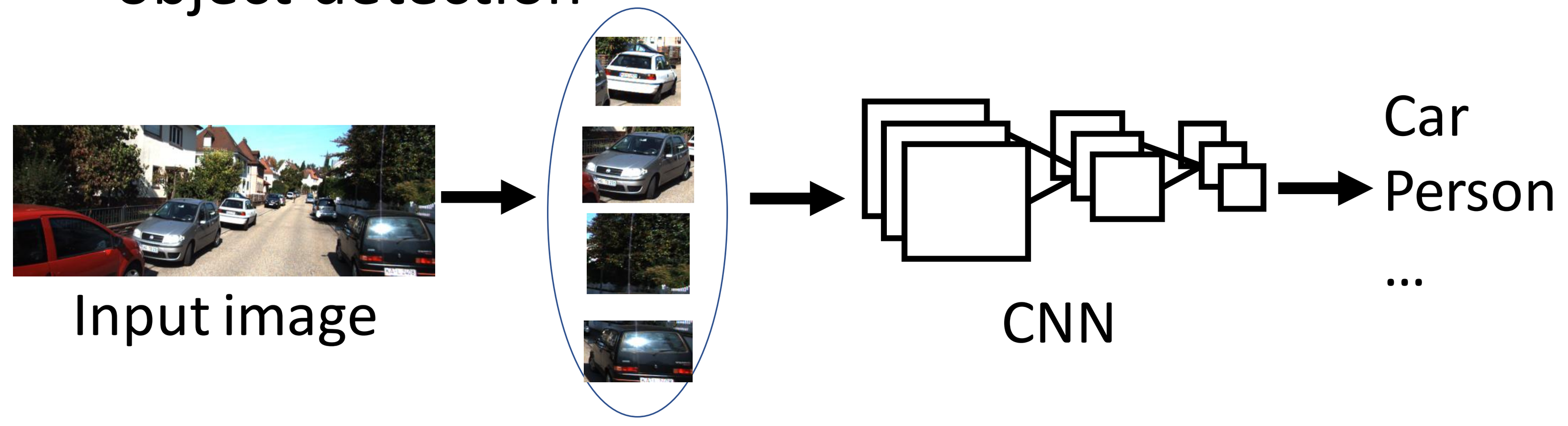
Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection

Yu Xiang[1], Wongun Choi[2], Yuanqing Lin[3] and Silvio Savarese[4]

[1]University of Washington, [2]NEC Laboratories America, Inc., [3]Baidu, Inc., [4]Stanford University

Introduction

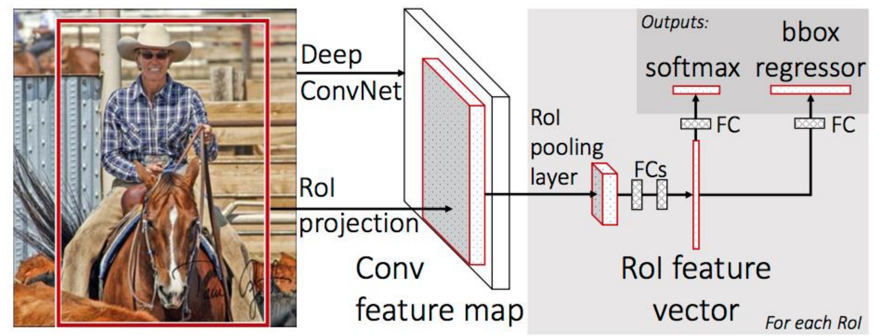
- Convolutional Neural Network (CNN) for object detection



- How to handle large scale change, occlusion and truncation?
- How to estimate detailed properties of objects (3D pose, 3D shape, 3D location)?
- We use subcategory information to help object proposal and detection in this work.

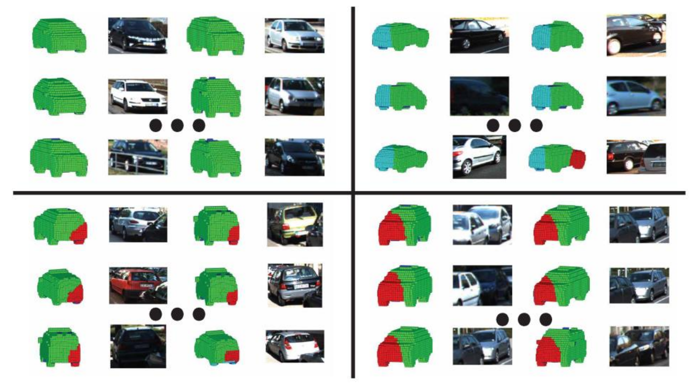
Related Work

- CNN-based object detection



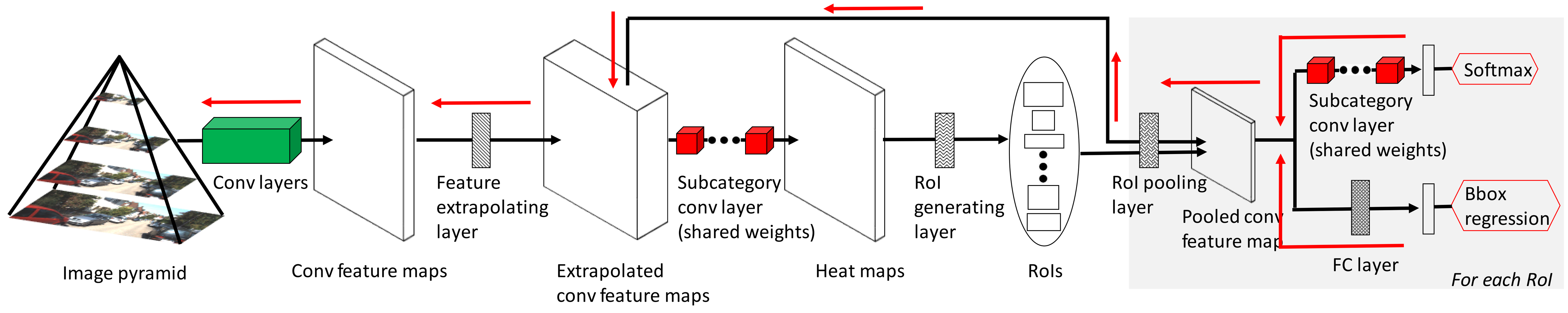
Fast RCNN. R. Girshick., ICCV'15.
Faster RCNN. Ren et al., NIPS'15.
YOLO. Redmon, et al., CVPR'16.
SSD. Liu et al., ECCV'16.

- Subcategory in object detection

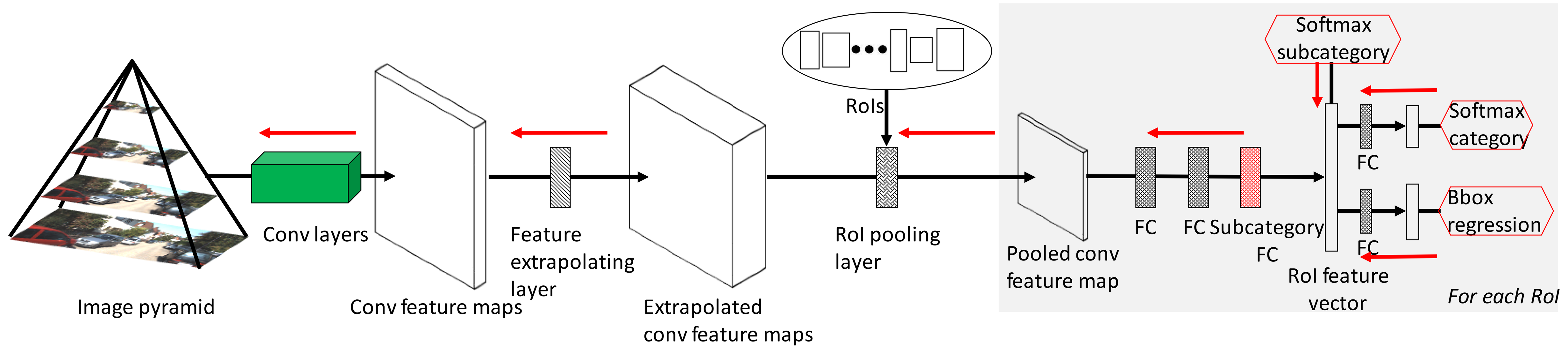


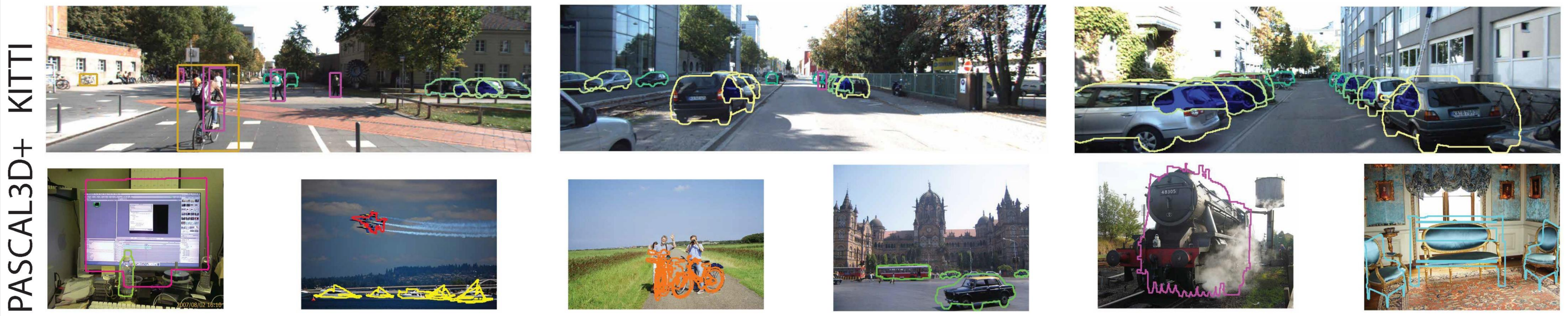
DPM. Felzenszwalb et al., TPAMI'10.
Gu & Ren, ECCV'10.
Ohn-Bar & Trivedi, ITS'15.
3DVP. Xiang et al., CVPR'15.

Subcategory-aware Region Proposal Network



Subcategory-aware Detection Network





[1] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. IJCV, 2013.
[2] C. L. Zitnick and P. Doll´ar. Edge boxes: Locating object proposals from edges. In ECCV, 2014.
[3] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2015
[4] P. Doll´ar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. TPAMI, 2014.
[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[6] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Multi-view and 3d deformable part models. TPAMI, 2015.
[7] B. Pepikj, M. Stark, P. Gehler, and B. Schiele. Occlusion patterns for object class detection. In CVPR, 2013..
[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.
[10] B. Li, T. Wu, and S.-C. Zhu. Integrating context and occlusion for car detection by hierarchical and-or model. In ECCV, 2014.
[11] Y. Xiang, W. Choi, Y. Lin, and S. Savarese. Data-driven 3d voxel patterns for object category recognition. In CVPR, 2015.
[12] X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun. 3d object proposals for accurate object class detection. In NIPS, 2015.
[13] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler and R. Urtasun: Monocular 3D Object Detection for Autonomous Driving. In CVPR, 2016.
[14] F. Yang, W. Choi, and Y. Lin. Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers. In CVPR, 2016.
[15] Z. Cai, Q. Fan, R. Feris, and N. Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In ECCV, 2016.
[16] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
[17] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In WACV, 2014.

Experiments

- Region proposal performance on KITTI [16]

Method	Easy	Moderate	Hard	Easy	Moderate	Hard	Easy	Moderate	Hard
	Car			Pedestrian			Cyclist		
SelectiveSearch [1]	58.17	42.12	37.62	68.95	57.65	52.57	57.05	49.59	49.44
EdgeBoxes [2]	81.40	61.84	55.68	86.15	71.88	65.39	56.11	46.52	45.72
RPN [3]	98.84	**97.37**	**95.31**	98.88	91.69	88.64	96.55	91.80	89.41
SubCNN	**99.27**	96.28	93.14	**99.44**	**93.46**	**91.02**	**99.67**	**93.03**	**91.64**

- Detection and Orientation Estimation on KITTI car

	Object Detection (AP)			Orientation Estimation (AOS)		
Method	Easy	Moderate	Hard	Easy	Moderate	Hard
ACF [4]	55.89	54.74	42.98	N/A	N/A	N/A
DPM [5]	68.02	56.48	44.18	67.27	55.77	43.59
DPM-VOC+VP [6]	74.59	64.71	48.76	72.28	61.84	46.54
OC-DPM [7]	74.94	65.95	53.86	73.50	64.42	52.40
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
AOG [10]	84.80	75.94	60.70	33.79	30.77	24.75
Faster R-CNN [3]	86.71	81.84	71.12	N/A	N/A	N/A
3DVP [11]	87.46	75.77	65.38	86.92	74.59	64.11
3DOP [12]	**93.04**	88.64	79.10	**91.44**	86.10	76.52
Mono3D [13]	92.33	88.66	78.96	91.01	86.62	76.84
SDP+RPN [14]	90.14	88.85	78.38	N/A	N/A	N/A
MS-CNN [15]	90.03	89.02	76.11	N/A	N/A	N/A
SubCNN-VGG16	90.74	88.55	77.95	90.49	87.88	77.10
SubCNN-GoogleNet	90.81	**89.04**	**79.27**	90.67	**88.62**	**78.68**

- Detection and Pose Estimation on PASCAL3D+ [17]

Method	DPM [5]	DPM-VOC+VP [6]	Ours w/o extra	Ours Full
Detection AP	29.6	28.3	58.8	**60.7**
Pose 4 views AVP	19.5	24.5	45.2	**47.5**
Pose 8 views AVP	18.7	22.2	28.6	**31.9**
Pose 16 views AVP	15.6	17.9	22.3	**24.5**
Pose 24 views AVP	12.1	14.4	17.9	**19.3**