Guided Image Filtering: A Survey and Evaluation Study

Weimin Yuan
yuanweimin@buaa.edu.cn
Image Processing Center, Beihang University
Beijing, China

Yinuo Wang
wangyinuo@buaa.edu.cn
Image Processing Center, Beihang University
Beijing, China

Cai Meng
Tsai@buaa.edu.cn
Image Processing Center, Beihang University
Beijing, China

Xiangzhi Bai
jackybxz@buaa.edu.cn
Image Processing Center, Beihang University
Beijing, China

ABSTRACT

In the past decade, there has been an increasing success of guided image filtering (GIF). Leveraging the guidance image as a prior and transferring the structural details to the target image, GIF has demonstrated its ability in faithfully preserving image edges while maintaining low computational complexity. Additionally, GIF exhibits good capability in extracting and characterizing images from various domains. Researchers have proposed large numbers of GIF-like variants. Nevertheless, limited effort has been devoted to a systematic review and evaluation of these methods. To fill this gap, this paper provides a comprehensive survey of existing GIF-like methods, including model- and deep learning-based approaches. Moreover, extensive experiments are conducted to compare the performance of 18 representative methods. Analysis of the qualitative and quantitative results reveals several observations concerning the current state of this area.

KEYWORDS

Guided Image Filtering; Depth Upsampling; Noise Reduction.

ACM Reference Format:
Weimin Yuan, Yinuo Wang, Cai Meng, and Xiangzhi Bai. 2023. Guided Image Filtering: A Survey and Evaluation Study. In *ACM Multimedia Asia Workshops (MMAsia '23 Workshops), December 6–8, 2023, Tainan, Taiwan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3611380.3629544

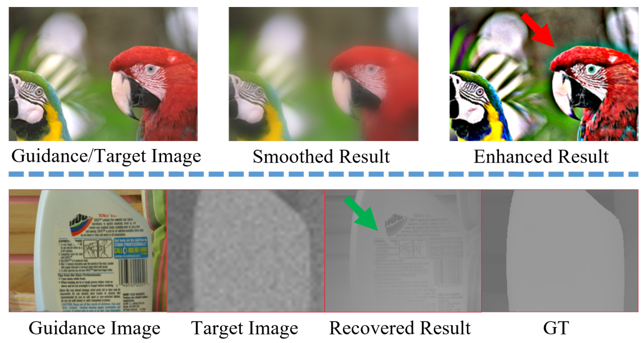

Figure 1: Two visual examples to show the limitations of GIF. Top row: Blurred edges in the smoothed result can lead to halo artifacts in final enhanced result, as indicated by red arrow. Bottom row: When the structures in guidance and target images are inconsistent, GIF would transfer extraneous or incorrect contents to the recovered result, causing texture-copy artifact, as indicated by green arrow.

1 INTRODUCTION

Image filtering, a crucial component in many image processing and computer vision tasks, is widely used to reduce unwanted signals (e.g., noise) while preserving desired ones (e.g., edges and structures). Conventional filters such as Gaussian, Laplacian and mean filters employ spatially invariant kernels, and thus lack consideration of the image content, leading to edge-blurring due to their content-blindness.

To address the challenge of content-blindness, Guided Image Filtering (GIF) [7] has received extensive attention from the research community. The key idea of GIF is to utilize an additional image as guidance to transfer the structure information of the guidance to the degraded target image, thus allowing the restoration of blurred edges or suppression of noise. This provides a new way to view the filtering process and has a wide range of applications. As shown in Table 1, we can categorize the applications into two classes based on whether the guidance and target images are derived from the same modality: self-guidance based and reference-guidance based tasks. However, two key assumptions of GIF, namely the locally linear model and the structure consistency, are often violated in certain scenes, leading to halo and texture-copy artifacts, respectively. Specifically, locally linear model assumes a local linear relationship between the output and guidance image. As a result, halo artifacts may occur around enhanced edges when the edges are blurred in the smoothed image, as illustrated in the top row of Fig. 1. Moreover, structure consistency between target and guidance images is often challenged when the two input streams are in different modalities (e.g., RGB and depth [10]), which can result in texture-copy visual artifacts, as shown in the bottom row of Fig. 1.

Motivated by the remarkable success of GIF and its two limitations, researchers have devoted much attention to this field. A

(Co-first authors: Weimin Yuan, Yinuo Wang; Corresponding author: Cai Meng.)

MMAsia '23 Workshops, December 6–8, 2023, Tainan, Taiwan

ACM ISBN 979-8-4007-0326-3/23/12…$15.00
https://doi.org/10.1145/3611380.3629544

Table 1: Examples of applications based on whether guidance and target images stem from the same modality.

Self-guidance tasks (Single-modality)	Image smoothing [26]
	Detail enhancement [14]
	Structure texture separation [6]
	Noise reduction [20]
Reference-guidance tasks (Cross-modality)	Depth upsampling [10]
	Haze removal [13, 22]
	Stereo matching [6, 17]
	Flash-No-flash denoising [15]

Method	Category	Year	Edge preserved	Structure Consistency	Source	Code
GIF [7]	L	2010	✗	✗	ECCV	✓
WMF [18]	L	2013	✗	✗	ICCV	✓
RGF [29]	L	2014	✓	✗	ECCV	✓
WGIF [14]	L	2014	✓	✗	TIP	✓
GDGIF [9]	L	2015	✓	✗	TIP	✓
FCGIF[2]	L	2015	✓	✗	ICCV	✗
MGIF [15]	L	2017	✓	✗	ICIP	✗
EGIF [17]	L	2018	✓	✗	SPL	✓
AWGIF [4]	L	2018	✓	✗	ICASSP	✓
IGIF [26]	L	2018	✓	✗	ICME	✗
SKWGIF [24]	L	2019	✓	✗	TIP	✓
SGIF [28]	L	2019	✓	✗	CVPR	✓
AnGIF [20]	L	2019	✓	✓	TIP	✓
AWGIF [12]	L	2022	✓	✗	PR	✓
FGMF [19]	L	2023	✓	✗	TIP	✓
MSJF [21]	G	2015	✗	✓	IJCV	✓
SDF [6]	G	2015	✓	✓	TPAMI	✓
MuGIF [5]	G	2017	✓	✓	TPAMI	✓
GGIF [13]	G	2018	✓	✓	TIP	✗
GSF [16]	G	2020	✓	✓	TPAMI	✓
DJF [10]	C	2016	✗	✓	ECCV	✓
DJFR [11]	C	2019	✗	✓	TPAMI	✓
DGF [25]	C	2018	✓	✓	CVPR	✓
SVLRM [3]	C	2019	✓	✓	TPAMI	✓
DKN [8]	C	2020	✓	✓	IJCV	✓
FDKN [8]	C	2020	✓	✓	IJCV	✓
FWM [27]	C	2020	✓	✓	ICML	✗
UnGIF [22]	C	2021	✓	✓	TIP	✓
DAGF [30]	C	2023	✓	✓	TNNLS	✓

Figure 2: A summary of the GIF-like methods. 'L', 'G' and 'C' indicates local, global CNN-based methods, respectively.

range of GIF-like variants, as listed in Table **??**, from model-based to deep learning-based methods, have been proposed. However, few efforts have been made in terms of a comprehensive survey, which can provide the community with a retrospective view of previous methods and a prospective outlook on future work. To the best of our knowledge, only one review [1] is related to this topic. Nonetheless, it only covers model-based GIF-like methods on a specialized task. In this paper, we conduct a comprehensive review and assessment of both existing model and deep learning-based GIF-like methods to evaluate and analyse their performance. In summary, the main contributions of this paper are:

1). To the best of our knowledge, this is the first systematical and holistic survey that focuses on both model and deep learning-based GIF-like methods.
2). We further evaluate 18 representative GIF-like methods on single-modality task image noise reduction and cross-modality task depth upsampling. Our evaluation and analysis reveal the strengths and limitations of current approaches, and bring some observations in this field.

2 RELATED WORKS

An overview of the GIF-like variants is given in Fig. 2.

• **Local methods**. Local methods, such as improved GIF [24], fully connected GIF [2], arbitrary window GIF [4], rolling GIF [27], side window GIF [26] and weighted median filter [17], compute weights from guidance image or different types of affinities among pixels to filter target image. However, the locally linear model [7] fails to preserve the edge sharpness, which is mainly attributed to the fixed regularizer and uniformly weighted aggregation (UWA). Thus, a series of adaptive regularization variants [9, 12, 14, 16, 18] and improved aggregation methods [22] have been proposed. An edge-aware regularizer which incorporates the average of local variances of pixels in the guidance image is proposed in effective GIF [16]. Moreover, the single-scale content-adaptive term is introduced in weighted GIF [12, 14] to avoid edge-blurring. Nevertheless, single-scale term [14] shows an unsatisfactory performance in complex scenarios. Steering kernel GIF [22], an improved version of [14], aims to suppress edge-blurring caused by UWA. By taking the edge direction into consideration, UWA is replaced by a weighted operation where the weights are obtained by steering kernel. Gradient domain GIF [9] extends the single-scale term [14, 22] to multi-scale by calculating the local variances of multiple radii, achieving better halo avoidance performance. The regularization term in anisotropic GIF [18] is formulated as approximate median variance divided by the local variance of guidance image, and UWA is reformulated as a weighted averaging scheme. Despite these efforts, local-based methods often introduce halo and texture-copying artifacts due to the locally linear model.

• **Global methods**. Many scenes, such as RGB/depth [5, 19], flash/no-flash [7], commonly contain inconsistent structures between guidance and target images, such as noise, holes, texture. Consequently, directly transferring inconsistent structural information from guidance to the target image in local-based methods may lead to significant errors, resulting in texture-copying artifacts. To address this problem, recent efforts have focused on considering common structures. These methods have formulated the filtering as a global objective function and have been solved in an iterative manner. For example, mutual-structure joint filtering (MSJF) [19] explores the relationship between the target and guidance images and further defines mutual, inconsistent structures and smooth regions. Static operator [6] convolves the target image with a weight function calculated from only the static guidance image, and the dynamic operator uses the weight function repeatedly calculated from the filtered output image, which is solved by majorize-minimization. Based on relative structure, mutually guided filter (MuGIF) [5] can jointly preserve consistent edges and avoid misleading information from inconsistent structures. By introducing truncated huber penalty, a generalized smoothing framework [15] achieves superior edge-preserving performance in a variety of vision tasks. However, the results are sensitive to the parameters. Inspired by [7, 14], globally GIF [13], consisting of a global structure transfer and an edge-aware smoothing operator, can effectively preserve fine structures and excel in haze removal. Compared to local methods, global methods can avoid the generation of texture-copying artifacts, but often come at the cost of increased computational complexity and the use of hand-crafted priors, which may not accurately represent images in complex scenes.

• **Deep learning-based Methods**. In recent years, convolutional neural networks (CNNs) have enabled significant advances in this field, with promising results reported in [3, 8, 10, 11, 20, 23, 25, 28]. These CNN-based methods can be divided into two categories, one employs filtering kernels constructed based on the guidance image, such as SVLRM [3], UnGIF [20]; and the other utilizes kernels constructed based on both the guidance and target images, such as DJF [10], DJFR [11], DKN [8], FDKN [8], DAGF [28].

SVLRM utilizes structural information from both the guidance and input images to estimate representation coefficients for the restoration of the target image. Motivated by unsharp masking, UnGIF proposes a successive guided filtering network, which explicitly transfers consistent structure by estimating a single coefficient.

A general lightweight deep joint filtering (DJF) consists of three sub-networks. The first two sub-networks act as feature extractors to determine features from target and guidance images. The extracted features are then concatenated and fed into the third sub-network to selectively transfer the consistent structures and reconstruct the filtered output. Subsequently, DJFR is presented by adding a skip connection between input and output images. Deformable kernel network (DKN) employs a similar network architecture as [10], but with an additional weight and offset learning module to learn the neighborhood for each pixel. These methods use a two-stream CNN architecture to extract the features from two different modalities and then combine them through another CNN to achieve the final reconstruction. However, the dependencies modelled by them are quite implicit. DGF [23] turns GIF into a differentiable guided filtering layer, and plugs it into CNN. Factorized weight matrix (FWM) [25] learns the weighted averaging process explicitly by the correlation of deep features. To efficiently utilize multi-modal information while avoiding the limitations of existing methods, DAGF uses an attention mechanism to adaptively fuse multi-modal information through dual regression.

3 EXPERIMENTS AND RESULTS

• **Experiment Settings**. We quantitatively and qualitatively evaluate 18 representative methods in this community. Specifically, 11 model-based methods, including 7 local-based methods (GIF, WGIF, GDGIF, EGIF, SKWGIF, SGIF and AnGIF) and 4 global-based methods (MSJF, SDF, MuGIF and GSF), and 7 CNN-based methods (DJF, DJFR, SVLRM, DKN, FDKN, UnGIF and DAGF) are tested on NYU-v2 [21] for two types of experiments: single-modality RGB image noise reduction (with noise levels 20, 40, and 60) and cross-modality depth upsampling (with upsampling factors 4×, 8× and 16×). Following [8], 1000 image pairs are used for training (for CNN-based methods) and the remainders for testing (for all methods). Default parameters of each method are used, and the average PSNR and RMSE (measured in centimeter) are reported for the noise reduction and depth upsampling, respectively.

• **Quantitative and Qualitative Performance**. As shown in Fig. 4, SVLRM and DAGF achieve top-3 performance on both noise reduction and depth upsampling, suggesting that the vast majority of testing methods can handle only halo or texture-copy artifacts to some extent. In particular, CNN-based methods DAGF, UnGIF and DKN rank top-3 in terms of RMSE on depth upsampling. Nevertheless, CNN-based methods do not completely overwhelm model-based counterparts in noise reduction, such as GSF, which even outperforms the majority of the CNN-based methods under heavy noise (noise levels 40 and 60).

Moreover, we draw the following observations from the visual comparison in Fig. 3. 1). Among traditional methods, to avoid halos caused by local characteristics, a range of adaptive edge recognition methods, such as WGIF, GDGIF, EGIF, SKWGIF, SGIF and AnGIF, produce over-smoothed and halo artifacts. This may be due to the inability of the locally linear model to capture consistent structures and the over-retention of inconsistent structures, especially at large scale factor (16×). 2). Compared to locally linear models, hand-crafted prior based methods such as MSJF, SDF, MuGIF and GSF are better suited for dealing with structural inconsistency, albeit at the expense of increased complexity. Nevertheless, these methods still fail to effectively recover the texture and edges when the hand-crafted priors are violated in some complex scenarios. 3). Compared to model-based counterparts, CNN-based methods can more effectively address halo artifacts around reconstructed edges and avoid texture-copying artifacts, thanks to their powerful learning and mapping capabilities. Specifically, DJF and DJFR combine the features of guidance and target image and then use regression to estimate the results. However, DJF and DJFR are less effective for edge recovered (Figs. 3(m), (n)). With the help of estimated spatially-variant kernels, SVLRM can transfer the common structures and preserve sharp edges as shown in Fig. 3(o). Nevertheless, the estimated kernels in SVLRM are small, which limits it from exploiting the pixels from a larger region. Therefore, the performance of SVLRM on depth upsampling at larger scales (8×, 16×) is significantly lower than that at smaller scale (4×), as shown in Fig. 4. DKN and FDKN show finer edges and more visual pleasant details without introducing texture-copy artifacts, as shown in Fig. 3(p) and Fig. 3(q). The superior performance of DAGF in Fig. 3(s) originates from the attentional kernel learning module and multi-scale strategy.

• **Running Time**. As labelled in Fig. 3, we compare the running time of 18 methods on image of size 640 × 480 using Intel Core i9-9900K with 3.60GHZ CPU and NVIDIA RTX 3090 GPU. Locally linear model-based methods implemented on CPU, such as GIF, WGIF, GDGIF and SGIF, can achieve fast speed, except that SKWGIF introduces a large computational cost. The running time of global-based methods are long due to the optimization process, which limit its application in scenarios where real-time perforamnce is required. The performance of DJF, DJFR, UnGIF and DAGF meet the real-time processing requirement (30 frames per second), thanks to the light-weight feed forward.

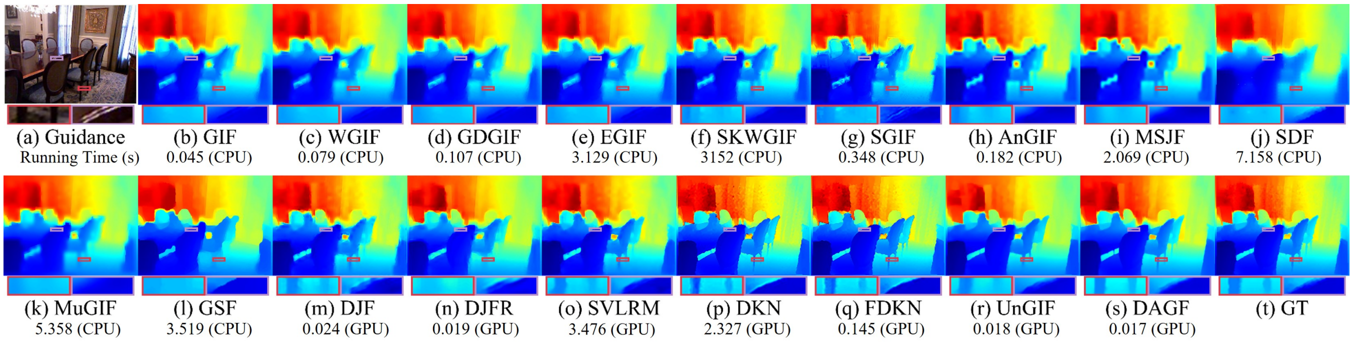

Figure 3: Visual comparison of 18 representative methods on depth upsampling (16×). (a) Guidance image, (b)-(s) Results by 18 representative methods. (t) GT. Please enlarge the PDF for more details.

Method	Noise Reduction			Depth Upsampling		
	PSNR (↑)			RMSE (↓)		
	σ=20	σ=40	σ=60	4×	8×	16×
Input	25.45	19.90	16.78	8.16	14.22	22.32
GIF [7]	27.99	22.81	18.93	7.32	13.62	22.03
WGIF [14]	27.22	22.99	19.15	7.56	13.82	21.03
GDGIF [9]	28.34	21.85	18.04	7.32	13.62	20.31
EGIF [17]	25.74	20.03	16.88	7.31	12.93	20.30
SKWGIF [24]	27.34	23.08	19.40	7.19	12.37	19.54
SGIF [28]	27.90	22.16	18.45	8.37	14.34	22.57
AnGIF [20]	31.10	26.81	23.90	7.37	12.79	19.67
MSJF [21]	28.88	23.52	21.43	7.25	13.06	20.40
SDF [6]	30.44	26.39	23.65	5.27	12.31	19.24
MuGIF [5]	29.40	25.81	22.95	7.51	12.66	19.77
GSF [16]	31.62	**27.49**	**24.42**	6.91	12.86	20.80
DJF [10]	31.38	27.30	**23.91**	3.54	6.20	10.21
DJFR [11]	**31.99**	25.63	20.98	3.38	5.86	10.11
SVLRM [3]	**32.69**	**27.48**	20.17	**1.74**	5.59	10.11
DKN [8]	25.49	20.03	16.94	2.46	**4.76**	**8.50**
FDKN [8]	25.48	20.04	16.95	2.62	4.99	8.67
UnGIF [22]	27.94	26.12	23.86	**2.33**	**3.58**	**6.07**
DAGF [30]	**32.91**	**29.12**	**26.06**	**2.35**	**4.62**	**7.81**

Figure 4: Quantitative comparison of 18 representative methods for noise reduction and depth upsampling on NYU-v2 [21]. The best, second and third results are highlighted in bold.

4 CONCLUSION

In this paper, a systematical survey on GIF-like variants from model-based to CNN-based methods is presented. Then, we conduct experiments to comprehensively evaluate the performance of 18 representative methods. The experiments reveal some valuable insights:
• Local-based methods such as WGIF, GDGIF and SGIF have shown their capability in enhancing edge preservation capability by introducing various adaptive regular terms. However, their transfer of inconsistent structures is weakened when applied to cross-modality tasks. SGIF calculates the filtering results in eight windows simultaneously and selects the one closest to the initial value as result. This idea fundamentally strengthens the edge-preserving ability of GIF, but ignores the suppression of noise and inconsistent information. How to incorporate the idea of side windows to suppress texture-copying is an interesting topic for future exploration.
• Global-based methods such as SDF and GSF, can achieve a good trade-off between single and cross-modality tasks, GSF even surpasses some CNN-based methods (eg., DJF, DJFR, DKN and FDKN) in image noise reduction. But their performance is sensitive to parameters. It is desired to design a complex scenes robust handcrafted prior with higher efficiency for future work.
• CNN-based methods have demonstrated promising performance for both single-modality and cross-modality tasks. However, the unsatisfactory performance of DKN and FDKN on noise reduction can be attributed to their complex matrix operations after deep convolution, resulting in confusion in the self-guidance noise modeling outcomes. For future work, it is worthwhile to adjust the interaction between weights and biases to adapt it to single-modality tasks.

This study provides an overview of the state-of-the-art advances in this domain, aiming to inspire further research into this rapidly developing field.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Grants Nos. 92148206, 61873010).

REFERENCES

[1] Usman Ali, Ik Hyun Lee, and Muhammad Tariq Mahmood. 2021. Guided image filtering in shape-from-focus: A comparative analysis. *Pattern Recognition.* 111 (2021), 107670.
[2] Longquan Dai, Mengke Yuan, Feihu Zhang, and Xiaopeng Zhang. 2015. Fully Connected Guided Image Filtering. In *ICCV*. 352–360.
[3] Jiangxin Dong, Jinshan Pan, Jimmy S. J. Ren, Liang Lin, Jinhui Tang, and Ming-Hsuan Yang. 2021. Learning Spatially Variant Linear Representation Models for Joint Filtering. *IEEE TPAMI* 44 (2021), 8355–8370.
[4] Norishige Fukushima, Kenjiro Sugimoto, and Sei-Ichiro Kamata. 2018. Guided Image Filtering with Arbitrary Window Function. In *ICASSP*. 1523–1527. https://doi.org/10.1109/ICASSP.2018.8462016
[5] Xiaojie Guo, Yu Li, and Jiayi Ma. 2017. Mutually Guided Image Filtering. *IEEE TPAMI* 42 (2017), 694–707.
[6] Bumsub Ham, Minsu Cho, and Jean Ponce. 2018. Robust Guided Image Filtering Using Nonconvex Potentials. *IEEE TPAMI* 40 (2018), 192–207.
[7] Kaiming He, Jian Sun, and Xiaoou Tang. 2010. Guided Image Filtering. In *ECCV*.

[8] Beomjun Kim, Jean Ponce, and Bumsub Ham. 2019. Deformable Kernel Networks for Joint Image Filtering. *IJCV* 129 (2019), 579–600.

[9] Fei Kou, Weihai Chen, Changyun Wen, and Zhengguo Li. 2015. Gradient Domain Guided Image Filtering. *IEEE TIP* 24 (2015), 4528–4539.

[10] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2016. Deep Joint Image Filtering. In *ECCV*.

[11] Yijun Li, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. 2017. Joint Image Filtering with Deep Convolutional Networks. *IEEE TPAMI* 41 (2017), 1909–1923.

[12] Yuwen Li, Zhengguo Li, Chaobing Zheng, and Shiqian Wu. 2022. Adaptive weighted guided image filtering for depth enhancement in shape-from-focus. *PR* 131 (2022), 108900. https://doi.org/10.1016/j.patcog.2022.108900

[13] Zhengguo Li and Jinghong Zheng. 2018. Single Image De-Hazing Using Globally Guided Image Filtering. *IEEE TIP* 27 (2018), 442–450.

[14] Zhengguo Li, Jinghong Zheng, Zijian Zhu, Wei Yao, and Shiqian Wu. 2015. Weighted Guided Image Filtering. *IEEE TIP* 24 (2015), 120–129.

[15] Wei Liu, Pingping Zhang, Yinjie Lei, Xiaolin Huang, Jie Yang, and Michael Ng. 2019. A Generalized Framework for Edge-Preserving and Structure-Preserving Image Smoothing. *IEEE TPAMI* 44 (2019), 6631–6648.

[16] Zongwei Lu, Bangyuan Long, Kang-Ming Li, and Fajin Lu. 2018. Effective Guided Image Filtering for Contrast Enhancement. *IEEE SPL* 25 (2018), 1585–1589.

[17] Ziyang Ma, Kaiming He, Yichen Wei, Jian Sun, and Enhua Wu. 2013. Constant Time Weighted Median Filtering for Stereo Matching and Beyond. In *ICCV*. 49–56.

[18] Carlo Noel Ochotorena and Yukihiko Yamashita. 2020. Anisotropic Guided Filtering. *IEEE TIP* 29 (2020), 1397–1412.

[19] Xiaoyong Shen, Chao Zhou, Li Xu, and Jiaya Jia. 2015. Mutual-Structure for Joint Filtering. *IJCV* 125 (2015), 19–33.

[20] Zenglin Shi, Yunlu Chen, Efstratios Gavves, Pascal Mettes, and Cees G. M. Snoek. 2021. Unsharp Mask Guided Filtering. *IEEE TIP* 30 (2021), 7472–7485.

[21] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. Indoor Segmentation and Support Inference from RGBD Images. In *ECCV*.

[22] Zhonggui Sun, Bo Han, Jie Li, Jin Zhang, and Xinbo Gao. 2020. Weighted Guided Image Filtering With Steering Kernel. *IEEE TIP* 29 (2020), 500–508.

[23] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. 2018. Fast End-to-End Trainable Guided Filter. In *CVPR*. 1838–1847.

[24] Jiafei Wu, Chong Wang, and Yongze Xu. 2018. An Improved Guided Filtering Algorithm for Image Enhancement. In *ICME*. 1–6.

[25] Xiangyu Xu, Yongrui Ma, and Wenxiu Sun. 2020. Learning Factorized Weight Matrix for Joint Filtering. In *ICML*.

[26] Hui Yin, Yuanhao Gong, and Guoping Qiu. 2019. Side Window Filtering. In *CVPR*. 8750–8758. https://doi.org/10.1109/CVPR.2019.00896

[27] Qi Zhang, Xiaoyong Shen, Li Xu, and Jiaya Jia. 2014. Rolling Guidance Filter. In *ECCV*.

[28] Zhiwei Zhong, Xianming Liu, Junjun Jiang, Debin Zhao, and Xiangyang Ji. 2023. Deep Attentional Guided Image Filtering. *IEEE TNNLS* (2023), 1–15.