Supplementary Materials for: Recovering 3D Planes from a Single Image via Convolutional Neural Networks

Fengting Yang and Zihan Zhou

The Pennsylvania State University
{fuy34, zzhou}@ist.psu.edu

In this document, we provide additional experiment results about our network design on SYNTHIA dataset.

Number of planes m. A key parameter in our network is the plane number m, which controls the maximum number of planes that can be detected. In this experiment, we vary m in our network and report the plane segmentation performance in Table 1(left). As one can see, our method achieves the best performance with $m = 5$. Further increasing m does not seem to help the results. This agrees with the fact that, based on the ground truth annotations, the top-5 planes cover 94.3% of the planar regions in all images. Further, the performance of our method is relatively stable w.r.t. m. For example, even with $m = 3$, our method outperforms the existing methods (see Table 1 of the paper).

Fig. 1 shows example plane segmentation results obtained by our method with different values of m. As one can see, when m is small, our method sometimes fails to detect all prominent planes in the scene. For example, in Fig. 1, first row, our method does not detect the building facade on the left when $m =$ 3 or 4. Meanwhile, increasing m enables our method to recover more planes, but also increases the risk of incorrectly dividing one plane into two or more.

Effect of semantic labels. Next, we study the effect of including semantic labels in our network. To this end, we train our network by replacing the regularization term Eq. (5) with Eq. (4) in our paper (we fix $m = 5$ in this experiment). As one can see in Table 1(right), the performance of our method degrades in the absence of semantic labels. We also note that, even without the labels, our method performs better than existing methods. This is significant as it shows that it is indeed feasible to "teach" neural networks to recognize mid/high-level scene structure with a purely geometric structure-induced loss.

Table 1. Plane segmentation results. **Left:** Comparison of difference choices of plane number m. **Right:** Effect of semantic labels.

Method	RI	VOI	SC
Ours ($m = 3$)	0.875	1.341	0.720
Ours ($m = 4$)	0.911	1.187	0.777
Ours ($m = 5$)	0.925	**1.129**	**0.797**
Ours ($m = 6$)	0.927	1.190	0.783
Ours ($m = 7$)	**0.928**	1.172	0.786

Method	RI	VOI	SC
Ours (no semantics)	0.874	1.464	0.708
Ours	**0.925**	**1.129**	**0.797**

Fig. 2 compares the results obtained by our method with and without the semantic labels. As one can see, even without the semantic labels, our method is able to distinguish most non-planar objects from planar surfaces. Nevertheless, the segmentation boundaries are not as accurate as those generated with the semantic labels. For example, in Fig. 2, second row, parts of the car wheels are included in the ground plane.

In last row of Fig. 2, we show an interesting case where the unpaved ground is classified as non-planar by our method when semantic labels are incorporated. In contrast, without the semantic labels, our method combines the unpaved ground and the paved road into a single plane – a reasonable result if only scene geometry is considered.

Fig. 1. Plane segmentation results with different plane numbers. **From left to right:** Input image, results of our method with $m = 3$, 4, 5, 6, and 7, respectively.

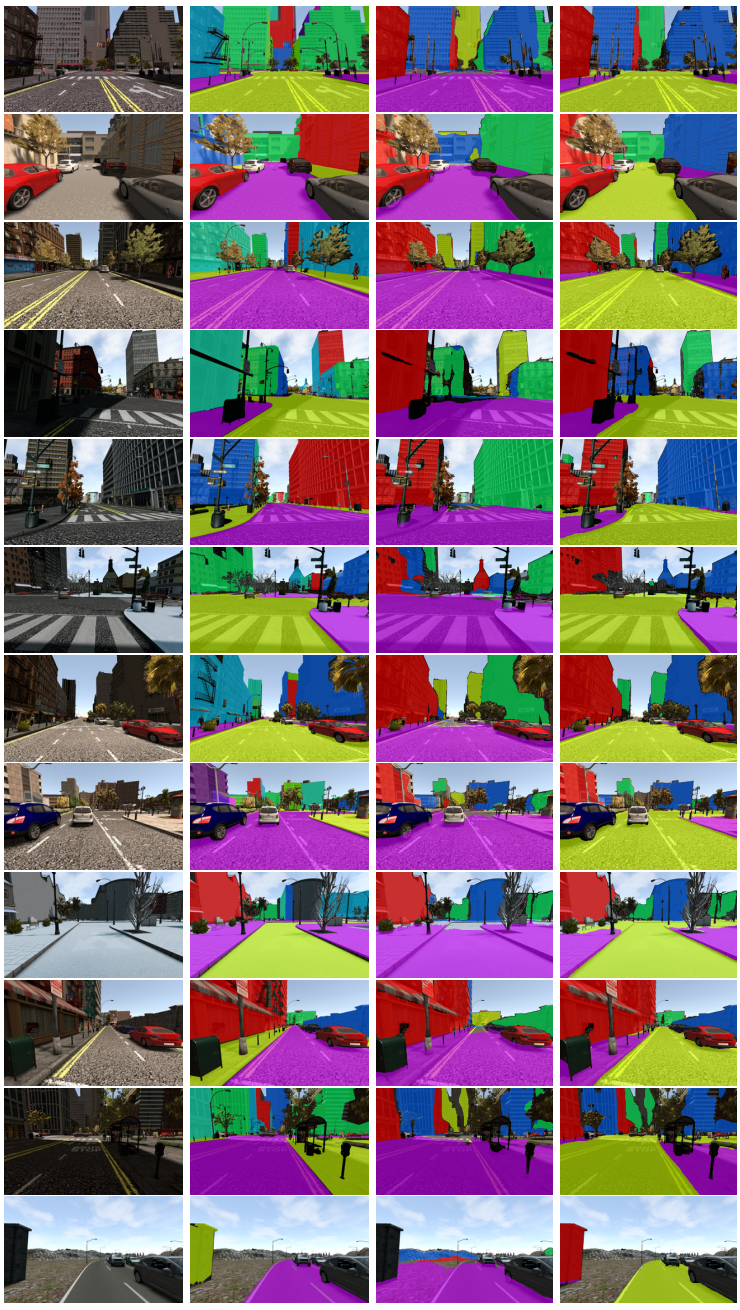

Fig. 2. Effect of semantic labels. **First column:** Input image. **Second column:** Ground truth annotation. **Third column:** Our method (without semantic labels). **Fourth column:** Our method.