PlaneRCNN: 3D Plane Detection and Reconstruction from a Single Image [2]

Jia Zheng

SIST, ShanghaiTech

January 3, 2019

Outline

Outline

Task

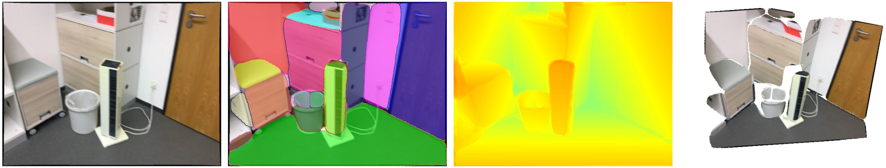


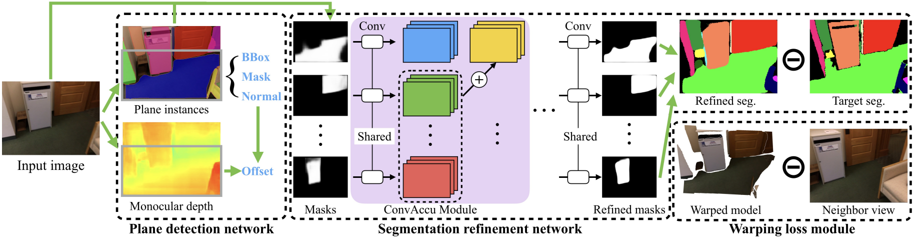
Figure: Piece-wise Planar 3D Reconstruction.

Limitations

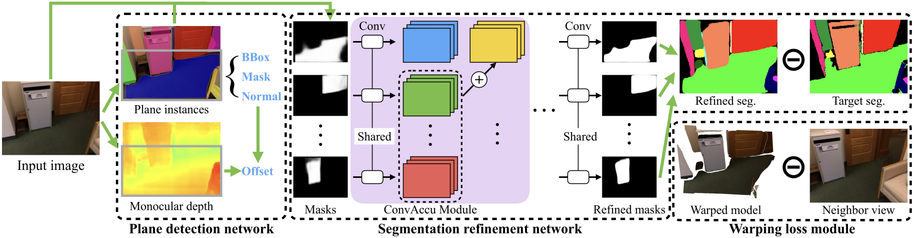
- Missing small surfaces.
- Requiring the maximum number of planes in a single image a priori.
- Poor generalization across domains (e.g., trained for indoors images and tested outdoors).

Outline

Pipeline

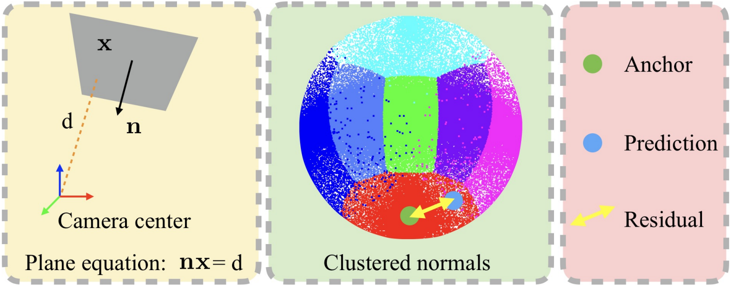

Input image
Plane instances
BBox
Mask
Normal
Monocular depth
Offset
Plane detection network
Masks
Conv
Shared
ConvAccu Module
Conv
Shared
Refined masks
Segmentation refinement network
Refined seg.
Target seg.
Warped model
Neighbor view
Warping loss module

Plane Detection Network



Predict a normal per planar instance, estimate a depth map for an entire image, use a simple algebraic formula to calculate the plane offset.

Plane normal estimation



Plane offset estimation

Given a plane normal n, it is straightforward to estimate the plane offset d

$$d = \frac{\sum_i m_i(\mathbf{n}^\intercal(z_i K^{-1}\mathbf{x_i}))}{\sum_i m_i} \tag{1}$$

where K is the 3×3 camera intrinsic matrix, $\mathbf{x_i}$ is the i-th pixel coordinate in a homogeneous representation, z_i is its predicted depth value, and m_i is an indicator variable, which becomes 1 if the pixel belongs to the plane.

Segmentation Refinement Network

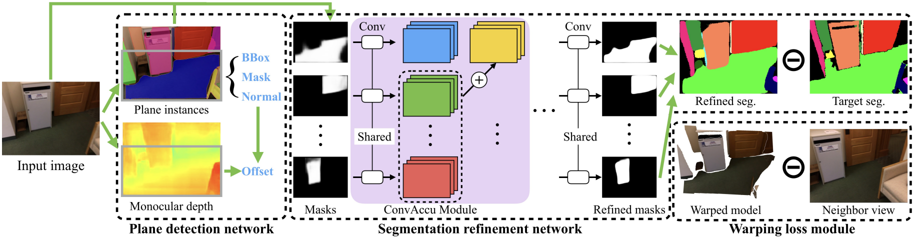
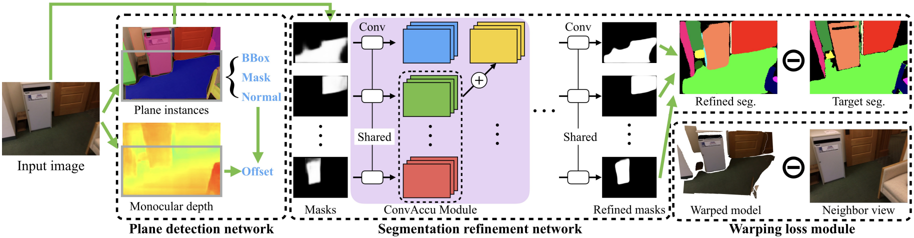


Figure: More details in Figure. 10.

The refinement network takes the original image, plane masks, the reconstructed depth map, and a 3D coordinate map for specific plane as input.

Warping Loss Module



The warping loss module enforces the consistency of reconstructed 3D planes with a nearby view (20 frames ahead) during training.

Outline

Modifications

1. Keep more small planar regions (threshold: 1% to 0.16% of image size, i.e., 500 pixels), do not drop small planes when the total number is larger than 10.
2. Skip merging process and keep all instance segmentation masks.
3. Remove some imprecision images.

This process leads to more fine-grained planar regions, yielding 14.7 plane instances per image on the average, which is more than double the PlaneNet [1] dataset containing 6.0 plane instances per image.

Outline

Qualitative Evaluations

The qualitative results on the ScanNet dataset are shown in Figure 5, Figure 11 and Figure 12. The comparisons against PlaneNet [1] and PlaneRecover [3] are shown in Figure 6, Figure 13 and Figure 14.

Plane Reconstruction Accuracy

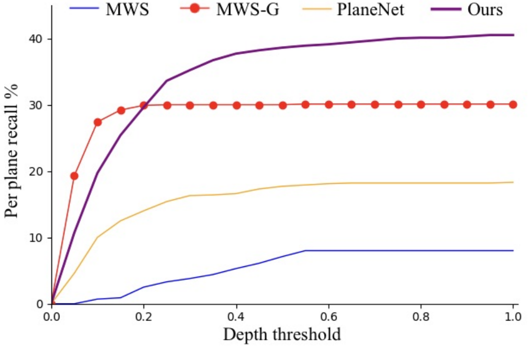


Figure: Per-Plane recall versus depth threshold.

Occlusion Reasoning

Add one more mask prediction module to PlaneRCNN to infer the complete mask for each plane instance.

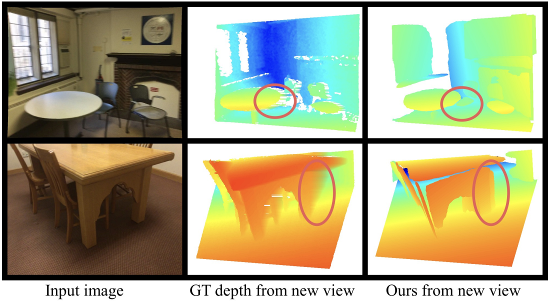


Figure: New view synthesis results with the layered depth map models.

Summary

This paper proposes a detection-based neural network for piecewise planar reconstruction from a single image.

Reference

- Chen Liu et al. "PlaneNet: Piece-wise Planar Reconstruction from a Single RGB Image". In: *CVPR*. 2018.
- Chen Liu et al. "PlaneRCNN: 3D Plane Detection and Reconstruction from a Single Image". In: *arXiv preprint arXiv:1812.04072* (2018).
- Fengting Yang and Zihan Zhou. "Recovering 3D Planes from a Single Image via Convolutional Neural Networks". In: *ECCV*. 2018.

Thanks

Thanks for Attention!