3D Object Recognition

Yu Xiang
University of Washington
Tutorial on 3DV 2016

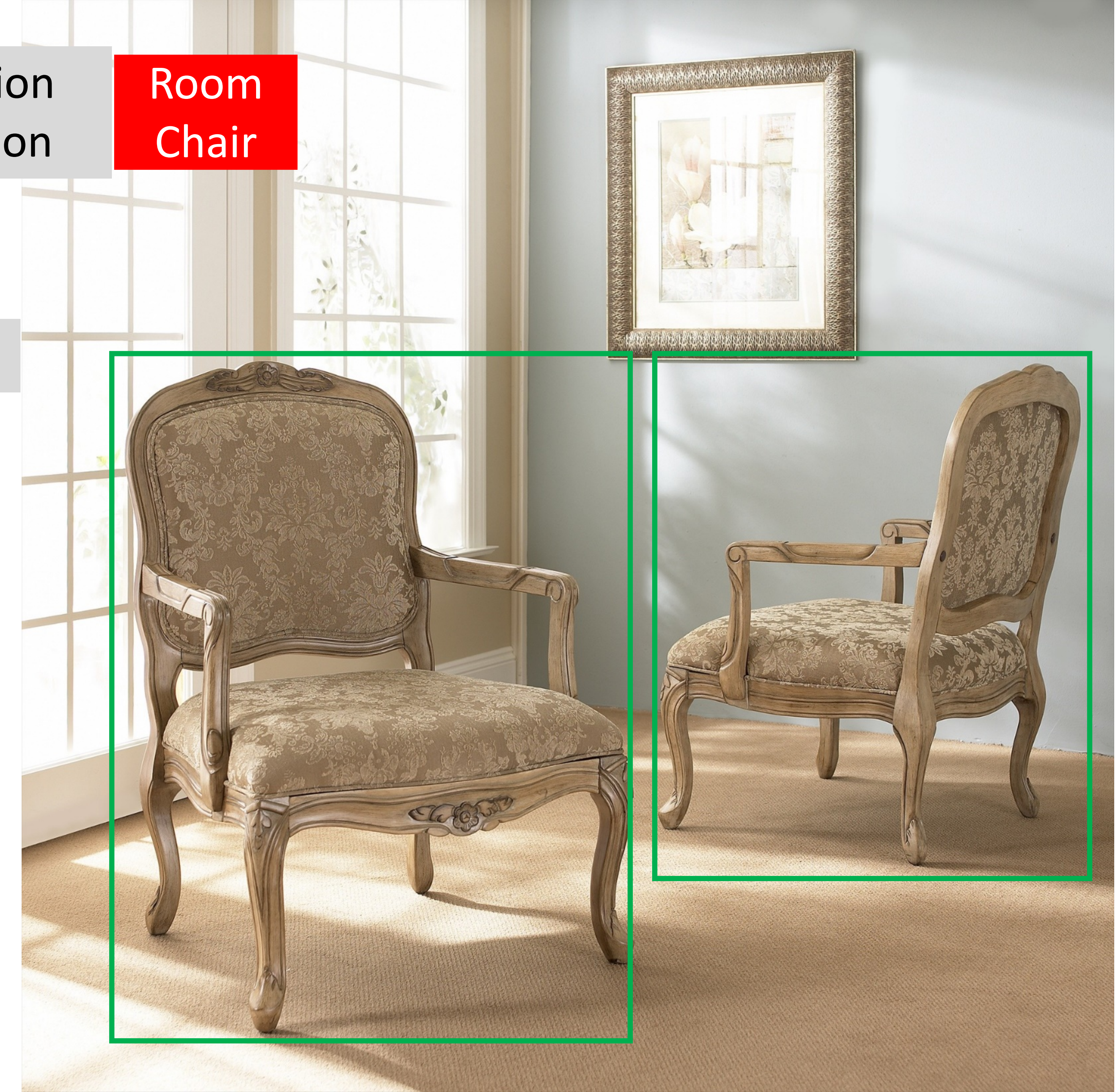
- Image classification tagging/annotation

Room
Chair

Fergus et al. CVPR’03
Fei-Fei et al. CVPRW’ 04
Chua et al. CIVR’09
Xiang et al. CVPR’10
Russakovsky et al. ECCV’12
Ordonez et al. ICCV’13
Deng et al. ECCV’14
…

- Image classification tagging/annotation

- Object detection

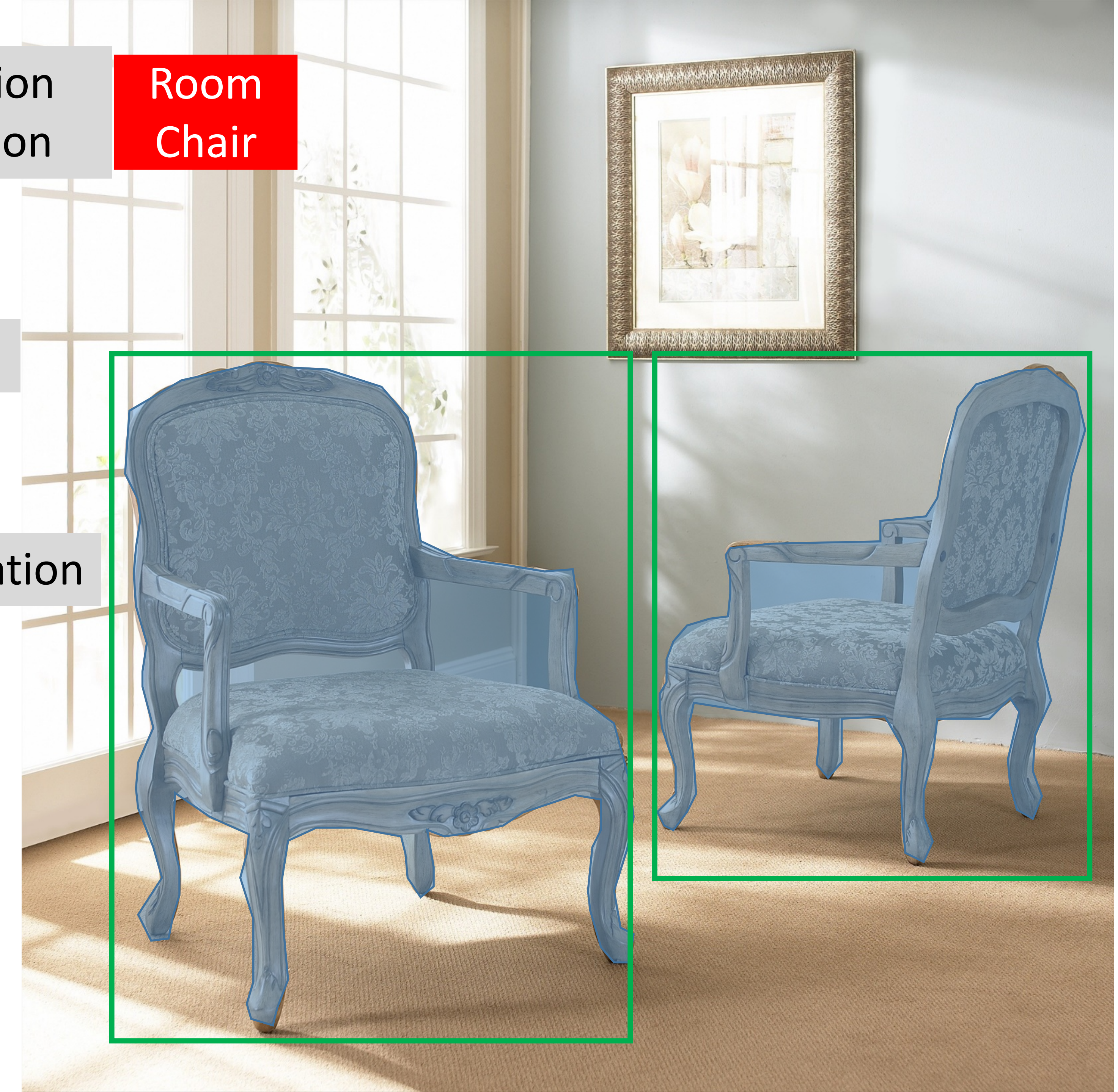


Viola & Jones. IJCV'04
Leibe et al. ECCVW'04
Dalal & Triggs. CVPR'05
Felzenszwalb et al. TPAMI' 10
Girshick et al. CVPR'14
Ren et al. NIPS'15
…

- Image classification tagging/annotation

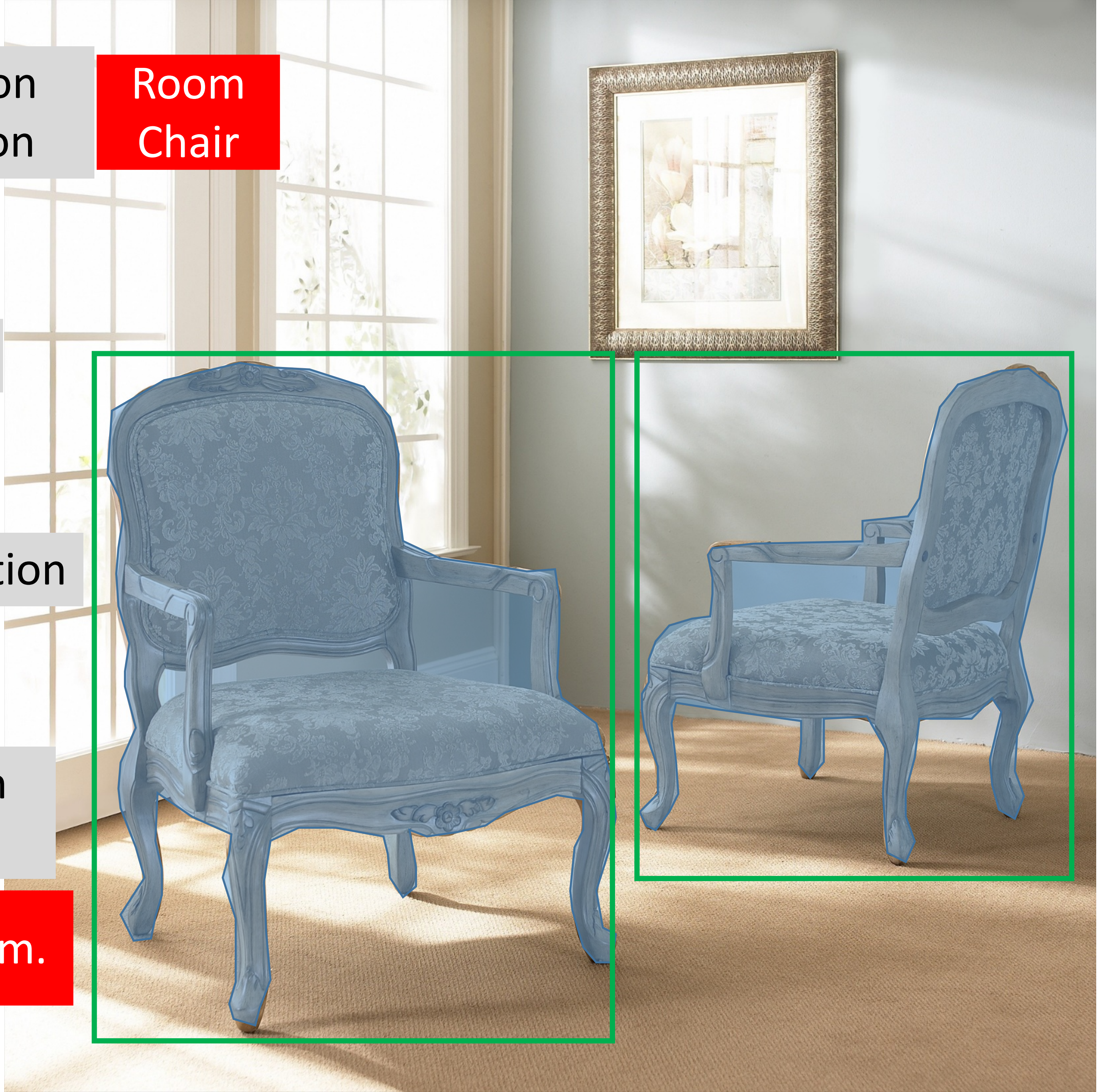
- Object detection

- Object segmentation



Shotton et al. IJCV'07
Pushmeet et al. IJCV'09
Ladicky et al. ECCV'10
Carreira et al. ECCV'12
Chen et al. ICLR'15
Long et al. CVPR'15
...

- Image classification tagging/annotation



- Object detection
- Object segmentation
- Image description generation

Two chairs in a room.

Kulkarni et al. CVPR’11
Karpathy & Fei-Fei. CVPR’15
Chen & Zitnik. CVPR’15
Gregor et al. ICML’15
Johnson et al. CVPR’16
…

- Image classification tagging/annotation

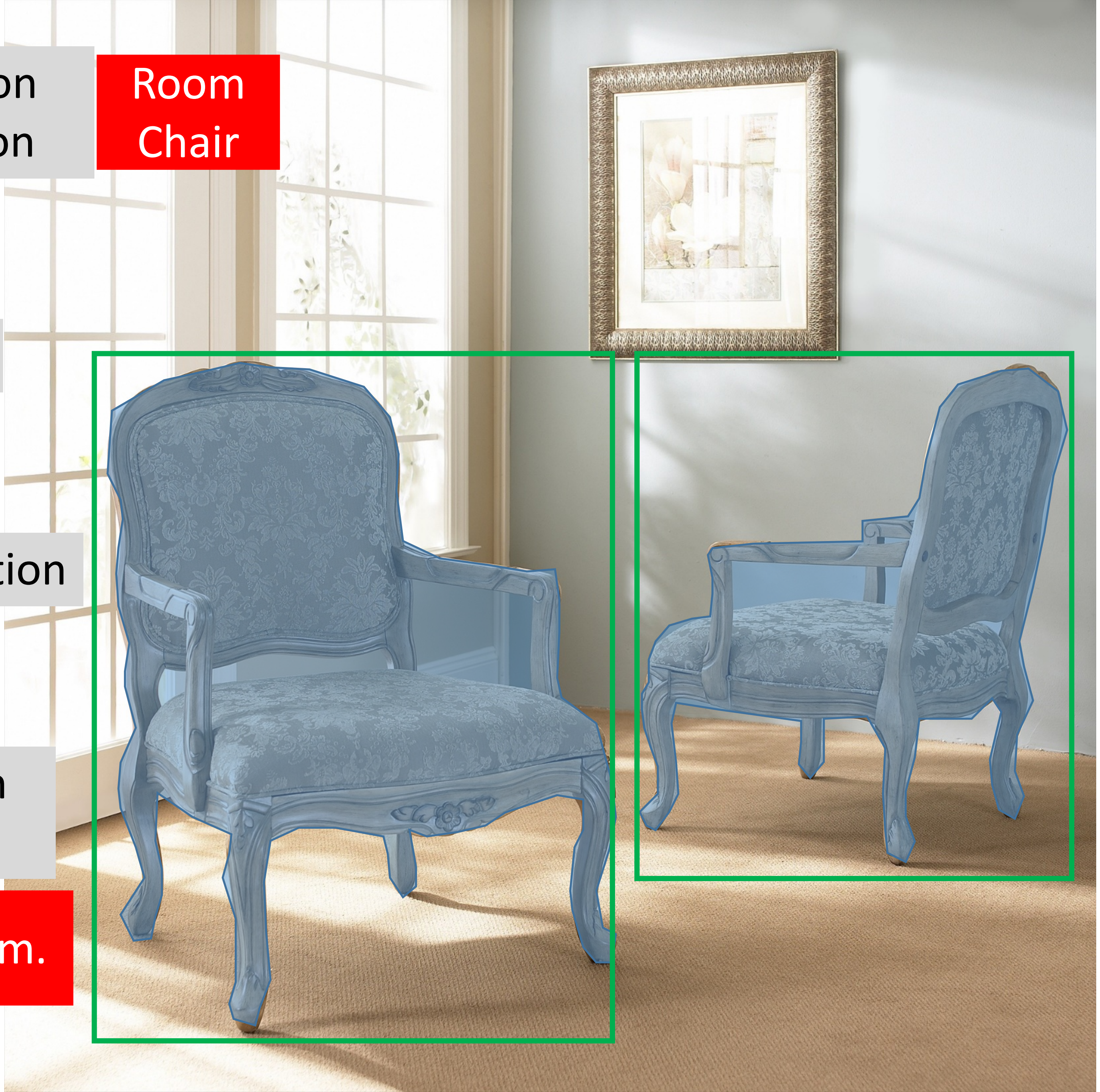
Room
Chair

- Object detection

- Object segmentation

- Image description generation

Two chairs in a room.



2D Recognition

- Image classification tagging/annotation

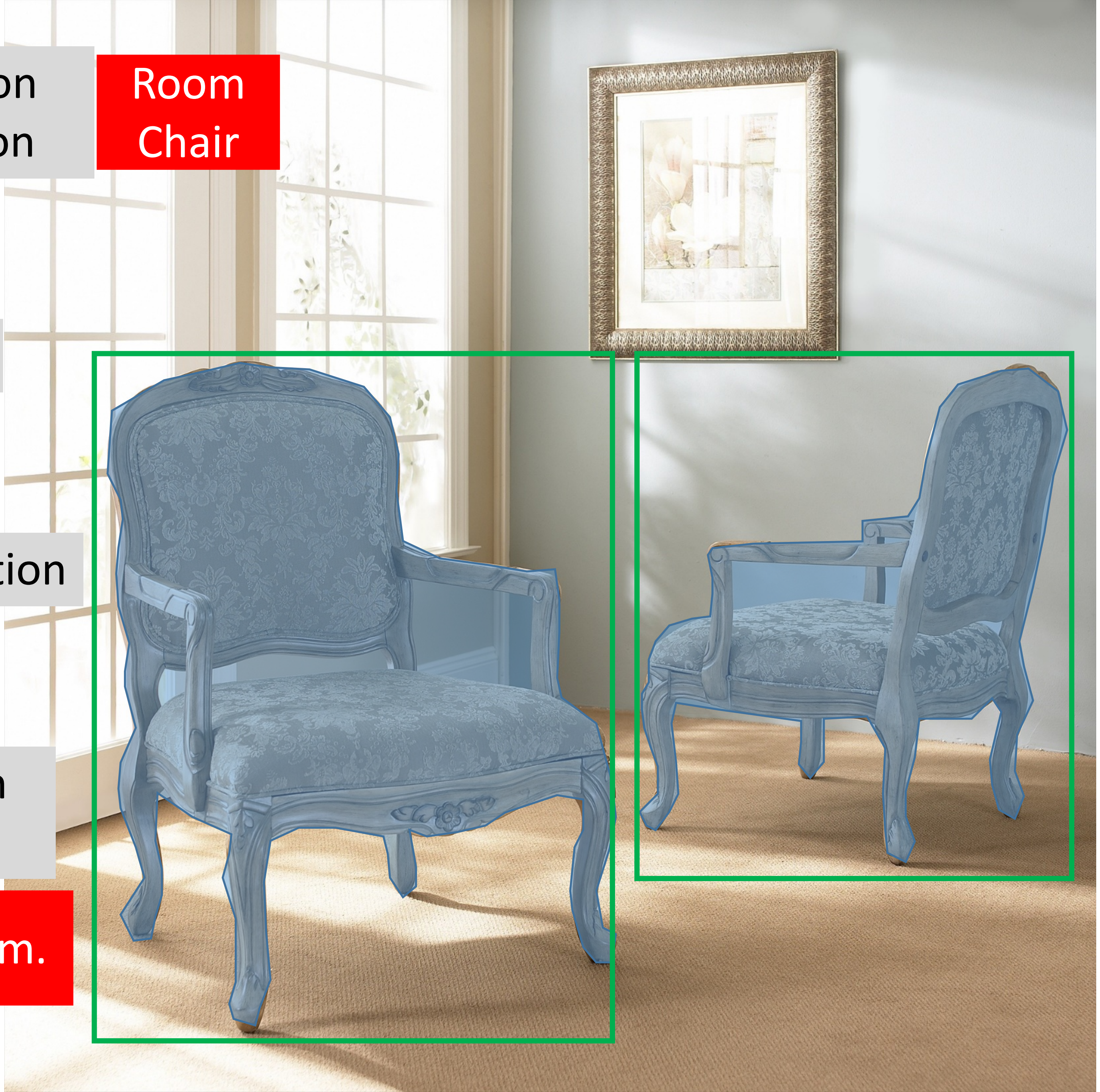
Room
Chair

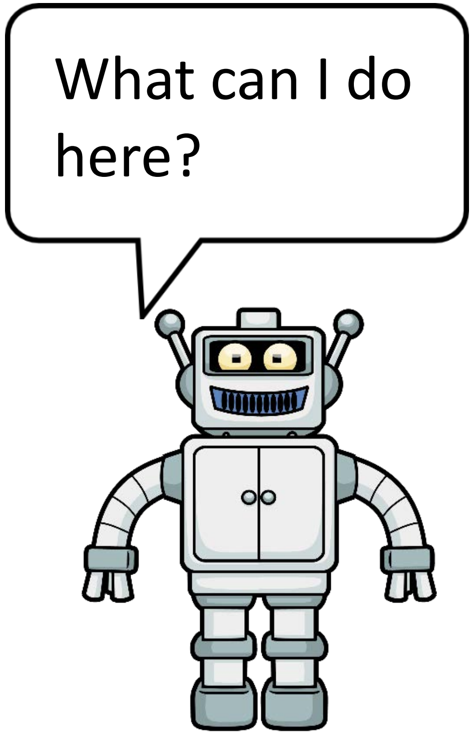
- Object detection

- Object segmentation

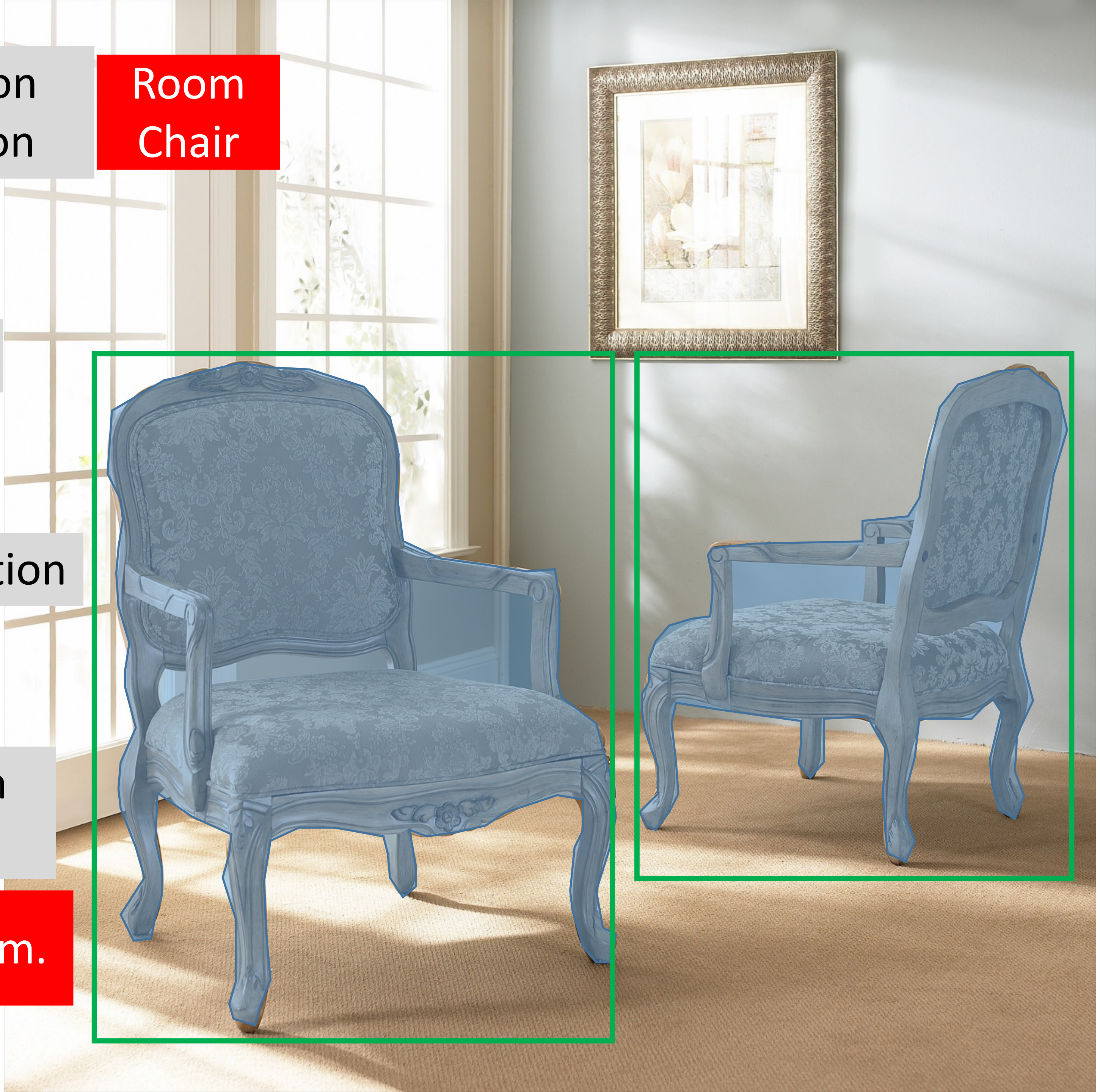
- Image description generation

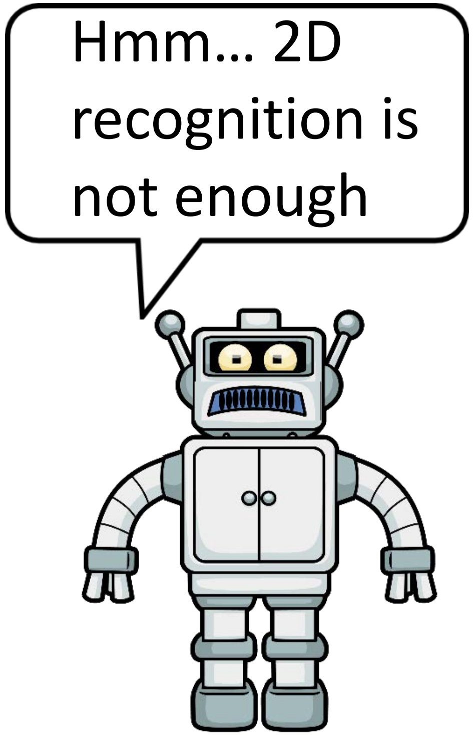
Two chairs in a room.





- Image classification tagging/annotation
- Object detection
- Object segmentation
- Image description generation

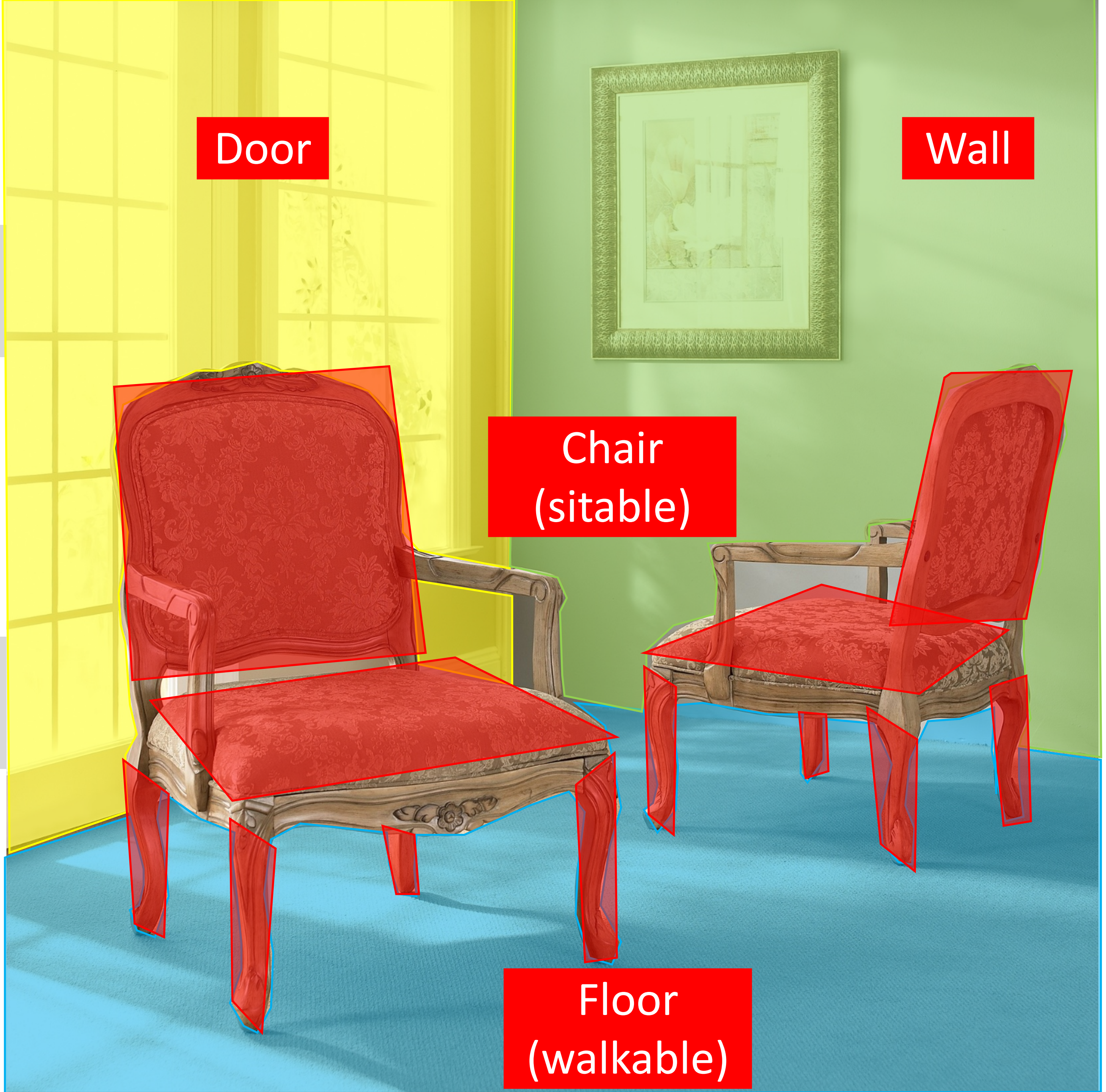




- 3D Scene Understanding

Hoiem et al., ICCV'05
Lee et al. CVPR'09
Hedau, el al., ICCV'09
Fouhey et al. ICCV'13
Schwing et al. ICCV'13
Lai, Bo & Fox. ICRA'14
Mallya & Lazebnik, ICCV'15
...

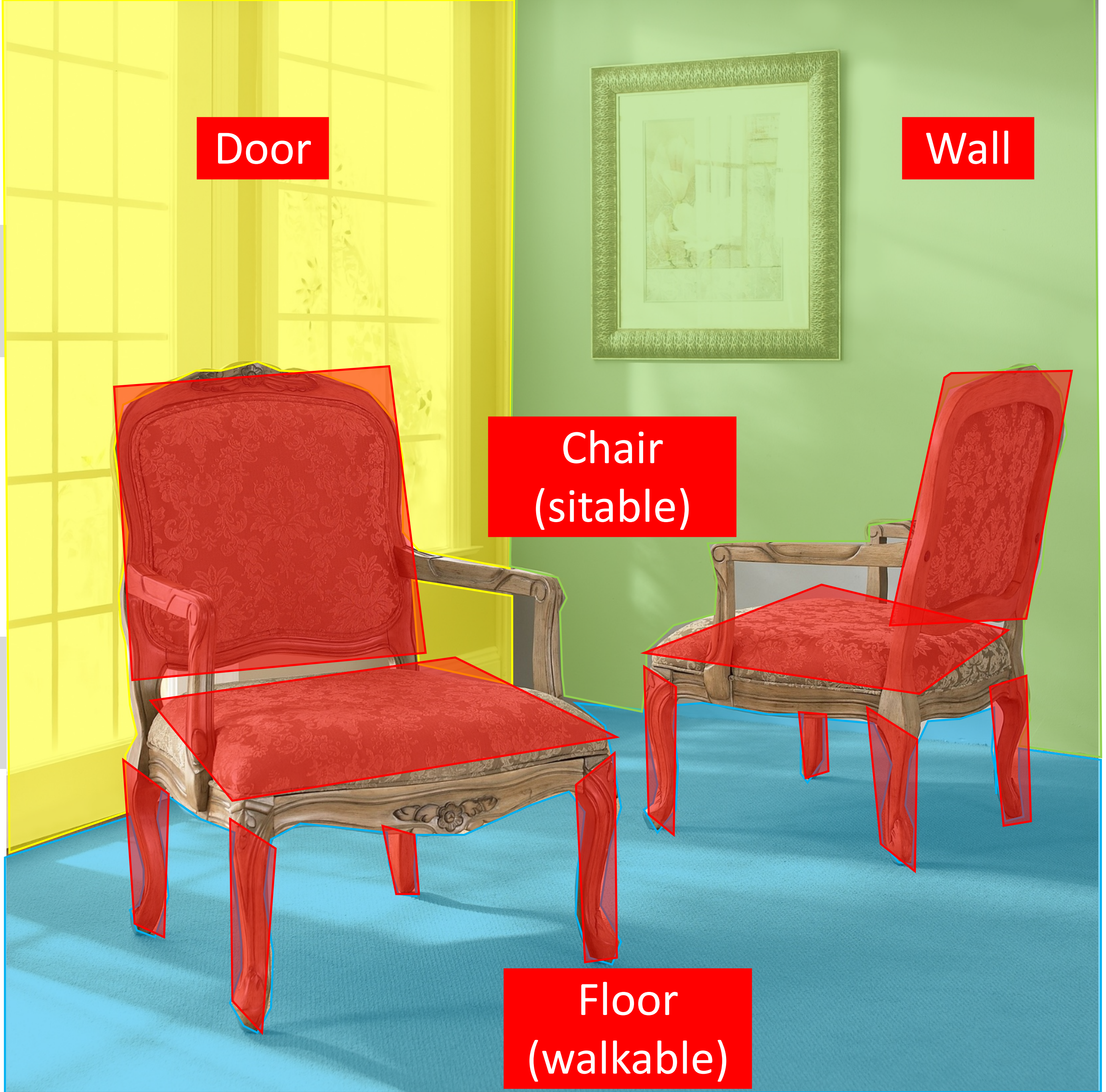
- 3D Scene Understanding

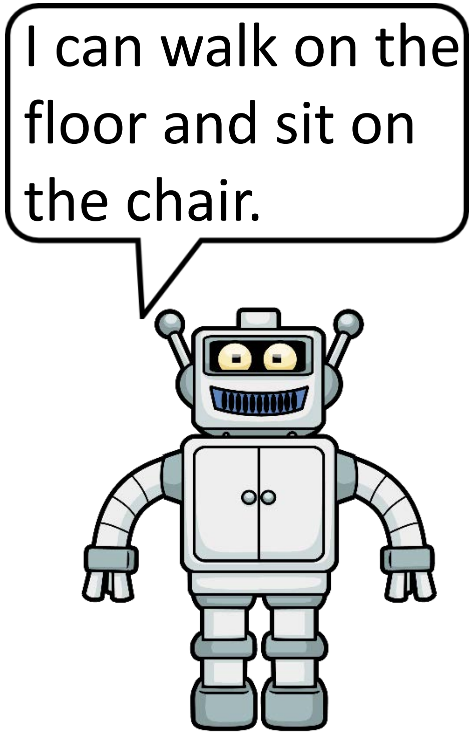


- 3D Object Recognition

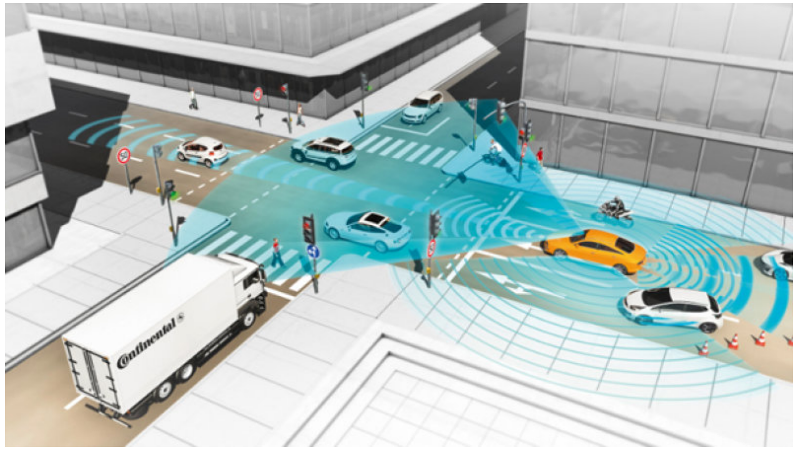
Savarese & Fei-Fei, ICCV’07
Sun et al. CVPR’09
Stark et al. BMVC’10
Glasner et al. ICCV’11
Pepik et al. CVPR’12
Xiang & Savarese, CVPR’12
Kar et al., ICCV’15
Tulsiani & Malik, CVPR’15
…

- 3D Scene Understanding
- 3D Object Recognition

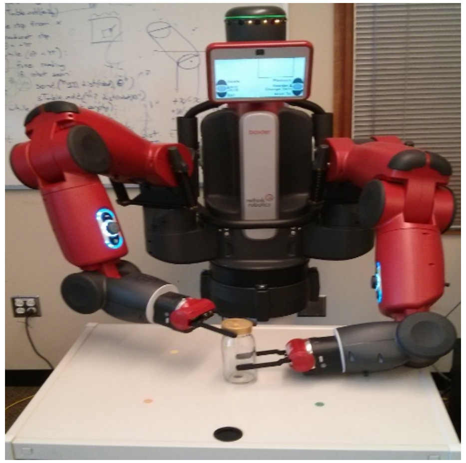




Applications that need 3D Object Recognition



Autonomous Driving



Robotics

Any application that requires interaction with the 3D world!

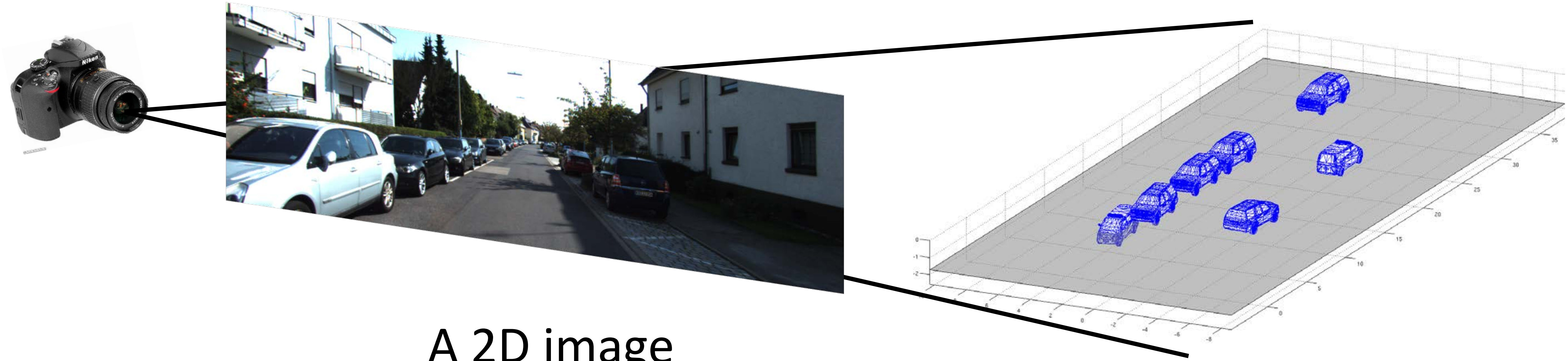
Augmented Reality

Gaming

Goal: Infer the 3D World

- Interaction
- Control
- Decision making
- Navigation

Etc.



A 2D image

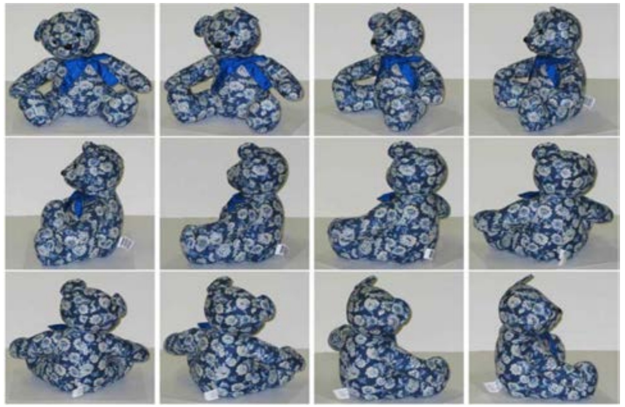
The 3D world

3D Object Recognition

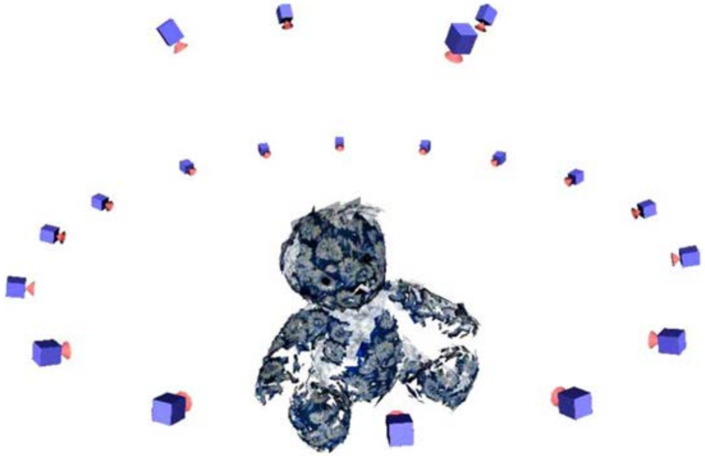
- Instance recognition
  - Training on images of an object instance
  - Testing on different images of the same object instance

- Category recognition
  - Training on images of different object instances
  - Testing on unseen object instances

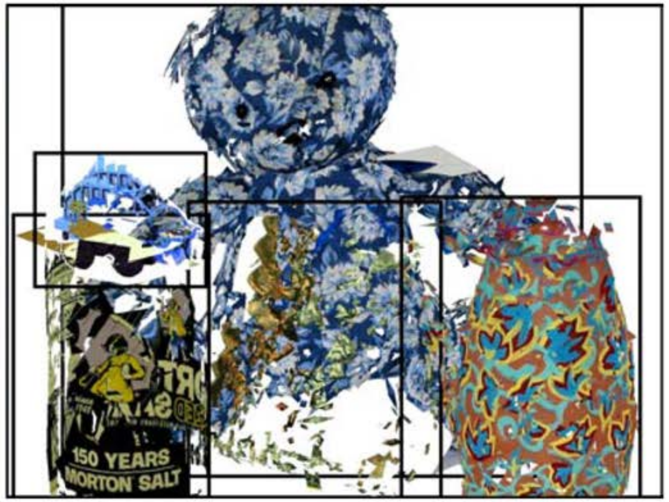
3D Object Instance Recognition



Training images

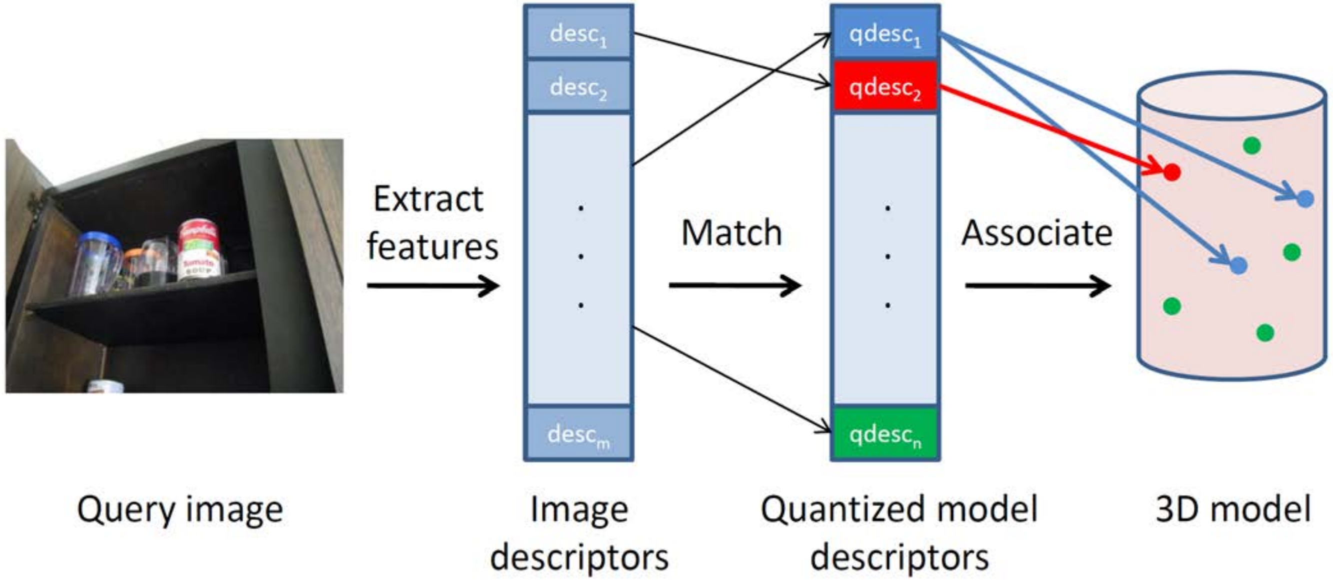


3D Object model



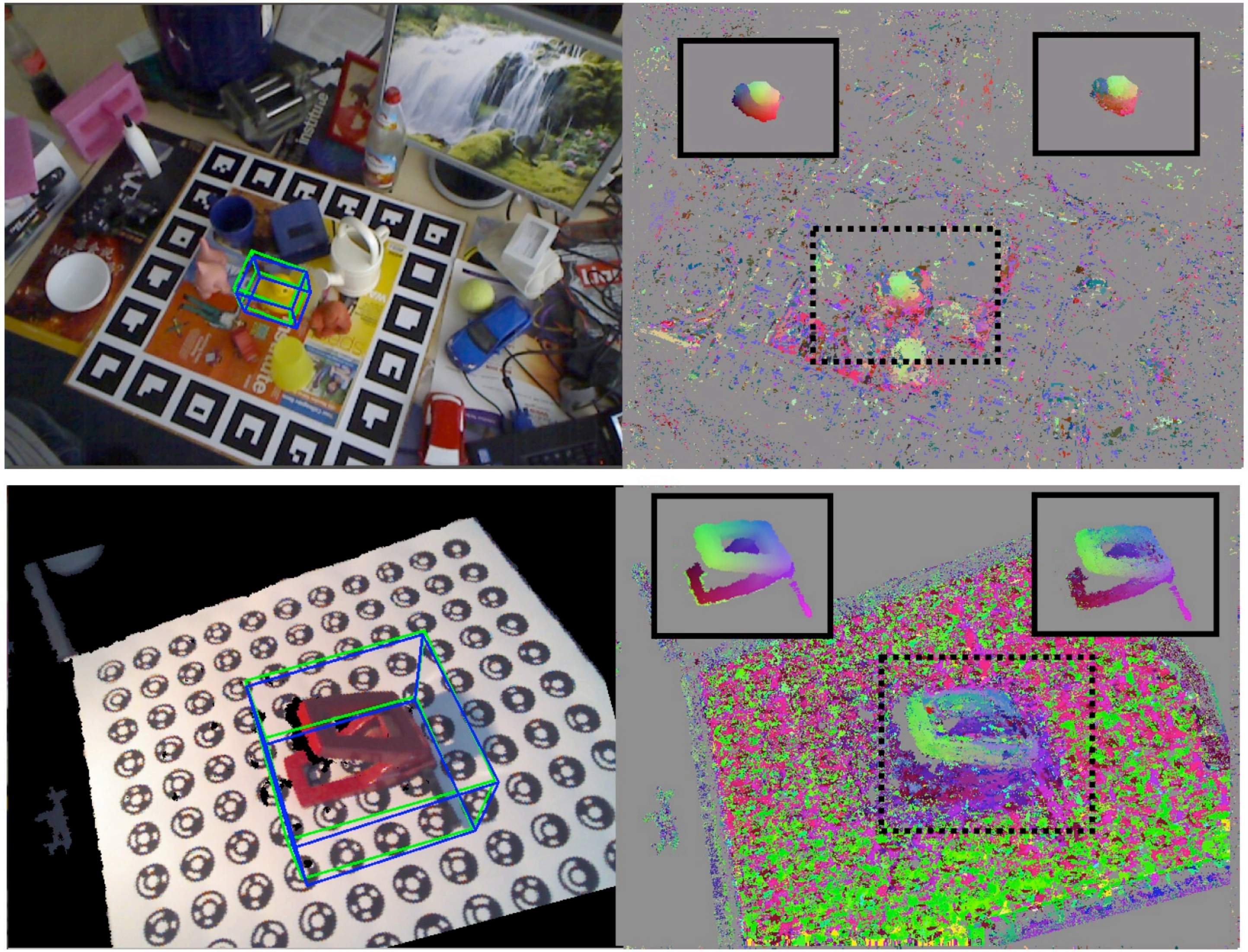
3D Object Modeling and Recognition Using Local Affine-Invariant Image Descriptors and Multi-View Spatial Constraints. *Rothganger et al., IJCV'06.*

3D Object Instance Recognition



Making specific features less discriminative to improve point-based 3D object recognition. Hsiao *et al., CVPR'10.*

3D Object Instance Recognition

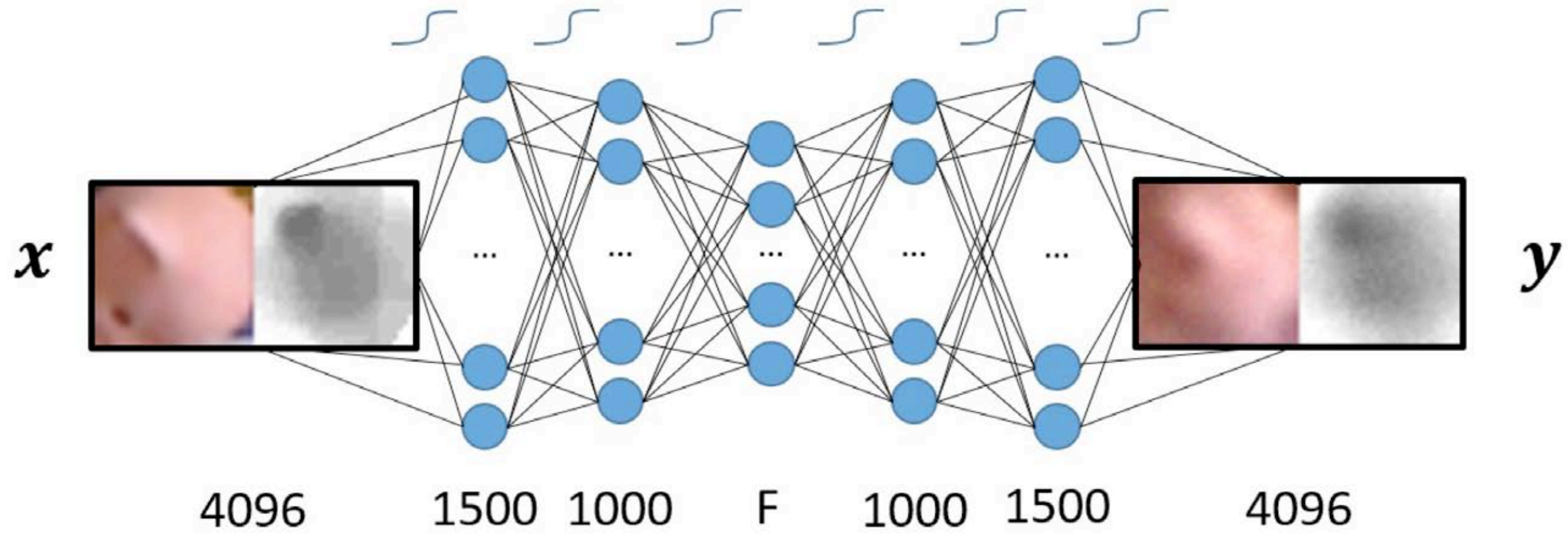


Using random forests, regress each pixel into

- class label
- 3D object coordinates

Learning 6D Object Pose Estimation using 3D Object Coordinates. Brachmann *et al., ECCV’14.*

3D Object Instance Recognition



Autoencoder for feature learning

Deep Learning of Local RGB-D Patches for 3D Object Detection and 6D Pose Estimation. Kehl et al., ECCV'16.

3D Object Instance Recognition

- Build a 3D representation of the object instance
  - 3D model from multi-view images
  - 3D CAD model
- Feature matching
  - Learning better features (e.g., deep learning)
  - Handle matching ambiguities
- Voting-based scheme
  - Vote for object label, object pose, object 3D coordinate, etc.
  - Model selection (e.g., RANSAC)

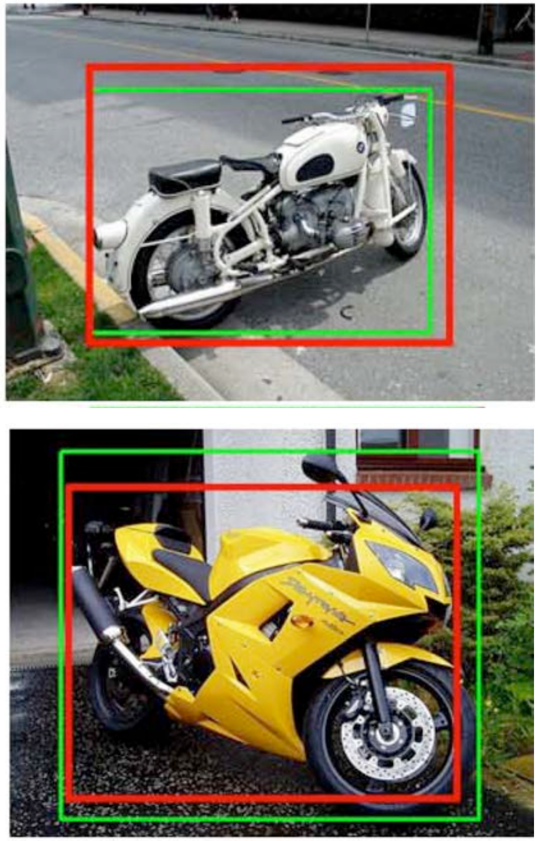
3D Object Category Recognition

- How to learn a 3D representation for an object category?

- How to use the learned 3D representation for recognition?

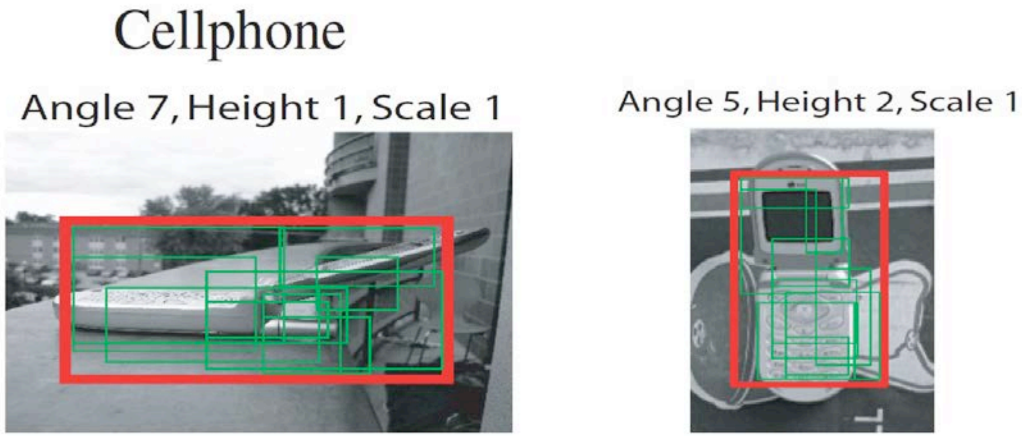
3D Object Category Recognition

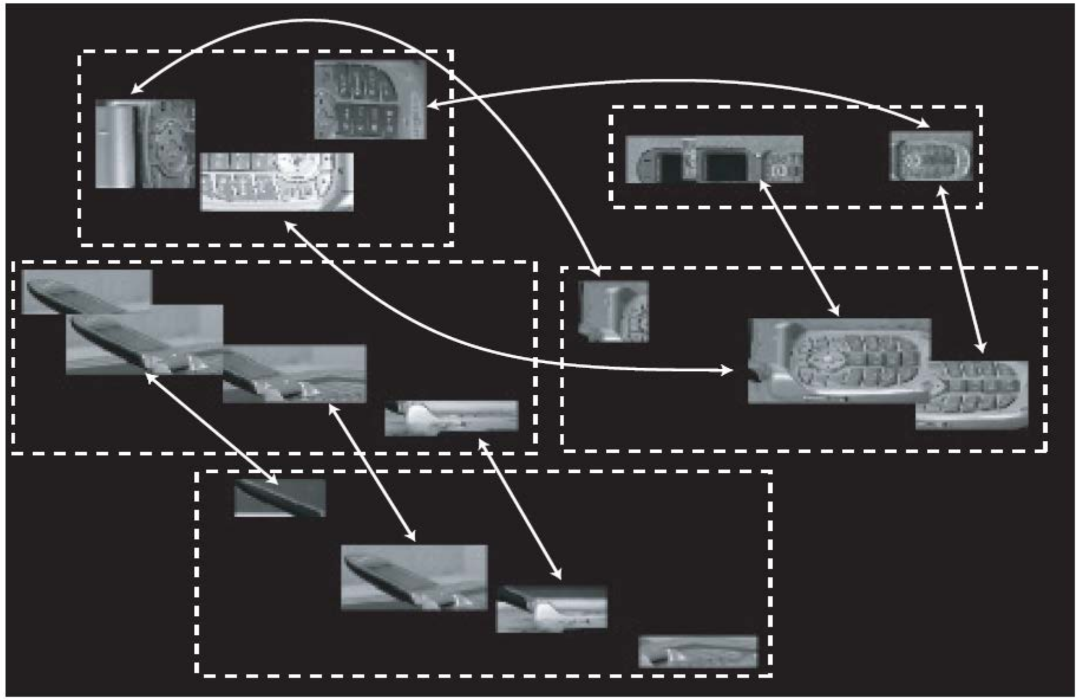
3D feature model



3D Model based Object Class Detection in An Arbitrary View. Yan et al., ICCV'07.

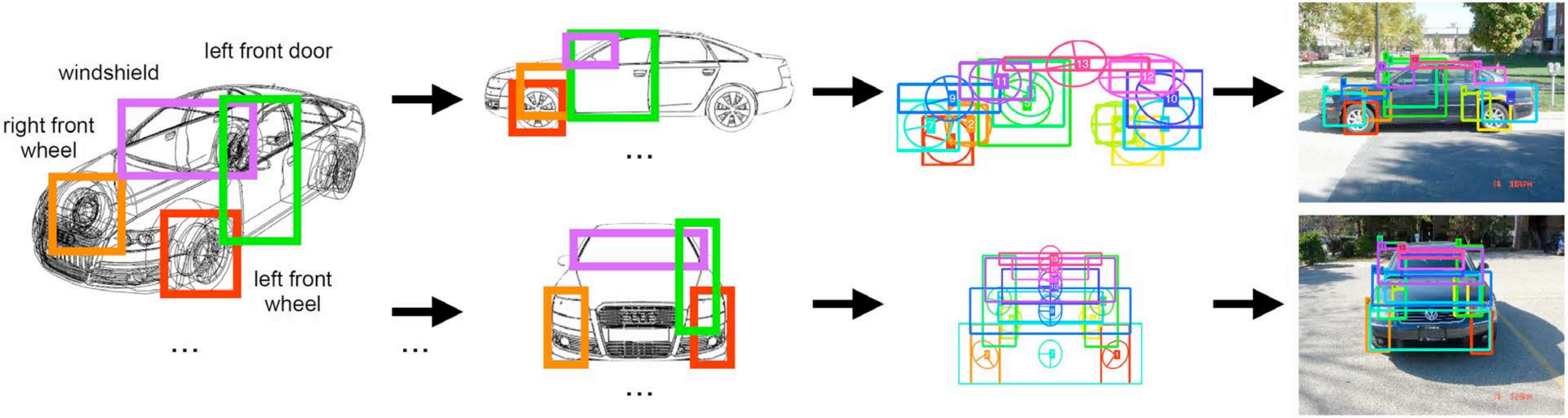
3D Object Category Recognition





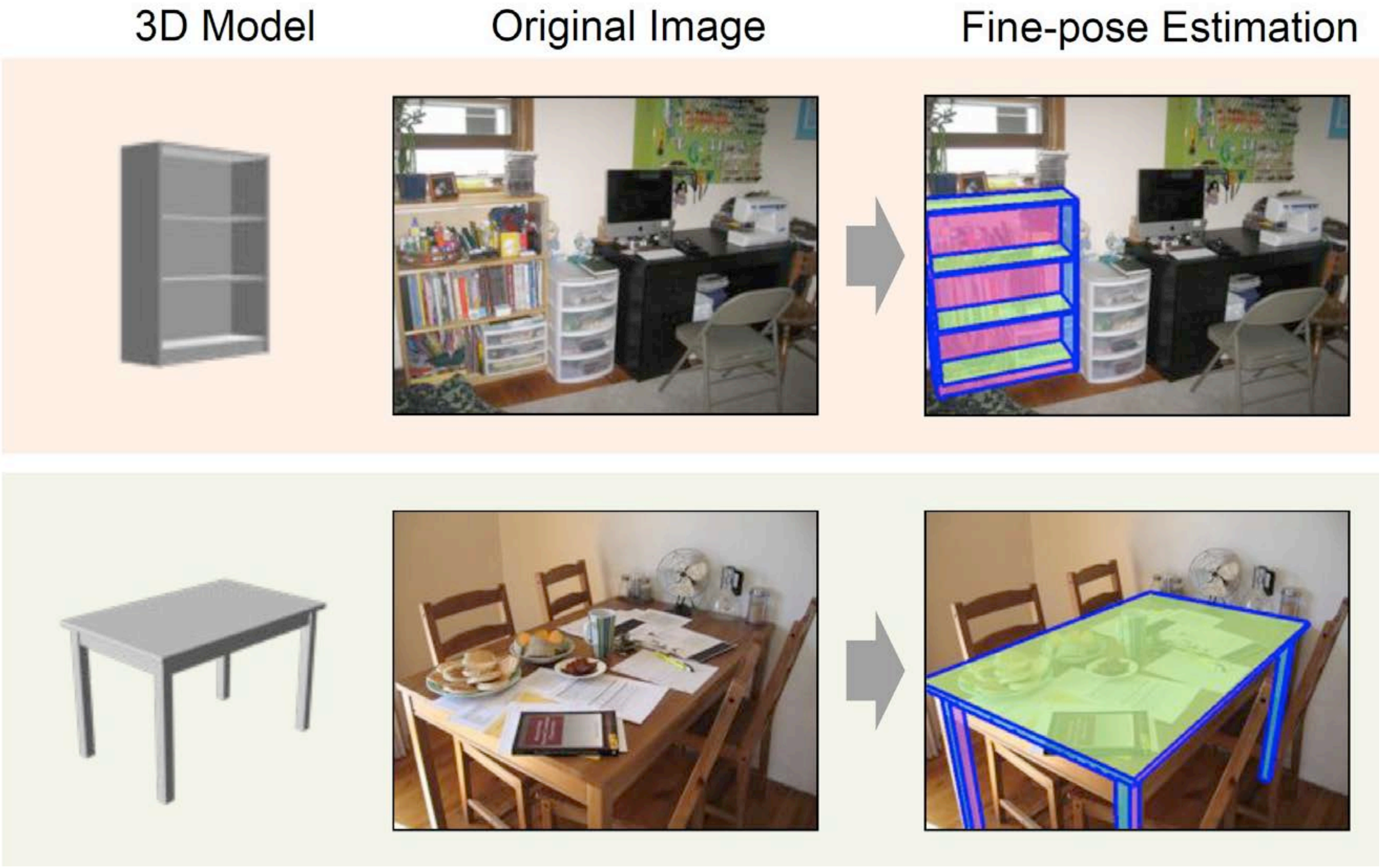
3D generic object categorization, localization and pose estimation. Savarese & Fei-Fei, ICCV’07.

3D Object Category Recognition



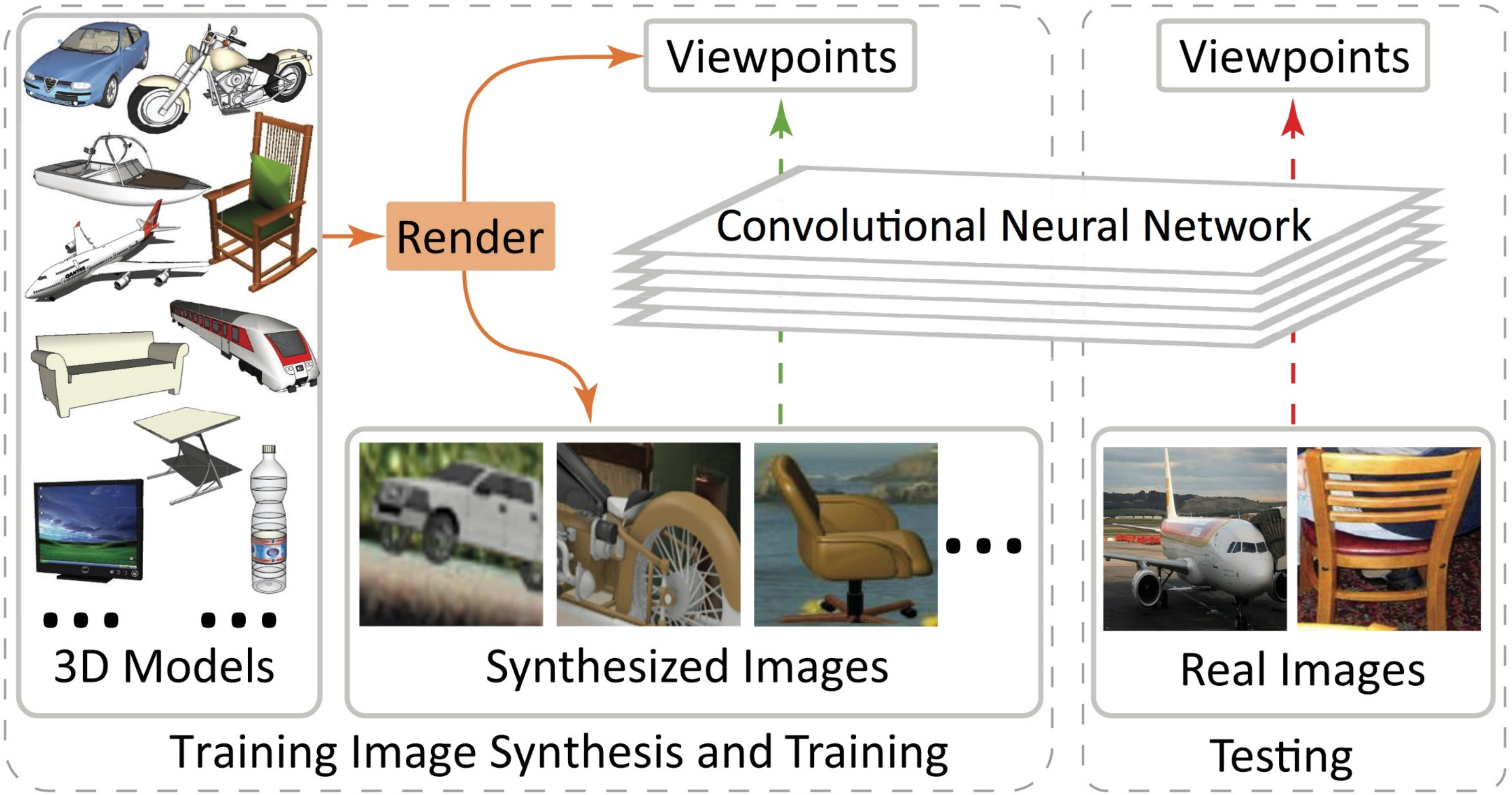
Back to the Future: Learning Shape Models from 3D CAD Data. Stark et al., BMVC'10.

3D Object Category Recognition



Parsing IKEA Objects: Fine Pose Estimation. Lim et al., ICCV'13.

3D Object Category Recognition



Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views. Su et al., ICCV'15.

3D Object Category Recognition

- Learning 3D object representations
  - Multi-view images or videos
  - 3D CAD models

- Challenges
  - Scalable to the number of categories
  - Handle appearance variations due to change in illumination, scale, 3D pose, 3D shape, occlusion and truncation
  - Recognize detailed properties of objects: 3D pose, 3D shape, 3D location

Outline

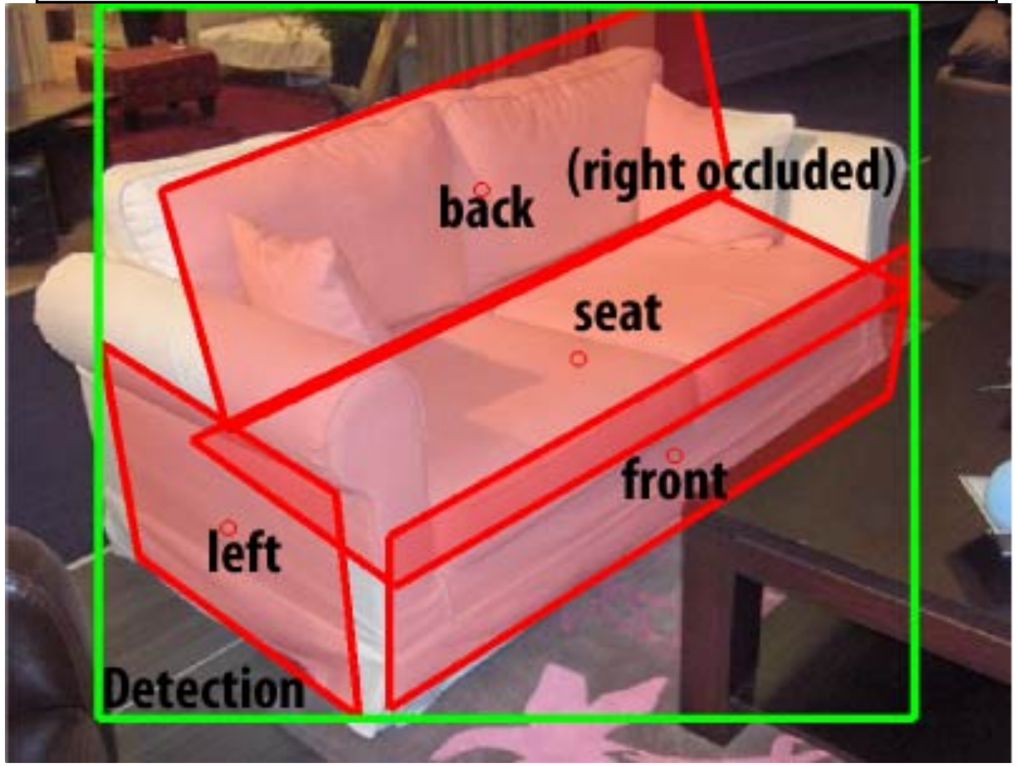
- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Summary

Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Summary

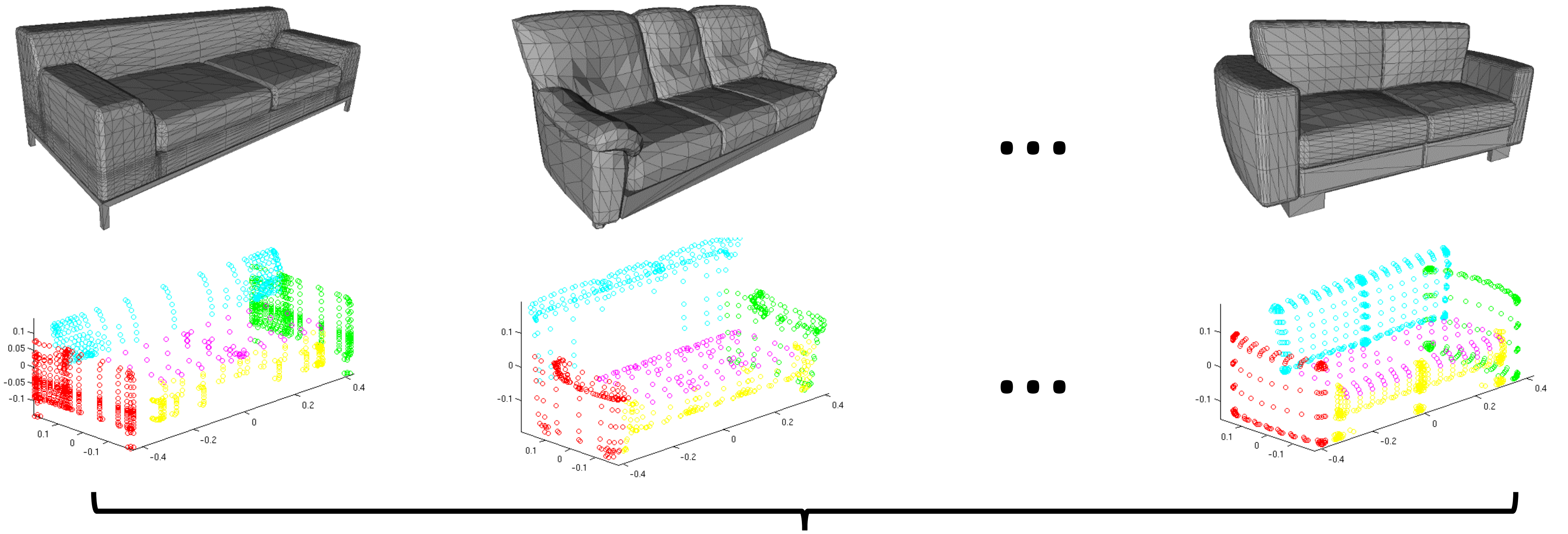
3D Aspect Part Representation

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

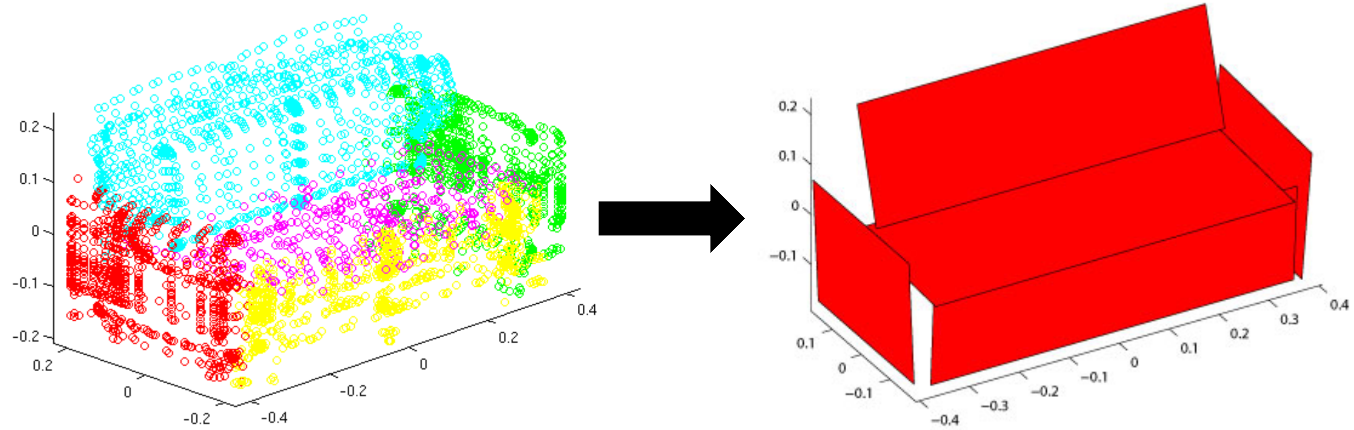


Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.

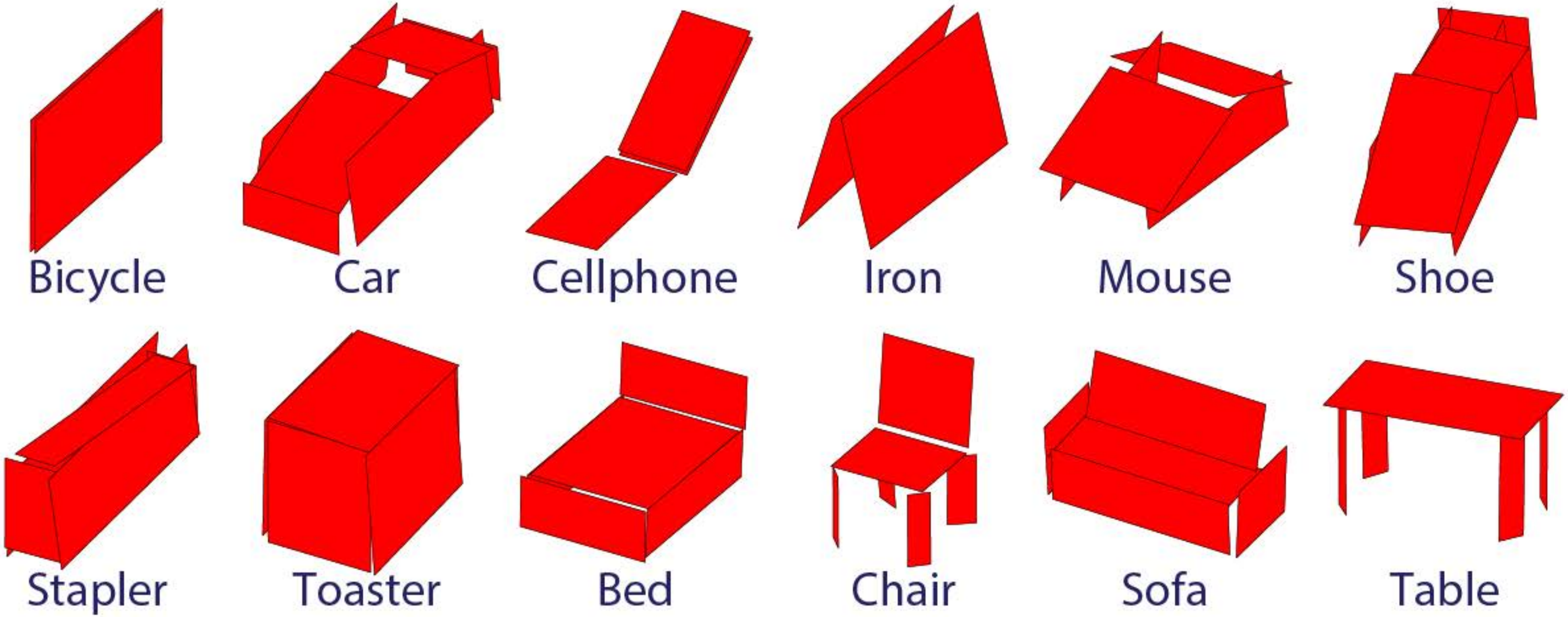
3D Aspect Parts from 3D CAD Models



Mean Shape

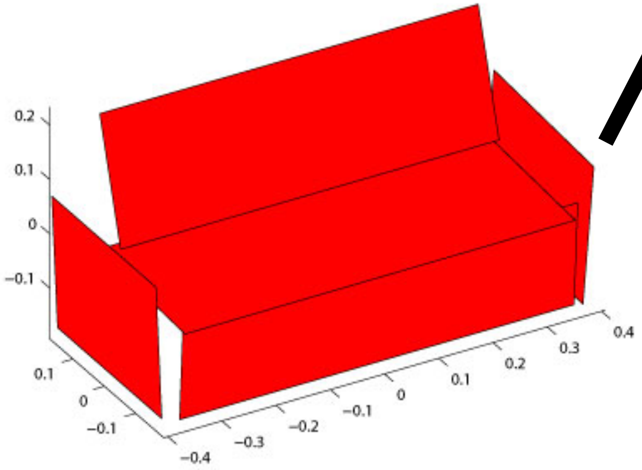


3D Aspect Part Representation



Aspect Layout Model

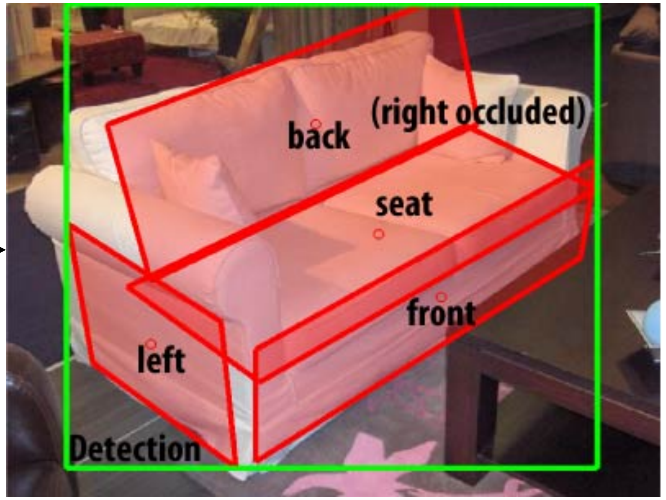
An input image



3D aspect part representation

Aspect Layout Model

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



Output

Aspect Layout Model

- Posterior distribution

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

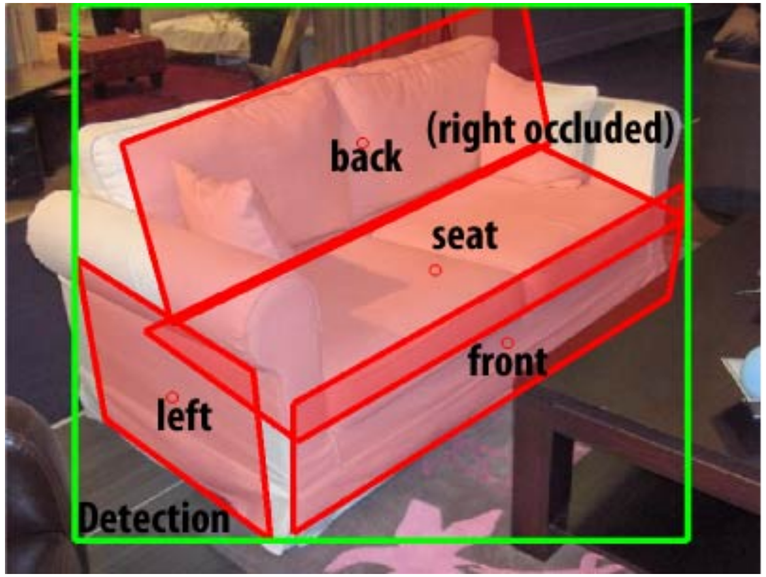
object label

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

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

unary potential

pairwise potential

Aspect Layout Model

- Unary potential

part template

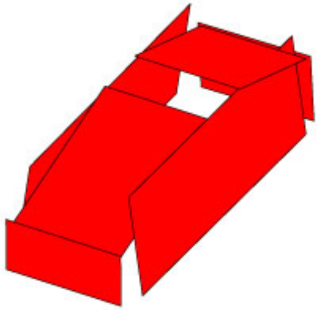
feature vector

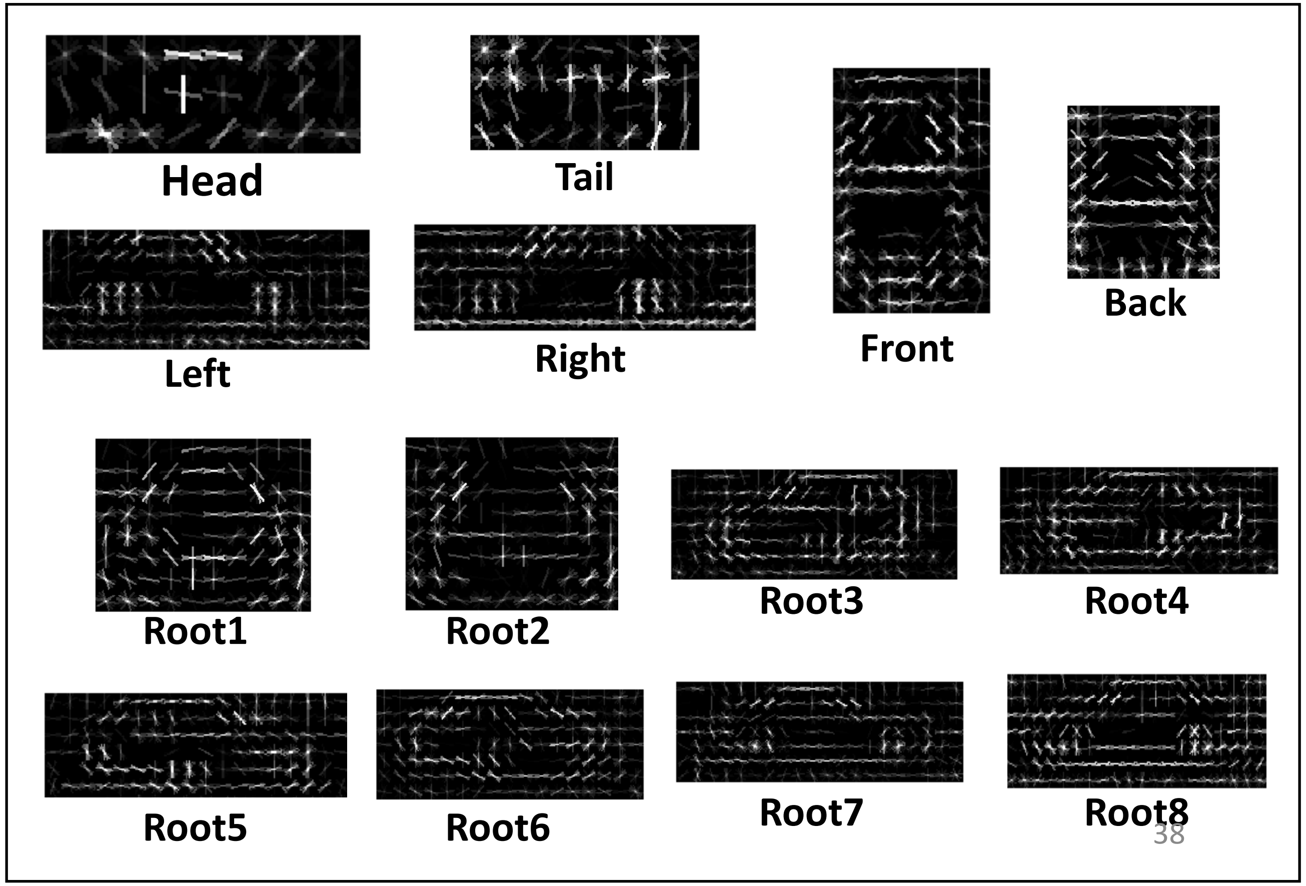
$$V_1(\mathbf{l}_i, O, V, I) = \begin{cases} \mathbf{w}_i^T \phi(\mathbf{l}_i, O, V, I), \text{ if unoccluded} \\ \alpha_i, \text{ if self-occluded} \end{cases}$$

self-occlusion weight

Aspect Layout Model

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

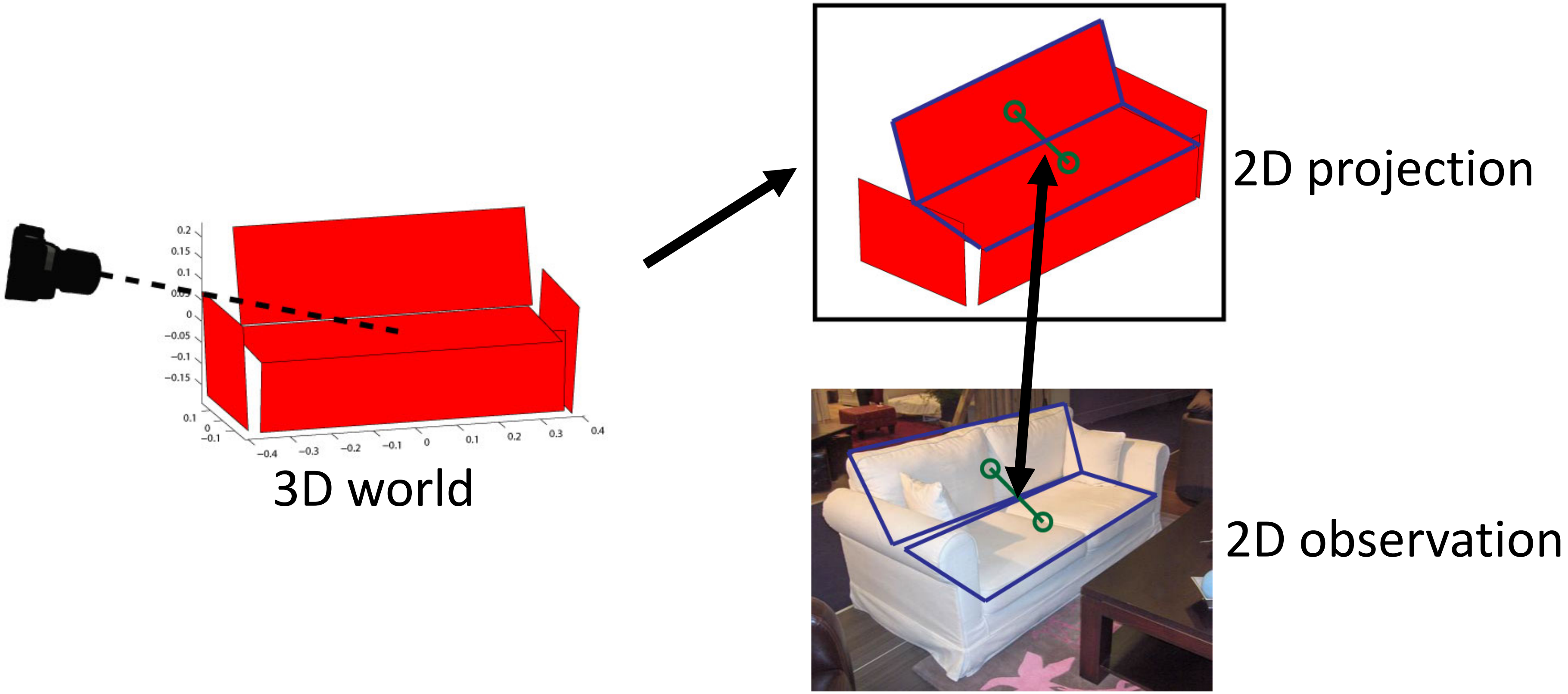




Aspect Layout Model

- 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

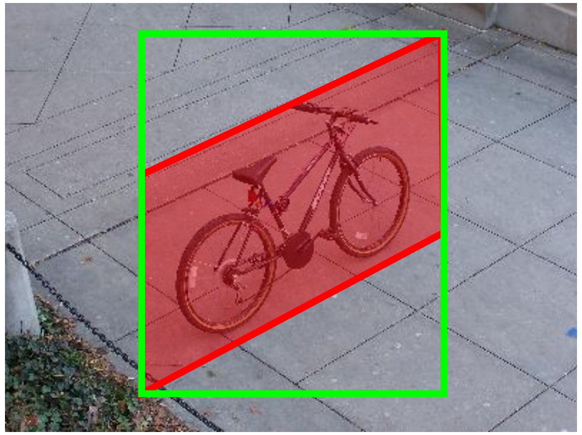
Chairs, tables, sofas and beds from IMAGE NET [Deng et al. CVPR'09]

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

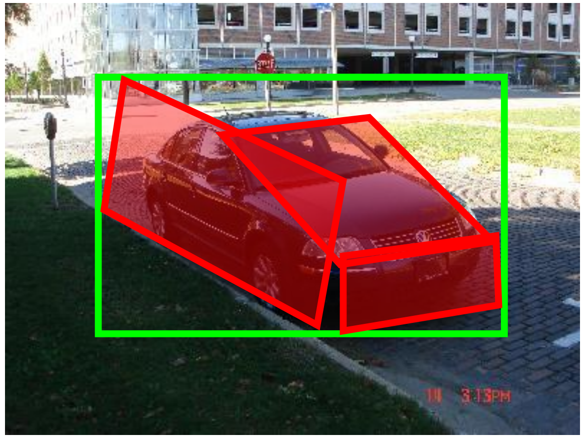
[1] N. Payet and S. Todorovic. From contours to 3d object detection and pose estimation. In ICCV, 2011.
[2] D. Glasner, M. Galun, S. Alpert, R. Basri, and G. Shakhnarovich. Viewpoint-aware object detection and pose estimation. In ICCV, 2011.
[3] M. Stark, M. Goesele, and B. Schiele. Back to the future: 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

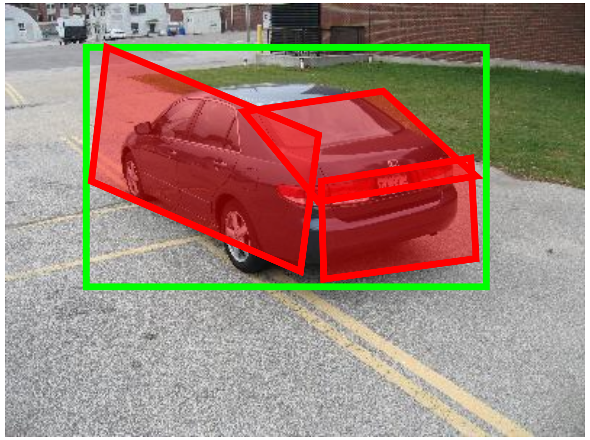
Prediction: a=225, e=30, d=7



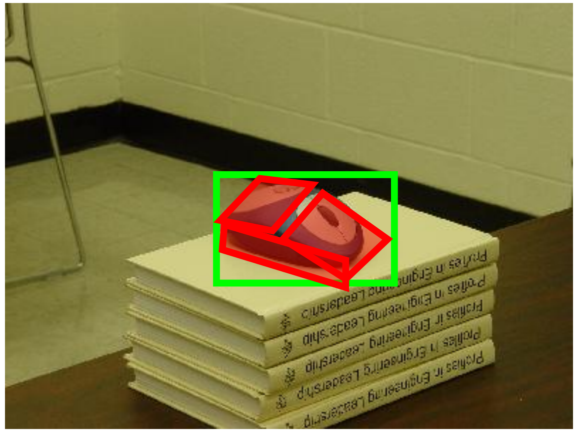
Prediction: a=330, e=15, d=7



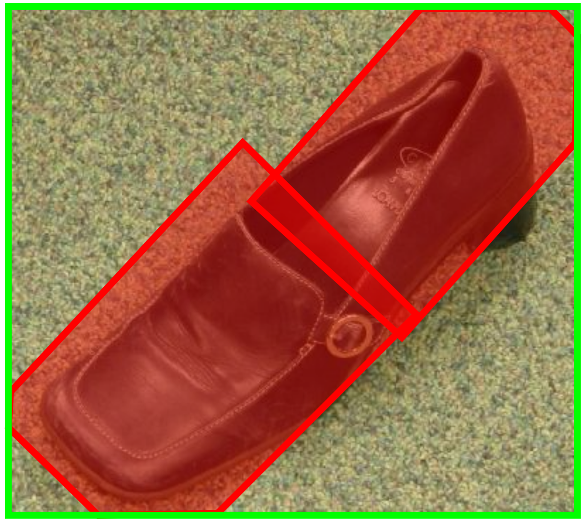
Prediction: a=150, e=15 d=7



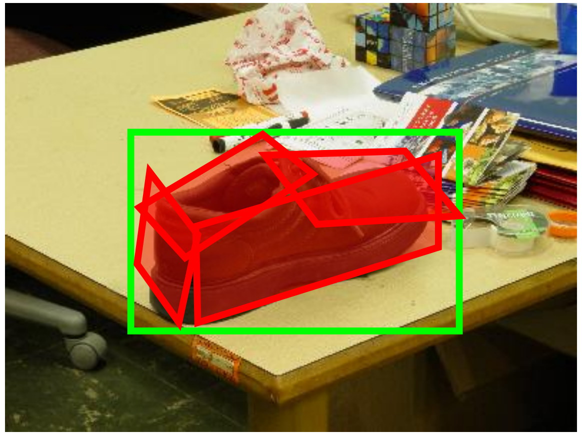
Prediction: a=300, e=45 d=23



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



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

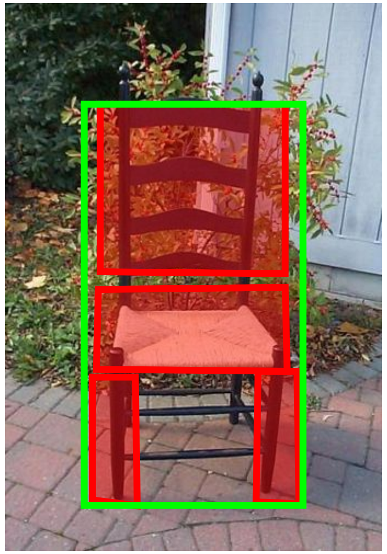


Aspect Layout Model

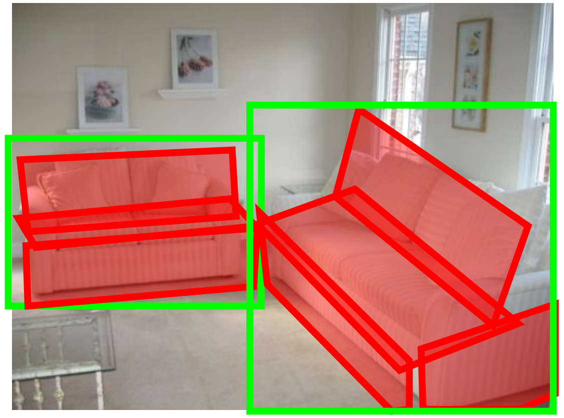
Prediction: a=30, e=15, d=2.5

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

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

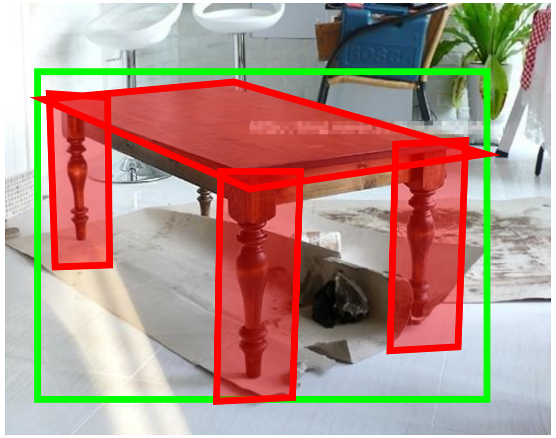


Prediction: a=345, e=15, d=3.5
a=60, e-30, d=2.5



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

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

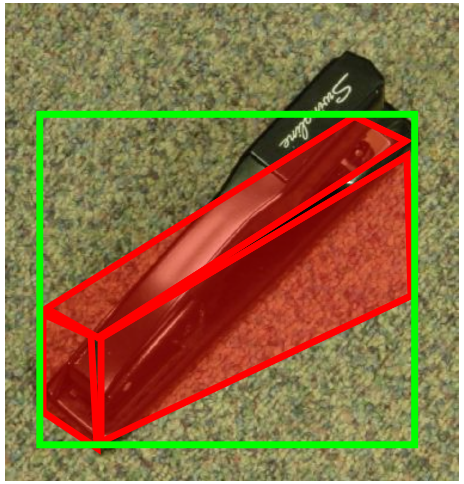


ImageNet dataset [Deng et al. 2010]

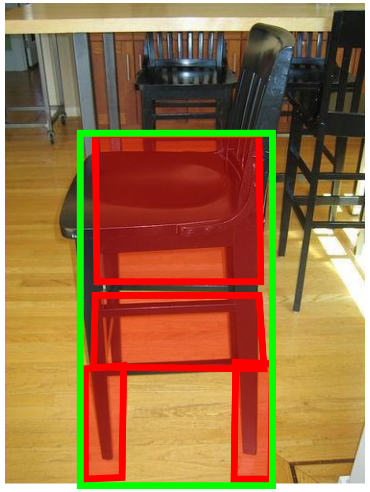
Wrong examples

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

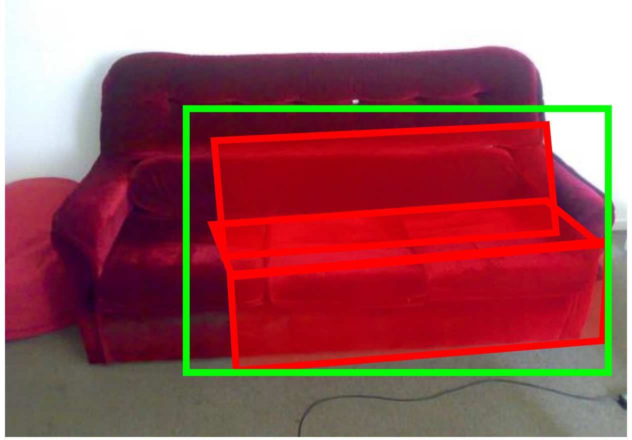
Prediction: a=225, e=30, d=7



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



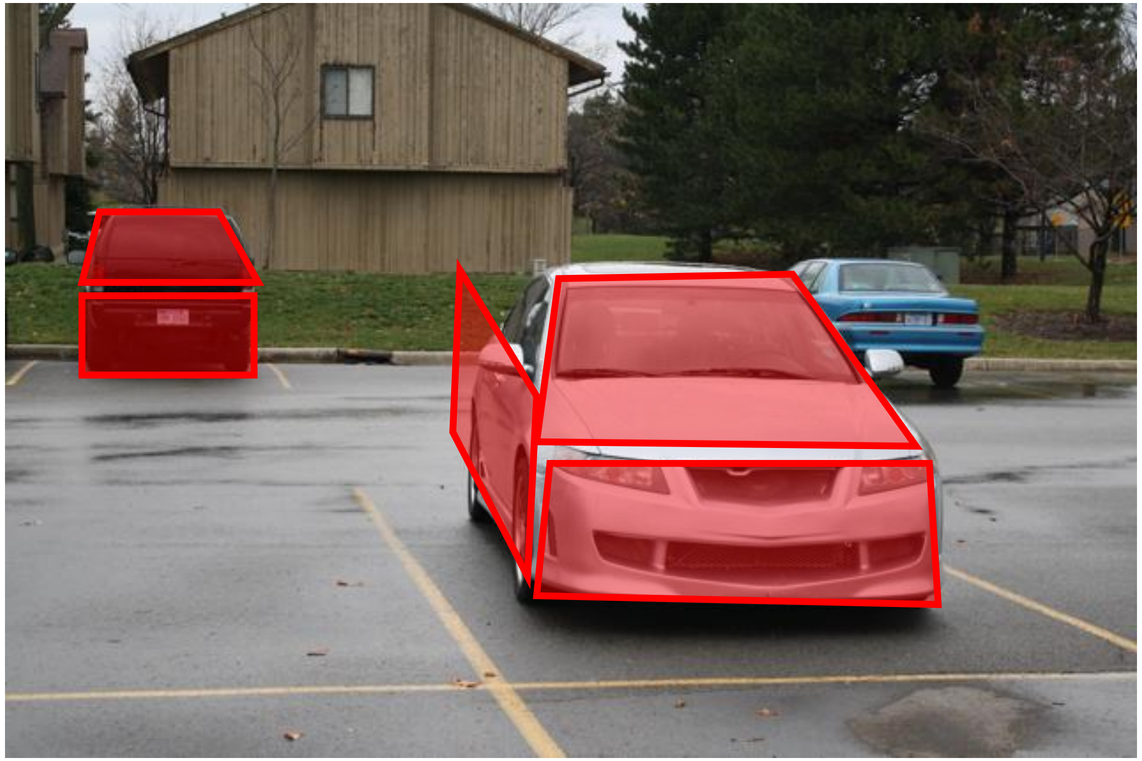
Prediction: a=345, e=15 d=2.5



Outline

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

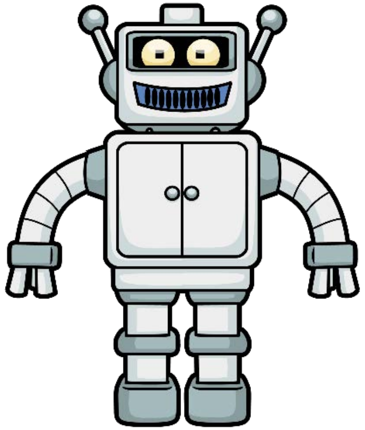
How to handle occlusion?

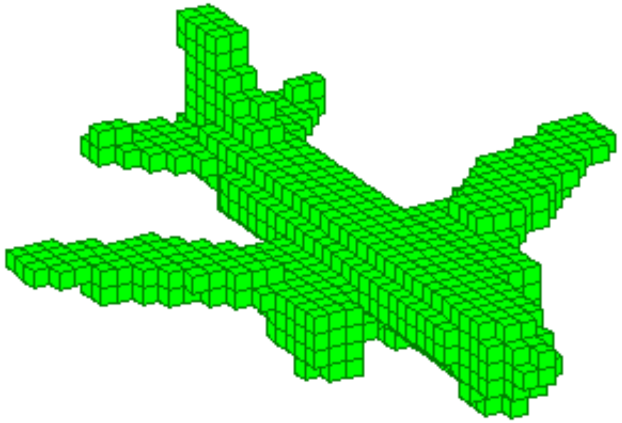


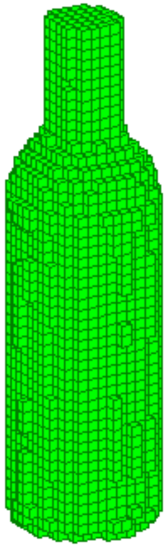
?

Occlusion changes the appearances of objects.

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



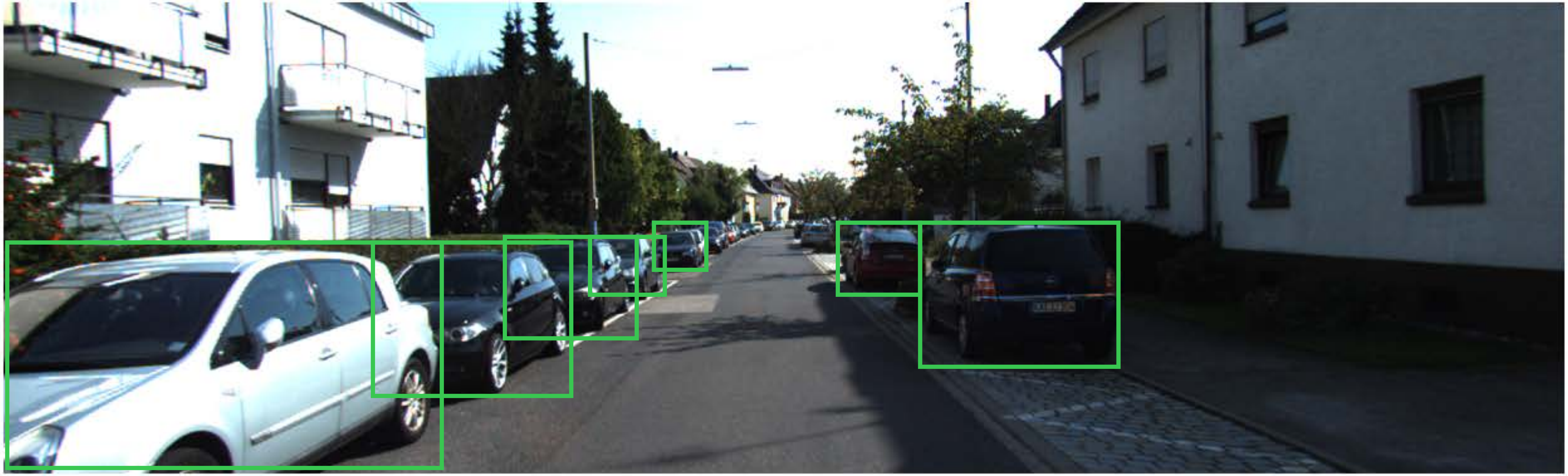




2D Object Detection

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

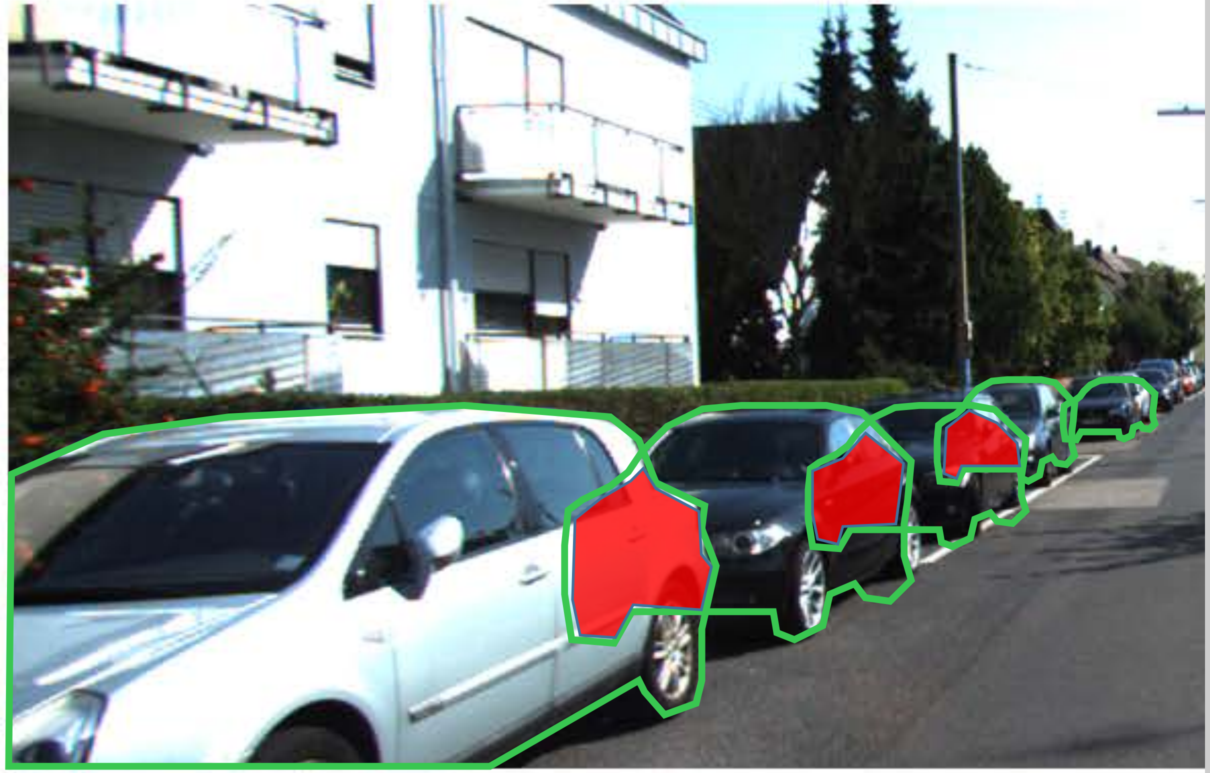
2D Object Detection

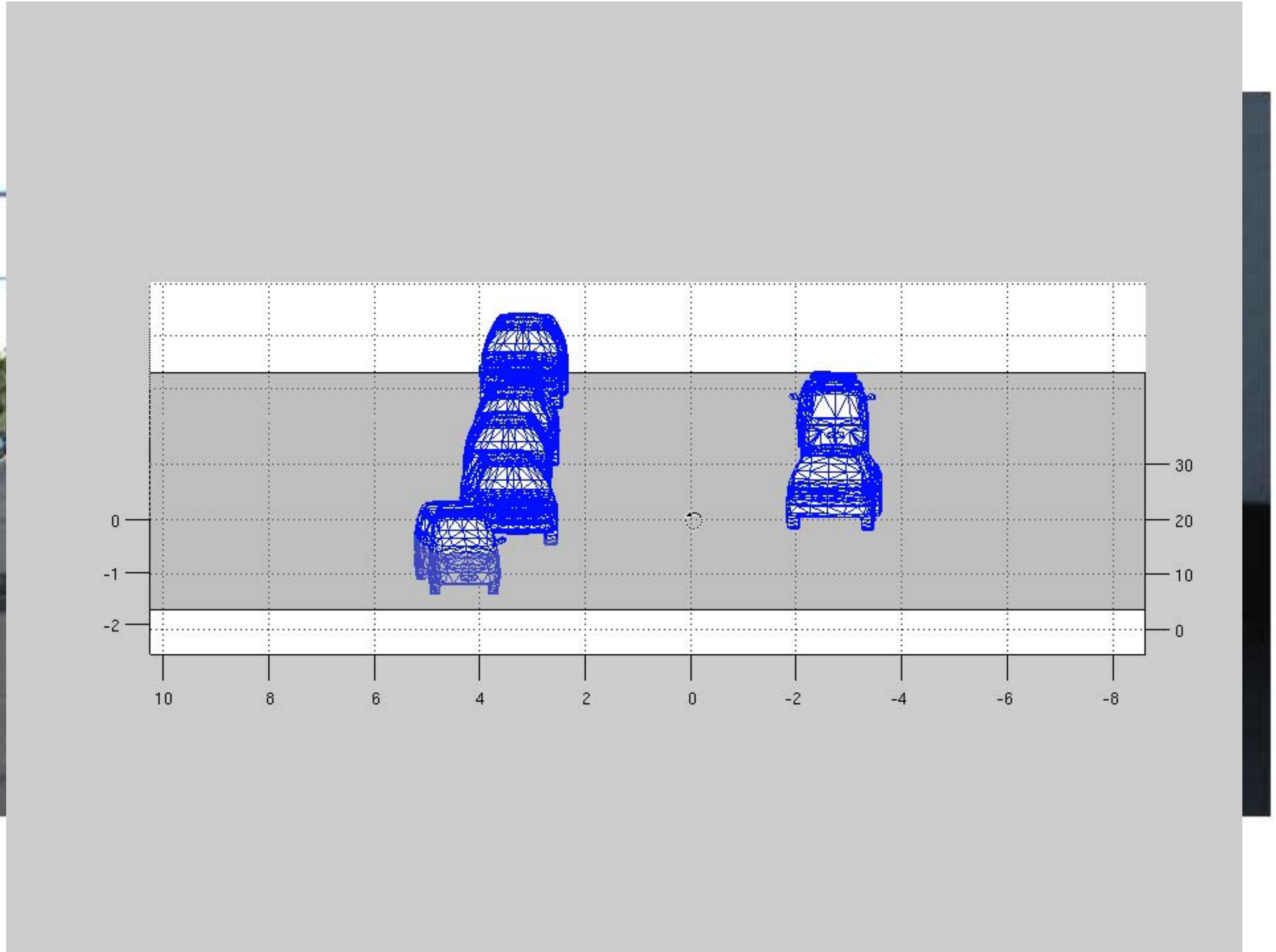


2D Segmentation and 3D Pose Estimation

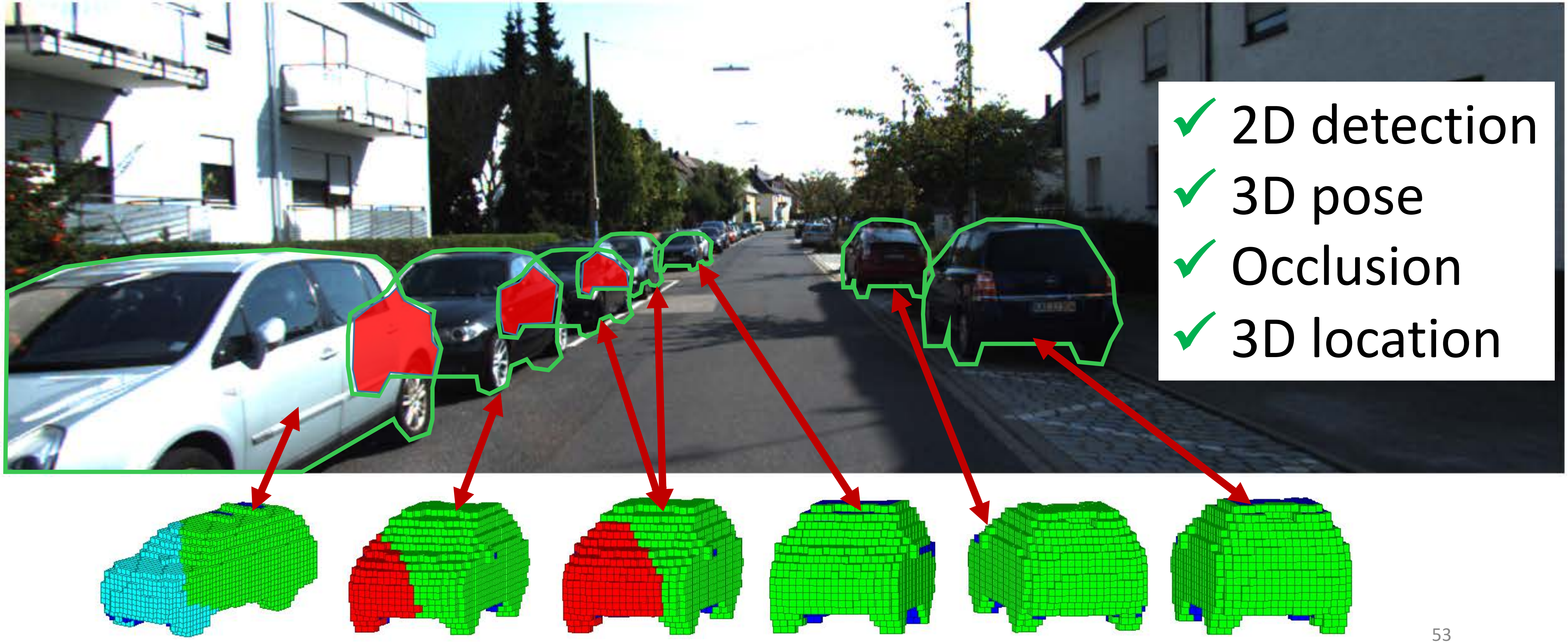
Occlusion Reasoning

3D Localization





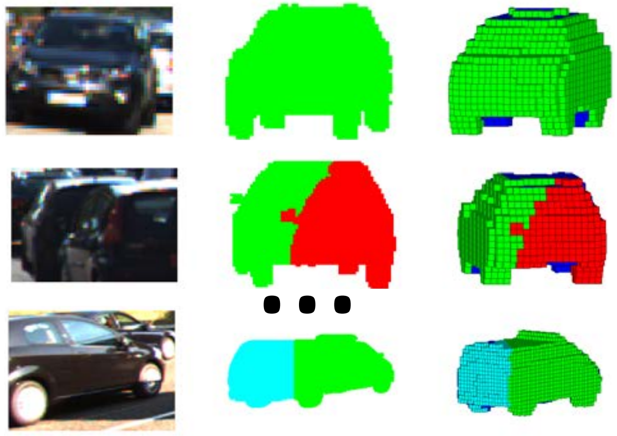
Data-Driven 3D Voxel Patterns



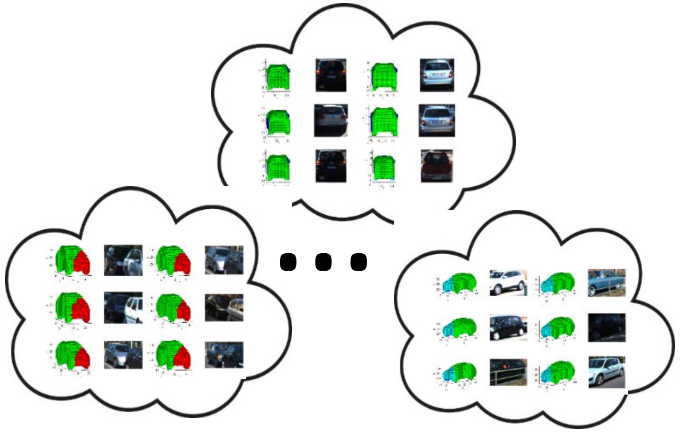
Y. Xiang, W. Choi, Y. Lin and S. Savarese. Data-Driven 3D Voxel Patterns for Object Category Recognition. In CVPR, 2015 (Oral).

Training Pipeline Overview

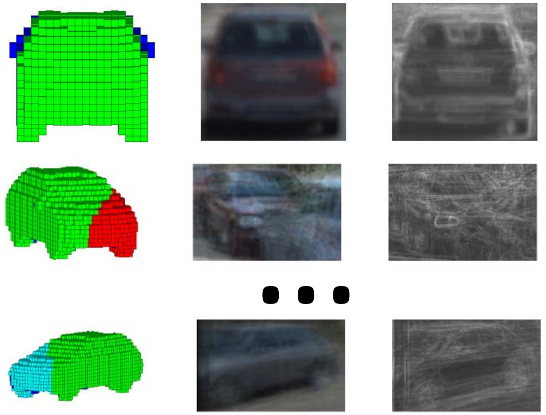
1. Align 2D images with 3D CAD models



2. 3D voxel exemplars

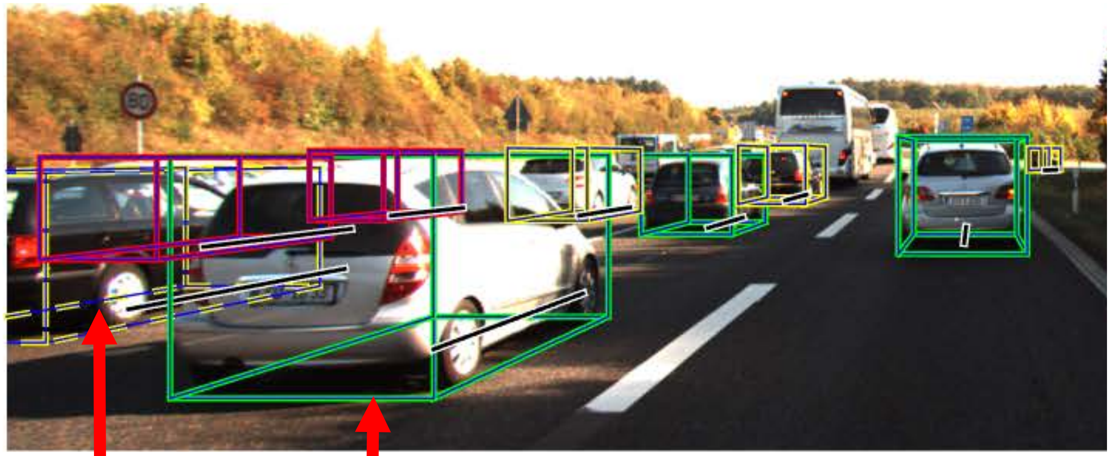


3. 3D voxel patterns



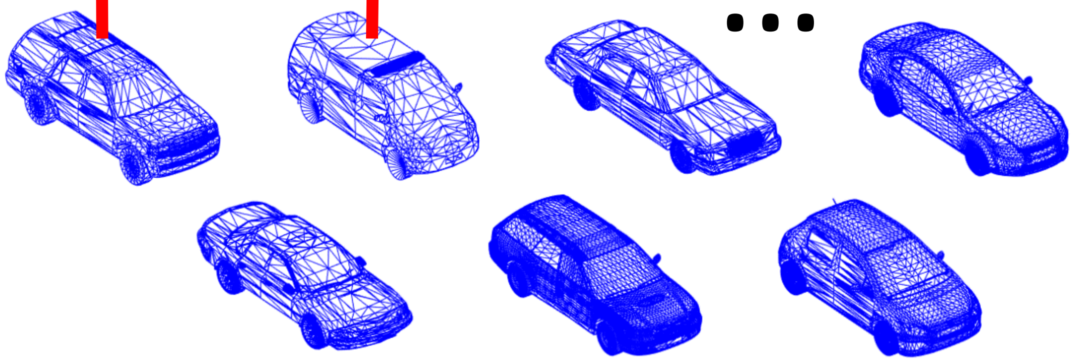
4. Training 3D voxel pattern detectors

1. Align 2D Images with 3D CAD Models



3D annotations (3D cuboids on point cloud)

Project of 3D CAD models

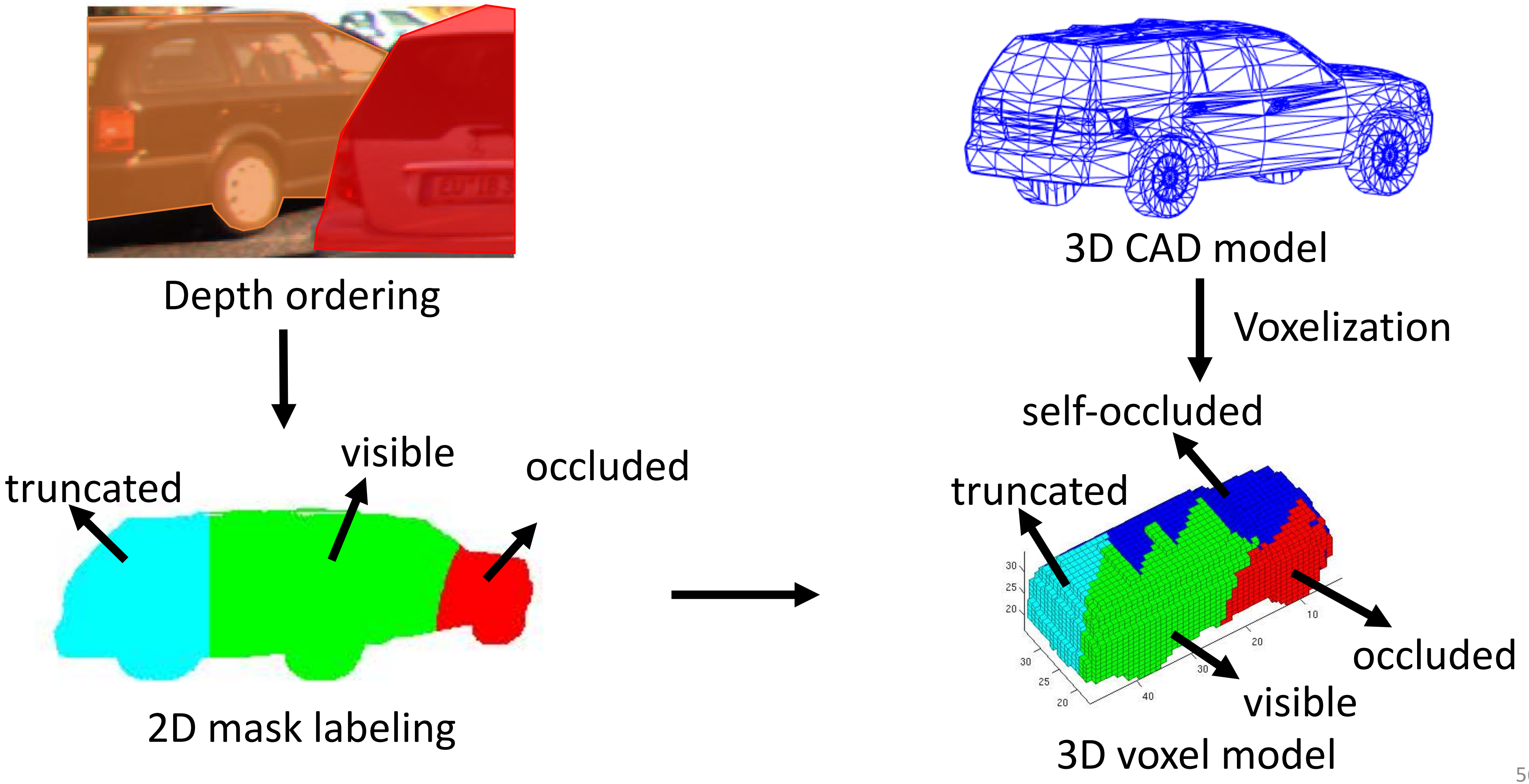


3D CAD models

Depth ordering

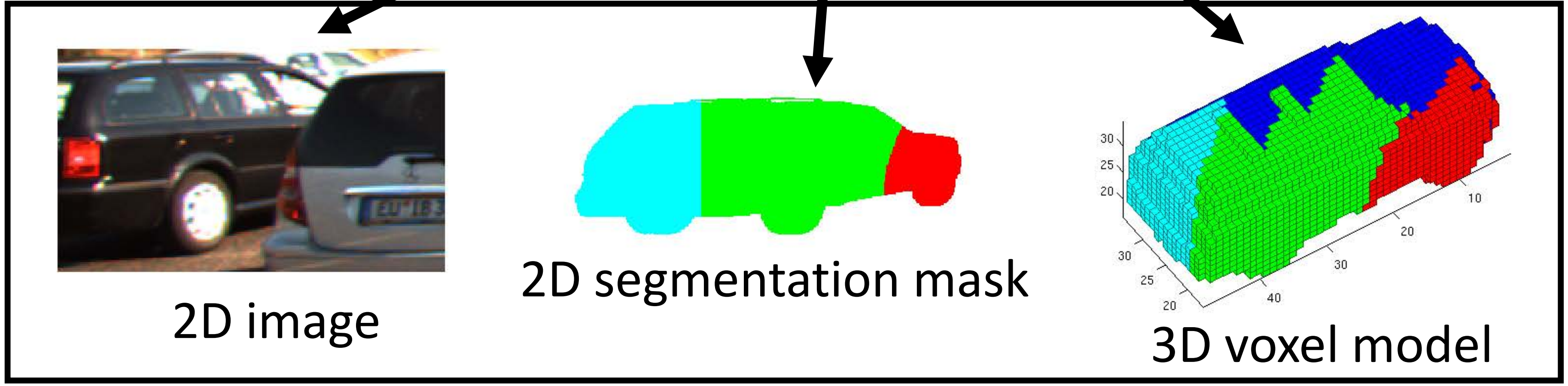
A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012

2. Building 3D Voxel Exemplars

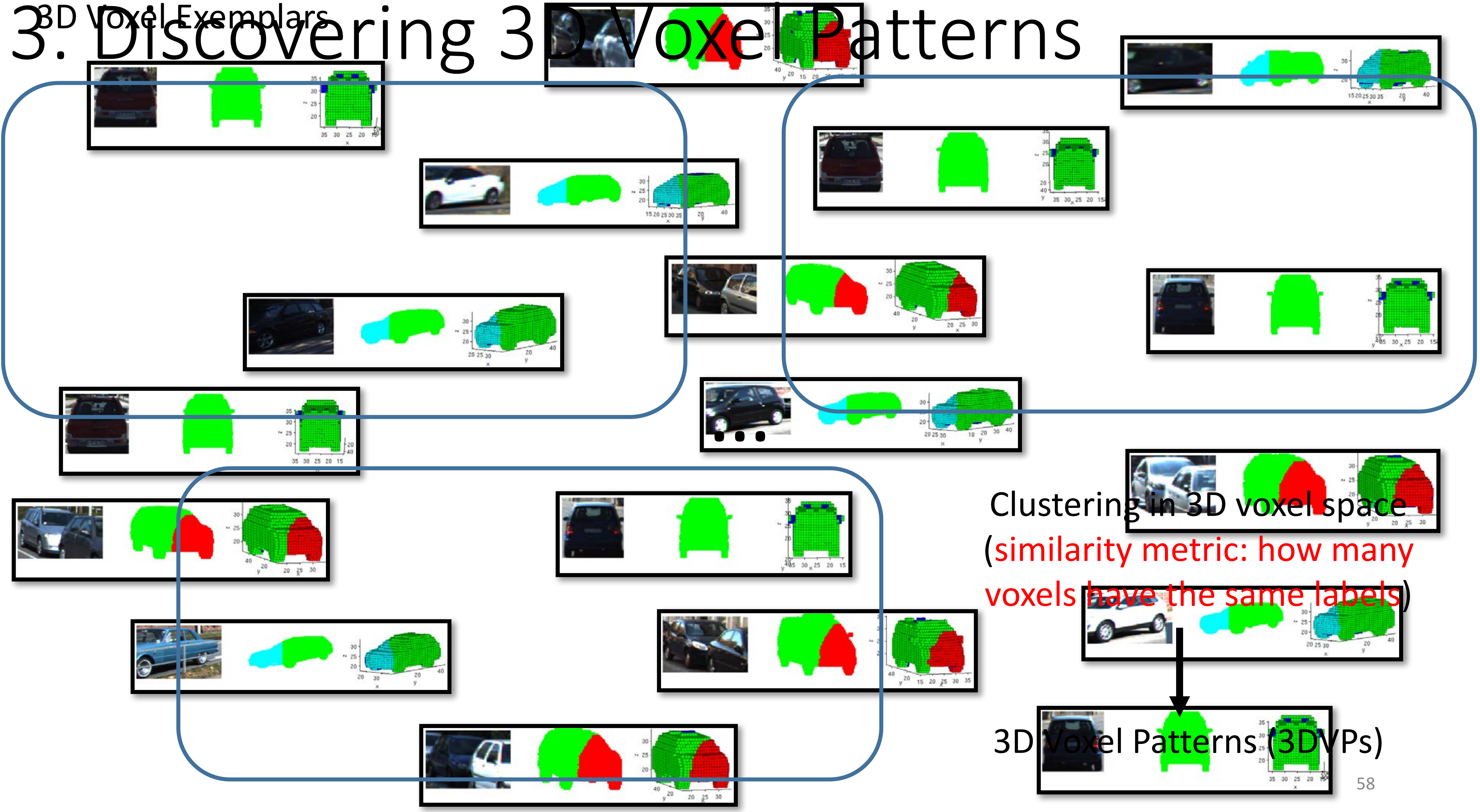


2. Building 3D Voxel Exemplars

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



3. Discovering 3D Voxel Patterns

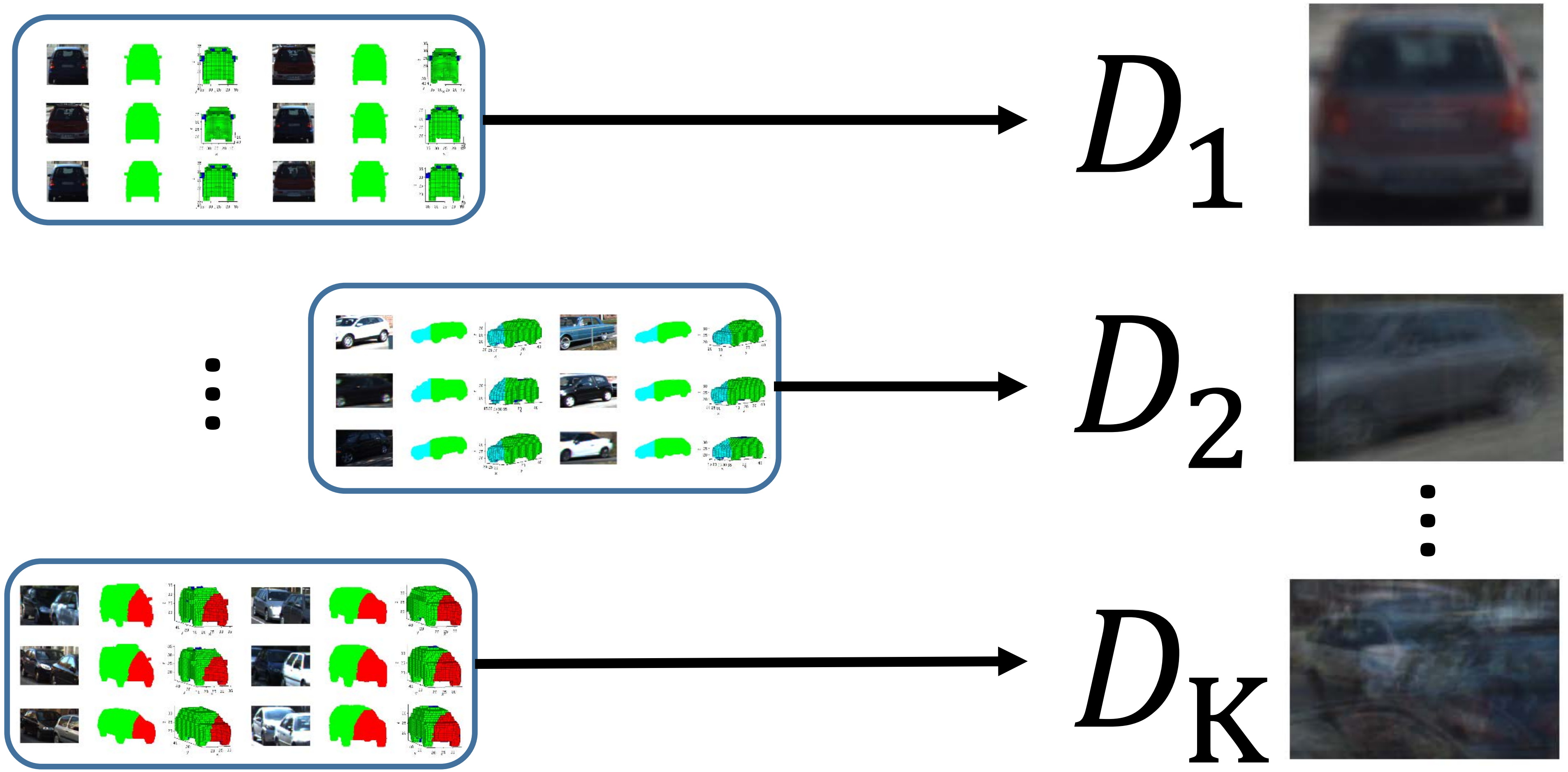


3D Voxel Exemplars

Clustering in 3D voxel space
(similarity metric: how many voxels have the same labels)

3D Voxel Patterns (3DVPs)

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.

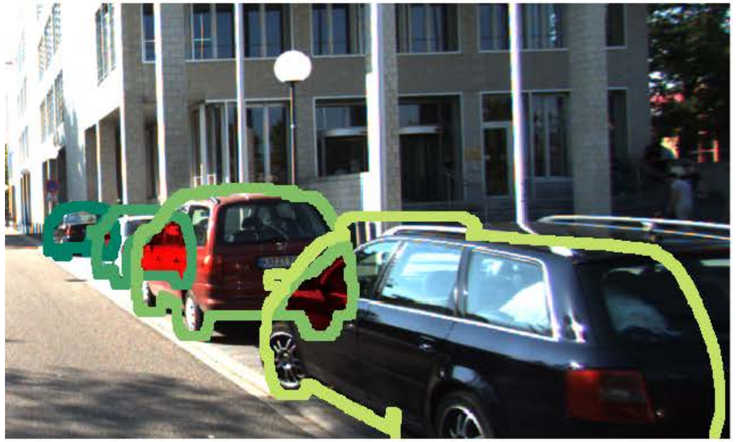
Testing Pipeline Overview

Input 2D image

1. Apply 3DVP detectors

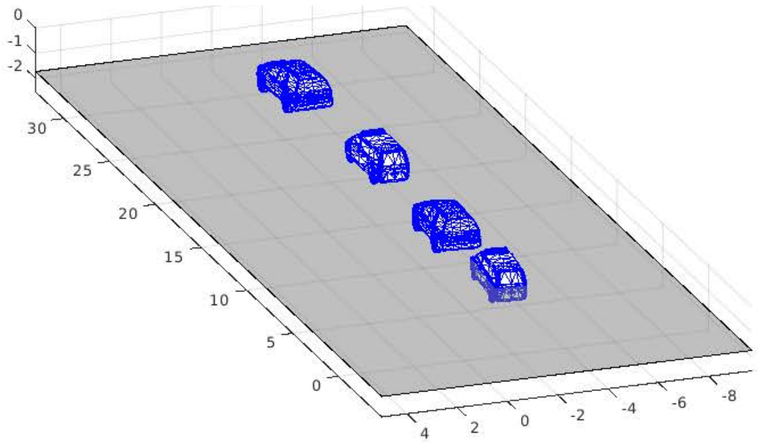
2D detection

2. Transfer meta-data
3. Occlusion reasoning



2D segmentation

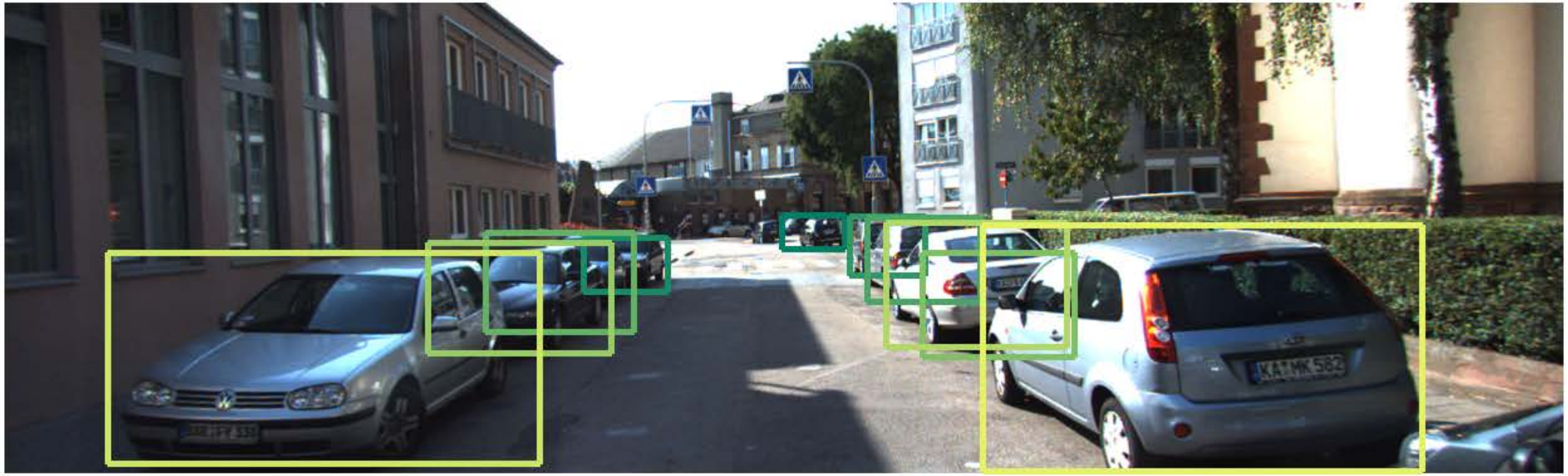
4. Backproject to 3D



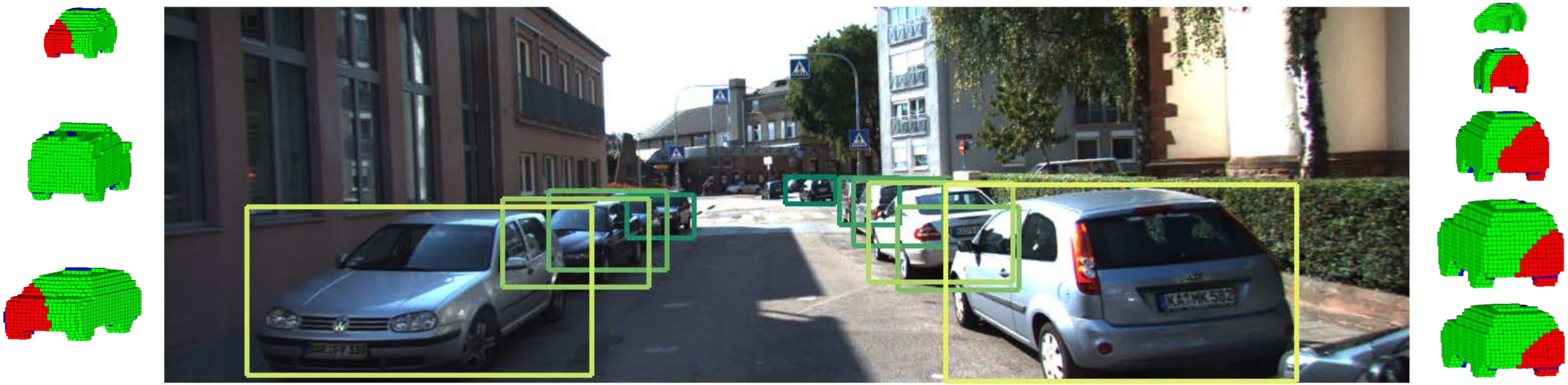
3D localization

1. Apply 3DVP Detectors

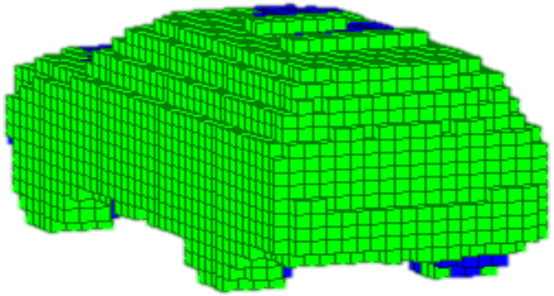
1. Apply 3DVP Detectors



2. Transfer Meta-Data (from cluster centers)



3D Voxel Patterns



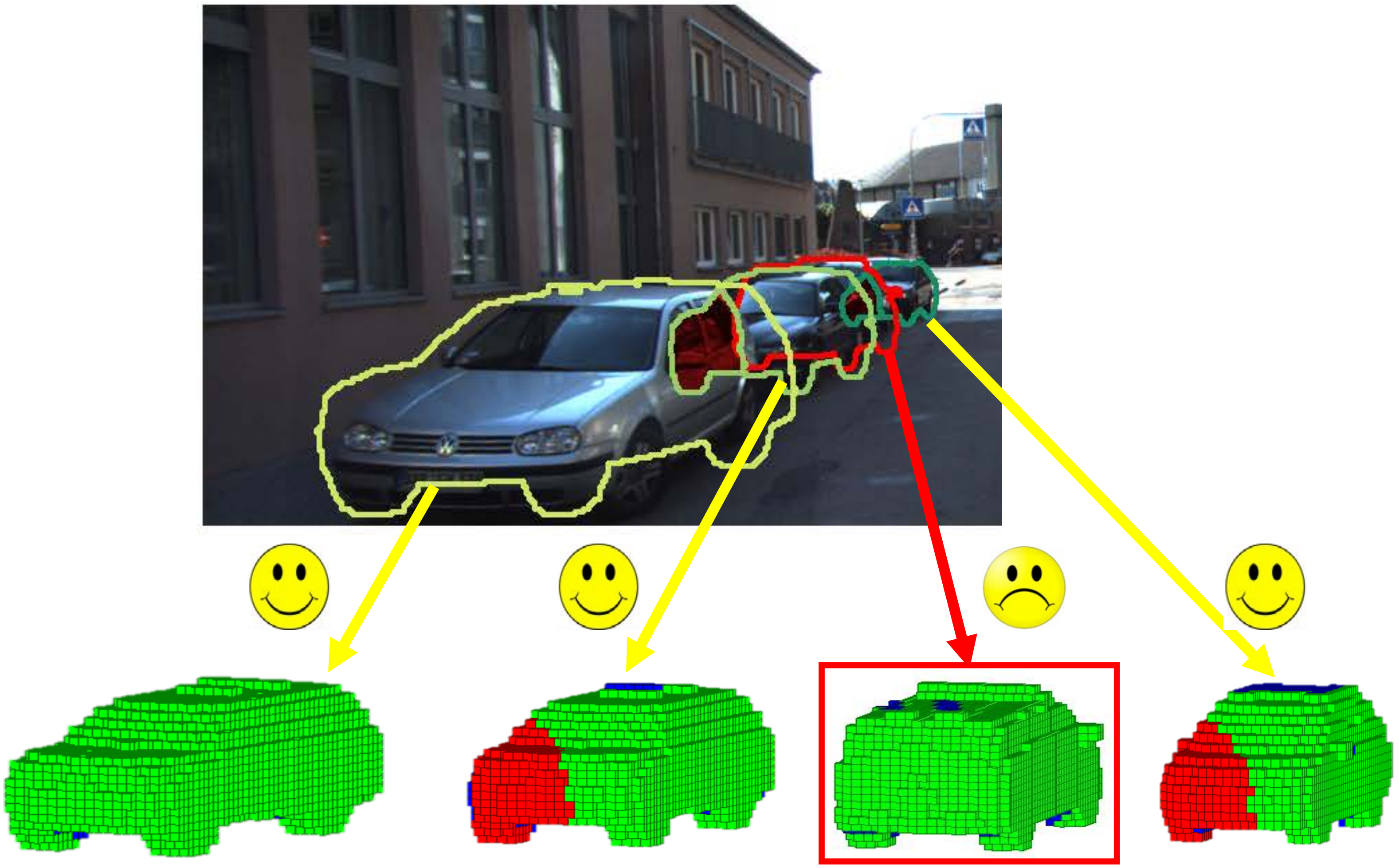
2. Transfer Meta-Data

3. Occlusion Reasoning

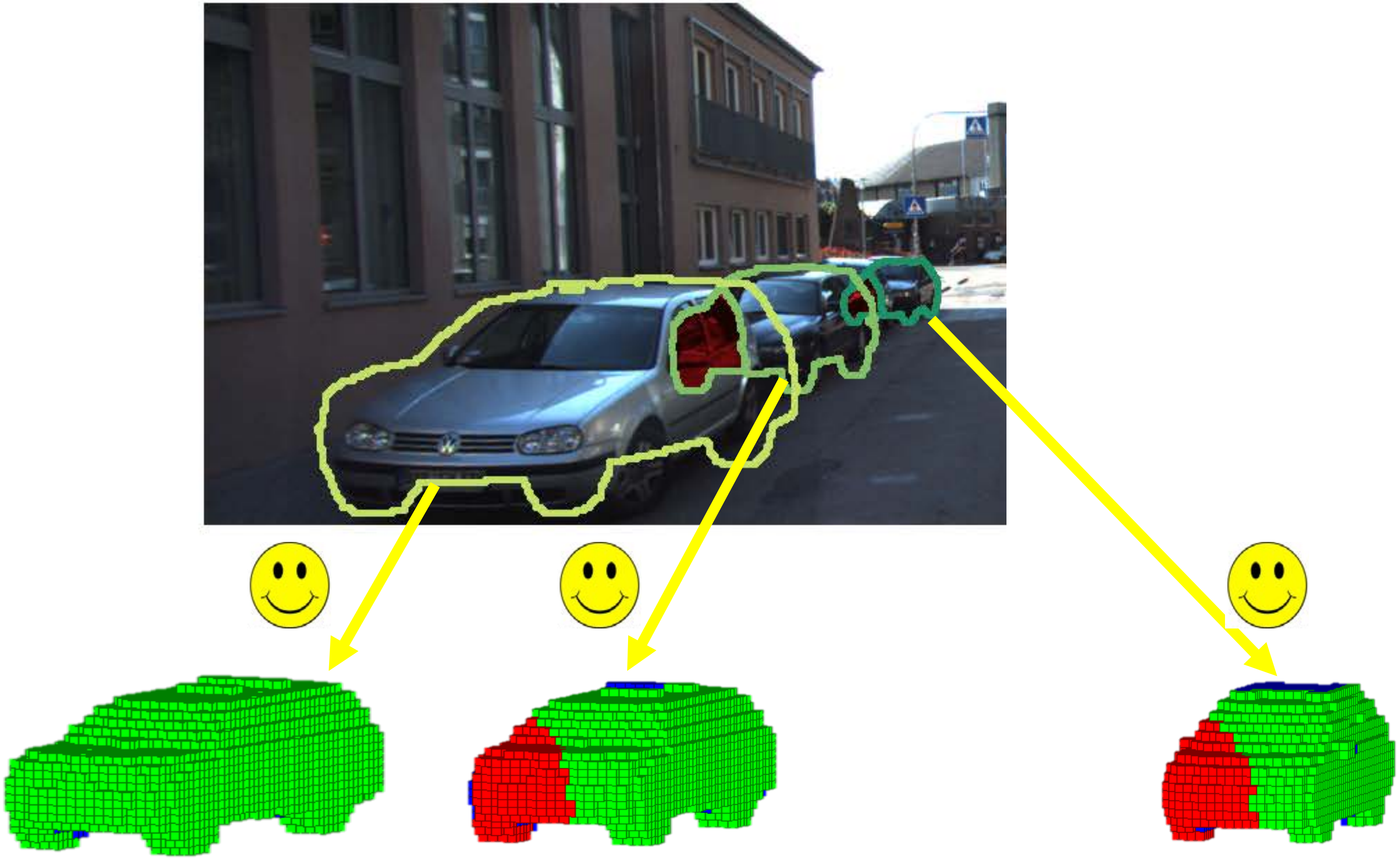
Occlusion reasoning: find a set of visibility-compatible detections

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

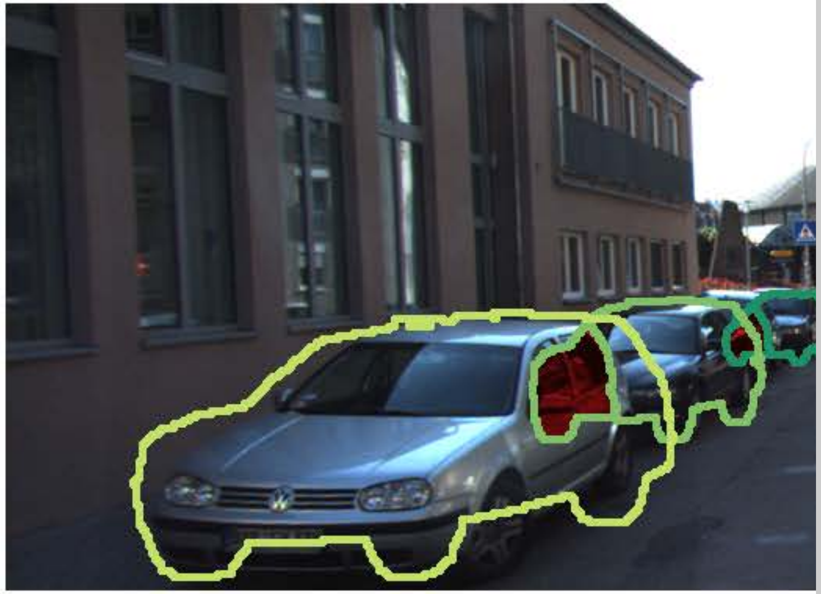
3. Occlusion Reasoning

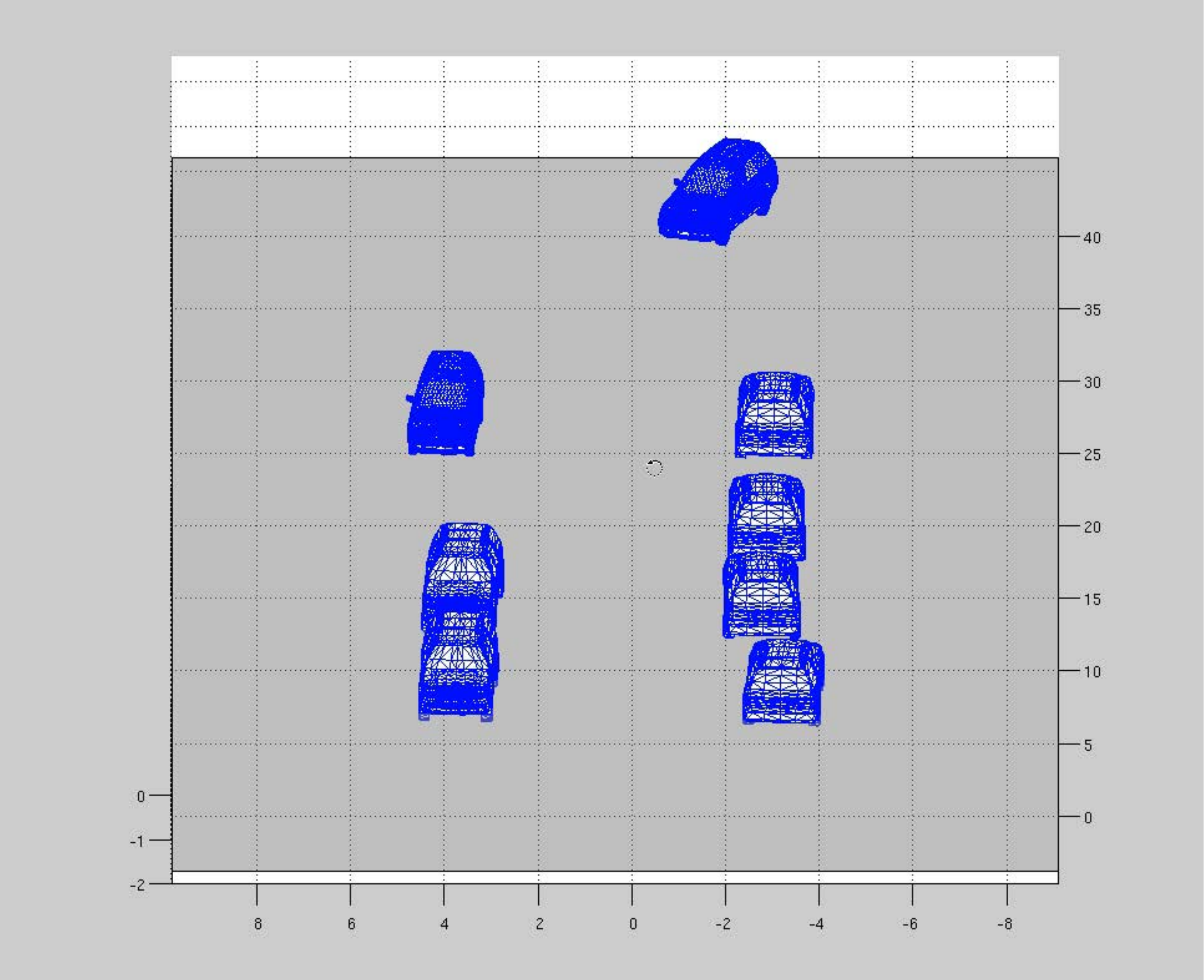


3. Occlusion Reasoning



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

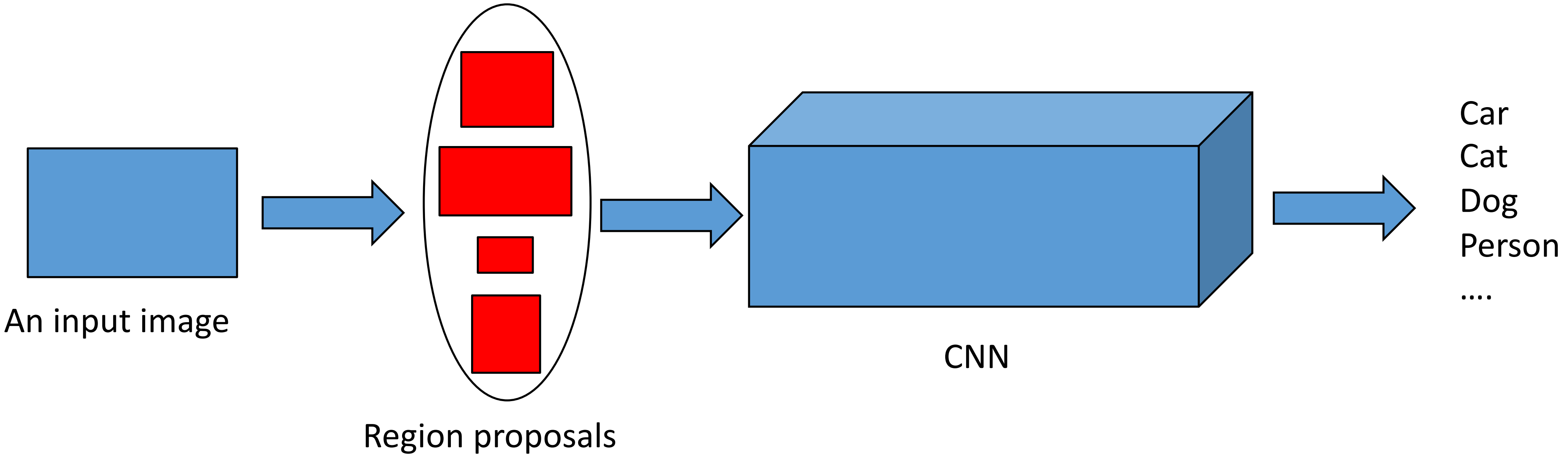
	Object Detection (AP)				Object Detection and Orientation estimation (AOS)		
Method	**Easy**	**Moderate**	**Hard**		**Easy**	**Moderate**	**Hard**
ACF [1]	55.89	54.77	42.98		N/A	N/A	N/A
DPM [2]	68.02	56.48	44.18		67.27	55.77	43.59
DPM-VOC+VP [3]	74.95	64.71	48.76		72.28	61.84	46.54
OC-DPM [4]	74.94	65.95	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.

Can we exploit 3D object representations in deep learning?

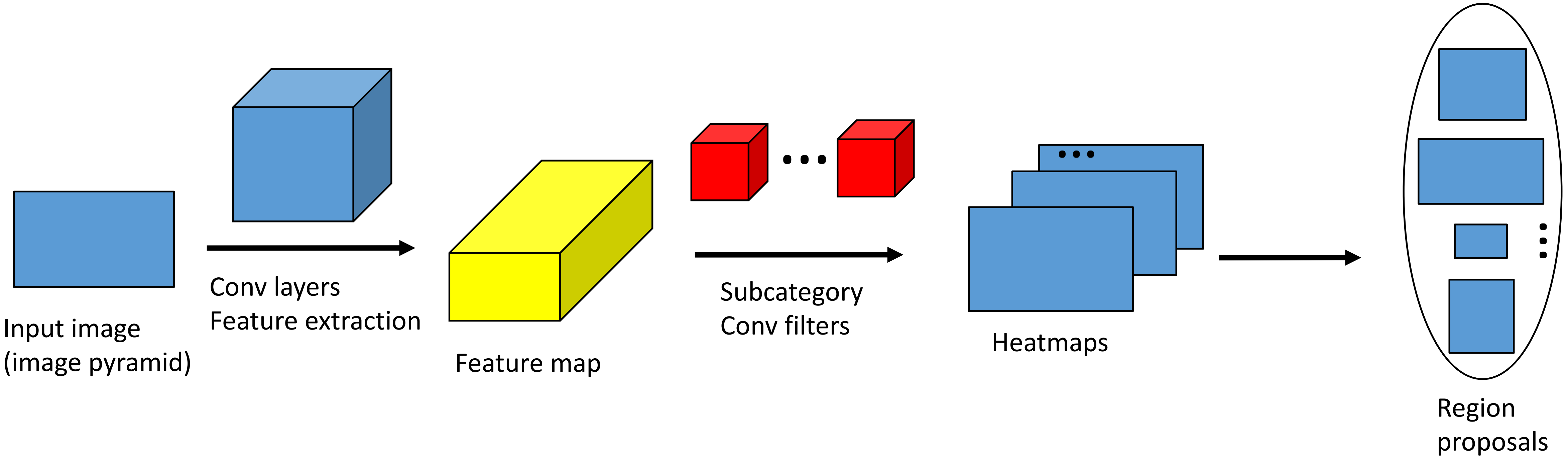
Our first trial: 3D voxel patterns as subcategories

Two-stage Object 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

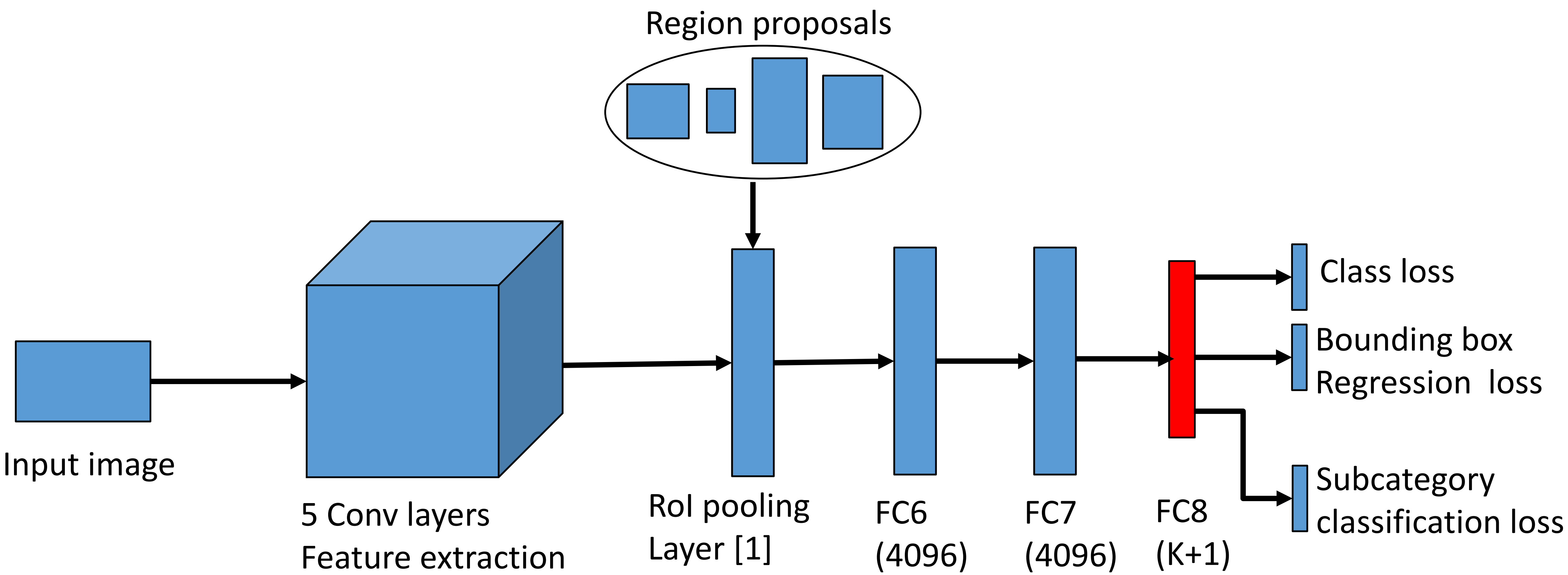


Feature Extrapolating Layer
- Generate features in nearby scales by extrapolating

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

Y. Xiang, W. Choi, Y. Lin and S. Savarese. Subcategory-aware CNNs for Object Proposals and Detection. arXiv:1604.04693, 2016.

Subcategory-ware Detection Network



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

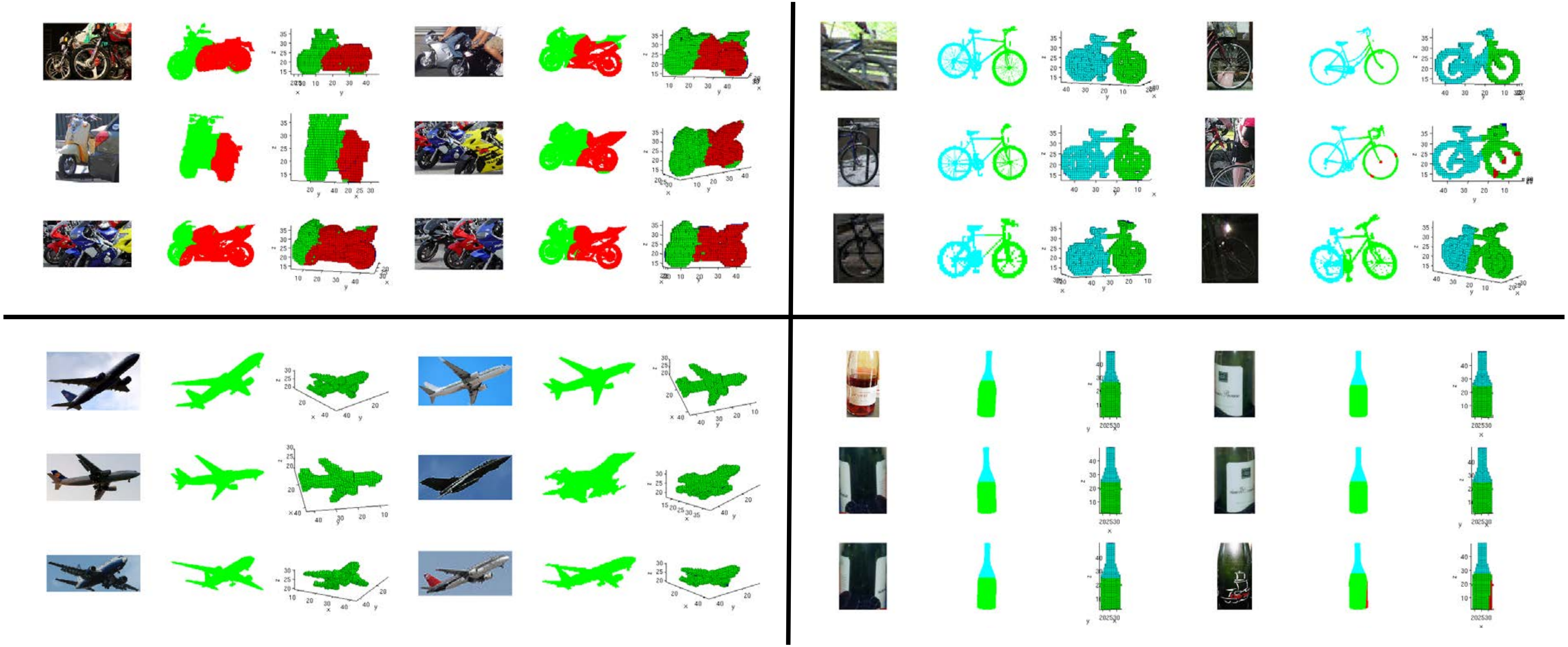
Car Detection and Orientation Estimation on KITTI

	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	90.81	**89.04**	**79.27**		90.67	**88.62**	**78.68**

[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 5
Pose : Rank 1

3D Voxel Patterns from PASCAL3D+ [1]



12 Rigid Categories

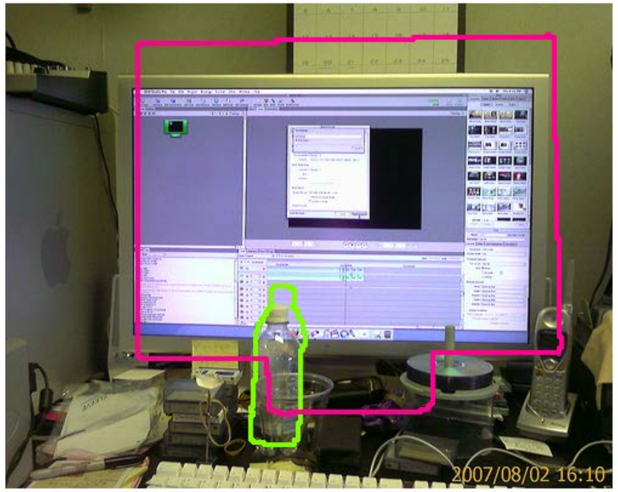
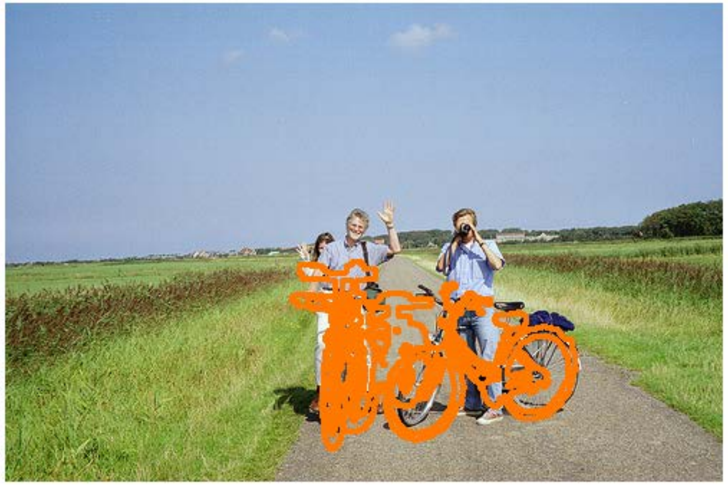
[1] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In WACV, 2014.

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**

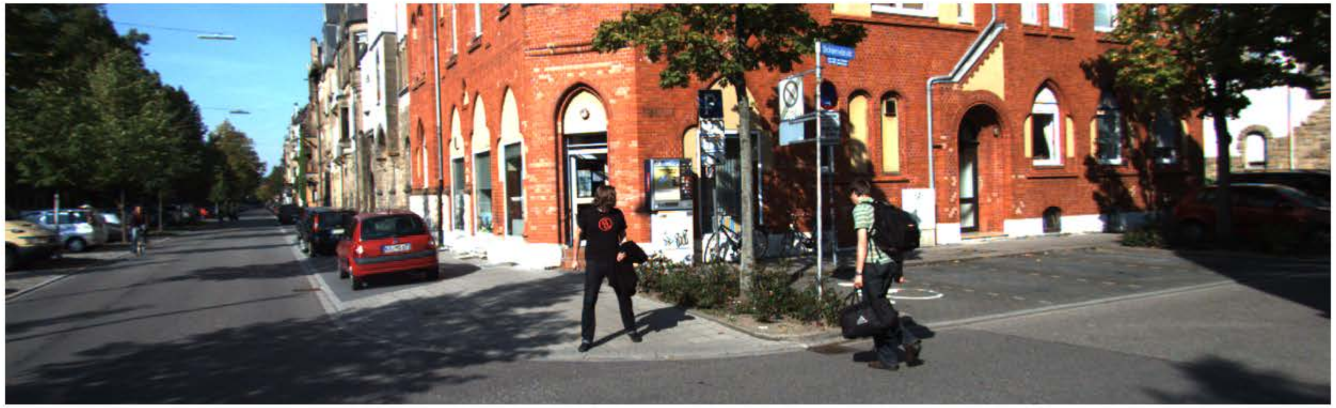
[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv preprint 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.

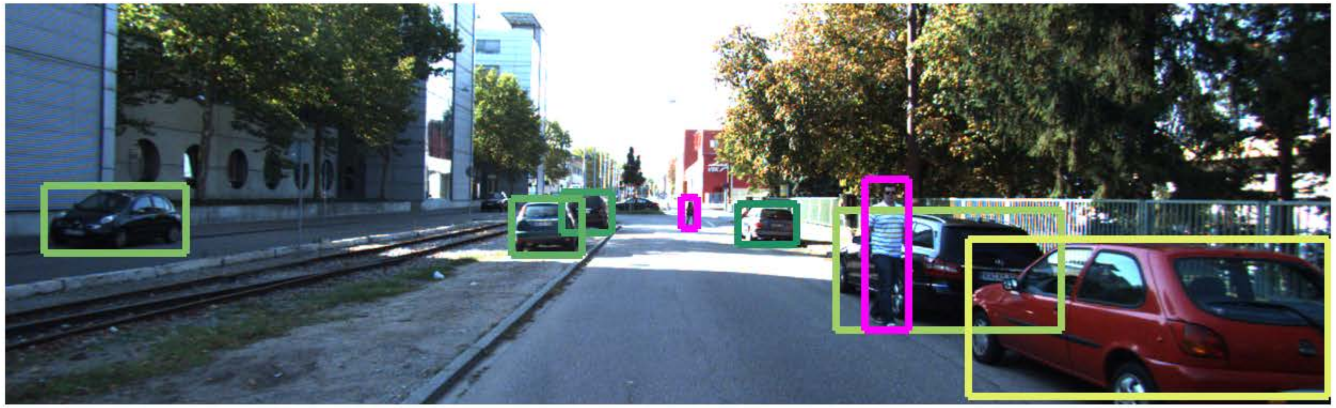

2007/08/02 16:10


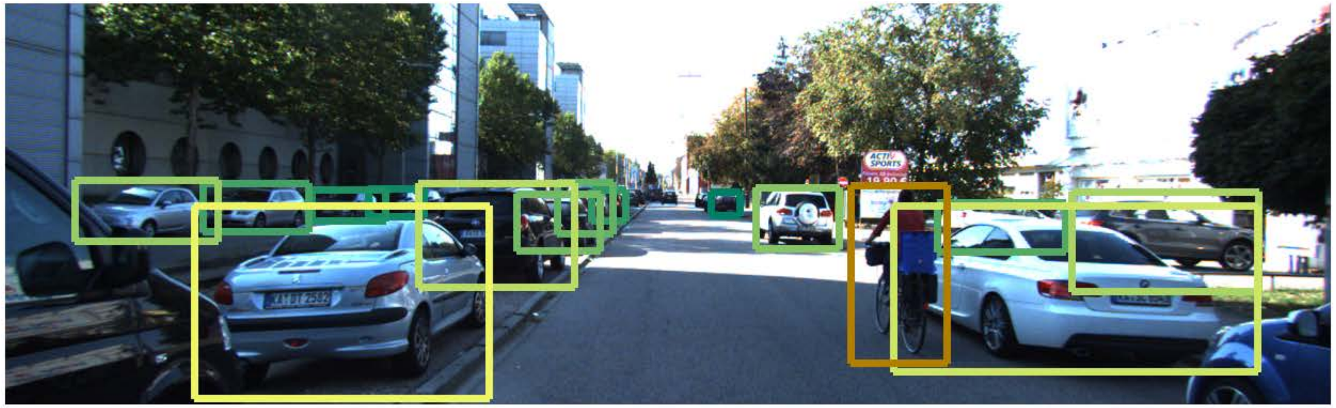


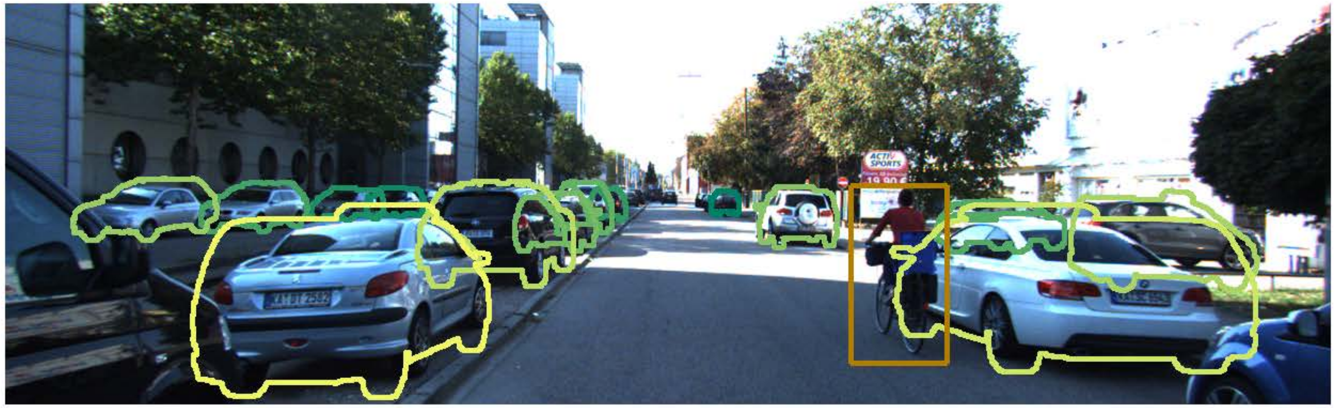

48305

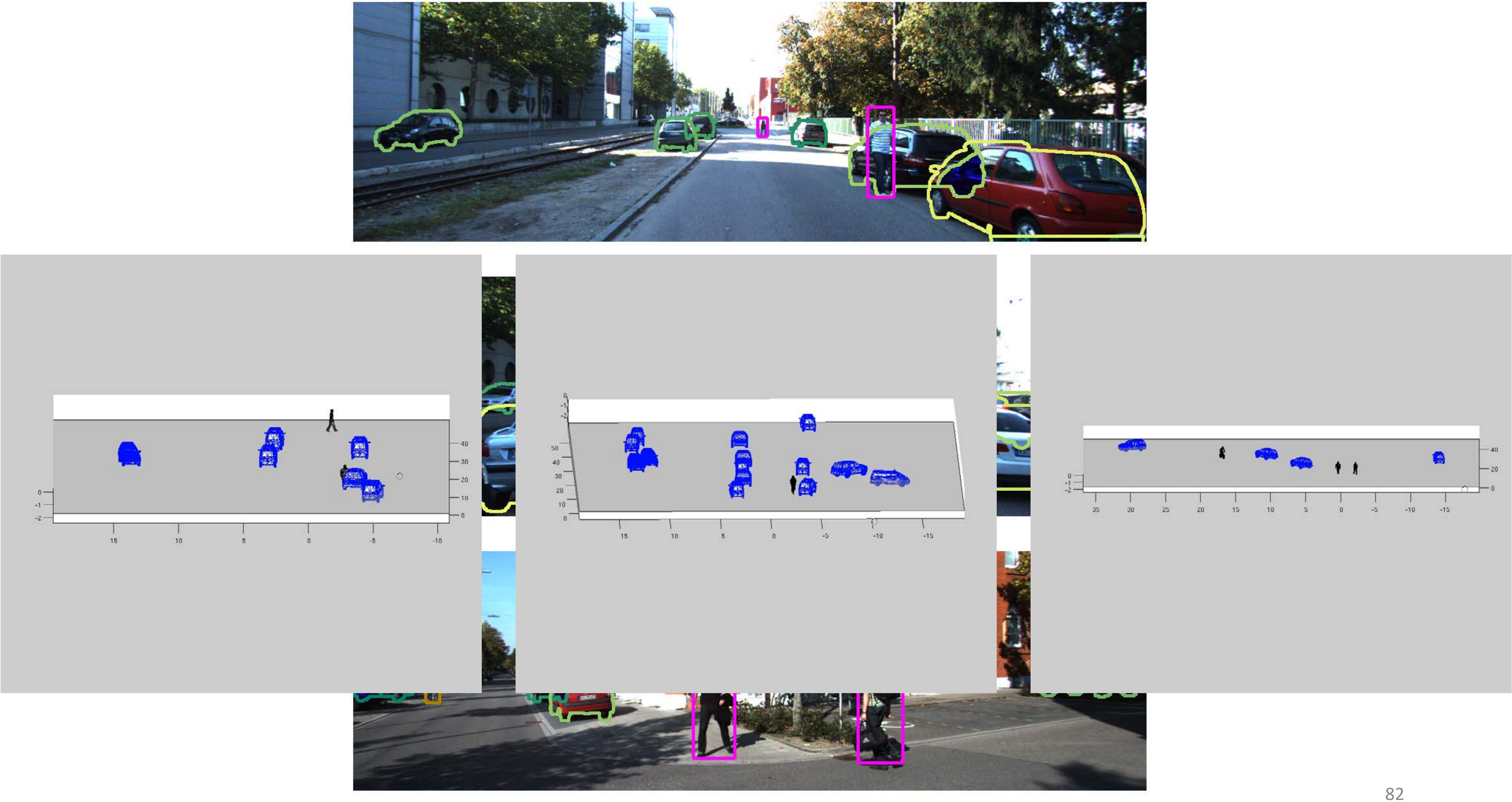

ACTIV SPORTS
19,90 €









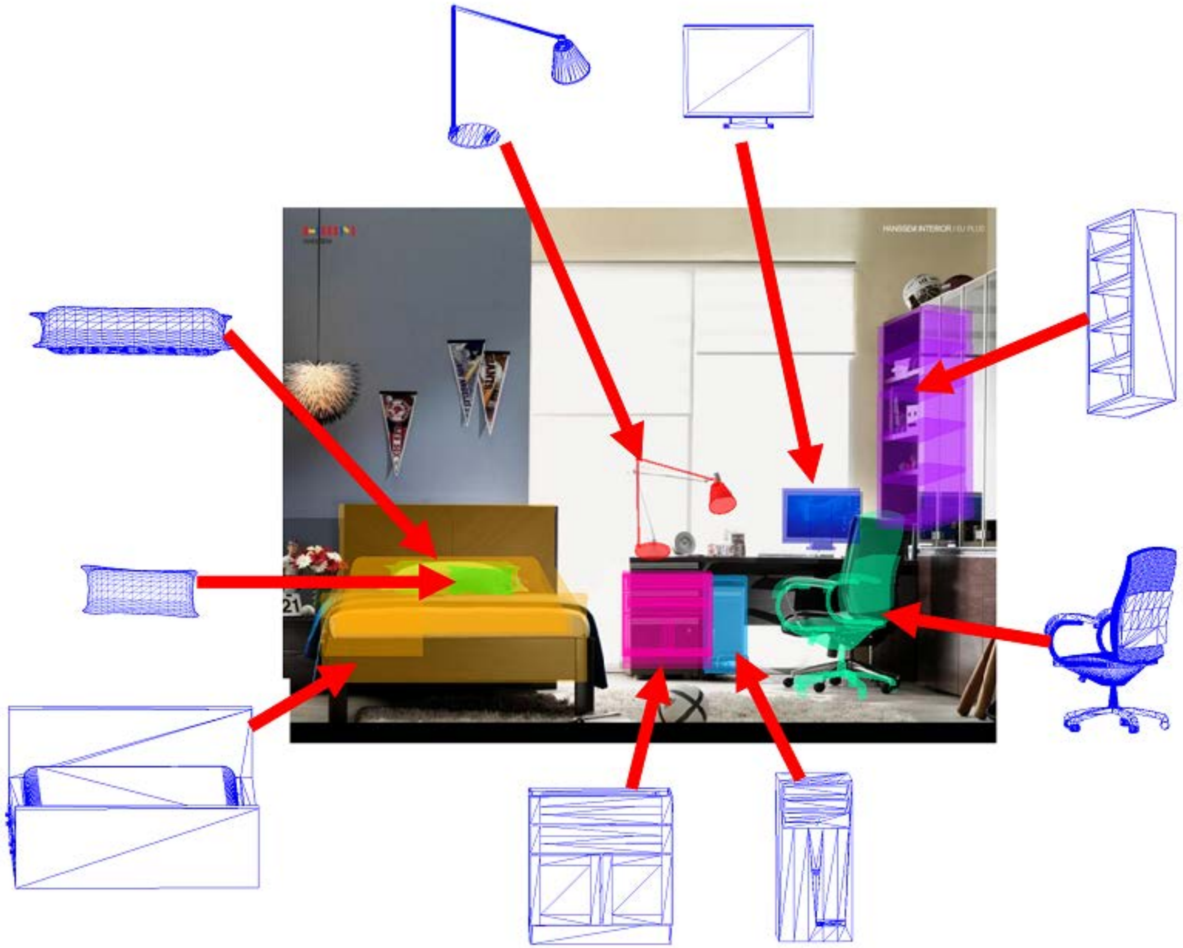


Outline

- 3D Aspect Part Representation
- 3D Voxel Pattern Representation
- A Benchmark for 3D Object Recognition in the Wild
- Summary

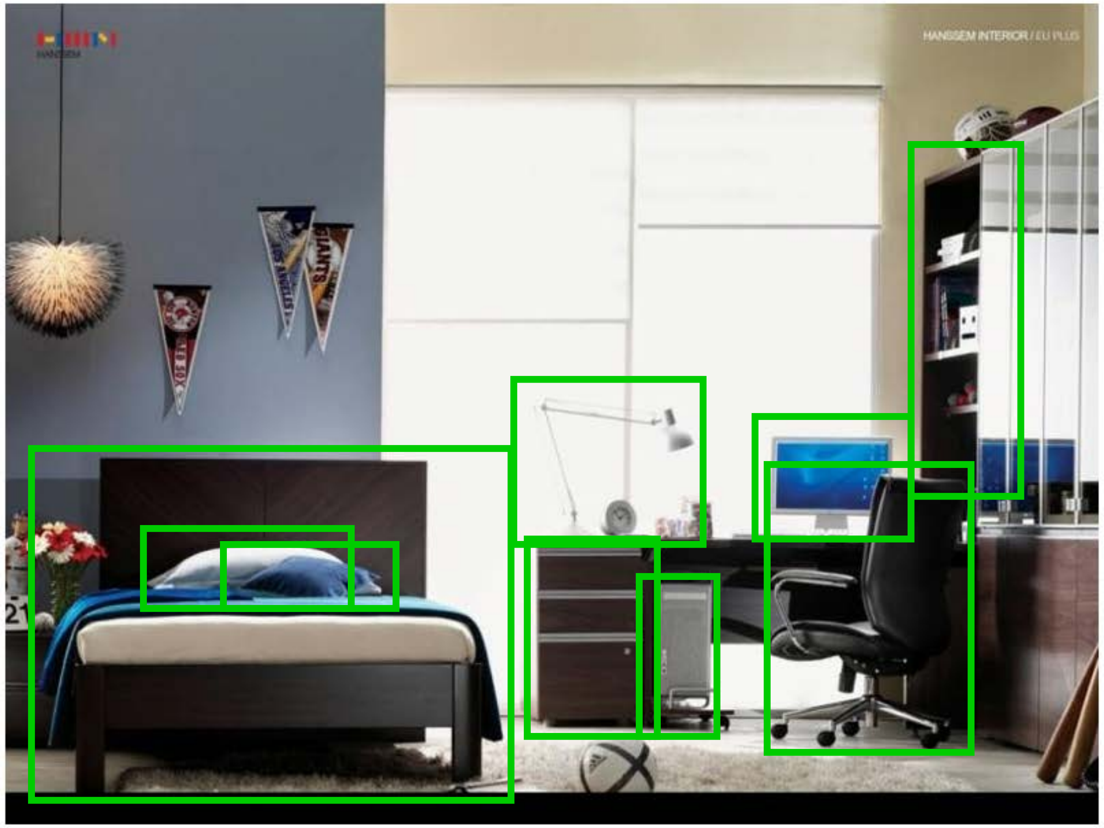
ObjectNet3D Database

- A large scale database for 3D object recognition

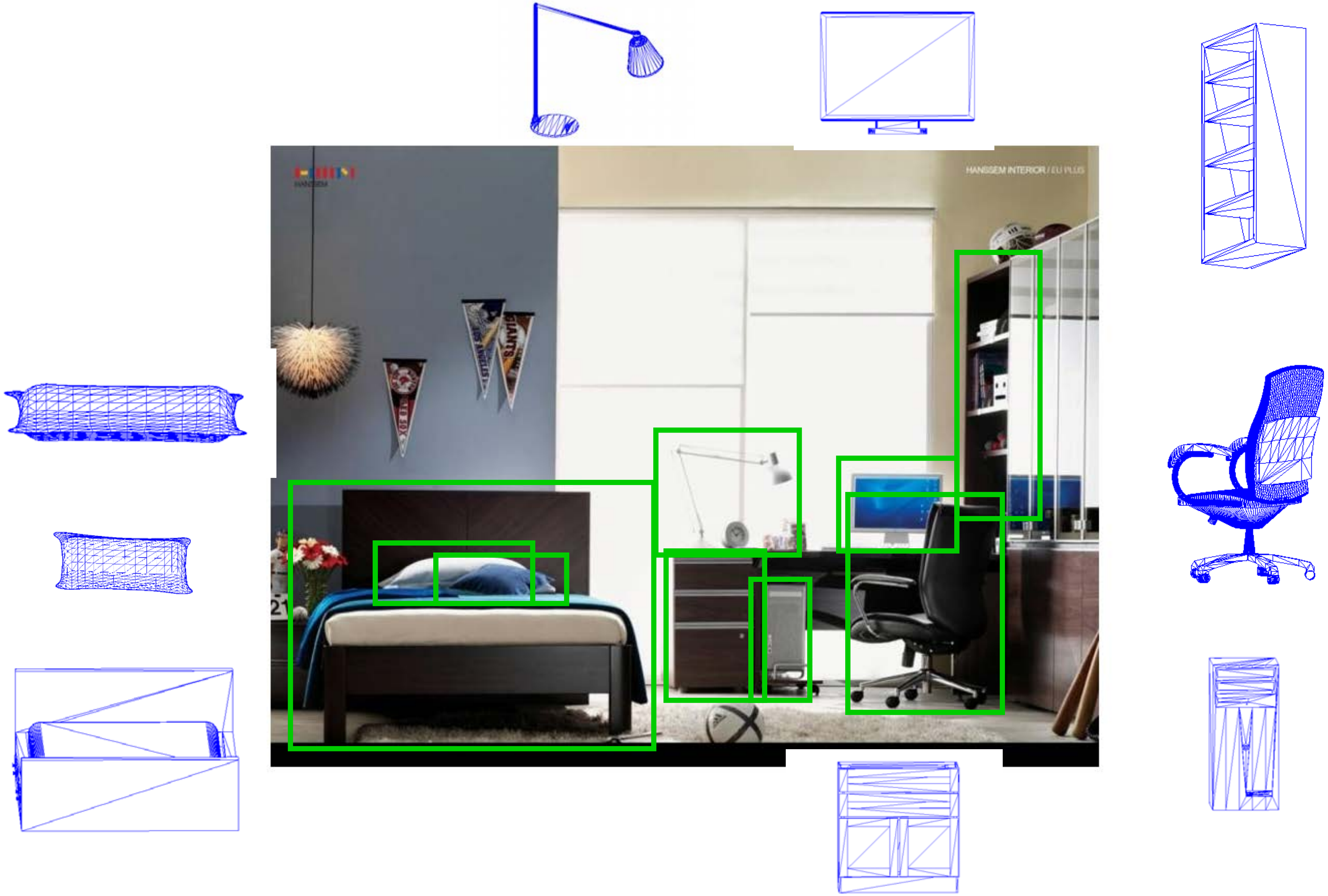


3D Annotation: 2D-3D Alignment

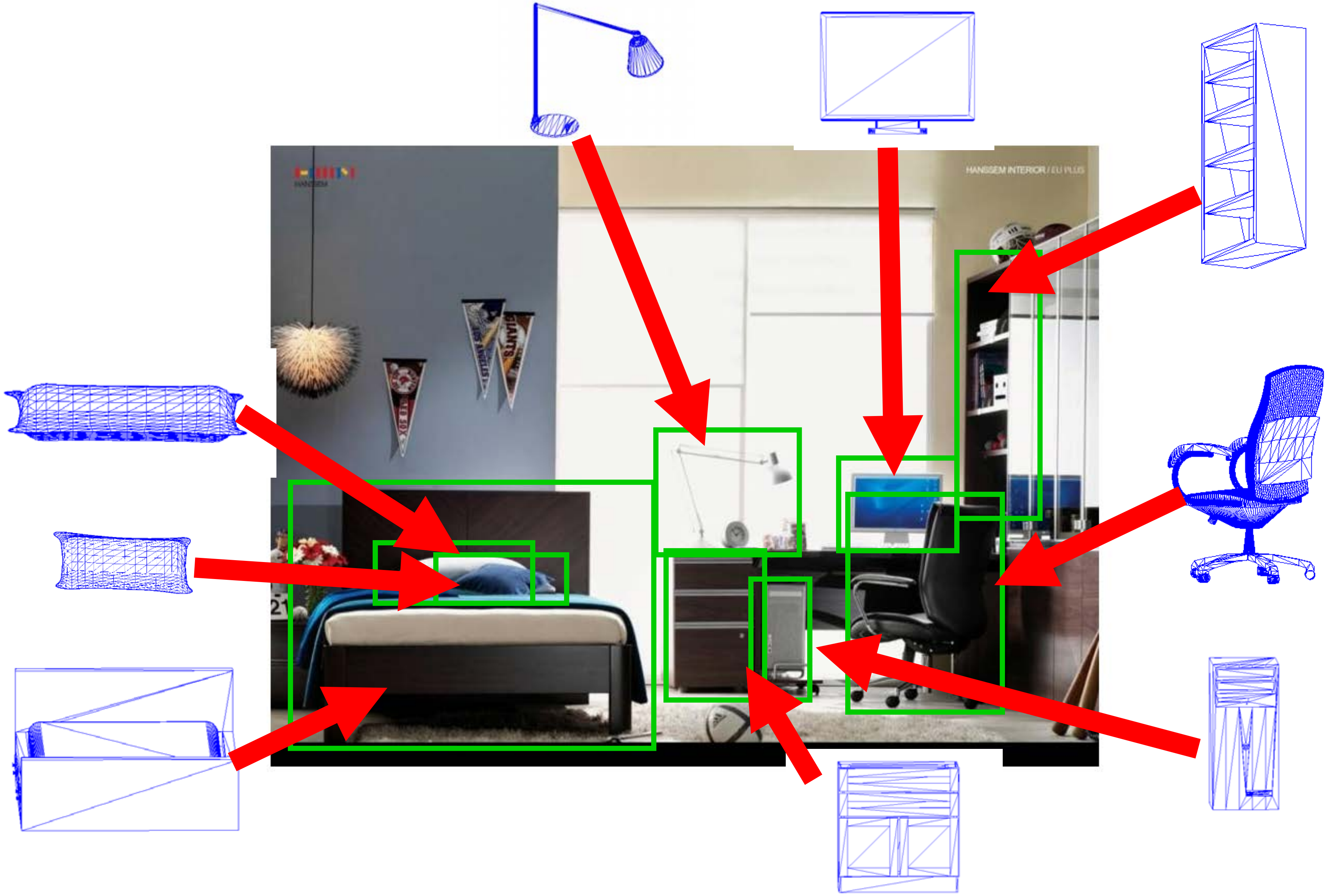
3D Annotation: 2D-3D Alignment



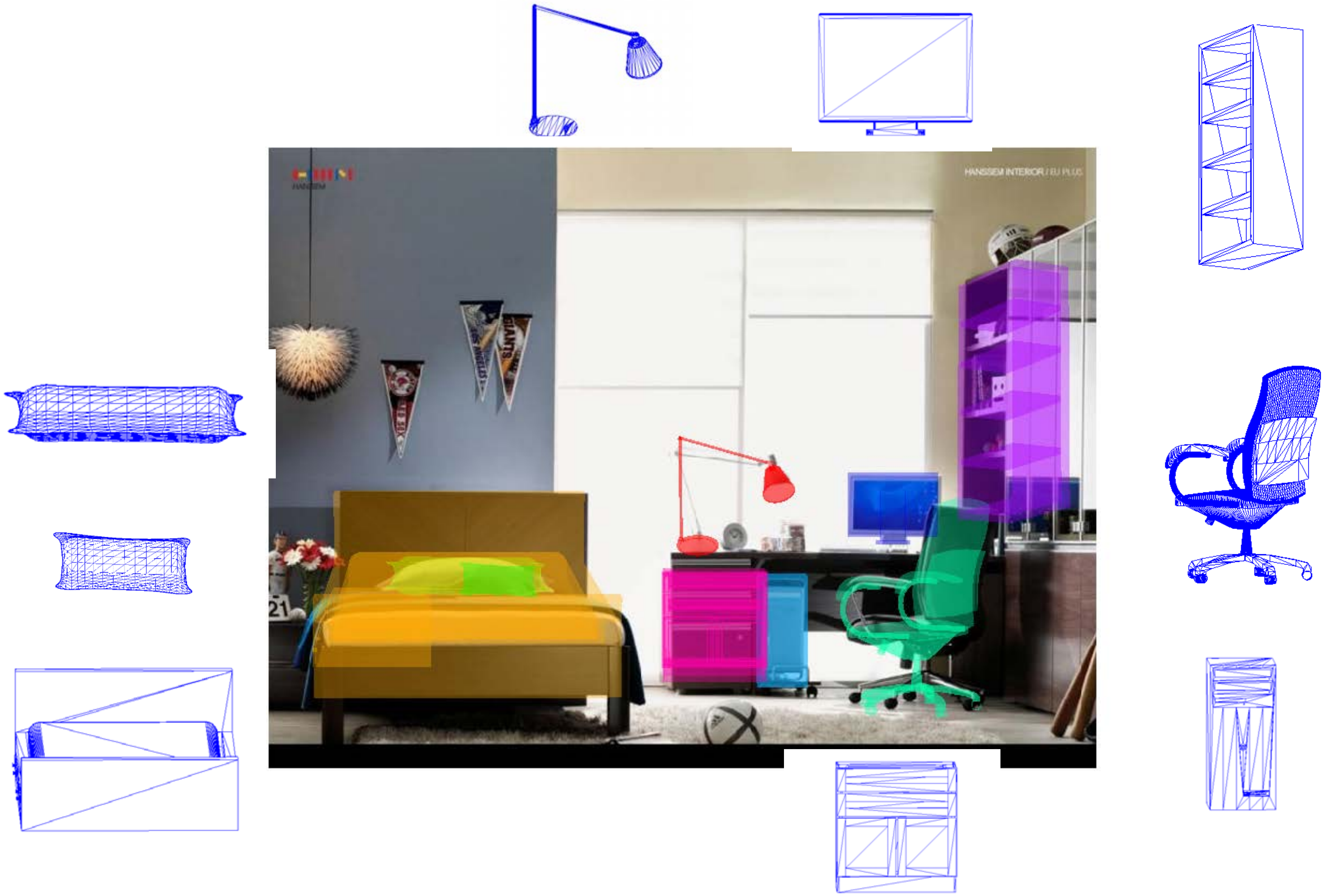
3D Annotation: 2D-3D Alignment



3D Annotation: 2D-3D Alignment



3D Annotation: 2D-3D Alignment



Comparison with Previous Datasets

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘
EPFL Car [2]	1	20	✘	✓	✘
RGB-D Object [3]	51	300	✘	✓	✘
PASCAL VOC [4]	20	27,450	✓	✘	✘
KITTI [5]	3	80,256	✓	✓	✘
PASCAL3D+ [6]	12	35,672	✓	✓	✓79

[1] S. Savarese and L. Fei-Fei. 3d generic object categorization, localization and pose estimation. In ICCV, 2007.
[2] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In CVPR, 2009.
[3] K. Lai, L. Bo, X. Ren and D. Fox. A large-scale hierarchical multi-view RGB-D object dataset. In ICRA, 2011.
[4] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 2010.
[5] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[6] Y. Xiang, R. Mottaghi and S. Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In WACV, 2014.

Comparison with Previous Datasets

	#category	#instance	Non-centered objects	Dense viewpoint	3D Shape
3D Object [1]	10	100	✘	✘	✘
EPFL Car [2]	1	20	✘	✓	✘
RGB-D Object [3]	51	300	✘	✓	✘
PASCAL VOC [4]	20	27,450	✓	✘	✘
KITTI [5]	3	80,256	✓	✓	✘
PASCAL3D+ [6]	12	35,672	✓	✓	✓ 79
ObjectNet3D	**100**	**201,888**	✓	✓	✓ **44,147**

[1] S. Savarese and L. Fei-Fei. 3d generic object categorization, localization and pose estimation. In ICCV, 2007.
[2] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In CVPR, 2009.
[3] K. Lai, L. Bo, X. Ren and D. Fox. A large-scale hierarchical multi-view RGB-D object dataset. In ICRA, 2011.
[4] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 2010.
[5] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[6] Y. Xiang, R. Mottaghi and S. Savarese. Beyond PASCAL: A benchmark for 3D object detection in the wild. In WACV, 2014.

Database Construction: Object Categories

- 100 rigid object categories

Aeroplane	Cap	Filing cabinet	Lighter	Remote control	Suitcase
Ashtray	Car	Fire extinguisher	Mailbox	Rifle	Teapot
Backpack	Cellphone	Fish tank	Microphone	Road pole	Telephone
Basket	Chair	Flashlight	Microwave	Satellite dish	Toaster
Bed	Clock	Fork	Motorbike	Scissors	Toilet
Bench	Coffee maker	Guitar	Mouse	Screwdriver	Toothbrush
Bicycle	Comb	Hair dryer	Paintbrush	Shoe	Train
Backboard	Computer	Hammer	Pan	Shovel	Trash bin
Boat	Cup	Headphone	Pen	Sign	Trophy
Bookshelf	Desk lamp	Helmet	Pencil	Skate	Tub
Bottle	Dining table	Iron	Piano	Skateboard	Tvmonitor
Bucket	Dishwasher	Jar	Pillow	Slipper	Vending machine
Bus	Door	Kettle	Plate	Sofa	Washing machine
Cabinet	Eraser	Key	Pot	Speaker	Watch
Calculator	Eyeglasses	Keyboard	Printer	Spoon	Wheelchair
Camera	Fan	Knife	Racket	Stapler	
Can	Faucet	Laptop	Refrigerator	Stove	

Database Construction: Object Categories

- 100 rigid object categories

Aeroplane	Cap	Filing cabinet	Lighter	Remote control	Suitcase
Ashtray	Car	Fire extinguisher	Mailbox	Rifle	Teapot
Backpack	Cellphone	Fish tank	Microphone	Road pole	Telephone
Bask		Fl	rowave	Satell	
Bed		Fo	orbike	Scisso	
Ben	aker	G	use	Screw	n
Bicycle	Comb	Hair dryer	Paintbrush	Shoe	Train
Backboard	Computer	Hammer	Pan	Shovel	Trash bin
Boat	Cup	Headphone	Pen	Sign	Trophy
Book	k lamp		cil		
Bottl	ing table		o		
Bucke	washer		w		ine
Bus	Door	Kettle	Plate	Sofa	Washing machine
Cabinet	Eraser	Key	Pot	Speaker	Watch
Calculator	Eyeglasses	Keyboard	Printer	Spoon	Wheelchair
Camera	Fan	Knife	Racket	Stapler	
Can	Faucet	Laptop	Refrigerator	Stove	

Vehicles

Furniture

Container

Tools

Electronics

Personal items

Database Construction: Images

- 2D images from the ImageNet database [1]

[1] http://image-net.org

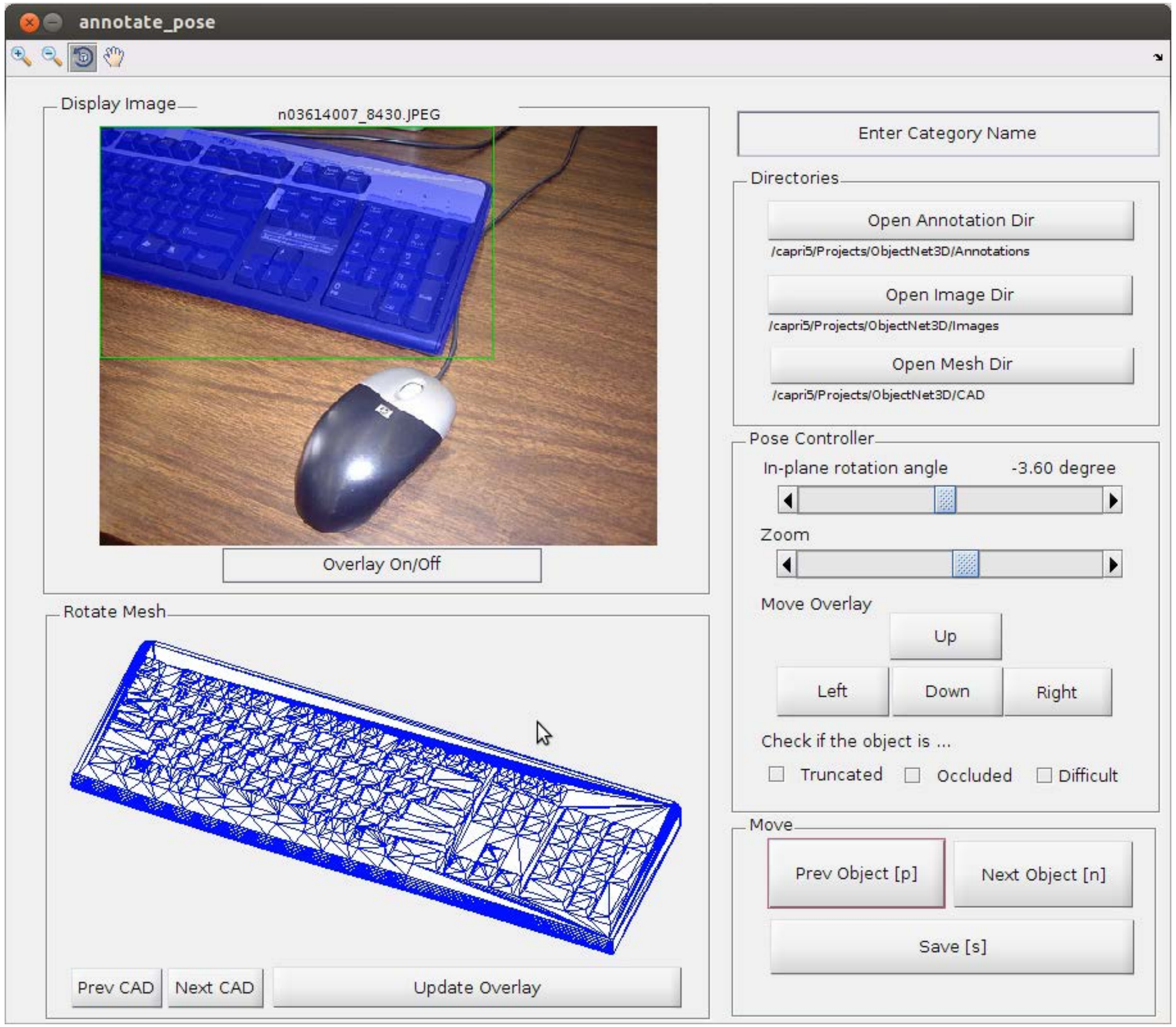
Database Construction: 3D Shapes

- Trimble 3D Warehouse [1]
- ShapeNet database [2]

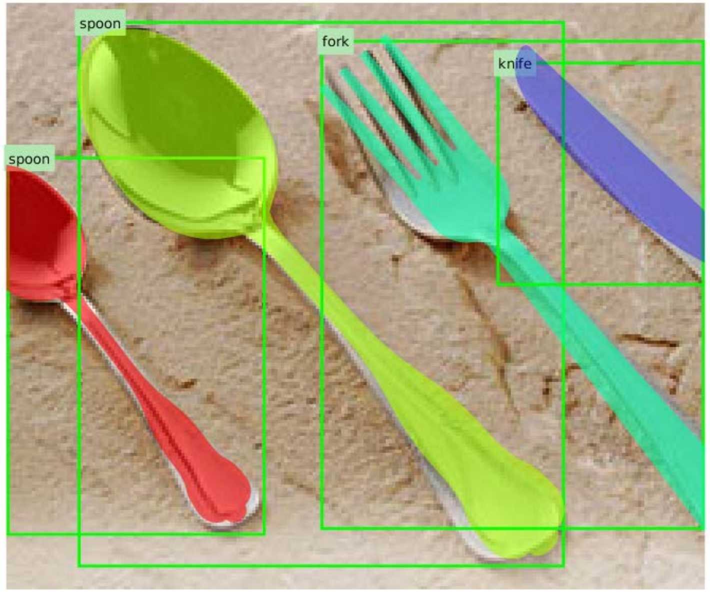
3D Shapes from Trimble 3D Warehouse
[1] https://3dwarehouse.sketchup.com

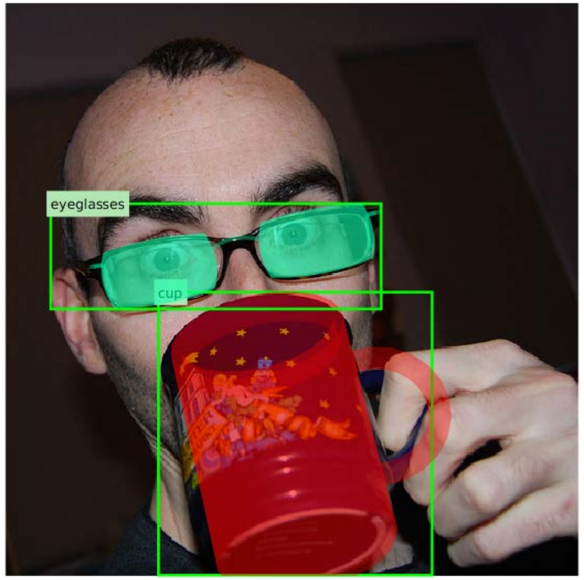
3D Shapes from ShapeNet
[2] http://shapenet.cs.stanford.edu/

Database Construction: Annotation Demo



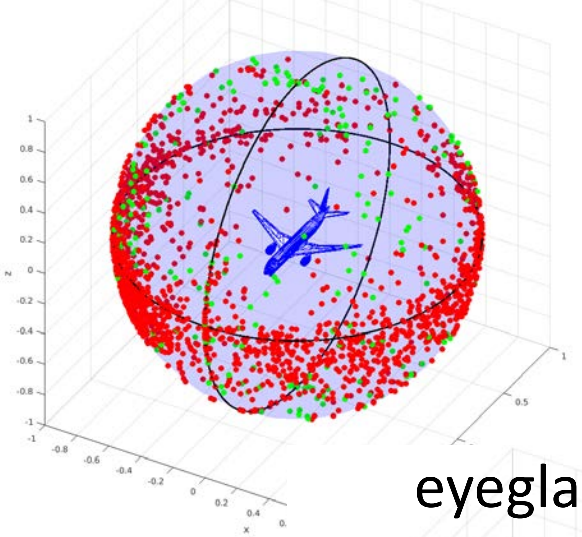
3D Pose Annotation Examples



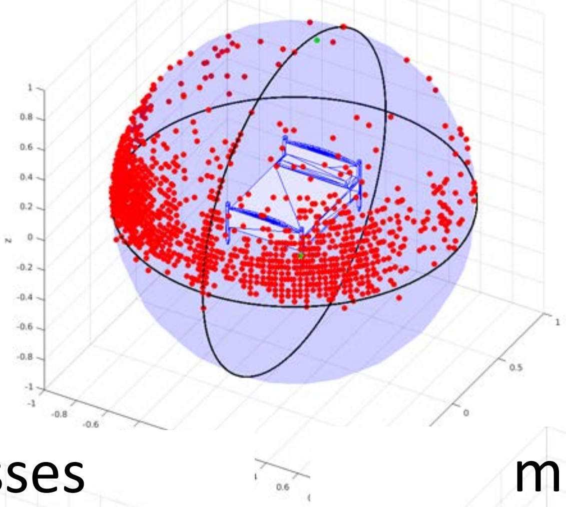


Viewpoint Distributions

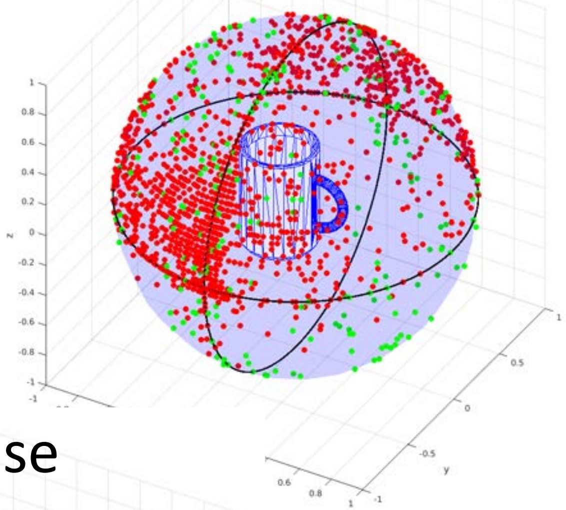
aeroplane



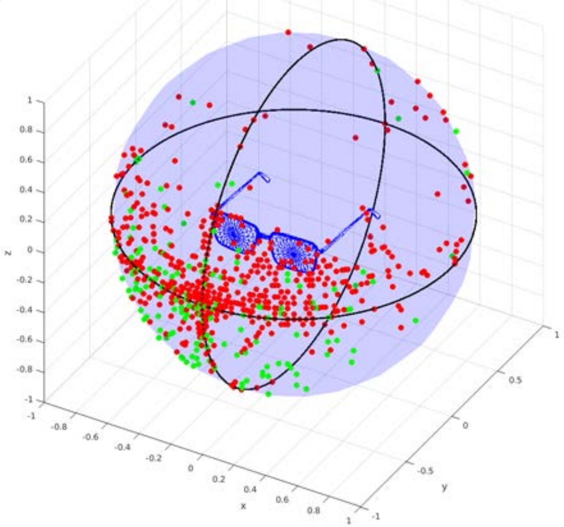
bed



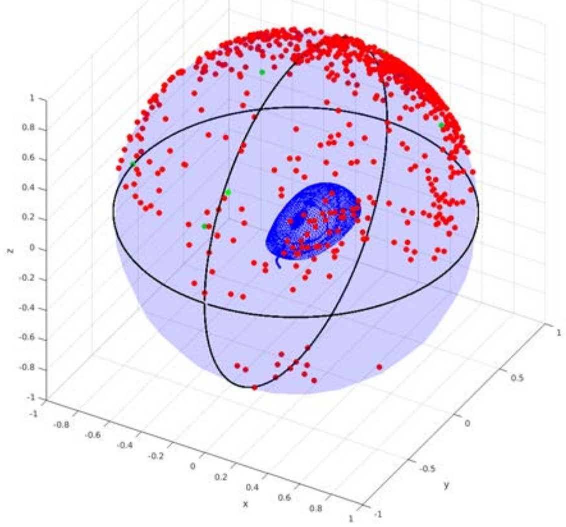
cup



eyeglasses

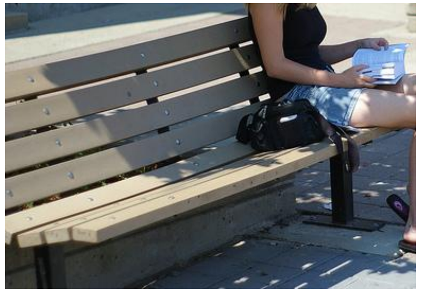


mouse



Database Construction: Image-based 3D Shape Retrieval

Test Object



Database Construction: Image-based 3D Shape Retrieval

Test Object | **Rank 1** | **Rank 2** | **Rank 3**

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

Database Construction: Image-based 3D Shape Retrieval

Test Object **Rank 1** **Rank 2** **Rank 3**

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

Baseline Experiments

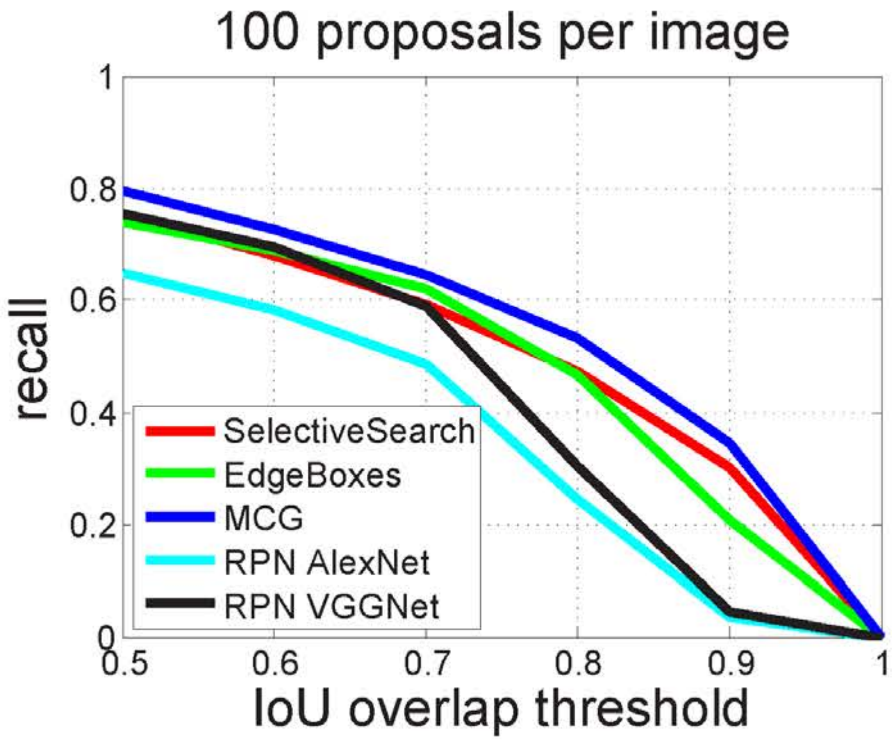
- Object proposal generation

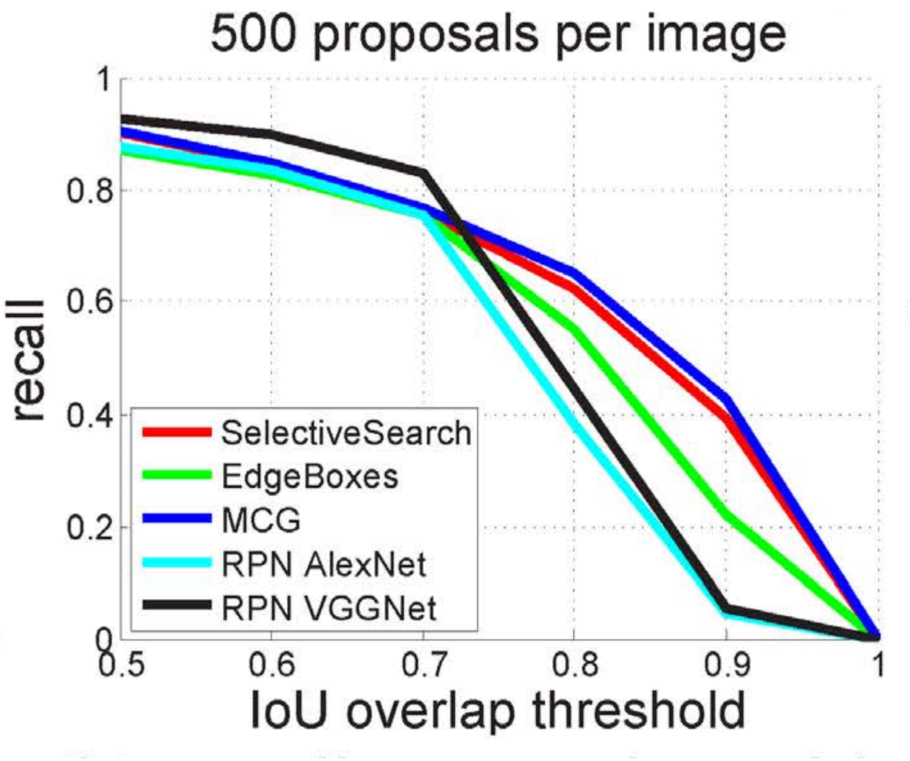
- 2D object detection

- Joint 2D detection and continuous 3D pose estimation

- Image-based 3D shape retrieval

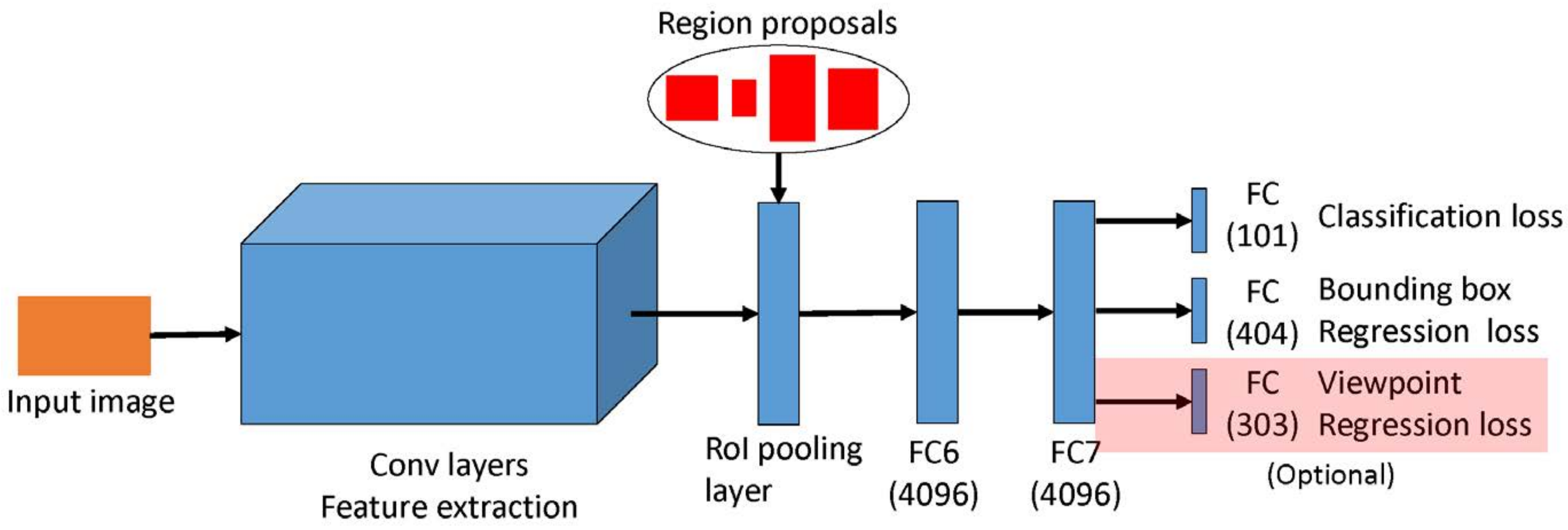
Object Proposal Generation





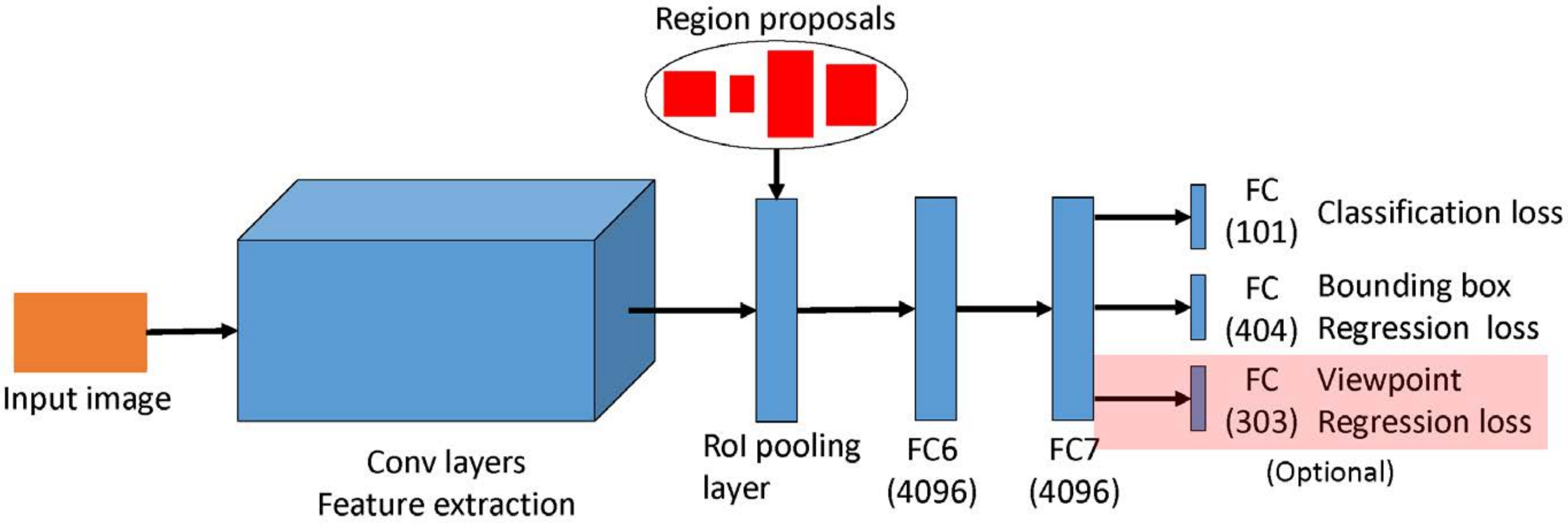
Selective Search: Uijlings et al., IJCV, 2013.
EdgeBoxes: Zitnick et al., ECCV, 2014.
MCG: Arbelaez et al., CVPR, 2014.
RPN: Ren et al., NIPS, 2015.

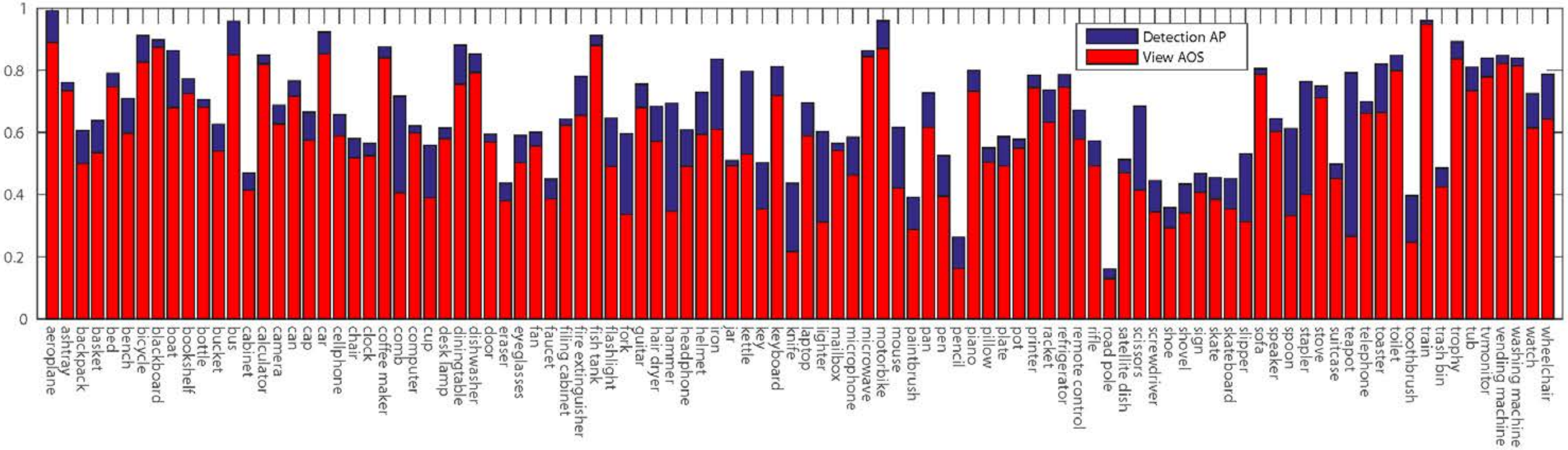
A Network for Object Detection and Pose Estimation



R. Girshick. Fast R-CNN. In ICCV, 2015.

A Network for Object Detection and Pose Estimation

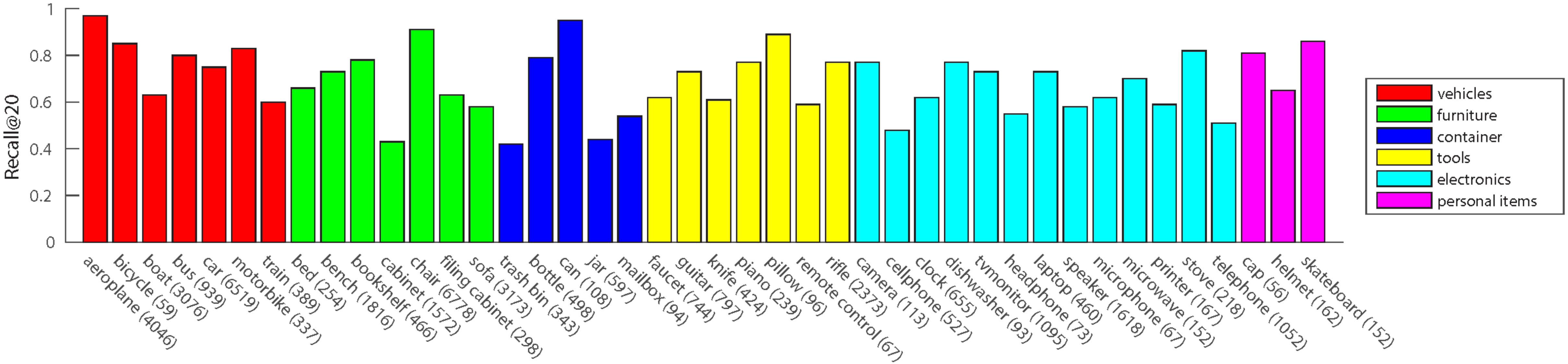




R. Girshick. Fast R-CNN. In ICCV, 2015.

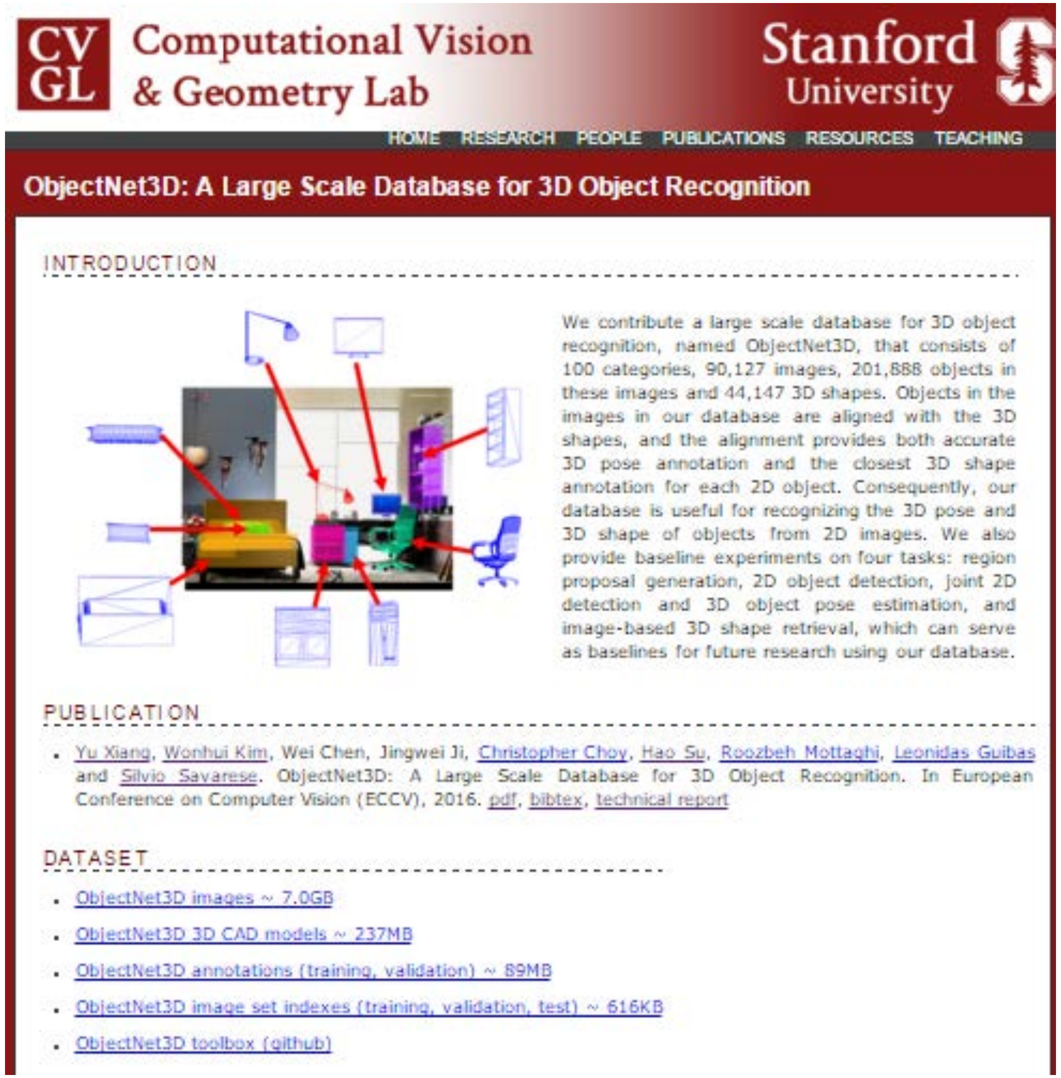
Image-based 3D Shape Retrieval

- User Study: 69.2% recall@20 for 42 categories



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

ObjectNet3D


CV GL Computational Vision & Geometry Lab

Stanford University

HOME RESEARCH PEOPLE PUBLICATIONS RESOURCES TEACHING

ObjectNet3D: A Large Scale Database for 3D Object Recognition

INTRODUCTION

We contribute a large scale database for 3D object recognition, named ObjectNet3D, that consists of 100 categories, 90,127 images, 201,888 objects in these images and 44,147 3D shapes. Objects in the images in our database are aligned with the 3D shapes, and the alignment provides both accurate 3D pose annotation and the closest 3D shape annotation for each 2D object. Consequently, our database is useful for recognizing the 3D pose and 3D shape of objects from 2D images. We also provide baseline experiments on four tasks: region proposal generation, 2D object detection, joint 2D detection and 3D object pose estimation, and image-based 3D shape retrieval, which can serve as baselines for future research using our database.

PUBLICATION

- Yu Xiang, Wonhui Kim, Wei Chen, Jingwei Ji, Christopher Choy, Hao Su, Roozbeh Mottaghi, Leonidas Guibas and Silvio Savarese. ObjectNet3D: A Large Scale Database for 3D Object Recognition. In European Conference on Computer Vision (ECCV), 2016. pdf, bibtex, technical report

DATASET

- ObjectNet3D images ~ 7.0GB
- ObjectNet3D 3D CAD models ~ 237MB
- ObjectNet3D annotations (training, validation) ~ 89MB
- ObjectNet3D image set indexes (training, validation, test) ~ 616KB
- ObjectNet3D toolbox (github)

- 100 object categories
- 90,127 images
- 201,888 objects
- 44,147 3D shapes
- 2D-3D alignments between 2D objects and 3D shapes
- Baseline experiments on different recognition tasks

Summary

- 3D object instance recognition vs. 3D object category recognition

- Learning 3D object representations
  - Multi-view images or videos
  - 3D CAD models
  - Deep learning

- Beyond 2D bounding boxes
  - Recognize detailed properties of objects: 3D pose, 3D shape, 3D location