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

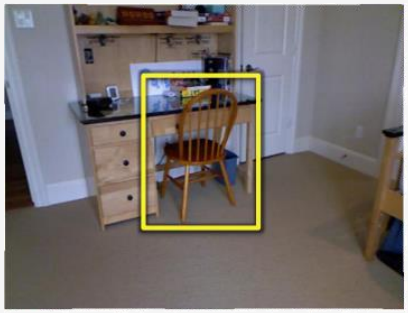
Aayush Bansal, Carnegie Mellon University

Bryan Russell, Adobe Research

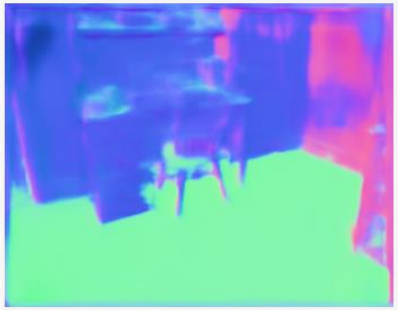
Abhinav Gupta, Carnegie Mellon University

CVPR 2016

Introduction



Input Image



Predicted Surface Normal

CAD Model Library

Aligned Models

Introduction

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

Introduction

- Why Marr's framework?
  - Most data for training 3D representations is **CAD data** (c.f. ShapeNet or ModelNet)
  - Big **gap** between CAD model renders and real 2D images
  - Marr's 2.5D representation helps to **bridge this gap**
  - Marr revisited
- Contributions
  - **Skip-network** architecture for surface normal estimation
  - **CNN** architecture for CAD retrieval combining **image** and predicted surface **normal**

Related work

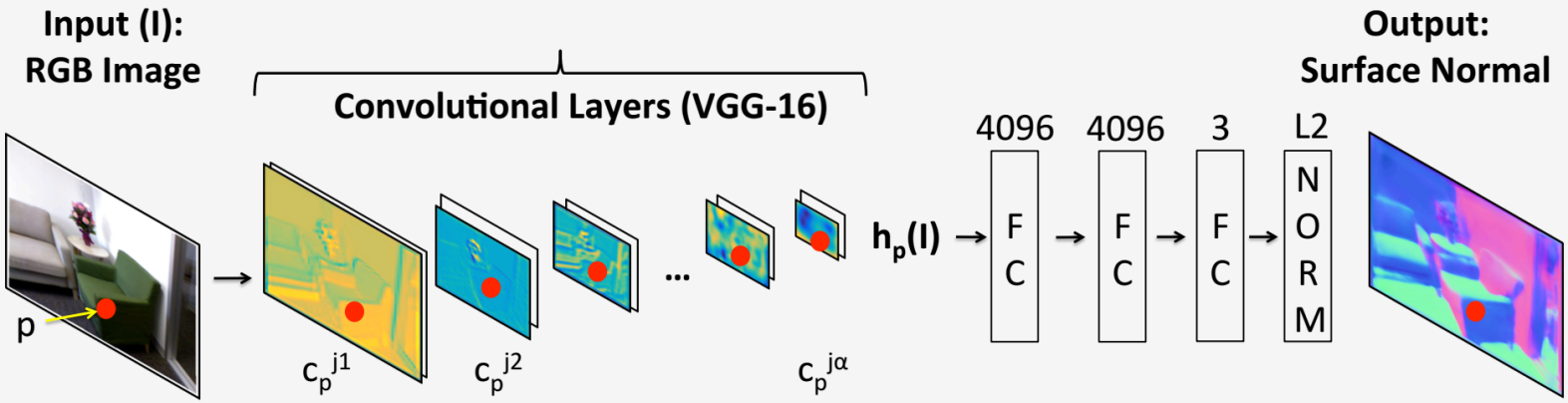
- 3D scene understanding
  - Recovering the 2.5D
    - Discriminative 3D primitives [Fouhey et al 2013]
    - Convex and concave edges [Fouhey et al 2014]
    - Discriminative learning with hand-crafted features [Ladicky et al 2014]
  - Recovering the 3D volumetric objects
    - Train CNNs to predict object class for CAD model alignment [Gupta et al 2015]
    - Train Siamese network modeling style similarity to retrieve product images having similar style as an object in an input photo [Bell et al 2015]

Skip-network architecture

2D ⟶ 2.5D

Surface Normal Estimation

Non-linear optimization of Hypercolumn features for fine details.



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

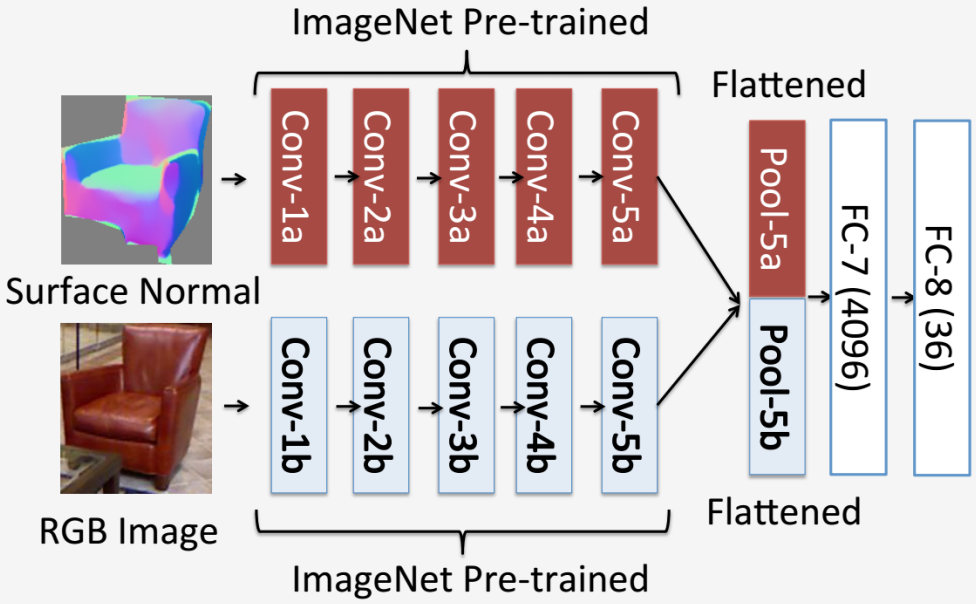
- During training, **N** (= 1500) pixels are sampled per image for optimization.
- At test time, entire image is fed-forward. Output from last layer are the predicted surface normal.

Networks for predicting pose and style

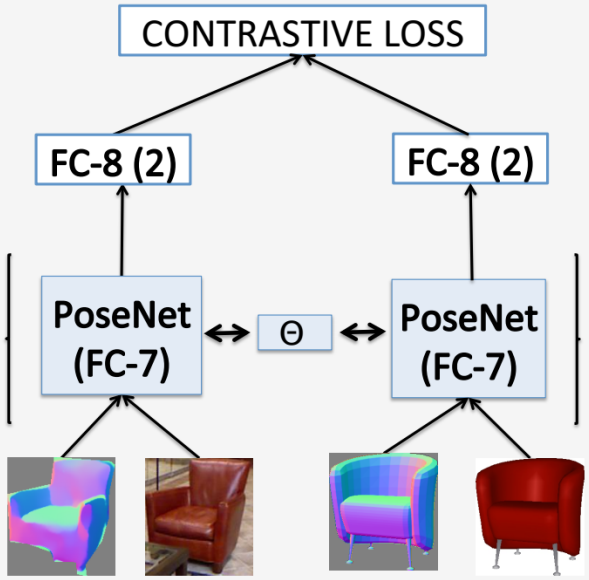
2.5D ⟶ 3D

Pose & Style Estimation

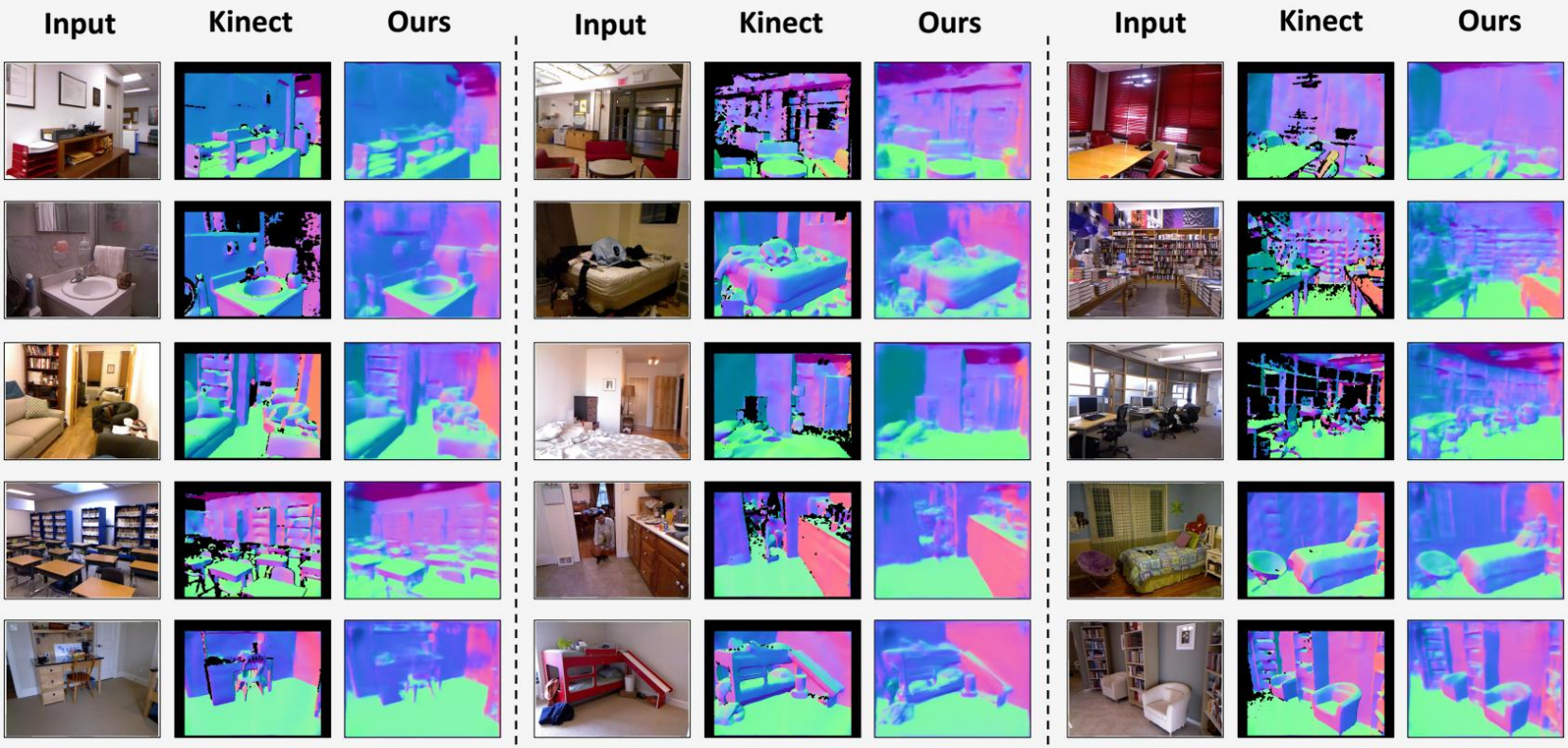
PoseNet: A 36-way pose classification.



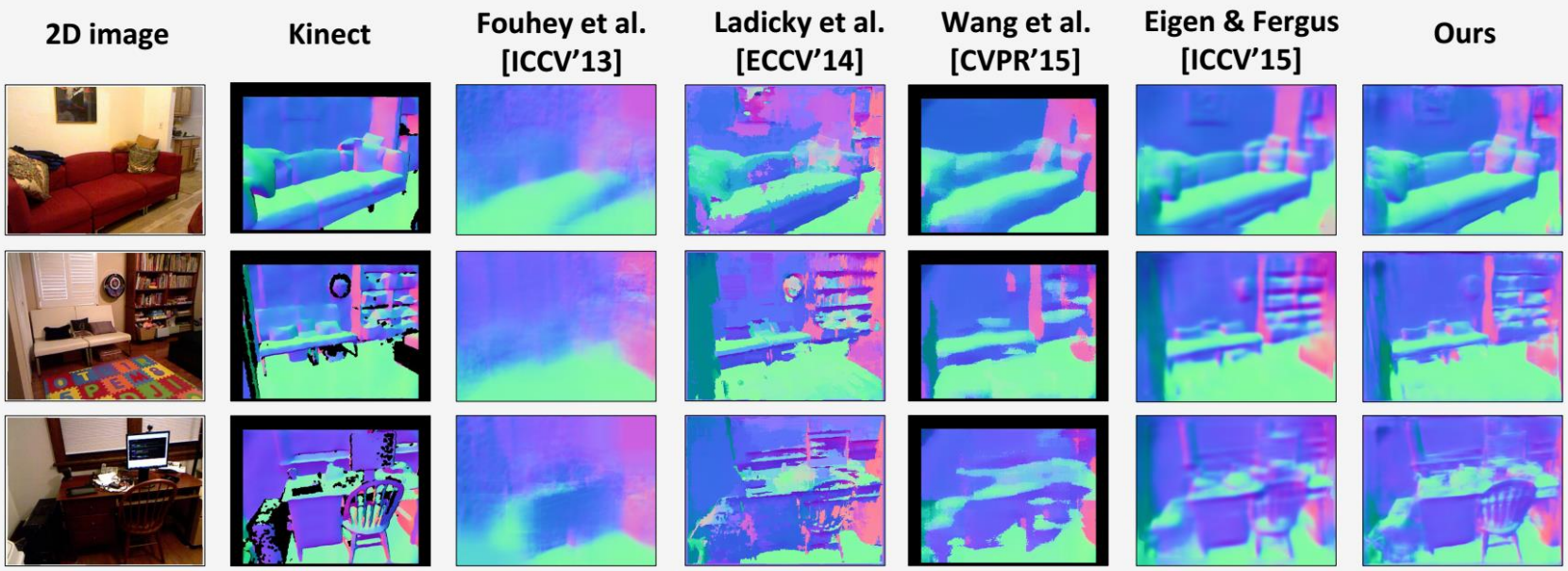
StyleNet: Are they similar?



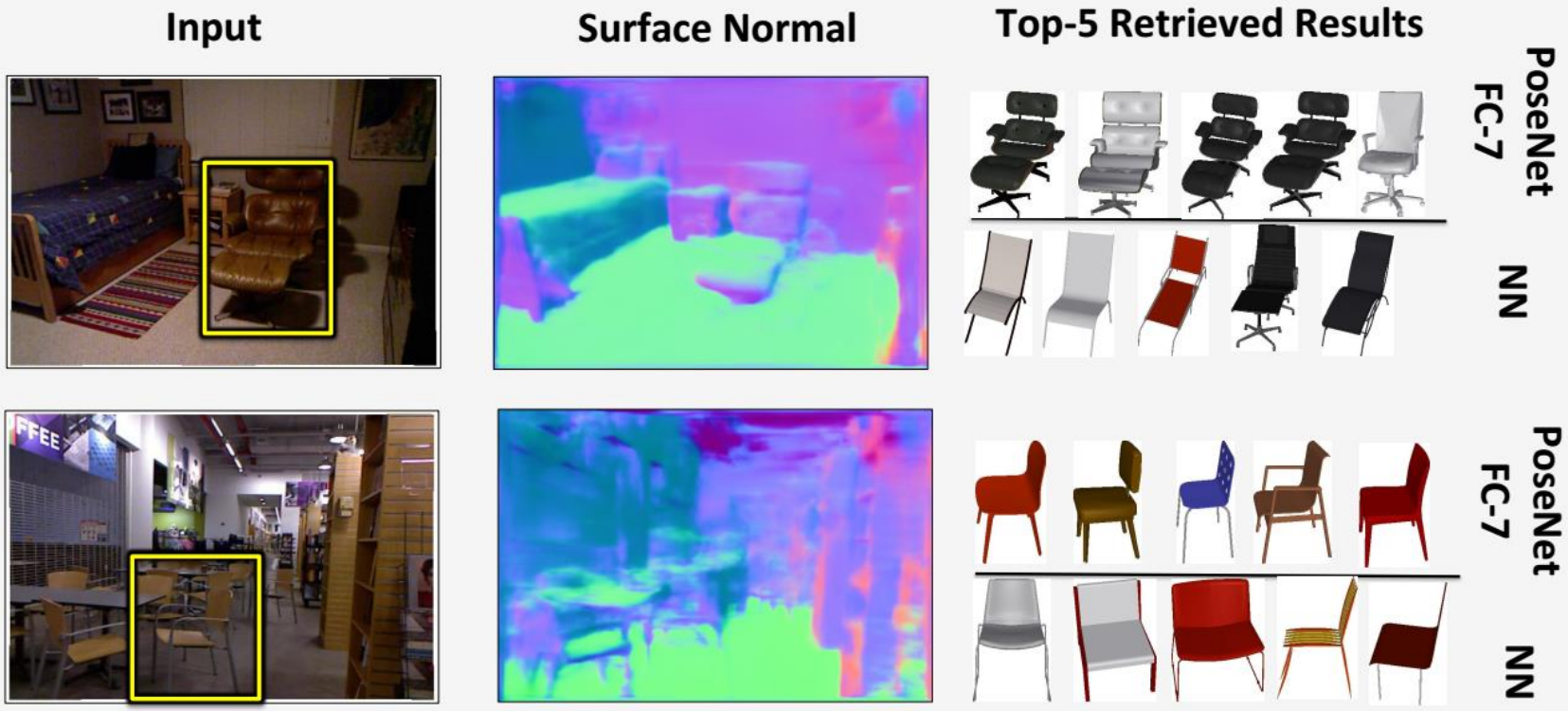
Surface Normal Estimation



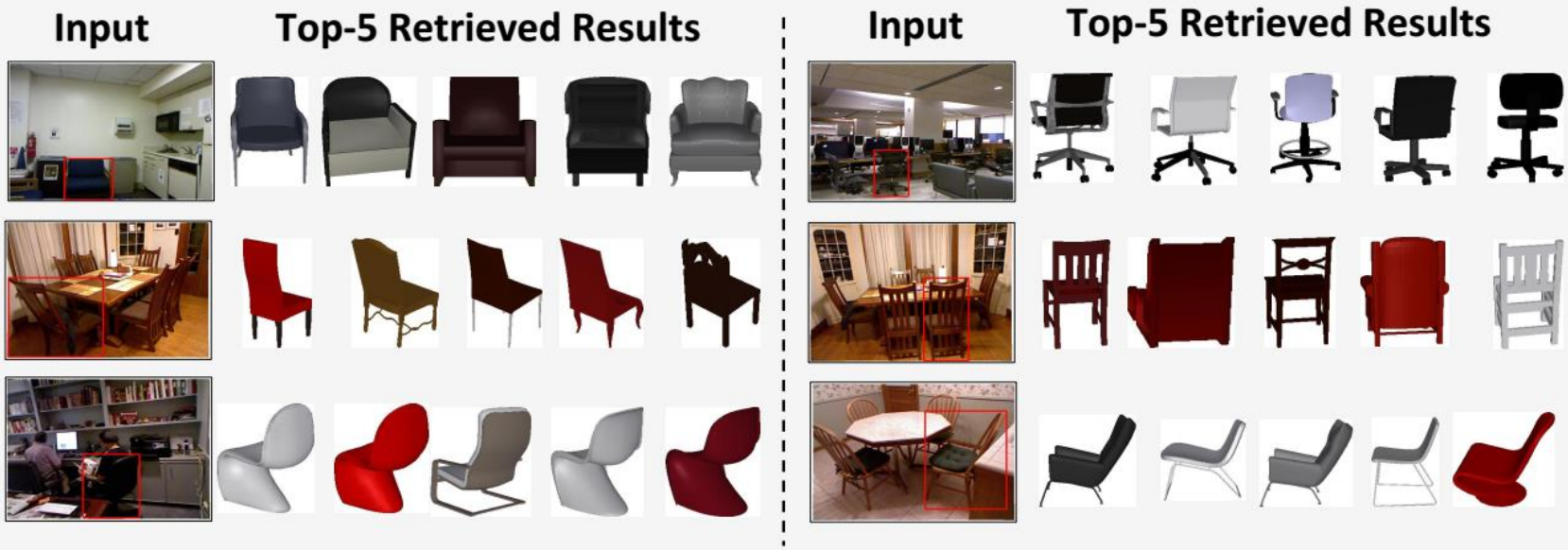
Qualitative Comparison



Pose Estimation



Pose Estimation



Style Estimation