Supplementary Materials for "When Fast Fourier Transform Meets Transformer for Image Restoration"

Xingyu Jiang, Xiuhui Zhang, Ning Gao, and Yue Deng[⋆]

School of Astronautics, Beihang University, Beijing, China

Table 1: Details of the datasets for image restoration tasks.

Task	Dataset	Train Number	Test Number	Testset Rename
Dehazing	RESIDE-ITS [29]	13990	500	SOTS-Indoor
	RESIDE-OTS [29]	313950	500	SOTS-Outdoor
	O-HAZE [7]	35	5	0-HAZE
	NH-HAZE [6]	45	5	NH-HAZE
	DENSE-HAZE [5]	45	5	DENSE-HAZE
Deraining	Rain200H [58]	1800	200	Rain200H
	Rain200L [58]	1800	200	Rain200L
	DID-Data [64]	12000	1200	DID-Data
	DDN-Data [20]	12600	1400	DDN-Data
	SPA-Data [52]	638492	1000	SPA-Data
	Raindrop [37]	861	58	Raindrop-A
	Raindrop [37]	0	239	Raindrop-B
Motion Deblurring	GoPro [35]	2103	1111	GoPro
	HIDE [43]	0	2025	HIDE
	RealBlur-R [42]	0	980	RealBlur-R
	RealBlur-J [42]	0	980	RealBlur-J
Defocus Deblurring	DPDD [2]	350	76	DPDD
Desnowing	CSD [13]	8000	2000	CSD(2000)
	SRRS [12]	15000	15000	SRRS(2000)
	Snow100K [32]	50000	50000	Snow100K(2000)
Underwater Enhancement	UIEB [30]	750	90	U-90
	LSUI [36]	3500	400	L-400
Low-light Enhancement	LOL-v1 [54]	485	15	LOL-v1
	LOL-v2-real [59]	689	100	LOL-v2-real
	LOL-v2-syn [59]	900	100	LOL-v2-syn
	MIT-Adobe FiveK [9]	4500	500	FiveK
Denoising	SIDD [1]	320	1280	SIDD
Super-Resolution	DIV2K [4]	800	0	-
	Set5 [8]	0	5	Set5
	Set14 [63]	0	14	Set14
	B100 [33]	0	100	B100
	Urban100 [22]	0	100	Urban100
	Manga109 [34]	0	109	Manga109

[⋆] Corresponding author

1 Datasets and Experimental Details

In tab.1, we list the datasets used for training and evaluation. Next, we describe them for each individual task.

1.1 Image Dehazing

We performing dehazing experiments both on the synthetic benchmark RESIDE [29] and real-world hazy dataset Dense-Haze [5], O-Haze [7] and NH-Haze [6]. For RESIDE, we train our model for the indoor and outdoor scenarios separately, and then test on the corresponding SOTS dataset. For the indoor experiment, ITS contains 13990 hazy/clear pairs for training and the SOTS-indoor contains 500 hazy/clear pairs for testing. For the outdoor experiment, OTS contains 313,950 hazy/clear pairs for training and the SOTS-outdoor contains 500 hazy/clear pairs for testing. SFHformer is trained for 600k steps both on ITS and OTS with a batch size of 24. O-HAZE, NH-HAZE and DENSE-HAZE are high-resolution real-world datasets. O-Haze consists 35 training images, 5 validation images and 5 test images. NH-HAZE is a non-homogeneous hazy dataset containing 45 training images, 5 validation images and 5 test images. DENSE-HAZE is a dense hazy dataset consists of 45 training images, 5 validation images, and 5 test images. For above three datasets, SFHformer is trained for 10k steps with the patch size of 800 × 800.

1.2 Image Deraining

Following previous work [14], we compare PSNR/SSIM on the Y channel in YCbCr color space. We perform the experiments on Rain200H [58], Rain200L [58], DID-Data [64], DDN-Data [20] and SPA-Data [52]datasets. Rain200H and Rain200L meanwhile contains 1800 synthetic rainy/clear image pairs for training and 200 ones for testing. DID-Data and DDN-Data respectively consist of 12000 and 12600 synthetic rainy/clear pairs varying in rain directions and density levels, with testing set of 1200 and 1400 pairs. SPA-Data is a large-scale real-world dataset containing 638492 rainy/clear pairs for training and 1000 ones for testing. SFHformer is trained for 300k steps.

1.3 Image Motion Deblurring

We evaluate SFHformer on GoPro [35], HIDE [43] and RealBlur [42] for single-image motion deblurring, following recent methods [18,61]. Gopro dataset contains 2103 blurry/clear training pairs and 1111 blurry/clear testing ones, which is obtained by a high-speed camera. To assess the robustness and generalizability of our approach, we conduct an evaluation by deploying the model trained on GoPro dataset directly onto the HIDE and RealBlur dataset. The HIDE dataset comprises 2025 pairs of blurry/clear images specifically curated for evaluation. The blurry images in both the GoPro and HIDE datasets are synthetically generated. As a complement, RealBlur dataset [42] is adopted, in which the blurry-sharp image pairs are acquired in real-world conditions. The RealBlur dataset has two subsets: (1)RealBlur-J contains 980 image pairs obtained directly as camera JPEG outputs, and (2)RealBlur-R is generated offline by applying white balance, demosaicking and denoising operations to the RAW images, which also has 980 images. SFHformer is trained for 600k steps with a batch size of 24.

1.4 Image Defocus Deblurring

We compare our proposed SFHformer with state-of-the-art approaches on DPDD [2] dataset for single-image defocus deblurring. DPDD comprises images spanning 500 distinct indoor and outdoor scenes, with each scene containing four images: right-view, left-view, center-view, and an associated all-in-focus ground truth image. Specifically, DPDD is partitioned into training, validation, and testing sets, containing 350, 74, and 76 scenes (37 indoor and 39 outdoor), respectively. In our work, we conduct SFHformer on the single-image defocus deblurring task, which involves training on the center-view image along with its corresponding ground truth. SFHformer is trained on DPDD for 150k steps.

1.5 Image Desnowing

We compare our method on CSD [13], SRRS [12], and snow100K [32] dataset with existing state-of-the-art methods for image desnowing. CSD is a large-scale snow dataset consisting of 8000 synthesized snow images. SRRS contains 15000 synthesized snow images and Snow100K has 100k synthesized snowy images. The dataset settings follow previous works, where we randomly sample 2500 image pairs from the training set for training, and 2000 images from testing set for evaluation. SFHformer is trained for 150k steps on each dataset.

1.6 Image Raindrop Removal

Following previous work [50], We perform experiments on Raindrop [37] for image raindrop removal. The Raindrop dataset contains 861 raindrop/clear pairs for training and, 58 ones of testset A and 239 ones of testset B for evaluation, respectively. SFHformer is trained for 60k steps.

1.7 Underwater Image Enhancement

We compare our method on UIEB [30] and LSUI [36] datasets with existing state-of-the-art methods for underwater image enhancement. The UIEB dataset contains 890 real underwater images with corresponding ground truths. We randomly selected 750 pairs for training, 50 pairs for validation, and 90 pairs for testing (U-90). LSUI, which builds in a similar method to UIEB but its scale is larger, contains 4279 image pairs. We randomly selected 3500 pairs for training, 379 pairs for validation, and 400 pairs for testing (L-400). SFHformer is trained for 40k and 120k steps on UIEB and LSUI, respectively.

1.8 Low-light Image Enhancement

We evaluate SFHformer on LOL-v1 [54], LOL-v2 [59] and FiveK [9] for low-light image enhancement, following recent methods [10]. The LOL dataset comprises versions v1 and v2. LOL-v2 is further categorized into real and synthetic subsets. The division of training and testing sets follows a ratio of 485:15, 689:100, and 900:100 for LOL-v1, LOL-v2-real, and LOL-v2-synthetic, respectively. The MIT-Adobe FiveK dataset, denoted as FiveK, is partitioned into training and testing sets, comprising 4500 and 500 pairs of low-/normal-light images, respectively. These images undergo manual adjustments by five photographers, labeled as A to E. The reference images used in our study are those adjusted by expert C, and the adopted output mode is sRGB. SFHformer is trained for 20k steps both on LOL-v1 and LOL-v2 with a patch size of 128 × 128. As for FiveK, SFHformer is trained for 150k steps.

1.9 Image Denoising

Following [61], we train our SFHformer on the SIDD dataset [1] for image denoising, which has 320 high-resolution images. With the SIDD-trained model, we evaluate our SFHformer on 1280 patches from the SIDD validation set [1].

1.10 Efficient Image Super-resolution

Following [15,27,48], we evaluate SFHformer on the DIV2K dataset [4] for efficient image super-resolution, which is composed of 800 high- and low-resolution image pairs. Meanwhile, five common benchmarks are used for evaluation, including Set5 [8], Set14 [63], BSD100 [33], Urban100 [22] and Manga109 [34] with two magnification factors: ×2 and × 4. In practice, we introduce two variants of different sizes: SFHformer-T and SFHformer-M. All PSNR and SSIM values are calculated on the Y channel of images transformed to YCbCr color space.

2 More Experimental Results

2.1 Additional Results for Low-light Enhancement

Table 2: Quantitative evaluations on FiveK [9] dataset.

Methods	DeepUPE [51]	URetinexNet [55]	Uformer [53]	MAXIM [50]	Restormer [61]	Retinexformer [10]	SFHformer
PSNR↑	23.04	23.51	23.89	24.64	24.52	24.94	**25.12**
SSIM↑	0.893	0.826	0.906	0.913	**0.926**	0.907	0.915
#Param.	1.02M	0.34M	20.60M	14.14M	26.10M	1.61M	3.87M
FLOPs	21.10G	57G	41.09G	216G	141.0G	15.57G	26.59G

Low-light Image Enhancement. As shown in Tab.2, we conduct additional experiments on the FiveK dataset [9] to verify the effectiveness of our SFHformer on the low-light enhancement task. Our SFHformer achieves the best performance of PSNR compared to various SOTA approaches. Although our model ranks second in the terms of SSIM, SFHformer has only 14.8% of the parameter size and 18.9% of the computational complexity compared to the best Restormer [61].

2.2 Additional Results for Motion Deblurring

Table 3: Quantitative results on RealBlur dataset [42].

Methods	RealBlur-R [42]	RealBlur-J [42]
(CVPR22)Restormer [61]	36.19/0.957	28.96/0.879
(ECCV22)Stripformer [49]	36.08/0.954	28.82/0.876
(CVPR23)FFTformer [24]	35.87/0.953	27.75/0.853
(Our)SFHformer	36.33/0.963	29.05/0.884

Motion Deblurring. As shown in Tab.3, following [61], we directly apply the GoPro-trained model on real-world motion deblurring dataset: RealBlur [42], to evaluate the generalization and effectiveness of our SFHformer in the real world. The experimental results indicate that our model obtains the best performance in terms of PSNR/SSIM compared with various related methods. In particular, compared to the similar method FFTformer [24] of extracting features in the frequency-domain, our SFHformer achieves a significant improvement in generalization, with PSNR increases of 0.46 and 0.30 in the RealBlur-R and RealBlur-J datasets [42], respectively.

2.3 Additional Results for Single-image Defocus Deblurring

Table 4: Quantitative evaluations on single-image defocus deblurring.

Method	Indoor Scenes				Outdoor Scenes				Combined			
	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓
(CVPR'19)DMENet [25]	25.50	0.788	0.038	0.298	21.43	0.644	0.063	0.397	23.41	0.714	0.051	0.349
(CVPR'15)JNB [44]	26.73	0.828	0.031	0.273	21.10	0.608	0.064	0.355	23.84	0.715	0.048	0.315
(ECCV'20)DPDNet [3]	26.54	0.816	0.031	0.239	22.25	0.682	0.056	0.313	24.34	0.747	0.044	0.277
(ICCV'21)KPAC [45]	27.97	0.852	0.026	0.182	22.62	0.701	0.053	0.269	25.22	0.774	0.040	0.227
(CVPR'21)IFAN [26]	28.11	0.861	0.026	0.179	22.76	0.720	0.052	0.254	25.37	0.789	0.039	0.217
(CVPR'22)Restormer [61]	28.87	**0.882**	0.025	**0.145**	23.24	**0.743**	0.050	**0.209**	25.98	**0.811**	0.038	**0.178**
(CVPR'23)NRKNet [40]	-				-				26.11	0.810	-	0.210
(ICCV'23)INIKNet [41]	-				-				26.01	0.803	-	0.185
(Ours)SFHformer	**28.95**	0.874	**0.024**	0.182	**23.44**	**0.743**	**0.049**	0.260	**26.12**	0.807	**0.037**	0.222

Single-image Defocus Deblurring. Tab.4 shows the quantitative results against SOTA defocus deblurring methods on DPDD [2]. Our model achieves the best PSNR and MAE against the previous SOTA methods.

Table 5: Quantitative evaluations on SIDD dataset [1] for denoising.

Methods	MPRNet [62]	UFormer [53]	Restormer [61]	SFHformer
PSNR↑	39.71	39.77	40.02	**40.19**
SSIM↑	0.958	0.959	0.960	**0.961**

2.4 Additional Results for Image Denoising

Image Denoising. As shown in Tab.5, we evaluate the effectiveness of our SFHformer on the SIDD dataset [1] for image denoising task. As suggested by the quantitative results against SOTA methods, our model obtains the superior performance.

2.5 Additional Results for Super-resolution

Table 6: Super-resolution results. Train on DIV2K [4] dataset and test on five common benchmarks. PSNR/SSIM is adopted for metrics.

Methods	Scale	Params	FLOPs	Set5 [8]	Set14 [63]	B100 [33]	Urban100 [22]	Manga109 [34]
(ICCV23)SAFMN [48]	×2	228K	52G	38.00/0.9605	33.54/0.9177	32.16/0.8995	31.84/0.9256	38.71/0.9771
(Our)SFHformer-T		310K	63G	38.02/0.9607	33.52/0.9186	32.21/0.9001	32.23/0.9295	38.82/0.9775
(AAAI23)HPUN-L [47]		714K	160G	38.09/0.9608	33.79/0.9198	32.25/0.9006	32.37/0.9307	39.07/0.9779
(CVPR23)NGswin [15]		998K	140G	38.05/0.9610	33.79/0.9199	32.27/0.9008	32.53/0.9324	38.97/0.9777
(ICCV23)CRAFT [27]		737K	173G	38.23/0.9615	33.92/0.9211	32.33/0.9016	32.86/0.9343	39.39/0.9786
(Our)SFHformer-M		919K	168G	38.24/0.9617	33.95/0.9216	32.38/0.9020	33.08/0.9364	39.33/0.9782
(ICCV23)SAFMN [48]	×4	240K	14G	32.18/0.8948	28.60/0.7813	27.58/0.7359	25.97/0.7809	30.43/0.9063
(Our)SFHformer-T		330K	16G	32.14/0.8943	28.48/0.7817	27.61/0.7371	26.10/0.7848	30.50/0.9073
(AAAI23)HPUN-L [47]		734K	40G	32.31/0.8962	28.73/0.7842	27.66/0.7386	26.27/0.7918	30.77/0.9109
(CVPR23)NGswin [15]		1019K	37G	32.33/0.8963	28.78/0.7859	27.66/0.7396	26.45/0.7963	30.80/0.9128
(ICCV23)CRAFT [27]		753K	44G	32.52/0.8989	28.85/0.7872	27.72/0.7418	26.56/0.7995	31.18/0.9168
(Our)SFHformer-M		939K	43G	32.45/0.8984	28.80/0.7864	27.77/0.7419	26.63/0.7998	31.09/0.9149

Efficient Image Super-resolution. To evaluate our SFHformer for super-resolution task, we conduct extensive experiments on DIV2K dataset [4] with two magnification factors: ×2 and ×4. For a more complete comparison with the recent approaches, we set up two configurations of different sizes: SFHformer-T and SFHformer-M. As shown in Tab.6, with similar model parameter size and computational complexity, our model achieves very competitive performance in terms of PSNR and SSIM among five common benchmarks.

2.6 Additional Results for Running time

Table 7: Run time in deraining.

Methods	Run time	PSNR	FLOPs
Restormer [61]	77.23ms	47.98	141.0G
IDT [56]	139.72ms	47.35	58.4G
DRSformer [14]	171.90ms	48.53	242.9G
SFHformer	**43.29ms**	**50.11**	**26.6G**

Table 8: Run time in dehazing.

Methods	Run time	PSNR	FLOPs
Dehazeformer [46]	37.39ms	38.46	48.6G
C^2PNet [65]	81.47ms	42.56	460.9G
MB-Taylorformer [39]	377.86ms	42.64	88.1G
SFHformer	**43.29ms**	**43.03**	**26.6G**

Running time vs. Performance. As shown in Tab.7 and Tab.8, we conduct additional experiments to evaluate our SFHformer's throughput and real-time performance efficiency on image deraining and image dehazing tasks. In practice, we measure the running time at one 256×256 image input under one NVIDIA 3090 GPU for both restoration tasks. The experimental results demonstrate that our model obtains faster running time with better performance.

3 Discussion and Future Work

While our proposed SFHformer has delivered competitive performance across a range of image restoration tasks and achieved a favorable balance between running time, model size, and computational cost, there remain areas that could benefit from further refinement and exploration. In the following discussion, we will delve into these aspects in greater detail.

(1) Modest improvements in some benchmarks. Despite performing well in tasks such as dehazing and deraining, our SFHformer exhibits less significant enhancements in others, notably deblurring. We infer that a key factor resulting in these modest gains in deblurring may primarily stem from the discrepancy between the training and testing image sizes. For instance, in tasks like deraining and debluring, we train the models at 256×256 size, yet test at 480×320 and 1280×720 sizes respectively. This larger size inconsistency tends to impact frequency domain operations more significantly than spatial domain operations, as the former are inherently more sensitive to changes in image dimensions. While solutions (e.g. TLC [16]) have been developed to address this challenge from a spatial perspective, exploring resolutions within the frequency domain presents an intriguing and promising avenue for future research.

(2) Depthwise Convolution vs. Pointwise Convolution. The global operation of the Fast Fourier Transform (FFT) consolidates the characteristics of the spatial domain into specific components within the frequency domain. This transformation principle leads to a lack of correlation and inductive bias among adjacent points in the frequency domain. Our network design confirms this intuition, showing that pointwise (PW) convolution for channel dimensions outperforms depthwise (DW) convolution for spatial dimensions. The introduction of DW convolution, while initially promising, actually increases the instability of network training and reduces model performance. Although our experiments demonstrate that directly applying DW convolution to extract frequency features is not effective, the presence of significant macroscopic patterns and representations of various degradation processes in the frequency domain, as shown in the introduction, suggests that exploring effective methods for extracting frequency features from spatial dimensions remains a valuable research avenue.

In summary, we aspire for our SFHformer to provide substantial insights from the frequency-domain perspective and to pioneer a new avenue in model design for the image restoration community.

4 More Visual Results

We provide images recovered by various methods for different image restoration tasks, organised as,

- Image dehazing: Fig.1, Fig.2, Fig.3.
- Image deraining: Fig.4, Fig.5.
- Image motion deblurring: Fig.6.
- Image raindrop removal: Fig.7.
- Image desnowing: Fig.8.
- Low-light image enhancement: Fig.9.
- Underwater image enhancement: Fig.10.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1692–1700 (2018) 1, 4, 6
2. Abuolaim, A., Brown, M.S.: Defocus deblurring using dual-pixel data. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16. pp. 111–126. Springer (2020) 1, 3, 5
3. Abuolaim, A., Brown, M.S.: Defocus deblurring using dual-pixel data. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16. pp. 111–126. Springer (2020) 5
4. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 126–135 (2017) 1, 4, 6
5. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In: 2019 IEEE international conference on image processing (ICIP). pp. 1014–1018. IEEE (2019) 1, 2
6. Ancuti, C.O., Ancuti, C., Timofte, R.: Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 444–445 (2020) 1, 2, 15
7. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 754–762 (2018) 1, 2, 14
8. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012) 1, 4, 6
9. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: CVPR 2011. pp. 97–104. IEEE (2011) 1, 4
10. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. arXiv preprint arXiv:2303.06705 (2023) 4, 21
11. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European Conference on Computer Vision. pp. 17–33. Springer (2022) 18, 20

12. Chen, W.T., Fang, H.Y., Ding, J.J., Tsai, C.C., Kuo, S.Y.: Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16. pp. 754–770. Springer (2020) 1, 3, 20
13. Chen, W.T., Fang, H.Y., Hsieh, C.L., Tsai, C.C., Chen, I., Ding, J.J., Kuo, S.Y., et al.: All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4196–4205 (2021) 1, 3
14. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5896–5905 (2023) 2, 6, 16, 17
15. Choi, H., Lee, J., Yang, J.: N-gram in swin transformers for efficient lightweight image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2071–2081 (2023) 4, 6
16. Chu, X., Chen, L., Chen, C., Lu, X.: Improving image restoration by revisiting global information aggregation. In: European Conference on Computer Vision. pp. 53–71. Springer (2022) 7
17. Cui, Y., Ren, W., Cao, X., Knoll, A.: Focal network for image restoration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13001–13011 (2023) 13, 20
18. Cui, Y., Tao, Y., Bing, Z., Ren, W., Gao, X., Cao, X., Huang, K., Knoll, A.: Selective frequency network for image restoration. In: The Eleventh International Conference on Learning Representations (2022) 2, 18
19. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2157–2167 (2020) 13, 14, 15
20. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3855–3863 (2017) 1, 2, 16
21. Fu, Z., Wang, W., Huang, Y., Ding, X., Ma, K.K.: Uncertainty inspired underwater image enhancement. In: European Conference on Computer Vision. pp. 465–482. Springer (2022) 22
22. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5197–5206 (2015) 1, 4, 6
23. Huo, F., Li, B., Zhu, X.: Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1944–1952 (2021) 22
24. Kong, L., Dong, J., Ge, J., Li, M., Pan, J.: Efficient frequency domain-based transformers for high-quality image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5886–5895 (2023) 5
25. Lee, J., Lee, S., Cho, S., Lee, S.: Deep defocus map estimation using domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12222–12230 (2019) 5
26. Lee, J., Son, H., Rim, J., Cho, S., Lee, S.: Iterative filter adaptive network for single image defocus deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2034–2042 (2021) 5

27. Li, A., Zhang, L., Liu, Y., Zhu, C.: Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12514–12524 (2023) 4, 6
28. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE international conference on computer vision. pp. 4770–4778 (2017) 13, 14, 15
29. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2018) 1, 2, 13
30. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing **29**, 4376–4389 (2019) 1, 3, 22
31. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7314–7323 (2019) 13, 14, 15
32. Liu, Y.F., Jaw, D.W., Huang, S.C., Hwang, J.N.: Desnownet: Context-aware deep network for snow removal. IEEE Transactions on Image Processing **27**(6), 3064–3073 (2018) 1, 3
33. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings eighth IEEE international conference on computer vision. ICCV 2001. vol. 2, pp. 416–423. IEEE (2001) 1, 4, 6
34. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia tools and applications **76**, 21811–21838 (2017) 1, 4, 6
35. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3883–3891 (2017) 1, 2, 18
36. Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. IEEE Transactions on Image Processing (2023) 1, 3
37. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2482–2491 (2018) 1, 3, 19
38. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 11908–11915 (2020) 13, 14, 15
39. Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H., Jin, Z.: Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12802–12813 (2023) 6
40. Quan, Y., Wu, Z., Ji, H.: Neumann network with recursive kernels for single image defocus deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5754–5763 (2023) 5
41. Quan, Y., Yao, X., Ji, H.: Single image defocus deblurring via implicit neural inverse kernels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12600–12610 (2023) 5
42. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Computer Vision–ECCV 2020: 16th European

Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. pp. 184–201. Springer (2020) 1, 2, 3, 5

43. Shen, Z., Wang, W., Lu, X., Shen, J., Ling, H., Xu, T., Shao, L.: Human-aware motion deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5572–5581 (2019) 1, 2
44. Shi, J., Xu, L., Jia, J.: Just noticeable defocus blur detection and estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 657–665 (2015) 5
45. Son, H., Lee, J., Cho, S., Lee, S.: Single image defocus deblurring using kernel-sharing parallel atrous convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2642–2650 (2021) 5
46. Song, Y., He, Z., Qian, H., Du, X.: Vision transformers for single image dehazing. IEEE Transactions on Image Processing **32**, 1927–1941 (2023) 6, 14, 15
47. Sun, B., Zhang, Y., Jiang, S., Fu, Y.: Hybrid pixel-unshuffled network for lightweight image super-resolution. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 2375–2383 (2023) 6
48. Sun, L., Dong, J., Tang, J., Pan, J.: Spatially-adaptive feature modulation for efficient image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13190–13199 (2023) 4, 6
49. Tsai, F.J., Peng, Y.T., Lin, Y.Y., Tsai, C.C., Lin, C.W.: Stripformer: Strip transformer for fast image deblurring. In: European Conference on Computer Vision. pp. 146–162. Springer (2022) 5
50. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.: Maxim: Multi-axis mlp for image processing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5769–5780 (2022) 3, 4, 19
51. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6849–6857 (2019) 4
52. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12270–12279 (2019) 1, 2
53. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17683–17693 (2022) 4, 6
54. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 (2018) 1, 4
55. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5901–5910 (2022) 4
56. Xiao, J., Fu, X., Liu, A., Wu, F., Zha, Z.J.: Image de-raining transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 6, 16, 17, 19
57. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022) 21
58. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1357–1366 (2017) 1, 2, 17

59. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021) 1, 4, 21
60. Yi, Q., Li, J., Dai, Q., Fang, F., Zhang, G., Zeng, T.: Structure-preserving deraining with residue channel prior guidance. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4238–4247 (2021) 16, 17
61. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022) 2, 4, 5, 6, 16, 17, 18
62. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14821–14831 (2021) 6, 16, 17, 18
63. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7. pp. 711–730. Springer (2012) 1, 4, 6
64. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 695–704 (2018) 1, 2, 16
65. Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y.: Curricular contrastive regularization for physics-aware single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5785–5794 (2023) 6

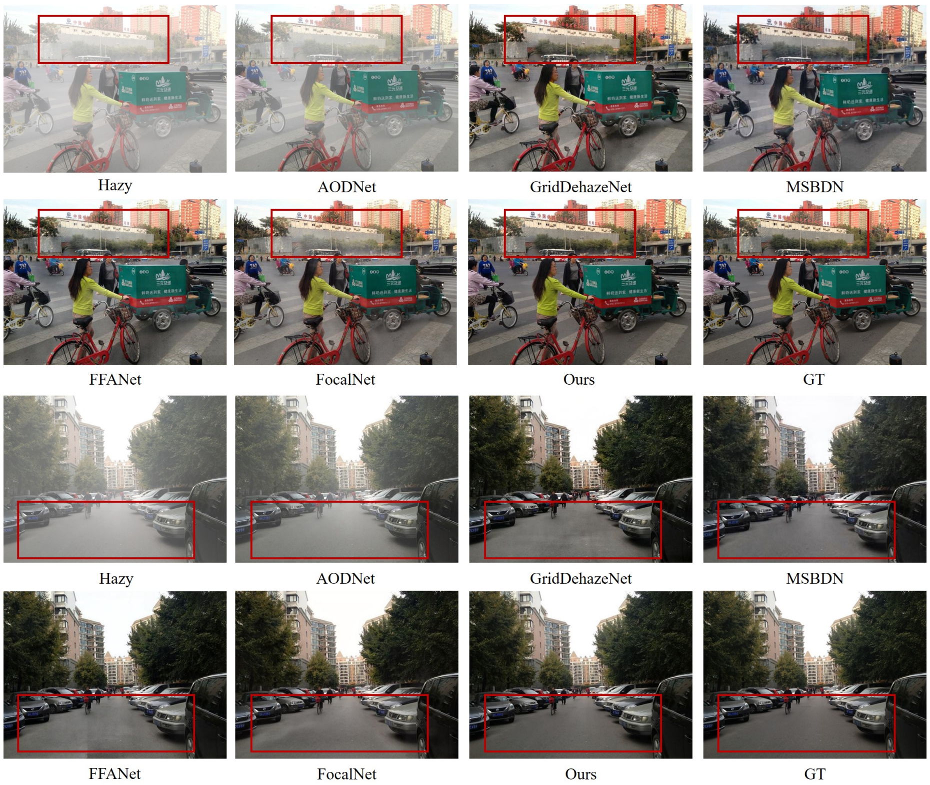


Fig. 1: Visual results for image dehazing on RESIDE [29] among AODNet [28], GridDehazeNet [31], MSBDN [19], FFANet [38], FocalNet [17] and ours.

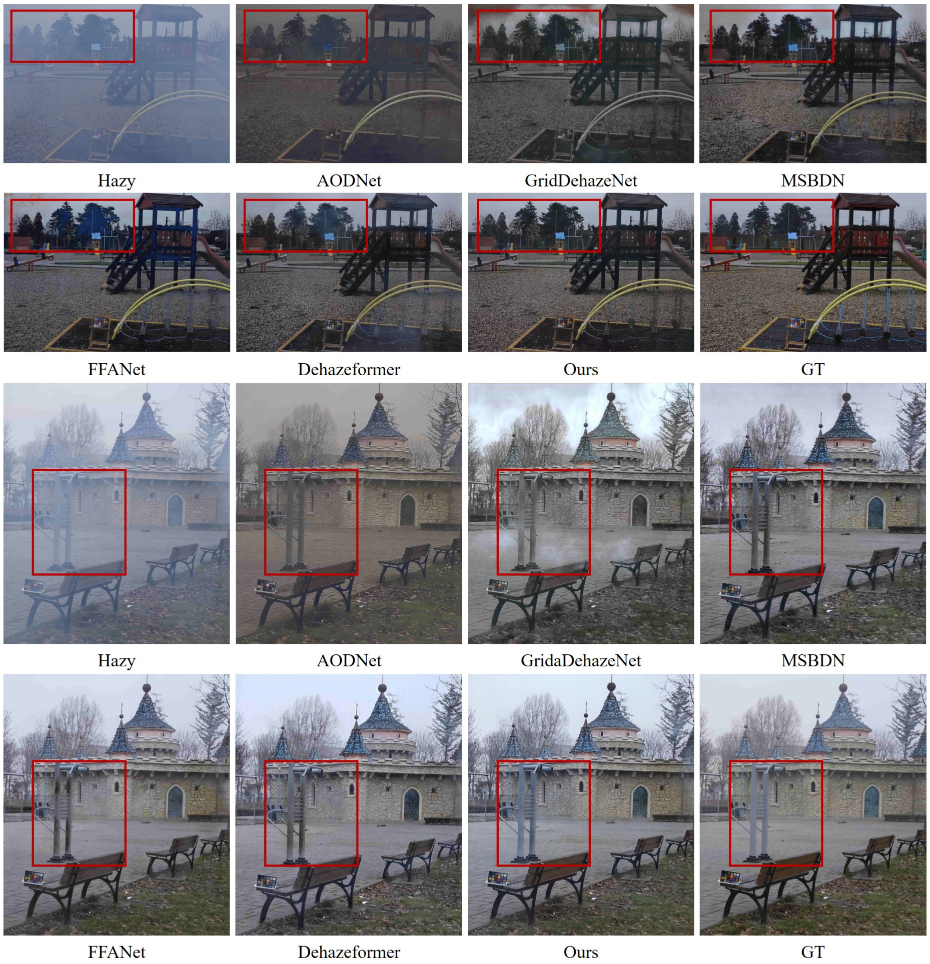


Fig. 2: Visual results for image dehazing on O-HAZE [7] among AODNet [28], GridDehazeNet [31], MSBDN [19], FFANet [38], Dehazeformer [46] and ours.

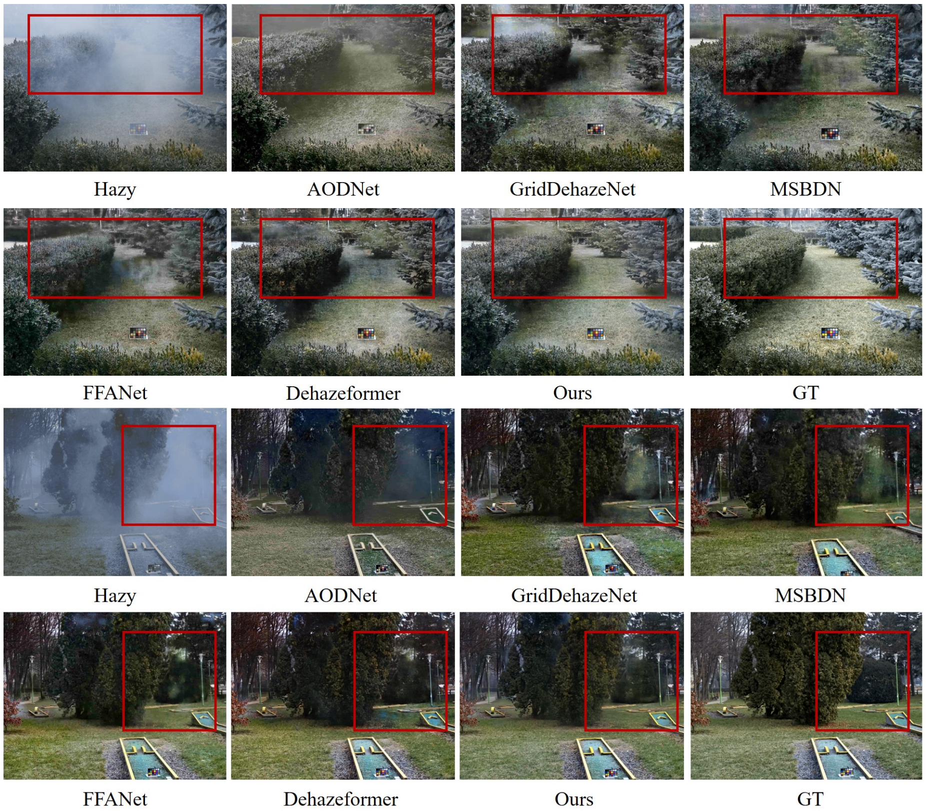


Fig. 3: Visual results for image dehazing on NH-HAZE [6] among AODNet [28], GridDehazeNet [31], MSBDN [19], FFANet [38], Dehazeformer [46] and ours.

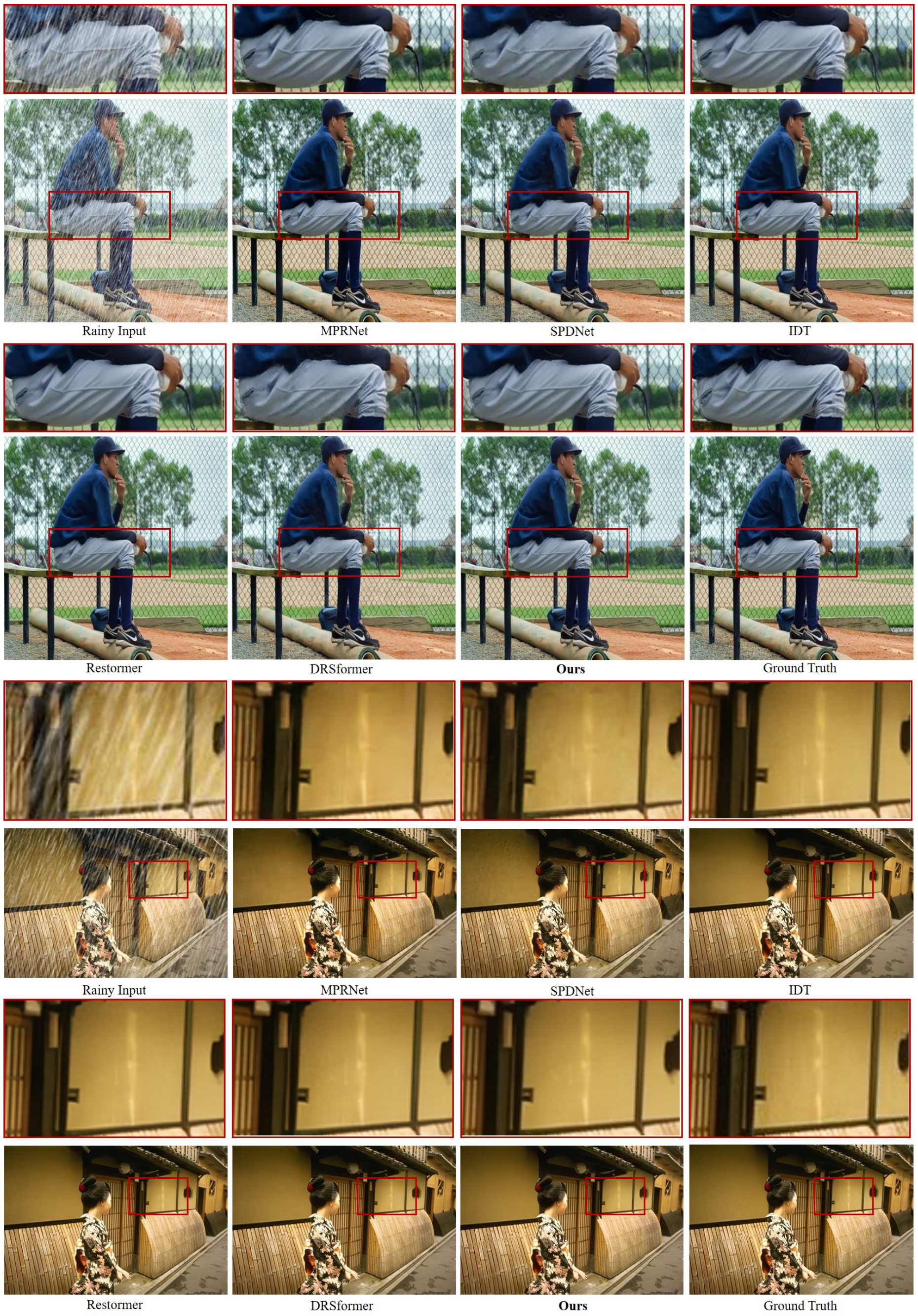


Fig. 4: Visual results for image deraining on DID-Data [64] and DDN-Data [20] among MPRNet [62], SPDNet [60], IDT [56] , Restormer [61], DRSformer [14] and ours.

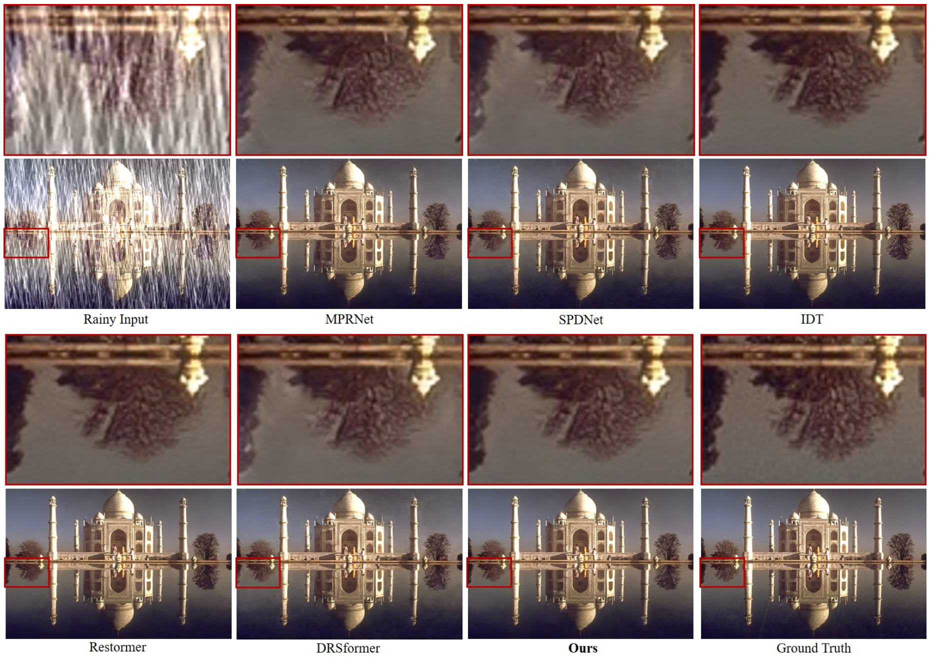


Fig. 5: Visual results for image deraining on Rain200H [58] among MPRNet [62], SPDNet [60], IDT [56] , Restormer [61], DRSformer [14] and ours.

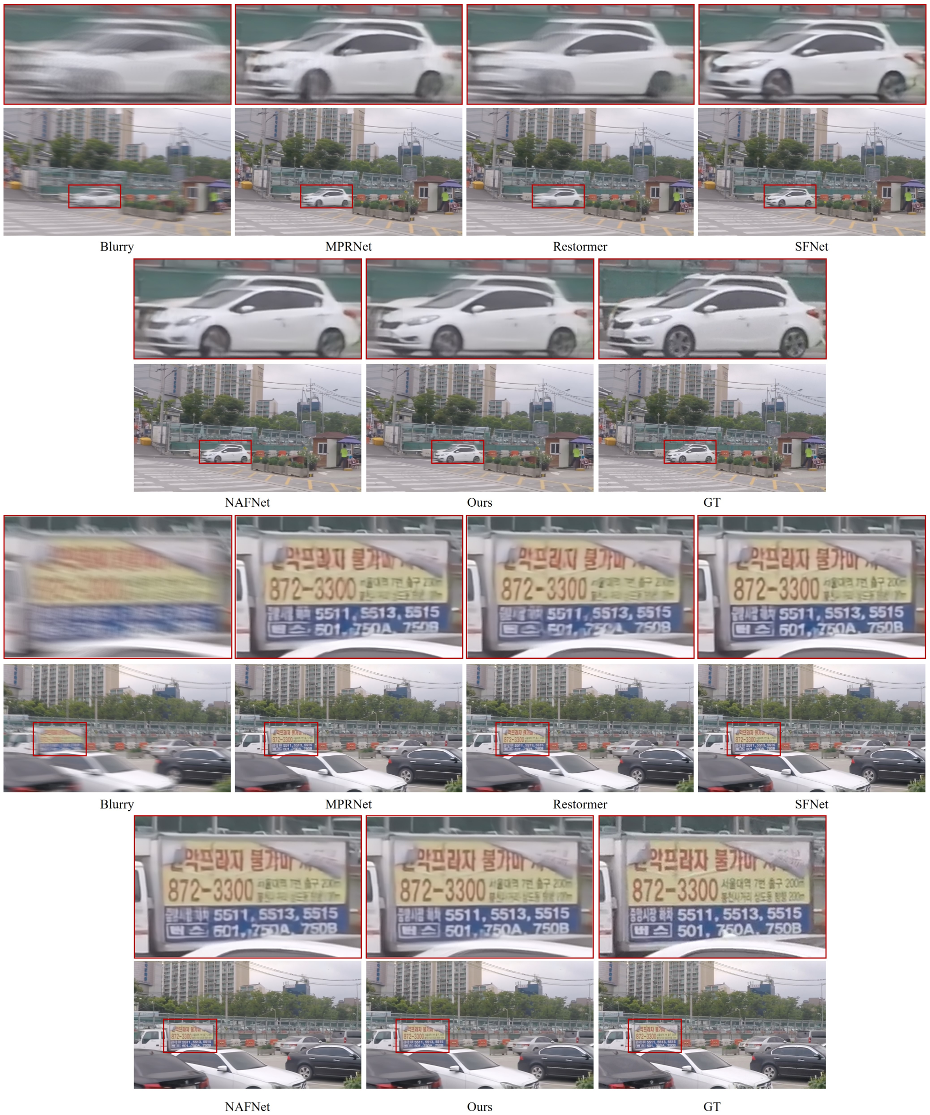


Fig. 6: Visual results for image motion deblurring on GoPro [35] among MPRNet [62], Restormer [61], SFNet [18], NAFNet [11] and ours.

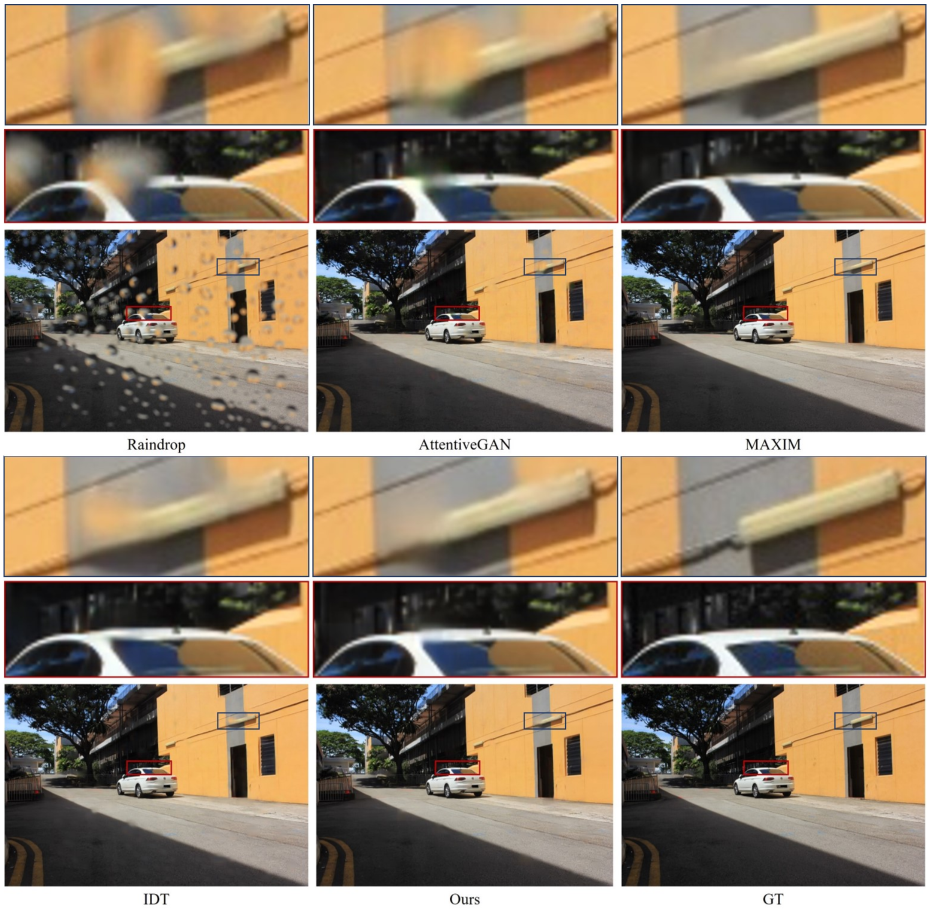


Fig. 7: Visual results for image raindrop removal on raindrop [37] among Attentive-GAN [37], MAXIM [50], IDT [56] and ours.

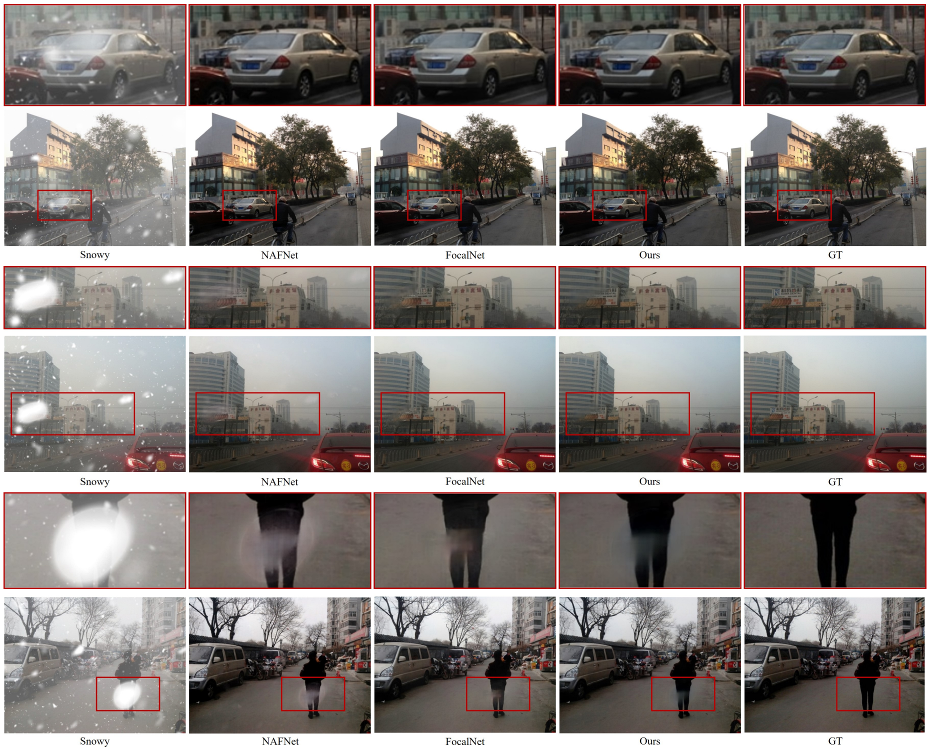
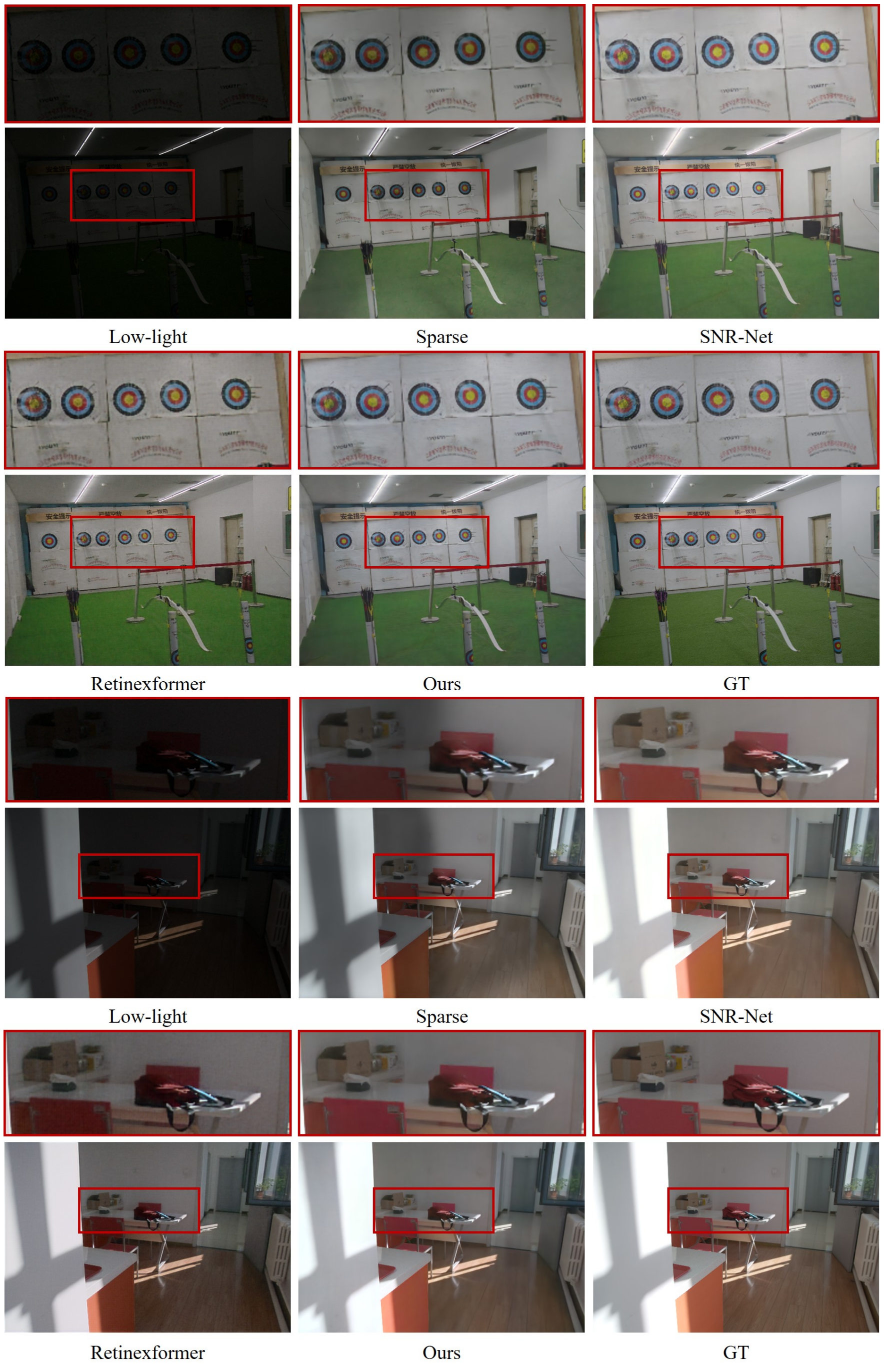


Fig. 8: Visual results for image desnowing on SRRS [12] among NAFNet [11], FocalNet [17] and ours.



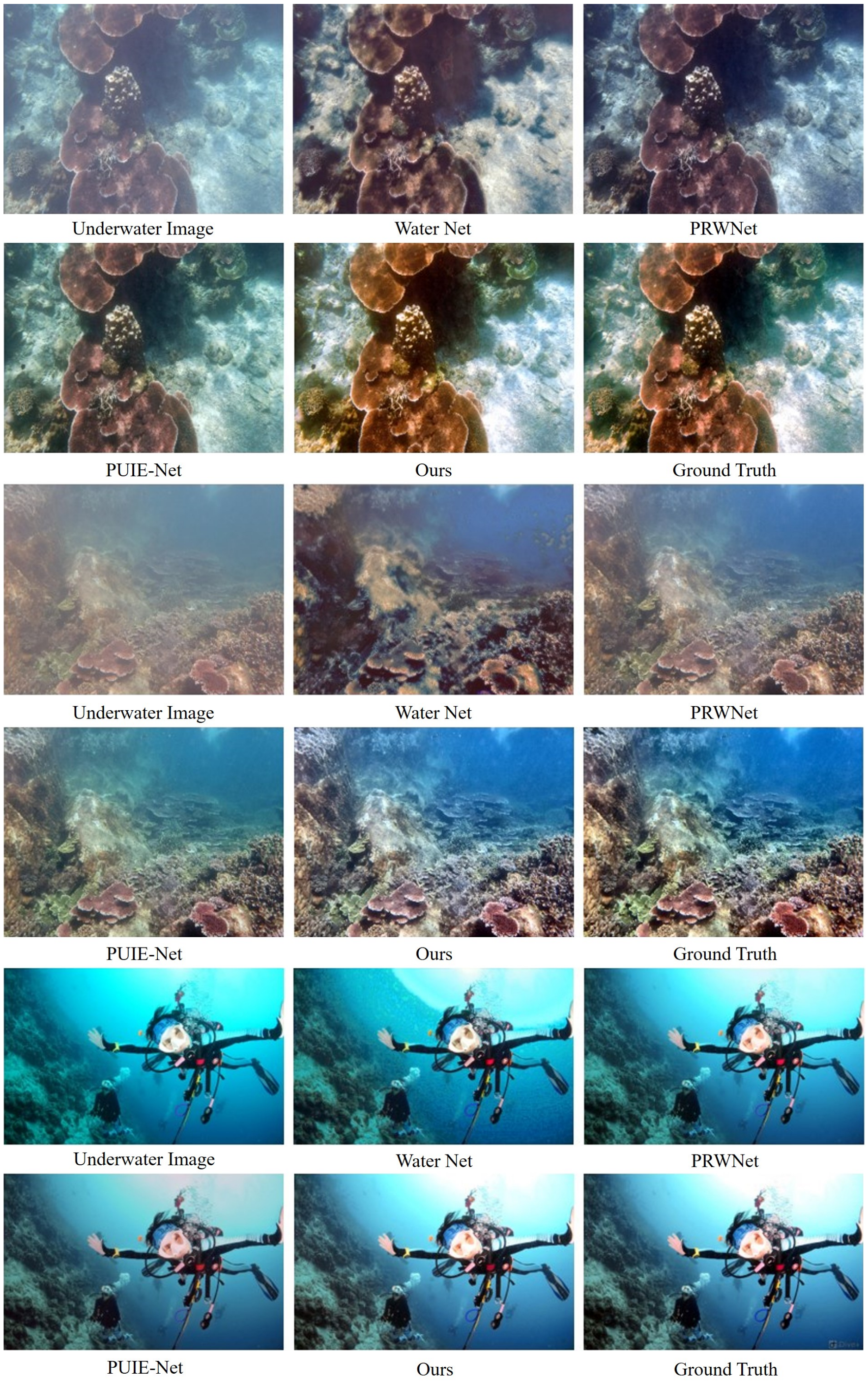
Low-light Sparse SNR-Net

Retinexformer Ours GT

Low-light Sparse SNR-Net

Retinexformer Ours GT

Fig. 9: Visual results for low-light image enhancement on LOL-v2 [59] among Spare [59], SNR-Net [57], Retinexformer [10] and ours.



Underwater Image Water Net PRWNet

PUIE-Net Ours Ground Truth

Underwater Image Water Net PRWNet

PUIE-Net Ours Ground Truth

Underwater Image Water Net PRWNet

PUIE-Net Ours Ground Truth

Fig. 10: Visual results for underwater image enhancement on UIEB [30] among WaterNet [30], PRWNet [23], PUIE-Net [21] and ours.