Motion-adaptive Separable Collaborative Filters for Blind Motion Deblurring

Supplementary Material

In this supplementary material, Sec. A1 provides the detailed explanations of our proposed MISC filter. Sec. A2 illustrates the detailed architecture of the motion estimation and residual reconstruction networks. Sec. A3 analyzes the limitations of our method. Finally, Sec. A4 shows more quantitative and qualitative comparison results.

A1. Detailed explanations

As described in Sec. 3.2 of the main paper, we propose a motion-guided alignment (MGA) module for aligning motion-induced blurring to the motion middle along the estimated flow direction and a separable collaborative filtering (SCF) module for predicting the parameters to filter the aligned image as output. In this section, we elaborate on the theoretical relationship between its filtering process and motion deblurring.

A1.1. Motion-guided alignment

Motion blur often occurs by the object displacement in an extremely short period. To generate a sharper image, we propose to predict the motion field that generates blurring by a flow estimator and use bi-directional warping to align the blur to the middle moment along the motion direction.

First, MGA predicts the motion field that produces blur by a flow estimator and utilizes bi-directional warping to align the blur to the middle moment along the motion direction. This practice extends the range of handling blur and mitigates the blur induced by fast motion between frames to some extent. At the same time, bi-directional alignment can aggregate the ghosting of moving objects and sharpen texture details in the image.

Second, to avoid the problem of pixel occlusion [3, 29] in different directions during the bi-directional warping, we incorporate a mask estimator to generate mask as a modulation mechanism to optimize bi-directional pixel synthesis. This practice of introducing a mask is intended to provide a second selection mechanism of pixels during the warping process to avoid generating incorrect textures.

A1.2. Separable collaborative filtering.

To alleviate the limitations of multiple degrees of freedom in filter parameter settings when capturing complex motions, we collaboratively obtain the filter parameters by using kernel, weight, and offset estimators. Finally, the blur is removed from the image by the filtering algorithm. This module aims to find reference pixels with blurred regions around the pixel to be reconstructed, thus recovering a sharper image. Unlike existing filtering algorithms that rely on pre-defined parameters, the predicted parameters (*i.e.*, kernel, weight, and offset) in our method are strongly correlated with the content of the input image and thus have a stronger motion capture capability.

Among these parameters, "kernel" is the initial weight of the filter, which is generated adaptively depending on the content of the blur in different regions. "Offset" is the vector direction that induces the motion of the local blur, and it locates the boundary of the blur around the pixel to be reconstructed and captures the blur's shape. "Weight" is similar to the attention mechanism and is used to weigh the contents of all reference pixels. This implies a quadratic adjustment of kernel content aggregation, increasing the nonlinear representation of the model.

A2. Architecture Details

As described in Sec. 3 of the main paper, our method introduces an additional motion estimation network similar to the residual reconstruction network for estimating the filter parameters. To ensure the non-local motion capture capability of the motion estimation network and the reconstruction capability of the residual reconstruction network, we use the same U-Net structure, frequency-domain learning scheme, and loss function as prior works [40].

In this section, we illustrate the detailed architecture of the lightweight U-Net as shown in Fig. A1. The entire network is based on the U-Net architecture. We follow the existing approach [40] to extract features by stacking some Res FFT-Conv Blocks on each scale. Each block contains a frequency-domain learning branch that learns non-local features in the frequency domain after the Fourier transform by 1×1 convolution layers without extending the receptive field of the block.

A3. Limitation

Although our MISC Filter can handle the complex motion of each region well, the number of filter parameters that need to be predicted increases when the image resolution is larger. Therefore, it is expected to build a filter bank to fix the number of operators so as to avoid increasing computational costs significantly.

A4. More Results

In this section, first, we provide more validation of the model generalization. Secondly, the performance of the proposed MISC filter on different residual reconstruction models is compared. Thirdly, we construct ablations for the residual reconstruction network. Finally, we show more

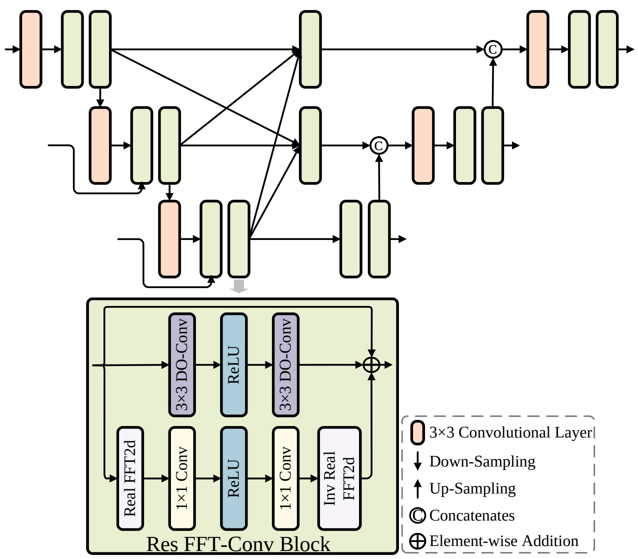

Figure A1. Structures of the motion estimation and residual reconstruction networks.

visualization results.

Validation of model generalizability. We further validate the generalization ability of our method on additional RealBlur-R [48] and RealBlur-J [48] datasets using GoPro trained models. As shown in Tab. A1, since we remove motion blur directly in the image space, our method achieves comparable performance to UFPNet [19] with fewer number of parameters. This superiority proves the generalization ability of our method.

Comparison on different residual reconstruction networks. As described in Sec. 3.2 of the main paper, our MISC filter can also be used for other methods in image space as a plug-and-play module. Therefore, we plug the proposed MISC filter into three state-of-the-art residual reconstruction networks [7, 12, 40] to verify its generalization. As shown in Tab. A2, our MISC filter delivers performance improvements of more than 0.3 dB on all three state-of-the-art deblurring models. This proves the generalizability of our method.

Ablation about residual reconstruction network. Both filtering in image space and residual reconstruction in feature space are important factors in restoring sharp images. In this section, we conduct ablation experiments on the residual reconstruction network to demonstrate the importance of the motion estimation network we used. As shown in Tab. A3, even without using the residual reconstruction network, our method can achieve a PSNR of 31.87 just by filtering the image space. When adding MISC filtering to the residual reconstruction network, the performance improves by 0.4 dB, demonstrating the filter's strong ability to remove motion blur.

Method	#P(M)	RealBlur-R	RealBlur-J
		PSNR/SSIM	PSNR/SSIM
SRN [56]	6.8	35.66/0.947	28.56/0.867
DeblurGAN [27]	-	33.79/0.903	27.97/0.834
DMPHN [67]	21.7	35.70/0.948	28.42/0.860
DeblurGANv2 [28]	60.9	35.26/ 0.944	28.70/ 0.866
MPRNet [65]	20.1	35.99/0.952	28.70/ 0.873
MIMO-Unet+ [12]	16.1	35.54/0.947	27.63/0.837
MAXIM [59]	-	35.78/0.947	28.83/0.875
Uformer-B [62]	50.9	36.19/0.956	29.09/0.886
Restormer [66]	26.1	36.19/0.957	28.96/0.879
MSDI-Net [30]	135.4	35.88/0.952	28.59/0.869
Stripformer [57]	26.1	36.07/0.952	28.82/0.876
NAFNet [7]	67.9	35.50/0.953	28.32/0.857
DeepRFT+ [40]	23	35.86/0.950	28.97/0.884
UFPNet [19]	80.3	36.25/0.953	29.87/0.884
MISC Filter(Ours)	16.0	36.26/0.957	29.35/0.886

Table A1. Quantitative comparison on the RealBlur-R [48], and RealBlur-J [48] dataset. We use the models trained on the GoPro [41] dataset. #P indicates the parametersy. Red indicates the best and blue indicates the second best performance (best viewed in color).

Method	GoPro	
	PSNR	SSIM
MIMO-Unet [12]	31.73	0.951
MIMO-Unet+MISC Filter	32.13	0.954
NAFNet [7]	31.79	0.951
NAFNet+MISC Filter	32.09	0.954
DeepRFT [40]	32.40	0.955
DeepRFT+MISC Filter	32.83	0.960

Table A2. Quantitative comparison on different residual reconstruction networks.

RCN	MEN	PSNR	SSIM
✓		32.40	0.955
	✓	31.87	0.945
✓	✓	**32.83**	**0.960**

Table A3. Results of ablation studies on residual reconstruction network. RCN: residual reconstruction network. MEN: motion estimation network. our MISC Filter can be interpreted as "RCN+MEN".

More Visualization Results. To further verify the effectiveness of our method, we show more comparison results among the proposed MISC filter and other advanced methods on three different benchmarks. The results on **RealBlur-J** [48], **GoPro** [41], and **HIDE** [50] are shown in Fig. A2, Fig. A3, and Fig. A4, respectively.

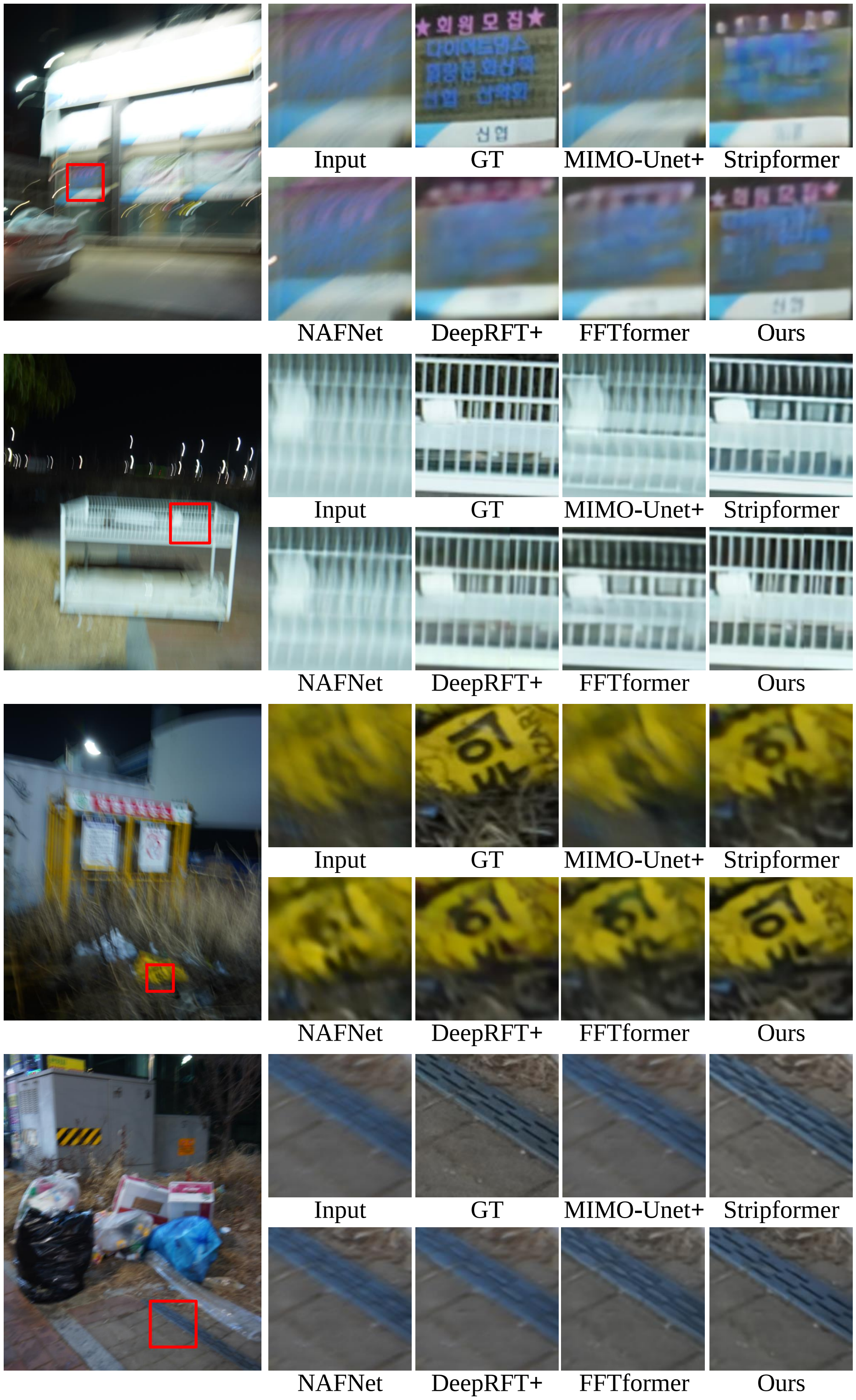

Figure A2. Visual results on RealBlur-J [48] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

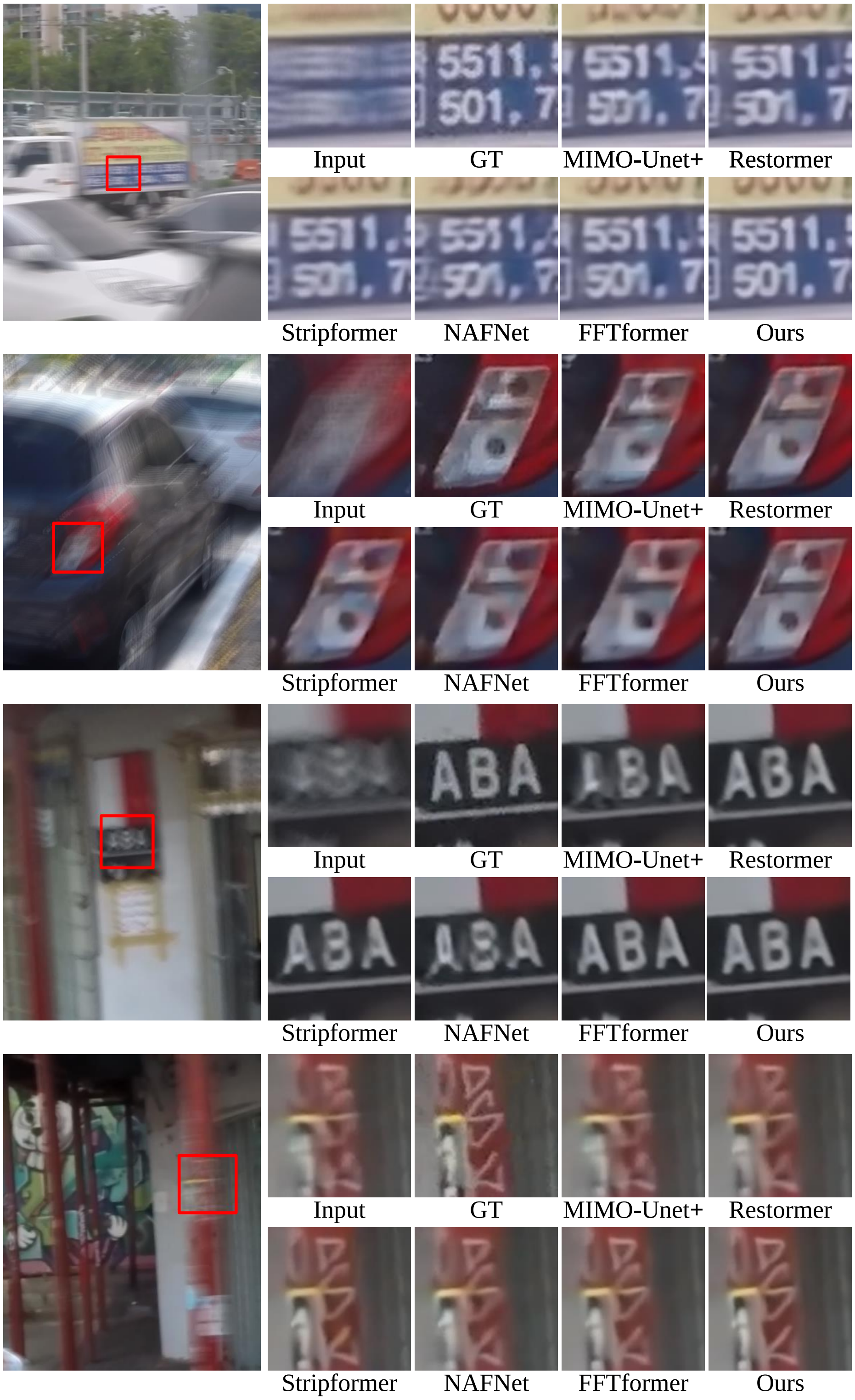

Figure A3. Visual results on GoPro [41] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

Figure A4. Visual results on HIDE [50] dataset. The method is shown at the bottom of each case. Zoom in to see better visualization.

References

[1] Suyash P Awate and Ross T Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE TPAMI*, 28(3):364–376, 2006. 2

[2] Yuval Bahat, Netalee Efrat, and Michal Irani. Non-uniform blind deblurring by reblurring. In *ICCV*, pages 3286–3294, 2017. 1, 2

[3] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-aware video frame interpolation. In *CVPR*, pages 3703–3712, 2019. 4, 1

[4] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *ECCV*, pages 221–235. Springer, 2016. 1, 2

[5] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. BasicVSR: The search for essential components in video super-resolution and beyond. In *CVPR*, pages 4947–4956, 2021. 4

[6] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *CVPRW*, pages 182–192, 2021. 2

[7] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pages 17–33. Springer, 2022. 1, 2, 5, 6

[8] Zheng Chen, Yulun Zhang, Ding Liu, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Hierarchical integration diffusion model for realistic image deblurring. *NeurIPS*, 36, 2024. 1, 2, 5

[9] Xianhang Cheng and Zhenzhong Chen. Video frame interpolation via deformable separable convolution. In *AAAI*, pages 10607–10614, 2020. 3, 4

[10] Xianhang Cheng and Zhenzhong Chen. Multiple video frame interpolation via enhanced deformable separable convolution. *IEEE TPAMI*, 44(10):7029–7045, 2021. 2, 3, 8

[11] Zhixiang Chi, Yang Wang, Yuanhao Yu, and Jin Tang. Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In *CVPR*, pages 9137–9146, 2021. 2

[12] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, pages 4641–4650, 2021. 1, 2, 5, 6

[13] Xiaojie Chu, Liangyu Chen, Chengpeng Chen, and Xin Lu. Improving image restoration by revisiting global information aggregation. In *ECCV*, pages 53–71. Springer, 2022. 5, 6

[14] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *ICCV*, pages 13001–13011, 2023. 2

[15] Yuning Cui, Yi Tao, Wenqi Ren, and Alois Knoll. Dual-domain attention for image deblurring. In *AAAI*, pages 479–487, 2023. 2

[16] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *ICCV*, pages 764–773, 2017. 2, 3, 8

[17] Aram Danielyan, Vladimir Katkovnik, and Karen Egiazarian. Bm3d frames and variational image deblurring. *IEEE TIP*, 21(4):1715–1728, 2011. 2

[18] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. In *ICCV*, pages 2758–2766, 2015. 4

[19] Zhenxuan Fang, Fangfang Wu, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Self-supervised non-uniform kernel estimation with flow-based motion prior for blind image deblurring. In *CVPR*, pages 18105–18114, 2023. 1, 2, 3, 5, 6

[20] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, pages 9935–9946, 2023. 2

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 2

[22] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *CVPR*, pages 2462–2470, 2017. 4

[23] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *NeurIPS*, 28, 2015. 4

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[25] Jaihyun Koh, Jangho Lee, and Sungroh Yoon. BNUDC: A two-branched deep neural network for restoring images from under-display cameras. In *CVPR*, pages 1950–1959, 2022. 8

[26] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, pages 5886–5895, 2023. 1, 2, 5, 6

[27] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *CVPR*, pages 8183–8192, 2018. 1, 2

[28] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, pages 8878–8887, 2019. 1, 2, 5, 6

[29] Hyeongmin Lee, Taeoh Kim, Tae-young Chung, Daehyun Pak, Yuseok Ban, and Sangyoun Lee. Adacof: Adaptive collaboration of flows for video frame interpolation. In *CVPR*, pages 5316–5325, 2020. 2, 3, 1

[30] Dasong Li, Yi Zhang, Ka Chun Cheung, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Learning degradation representations for image deblurring. In *ECCV*, pages 736–753. Springer, 2022. 5, 6, 2

[31] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *CVPR*, pages 18278–18289, 2023. 2

[32] Chengxu Liu, Huan Yang, Jianlong Fu, and Xueming Qian. Learning trajectory-aware transformer for video super-resolution. In *CVPR*, pages 5687–5696, 2022. 4

[33] Chengxu Liu, Xuan Wang, Shuai Li, Yuzhi Wang, and Xueming Qian. FSI: Frequency and spatial interactive learning for image restoration in under-display cameras. In *ICCV*, pages 12537–12546, 2023. 1, 8

[34] Chengxu Liu, Huan Yang, Jianlong Fu, and Xueming Qian. 4D LUT: learnable context-aware 4d lookup table for image enhancement. *IEEE TIP*, 32:4742–4756, 2023. 2

[35] Chengxu Liu, Huan Yang, Jianlong Fu, and Xueming Qian. TTVFI: Learning trajectory-aware transformer for video frame interpolation. *IEEE TIP*, 32:4728–4741, 2023. 4

[36] Chengxu Liu, Xuan Wang, Yuanting Fan Fan, Shuai Li, and Xueming Qian. Decoupling degradations with recurrent network for video restoration in under-display camera. In *AAAI*, 2024. 2

[37] Xina Liu, Jinfan Hu, Xiangyu Chen, and Chao Dong. Udc-unet: Under-display camera image restoration via u-shape dynamic network. In *ECCVW*, pages 113–129. Springer, 2022. 8

[38] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 5

[39] Guo Lu, Wanli Ouyang, Dong Xu, Xiaoyun Zhang, Zhiyong Gao, and Ming-Ting Sun. Deep kalman filtering network for video compression artifact reduction. In *ECCV*, pages 568–584, 2018. 2, 3

[40] Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *AAAI*, pages 1905–1913, 2023. 1, 2, 5, 6

[41] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 3883–3891, 2017. 5, 6, 7, 2, 3

[42] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *ICCV*, pages 261–270, 2017. 4

[43] Hrishikesh Panikkasseril Sethumadhavan, Densen Puthussery, Melvin Kuriakose, and Jiji Charangatt Victor. Transform domain pyramidal dilated convolution networks for restoration of under display camera images. In *ECCVW*, pages 364–378. Springer, 2020. 8

[44] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *ECCV*, pages 327–343. Springer, 2020. 2

[45] Dongwon Park, Byung Hyun Lee, and Se Young Chun. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *CVPR*, pages 5815–5824. IEEE, 2023. 2, 3

[46] Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In *ICCV*, pages 2309–2319, 2021. 2

[47] Roberto Rigamonti, Amos Sironi, Vincent Lepetit, and Pascal Fua. Learning separable filters. In *CVPR*, pages 2754–2761, 2013. 4

[48] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, pages 184–201. Springer, 2020. 3, 5, 6, 7, 2, 4

[49] Christian J Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to deblur. *IEEE TPAMI*, 38 (7):1439–1451, 2015. 1, 2

[50] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *ICCV*, pages 5572–5581, 2019. 5, 6, 7, 3

[51] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *CVPR*, pages 11166–11175, 2019. 2, 3, 8

[52] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *CVPR*, pages 3606–3615, 2020. 6

[53] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *CVPR*, pages 8934–8943, 2018. 4

[54] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, pages 769–777, 2015. 1, 2

[55] Varun Sundar, Sumanth Hegde, Divya Kothandaraman, and Kaushik Mitra. Deep atrous guided filter for image restoration in under display cameras. In *ECCVW*, pages 379–397. Springer, 2020. 8

[56] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, pages 8174–8182, 2018. 5, 6, 2

[57] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *ECCV*, pages 146–162. Springer, 2022. 1, 2, 5, 6

[58] Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Yen-Yu Lin, and Chia-Wen Lin. BANet: a blur-aware attention network for dynamic scene deblurring. *IEEE TIP*, 31:6789–6799, 2022. 5

[59] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. MAXIM: Multi-axis mlp for image processing. In *CVPR*, pages 5769–5780, 2022. 1, 2, 5, 6

[60] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPRW*, pages 0–0, 2019. 2

[61] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 5

[62] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 1, 2, 5, 6

[63] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, pages 13095–13105, 2023. 2

[64] Qirui Yang, Yihao Liu, Jigang Tang, and Tao Ku. Residual and dense unet for under-display camera restoration. In *ECCVW*, pages 398–408. Springer, 2021. 8

[65] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021. 1, 2, 5, 6

[66] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 1, 2, 5, 6

[67] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *CVPR*, pages 5978–5986, 2019. 2

[68] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *CVPR*, pages 2737–2746, 2020. 2

[69] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *CVPR*, pages 9179–9188, 2021. 8

[70] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable convnets v2: More deformable, better results. In *CVPR*, pages 9308–9316, 2019. 2