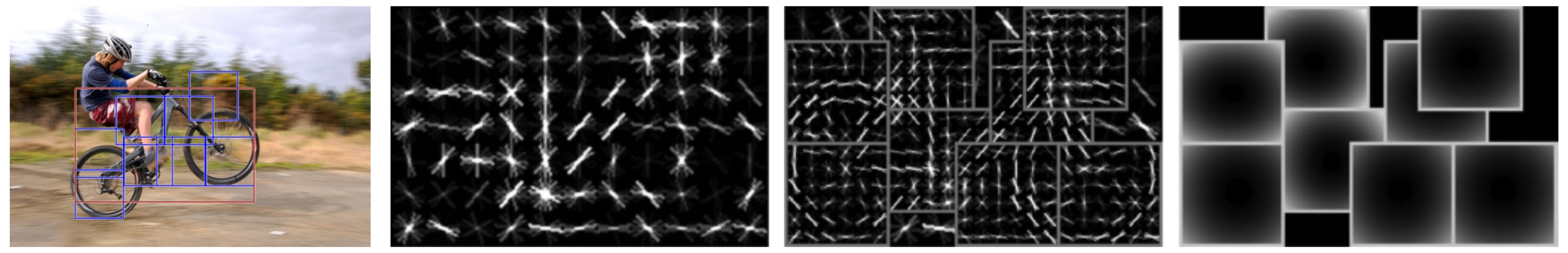
Building Part-based Object Detectors via 3D Geometry

Abhinav Shrivastava and Abhinav Gupta

THE ROBOTICS INSTITUTE

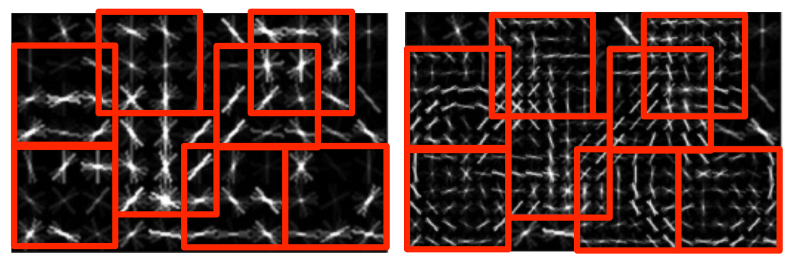
Carnegie Mellon University, Pittsburgh Pennsylvania 1900

Discriminative Part-based Models

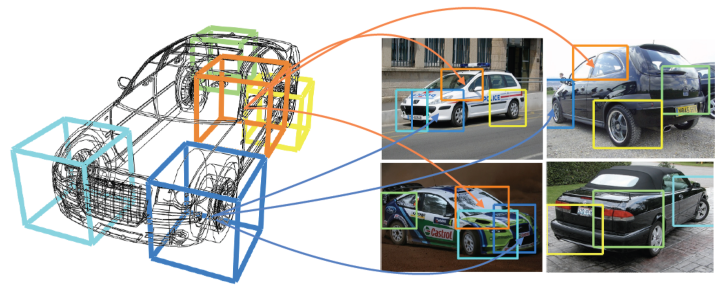


Unsupervised Parts

Heuristic initialization, e.g., gradient magnitudes.



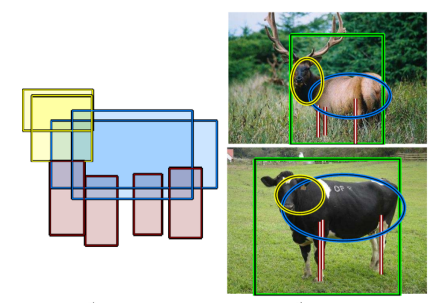
[Felzenszwalb et al. 2009]



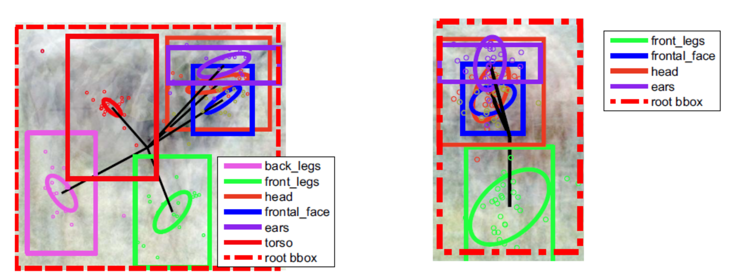
[Pepik et al. 2012]

Supervised Parts

Key-point/part annotation, e.g., anatomical.



[Endres et al. 2012]

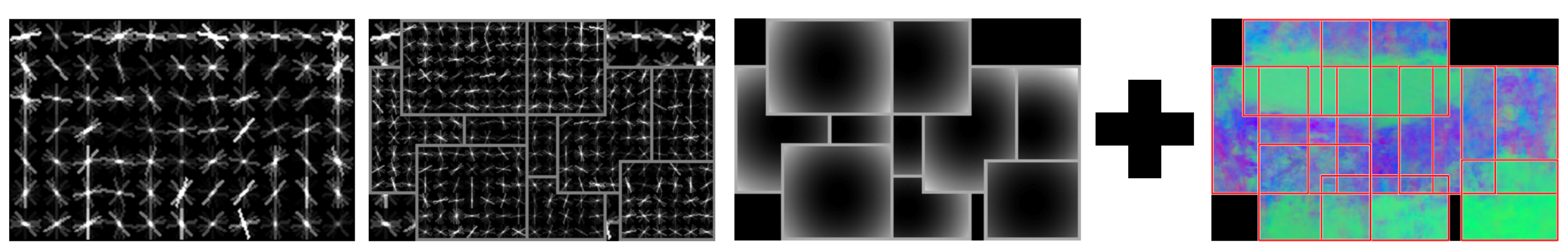


[Azizpour et al. 2012]

What is the right way to select parts?

Geometry-driven Parts (gDPM)

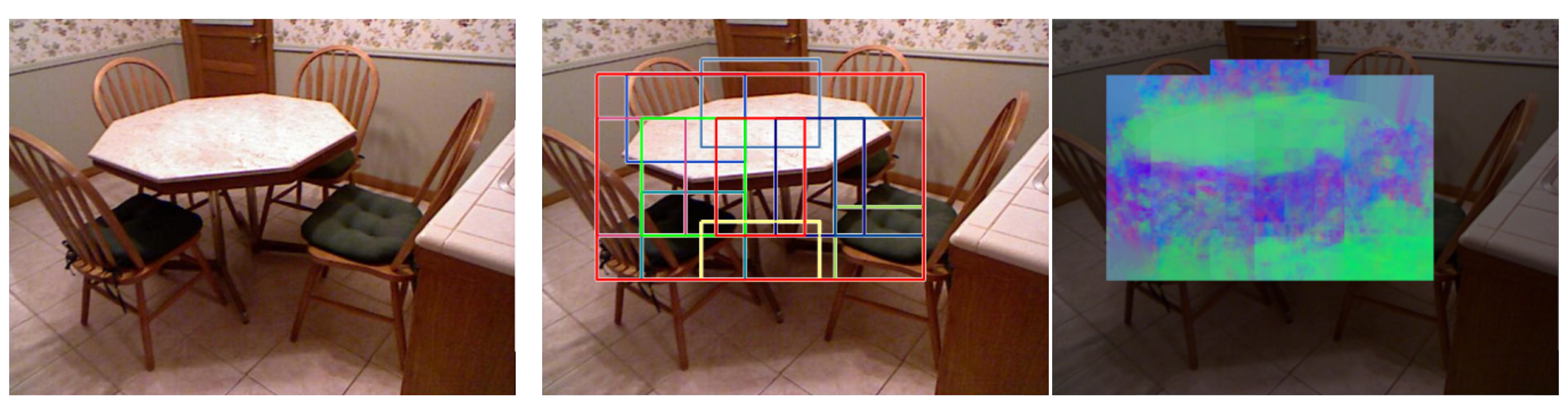
Parts based on consistent underlying 3D geometry



Advantages:

a) Better optimization and learning
   - Leverage large-scale RGBD data for constraints

b) 3D Scene Understanding



RGB Input | gDPM Detection | Predicted Geometry

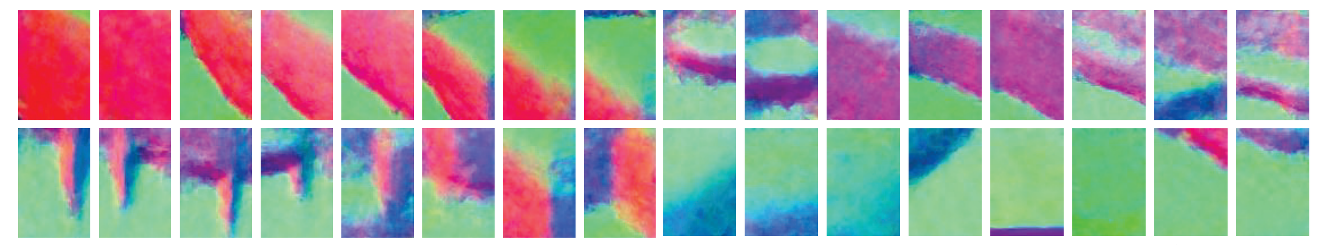
Our Approach

Effectiveness of **DPMs** + Richness of **Geometric Representation**

Geometry-driven Dictionary of 3D Elements

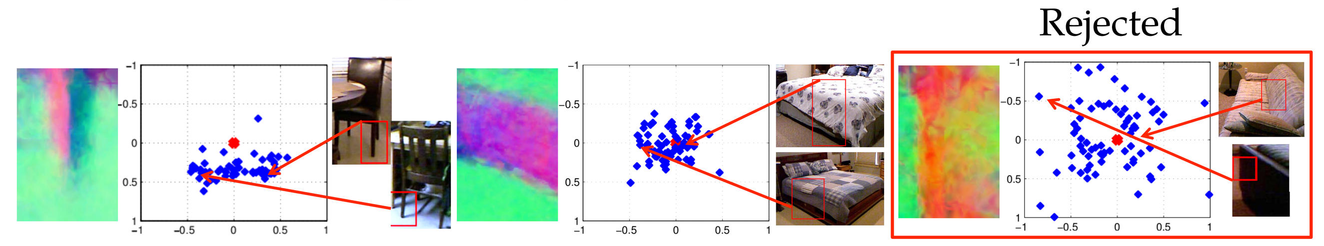
Desired properties:

- Representative: frequent among objects
- Spatially consistent w.r.t the object



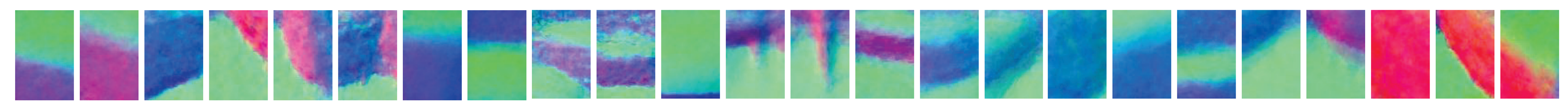
K-Means on Surface Normal and Relative Depth (10000s clusters) (3 Aspect Ratios)

Rejected



Good Clusters — Increasing $\min(\sigma_x^2, \sigma_y^2)$ → Bad Clusters

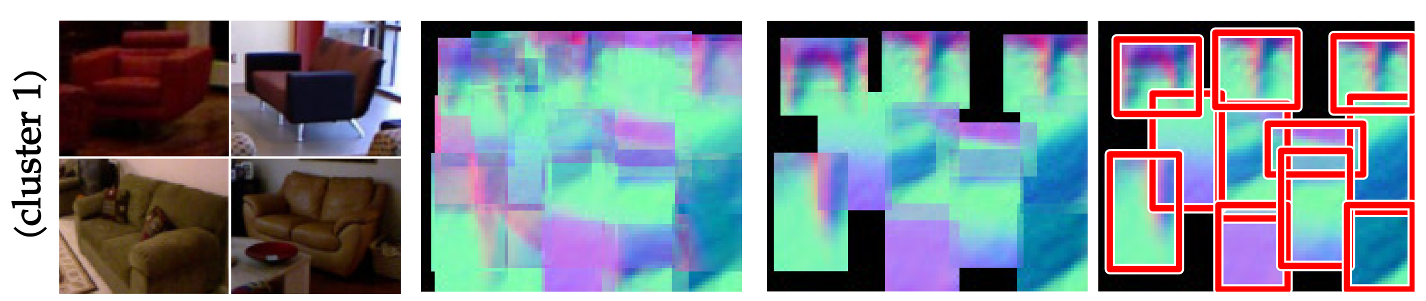
Check Geometric Consistency and Merge.



Dictionary ($\mathcal{D}$) of Geometrically-consistent 3D elements (100s clusters, 3 Aspect Ratio)

From 3D Parts to Object Hypothesis

Greedily select N parts based on frequency and geometric consistency to make an object hypothesis: $\mathbf{p} = [p^1, \ldots, p^N]$, where $p^j : (e_i, l^j)$. (N ~ [6, 12])



Sofa (cluster 1)

Sample Objects | Frequent Parts | Selected Consistent Parts

Learning gDPM

Enforcing **geometric constraints** using large-scale RGBD data.

Input - $x = \{I, I^G, l\},\ y \in \{-1, 1\}$

I - RGB Image, I^G – Surface Normal Map, l - bounding-box location.

$$\text{Minimize: } L_D(\beta) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N} \max(0, 1 - y_i f_\beta(x_i))$$

β - Model Parameters, z - Root & Part Locations, c - Cluster Membership. — Latent

Standard Negative Hard-mining

$$f_\beta(x) = \begin{cases} S_{\text{Appearance}}(I, z, \beta_c) & \text{if } y = -1, \\ \underbrace{S_{\text{Appearance}}(I, z, \beta_c)}_{\text{Standard Appearance Score of root and part filters.}} + \lambda \underbrace{\sum_{i=1}^{n_c} S_G(e^i, \omega(I^G, l_i))}_{\text{Geometry constraints on part placements.}}, & \text{if } y = +1. \end{cases}$$

Geometry constrained Positive Latent Update

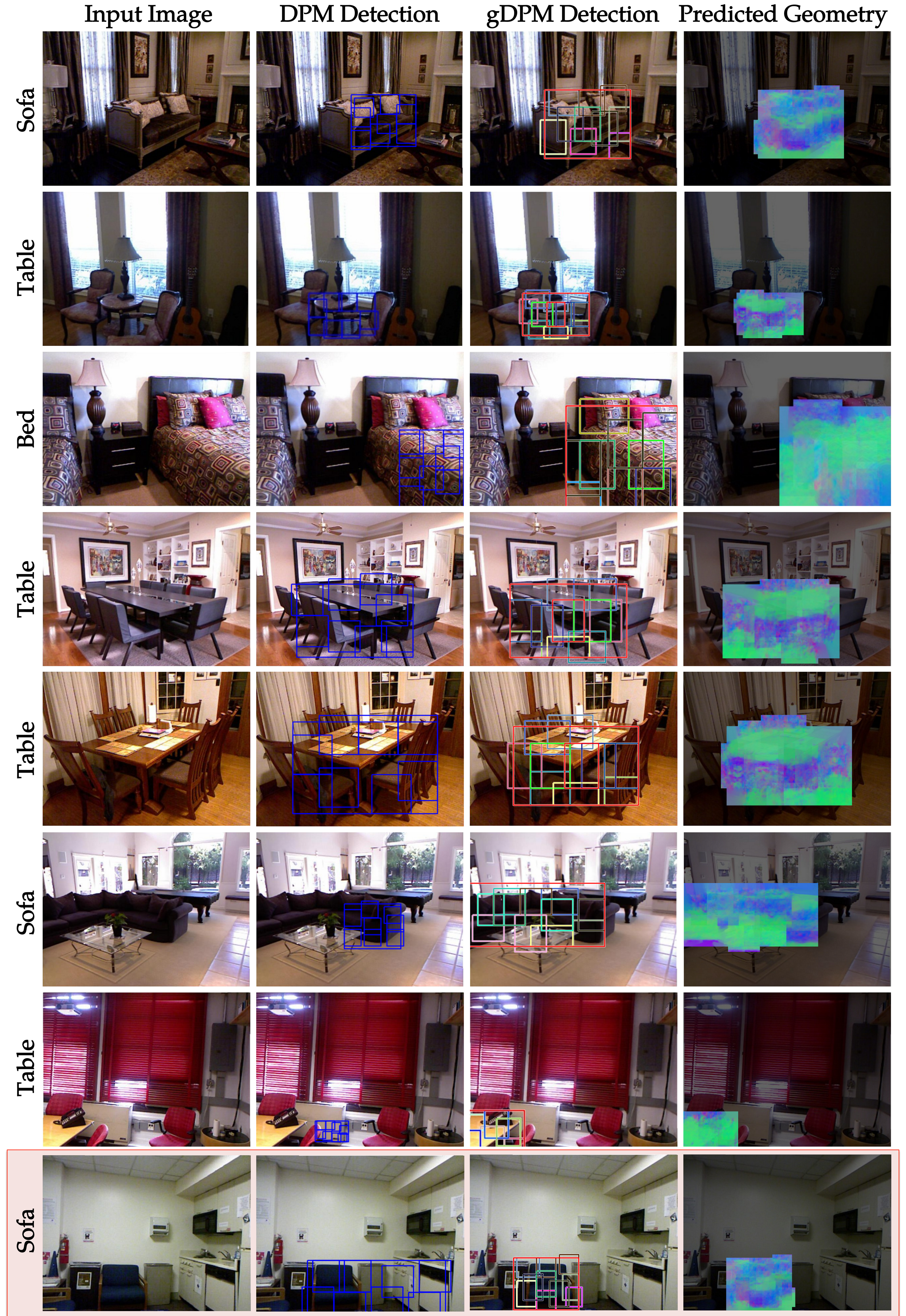
$\omega(\cdot)$ – Raw surface normal at

$S_G(\cdot)$ – Similarity function between surface normal maps

Experimental Results

Qualitative Results

Test Set: NYU v2 RGB Images



Quantitative Results

Test Set: NYU v2 RGB Images

Table 1. AP performance on the task of object detection.

	Bed	Chair	M.+TV	Sofa	Table
DPM (No Parts)	20.94	10.69	6.38	5.51	2.73
DPM	22.39	**14.44**	8.10	7.16	3.53
DPM (Our Parts, No Latent)	26.59	5.71	2.35	6.82	3.41
DPM (Our Parts)	29.15	11.43	4.17	8.30	1.76
gDPM	**33.39**	13.72	**9.28**	**11.04**	**4.05**