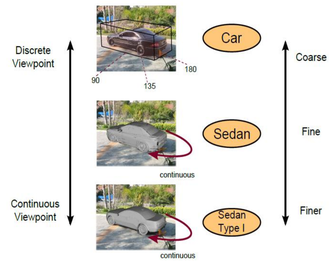
A Coarse-to-Fine Model for 3D Pose Estimation and Sub-Category Recognition

Roozbeh Mottaghi[1], Yu Xiang[2,3], Silvio Savarese[3]

[1] Allen Institute for AI [2] University of Michigan [3] Stanford University

Computational Vision & Geometry Lab

Summary



• A hierarchical representation for objects where each level represents a different level of granularity.
• Joint modeling of **object detection**, **3D pose estimation** and **sub-category recognition**.

Advantages of Hierarchy

• Tasks at different levels of granularity can benefit from each other.

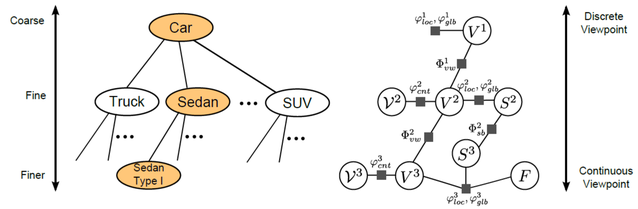
• Different types of features are required for different tasks.

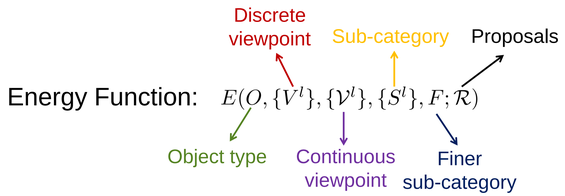
• We can better leverage the structure of the parameters so the performance does not drop as we increase complexity.

Model

• Our model is a hybrid random field, which consists of a mixture of continuous and discrete random variables.

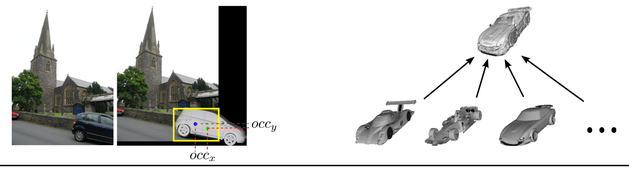
• We formulate the problem as structured prediction, where we jointly optimize the parameters of all layers.





• **Global shape:** HOG templates (φ_{glb})
• **Local Features:** CNN features (φ_{loc})
• **Consistency across layers** (Φ_{vw})
• **Relationship between discrete and continuous viewpoints** (φ_{cnt})

Continuous viewpoint consists of azimuth, elevation, distance, and occlusion.



Learning & Inference

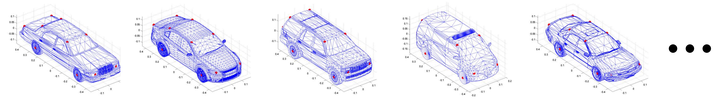
• **Learning:** We use structured SVM, where each layer has its own loss. The loss functions penalize mispredictions in viewpoint, sub-category, and finer-sub-category.

• **Inference:** Our inference shares similarities with Particle Convex Belief Propagation (PCBP).

We draw multiple samples from the continuous viewpoint variables. After sampling, the model can be considered as a fully discrete MRF. We perform exact inference by enumerating all possibilities.

Results

• **Dataset:** We augment 3 categories of PASCAL 3D+ [1] with annotations for sub-category and finer-sub-category.



Segmentation results:

2D Segmentation	Discrete	Continuous
Aeroplane	50.5	51.5
Boat	35.7	40.3
Car	60.4	64.4

Detection results:

	Bounding Box	All	Sub-category + Viewpoint	Sub-category	Viewpoint
RCNN [2]	51.4	X	X	X	X
DPM-VOC+VP [3]	29.5	X	X	X	21.8
V-DPM [4]	27.6	X	X	X	16.2
SV-DPM [4]	27.8	X	8.4	13.8	18.2
FSV-DPM [4]	25.8	0.35	7.9	12.7	16.1

	Bounding Box	All	Sub-category + Viewpoint	Sub-category	Viewpoint
1-layer	49.5	X	X	X	28.9
2-layer	51.0	X	16.0	27.5	29.5
3-layer	51.6	3.2	17.6	30.6	29.5
Flat model	51.6	2.6	14.8	27.8	26.3

References:
[1] Y. Xiang, R. Mottaghi, and S. Savarese. Beyond pascal: A benchmark for 3d object detection in the wild. In WACV, 2014.
[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.
[3] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Teaching 3d geometry to deformable part models. In CVPR, 2012.
[4] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. PAMI, 2010.