Appendix: Guided Image Filtering-Conventional to Deep Models: A Review and Evaluation Study

Weimin Yuan, Yinuo Wang, Cai Meng, Xiangzhi Bai

Image Processing Center, Beihang University, Beijing 100191, China

1. SSIM results on Depth Upsampling

Depth upsampling is chosen as the evaluation task for GIF methods due to its practical relevance in applications such as 3D reconstruction and virtual reality, where enhancing depth image resolution while preserving edge sharpness is critical. This task effectively tests the edge-preserving and structure-transferring capabilities of GIF algorithms under challenging conditions. Moreover, depth upsampling is a standard benchmark, facilitating meaningful comparisons with existing methods and providing both quantitative and qualitative metrics to assess the performance of different methods.

We evaluate the performance of 20 representative approaches, including 8 local methods: GIF [1] [1], WGIF [2] [2], GDGIF [3] [3], EGIF [4] [4], SKWGIF [5] [5], SGIF [6] [6], AnGIF [7] [7] and WSGGF [8] [8], 4 global methods: MSJF [9] [9], SDF [10] [10], MuGIF [11] [11] and GSF [12] [12], and 8 deep learning-based

[1] https://kaiminghe.github.io/
[2] http://koufei.weebly.com
[3] http://koufei.weebly.com
[4] https://github.com/bangyuanl/egif
[5] https://github.com/altlp/SKWGIF
[6] http://www.vista.ac.cn/side-window-filtering/
[7] https://codeocean.com/capsule/0267263/tree
[8] https://github.com/weimin581/WSGGF
[9] http://www.cse.cuhk.edu.hk/leojia/projects/mutualstructure/
[10] http://www.di.ens.fr/willow/research/sdfilter
[11] https://sites.google.com/view/xjguo/mugif
[12] https://github.com/wliusjtu/Generalized-Smoothing-Framework

methods: DJF [13] [13], DJFR [14] [14], SVLRM [15] [15], DKN [16] [16], FDKN [16] [17], CUNet [17][18], UnGIF [18] [19] and DAGF [19] [20]. For a fair comparison, the results of these methods are obtained by using the source codes released by authors with default parameters.

We evaluate the methods using 3 commonly employed RGB-D datasets: NYU-2 [20] [21], Middlebury [21] [22] and Lu [22] [23]. We follow the setup of prior works [13, 14, 16, 19], including data pre-processing, and train/test splits. Specifically, NYU-2 [20] contains 1449 RGB/depth pairs captured with Microsoft Kinect using structured light. The first 1000 paired images are used as training set for deep-learning GIF methods, and the remaining 449 image pairs to evaluate the performance of all compared methods. In addition, the pre-trained model on NYU-2 of deep learning-based methods are evaluated on Middlebury [21], contains 30 RGB/depth image pairs from 2001-2006 datasets, and Lu [22], contains 6 image pairs captured by ASUS Xtion Pro camera. Following [13, 14, 16, 19, 23], we use the root mean square error (RMSE) as evaluation metric, where a lower RMSE indicates better quality. For the NYU-2 dataset, RMSE is computed in centimeters, whereas for the Middlebury and Lu datasets, RMSE is calculated with upsampled depth scaled to the range [0,255].

Besides RMSE, SSIM [24] is also employed for evaluating the performance. SSIM focuses on structural similarity and edge preservation, it serves as a crucial metric for evaluating the quality of depth map restoration.

As SSIM focuses on structural similarity and edge preservation, it serves as a crucial metric for evaluating the quality of depth map restoration, es-

[13] https://github.com/Yijunmaverick/DeepJointFilter
[14] https://github.com/Yijunmaverick/DeepJointFilter
[15] https://github.com/curlyqian/SVLRM
[16] https://github.com/cvlab-yonsei/dkn
[17] https://github.com/cvlab-yonsei/dkn
[18] https://github.com/cindydeng1991/TPAMI-CU-Net
[19] https://github.com/shizenglin/Unsharp-Mask-Guided-Filtering
[20] https://github.com/zhwzhong/DAGF
[21] https://cs.nyu.edu/~fergus/datasets/nyu_depth_v2.html
[22] https://vision.middlebury.edu/stereo/data/
[23] https://web.cecs.pdx.edu/~fliu/project/depth-enhance/

Table 1 – The SSIM [24] on depth upsampling of Middlebury [21], NYU-v2 [20] and Lu [22] for scale factors 4×, 8× and 16×. Higher SSIM indicates better quality. Best result in bold, second-best is underlined.

Method	NYU-2[20]			Middlebury[21]			Lu[22]			Average		
	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×
Input	0.9338	0.9202	0.9097	0.9545	0.9226	0.9051	0.9442	0.9214	0.9074	0.9442	0.9214	0.9074
GIF[1]	0.9422	0.9349	0.9106	0.9614	0.9368	0.9133	0.9518	0.9359	0.9120	0.9518	0.9359	0.9119
WGIF[2]	0.9429	0.9362	0.9118	0.9673	0.9407	0.9152	0.9551	0.9384	0.9134	0.9551	0.9384	0.9135
GDGIF[3]	0.9438	0.9366	0.9111	0.9671	0.9412	0.9133	0.9555	0.9390	0.9123	0.9555	0.9390	0.9122
EGIF[4]	0.9442	0.9367	0.9128	0.9679	0.9457	0.9108	0.9561	0.9412	0.9118	0.9561	0.9412	0.9118
SKWGIF[5]	0.9451	0.9370	0.9141	0.9671	0.9461	0.9119	0.9561	0.9416	0.9131	0.9561	0.9416	0.9130
SGIF[6]	0.9382	0.9303	0.9108	0.9560	0.9604	0.9078	0.9471	0.9409	0.9049	0.9471	0.9439	0.9078
AnGIF[7]	0.9450	0.9397	0.9131	0.9687	0.9475	0.9118	0.9569	0.9436	0.9125	0.9569	0.9769	0.9125
WSGGF[8]	0.9444	0.9380	0.9135	0.9584	0.9628	0.9108	0.9514	0.9460	0.9077	0.9514	0.9489	0.9107
MSJF[9]	0.9411	0.9412	0.9121	0.9631	0.9401	0.9126	0.9521	0.9406	0.9123	0.9521	0.9406	0.9123
SDF[10]	0.9552	0.9643	0.9383	0.9631	0.9368	0.9076	0.9592	0.9414	0.9138	0.9592	0.9475	0.9199
MuGIF[11]	0.9425	0.9504	0.9290	0.9540	0.9370	0.9019	0.9483	0.9359	0.9076	0.9483	0.9411	0.9128
GSF[12]	0.9547	0.9488	0.9157	0.9529	0.9321	0.9028	0.9538	0.9404	0.9092	0.9538	0.9404	0.9092
DJF[13]	0.9934	0.9801	0.9548	0.9927	0.9837	0.9547	0.9931	0.9819	0.9548	0.9931	0.9819	0.9548
DJFR[14]	0.9937	0.9807	0.9551	0.9930	0.9839	0.9550	0.9934	0.9823	0.9551	0.9934	0.9823	0.9551
DGF[25]	0.9897	0.9776	0.9568	0.9934	0.9842	0.9692	0.9916	0.9809	0.9630	0.9916	0.9809	0.9630
SVLRM[15]	0.9946	0.9835	0.9616	0.9921	0.9845	0.9567	0.9934	**0.9840**	0.9591	0.9934	0.9840	0.9591
DKN[16]	0.9940	0.9829	0.9607	0.9909	0.9838	0.9559	0.9925	0.9834	0.9583	0.9925	0.9834	0.9583
FDKN[16]	0.9937	0.9821	0.9577	0.9903	0.9837	0.9554	0.9920	0.9829	0.9565	0.9920	0.9829	0.9565
CUNet[17]	0.9948	0.9812	0.9607	0.9941	0.9846	0.9744	0.9945	0.9829	0.9676	0.9945	0.9829	0.9676
UnGIF[18]	**0.9949**	**0.9863**	**0.9706**	0.9943	**0.9864**	**0.9754**	0.9945	0.9829	0.9676	**0.9946**	**0.9852**	**0.9712**
DAGF[19]	0.9947	0.9830	0.9681	**0.9945**	0.9840	0.9704	**0.9946**	0.9835	**0.9693**	**0.9946**	0.9835	0.9693

pecially when compared to RMSE. Table 1 presents the SSIM for depth upsampling on Middlebury, NYU-v2. From Table 1, we can draw the following conclusions:

1. Local methods such as WGIF, GDGIF, EGIF, and AnGIF show relatively consistent SSIM performance, maintaining good edge preservation. For example, AnGIF achieves competitive SSIM scores across all scales. However, these methods tend to degrade at larger scales, as seen at 16×, where SSIM values slightly drop, indicating the difficulty of transferring fine structural information from RGB guidance to depth maps at high upsampling ratios. WSGGF also exhibits similar trends, with good edge preservation at

smaller scales but a slight decrease in SSIM at larger scales.

2. Global methods, such as MSJF, SDF, and MuGIF, generally outperform local methods in terms of SSIM, particularly at higher upsampling factors. SDF, in particular, demonstrates strong edge and structural preservation at larger scales, outperforming local methods. The MSJF and MuGIF methods also perform well, although not as robustly as SDF at larger scales. These results suggest that global methods, which optimize depth upsampling through iterative refinement, can better preserve structural consistency and edge information over large upsampling ratios.

3. Data-driven based DJF, DJFR, and DGF achieve higher SSIM values across all scale factors. The SSIM scores at $4\times$ and $8\times$ highlight its strong ability to preserve both fine details and structural integrity. Linear models like SVLRM, CUNet, UnGIF, and DAGF also perform strongly in terms of SSIM. Notably, UnGIF achieves the highest SSIM scores across most datasets and scales, outperforming other methods. SVLRM shows a more moderate but consistent performance, with values close to the best-performing methods, especially at $4\times$ and $16\times$ scales. These methods strike a good balance between edge preservation and structural consistency, making them competitive for depth upsampling tasks.

The SSIM results further corroborate the effectiveness of global and data-driven methods in maintaining structural integrity and edge preservation during depth upsampling. Overall, SSIM serves as a valuable complementary metric to RMSE, providing a more detailed understanding of how well each method preserves structural and edge information in depth upsampling tasks.

References

[1] K. He, J. Sun, X. Tang, Guided image filtering, in: European Conference on Computer Vision (ECCV), Springer, 2010, pp. 1–14.

[2] Z. Li, J. Zheng, Z. Zhu, W. Yao, S. Wu, Weighted guided image filtering, IEEE Transactions on Image processing 24 (1) (2015) 120–129.

[3] F. Kou, W. Chen, C. Wen, Z. Li, Gradient domain guided image filtering, IEEE Transactions on Image processing 24 (2015) 4528–4539.

[4] Z. Lu, B. Long, K.-M. Li, F. Lu, Effective guided image filtering for contrast enhancement, IEEE Signal Processing Letter 25 (2018) 1585–1589.

[5] Z. Sun, B. Han, J. Li, J. Zhang, X. Gao, Weighted guided image filtering with steering kernel, IEEE Transactions on Image processing 29 (2020) 500–508.

[6] H. Yin, Y. Gong, G. Qiu, Side window filtering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 8758–8766.

[7] C. N. Ochotorena, Y. Yamashita, Anisotropic guided filtering, IEEE Transactions on Image Processing 29 (2020) 1397–1412.

[8] W. Yuan, C. Meng, X. Bai, Weighted side-window based gradient guided image filtering, Pattern Recognition (2023) 110006.

[9] X. Shen, C. Zhou, L. Xu, J. Jia, Mutual-structure for joint filtering, International Journal of Computer Vision 125 (2015) 19–33.

[10] B. Ham, M. Cho, J. Ponce, Robust guided image filtering using nonconvex potentials, IEEE transactions on pattern analysis and machine intelligence 40 (2018) 192–207.

[11] X. Guo, Y. Li, J. Ma, H. Ling, Mutually guided image filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (3) (2020) 694–707.

[12] W. Liu, P. Zhang, Y. Lei, X. Huang, J. Yang, M. Ng, A generalized framework for edge-preserving and structure-preserving image smoothing, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2019) 6631–6648.

[13] Y. Li, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep joint image filtering, in: Computer Vision–ECCV 2016: 14th European Conference, Springer, 2016, pp. 154–169.

[14] Y. Li, J.-B. Huang, N. Ahuja, M.-H. Yang, Joint image filtering with deep convolutional networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2017) 1909–1923.

[15] J. Dong, J. Pan, J. S. J. Ren, L. Lin, J. Tang, M.-H. Yang, Learning spatially variant linear representation models for joint filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2021) 8355–8370.

[16] B. Kim, J. Ponce, B. Ham, Deformable kernel networks for joint image filtering, International Journal of Computer Vision 129 (2) (2021) 579–600.

[17] X. Deng, P. L. Dragotti, Deep convolutional neural network for multi-modal image restoration and fusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (10) (2021) 3333–3348.

[18] Z. Shi, Y. Chen, E. Gavves, P. Mettes, C. G. M. Snoek, Unsharp mask guided filtering, IEEE Transactions on Image Processing 30 (2021) 7472–7485.

[19] Z. Zhong, X. Liu, J. Jiang, D. Zhao, X. Ji, Deep attentional guided image filtering, IEEE Transactions on Neural Networks and Learning Systems (2023).

[20] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from rgbd images, in: European Conference on Computer Vision (ECCV), Springer, 2012, pp. 746–760.

[21] D. Scharstein, C. Pal, Learning conditional random fields for stereo, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2007, pp. 1–8.

[22] S. Lu, X. Ren, F. Liu, Depth enhancement via low-rank matrix completion, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3390–3397.

[23] Z. Wang, Z. Yan, M.-H. Yang, J. Pan, J. Yang, Y. Tai, G. Gao, Scene prior filtering for depth map super-resolution, arXiv preprint arXiv:2402.13876 (2024).

[24] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.

[25] H. Wu, S. Zheng, J. Zhang, K. Huang, Fast end-to-end trainable guided filter, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1838–1847.