Deep Joint Image Filtering
Supplementary Material

Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang

https://sites.google.com/site/yijunlimaverick/deepjointfilter

1 The output channels n_3 of $\mathrm{CNN_T}$ and $\mathrm{CNN_G}$

In Section 5 of the paper, we set the output feature maps extracted from the sub-network $\mathrm{CNN_T}$ and $\mathrm{CNN_G}$ as one single channel ($n_3 = 1$). Here we test three cases of using multiple channels. We show the results in Table 1. We observe that using multi-dimensional feature maps does not yield improved performance. As a result, we choose $n_3 = 1$ in our model for computational efficiency.

Table 1. Quantitative results for depth image upsampling (RMSE in centimeters) of using different filter numbers in the 3rd layer of $\mathrm{CNN_T}$ and $\mathrm{CNN_G}$.

n_3	1	16	32	64
$upscale = 8$	6.20	6.40	6.24	6.34

2 Applications of Joint Image Filtering

In this section, we present more results on processing visual signals in various types of domains and compare with the state-of-the-art approaches. Specifically, we validate the effectiveness of our model in seven computer vision and computational photography tasks as shown in Table 2.

Table 2. Applications of joint image filtering

Application	Figures	Source of test images
(a) Joint depth map upsampling	Figure 1 – Figure 3	[1,2,3,4,5]
(b) Chromaticity map upsampling	Figure 4 – Figure 5	[6]
(c) Saliency map upsampling	Figure 6 – Figure 9	[7]
(d) Cross-modal noise reduction	Figure 10 – Figure 11	[8]
(e) Inverse halftoning	Figure 12 – Figure 13	[9]
(f) Tone mapping	Figure 14 – Figure 15	[10]
(g) Texture removal	Figure 16 – Figure 17	[11]

Joint depth map upsampling: We compare our approach against several state-of-the-art joint image filters for depth map upsampling. Among them, JBU [12] and GF [13] are generic methods for joint image upsampling while TGV [14] and Park [15] are algorithms specifically designed for image guided depth upsampling.

Chromaticity map upsampling: We compare our approach with three different upsampling methods including Bicubic, GF [13], Ham [16] as well as the direct solution of [6] on the high-resolution image. Figure 4–5 show that our results in general have less color bleeding artifacts. We also report the combined running time, which consists of two parts: (1) solving chromaticity solution map on a low-resolution image [6] and (2) upsampling using joint image filter.

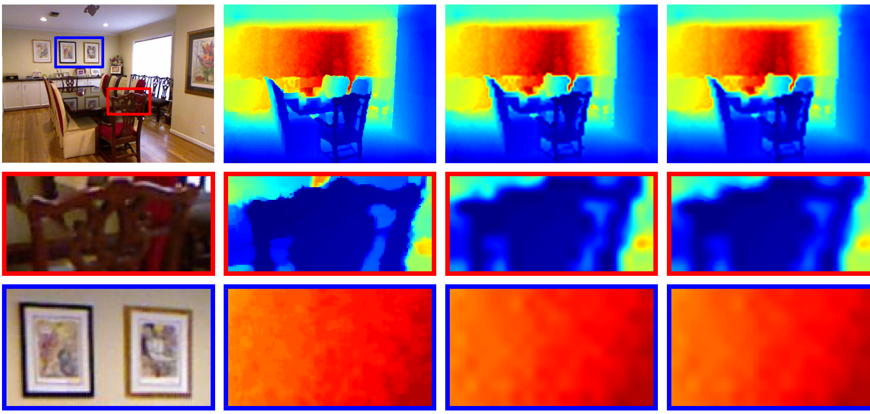
Saliency map upsampling: We show an application of joint image filter for saliency map upsampling. We first obtain the low-resolution saliency map using [7] on the low-resolution image ($10\times$) and then upsample it to the original spatial resolution. We compare our approach with Bicubic, GF [13], Ham [16] as well as the direct solution of [7] on the high-resolution image. Figure 6–9 show that our results in general could better preserve the boundary under a large upsampling scale.

Cross-modal noise reduction: Cross-modal images can be obtained by different sensors or under different imaging settings, such as Flash/Non-Flash, RGB/NIR. We compare our filter with GF [13], Ham [16] and a specialized cross-modality image restoration algorithm [8] on filtering cross-modality image pairs. In Figure 10-11, compared with GF [13] and Ham [16], we show that our filter could successfully suppress noises while preserving main salient structures.

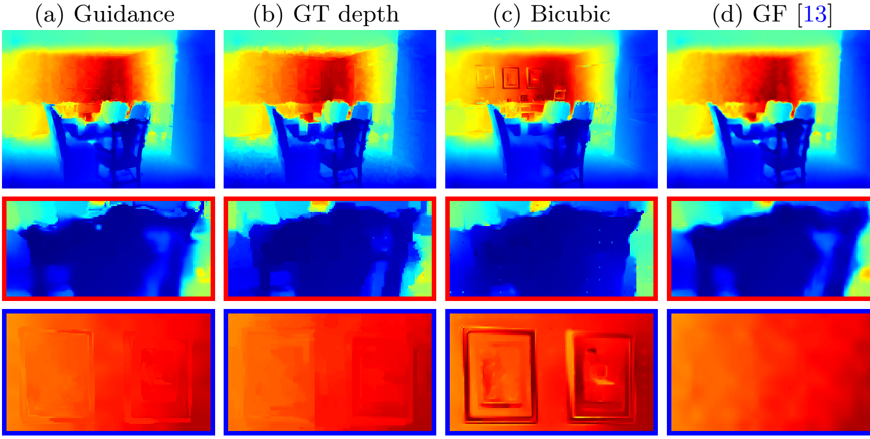
Inverse halftoning: We compare our results with that from RGF [17], L0 [18], Xu [19] and a specific algorithm for reconstructing halftoned images Kopf [9]. We show in Figure 12-13 that our filter can preserve edges well and achieve comparable performance compared to the state-of-the-art techniques.

Tone mapping: In Figure 14-15, we use our joint filter as an edge-preserving filter for tone mapping. We show that we achieve comparable results compared with BF [20] and GF [13].

Texture removal: We apply the proposed deep joint filter for removing textures. We show in Figure 16-17 that our filter is capable of removing textures and rendering comparable results with RGF [17], L0 [18], Xu [11] and Ham [16].



(a) Guidance (b) GT depth (c) Bicubic (d) GF [13]



(e) JBU [12] (f) Park [15] (g) TGV [14] (h) Ours

Fig. 1. Qualitative comparisons of joint depth map upsampling results ($8\times$) from the NYU v2 dataset [2].

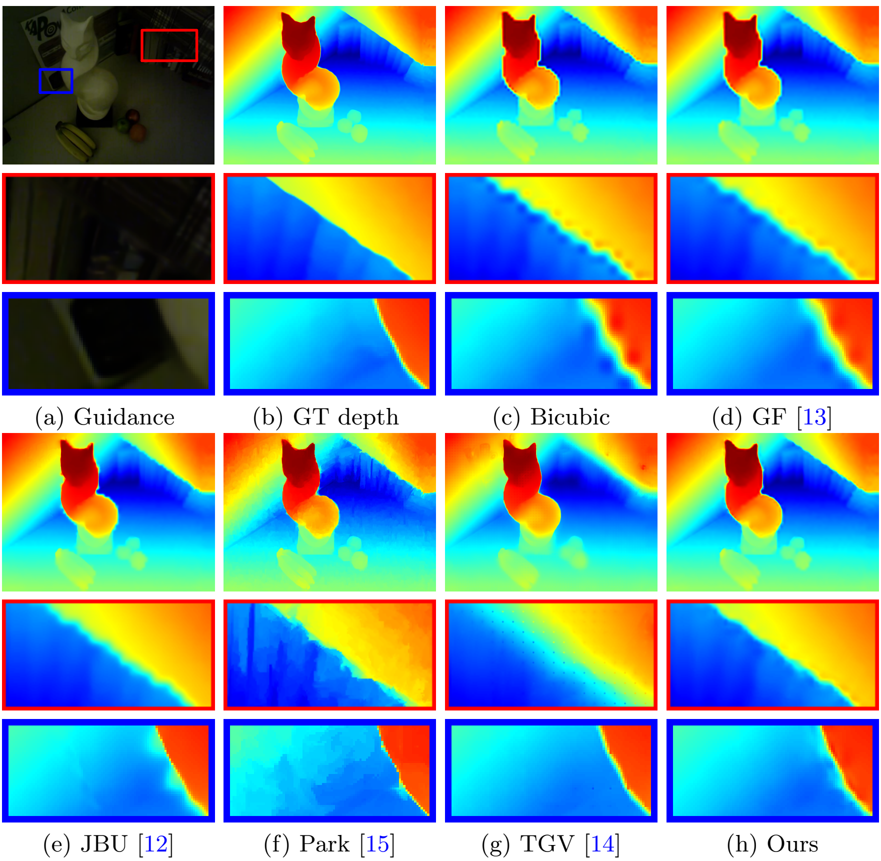


Fig. 2. Qualitative comparisons of joint depth map upsampling results ($8\times$) from the dataset [1] .

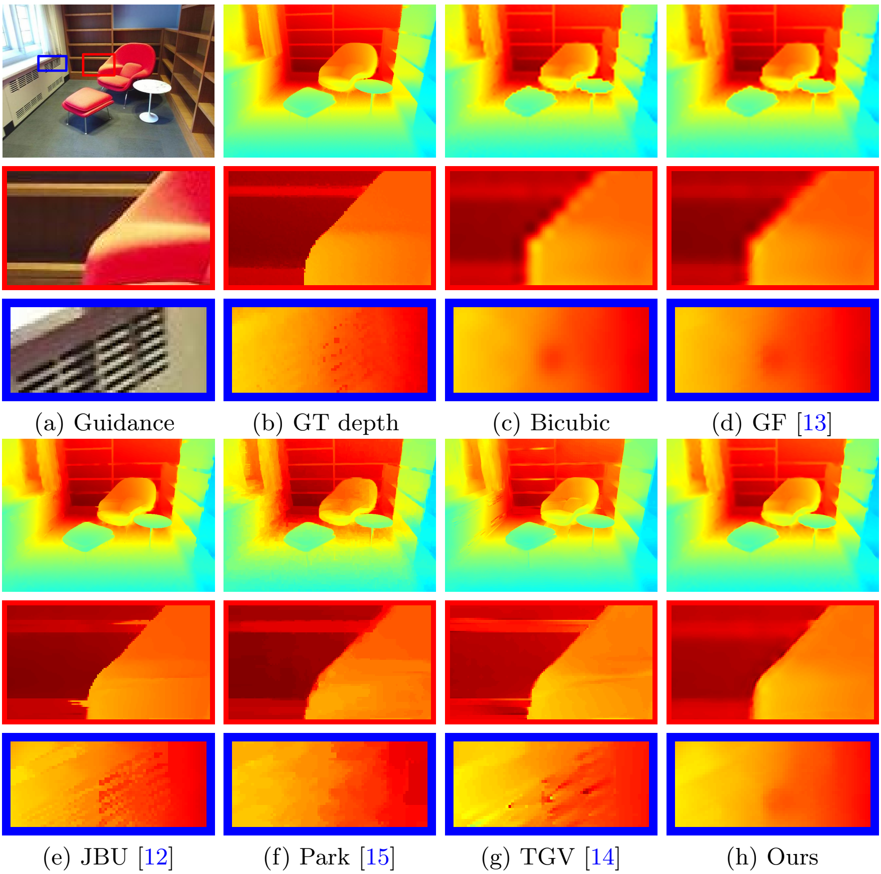
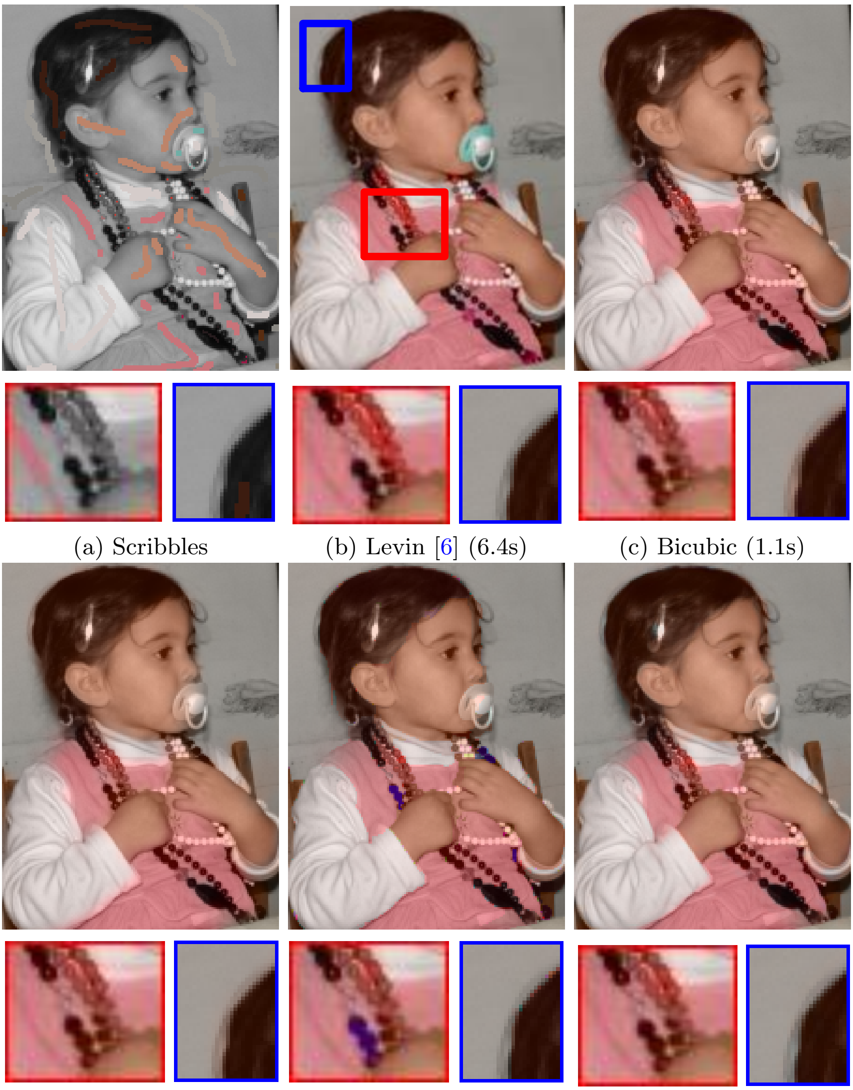


Fig. 3. Qualitative comparisons of joint depth map upsampling results (8×) from the SUN RGBD dataset [3].

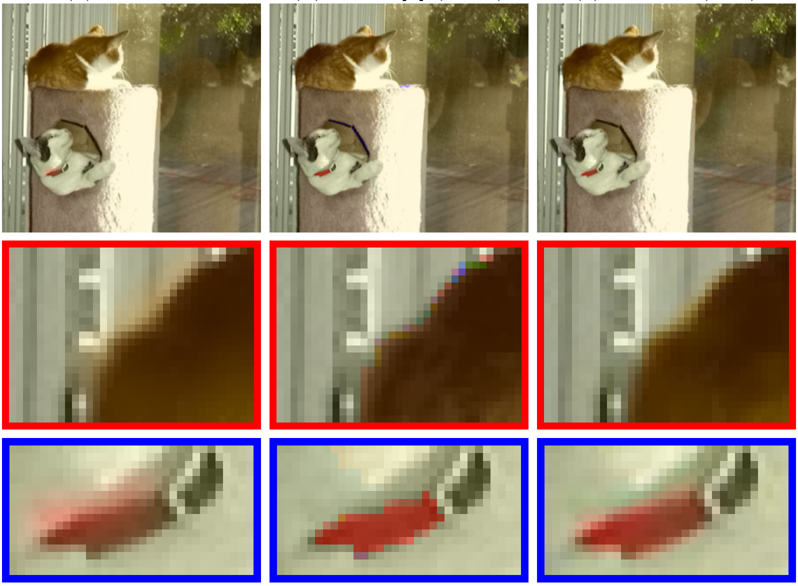


(a) Scribbles (b) Levin [6] (6.4s) (c) Bicubic (1.1s)

(d) GF [13] (1.3s) (e) Ham [16] (22.3s) (f) Ours (2.1s)

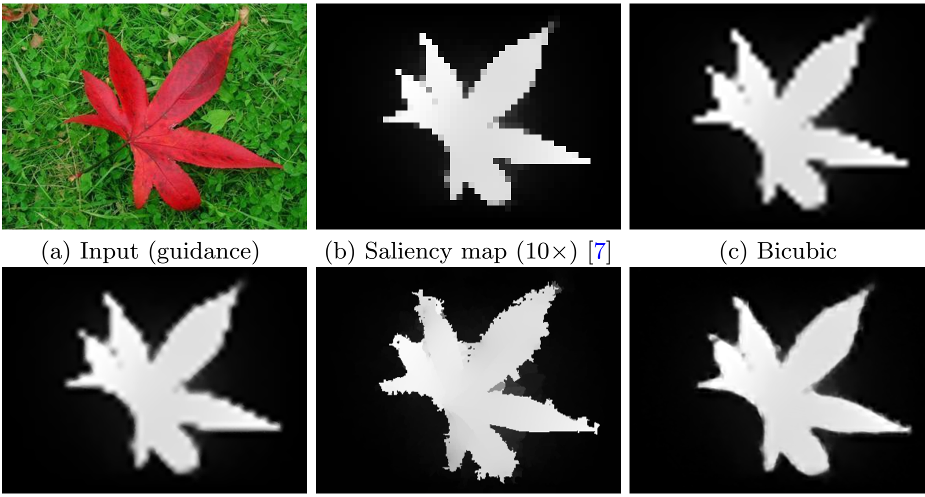
Fig. 4. Qualitative comparisons of chromaticity map upsampling (4×) .

(a) Scribbles (b) Levin [6] (11.2s) (c) Bicubic (1.8s)



(a) GF [13] (2.1s) (b) Ham [16] (26.8s) (c) Ours (3.6s)

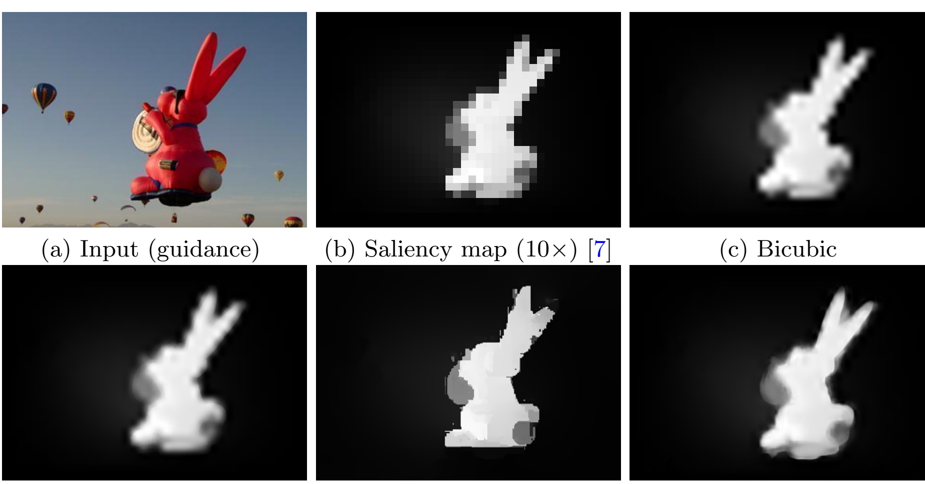
Fig. 5. Qualitative comparisons of chromaticity map upsampling (4×).



(a) Input (guidance) (b) Saliency map (10×) [7] (c) Bicubic

(d) GF [13] (e) Ham [16] (f) Ours

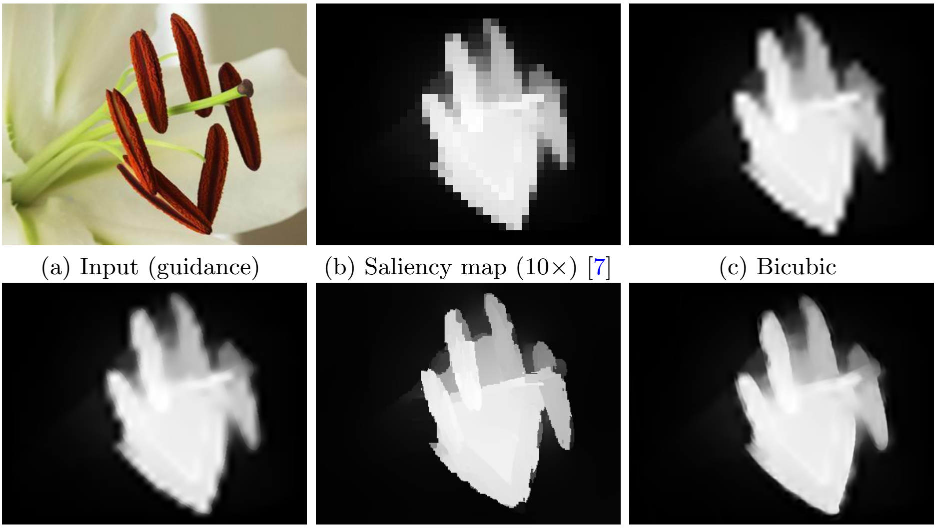
Fig. 6. Visual comparisons of saliency map upsampling results (10×).



(a) Input (guidance) (b) Saliency map (10×) [7] (c) Bicubic

(d) GF [13] (e) Ham [16] (f) Ours

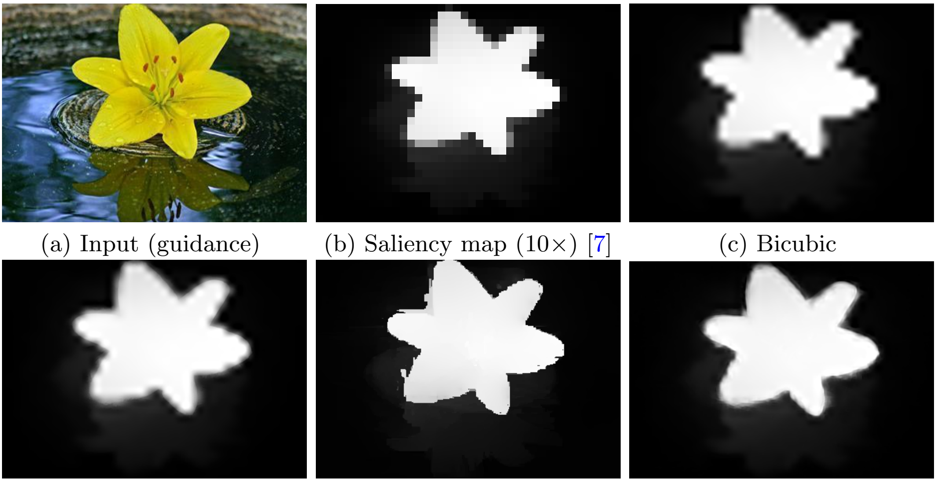
Fig. 7. Visual comparisons of saliency map upsampling results (10×).



(a) Input (guidance) (b) Saliency map (10×) [7] (c) Bicubic

(d) GF [13] (e) Ham [16] (f) Ours

Fig. 8. Visual comparisons of saliency map upsampling results (10×).

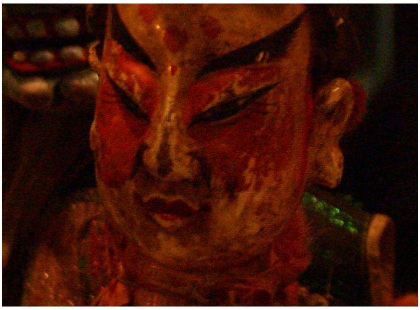


(a) Input (guidance) (b) Saliency map (10×) [7] (c) Bicubic

(d) GF [13] (e) Ham [16] (f) Ours

Fig. 9. Visual comparisons of saliency map upsampling results (10×).

(a) Guidance Flash



(b) Target Non-Flash

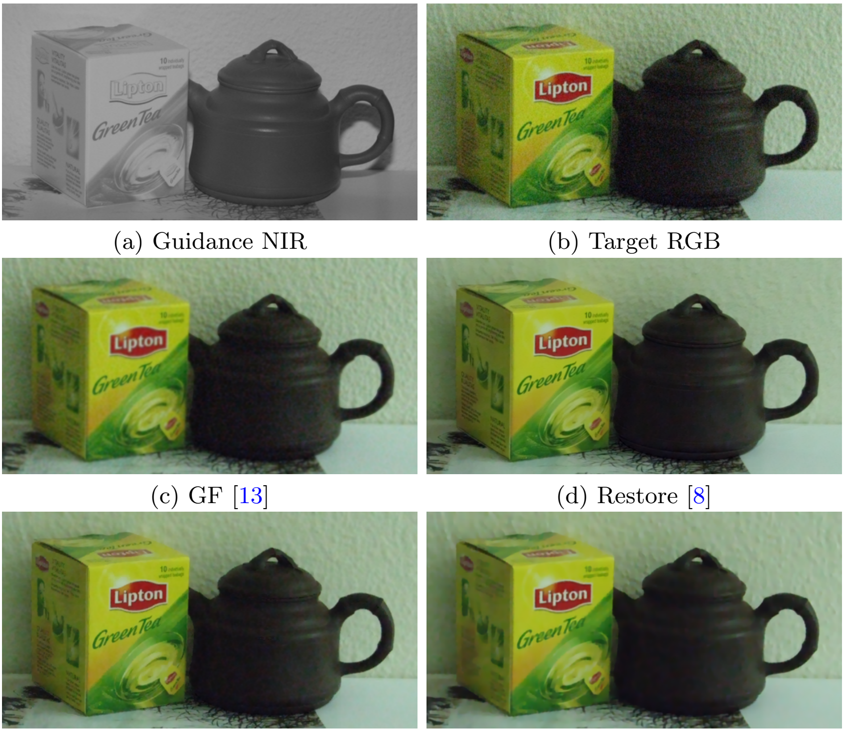
(c) GF [13]

(d) Restore [8]

(e) Ham [16]

(f) Ours

Fig. 10. Comparisons of noise reduction results using a pair of Flash/Non-Flash images.



(a) Guidance NIR (b) Target RGB

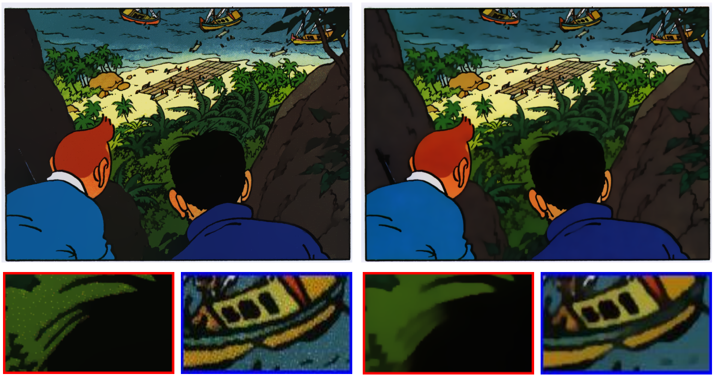
(c) GF [13] (d) Restore [8]

(e) Ham [16] (f) Ours

Fig. 11. Comparisons of noise reduction results using a pair of RGB/NIR images.

(a) Input

(b) RGF [17]



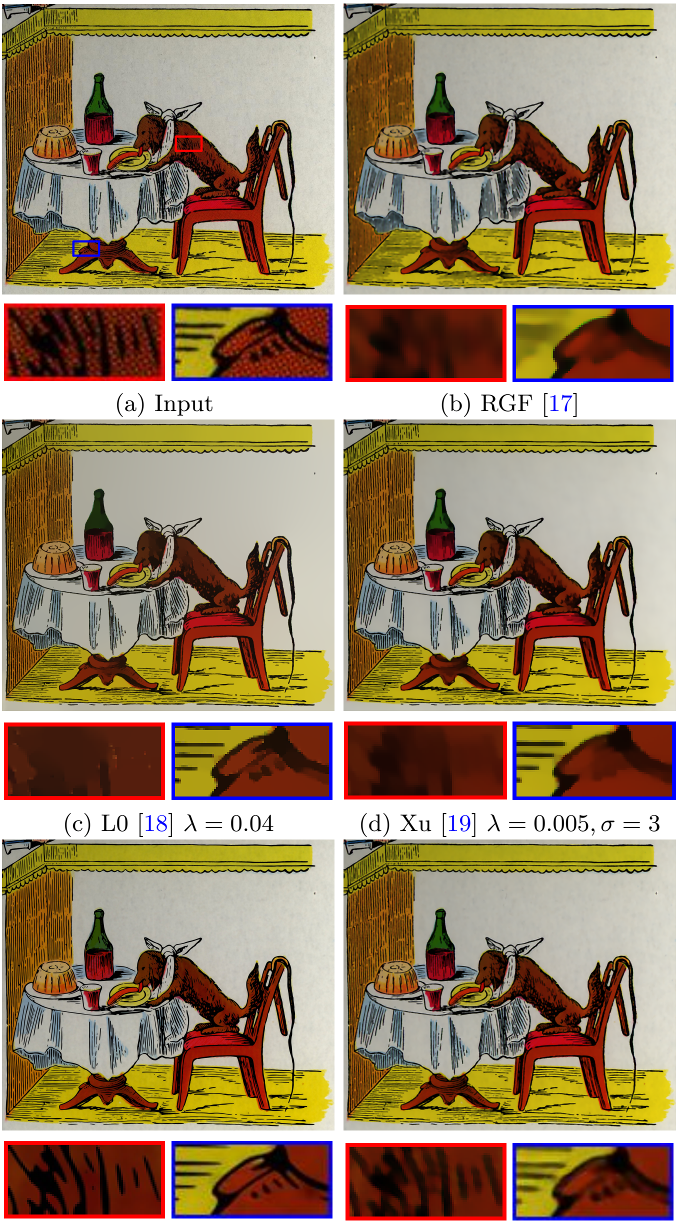
(c) L0 [18] $\lambda = 0.015$

(d) Xu [19] $\lambda = 0.005, \sigma = 3$

(e) Kopf [9]

(f) Ours

Fig. 12. Comparisons of inverse halftoning results.



(a) Input (b) RGF [17]
(c) L0 [18] $\lambda = 0.04$ (d) Xu [19] $\lambda = 0.005, \sigma = 3$
(e) Kopf [9] (f) Ours

Fig. 13. Comparisons of inverse halftoning results.

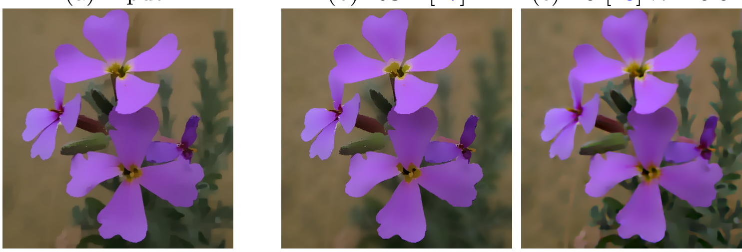
(a) BF [20] (b) GF [13] (c) Ours

Fig. 14. Visual comparisons of tone mapping results on the *memorial* HRD image.

(a) BF [20] (b) GF [13] (c) Ours

Fig. 15. Visual comparisons of tone mapping results on the *bigFogMap* HRD image.

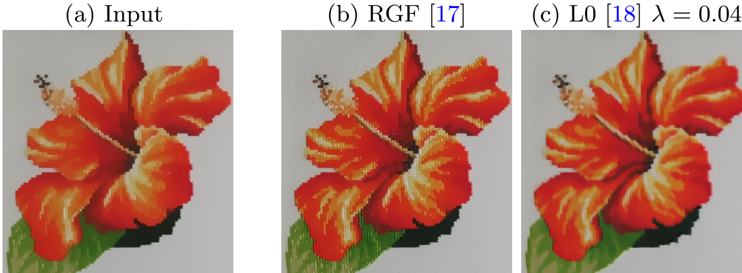
(a) Input (b) RGF [17] (c) L0 [18] $\lambda = 0.04$



(d) Xu [19] $\lambda = 0.01, \sigma = 0.5$ (e) Ham [16], $\lambda = 30$ (f) Ours

Fig. 16. Visual comparisons of texture removal methods.

(a) Input (b) RGF [17] (c) L0 [18] $\lambda = 0.04$



(d) Xu [19] $\lambda = 0.01, \sigma = 0.5$ (e) Ham [16], $\lambda = 30$ (f) Ours

Fig. 17. Visual comparisons of texture removal methods.

References

1. Lu, S., Ren, X., Liu, F.: Depth enhancement via low-rank matrix completion. In: CVPR. (2014) 1, 4
2. Nathan Silberman, Derek Hoiem, P.K., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: ECCV. (2012) 1, 3
3. Song, S., Lichtenberg, S.P., Xiao, J.: Sun rgb-d: A rgb-d scene understanding benchmark suite. In: CVPR. (2015) 1, 5
4. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: CVPR. (2007) 1
5. Hirschmüller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: CVPR. (2007) 1
6. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: SIGGRAPH. (2004) 1, 2, 6, 7
7. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: CVPR. (2013) 1, 2, 8, 9
8. Yan, Q., Shen, X., Xu, L., Zhuo, S., Zhang, X., Shen, L., Jia, J.: Cross-field joint image restoration via scale map. In: ICCV. (2013) 1, 2, 10, 11
9. Kopf, J., Lischinski, D.: Digital reconstruction of halftoned color comics. In: SIGGRAPH. (2012) 1, 2, 12, 13
10. Fattal, R., Lischinski, D., Werman, M.: Gradient domain high dynamic range compression. ACM Transactions on Graphics **21**(3) (2002) 249–256 1
11. Xu, L., Ren, J., Yan, Q., Liao, R., Jia, J.: Deep edge-aware filters. In: ICML. (2015) 1, 2
12. Kopf, J., Cohen, M.F., Lischinski, D., Uyttendaele, M.: Joint bilateral upsampling. In: SIGGRAPH. (2007) 2, 3, 4, 5
13. He, K., Sun, J., Tang, X.: Guided image filtering. PAMI **35**(6) (2013) 1397–1409 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14
14. Ferstl, D., Reinbacher, C., Ranftl, R., Rüther, M., Bischof, H.: Image guided depth upsampling using anisotropic total generalized variation. In: ICCV. (2013) 2, 3, 4, 5
15. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.: High quality depth map upsampling for 3d-tof cameras. In: ICCV. (2011) 2, 3, 4, 5
16. Ham, B., Cho, M., Ponce, J.: Robust image filtering using joint static and dynamic guidance. In: CVPR. (2015) 2, 6, 7, 8, 9, 10, 11, 15
17. Zhang, Q., Shen, X., Xu, L., Jia, J.: Rolling guidance filter. In: ECCV. (2014) 2, 12, 13, 15
18. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via ℓ_0 gradient minimization. In: ACM SIGGRAPH ASIA. (2011) 2, 12, 13, 15
19. Xu, L., Yan, Q., Xia, Y., Jia, J.: Structure extraction from texture via relative total variation. ACM Transactions on Graphics **31**(6) (2012) 139 2, 12, 13, 15
20. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV. (1998) 2, 14