Depth Map Super-Resolution by Deep Multi-Scale Guidance: Supplementary material

Tak-Wai Hui[1], Chen Change Loy[1,2], Xiaoou Tang[1,2]

[1]Department of Information Engineering, The Chinese University of Hong Kong
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
{twhui,ccloy,xtang}@ie.cuhk.edu.hk

S1 Introduction

In this supplementary material, discussions of backwards convolution and more results are presented for the paper Depth Map Super-Resolution by Deep Multi-Scale Guidance [1].

S2 Filter size of backwards convolution

We intend to show that the optimal filter size of backwards convolution (or deconvolution (**deconv**)) for upsampling is closely related to the upscaling factor s. For conciseness, we consider a single-scale network (SS-Net(ord)) trained in an ordinary domain for upsampling a LR depth map with an upscaling factor $s = 4$. Figure S1 shows an overview of SS-Net(ord). Specifically, the first and third layers perform convolution, whereas the second layer performs backwards strided convolution. Activation function PReLU is used in SS-Net(ord) except the last layer. We set the network parameters: $n_1 = 64, n_2 = 32, n_3 = 1$ and $f_1 = f_3 = 5$. We evaluate the super-resolving performance of SS-Net(ord) by using different deconv filter sizes $f_2 \times f_2$. Figure S2 shows the convergence curves using $f_2 \in (3, 9, 11)$. It can be shown that upsampling accuracy increases with f_2 until it reaches $2s+1$ i.e. $f_2 = 9$. In a compromise between computation efficiency and upsampling performance, we choose deconv filter size to $(2s+1) \times (2s+1)$.

S3 Number of backwards convolution parameters: MS-Net vs SS-Net

We discuss two possible network structures for deconv network with an upscaling factor $s = 2^m$. First, we use a single deconv layer with large output stride ($= 2^m$). Second, we consider a cascade of m deconv layers with small output stride ($= 2$). As presented in Section 3.2 of the main paper, the number of kernel parameters of the former case (SS-Net) is indeed more than that of the latter case (MS-Net). Here, we summarize the reduction ratio of deconv parameters of MS-Net relative to SS-Net in Table S1.

Table S1: The reduction ratio of the number of kernel parameters used in a single large-stride deconv layer against a cascade of multiple small-stride deconv layers

Upsampling factor	2×	4×	8×	16×
Reduction ratio	1.0	1.62	3.85	10.89

S4 Additional results

We provide more visual evaluations on dataset B in Figures S3 to S5 and dataset C[1] in Figures S6 to S9. Error maps are trimmed to the range (0, 30) for better visualization. Hot color represents large depth error and cold color represents small depth error.

[1] We have a set of upsampled depth maps (*Tsukuba*, *Venus*, *Cones* and *Teddy* for the upscaling factors 2, 4 and 8) which was downloaded from the project page of Kwon *et al.* [2] in 2015. We are not able to compute the same RMSE values as Table 1 in their supplementary material. By careful inspection, we have noted: (1) the depth maps for the upscaling factor 8 labeled as "Ours" in Figures 2 to 4 of their supplementary material have same perceptual quality as the upsampled depth maps for the upscaling factor 4 in the evaluation package. (2) These depth maps are "brighter" in intensity than the ground-truths. It is likely that the authors wrongly converted and displayed their upsampled depth maps for the upscaling factor 4 (*not 8*) in an extended depth range.

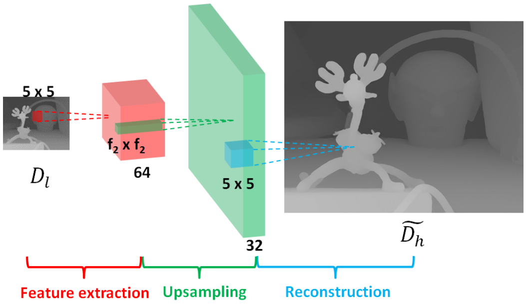


Fig. S1: The network architecture of SS-Net(ord) for single-image super resolution.

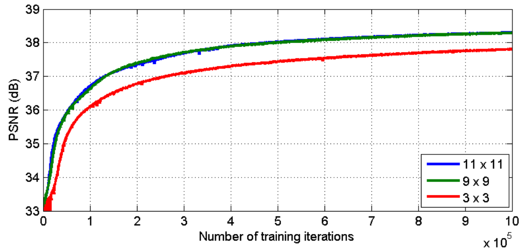
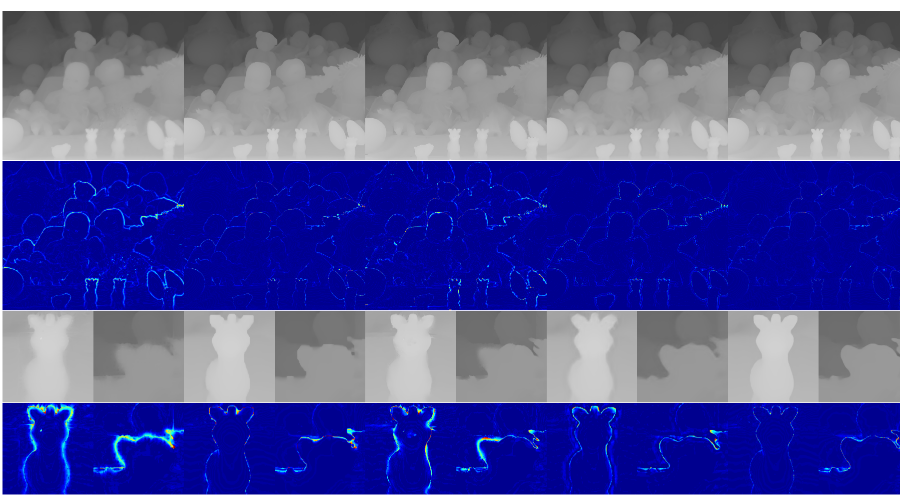


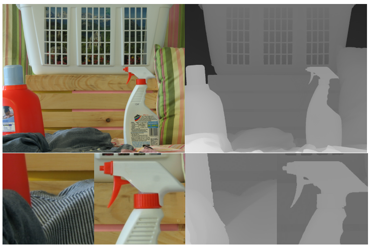
Fig. S2: The convergence curves of SS-Net(ord) using different deconv filter sizes.

RGB image Ground-truth

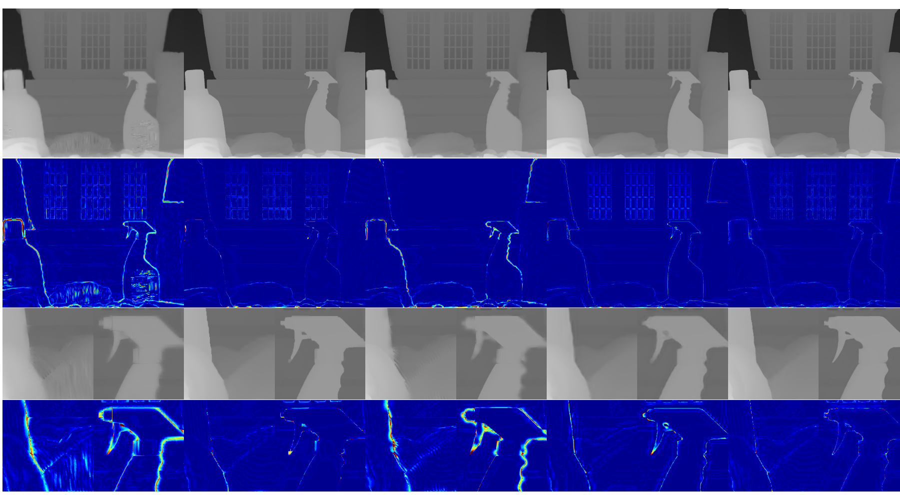


Ferstl *et al.* [3] Kiechle *et al.* [4] Yang *et al.* [5] SRCNN [6] MSG-Net [1]

Fig. S3: Upsampled depth maps and error maps with the upscaling factor 8 for *Dolls* in dataset *B*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.



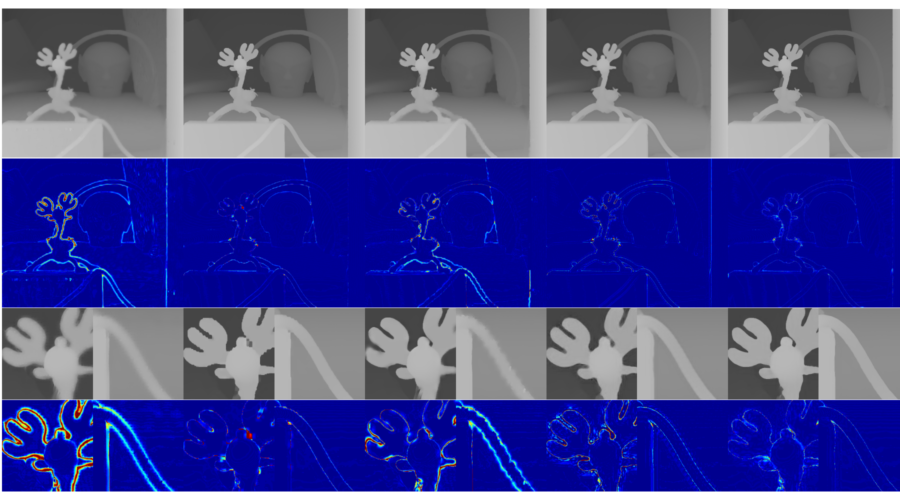
RGB image Ground-truth



Ferstl *et al.* [3] Kiechle *et al.* [4] Yang *et al.* [5] SRCNN [6] MSG-Net [1]

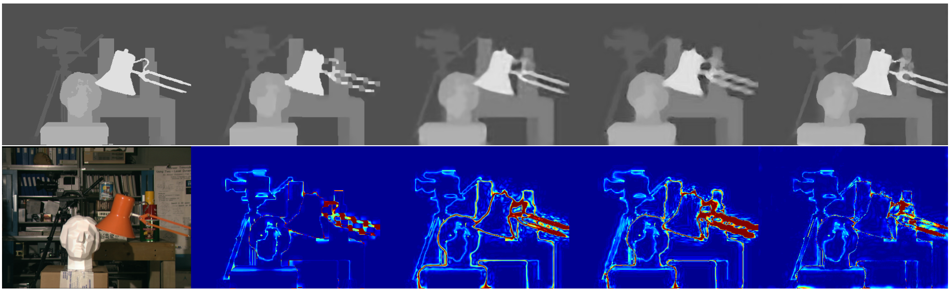
Fig. S4: Upsampled depth maps and error maps with the upscaling factor 8 for *Laundry* in dataset *B*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.

RGB image Ground-truth



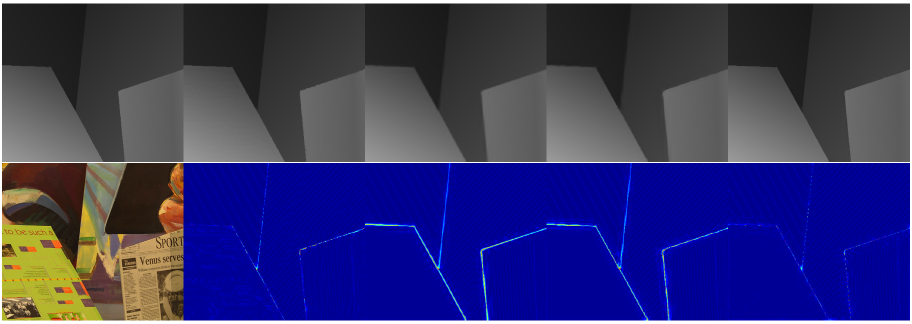
Ferstl *et al.* [3] Kiechle *et al.* [4] Yang *et al.* [5] SRCNN [6] MSG-Net [1]

Fig. S5: Upsampled depth maps and error maps with the upscaling factor 8 for *Reindeer* in dataset B. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.



Ground-truth | Kiechle *et al.* [4] | Wang *et al.* [7] | SRCNN [6] | MSG-Net [1]

Fig. S6: Upsampled depth maps and error maps with the upscaling factors: 8 (top) and 4 (bottom), for *Tsukuba* in dataset *C*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.



Ground-truth | Kiechle *et al.* [4] | Wang *et al.* [7] | SRCNN [6] | MSG-Net [1]

Fig. S7: Upsampled depth maps and error maps with the upscaling factors: 8 (top) and 4 (bottom), for *Venus* in dataset *C*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.

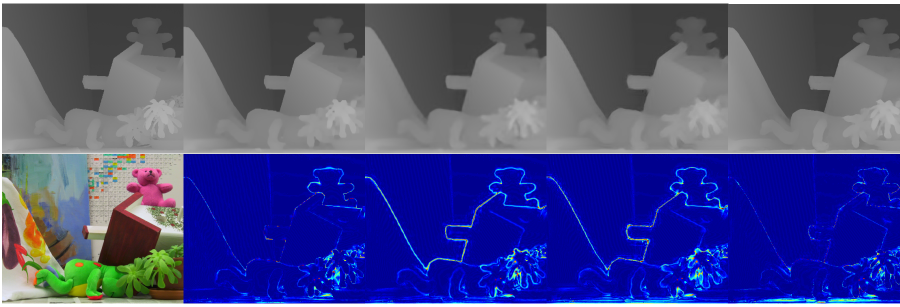


Fig. S8: Upsampled depth maps and error maps with the upscaling factors: 8 (top) and 4 (bottom), for *Teddy* in dataset *C*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.

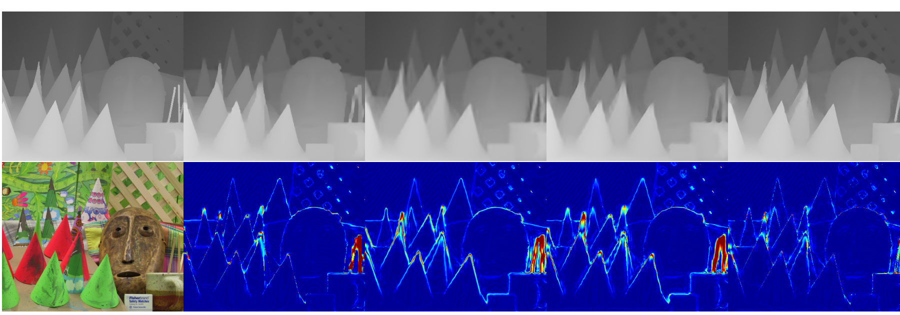


Fig. S9: Upsampled depth maps and error maps with the upscaling factors: 8 (top) and 4 (bottom), for *Cones* in dataset *C*. For the best visual evaluation, we recommend readers to enlarge the figure in the electronic version of this article.

References

1. Hui, T.W., Loy, C.C., Tang, X.: Depth map super-resolution by deep multi-scale guidance. ECCV (2016) 353–369 1, 3, 4, 5, 6, 7
2. Kwon, H., Tai, Y.W., Lin, S.: Data-driven depth map refinement via multi-scale sparse representation. CVPR (2015) 159–167 2
3. Ferstl, D., Reinbachera, C., Ranftl, R., M. Rüther, H.B.: Image guided depth up-sampling using anisotropic total generalized variation. ICCV (2013) 993–1000 3, 4, 5
4. Kiechle, M., Hawe, S., Kleinsteuber, M.: A joint intensity and depth co-sparse analysis model for depth map super-resolution. ICCV (2013) 1545–1552 3, 4, 5, 6, 7
5. Yang, J., Ye, X., Li, K., Hou, C., Wang, Y.: Color-guided depth recovery from RGB-D data using an adaptive autoregressive model. TIP **23**(8) (2014) 3962–3969 3, 4, 5
6. Dong, C., Loy, C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. PAMI (2015) 3, 4, 5, 6, 7
7. Wang, Z., Liu, D., Yang, J., Han, W., Huang, T.: Deep networks for image super-resolution with sparse prior. ICCV (2015) 370–378 6, 7