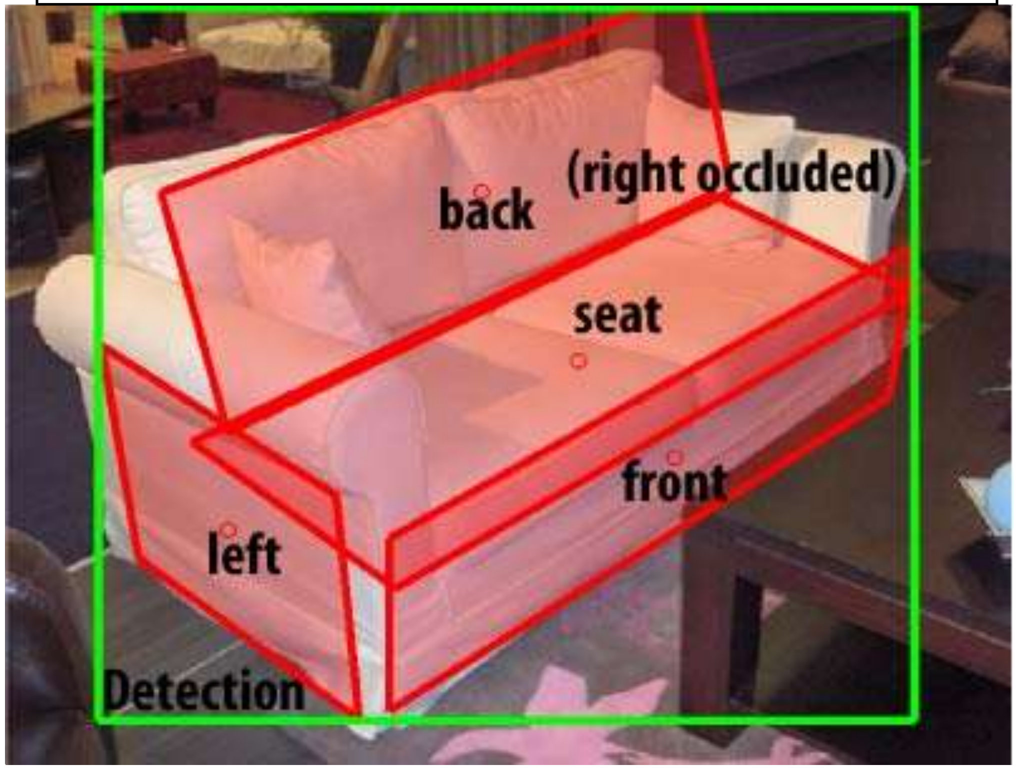
3D Aspect Part Representation

Viewpoint Variation

3D Aspect Part Representation

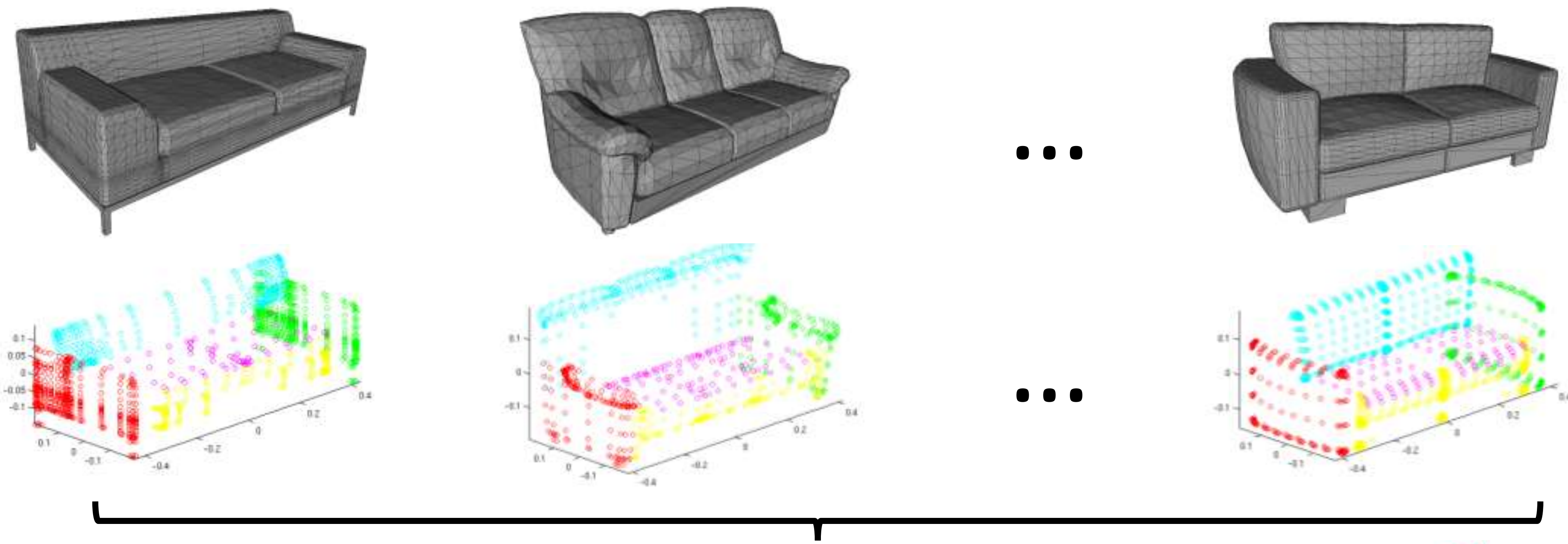


Viewpoint: Azimuth 315˚, Elevation 30˚, Distance 2

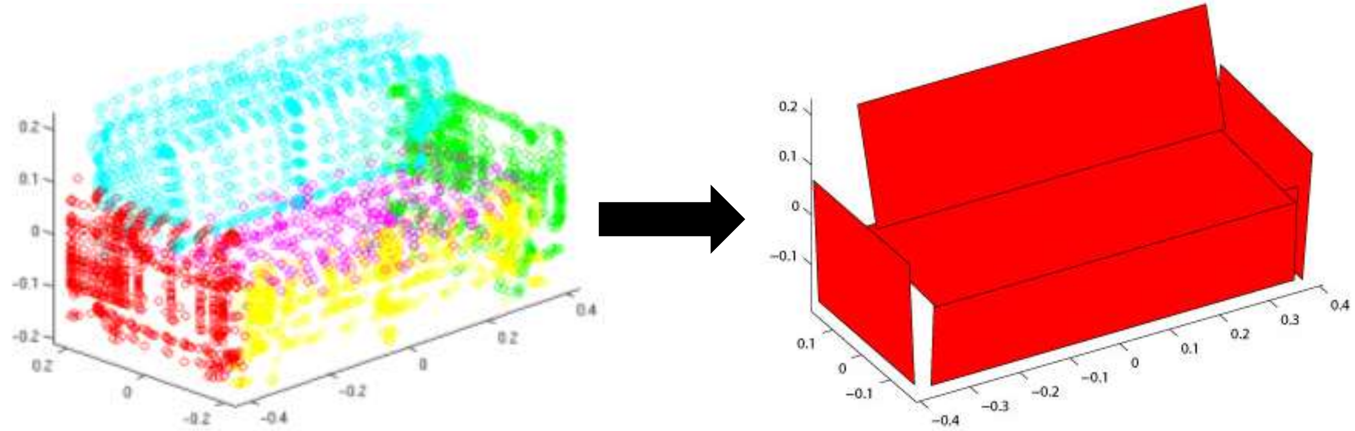


Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.

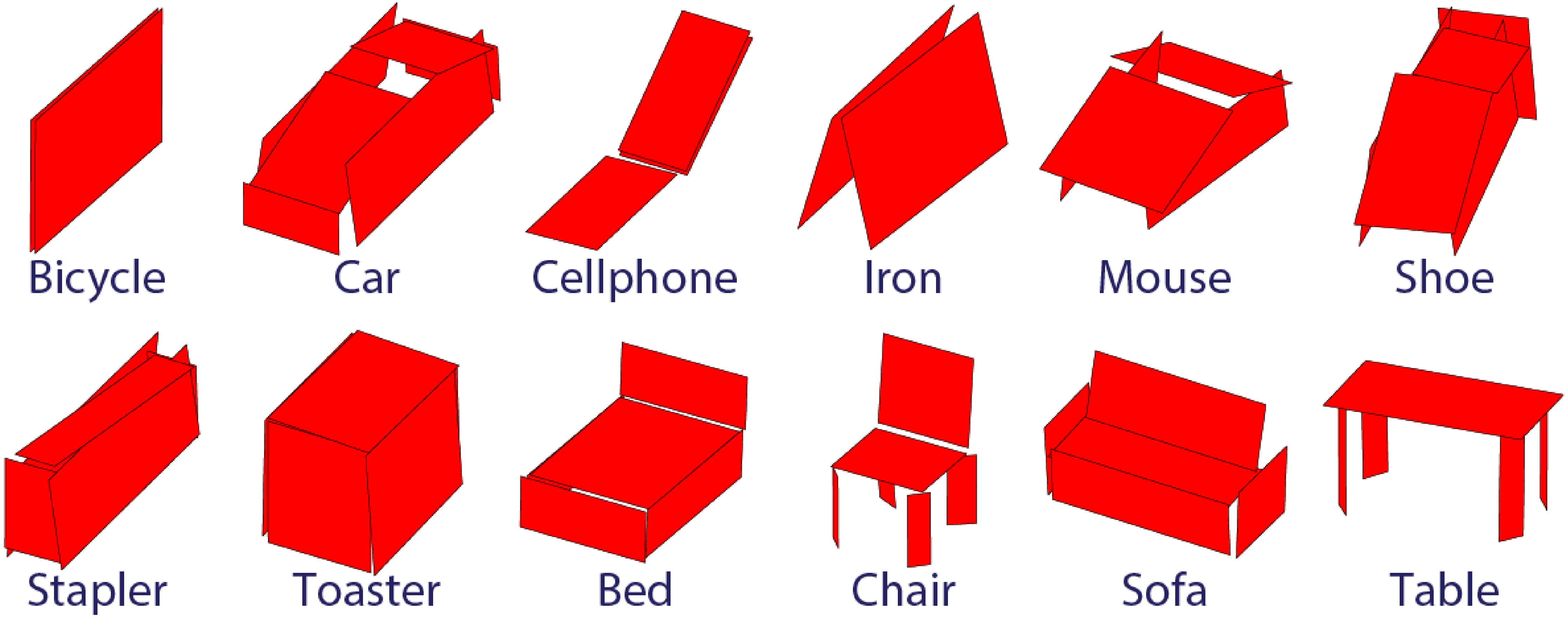
3D Aspect Parts from 3D CAD Models



Mean Shape

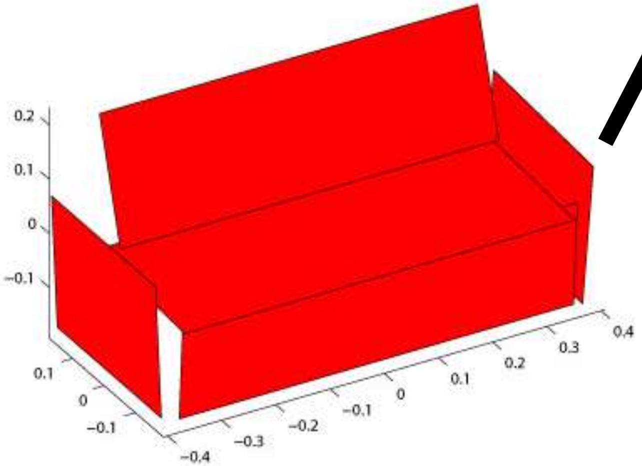


3D Aspect Part Representation



Aspect Layout Model

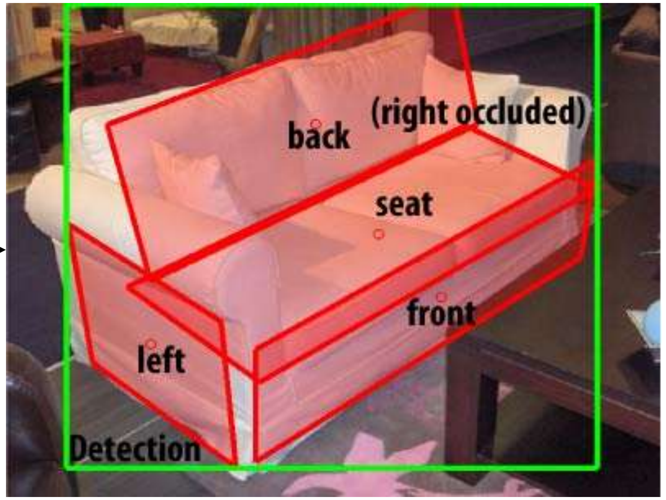
An input image



3D aspect part representation

Aspect
Layout
Model

Viewpoint: Azimuth 315˚,
Elevation 30˚, Distance 2



Output

Aspect Layout Model

- Posterior distribution

$$P(Y, L, O, V | I) \propto \exp(E(Y, L, O, V, I))$$

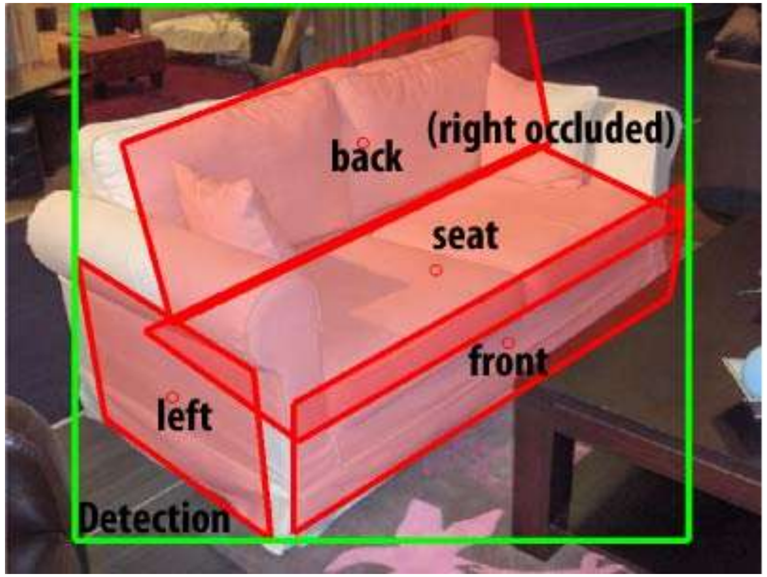
object label

2D part center coordinates

3D object

viewpoint

$$L = (\mathbf{l}_1, \ldots, \mathbf{l}_n),\ \mathbf{l}_i = (x_i, y_i)$$



Aspect Layout Model

- Energy function

$$E(Y,L,O,V,I)=\begin{cases}\displaystyle\sum_i \underbrace{V_1(\mathbf{l}_i,O,V,I)}_{\text{unary potential}}+\sum_{(i,j)}\underbrace{V_2(\mathbf{l}_i,\mathbf{l}_j,O,V)}_{\text{pairwise potential}}, & \text{if } Y=+1\\ 0, & \text{if } Y=-1\end{cases}$$

Aspect Layout Model

- Unary potential

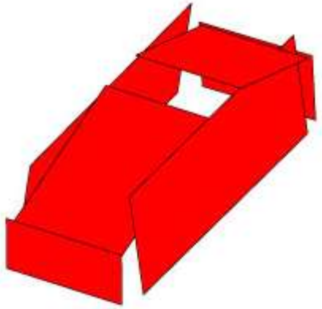
part template

feature vector

$$V_1(\mathbf{l}_i, O, V, I) = \begin{cases} \mathbf{w}_i^T \phi(\mathbf{l}_i, O, V, I), & \text{if unoccluded} \\ \alpha_i, & \text{if self-occluded} \end{cases}$$

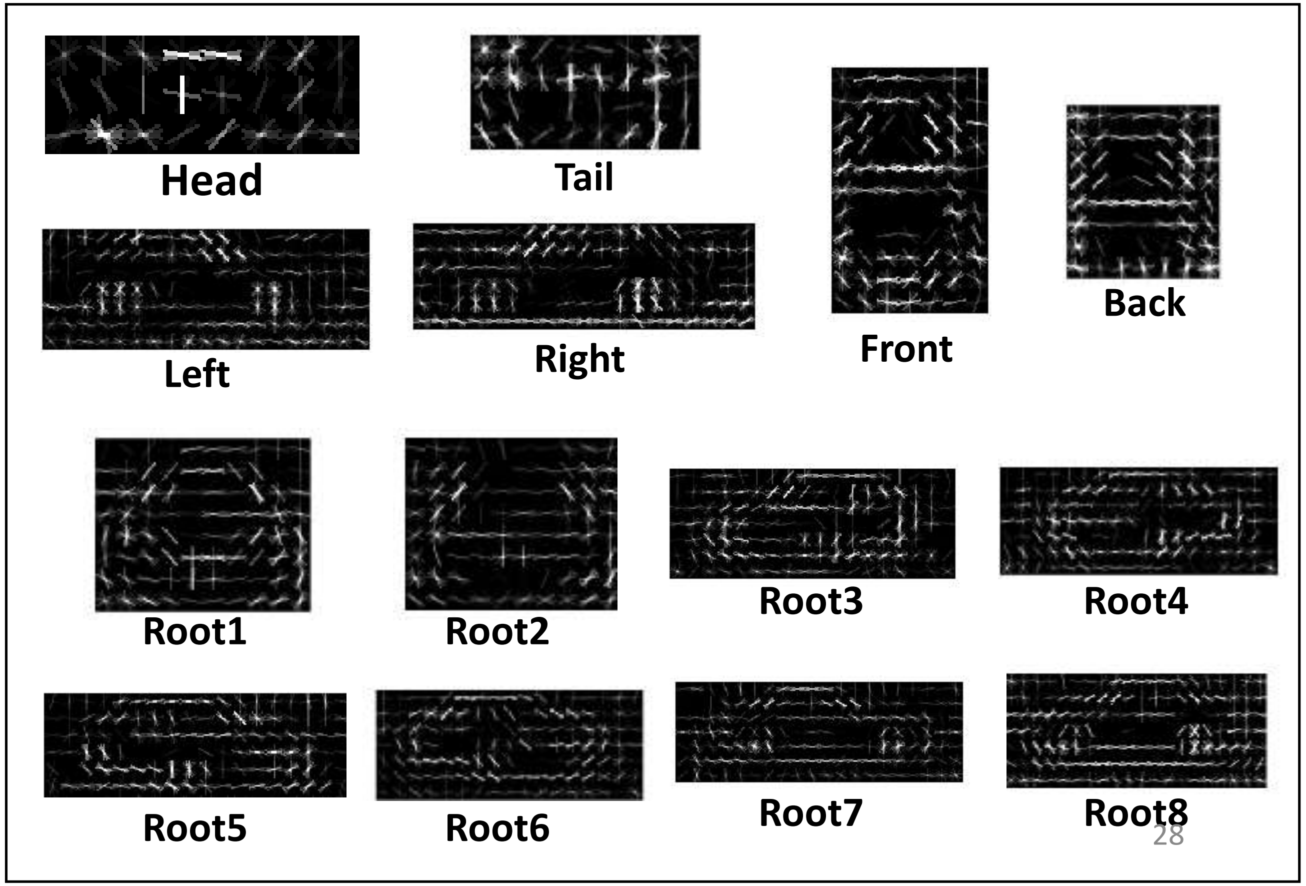
self-occlusion weight

Aspect Layout Model



Car

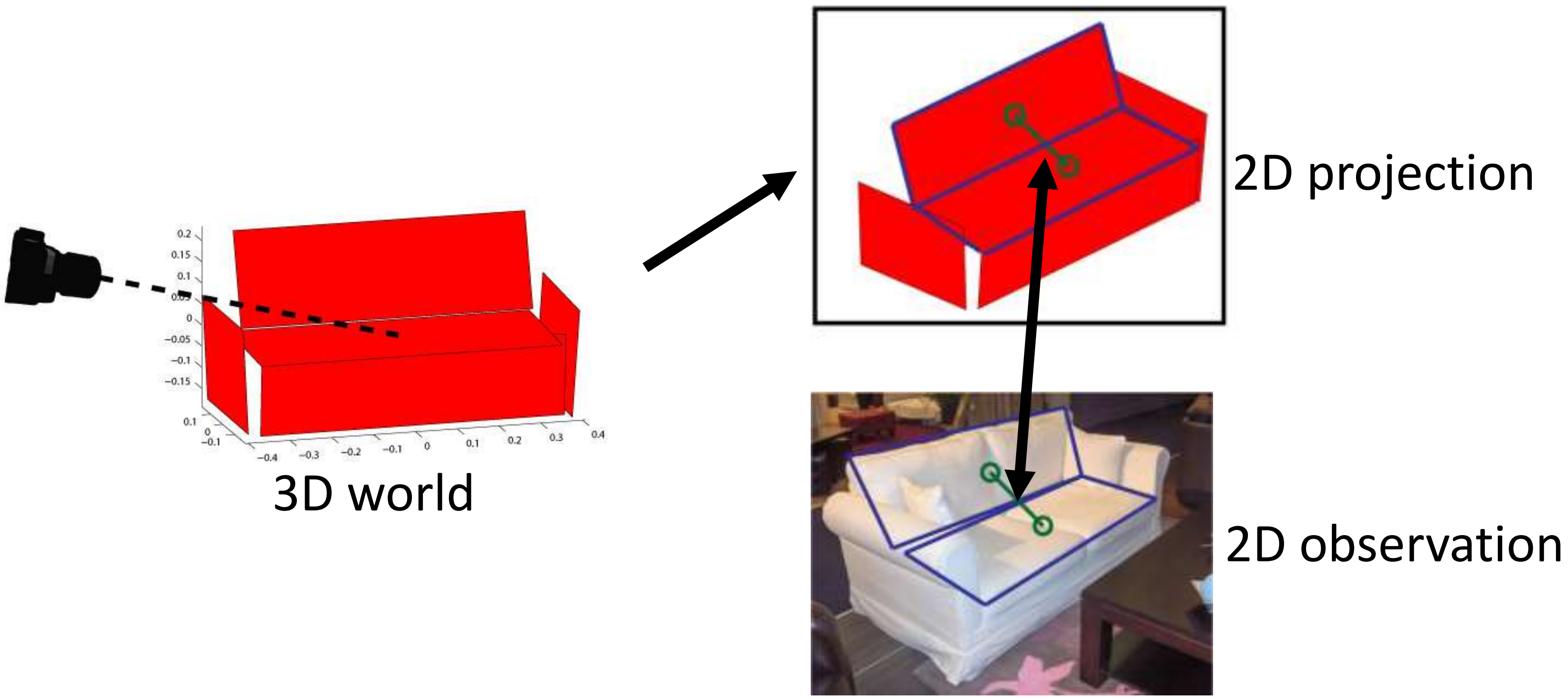
$$V_1(\mathbf{l}_i, O, V, I) = \begin{cases} \mathbf{w}_i^T \phi(\mathbf{l}_i, O, V, I), \text{ if unoccluded} \\ \alpha_i, \text{ if occluded} \end{cases}$$



Aspect Layout Model

- Pairwise potential

$$V_2(\mathbf{l}_i, \mathbf{l}_j, O, V) = -w_x (x_i - x_j + d_{ij,O,V} \cos(\theta_{ij,O,V}))^2 - w_y (y_i - y_j + d_{ij,O,V} \sin(\theta_{ij,O,V}))^2$$



Aspect Layout Model

- Training with Structural SVM [1]

$$\min_{\theta} \frac{1}{2}\|\theta\|^2 + \lambda \sum_{t=1}^{N} \left[\max_{Y,L,O,V} \left[\theta^T \Psi_{t,Y,L,O,V} + \Delta_{t,Y,L,O,V} \right] - \theta^T \Psi_{t,Y^t,L^t,O^t,V^t} \right]$$

- Inference $(Y^*, L^*, O^*, V^*) = \arg\max_{Y,L,O,V} E(Y, L, O, V, I|\theta)$
  - Loop over discretized viewpoints
  - Run Belief Propagation [2] under each viewpoint to predict part locations

[1] I. Tsochantaridis, T. Hofmann, T. Joachims and Y. Altun. Support vector machine learning for interdependent and structured output spaces. In ICML, 2004.
[2] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In Exploring artificial intelligence in the new millennium, 2003.

Aspect Layout Model

- Best results upon publication in pose estimation and 3D part estimation

Cars from 3D Object dataset [Savarese & Fei-Fei ICCV'07]

Method	Ours	[1]	[2]	[3]	[4]	[5]	[6]
Viewpoint (cars)	**93.4%**	85.4	85.3	81	70	67	48.5

Cars from EPFL dataset [Ozuysal et al. CVPR'09]

Method	Ours	Ours - baseline	DPM [7]	[8]
Viewpoint (cars)	**64.8%**	58.1	56.6	41.6

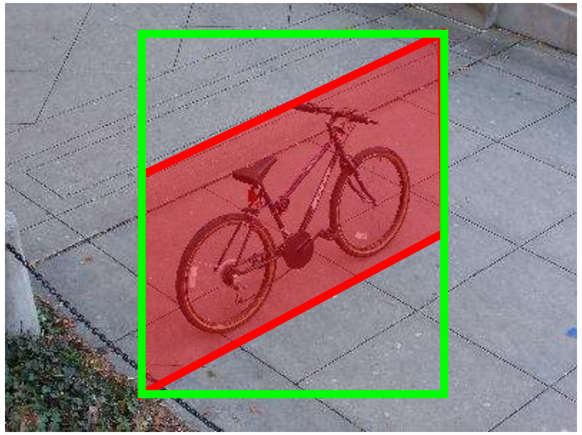
Chairs, tables, sofas and beds from IMAGE NET [Deng et al. CVPR'09]

Method	Ours	Ours - baseline	DPM [7]
Viewpoint	**63.4%**	34.0	49.5

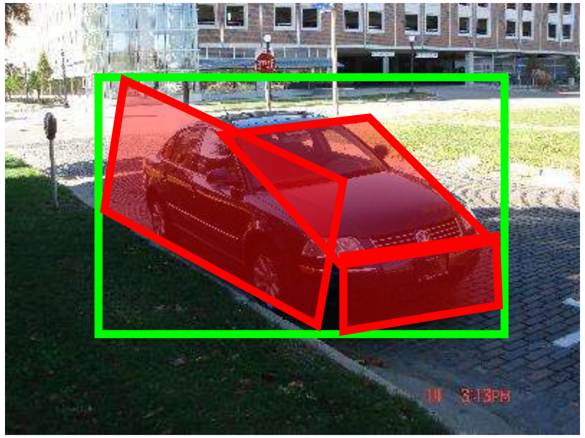
[1] N. Payet and S. Todorovic. From contours to 3d object detection and pose estimation. In ICCV, 2011.
[2] D. Glasner, M. Galun, S. Alpert, R. Basri, and G. Shakhnarovich. Viewpoint-aware object detection and pose estimation. In ICCV, 2011.
[3] M. Stark, M. Goesele, and B. Schiele. Back to the future: Learning shape models from 3d cad data. In BMVC, 2010.
[4] J. Liebelt and C. Schmid. Multi-view object class detection with a 3D geometric model. In CVPR, 2010.
[5] H. Su, M. Sun, L. Fei-Fei, and S. Savarese. Learning a dense multiview representation for detection, viewpoint classification. In ICCV, 2009.
[6] M. Arie-Nachimson and R. Basri. Constructing implicit 3d shape models for pose estimation. In ICCV, 2009.
[7] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[8] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In CVPR, 2009.

Aspect Layout Model

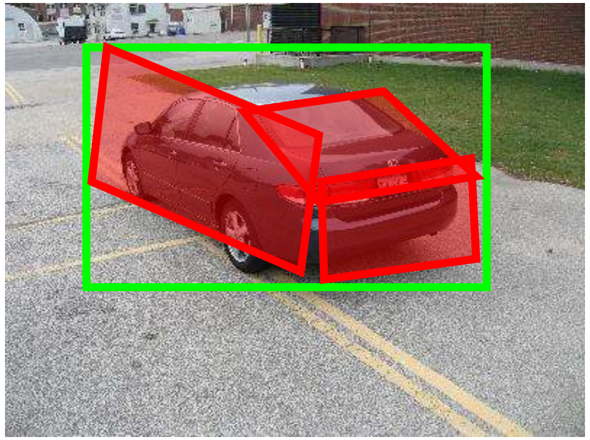
Prediction: a=225, e=30, d=7



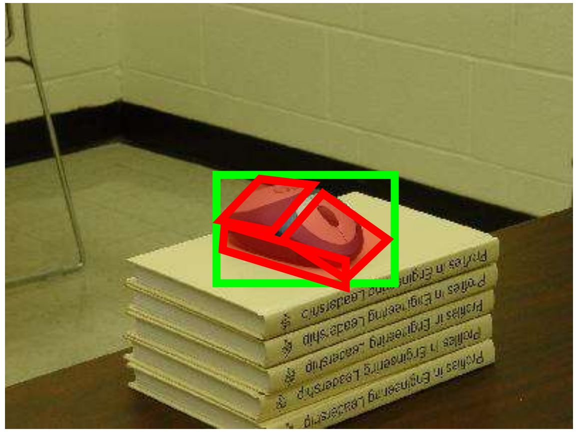
Prediction: a=330, e=15, d=7



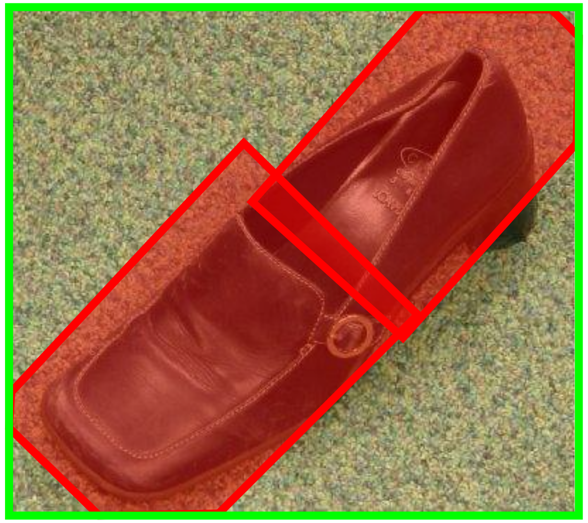
Prediction: a=150, e=15 d=7



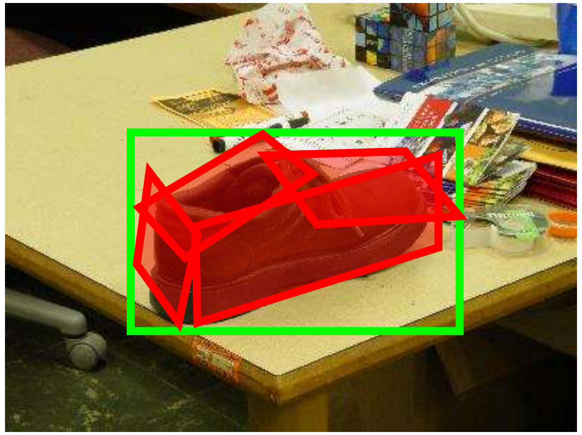
Prediction: a=300, e=45 d=23



Prediction: a=45, e=90, d=5



Prediction: a=240, e=45, d=11

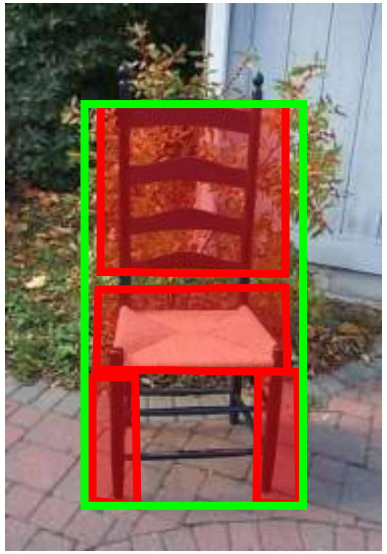


Aspect Layout Model

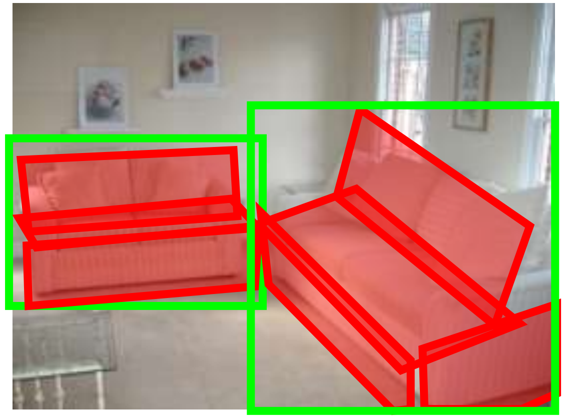
Prediction: a=30, e=15, d=2.5

Prediction: a=0, e=15, d=1.5

Prediction: a=0, e=30, d=7

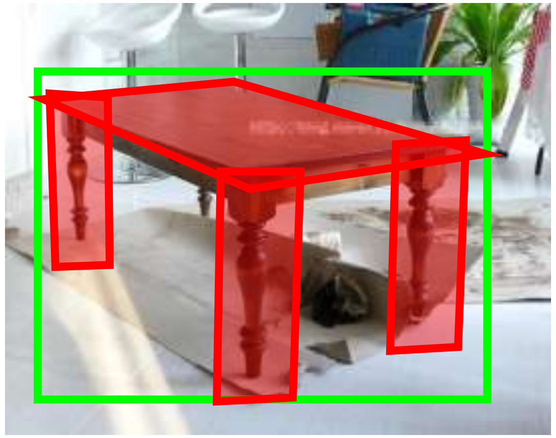


Prediction: a=345, e=15, d=3.5
a=60, e-30, d=2.5



Prediction: a=315, e=30, d=2

Prediction: a=60, e=15, d=2



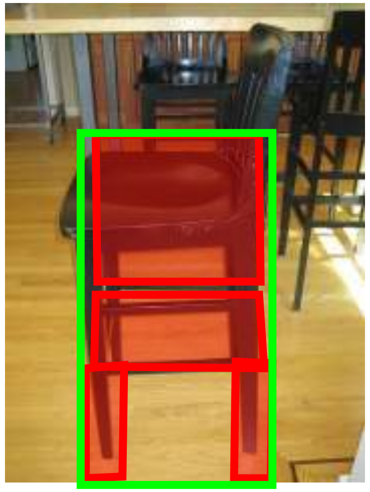
ImageNet dataset [Deng et al. 2010]

Wrong examples

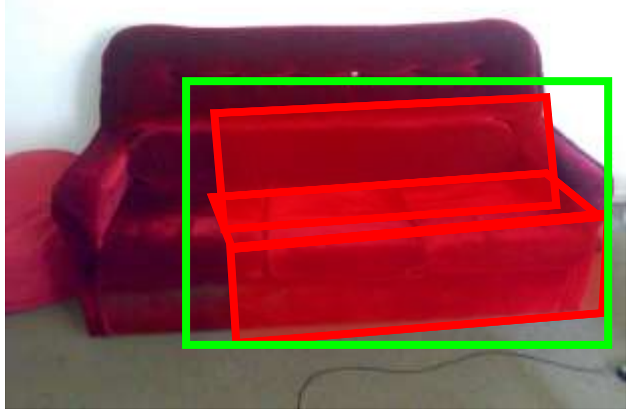
Prediction: a=45, e=15, d=1.5

Prediction: a=225, e=30, d=7

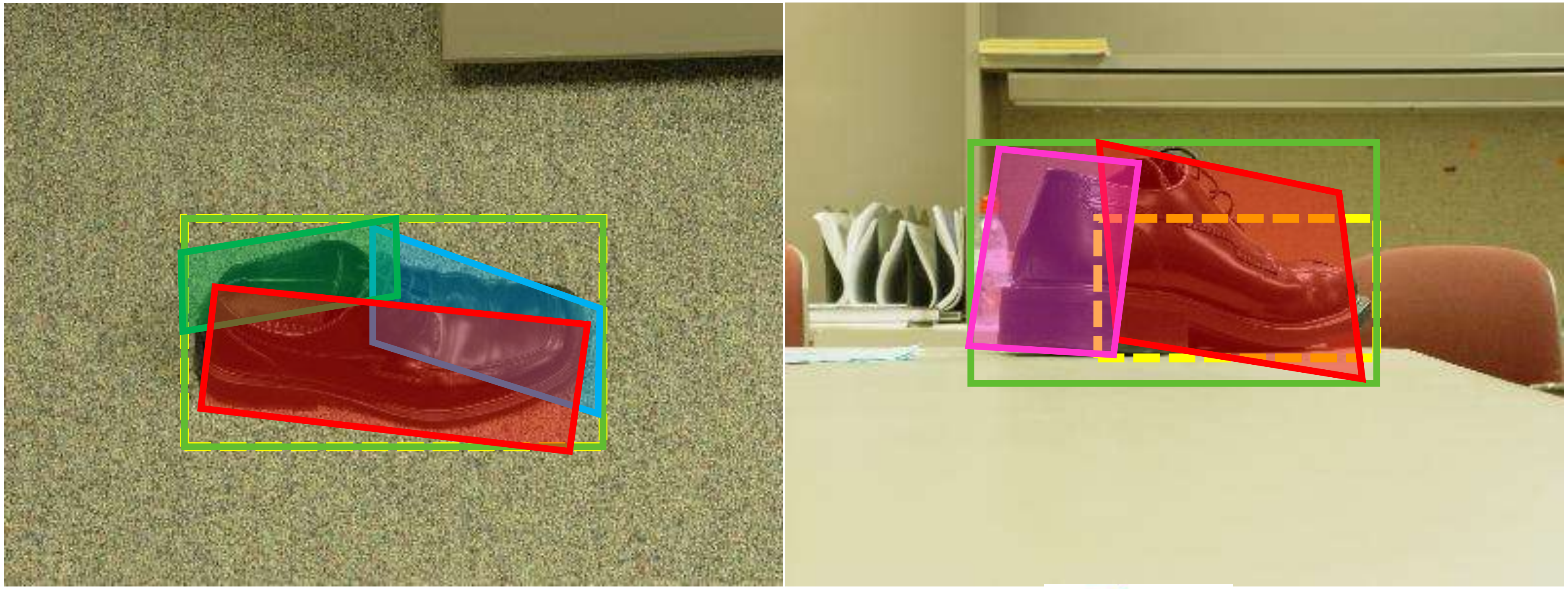
Prediction: a=0, e=30, d=7



Prediction: a=345, e=15 d=2.5

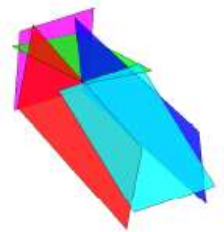


Application I: Object Co-detection with 3D Aspect Parts



Single Image Detector

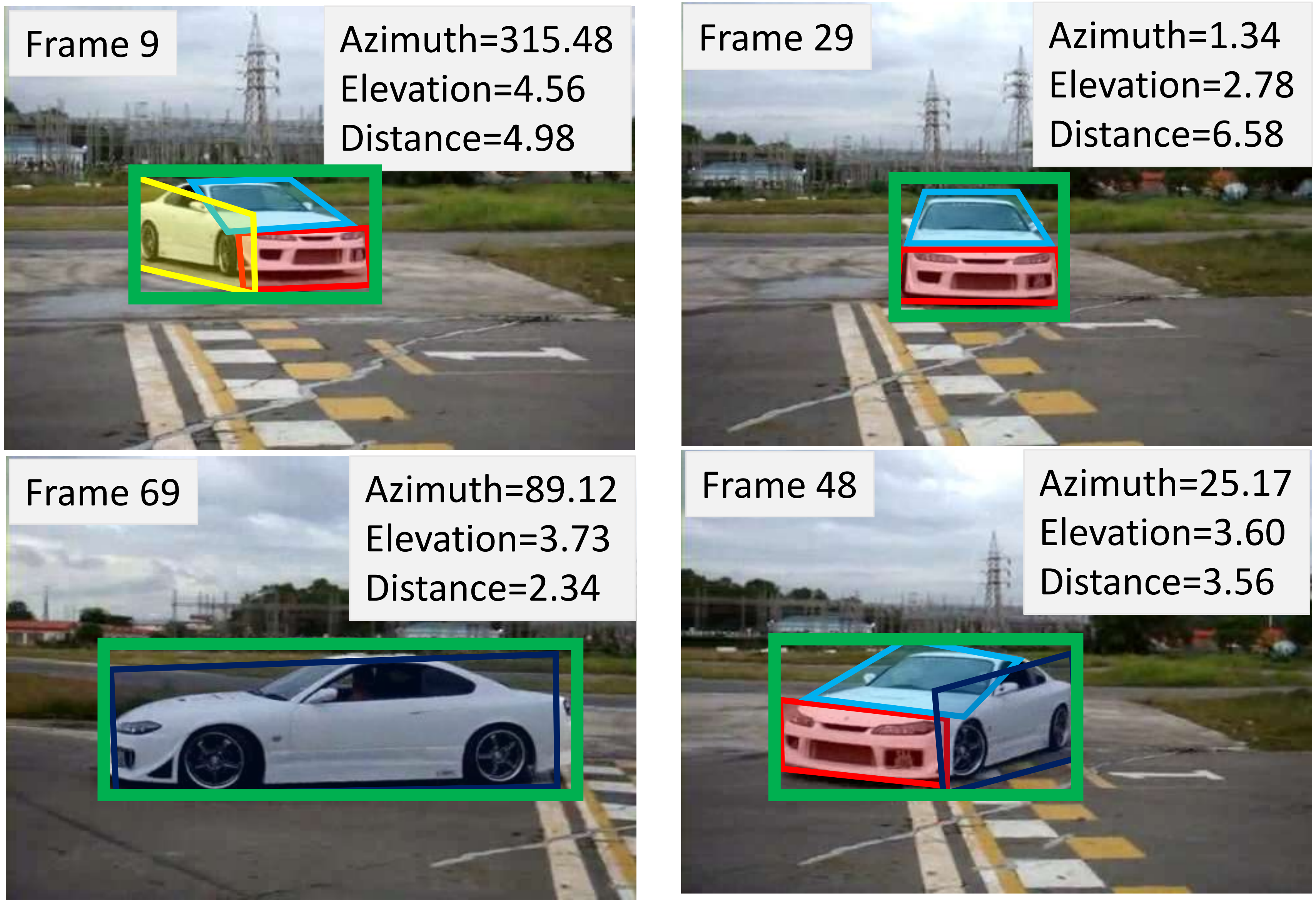
Co-detector



Shoe model

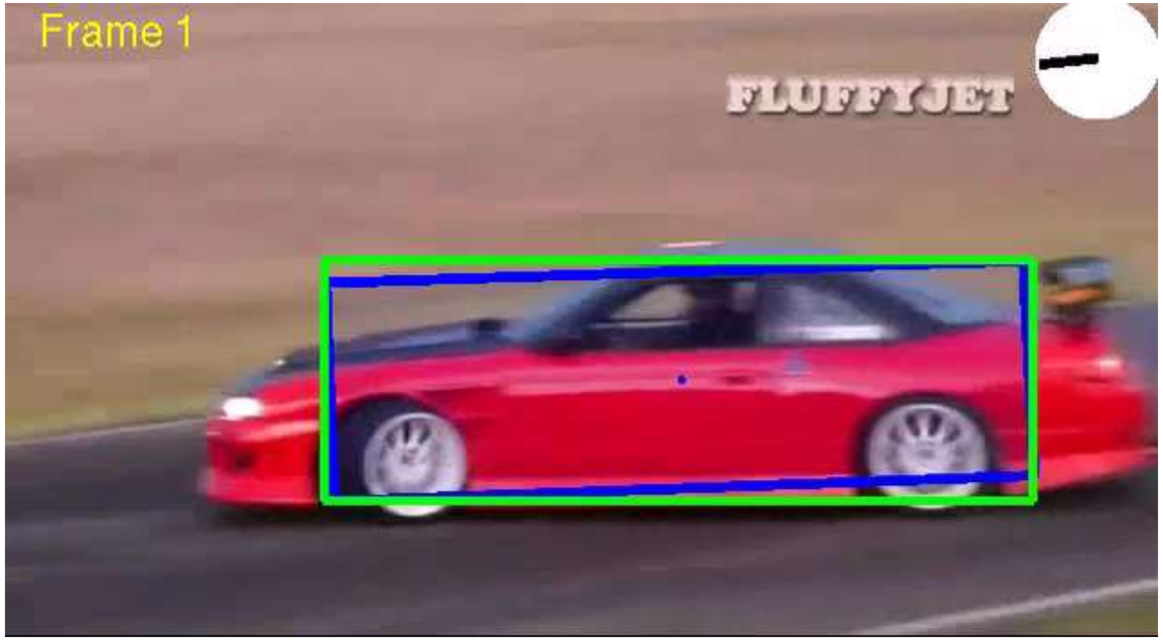
S. Bao, Y. Xiang and S. Savarese. Object Co-detection. In ECCV, 2012.

Application II: Multiview Object Tracking with 3D Aspect Parts

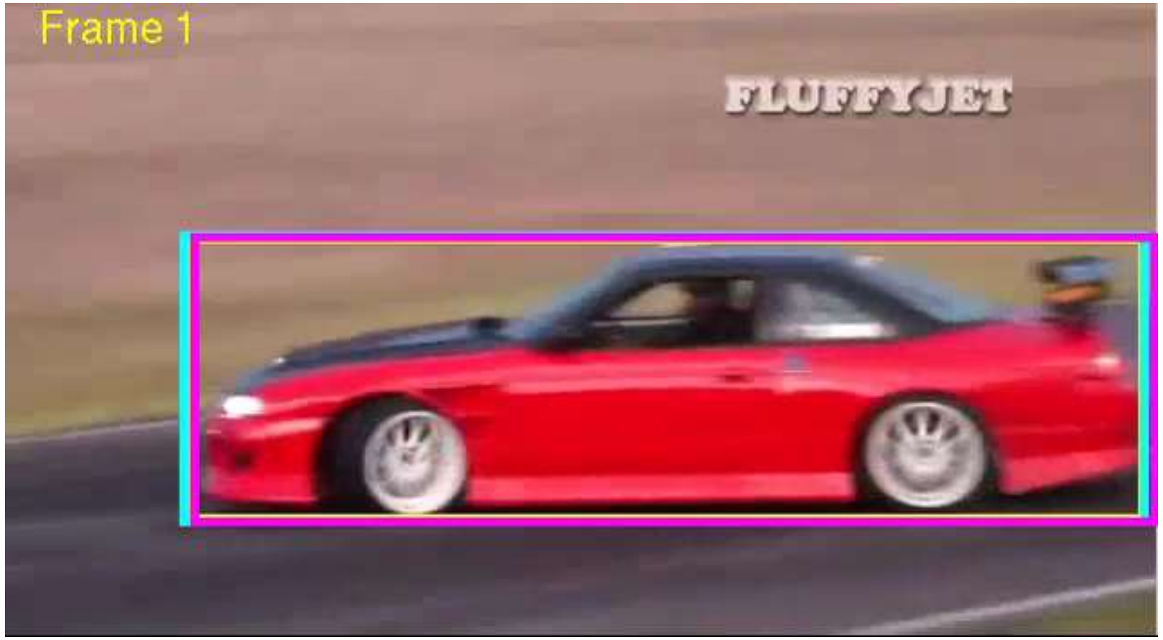


Y. Xiang, C. Song, R. Mottaghi and S. Savarese. Monocular Multiview Object Tracking with 3D Aspect Parts. In ECCV, 2014.

Application II: Multiview Object Tracking with 3D Aspect Parts



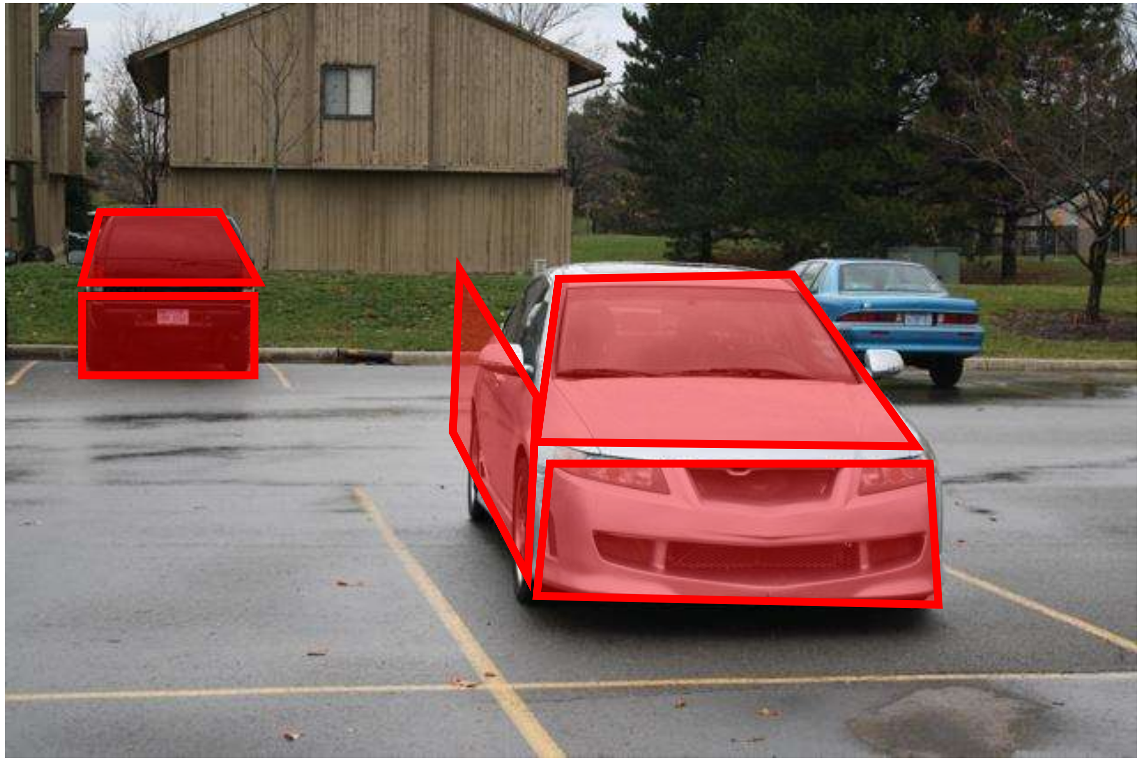
Ours: Multiview tracker



MIL L1 TLD Struct

[MIL] Babenko, B., Yang, M.H., Belongie, S.: Robust object tracking with online multiple instance learning. TPAMI, 2011.
[L1] Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust l1 tracker using accelerated proximal gradient approach. In CVPR, 2012.
[TLD] Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. TPAMI, 2012.
[Struct] Hare, S., Saari, A., Torr, P.H.: Struck: Structured output tracking with kernels. In ICCV, 2011.

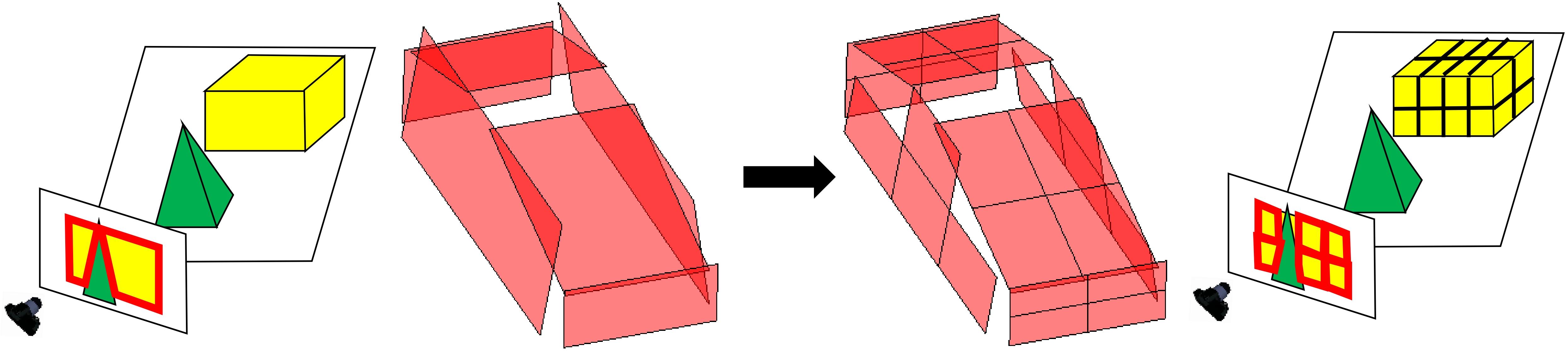
How to handle occlusion?



?

Occlusion changes the appearances of objects.

3D Aspectlet Representation

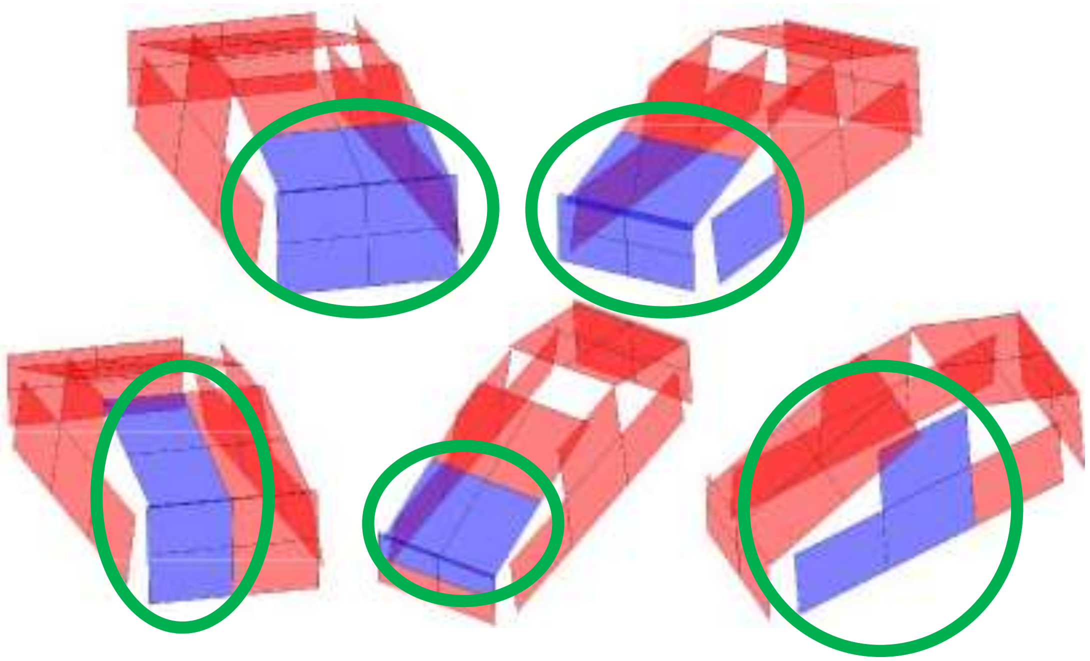


3D Aspect Parts

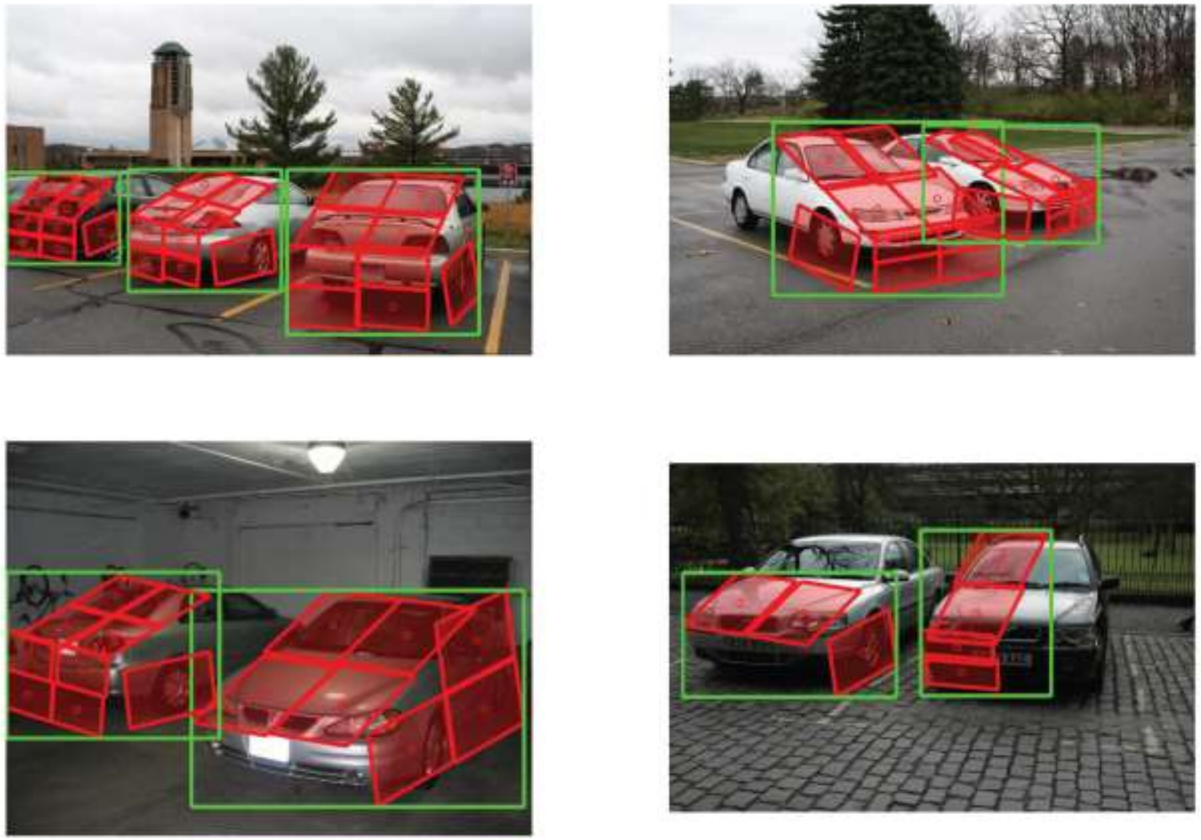
Atomic 3D Aspect Parts

Y. Xiang and S. Savarese. Object Detection by 3D Aspectlets and Occlusion Reasoning. In 3dRR, 2013.

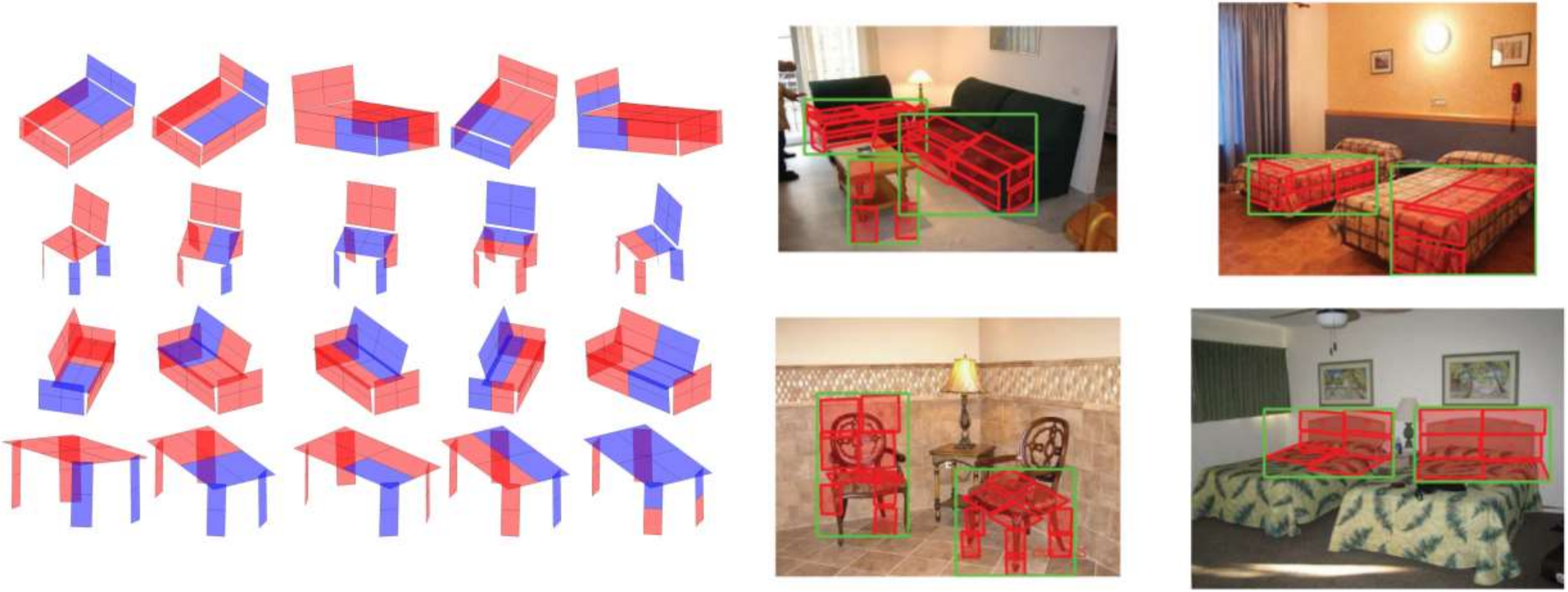
3D Aspectlet Representation



3D Aspectlets



3D Aspectlet Representation

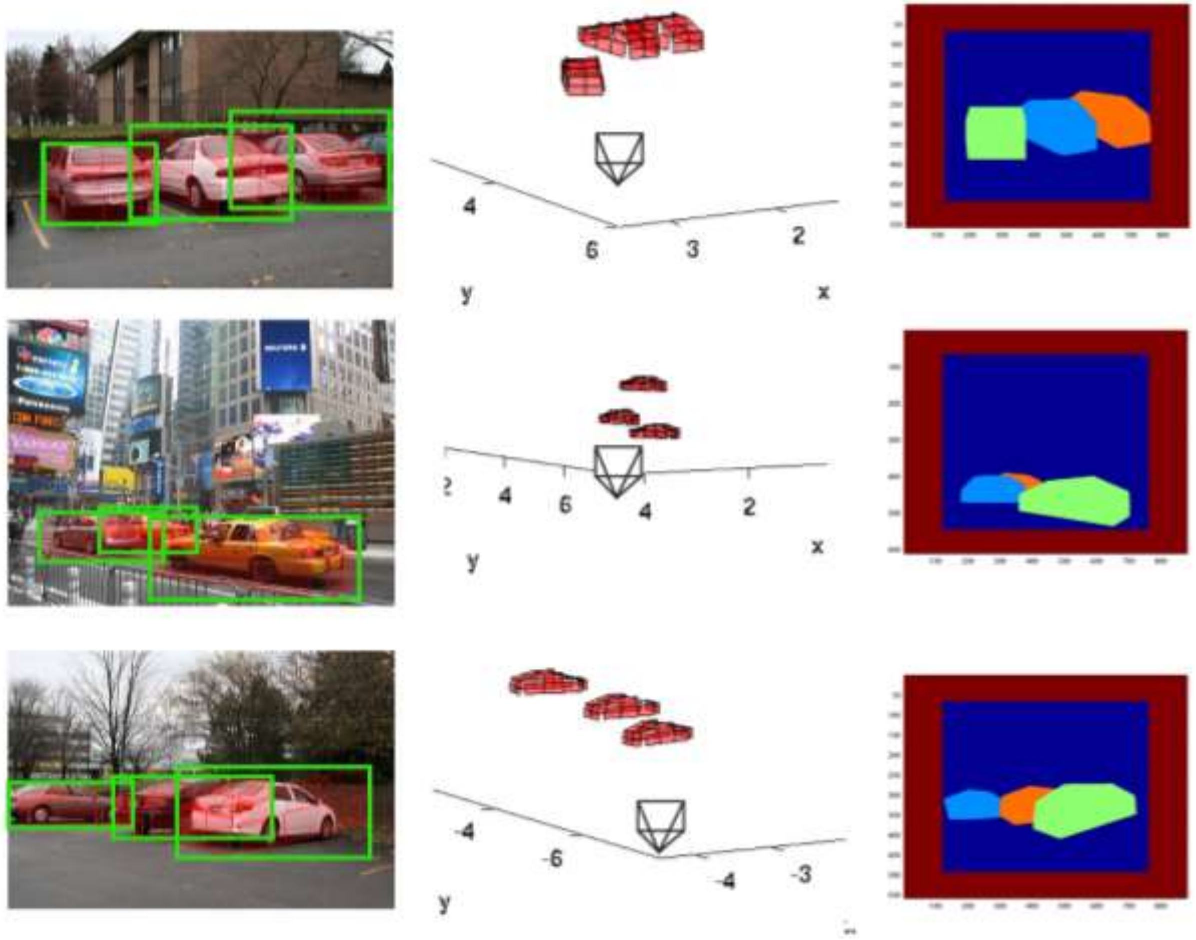


Object Detection Experiments

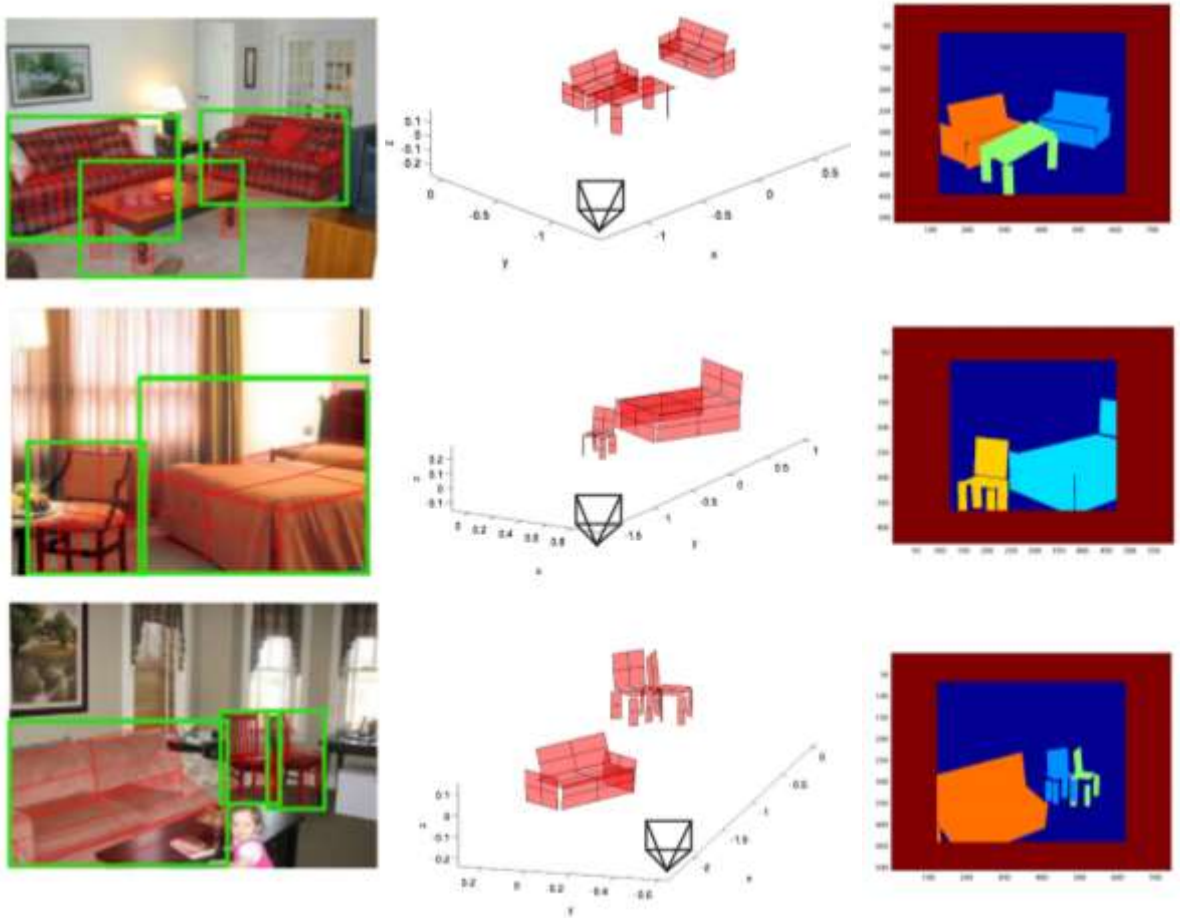
Dataset	Outdoor-scene			Indoor-scene		
% occlusion	< 0.3	0.3 – 0.6	> 0.6	<0.2	0.2-0.4	>0.4
# images	66	68	66	77	111	112
ALM [1]	72.3	42.9	35.5	38.5	25.0	20.2
DPM [2]	75.9	58.6	44.6	38.0	22.9	21.9
Ours 3D Aspectlets	**80.2**	**63.3**	**52.9**	**45.9**	**34.5**	**28.0**

[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.

Object Detection Experiments



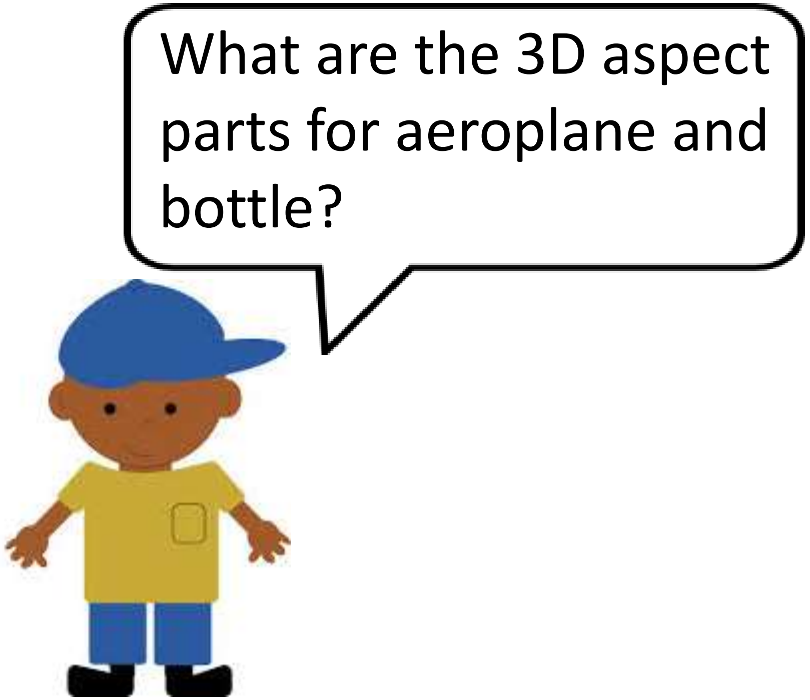
Outdoor Scenes



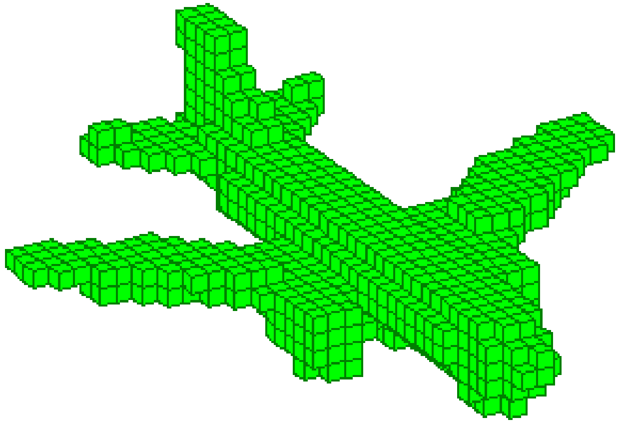
Indoor Scenes

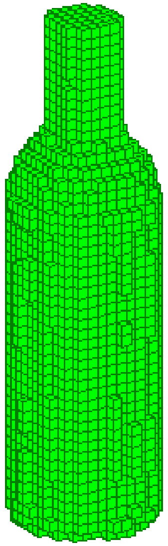
Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Conclusion and Future Work

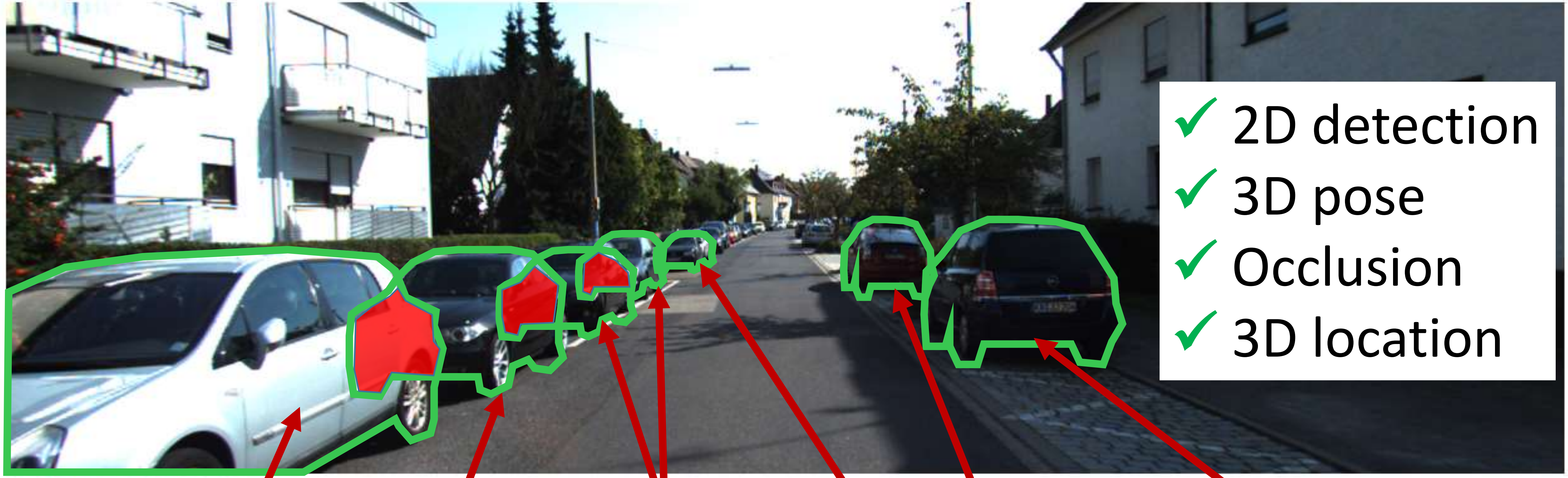

What are the 3D aspect parts for aeroplane and bottle?

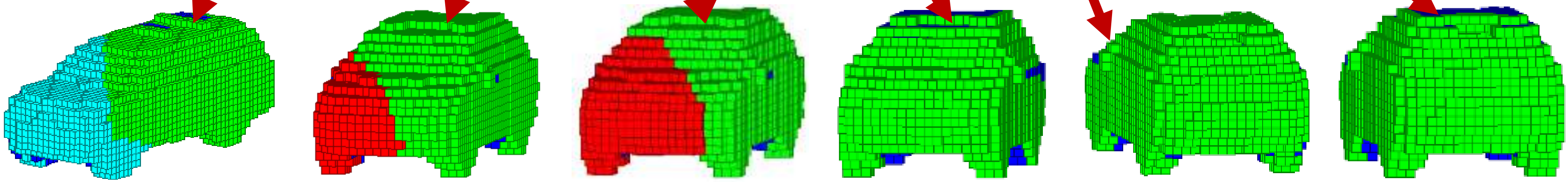

US AIRWAYS





Data-Driven 3D Voxel Patterns



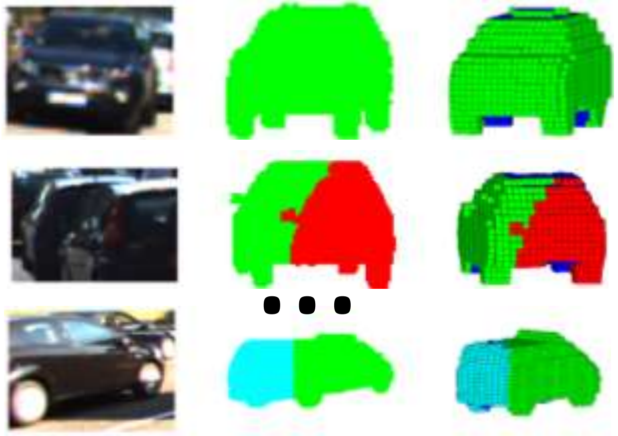


- ✓ 2D detection
- ✓ 3D pose
- ✓ Occlusion
- ✓ 3D location

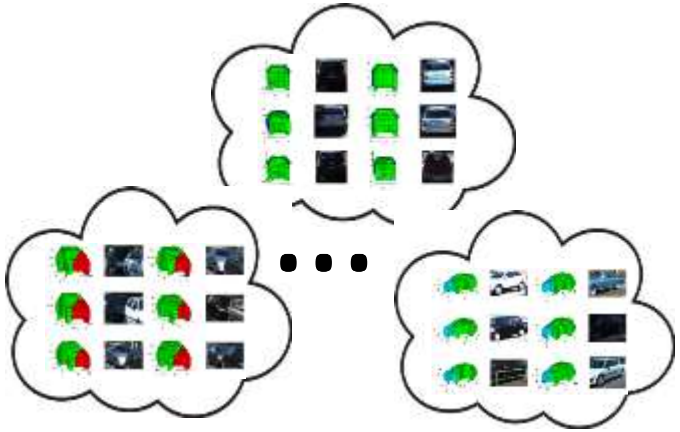
Y. Xiang, W. Choi, Y. Lin and S. Savarese. Data-Driven 3D Voxel Patterns for Object Category Recognition. In CVPR, 2015.

Training Pipeline Overview

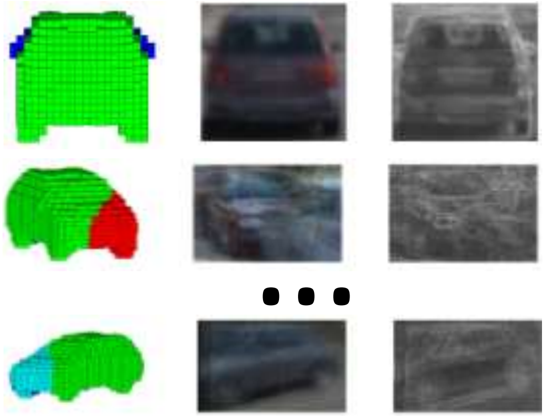
1. Align 2D images with 3D CAD models



2. 3D voxel exemplars

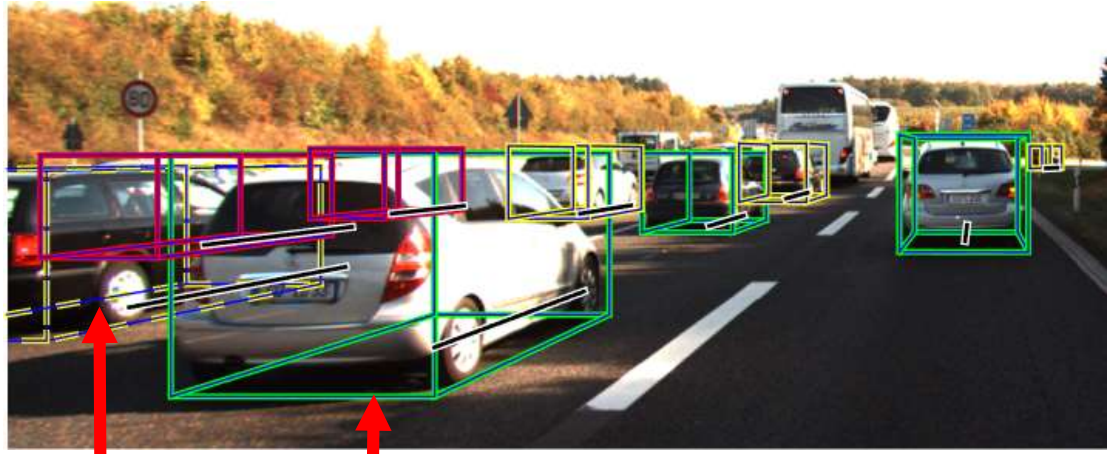


3. 3D voxel patterns

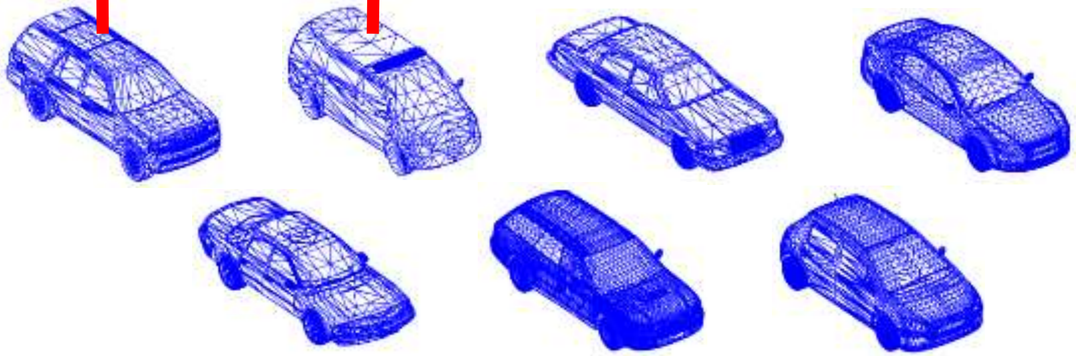


4. Training 3D voxel pattern detectors

1. Align 2D Images with 3D CAD Models

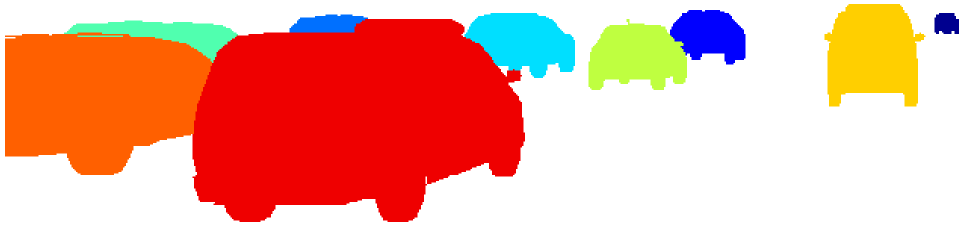


3D annotations • • •



3D CAD models

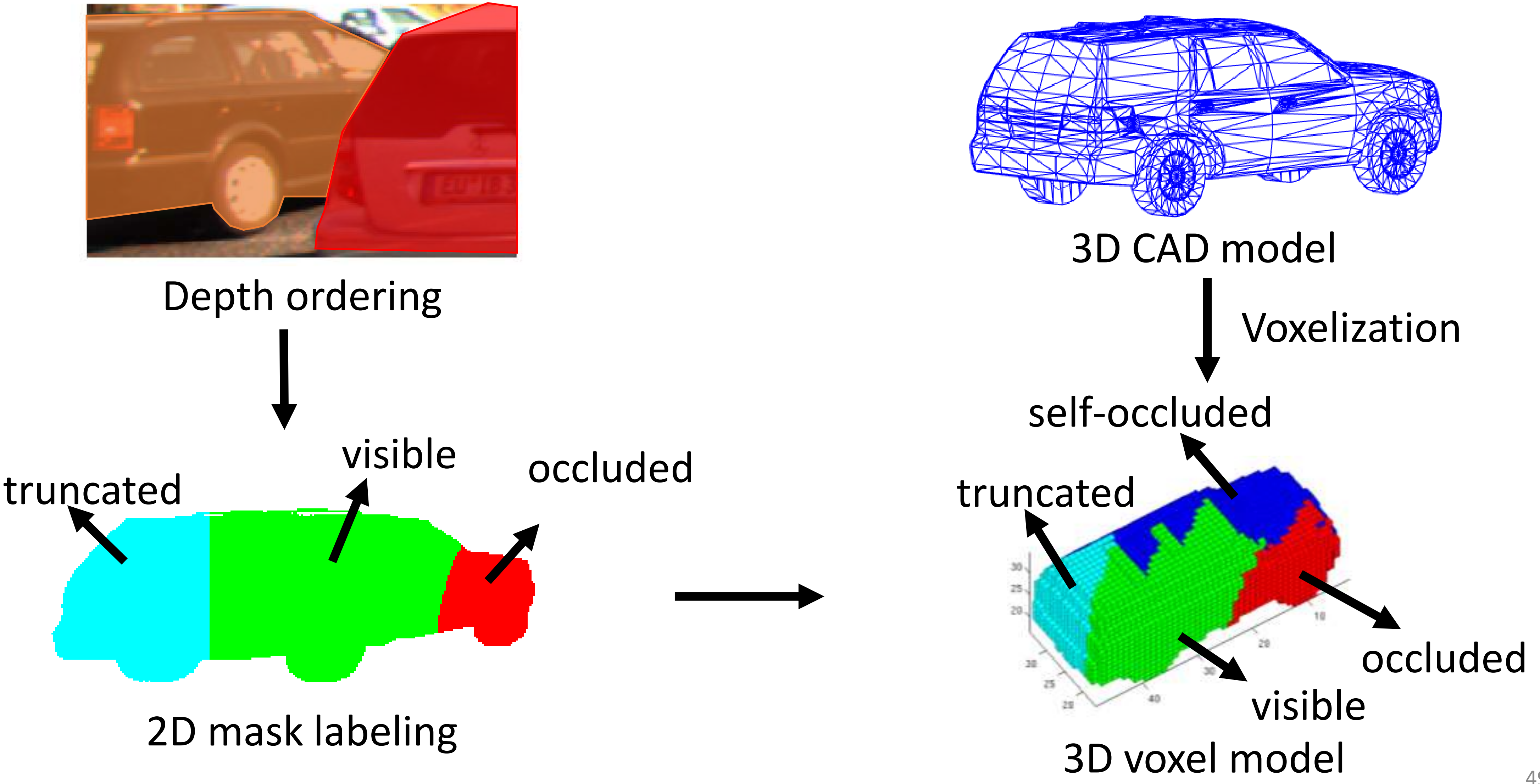
Project of 3D CAD models



Depth ordering

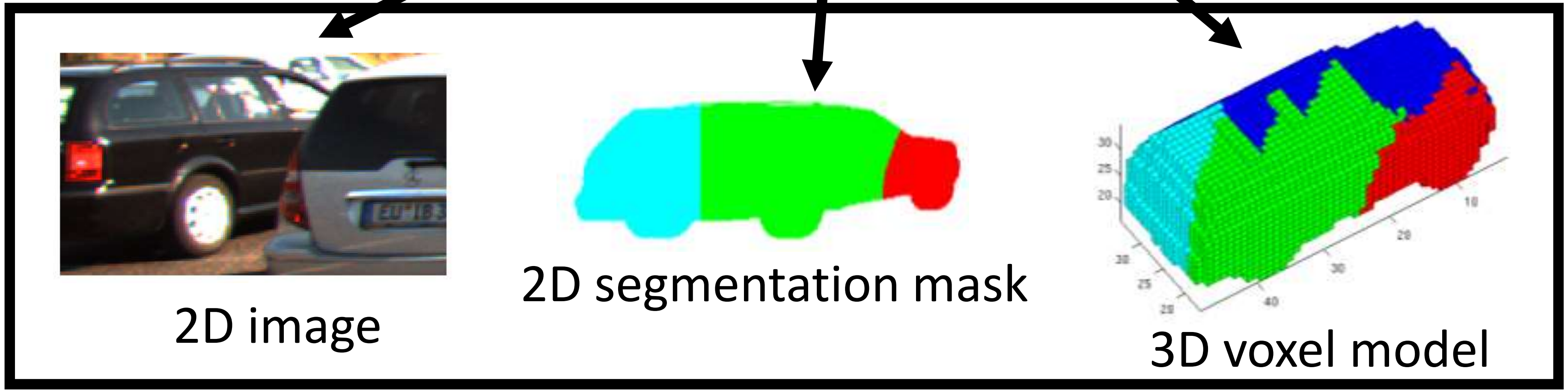
A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012

2. Building 3D Voxel Exemplars

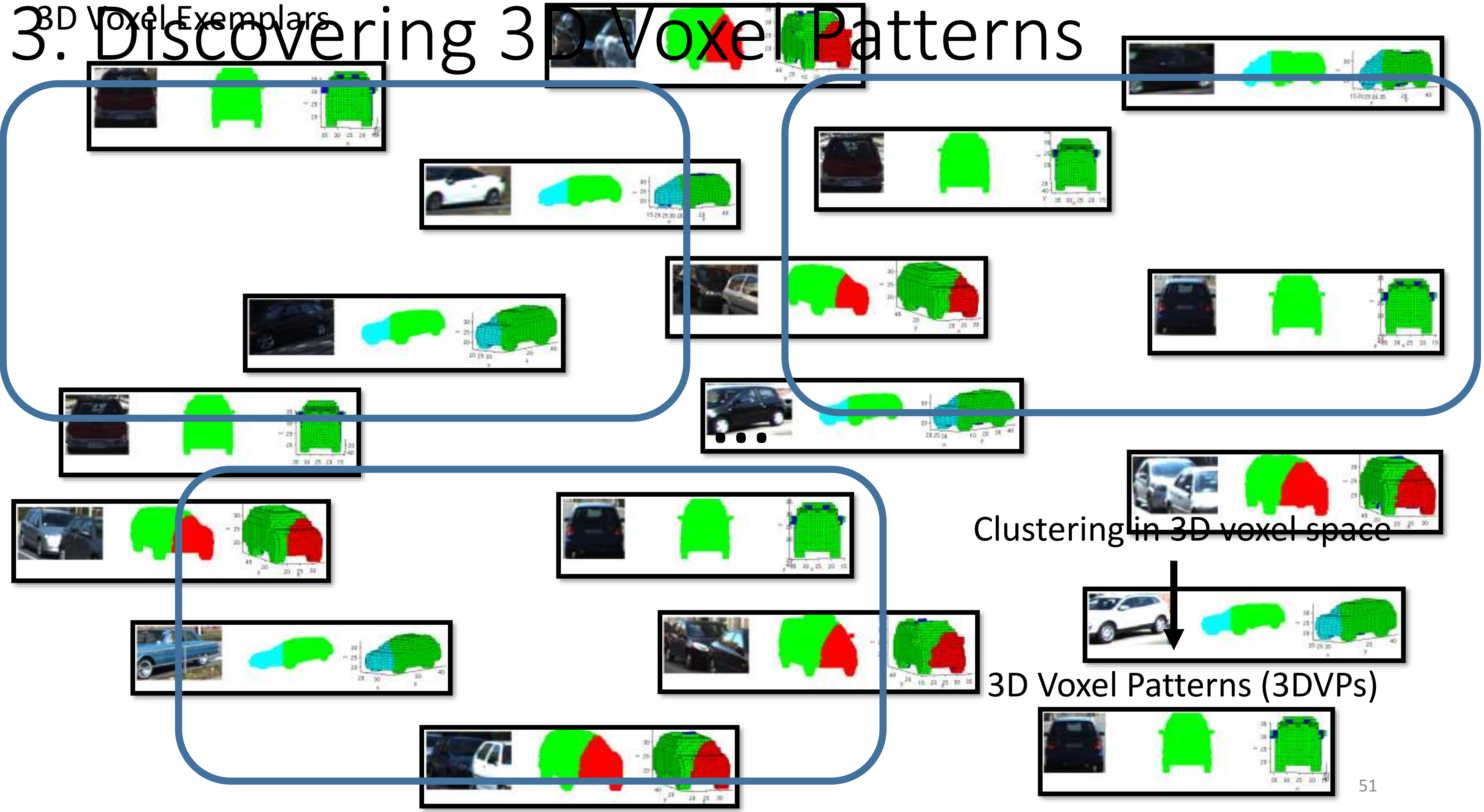


2. Building 3D Voxel Exemplars

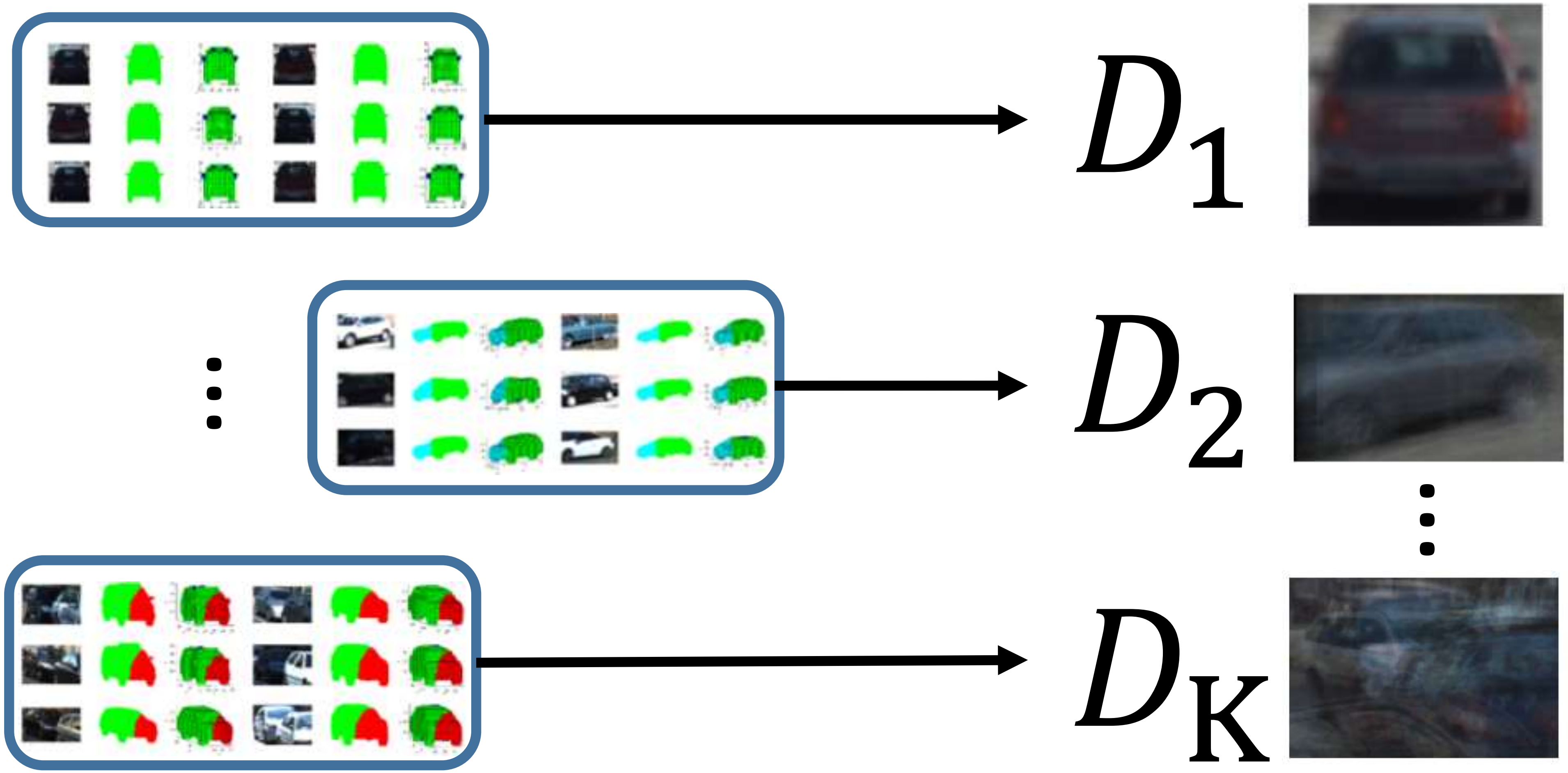
A 3D voxel exemplar $E_i = (I_i, M_i, V_i)$



3. Discovering 3D Voxel Patterns



4. Training 3D Voxel Pattern detectors



- Train a ACF detector for each 3DVP.

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

Testing Pipeline Overview

Input 2D image

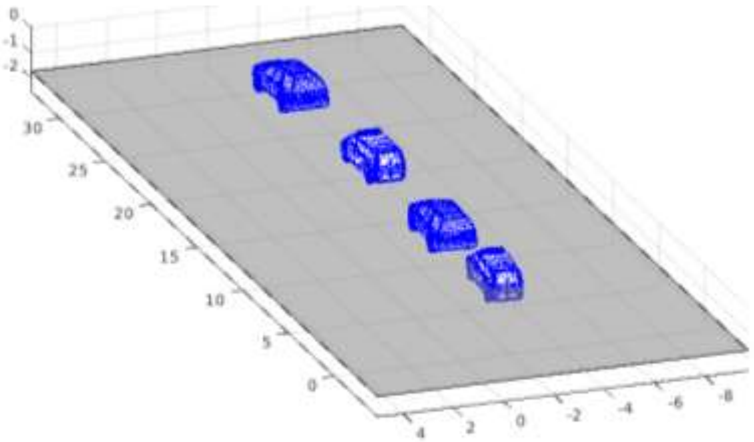
1. Apply 3DVP detectors

2D detection

2. Transfer meta-data
3. Occlusion reasoning

2D segmentation

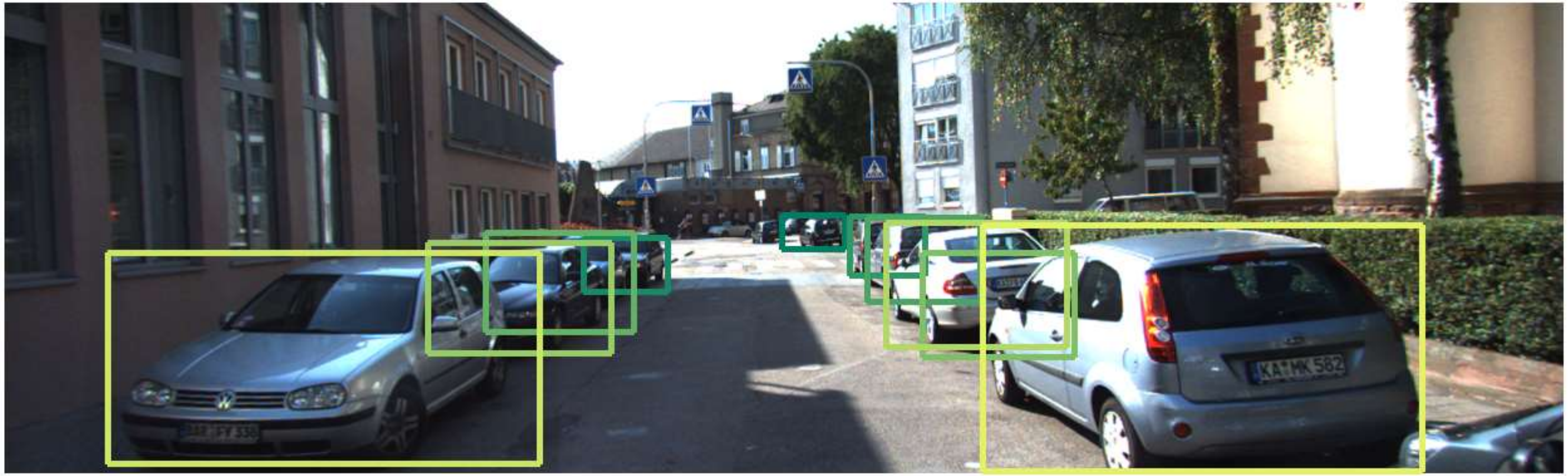
4. Backproject to 3D



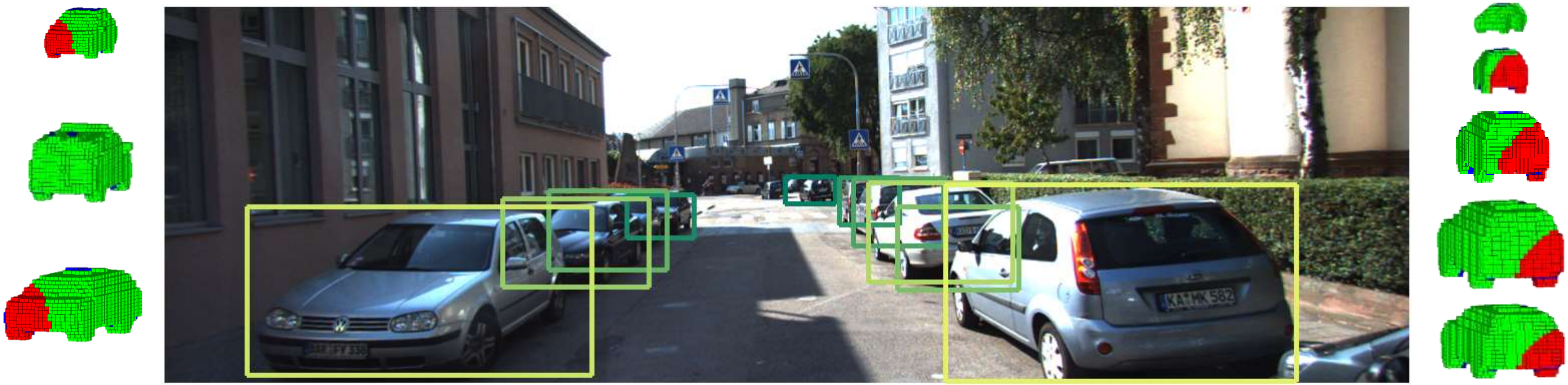
3D localization

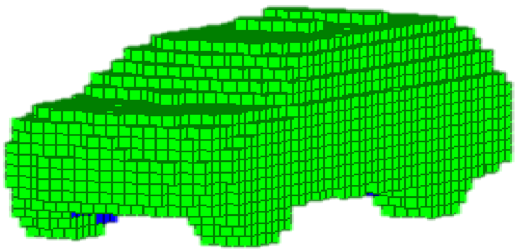
1. Apply 3DVP Detectors

1. Apply 3DVP Detectors

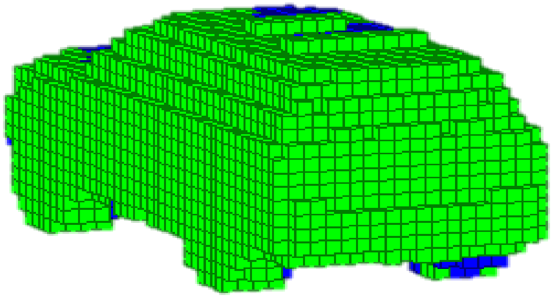


2. Transfer Meta-Data





3D Voxel Patterns



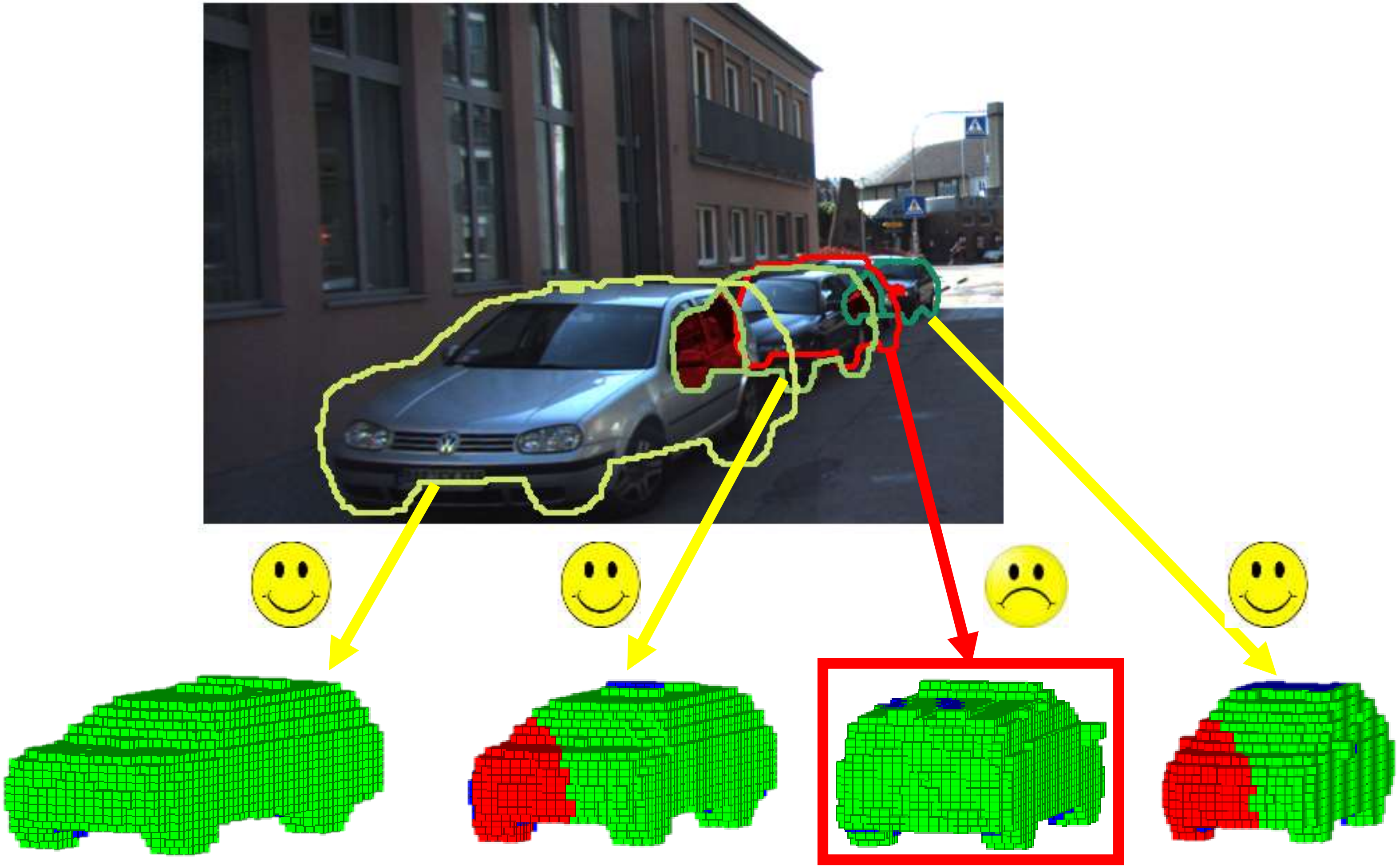
2. Transfer Meta-Data

3. Occlusion Reasoning

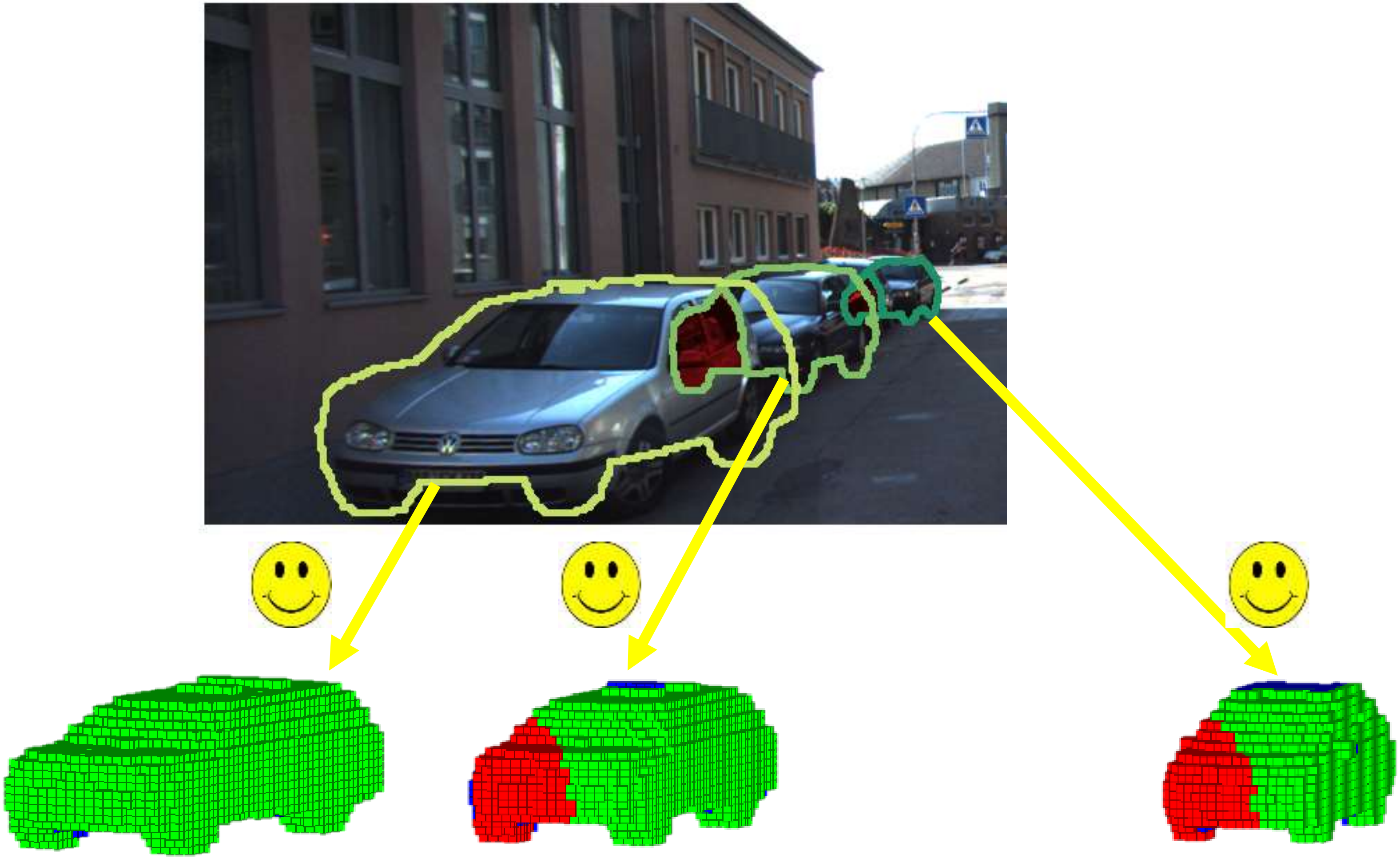
Occlusion reasoning: find a set of visibility-compatible detections

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

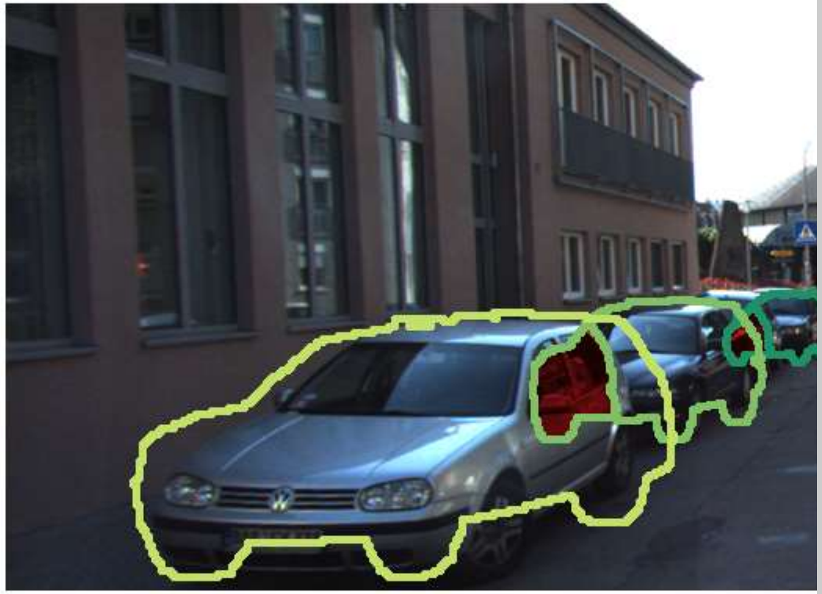
3. Occlusion Reasoning



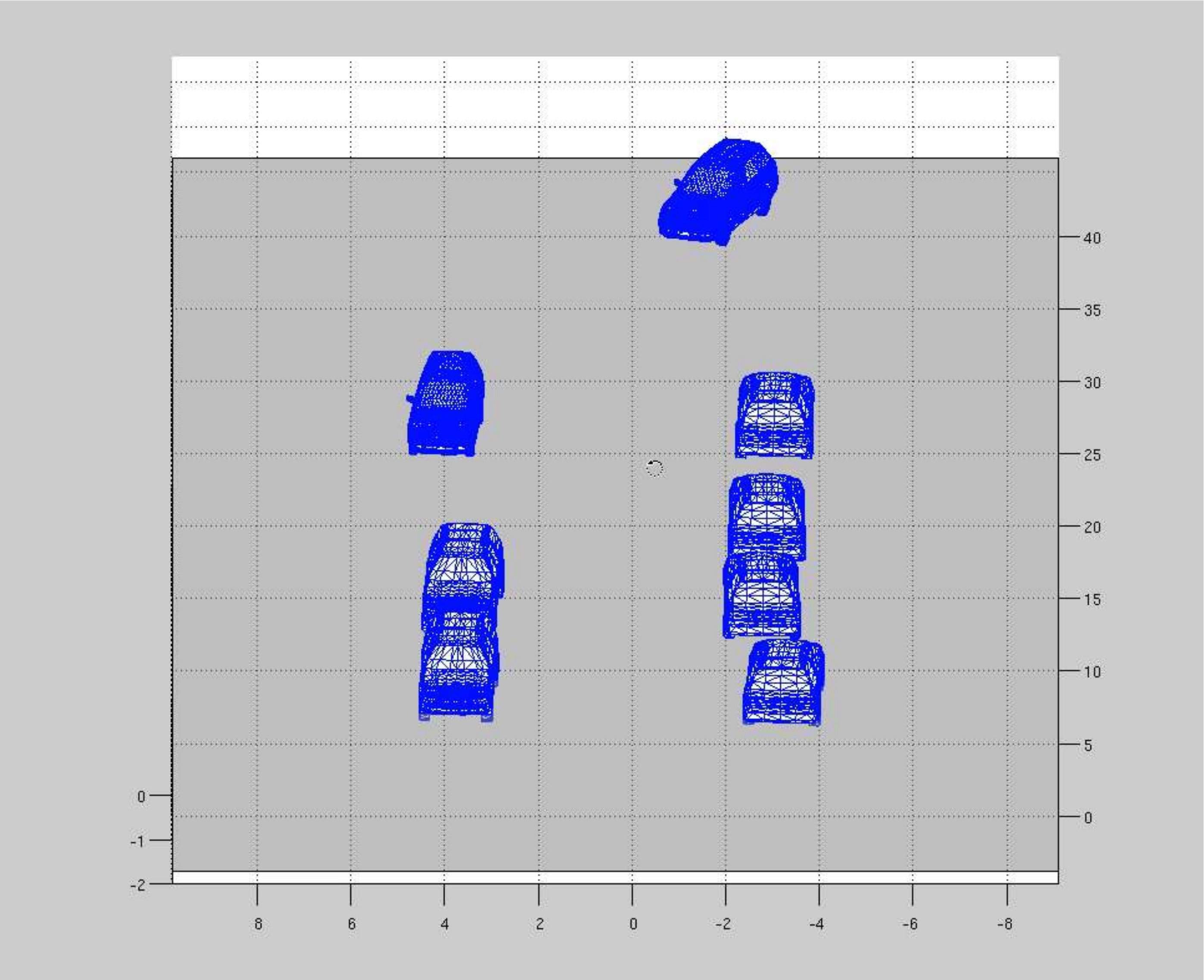
3. Occlusion Reasoning



4. 3D Localization



Backprojection



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]	68.02	56.48	44.18		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]	84.14	75.46	59.71		83.41	**74.42**	58.83
Regionlets [6]	84.75	**76.45**	59.70		N/A	N/A	N/A
AOG [7]	84.80	75.94	60.70		33.79	30.77	24.75
Ours 3DVP	**84.81**	73.02	**63.22**		**84.31**	71.99	**62.11**

[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. Learning to detect vehicles by clustering appearance patterns. T-ITS, 2015.
[6] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013.
[7] B. Li, T. Wu, and S.-C. Zhu. Integrating context and occlusion for car detection by hierarchical and-or model. In ECCV, 2014.

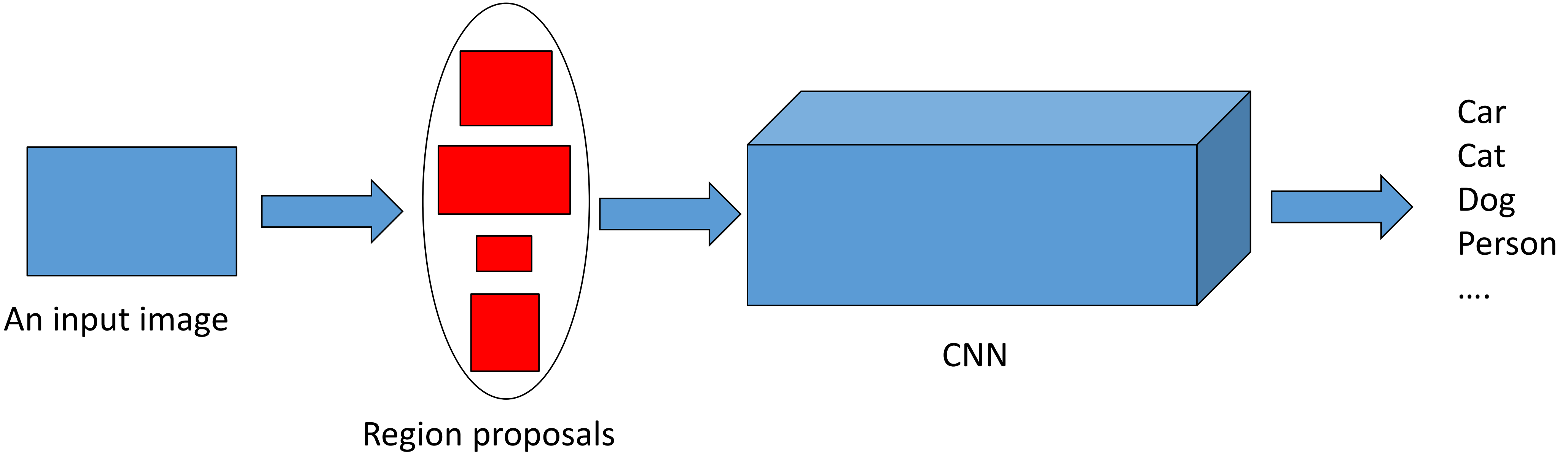
Car Detection and Orientation Estimation on KITTI

	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]	68.02	56.48	44.18		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]	84.14	75.46	59.71		83.41	74.42	58.83
Regionlets [6]	84.75	**76.45**	59.70		N/A	N/A	N/A
AOG [7]	84.80	75.94	60.70		33.79	30.77	24.75
Ours 3DVP	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. Learning to detect vehicles by clustering appearance patterns. T-ITS, 2015.
[6] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013.
[7] B. Li, T. Wu, and S.-C. Zhu. Integrating context and occlusion for car detection by hierarchical and-or model. In ECCV, 2014.

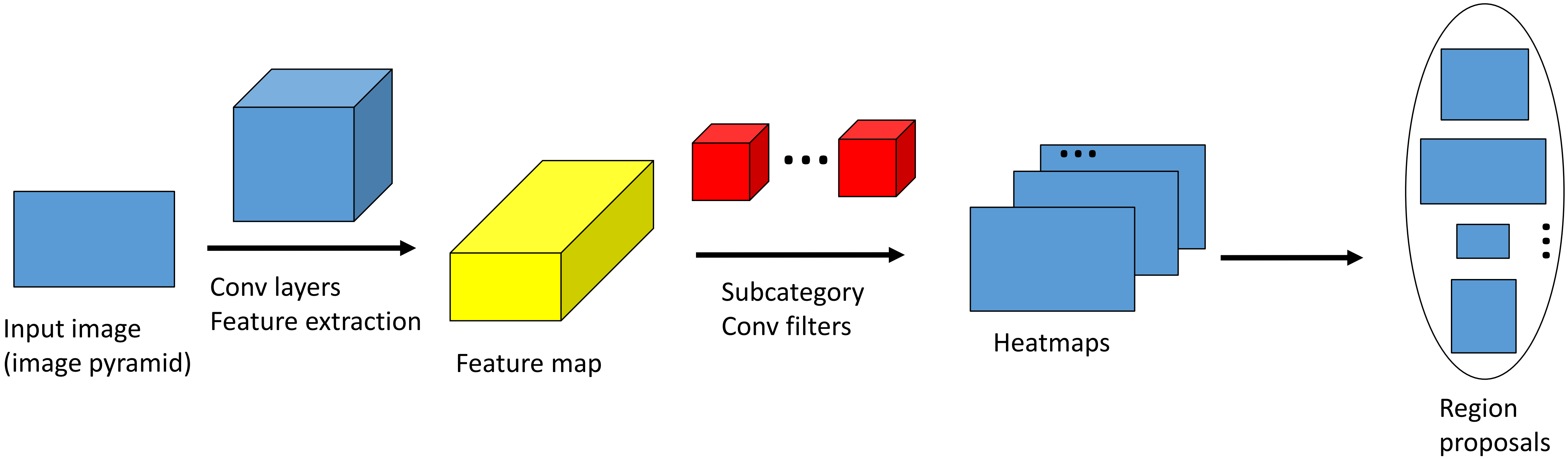
Apply 3DVPs in CNN-based Object Detection

- 3DVPs as subcategories
- Two-stage detection framework



- R. Girshick et al., CVPR’14
- R. Girshick, ICCV’15
- S. Ren et al., NIPS’15
- S. Gidaris and N. Komodakis, CoRR’15

Subcategory-ware Region Proposal Network

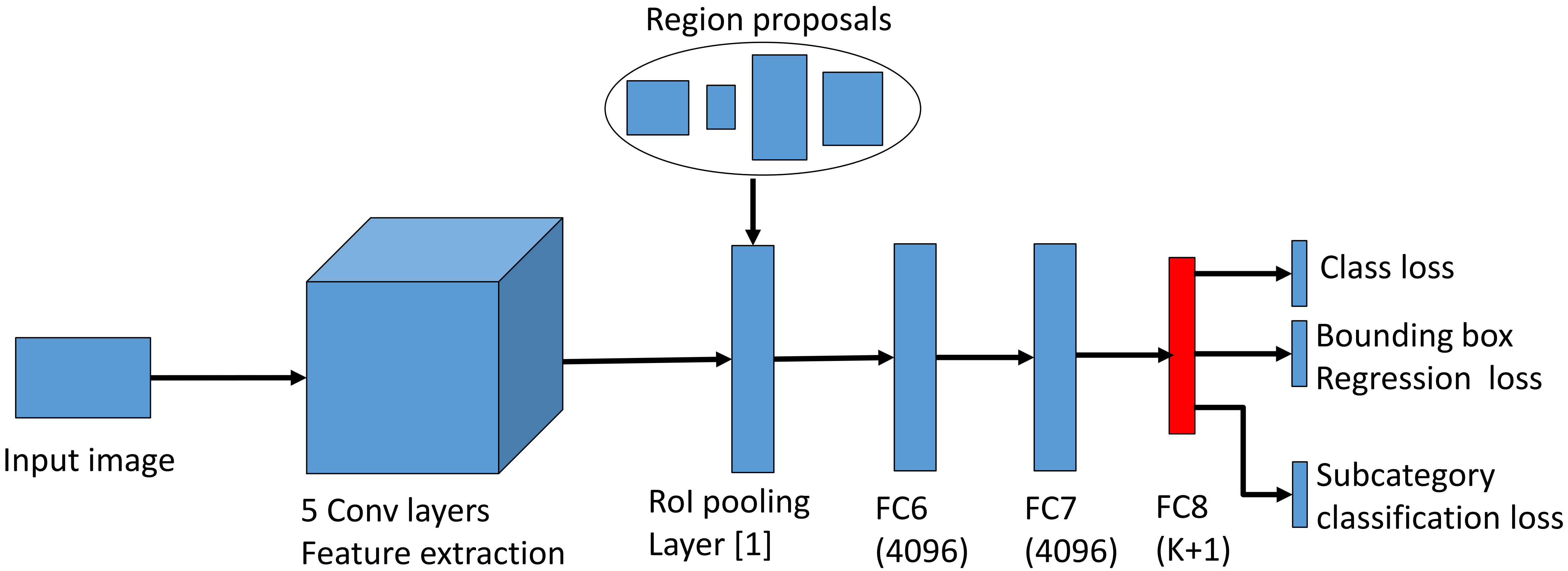


Feature Extrapolating Layer
- Generate features in nearby scales by extrapolating

RoI Generating Layer
- Training: hard positives and hard negatives
- Testing: high score boxes

arXiv:1604.04693

Subcategory-ware Detection Network



arXiv:1604.04693

[1] R. Girshick. Fast R-CNN. ICCV, 2015.

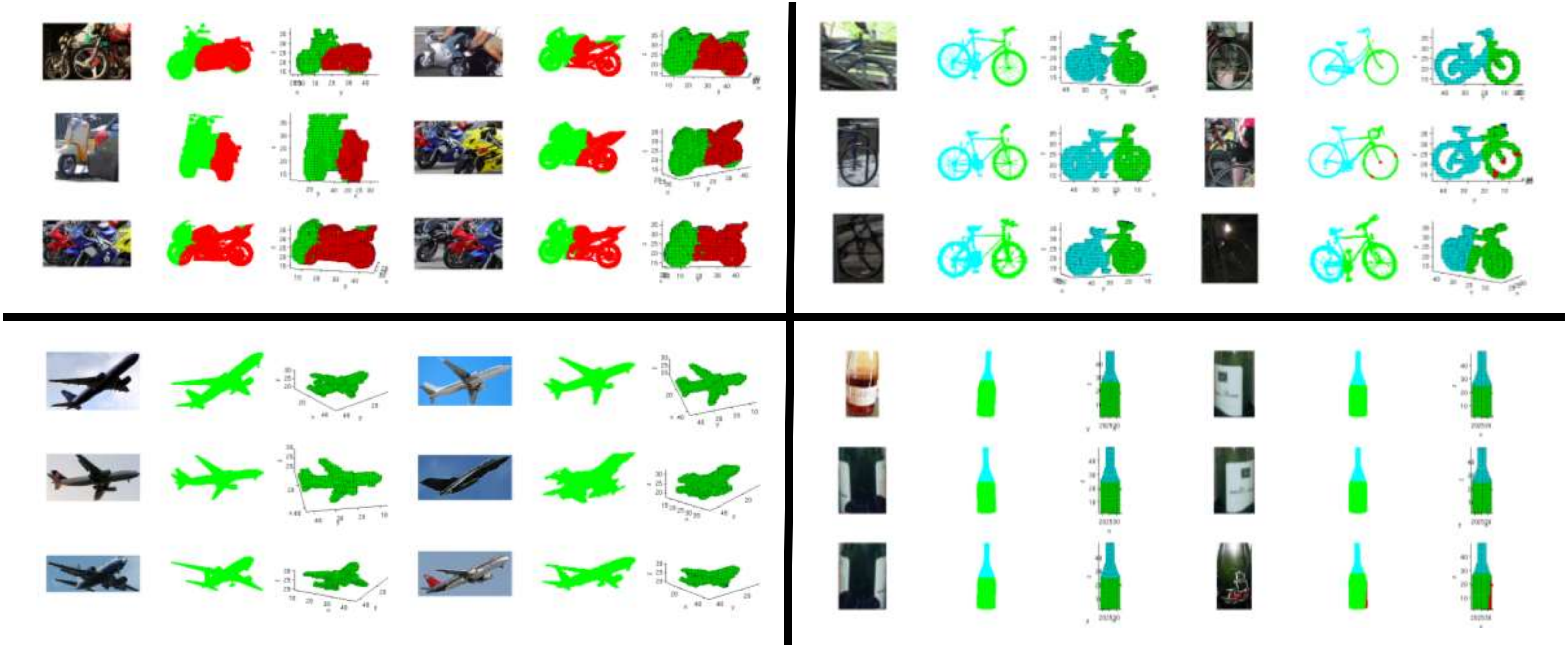
Car Detection and Orientation Estimation on KITTI

	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]	68.02	56.48	44.18	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]	84.14	75.46	59.71	83.41	74.42	58.83
Regionlets [6]	84.75	76.45	59.70	N/A	N/A	N/A
AOG [7]	84.80	75.94	60.70	33.79	30.77	24.75
Mono3D [8]	**92.33**	88.66	78.96	**91.01**	86.62	76.84
Ours 3DVP	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
Ours CNN	90.86	**89.32**	**79.33**	90.72	**88.97**	**78.83**

[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. Learning to detect vehicles by clustering appearance patterns. T-ITS, 2015.
[6] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013.
[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] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, R. Urtasun. Monocular 3D Object Detection for Autonomous Driving, in CVPR, 2016.

Detection: Rank 3
Pose : Rank 1

3D Voxel Patterns from PASCAL3D+ [1]



12 Rigid Categories

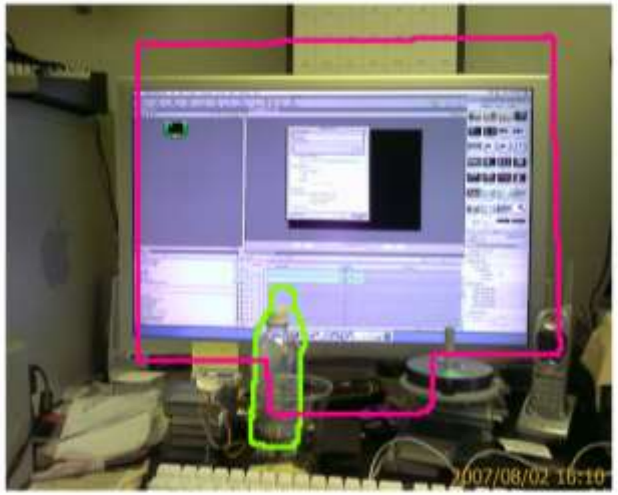
[1] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In WACV, 2014.

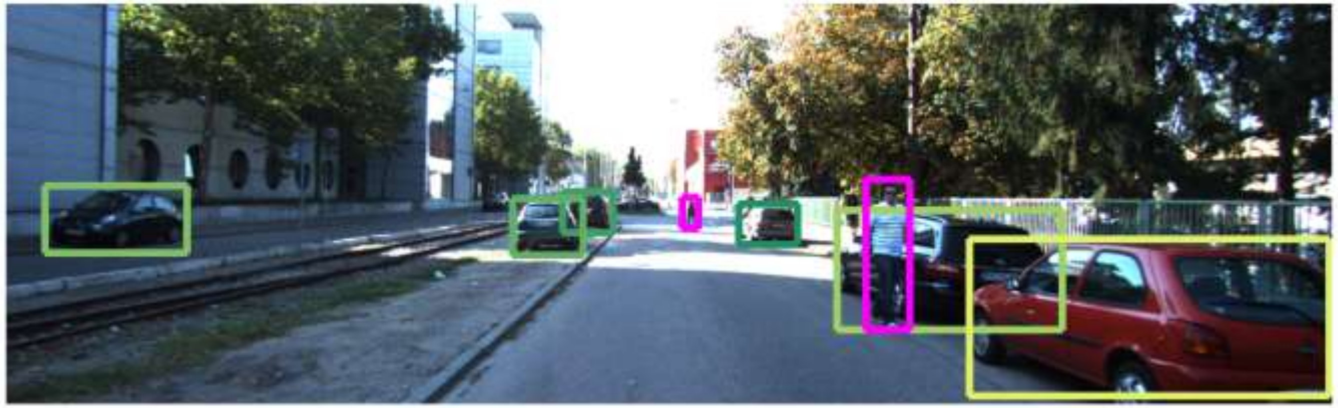
Detection and Pose Estimation on PASCAL3D+

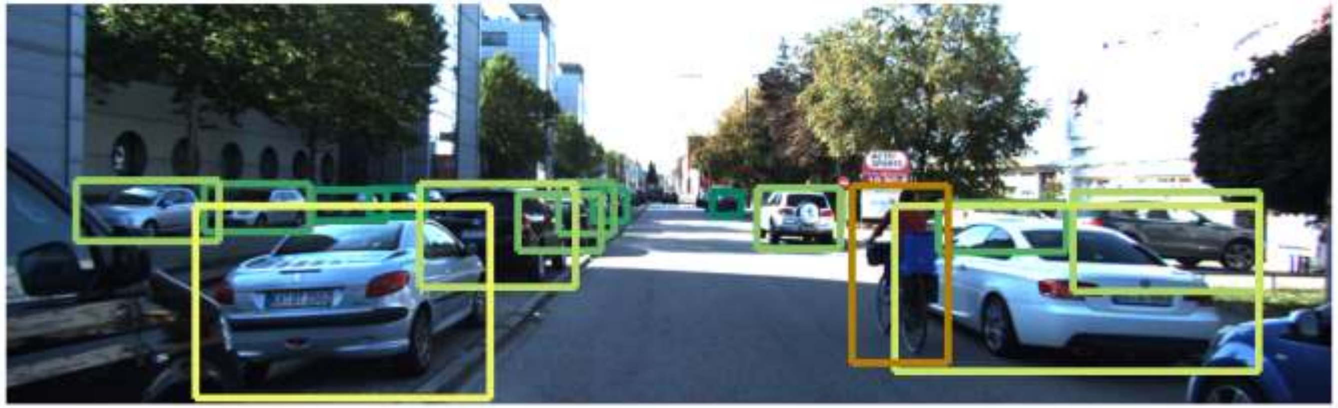
Method	Detection (AP)
DPM [1]	29.6
R-CNN [2]	56.9
Ours CNN	**60.7**

Method	4 Views (AVP)	8 Views (AVP)	16 Views (AVP)	24 Views (AVP)
VDPM [3]	19.5	18.7	15.6	12.1
DPM-VOC+VP [4]	24.5	22.2	17.9	14.4
Ours CNN	**47.5**	**31.9**	**24.5**	**19.3**

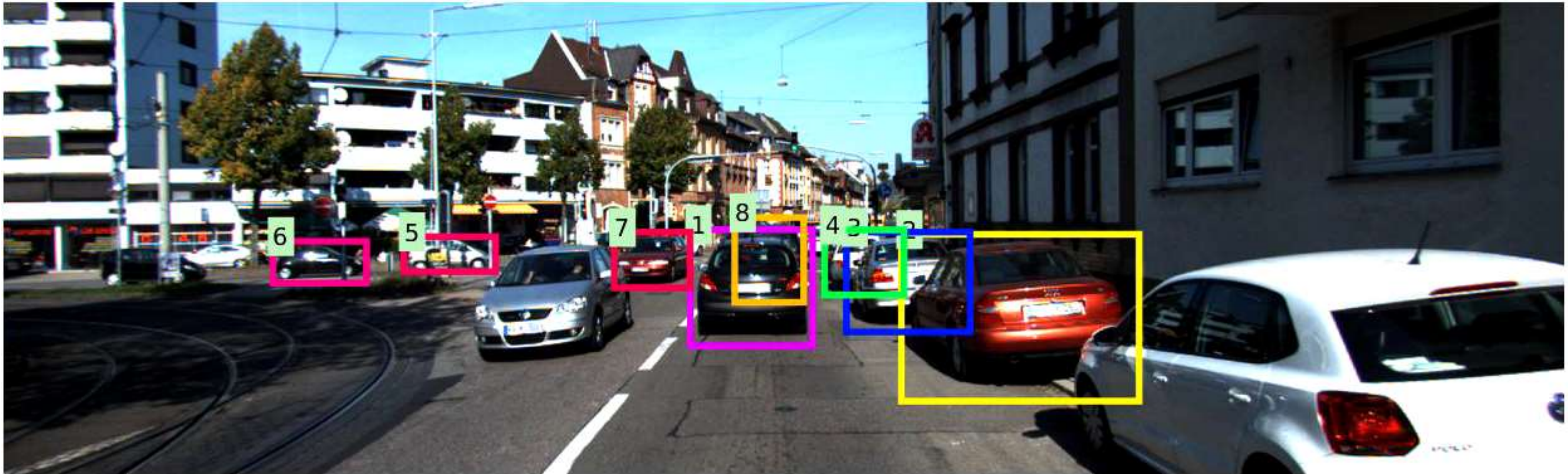
[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv preprint arXiv:1311.2524, 2013.
[3] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In WACV, 2014.
[4] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Multi-view and 3d deformable part models. TPAMI, 2015.







Application: Online Multi-Object Tracking



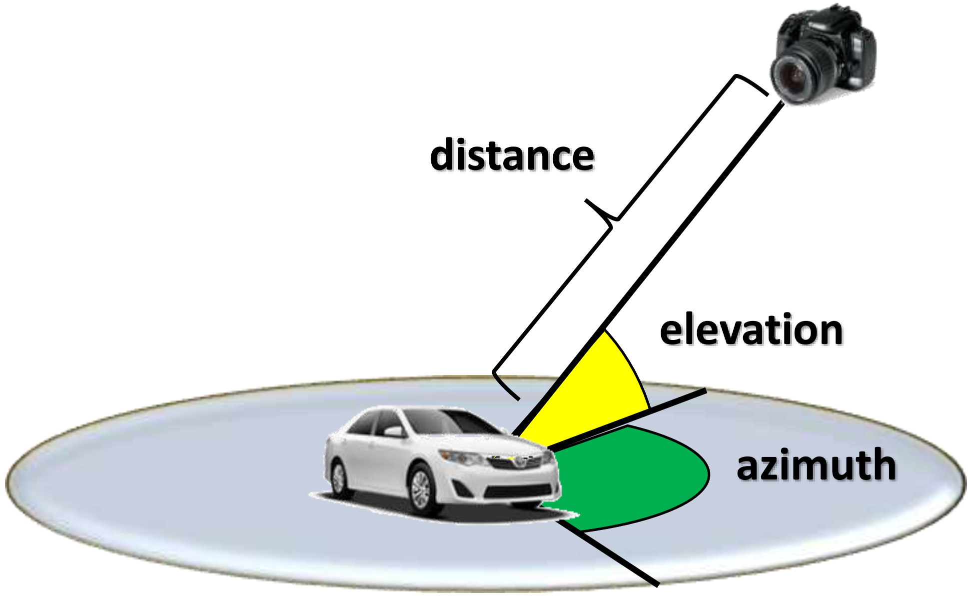
Y. Xiang, A. Alahi and S. Savarese. Learning to Track: Online Multi-Object Tracking by Decision Making. In ICCV, 2015.

Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Conclusion and Future Work

Goal

- Build a large scale dataset for 3D object recognition in the wild



3D Object Dataset

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘

[1] S. Savarese and L. Fei-Fei. 3d generic object categorization, localization and pose estimation. In ICCV, 2007.

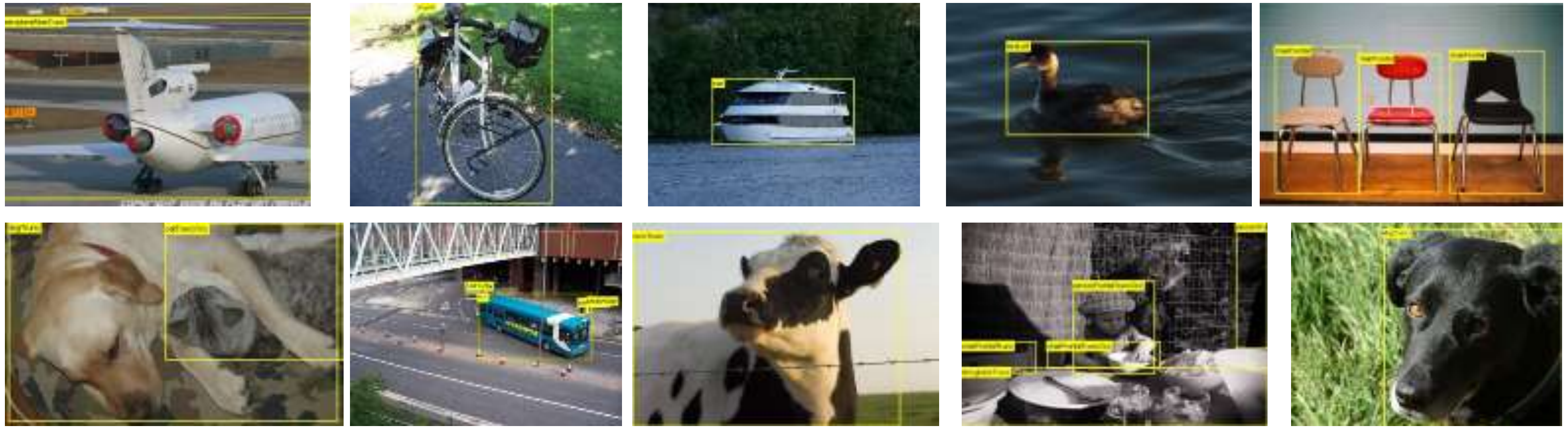
EPFL Car Dataset

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘
EPFL Car [2]	1	20	✘	✓	✘

[2] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In CVPR, 2009.

PASCAL VOC Dataset

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘
EPFL Car [2]	1	20	✘	✓	✘
PASCAL VOC [3]	20	27,450	✓	✘	✘



[3] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 2010.

KITTI Dataset

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘
EPFL Car [2]	1	20	✘	✓	✘
PASCAL VOC [3]	20	27,450	✓	✘	✘
KITTI [4]	3	80,256	✓	✓	✘

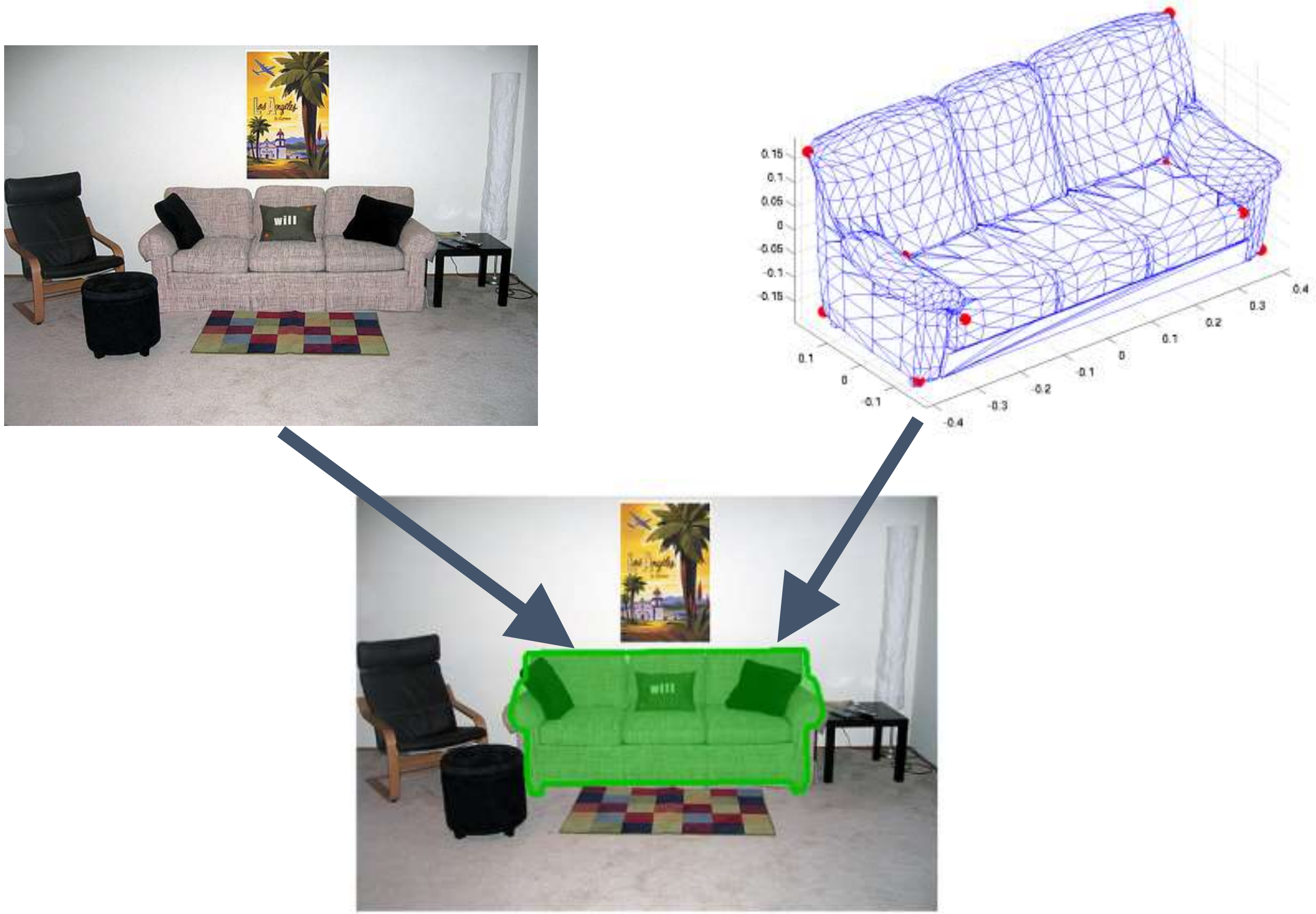
[4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.

Our Contribution: PASCAL3D+

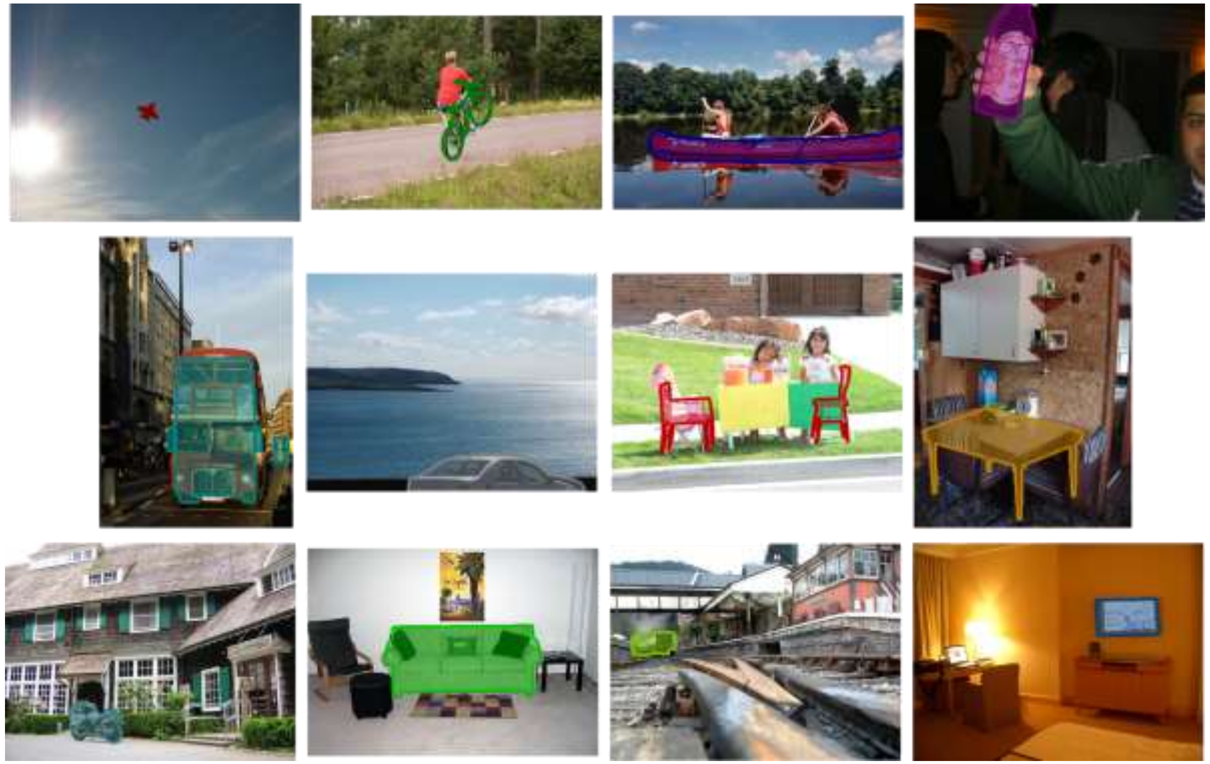
	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✗	✗	✗
EPFL Car [2]	1	20	✗	✓	✗
PASCAL VOC [3]	20	27,450	✓	✗	✗
KITTI [4]	3	80,256	✓	✓	✗
PASCAL3D+ (Ours)	**12**	**35,672**	✓	✓	✓

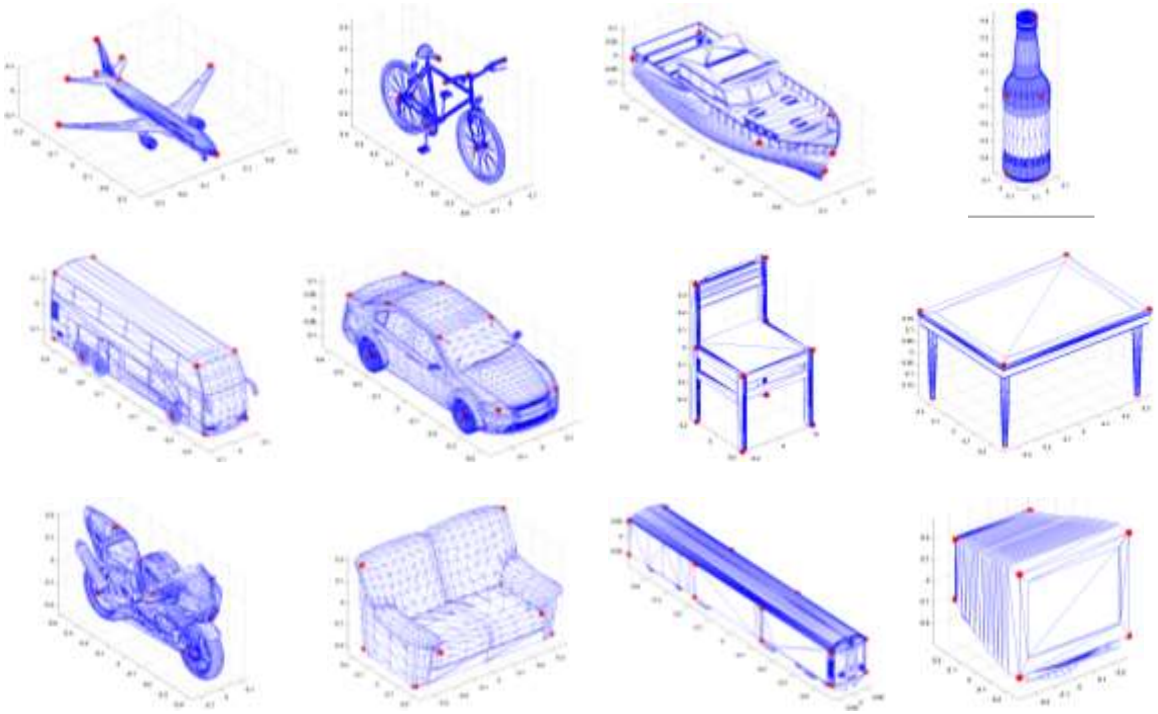
Y. Xiang, R. Mottaghi and S. Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In WACV, 2014.

3D Annotation: 2D-3D Alignment



PASCAL3D+: A Benchmark for 3D Object Recognition





Our Contribution: ObjectNet3D

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✗	✗	✗
EPFL Car [2]	1	20	✗	✓	✗
PASCAL VOC [3]	20	27,450	✓	✗	✗
KITTI [4]	3	80,256	✓	✓	✗
PASCAL3D+ (Ours)	12	35,672	✓	✓	✓79
ObjectNet3D (Ours)	**100**	**178,633**	✓	✓	✓**44,147**

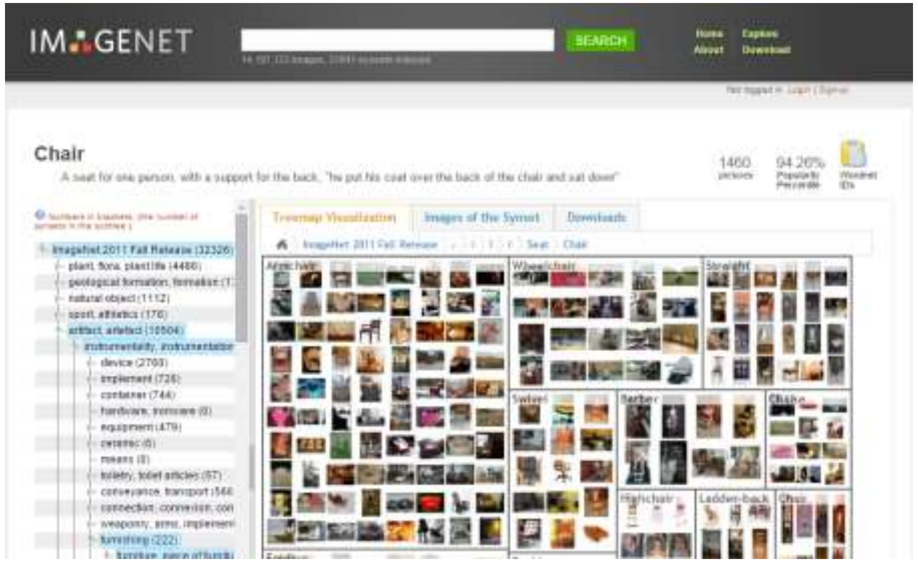




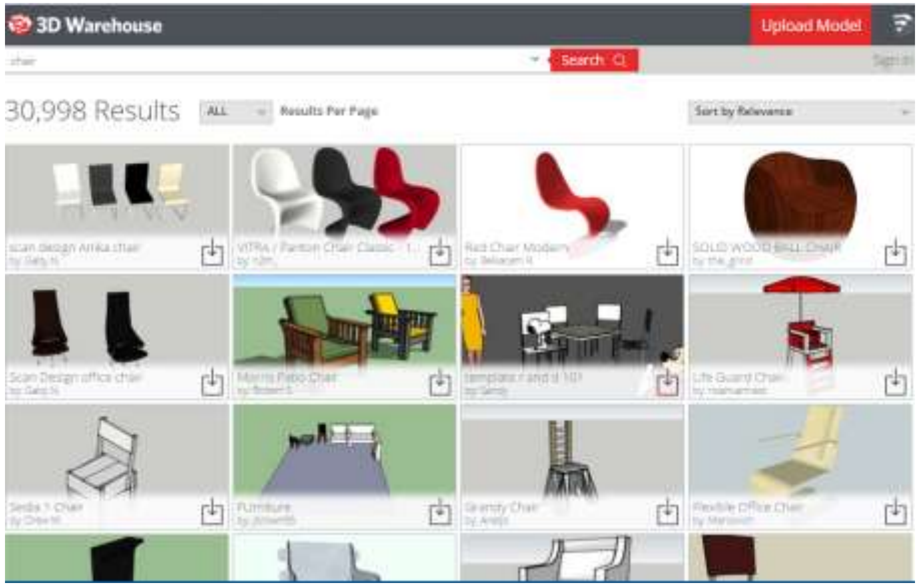


Under Review

ObjectNet3D: A Large Scale Database for 3D Object Recognition



2D-3D alignment

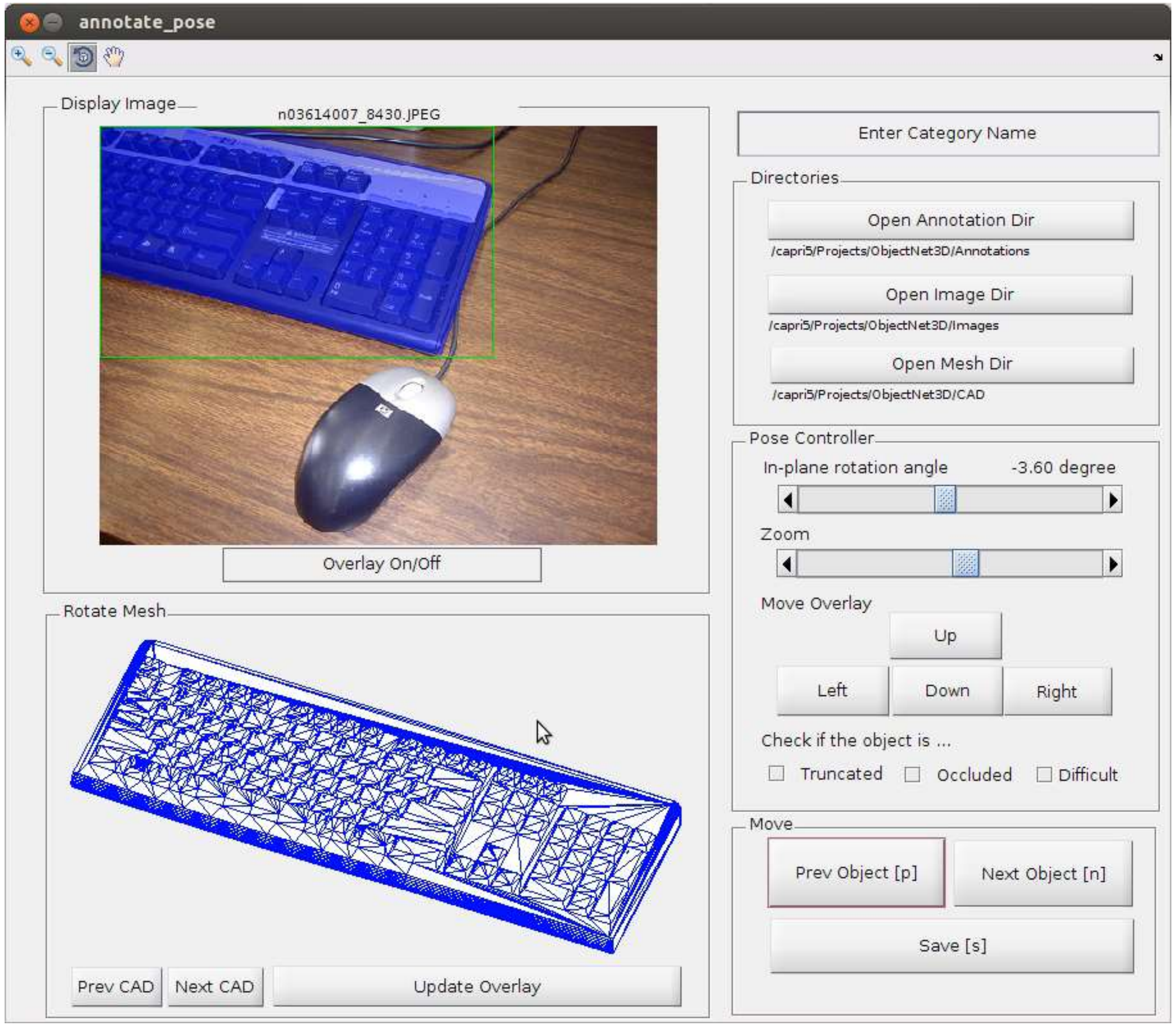


Images from ImageNet

3D Shapes from 3D Warehouse and ShapeNet

100 rigid object categories

Annotation Demo



Viewpoint Distribution

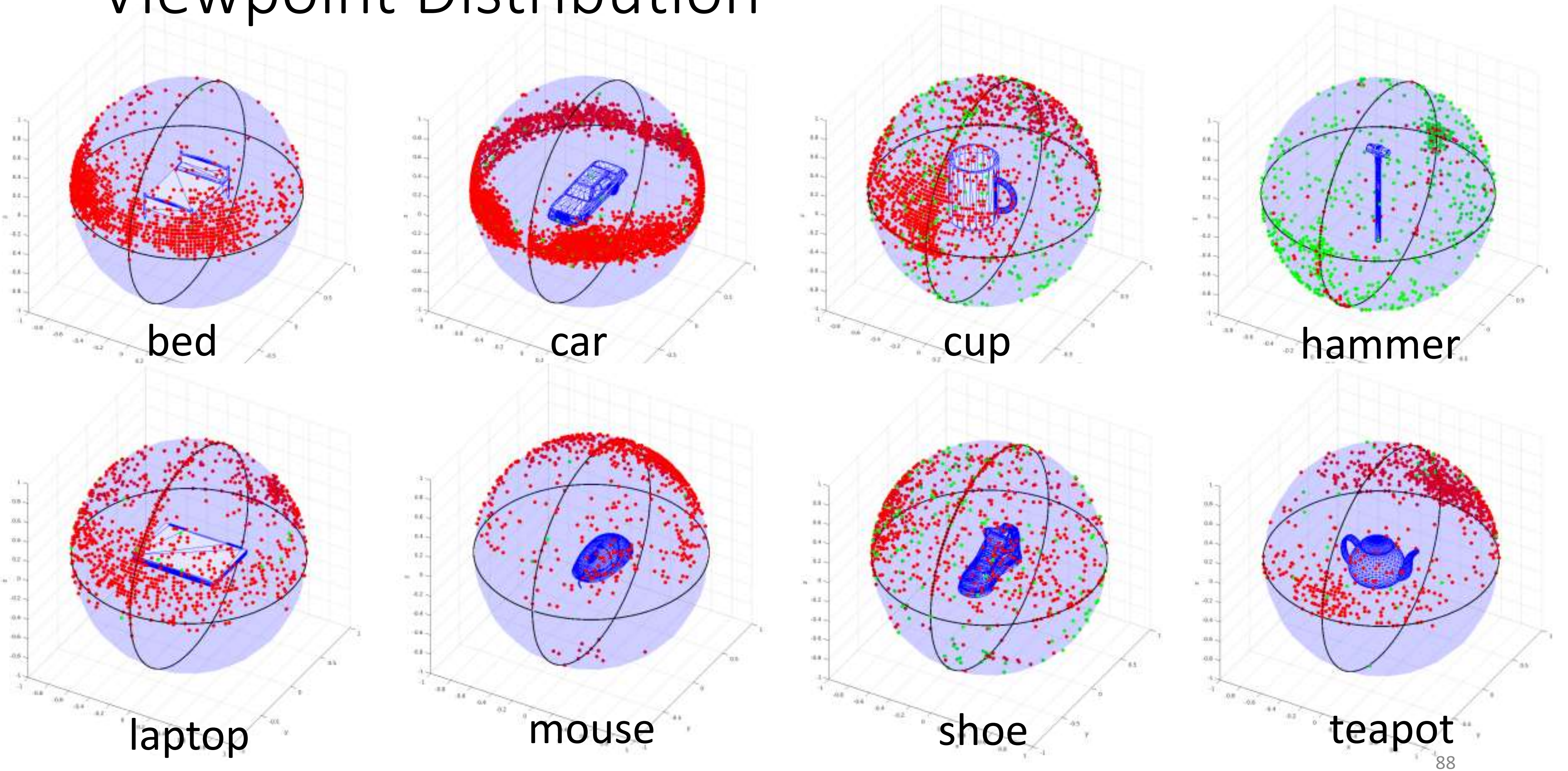
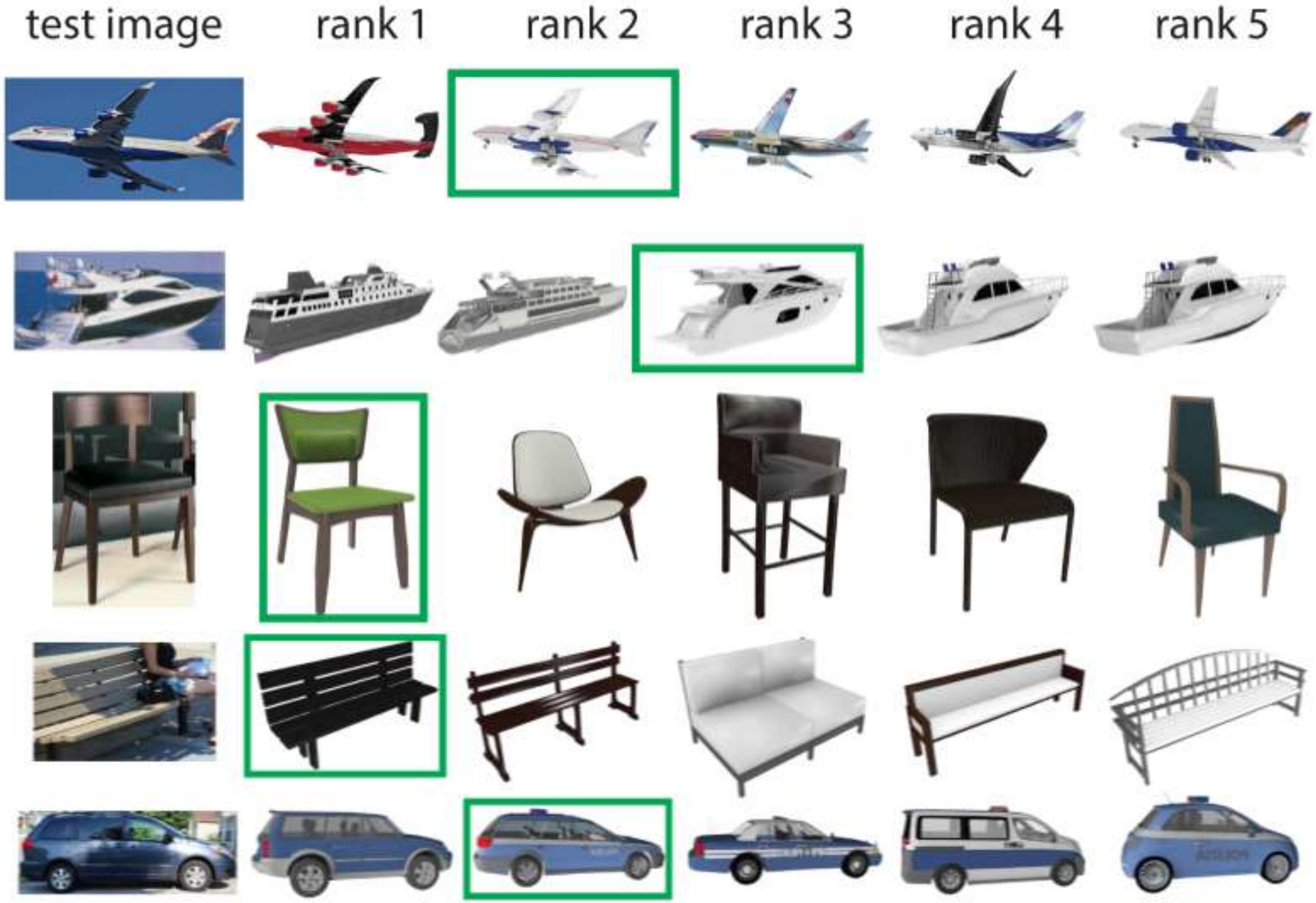


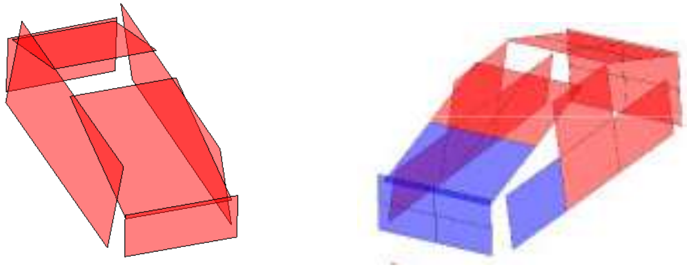
Image-based 3D Shape Retrieval



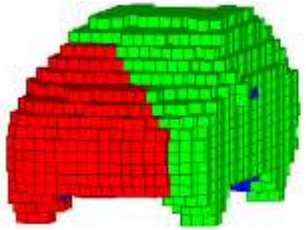
H.O. Song, Y. Xiang, S. Jegelka and S. Savarese. Deep Metric Learning via Lifted Structured Feature Embedding. In CVPR, 2016.

Conclusion

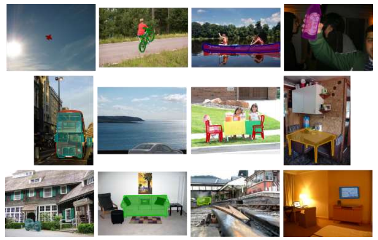
- 3D Aspect Part Representation

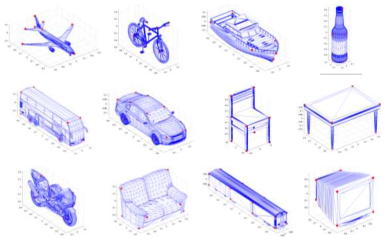


- 3D Voxel Pattern Representation

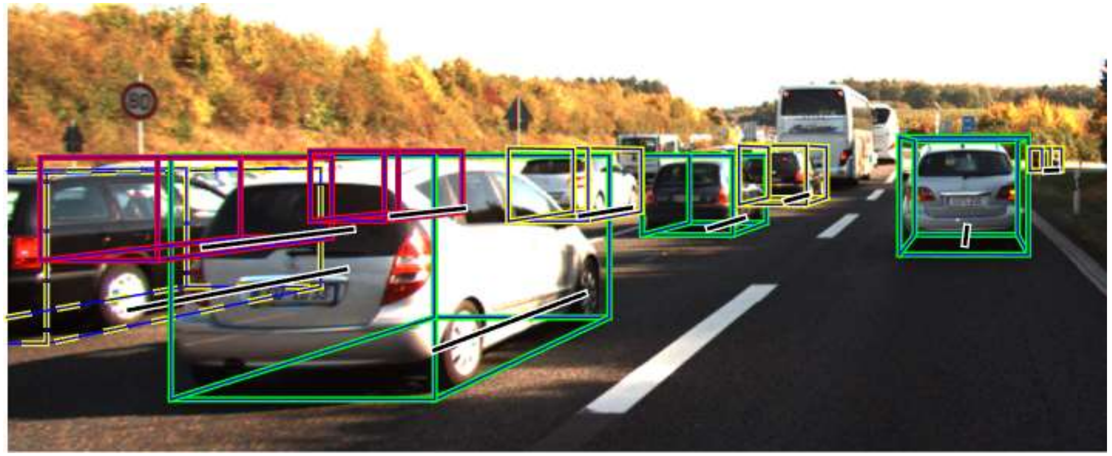


- A Benchmark for 3D Object Recognition in the Wild
  - PASCAL3D+ and ObjectNet3D





Future Work: Generalization of Object Recognition

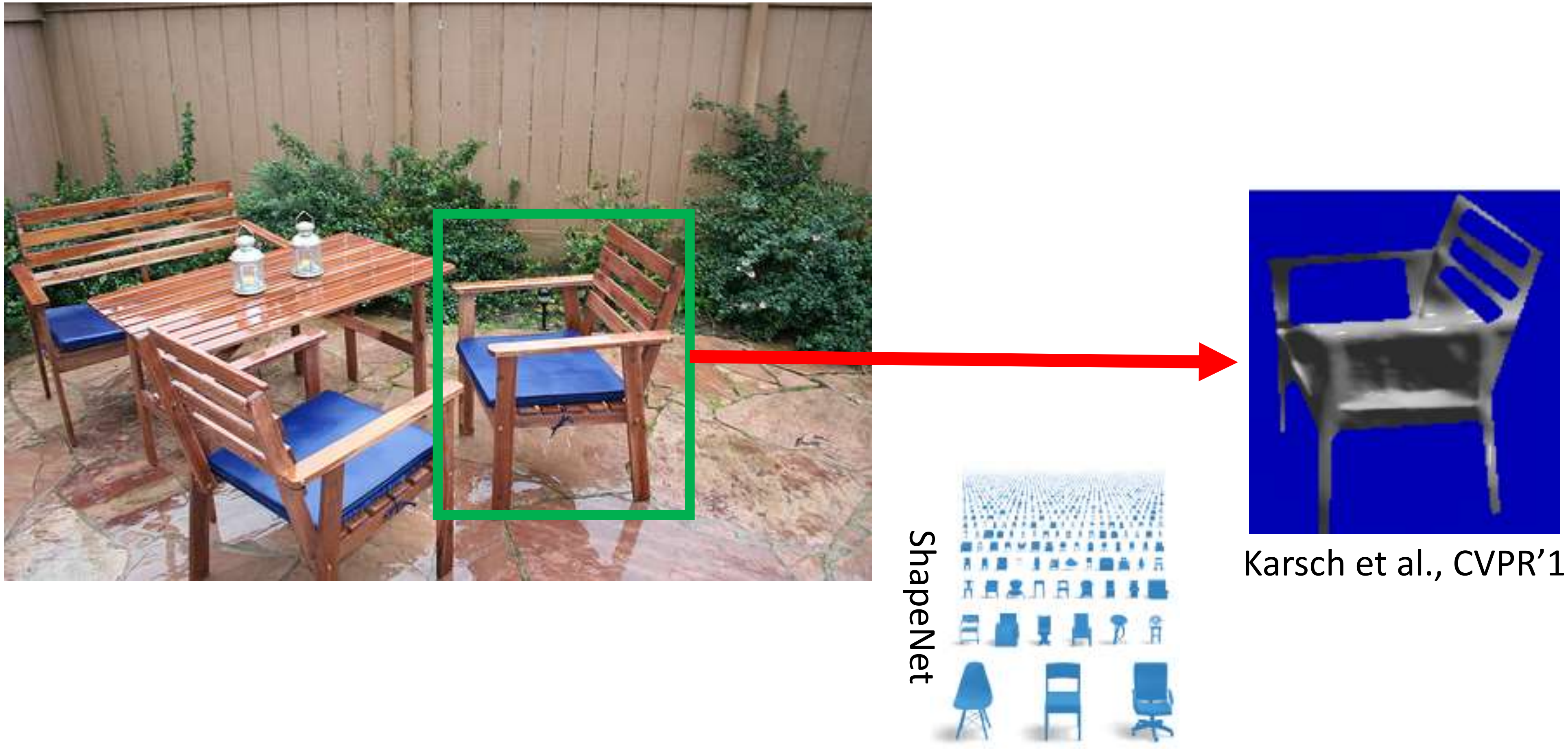


Training Data
(with annotations)

Testing Data
(with few annotations or even no annotation)

- How to achieve generalization across domains?
- Can 3D object representations improve generalization?
- Can we find better 3D object representations for recognition?

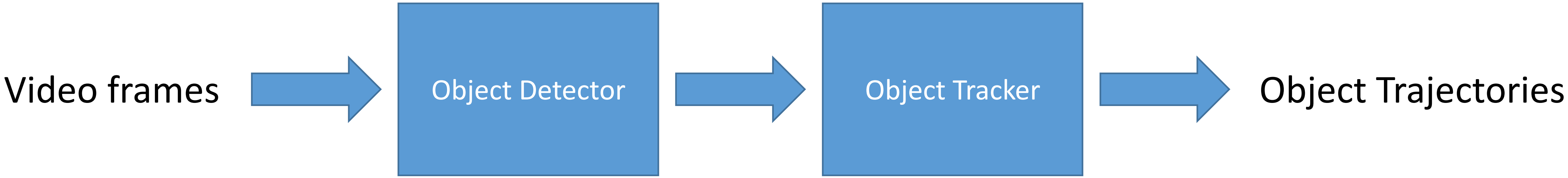
Future Work: Joint Recognition and Reconstruction



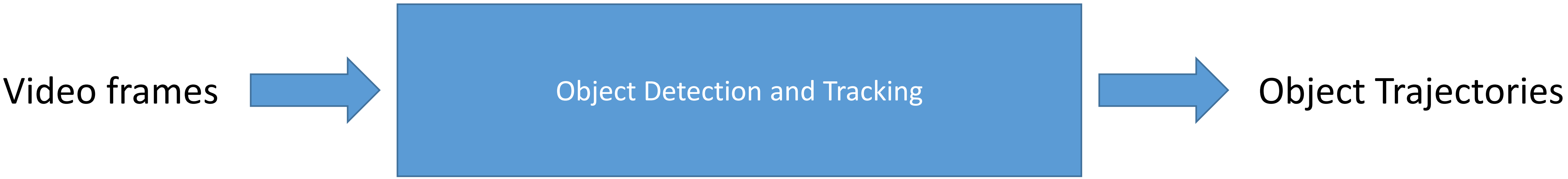
Karsch et al., CVPR’13

Utilize large scale 3D shape data

Future Work: End-to-End Multi-Object Tracking



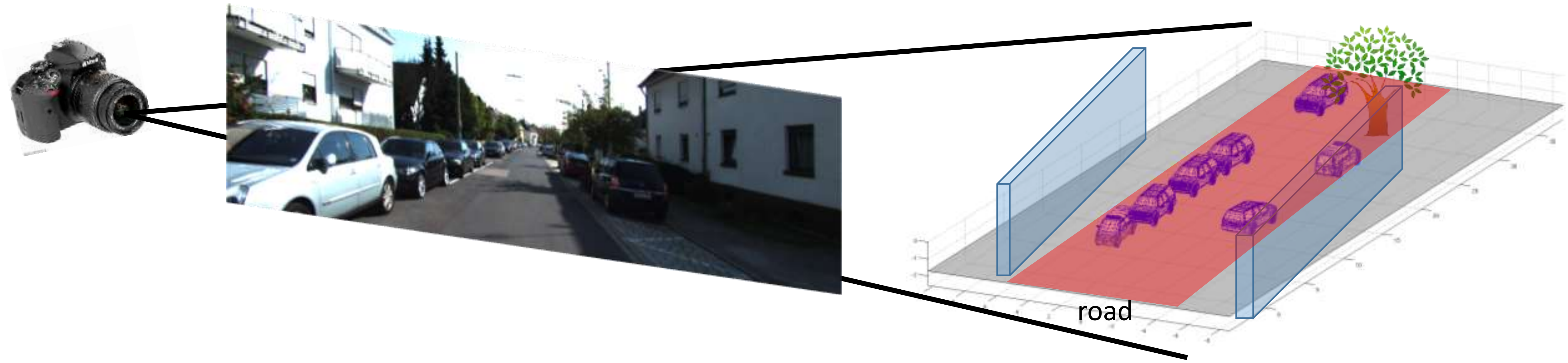
Tracking by Detection



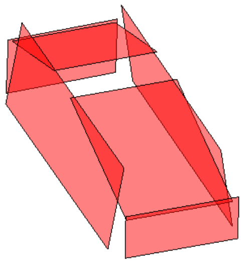
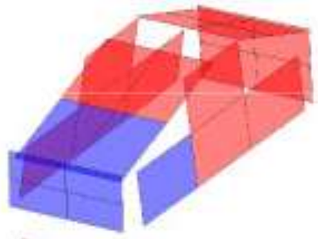
Learning to Detect and Track

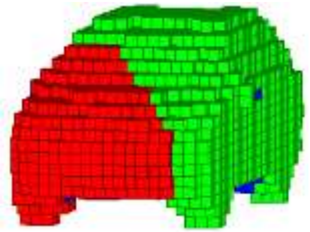
Future Work: Putting Objects in the Scene

- 3D object recognition and scene geometry understanding
  - Holistic 3D scene understanding



Conclusion

- 3D aspect part representation  

- 3D voxel pattern representation 

- A Benchmark for 3D Object Recognition in the Wild

Thank you!