Single Image 3D Without A Single 3D Image

David Fouhey[1], Wajahat Hussain[2], Abhinav Gupta[1], Martial Hebert[1]

CMU Robotics Institute[1], Universidad de Zaragoza[2]

Carnegie Mellon University — THE ROBOTICS INSTITUTE

Universidad Zaragoza — 1542

Task: 3D Understanding

Image → Surface Normals
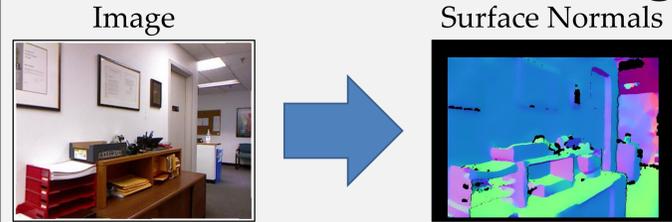
Do we *really* need supervision?

Quiz: Which is Unsupervised?

A B
C D

Factorization
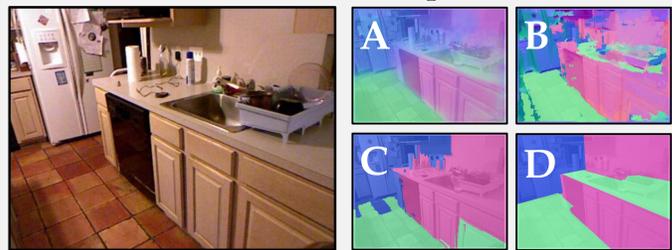
Image = 3D Structure x Style
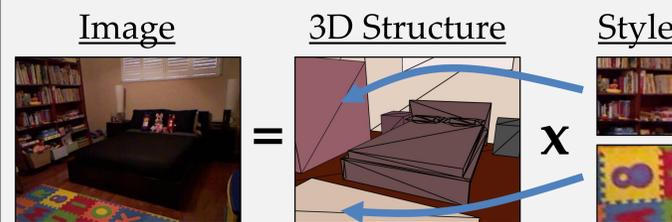
Style → Sharing Across Viewpoints

Image + Style → 3D Structure

ELDA Detector

Rectified Images / Style (Correspond to Directions)
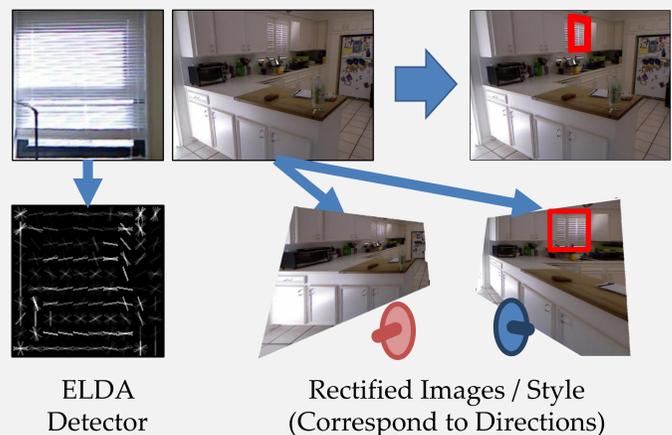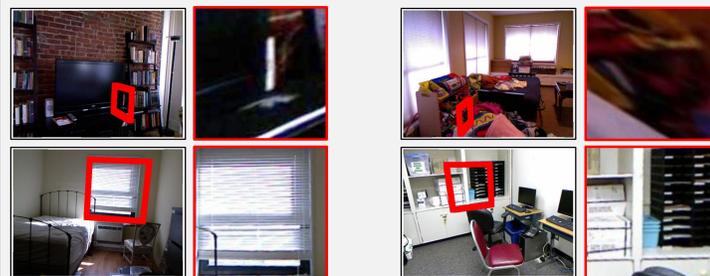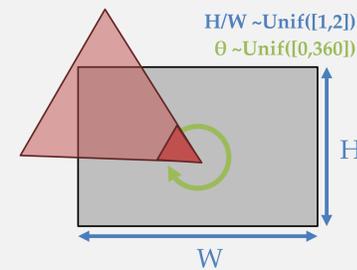
Learning

Idea: Hypothesize style elements; compare implied 3D structure with prior

Hypotheses

Prior

H/W ~Unif([1,2])
θ ~Unif([0,360])

H
W

Comparison with Prior

Prior

Surface Orientation

X Location

Inference

Image → Rectified Images → Detections+Prior → Per-Pixel → +Segments

Learned Elements
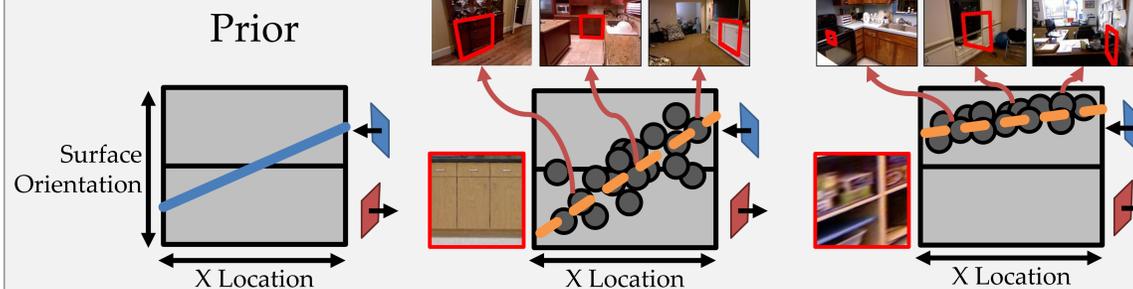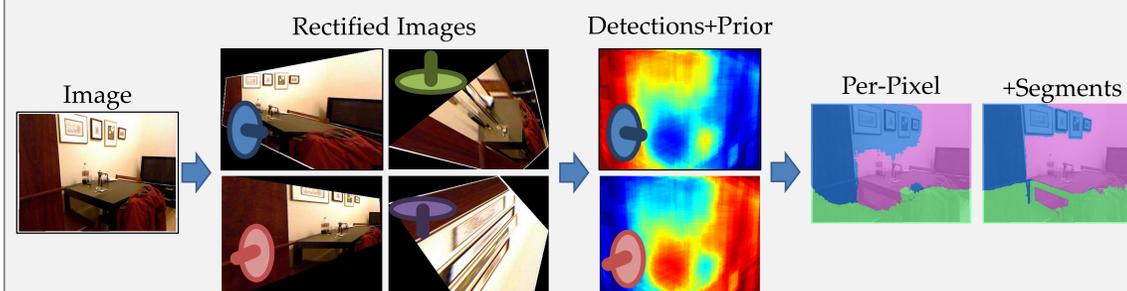
Element	Detections	Element	Detections	Element	Detections
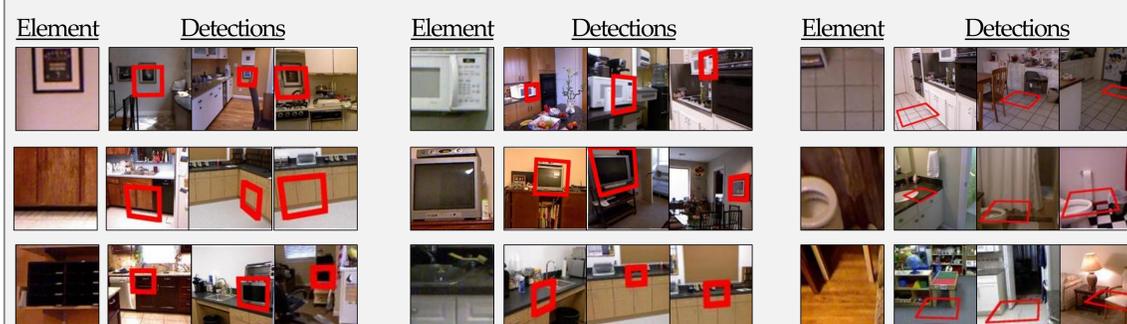
Qualitative Results
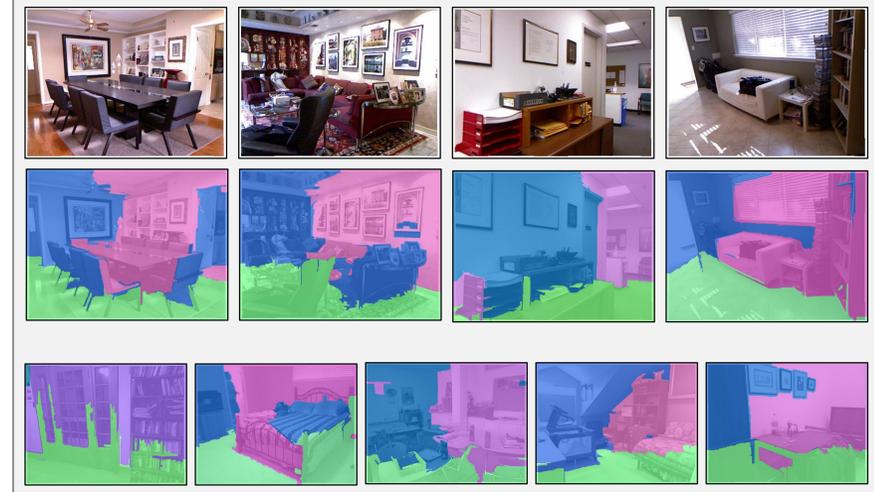
Internet Image Results

Elements   Predictions
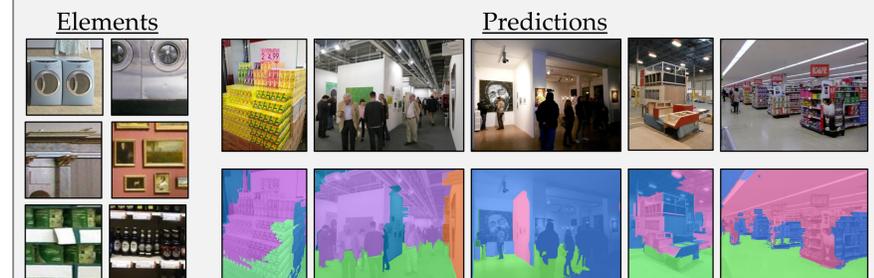
Quantitative Results

NYUv2: Standard train/test, Ladicky '14 GT. Per-pixel Criterion: angular error

	Full Image Results			Vertical Pixels	
	Median	%<11.25	%<30	Median	%<30
Proposed	21.7	36.7	55.4	19.7	59.7
3DP (MW) ICCV13	19.2	39.2	57.8	19.9	58.8
Fouhey ECCV14	17.9	40.5	58.9	--	--
Ladicky ECCV14	23.1	27.7	58.7	--	--
Using Video Dataset For Training					
	Median	%<11.25	%<30		
*Wang CVPR15	14.8	42.0	68.2		
*Eigen (VGG) ICCV15	13.2	44.4	75.9		

Places-205: Discretized ground-truth. Per-pixel criterion: accuracy

	Proposed	3DP (MW)	Wang et al.
Accuracy	62.9	59.2	63.7
# Categories Won	4	1	5