Unsupervised Learning of Monocular Depth Estimation and Visual Odometry with Deep Feature Reconstruction (Supplementary Material)

Huangying Zhan[1,2], Ravi Garg[1,2], Chamara Saroj Weerasekera[1,2], Kejie Li[1,2], Harsh Agarwal[3], Ian Reid[1,2]
[1]The University of Adelaide
[2]Australian Centre for Robotic Vision
[3]Indian Institute of Technology (BHU)
{huangying.zhan, ravi.garg, chamara.weerasekera, kejie.li, ian.reid}@adelaide.edu.au
harsh.agarwal.eee14@iitbhu.ac.in

1. Network Architecture

1.1. Depth Network

We have introduced two network architectures in this paper. Both networks use same encoder network (ResNet50-1by2) but with different decoder networks. The ResNet50-1by2 architecture can be found in [**?**]. Only the fully convolutional network of ResNet50-1by2 is used (fully connected layers are not included). The main network architecture (Figure 1 (a)) in this work follows [2] closely, which uses a bilinear upsampler with skip connection as decoder. Moreover, we design a learnable deconv network (Figure 1 (b)) for the purpose of using self-embedded depth feature.

1.2. Odometry Network

The visual odometry network is shown in Figure 2.

2. More Visual Odometry Results

In the paper, we use KITTI Odometry Split for training (Seq. 00-08) and evaluation (Seq.09-10) of visual odometry. As mentioned in the paper, Eigen Split and Odometry Split have overlapping scenes such that finetuning/testing models trained in any split to another split is not allowable. However, there are some sequences in Odometry Split that are not appeared in Eigen Split training set such that the sequences can be used to evaluate the odometry performance of models trained in Eigen Split. The sequences are 00,04,05,and 07. The odometry results of Eigen Split model is shown in Table 1.

Method	Seq. 00		Seq. 04		Seq. 05		Seq. 07	
	$t_{err}(\%)$	$r_{err}(^\circ/100m)$	$t_{err}(\%)$	$r_{err}(^\circ/100m)$	$t_{err}(\%)$	$r_{err}(^\circ/100m)$	$t_{err}(\%)$	$r_{err}(^\circ/100m)$
Ours	25.08	7.91	12.58	3.91	8.01	2.54	13.91	5.63

Table 1. Additional visual odometry results evaluated on Sequence 00, 04, 05, 07. The model is trained in KITTI Eigen Split only.

3. More Depth Estimation Results

3.1. More Qualitative Results

In here we show more qualitative results for depth estimation with different methods we proposed in the paper. The results are showed in Figure 3.

3.2. Generalization Ability

We illustrate some examples of our model generalizaing to other datasets, including Cityscapes dataset and Make3D dataset. The examples are shown in Figure 4 and Figure 5.

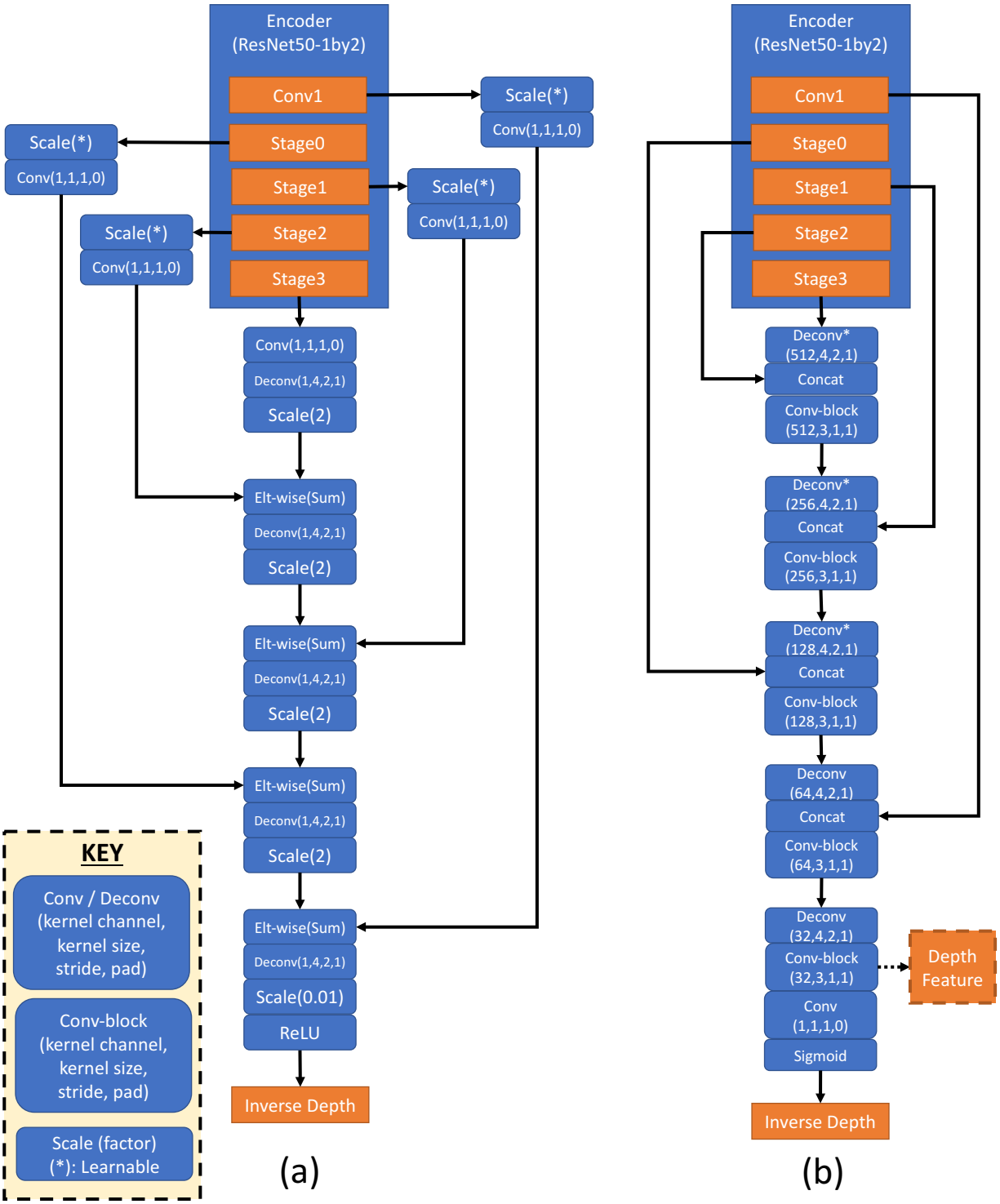

Figure 1. Depth network architectures. (a): ResNet50-1by2 as encoder; Bilinear upsampler as decoder. (b): ResNet50-1by2 as encoder; Learnable upsampler ("Deconv*" is learnable) as decoder. Conv-block includes a convolutional layer, a batch normalization layer, a scaling layer and a ReLU layer.

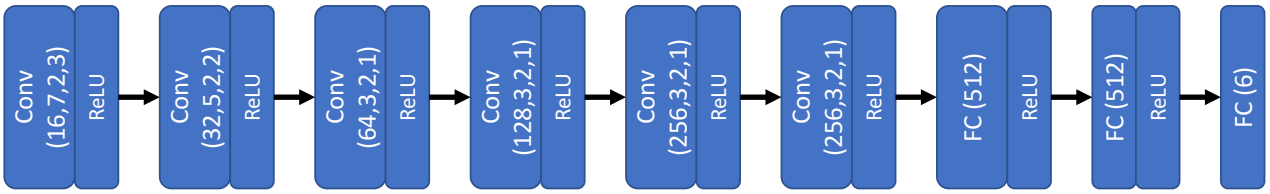

Figure 2. Visual odometry network architecture.

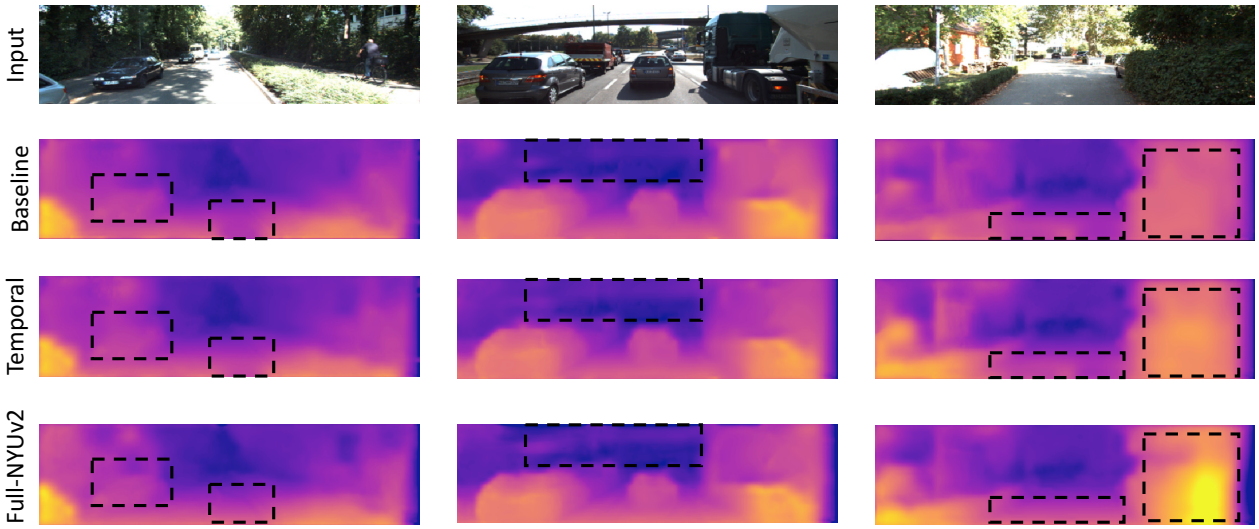

Figure 3. Qualitative results on the KITTI Eigen Split with our methods. The Full-NYUv2 model performs better in ambiguous regions (e.g. road) and shows sharper depths than others.

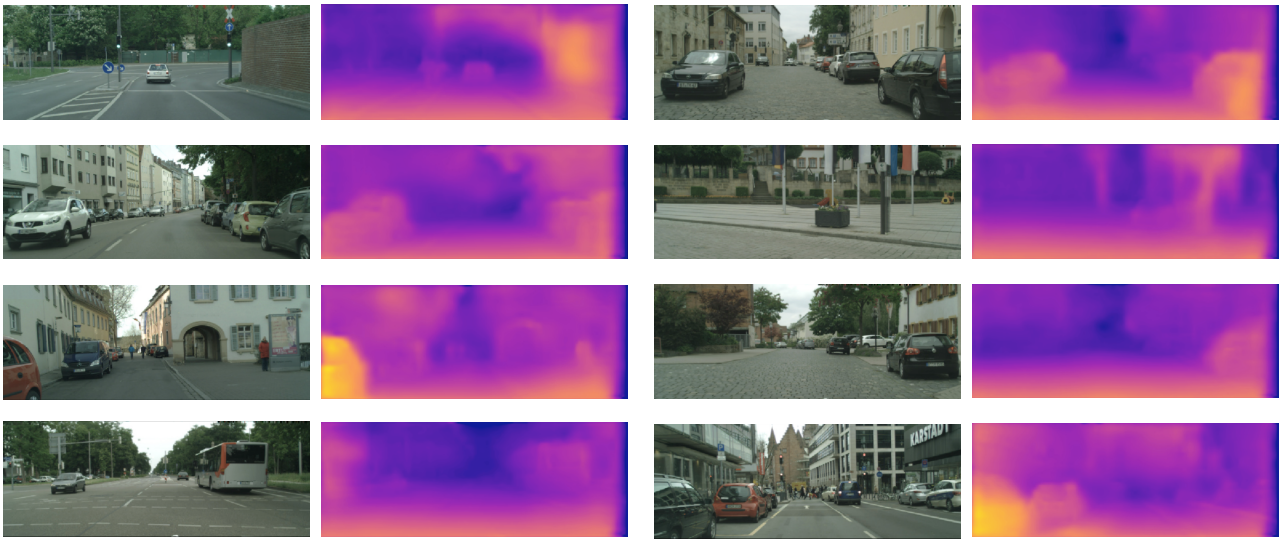

Figure 4. Qualitative results on Cityscapes[1]

Figure 5. Qualitative results on Make3D[3]

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] R. Garg, V. K. B G, G. Carneiro, and I. Reid. Unsupervised cnn for single view depth estimation: Geometry to the rescue. In *European Conference on Computer Vision*, pages 740–756. Springer, 2016.

[3] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2009.