Deep Bilateral Learning for Stereo Image Super-Resolution

Qingyu Xu, Longguang Wang, Yingqian Wang, Weidong Sheng, and Xinpu Deng

***Abstract*—Bilateral filter has demonstrated its effectiveness in many traditional methods for image restoration tasks. In this letter, we incorporate the idea of bilateral grid processing in a CNN framework and propose a bilateral stereo super-resolution network (BSSRnet). Specifically, we use a parallax-attention module to incorporate information from left and right views to learn content-aware bilateral filters. Then, these bilateral filters are used to recover missing details at different spatial locations while preserving stereo consistency. Our network is fully differentiable and is robust to both content and disparity variations. Comparative results show that our BSSRnet achieves state-of-the-art performance on the Flickr1024, Middlebury, KITTI 2012 and KITTI 2015 datasets. Source code is available at.**

***Index Terms*—Bilateral filter, recursive, stereo image, super-resolution.**

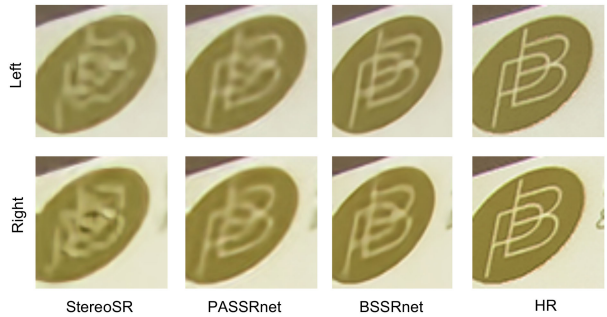

Fig. 1. Visual results achieved by StereoSR [6], PASSRnet [7], and our BSSRnet for 4× SR on "test image 002" of the Flickr1024 dataset [10].

I. Introduction

STEREO image super-resolution (SR) aims at reconstructing high-resolution (HR) images from a pair of low-resolution (LR) images. With rapid development of dual cameras and 3D devices, stereo image super-resolution has recently received increasing attention in the computer vision community.

A straightforward way to achieve stereo image SR is to perform single image SR (SISR) on stereo image pairs separately. Recent years have witnessed the great advances of SISR. Specifically, as one of the seminal work, SRCNN [1] achieved superior performance with a three-layer convolutional neural network. The subsequent SISR methods used advanced CNN architectures to improve SR performance (*e.g.*, residual connection [2], dense connection [3] and multi-scale structure [4]). However, super-resolving a stereo image pair as two separate images cannot fully use the beneficial cross-view information and thus suffers inferior performance [5].

To incorporate information from stereo images, Jeon *et al.* [6] proposed StereoSR to learn a parallax prior by jointly training two cascaded sub-networks. In StereoSR [6], the right image was first shifted with different intervals and then concatenated with the left image to obtain an image volume, which was used to reconstruct an HR left image. Later, Wang *et al.* [7], [8] proposed a parallax-attention mechanism to capture stereo correspondence under large disparity variations. Their parallax-attention mechanism can effectively incorporate information from both sides of views to achieve improved SR performance. Recently, Song *et al.* [9] proposed a SPAMnet to integrate self-attention and parallax-attention mechanisms to incorporate both intra-view and cross-view information.

Since the content of an image varies at different spatial locations, it is important and beneficial to make the network be aware of this variation. However, existing stereo image SR methods use fixed convolution kernels at all spatial locations, which limits their capability to handle different contents. Traditional bilateral grid [11], [12] was proposed for fast edge-aware image processing, and was demonstrated effective in image denoising [13], [14], image deblurring [15], [16], and super-resolution [17]–[19]. These methods can adapt their filters to different local contents. Recently, several efforts have been made to combine bilateral filter with neural networks for image enhancement [20] and image style transfer [21] tasks. In this letter, we propose a bilateral stereo super-resolution network (BSSRnet) for stereo image SR. Our BSSRnet can incorporate stereo correspondence to dynamically generate content-aware bilateral filters for different spatial locations. Consequently, our network is more robust to content variations and can better recover missing details. As shown in Fig. 1, our BSSRnet produces SR results with higher perceptual quality.

In summary, the contributions of this letter are as follows: 1) We propose BSSRnet to learn dynamic bilateral filters for stereo image SR. Our BSSRnet is fully differentiable and is flexible to content variations. 2) Our BSSRnet enforces stereo consistency by incorporating cross-view information to generate bilateral

Manuscript received January 7, 2021; revised March 4, 2021; accepted March 9, 2021. Date of publication March 22, 2021; date of current version April 7, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61921001 and Grant 62001478. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Minglun Gong. *(Corresponding author: Longguang Wang.)*

The authors are with the College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China (e-mail: xuqingyu@nudt.edu.cn; wanglongguang15@nudt.edu.cn; wangyingqian16@nudt.edu.cn; shengweidong111@sohu.com; dengxinpu@nudt.edu.cn).

Digital Object Identifier 10.1109/LSP.2021.3066125

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

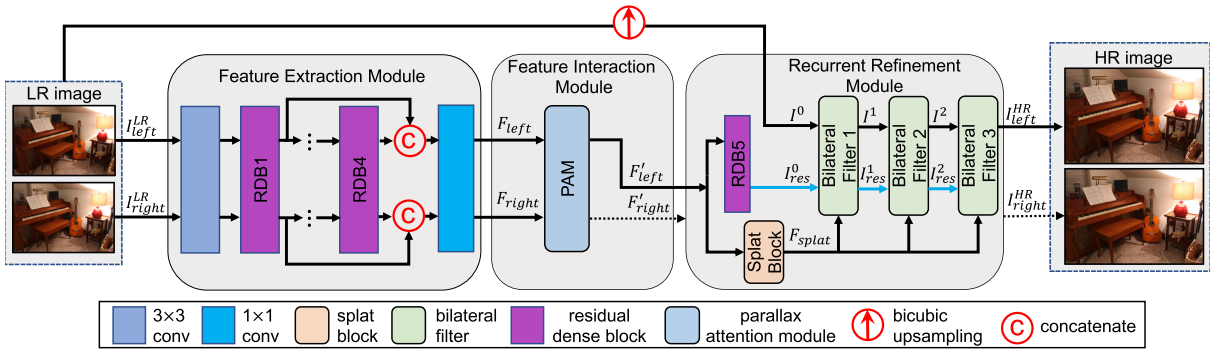

Fig. 2. The network architecture of the proposed BSSRnet framework.

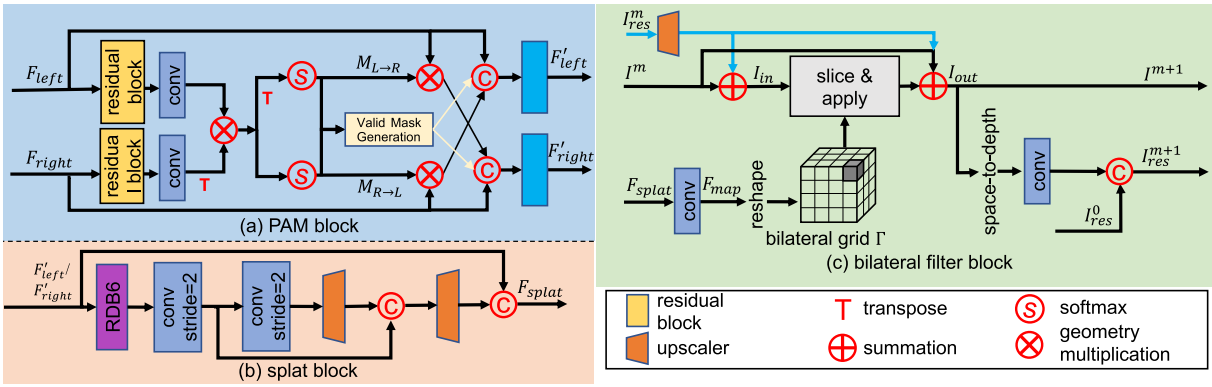

Fig. 3. An illustration of the PAM block, the splat block, and the bilateral filter block.

filters. 3) Extensive experiments demonstrate the state-of-the-art SR performance and high computational efficiency of our method.

II. Methodology

An overview of our BSSRnet is illustrated in Fig. 2. Our network consists of a feature extraction module, a feature interaction module and a recurrent refinement module.

A. Feature Extraction Module

The feature extraction module is used to extract hierarchical features from input stereo images. First, the input image pair I_{left}^{LR} and I_{right}^{LR} are fed to a convolutional layer to generate initial features. Then, four cascaded residual dense blocks (RDB) [22] are used for deep feature extraction. Within each RDB, we use 4 convolutions with a growth rate of 24. After 4 cascaded RDBs, the resultant features are fed to a 1×1 convolutional layer to obtain fused features F_{left} and F_{right}.

B. Feature Interaction Module

After feature extraction, F_{left} and F_{right} are fed to the feature interaction module to achieve cross-view information interaction. Since disparities can vary significantly for stereo cameras with different baselines, focal lengths and resolutions, we use the parallax attention module (PAM) [7], [8] to capture stereo correspondence, as shown in Fig. 3(a). Parallax-attention map $\mathcal{M}_{R \to L}$ is first obtained from the input features. Then, we transpose the right-to-left attention map $\mathcal{M}_{R \to L}$ to produce $\mathcal{M}_{L \to R}$. Consequently, our feature interaction module can interact information for both sides of views simultaneously. Next, valid masks (V_L, V_R) are generated and concatenated with attention maps ($\mathcal{M}_{R \to L}$, $\mathcal{M}_{L \to R}$) and input features (F_{left}, F_{right}) to produce the output features (F'_{left}, F'_{right}).

C. Recurrent Refinement Module

With F'_{left} and F'_{right}, three bilateral filters are used for refinement. Specifically, F'_{left} and F'_{right} are first fed to an RDB to generate residual image I^0_{res}. Meanwhile, these features are passed to the splat block to produce dynamic bilateral filters, as shown in Fig. 3(b). Then, the resultant bilateral filters are used to process the upscaled input image in a recursive manner. Without loss of generality, we take F'_{left} as an example and our refinement module processes the left and right features following the same manner. The details are as follows:

1) Residual Image Generation: Following [23], F'_{left} is first fed to an RDB block to generate a residual image to enhance high frequency details. For the m^{th} bilateral filter block, the residual image $I^m_{res} \in \mathbb{R}^{w \times h \times 64}$ is upscaled before adding to the output of the previous bilateral filter block $I^m \in \mathbb{R}^{w \times h \times 3}$, as shown in Fig. 3(c).

2) Bilateral Grid Generation: Since F'_{left} integrate information from both sides of views, the splat block is further used to learn a multi-scale distribution F_{splat} between stereo images, as shown in Fig. 3(b). Then, we use a 3×3 convolution on F_{splat} for channel fusion. Finally, we reshape the learned $F_{map} \in \mathbb{R}^{w \times h \times 648}$ to encode transformations into an affine

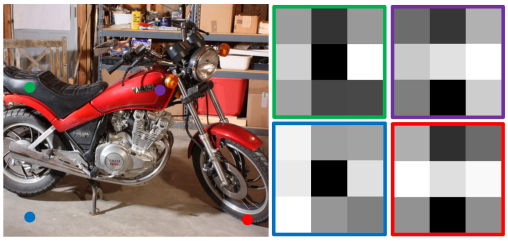

Fig. 4. A visualization of our bilateral filters at four different locations.

TABLE I
RESULTS (×4) ACHIEVED BY BSSRNET WITH DIFFERENT SETTINGS ON THE FLICKR1024 DATASET [10]. THE RUNNING TIME AND WARPING ERROR [9] ARE EVALUATED ON THE MIDDLEBURY DATASET

Model	PSNR	SSIM	Params	Time	Warping Error ($\times 10^{-3}$)
$M_{w/oBF}$	23.27	0.722	2.38M	5.34ms	14.68
$M_{f=1}$	23.32	0.725	1.82M	6.31ms	14.70
$M_{f=2}$	23.35	0.727	1.86M	7.25ms	14.63
$M_{f=3}$	**23.37**	**0.729**	1.91M	8.12ms	**14.62**
$M_{retrained}$	23.34	0.727	1.91M	8.12ms	14.67
PASSRnet	23.21	0.719	1.42M	16.29ms	14.86

bilateral grid $\Gamma \in R^{w\times h\times d\times g}$, where $d = 8$ is the depth of grid and $g = 81$ is the number of parameters in each grid cell.

3) Bilateral Filtering: We refine the image in HR space. First, we upsample the residual feature I^m_{res}, and add it to I^m and I^{m+1}. For each pixel $p(x, y)$ in the compensated image I_{in}, we use the learned 3×3 convolutional layer to obtain value $z = g(p)$ and slice out a kernel $K_c = \Gamma(x/w, y/h, d/z) \in \mathbb{R}^{3\times3\times3\times3}$ using trilinear interpolation. Next, we use K_c to process a 3×3 neighborhood patch $I_{patch} \in \mathbb{R}^{3\times3\times3}$ centered at p and add the result to I^m to produce the output image I_{out}. We visualize four kernels at different locations of an example image in Fig. 4. It can be observed that our bilateral filters can adapt to different local contents.

D. Loss Function

Following [7], the loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_{SR} + \mathcal{L}_{pam} + \mathcal{L}_{stereo}, \tag{1}$$

where $\mathcal{L}_{SR}$ is the SR loss, $\mathcal{L}_{pam}$ is the parallax-attention loss, and $\mathcal{L}_{stereo}$ is the stereo consistency loss.

1) SR Loss: The L_1 loss between the super-resolved and groundtruth stereo images is used as the SR loss.

2) Parallax-Attention Loss: We formulate the parallax-attention loss as a combination of photometric, smoothness and cycle terms. That is, $\mathcal{L}_{pam} = \lambda(\mathcal{L}_{photo} + \mathcal{L}_{smooth} + \mathcal{L}_{cyc})$, where λ is empirically set to 0.1. Please refer to [7] for details.

3) Stereo-Consistency Loss: To enforce stereo consistency in the SR stereo images, we employed the stereo consistency loss in a residual manner [24]. We define the LR residuals Y_L and Y_R between the super-resolved images and groundtruth images as follows:

$$Y_L = |\, I^{HR}_L - I^{SR}_L \,|\downarrow, Y_R = |\, I^{HR}_R - I^{SR}_R \,|\downarrow, \tag{2}$$

where $\downarrow$ represents bicubic downsampling.

TABLE II
COMPARATIVE RESULTS (PSNR/SSIM) ACHIEVED BY BSSRNET USING DIFFERENT RESIDUAL LEARNING MANNERS

Models	KITTI2012	KITTI2015	Middlebury	Flickr1024	Average
$M_{no\ res}$	26.06/0.791	25.17/0.774	28.52/0.820	22.94/0.707	25.67/0.773
$M_{in\ res}$	26.41/**0.801**	25.53/0.786	29.12/0.834	**23.37**/0.724	26.12/**0.788**
$M_{out\ res}$	26.30/0.798	25.42/0.783	29.06/0.834	23.29/0.724	26.01/0.784
$M_{all\ res}$	**26.46/0.801**	**25.59/0.787**	**29.13/0.835**	**23.37/0.727**	**26.14/0.788**

The stereo consistency loss can be formulated as:

$$\begin{aligned}\mathcal{L}_{stereo} = &\parallel V_L \odot (Y_L - \mathcal{M}_{R\to L} \otimes Y_R) \parallel_1 \\ &+\parallel V_R \odot (Y_R - \mathcal{M}_{L\to R} \otimes Y_L) \parallel_1.\end{aligned} \tag{3}$$

III. EXPERIMENTS

A. Datasets and Implementation Details

1) Datasets: During the training phase, we used 800 images from the training set of the Flickr1024 dataset [10] and 60 images from the Middlebury dataset [25]. Following [6][7], we perform bicubic downsampling with a factor of 2 on images from the Middlebury dataset to generate HR images. For test, we followed [6], [7], [9] to use 5 images from the Middlebury dataset, 20 images from the KITTI 2012 [26] dataset, 20 images from the KITTI 2015 [27] dataset, and 112 images from the test set of the Flickr1024 dataset [10].

2) Implementation Details: We first bicubicly downsampled HR images to generate LR images. Then, we cropped 32×92 patches with a stride of 20 from these LR images. Their HR counterparts were also cropped correspondingly. These patches were randomly flipped horizontally and vertically for data augmentation. PSNR and SSIM were used as quantitative metrics in this letter. We use the warping error, which measures the mean square error between SR left and warped SR right images, to evaluate stereo consistency. Our BSSRnet was implemented in Pytorch on a PC with an Nvidia GTX 3090 GPU. All models were optimized using the Adam method with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a batch size of 32. The initial learning rate was set to 2×10^{-4} and halved every 30 epochs. The training was stopped after 120 epochs since more epochs do not provide further consistent improvement.

B. Ablation Study

1) Bilateral Filter: To investigate the effectiveness of the bilateral filter, we first introduce a network variant $M_{w/oBF}$ by replacing the recurrent refinement module with cascaded residual blocks. Then, we introduce 3 variants with different numbers of bilateral filters. The results are shown in Table I. Note that, we increase the number of channels in $M_{w/oBF}$ to make the variant size comparable to BSSRnet. Model $M_{f=1}$ achieves an improvement of 0.05/0.003 in terms of PSNR/SSIM over model $M_{w/oBF}$ with a smaller model size. Furthermore, The performance consistently improves as filter number increases. For example, an improvement of 0.05/0.004 in PSNR/SSIM can be achieved by $M_{f=3}$ as compared to $M_{f=1}$. We set $f = 3$ in our proposed network for a good trade-off between performance and efficiency.

2) Residual Learning Manner: In our experiments, we observed that the residual learning manner in our bilateral filter

TABLE III
QUANTITATIVE RESULTS ACHIEVED BY DIFFERENT METHODS ON THE KITTI 2012, KITTI 2015, MIDDLEBURY AND FLICKR1024 DATASETS. THE RUNNING TIME AND WARPING ERROR [9] ARE EVALUATED ON THE MIDDLEBURY DATASET

Metric	Dataset	Scale	Single Image SR			Stereo Image SR			
			VDSR	LapSRN	DRRN	StereoSR	PASSRnet	SRRes+SAM	Ours
PSNR/SSIM	Middlebury	×2	32.66/0.910	32.75/0.940	33.44/0.932	33.23/0.934	34.05/0.960	-	**34.73/0.947**
	KITTI2012		30.17/0.906	30.10/0.905	30.23/0.909	29.52/0.905	30.65/0.916	-	**30.98/0.922**
	KITTI2015		28.99/0.904	28.97/0.903	29.07/0.907	28.62/0.905	29.78/0.919	-	**30.04/0.925**
	Flickr1024		25.46/0.850	25.47/0.851	27.08/0.876	25.96/0.861	28.28/0.904	-	**28.52/0.909**
	Middlebury	×4	27.73/0.794	28.17/0.809	28.00/0.794	27.79/0.804	28.75/0.824	28.89/0.829	**29.13/0.835**
	KITTI2012		25.61/0.769	26.02/0.785	25.92/0.769	24.57/0.753	26.34/0.794	26.43/0.798	**26.46/0.801**
	KITTI2015		24.73/0.747	25.06/0.766	24.99/0.748	23.72/0.729	25.47/0.779	25.55/0.784	**25.59/0.787**
	Flickr1024		22.43/0.672	22.71/0.690	22.68/0.678	21.68/0.646	23.21/0.719	23.24/0.723	**23.37/0.727**
Warping Error ($\times 10^{-3}$)	Middlebury	×4	15.004	15.113	14.987	15.634	14.869	14.803	**14.623**
Running Time			33.07ms	31.02ms	102.67ms	184.31ms	16.29ms	25.49ms	**8.12ms**

Note: We do not present 2×SR results of SRRes+SAM since their models are unavailable.

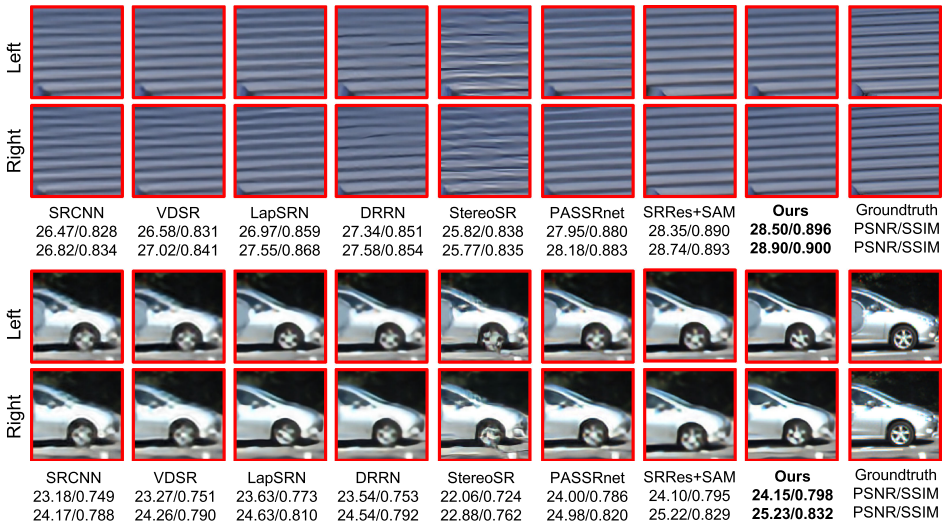

Fig. 5. Visual comparison for 4× SR. These results are achieved on "test image 089" of the Flickr1024 dataset [10] and "test image 007" of the KITTI 2015 dataset [27].

block is crucial to the SR performance. We first introduced a variant M_{nores} without residual connections. Then, we further proposed three variants by adding the residual image to the input image (M_{inres}), the output image (M_{outres}) and both of them (M_{allres}), respectively. Table II shows the quantitative results of different variants. Compared to other model variants, M_{allres} achieves the highest PSNR and SSIM values on all the four datasets, which demonstrates the effectiveness of our proposed residual learning manner.

3) Stereo Consistency Loss: We retrained BSSRnet without stereo consistency loss to validate its effectiveness. The results in Table I demonstrate that the loss function can facilitate our network to recover stereo consistent details in the SR results.

C. Comparison to the State-of-The-Arts

In this section, we compare our BSSRnet with 3 SISR methods (*i.e.*, *VDSR* [28], *LapSRN* [29], *DRRN* [30]) and 3 stereo image SR methods (*i.e.*, *StereoSR* [6], *PASSRnet* [7], *SRRes+SAM* [5]). For fair comparison, we use their officially released codes to generate SR results for evaluation. Quantitative results are listed in Table III. It can be observed that our network produces much better results than other SISR methods [28][29] [30] and stereo image SR methods [6][7] [5] with superior efficiency. Specifically, our BSSRnet outperforms PASSRnet by 0.38 dB on Middlebury with shorter running time (8.12 ms vs 16.29 ms). Moreover, our method achieves the lowest warping error, which demonstrates that the SR stereo images obtained by our method are more stereo-consistent.

Qualitative results are shown in Fig. 5. It can be observed from the zoom-in regions that the compared methods cannot generate consistent fine details of stereo pairs. In contrast, our method produces results with better visual quality and fewer artifacts.

IV. CONCLUSION

In this letter, we have proposed a bilateral learning network for stereo image super-resolution. Specifically, our network integrates information from both left and right images to learn content-aware bilateral filters. Our bilateral filter can recover missing details based on local content while preserving stereo consistency. Experimental results have demonstrated that our BSSRnet can effectively use the cross-view information for stereo image SR and achieve the state-of-the-art SR performance.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[2] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.

[3] R. Wen, K. Fu, H. Sun, X. Sun, and L. Wang, "Image superresolution using densely connected residual networks," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1565–1569, Oct. 2018.

[4] Y. Yang, D. Zhang, S. Huang, and J. Wu, "Multilevel and multiscale network for single-image super-resolution," *IEEE Signal Process. Lett.*, vol. 26, no. 12, pp. 1877–1881, Dec. 2019.

[5] X. Ying, Y. Wang, L. Wang, W. Sheng, W. An, and Y. Guo, "A stereo attention module for stereo image super-resolution," *IEEE Signal Process. Lett.*, vol. 27, pp. 496–500, 2020.

[6] D. S. Jeon, S.-H. Baek, I. Choi, and M. H. Kim, "Enhancing the spatial resolution of stereo images using a parallax prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1721–1730.

[7] L. Wang *et al.*, "Learning parallax attention for stereo image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12 250–12259.

[8] L. Wang *et al.*, "Parallax attention for unsupervised stereo correspondence learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Sep. 25, 2020, doi: 10.1109/TPAMI.2020.3026899.

[9] W. Song, S. Choi, S. Jeong, and K. Sohn, "Stereoscopic image super-resolution with stereo consistent feature." in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 12031–12038.

[10] Y. Wang, L. Wang, J. Yang, W. An, and Y. Guo, "Flickr1024: A large-scale dataset for stereo image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 3852–3857.

[11] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 568–580.

[12] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 103–es, 2007.

[13] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[14] M. Zhang and B. K. Gunturk, "Multiresolution bilateral filtering for image denoising," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2324–2333, Dec. 2008.

[15] A. Gupta and R. Kumar, "Design and analysis of an algorithm for image deblurring using bilateral filer," *Int. J. of Sci. Emerg. Trends Latest Trends*, vol. 5, no. 1, pp. 28–34, 2013.

[16] J. Iyer, E. Chitra, V. Maik, S. Padhi, S. Gupta, and S. Honawad, "Image deblurring and super resolution using bilateral filter and sparse representation," *Mater. Today: Proc.*, vol. 33, pp. 3922–3929, 2020.

[17] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 428–438, Jan. 2016.

[18] A. Laghrib, A. Hakim, and S. Raghay, "A combined total variation and bilateral filter approach for image robust super resolution," *EURASIP J. Image Video Process.*, vol. 2015, no. 1, pp. 1–10, 2015.

[19] J. T. Barron and B. Poole, "The fast bilateral solver," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 617–632.

[20] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, 2017.

[21] X. Xia *et al.*, "Joint bilateral learning for real-time universal photorealistic style transfer," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 327–342.

[22] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.

[23] Y. Jo, S. Wug Oh, J. Kang, and S. Joo Kim, "Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3224–3232.

[24] Y. Wang, X. Ying, L. Wang, J. Yang, W. An, and Y. Guo, "Symmetric parallax attention for stereo image super-resolution," 2020, *arXiv:2011.03802*.

[25] D. Scharstein *et al.*, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Proc. German Conf. Pattern Recognit.*, 2014, pp. 31–42.

[26] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.

[27] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3061–3070.

[28] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[29] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 624–632.

[30] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.