Marr Revisited: 2D-3D Model Alignment via Surface Normal Prediction

Aayush Bansal [1] Bryan Russell [2] Abhinav Gupta [1]

[1] Carnegie Mellon University [2] Adobe Research

Task: 2D ⟶ 2.5D ⟶ 3D

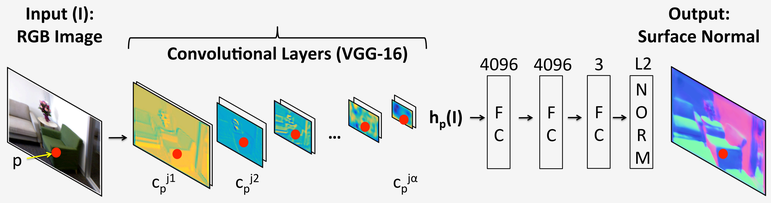
- Surface normal (2.5D) from a single 2D image.
- Pose & style of objects from RGB + 2.5D cues.

Input Image | Predicted Surface Normal | CAD Model Library (3D ShapeNets, Wu et al. [CVPR'15]) | Aligned CAD Model

2D ⟶ 2.5D

Surface Normal Estimation

Non-linear optimization of Hypercolumn features for fine details.



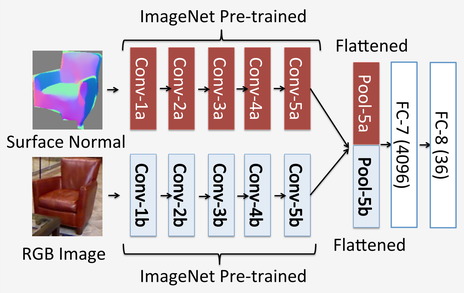
Hypercolumn (Hariharan et al. [CVPR'15]): $\mathbf{h_p(I)} = [c_p^{j1}, c_p^{j2} \dots, c_p^{j\alpha}]$

- During training, **N** (= 1500) pixels are sampled per image for optimization.
- At test time, entire image is fed-forward. Output from last layer are the predicted surface normal.

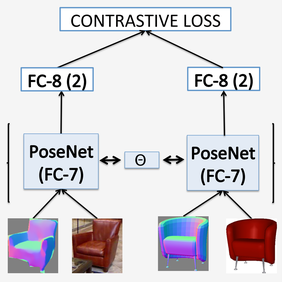
2.5D ⟶ 3D

Pose & Style Estimation

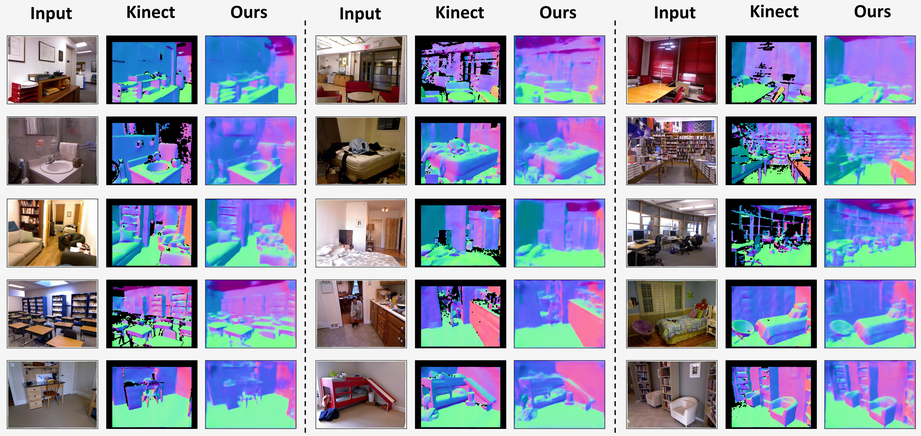
PoseNet: A 36-way pose classification.



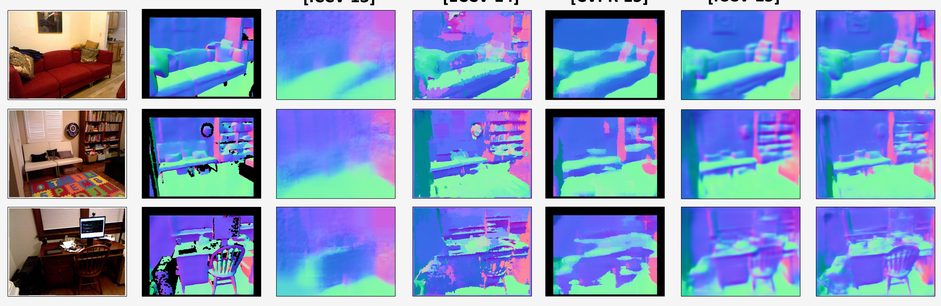
StyleNet: Are they similar?



Surface Normal Estimation



Qualitative Comparison



Quantitative Results

Analysis done on NYUD-v2 test set (Silberman et al. [ECCV'12]).

Comparison with previous state-of-the-art approaches:

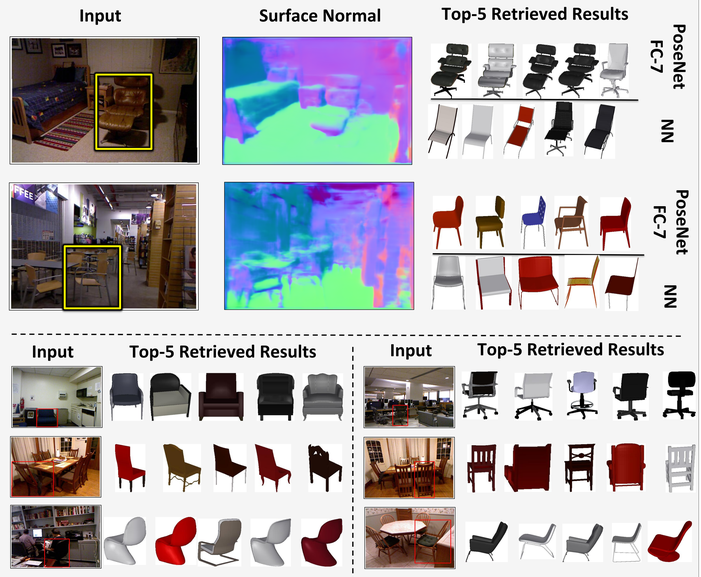
Approach	Mean	Median	11.25	22.5	30
Fouhey et al.	35.3	31.2	16.4	36.6	48.2
E-F (AlexNet)	23.7	15.5	39.2	62.0	71.1
E-F (VGG-16)	20.9	13.2	44.4	67.2	75.9
Ours	**19.8**	**12.0**	**47.9**	**70.0**	**77.8**
with Manhattan World					
Fouhey et al.	36.3	19.2	39.2	52.9	57.8
Wang et al.	26.9	14.8	42.0	61.2	68.2
Ours	**23.9**	**11.9**	**48.4**	**66.0**	**72.7**

Ablative Analysis using diff. conv-layers from VGG-16:

Conv-layers	Mean	Median	11.25	22.5	30
$\{1_1, 1_2\}$	44.4	42.7	04.1	16.5	28.2
$\{1_1, 1_2, 3_3\}$	30.2	24.7	23.1	46.2	58.4
$\{1_1, 1_2, 5_3\}$	22.6	15.3	39.1	63.4	73.1
$\{1_1, 1_2, 3_3, 5_3\}$	21.3	13.9	42.3	67.0	76.0
$\{1_2, 3_3, 5_3\}$	21.3	14.0	42.0	66.7	75.8
$\{1_2, 2_2, 3_3, 4_3, 5_3\}$	20.9	13.6	43.1	67.9	77.0
$\{1_2, 2_2, 3_3, 4_3, 5_3, 7\}$	**19.8**	**12.0**	**47.9**	**70.0**	**77.8**

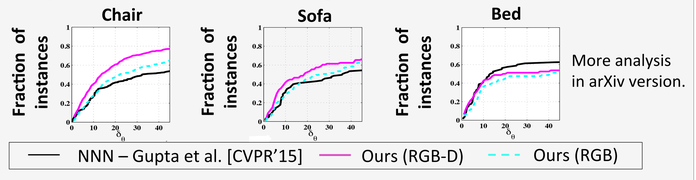
VGG-16 has 13 conv-layers represented as $\{1_1, 1_2, 2_1, 2_2, 3_1, 3_2, 3_3, 4_1, 4_2, 4_3, 5_1, 5_2, 5_3\}$. FC-7 of VGG-16 is referred as conv-7 here.

Pose Estimation



Quantitative Results

Analysis done on NYUD-v2 val set using annotations from Guo & Hoiem [ICCV'13].



Style Estimation