Task: 2D $\rightarrow$ 2.5D $\rightarrow$ 3D

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

2D $\rightarrow$ 2.5D

Surface Normal Estimation

Non-linear optimization of Hypercolumn features for fine details.

Input (I): RGB Image

Convolutional Layers (VGG-16)

Output: Surface Normal

Hypercolumns (Harikaran et al. [CVPR'15]): $h_{j0}, h_{j1}, h_{j2}, \ldots, h_{jN}$

- During training, $N \approx 1500$ 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 $\rightarrow$ 3D

Pose & Style Estimation

PoseNet: A 36-way pose classification.

StyleNet: Are they similar?

Surface Normal Estimation

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

**Approach** | **Mean** | **Median** | **11.25** | **22.5** | **30**
--- | --- | --- | --- | --- | ---
Fouhey et al. (E-F [v2]) | 35.3 | 31.2 | 16.4 | 36.6 | 48.2
Fouhey et al. (E-F [v1]) | 23.7 | 15.5 | 39.2 | 62.0 | 71.1
Fouhey et al. (E-F [VGG-16]) | 20.9 | 13.2 | 44.4 | 67.2 | 75.9
Ours (RGB) | 19.8 | 12.0 | 47.9 | 70.0 | 77.8

with Manhattan World

Fouhey et al. (RGB) | 36.3 | 19.2 | 39.2 | 52.9 | 75.8
Wang et al. (RGB) | 26.9 | 14.8 | 42.0 | 61.2 | 68.2
Ours (RGB) | 23.9 | 11.9 | 48.4 | 66.0 | 72.7

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. (E-F [v2]) | 44.4 | 42.7 | 40.4 | 16.5 | 28.2
Fouhey et al. (E-F [v1]) | 32.0 | 24.7 | 23.1 | 46.2 | 58.4
Fouhey et al. (E-F [VGG-16]) | 22.6 | 15.3 | 39.1 | 63.4 | 73.1
Wang et al. (RGB) | 21.3 | 13.9 | 42.3 | 67.0 | 76.0
Ours (RGB) | 21.3 | 14.1 | 42.0 | 66.7 | 75.8

with Manhattan World

Fouhey et al. (RGB) | 20.9 | 13.6 | 43.1 | 67.9 | 77.0
Wang et al. (RGB) | 19.8 | 12.0 | 47.9 | 70.0 | 77.8
Ours (RGB) | 19.8 | 12.0 | 47.9 | 70.0 | 77.8

- VGG-16 has 33 conv-layers represented as $\{1, 2, 3, 4, 5, \ldots, 33\}$. FC-7 of VGG-16 is referred as conv-7 here.