Boosting Image Restoration via Priors from Pre-trained Models

Xiaogang Xu[1,2,3] Shu Kong[5,6,7] Tao Hu[3,8] Zhe Liu[1*] Hujun Bao[1,4]
[1] Zhejiang Lab [2] CUHK [3] RealityEdge [4] Zhejiang University [5] University of Macau
[6] Institute of Collaborative Innovation [7] Texas A&M University [8] National University of Singapore
xiaogangxu00@gmail.com, skong@um.edu.mo, yihouxiang@gmail.com
zhe.liu@zhejianglab.com, bao@cad.zju.edu.cn

CVPR 2024

Presenter: Hao Wang

Advisor: Prof. Chia-Wen Lin

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Introduction

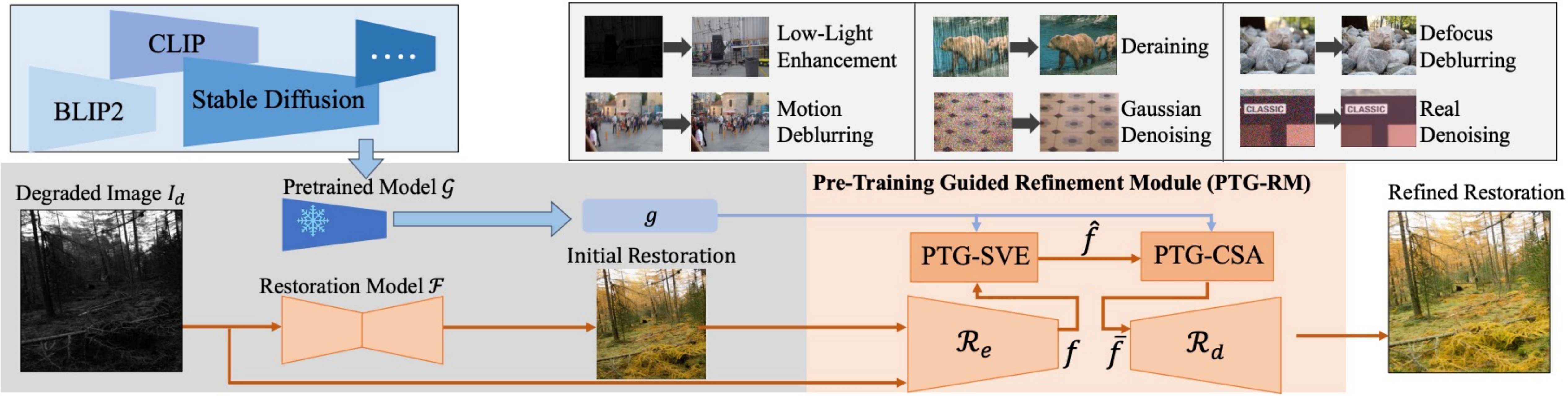
- Present a novel and general method that **leverages pre-trained models** to enhance various restoration tasks.
- propose a novel paradigm that leverages **pre-trained priors** to formulate effective **neural operation ranges** and **attention mechanisms**.
- We validate our method through extensive experiments on different datasets, networks, and tasks, and show **remarkable improvements** over prior methods.

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Framework

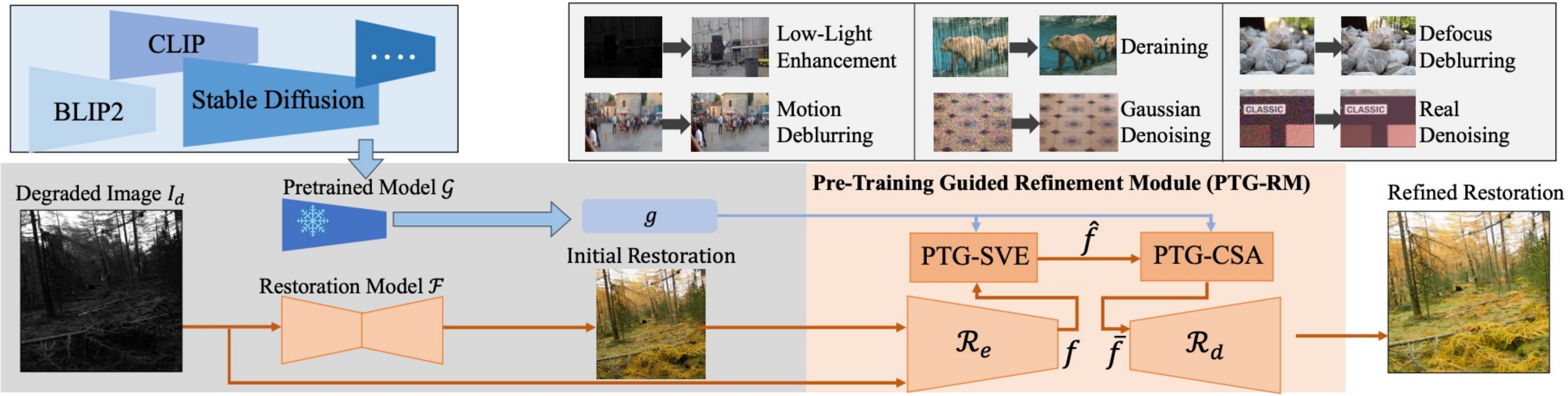
Using a pre-trained model $\mathcal{G}$ to boost image restoration



Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Overview of Refinement Module



$$\hat{I}_c = \mathcal{F}(I_d)$$

$$g = \mathcal{G}(I_d)$$

$$\bar{I}_c = \mathcal{R}(\hat{I}_c, I_d, g)$$

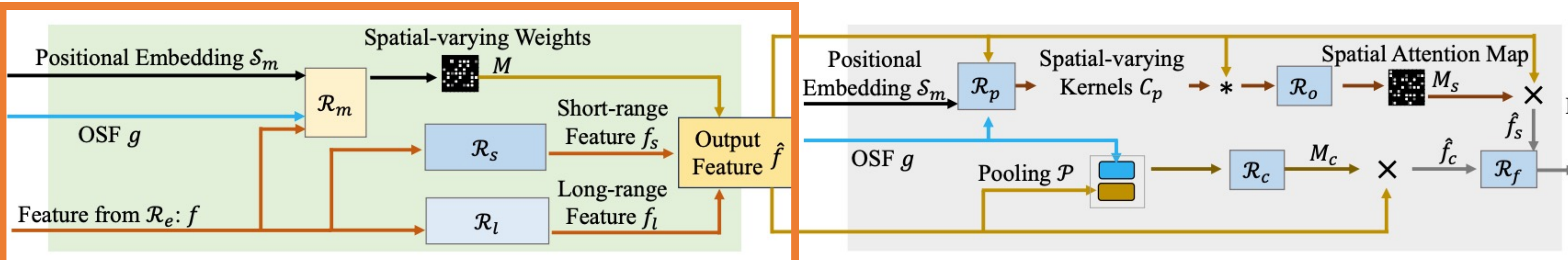
$$f = \mathcal{R}_e(\hat{I}_c \oplus I_d)$$

$$\bar{f} = \mathcal{C}(\mathcal{A}(f, g), g)$$

$$[I_m, I_r] = \mathcal{R}_d(\hat{f})$$

$$\bar{I}_c = I_d + (\hat{I}_c - I_d) \times I_m + I_r$$

Pre-Train-Guided Spatial-Varying Operations



(a) Pre-Train-Guided Spatial-Varying Enhancement $\mathcal{A}$

(b) Pre-Train-Guided Channel-Spatial Attention $\mathcal{C}$

$$M = \mathcal{R}_m(f \oplus \mathcal{T}_m(g) \oplus \mathcal{S}_m).$$

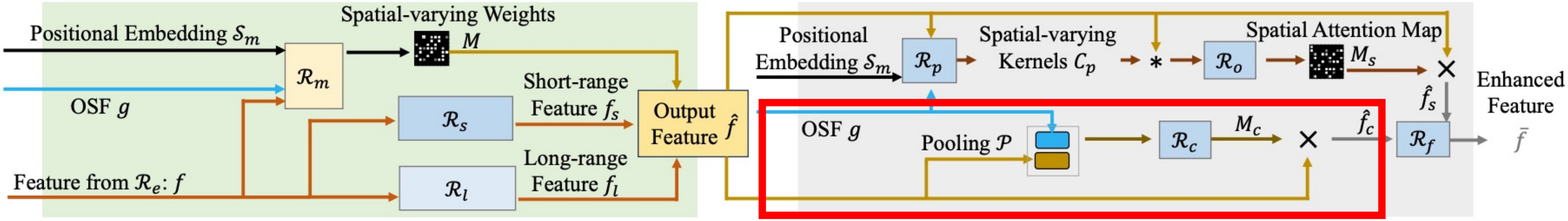
$$f_s = \mathcal{R}_s(f),\ f_l = \mathcal{R}_l(f)$$

$$\hat{f} = M \times f_s + (1 - M) \times f_l.$$

- R_s Short range
  - CNN
- R_l Long range
  - Transformer

Pre-Train-Guided Attention

- **channel-attention**



(a) Pre-Train-Guided Spatial-Varying Enhancement $\mathcal{A}$

(b) Pre-Train-Guided Channel-Spatial Attention $\mathcal{C}$

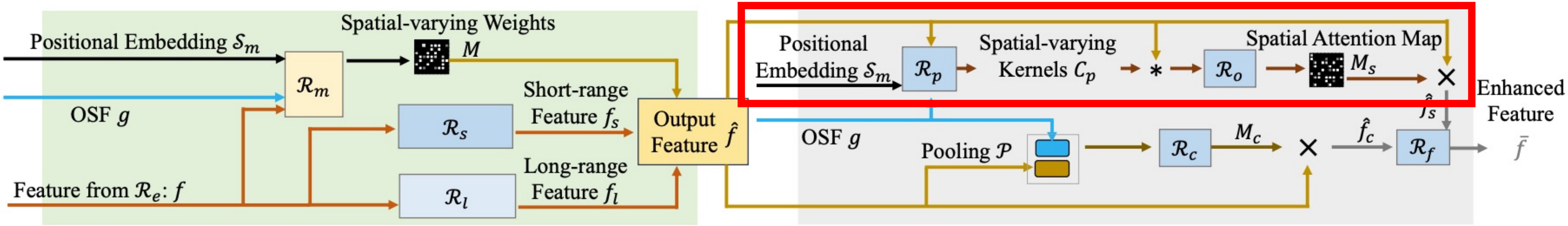
$$M_c = \mathcal{R}_c(\mathcal{O}(\hat{f}) \oplus \mathcal{T}_c(g))$$

$$\mathcal{M}_c \in \mathbb{R}^c$$

$$\hat{f}_c = \hat{f} \times M_c, \qquad (4)$$

Pre-Train-Guided Attention

- **spatial-attention**
  - similar condition for neighboring features, limiting the elimination of spatial artifacts.



(a) Pre-Train-Guided Spatial-Varying Enhancement $\mathcal{A}$

(b) Pre-Train-Guided Channel-Spatial Attention $\mathcal{C}$

$$\mathcal{C}_p = \mathcal{R}_p(\hat{f}, \mathcal{T}_c(g), \mathcal{S}_c)$$

$$\mathcal{C}_p \in \mathbb{R}^{h \times w \times (k_h \times k_w \times c)}$$

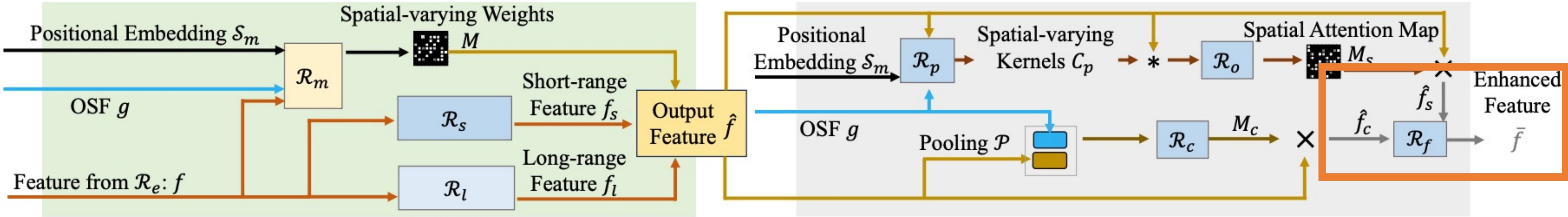
$$\hat{M}_s = \hat{f} * \mathcal{C}_p$$

$$M_s = \mathcal{R}_o(\hat{M}_s)$$

- maps the feature channel c to 1

$$\hat{f}_s = \hat{f} \times M_s$$

Fusion module



$$\bar{f} = \mathcal{R}_f(\hat{f}_c \oplus \hat{f}_s)$$

$$[I_m, I_r] = \mathcal{R}_d(\hat{f})$$

$$\bar{I}_c = I_d + (\hat{I}_c - I_d) \times I_m + I_r$$

Loss function

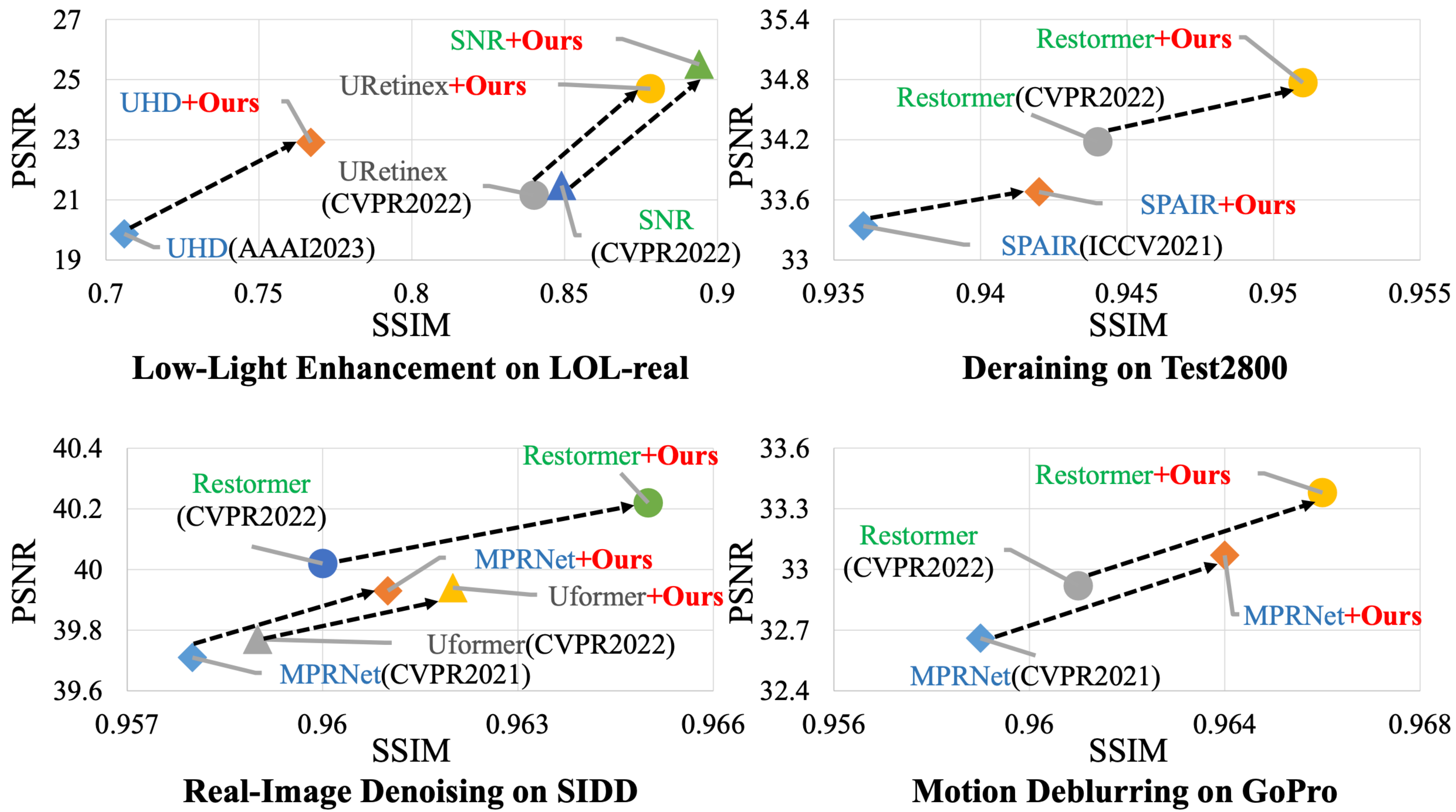
$$\mathcal{L}_g(\hat{I}_c, \mathcal{I}_c) + \lambda_1 \mathcal{L}_g(\bar{I}_c, \mathcal{I}_c)$$

- R can be jointly trained with the model F
- Loss
  - reconstruction loss in the pixel level
  - perceptual loss

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Result



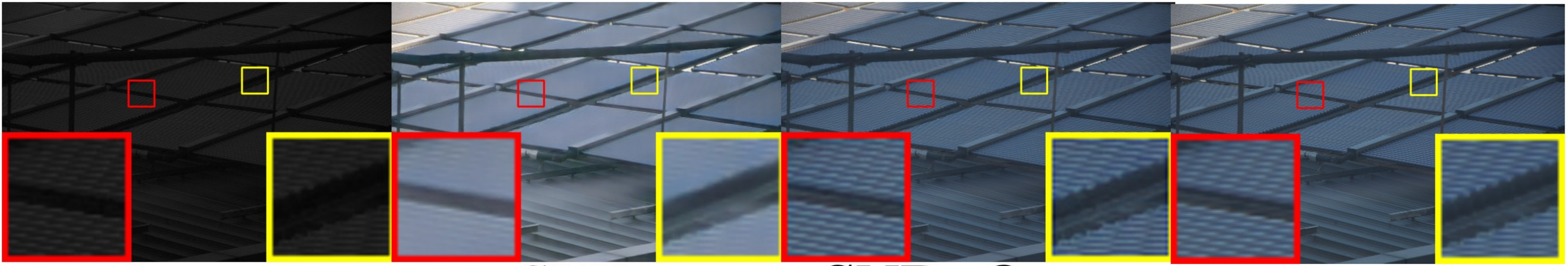
SNR+Ours
URetinex+Ours
UHD+Ours
URetinex
(CVPR2022)
SNR
(CVPR2022)
UHD(AAAI2023)
PSNR
SSIM
Low-Light Enhancement on LOL-real
Restormer+Ours
Restormer(CVPR2022)
SPAIR+Ours
SPAIR(ICCV2021)
PSNR
SSIM
Deraining on Test2800
Restormer+Ours
Restormer
(CVPR2022)
MPRNet+Ours
Uformer+Ours
Uformer(CVPR2022)
MPRNet(CVPR2021)
PSNR
SSIM
Real-Image Denoising on SIDD
Restormer+Ours
Restormer
(CVPR2022)
MPRNet+Ours
MPRNet(CVPR2021)
PSNR
SSIM
Motion Deblurring on GoPro

Result (low-light enhancement)

Datasets	Methods	Original		+Ours-c		+Ours-b		+Ours-s		+Ours-r		+Ours-f	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LOL	UHD	19.87	0.706	22.91 (**+3.04**)	0.767 (**+6.1**)	21.83 (**+1.96**)	0.732 (**+2.6**)	22.35 (**+2.48**)	0.758 (**+5.2**)	21.71 (**+1.84**)	0.737 (**+3.1**)	22.74 (**+2.87**)	0.764 (**+5.8**)
	URetinex	21.16	0.840	24.70 (**+3.54**)	0.878 (**+3.8**)	23.57 (**+2.41**)	0.869 (**+2.9**)	24.23 (**+3.07**)	0.866 (**+2.6**)	23.99 (**+2.83**)	0.862 (**+2.2**)	24.56 (**+3.40**)	0.870 (**+3.0**)
	SNR	21.48	0.849	25.50 (**+4.02**)	0.892 (**+4.3**)	25.61 (**+4.13**)	0.891 (**+4.2**)	25.19 (**+3.71**)	0.874 (**+2.5**)	25.24 (**+3.76**)	0.887 (**+3.8**)	24.90 (**+3.42**)	0.888 (**+3.9**)
SID	UHD	20.46	0.614	20.99 (**+0.53**)	0.616 (**+0.2**)	21.06 (**+0.60**)	0.619 (**+0.5**)	22.34 (**+1.88**)	0.625 (**+1.1**)	21.11 (**+0.65**)	0.618 (**+0.4**)	21.08 (**+0.62**)	0.619 (**+0.5**)
	URetinex	21.56	0.619	22.34 (**+0.78**)	0.623 (**+0.4**)	22.02 (**+0.46**)	0.621 (**+0.2**)	22.21 (**+0.65**)	0.623 (**+0.4**)	22.17 (**+0.61**)	0.625 (**+0.6**)	22.40 (**+0.84**)	0.626 (**+0.7**)
	SNR	22.87	0.625	23.34 (**+0.47**)	0.630 (**+0.5**)	23.15 (**+0.28**)	0.627 (**+0.2**)	23.08 (**+0.21**)	0.631 (**+0.6**)	23.06 (**+0.19**)	0.632 (**+0.7**)	23.17 (**+0.30**)	0.636 (**+1.1**)

- Comparisons on LOL-real and SID dataset.
  - −c, using CLIP
  - −b, using BLIP2
  - −s, using Stable Diffusion
  - −r, using restoration models trained on SDSD
  - −f, denotes applying refinement on the features of F

Result (low-light enhancement)



Input SNR **SNR+Ours** Ground-truth

Input SNR **SNR+Ours** Ground-truth

Result

- Comparison with Other Priors

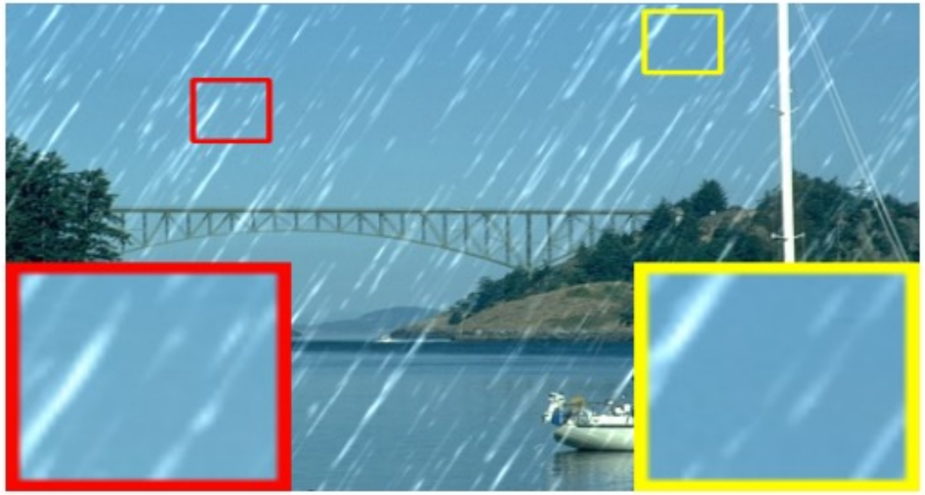
Methods	SNR	+SKF	+SMG	+SMG(dep)	+Ours-c
PSNR	21.48	23.05	24.84	24.12	**25.50**
SSIM	0.849	0.853	0.880	0.851	**0.892**
Methods	URetinex	+SKF	+SMG	+SMG(dep)	+Ours-c
PSNR	21.16	23.51	23.74	23.25	**24.70**
SSIM	0.840	0.856	0.852	0.849	**0.878**
+Params	0	2.15M	16.76M	16.76M	0.67M

- SKF and SMG, utilize additional information, requiring supervision with paired multi-modal information
  - semantic maps
  - edge maps
  - depth maps to enhance
- Better performance and lesser parameter

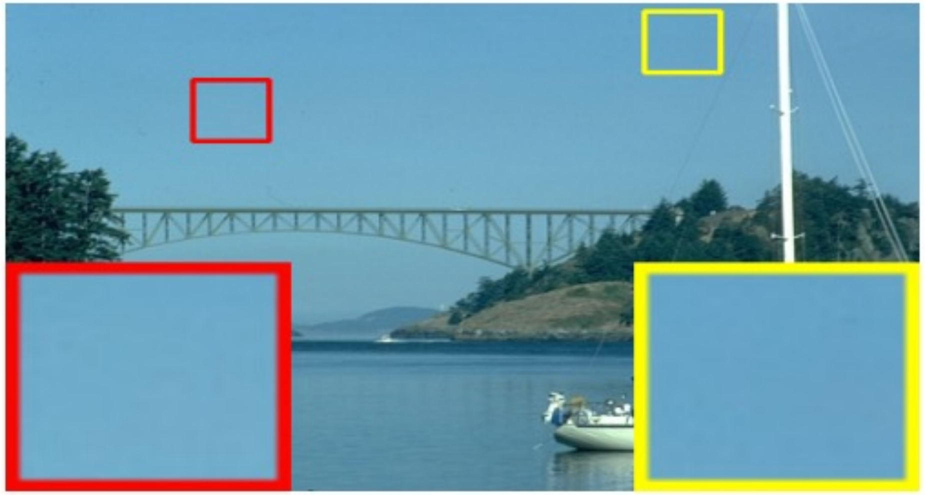
Result (Deraining)

Method	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
	Test100		**Rain100H**		**Rain100L**	
SPAIR	30.35	0.909	30.95	0.892	36.93	0.969
SPAIR+Ours-c	**30.62**	**0.917**	**31.20**	**0.901**	**37.26**	**0.973**
Restormer	32.00	0.923	31.46	0.904	38.99	0.978
Restormer+Ours-c	**32.30**	**0.934**	**31.77**	**0.913**	**39.27**	**0.985**
	Test2800		**Test1200**		**Average**	
SPAIR	33.34	0.936	33.04	0.922	32.91	0.926
SPAIR+Ours-c	**33.58**	**0.942**	**33.35**	**0.924**	**33.16**	**0.932**
Restormer	34.18	0.944	33.19	0.926	33.96	0.935
Restormer+Ours-c	**34.47**	**0.951**	**33.48**	**0.929**	**34.24**	**0.943**

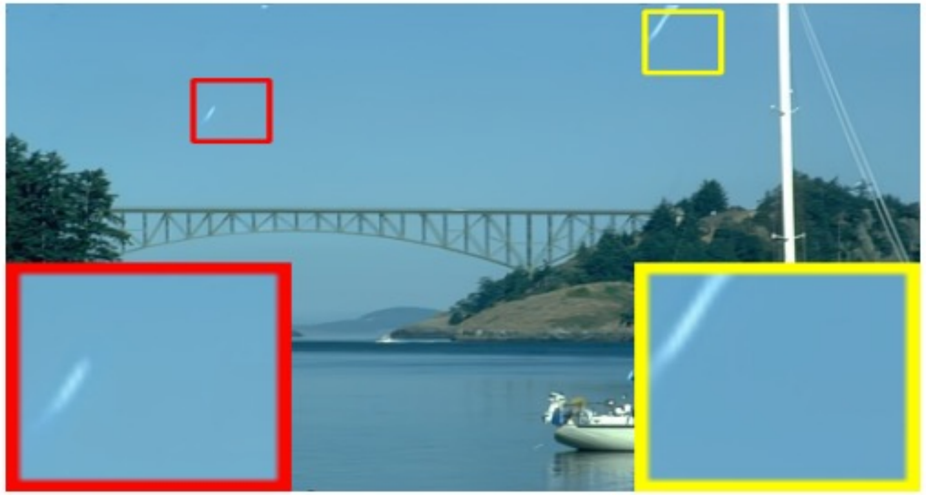
Result (Deraining)



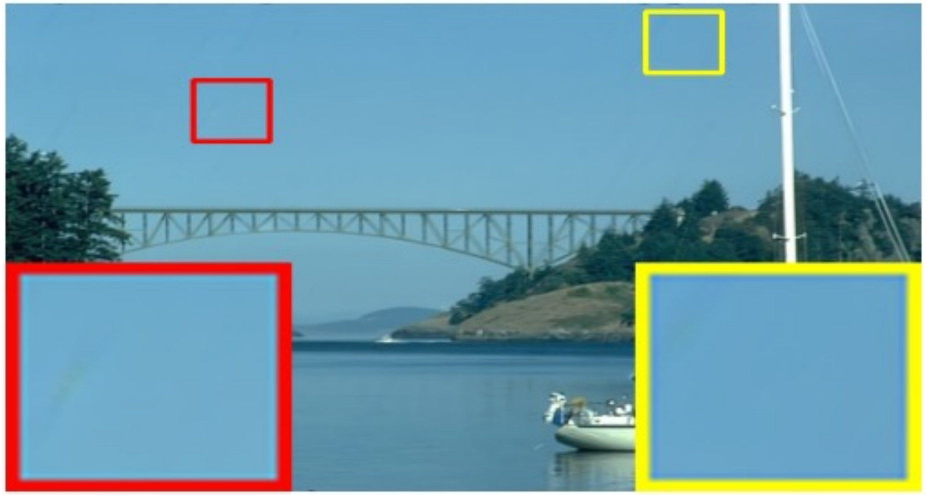
Input



Ground-truth



Restormer



Restormer+Ours

Result (Motion Deblurring)

Method	GoPro		HIDE		RealBlur-R		RealBlur-J	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MPRNet	32.66	0.959	30.96	0.939	35.99	0.952	28.70	0.873
MPRNet+Ours-c	**32.87**	**0.964**	**31.19**	**0.943**	**36.25**	**0.960**	**28.98**	**0.881**
Restormer	32.92	0.961	31.22	0.942	36.19	0.957	28.96	0.879
Restormer+Ours-c	**33.18**	**0.966**	**31.51**	**0.950**	**36.47**	**0.962**	**29.21**	**0.883**

Result (Motion Deblurring)

Input

Ground-truth

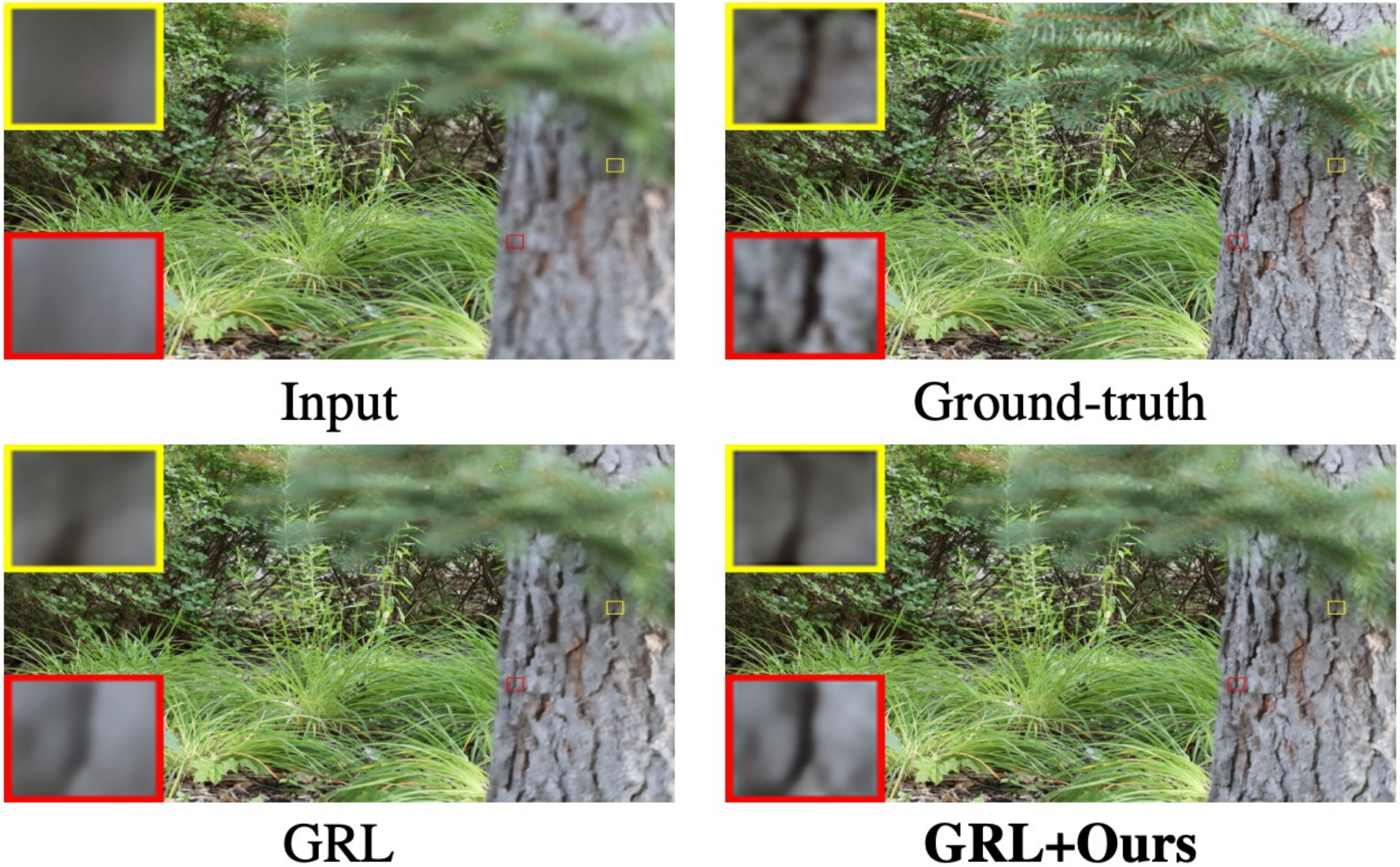
Restormer

Restormer+Ours

Result (Defocus Deblurring)

Method	Indoor Scenes			Outdoor Scenes			Combined		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
IFAN_S	28.11	0.861	0.179	22.76	0.720	0.254	25.37	0.789	0.217
IFAN_S+Ours-c	**28.32**	**0.870**	**0.171**	**23.08**	**0.727**	**0.248**	**25.72**	**0.795**	**0.213**
Restormer_S	28.87	0.882	0.145	23.24	0.743	0.209	25.98	0.811	0.178
Restormer_S+Ours-c	**29.17**	**0.890**	**0.141**	**23.43**	**0.749**	**0.206**	**26.13**	**0.816**	**0.165**
GRL_S-B	29.06	0.886	0.139	23.45	0.761	0.196	26.18	0.822	0.168
GRL_S-B +Ours-c	**29.30**	**0.894**	**0.133**	**23.67**	**0.768**	**0.189**	**26.45**	**0.828**	**0.161**
IFAN_D	28.66	0.868	0.172	23.46	0.743	0.240	25.99	0.804	0.207
IFAN_D+Ours-c	**28.94**	**0.875**	**0.167**	**23.70**	**0.748**	**0.235**	**26.20**	**0.811**	**0.203**
Restormer_D	29.48	0.895	0.134	23.97	0.773	0.175	26.66	0.833	0.155
Restormer_D+Ours-c	**29.79**	**0.902**	**0.131**	**24.23**	**0.778**	**0.155**	**26.89**	**0.840**	**0.153**
GRL_D-B	29.83	0.903	0.114	24.39	0.795	0.150	27.04	0.847	0.133
GRL_D-B+Ours-c	**29.96**	**0.911**	**0.110**	**24.62**	**0.803**	**0.145**	**27.27**	**0.855**	**0.128**

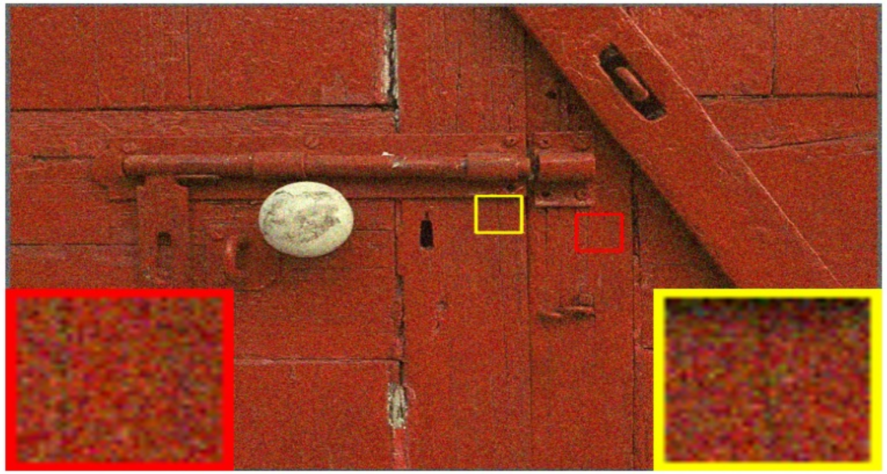
Result (Defocus Deblurring)



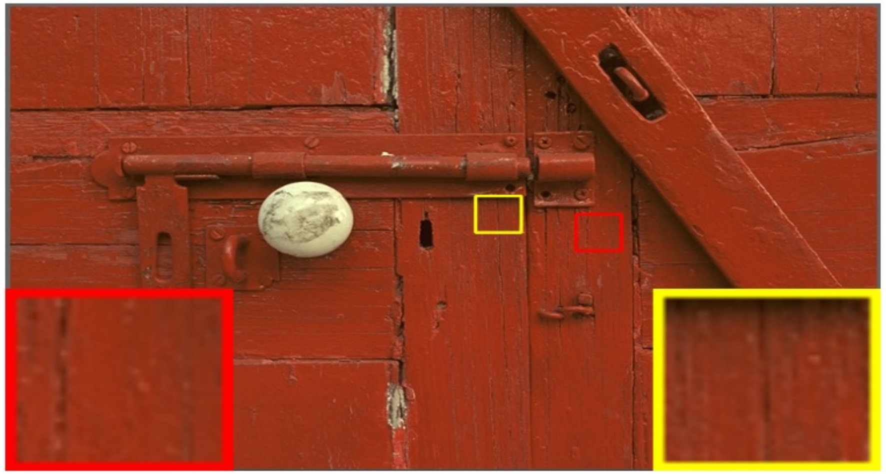
Result (Gaussian Denoising)

Method	Set12			BSD68			Urban100		
	σ=15	σ=25	σ=50	σ=15	σ=25	σ=50	σ=15	σ=25	σ=50
DRUNet	33.25	30.94	27.90	31.91	29.48	26.59	33.44	31.11	27.96
DRUNet+Ours-c	**33.51**	**31.18**	**28.27**	**32.20**	**29.73**	**26.84**	**33.65**	**31.34**	**28.16**
Restormer	33.35	31.04	28.01	31.95	29.51	26.62	33.67	31.39	28.33
Restormer+Ours-c	**33.57**	**31.28**	**28.36**	**32.11**	**29.78**	**26.91**	**33.96**	**31.67**	**28.58**
Restormer	33.42	31.08	28.00	31.96	29.52	26.62	33.79	31.46	28.29
Restormer+Ours-c	**33.70**	**31.29**	**28.35**	**32.24**	**29.81**	**26.86**	**33.97**	**31.73**	**28.58**
GRL-B	33.47	31.12	28.03	32.00	29.54	26.60	34.09	31.80	28.59
GRL-B+Ours-c	**33.74**	**31.30**	**28.37**	**32.29**	**29.76**	**26.91**	**34.22**	**31.95**	**28.74**

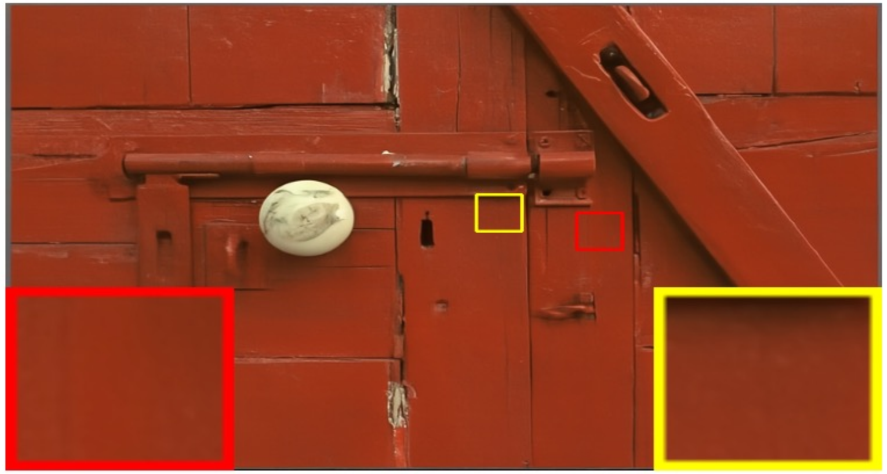
Result (Gaussian Denoising)



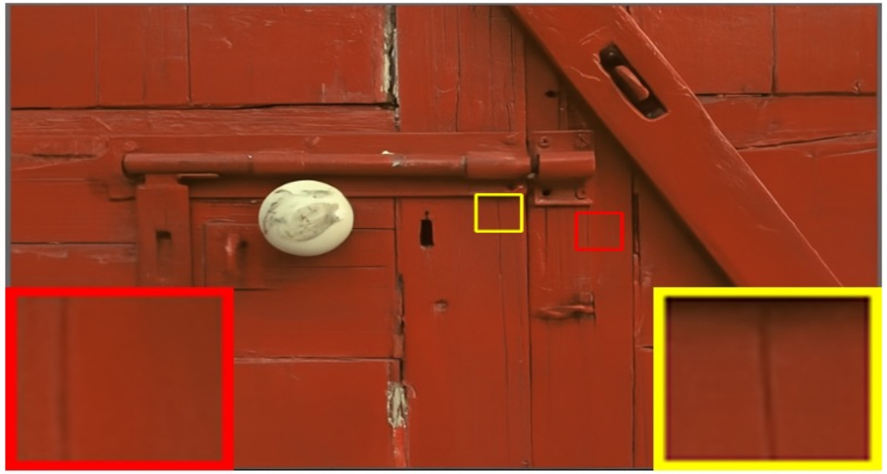
Input



Ground-truth



GRL



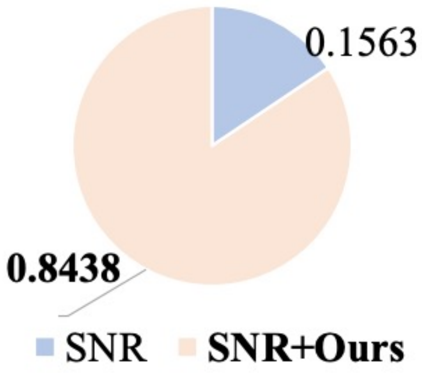
GRL+Ours

Result (Real Denoising)

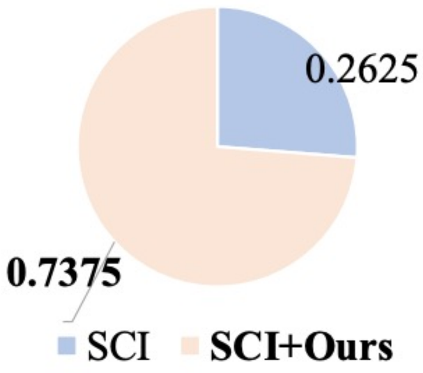
Dataset	**Method**	MPRNet	MPRNet + Ours-c	Uformer	Uformer + Ours-c	Restormer	Restormer + Ours-c
SIDD	PSNR ↑	39.71	**39.93**	39.77	**39.94**	40.02	**40.22**
	SSIM ↑	0.958	**0.961**	0.959	**0.962**	0.960	**0.965**

Input Ground-truth Restormer **+Ours**

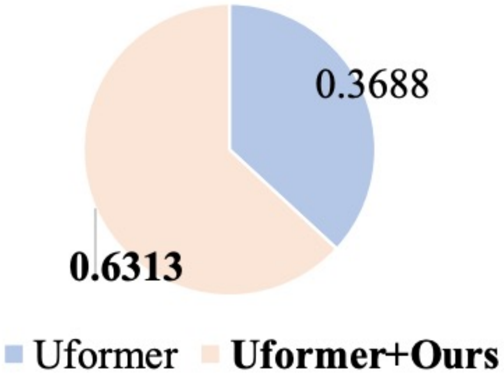
Result (User Study)



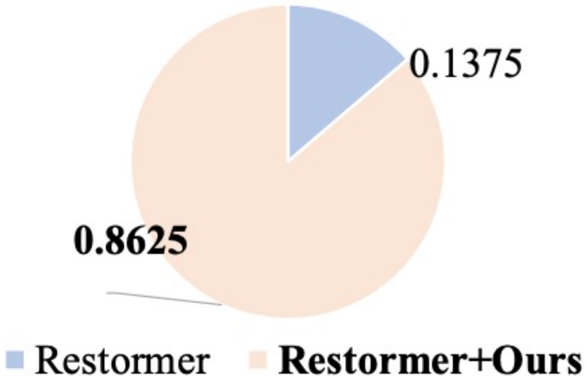
Low-light Enhancement



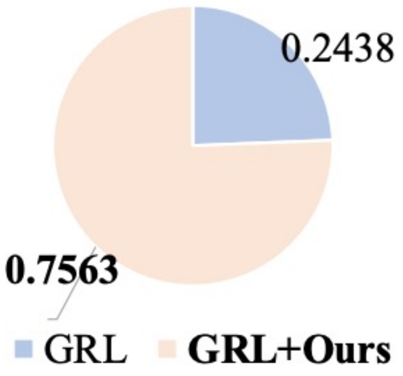
Low-light Enhancement



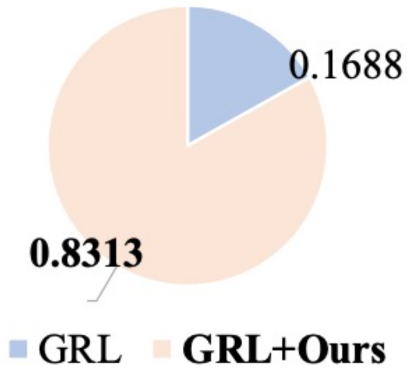
Real-Image Denoising



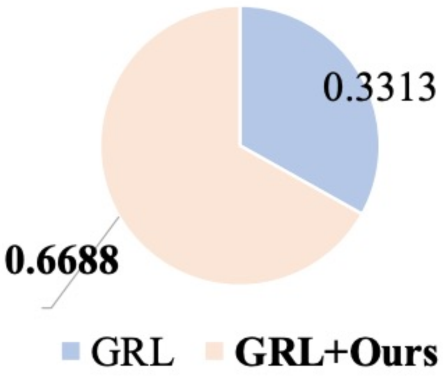
Motion Deblurring



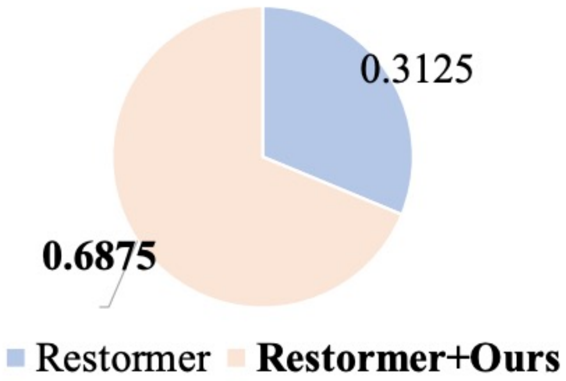
Single-Image Defocus Deblurring



Dual-Pixel Defocus Deblurring



Gaussian Denoising



Deraining

Ablation study

	LOL-real				SID			
	URetinex		SNR		URetinex		SNR	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
w/o SP, with CA and SA	23.45	0.868	24.25	0.886	21.98	0.619	23.02	0.620
with SP, w/o CA, with SA	22.10	0.856	24.05	0.875	22.05	**0.623**	22.93	0.624
with SP and CA, w/o SA	23.76	0.850	23.86	0.879	21.92	0.620	23.07	0.621
Large $\mathcal{R}$ w/o SP/CA/SA	22.74	0.857	24.51	0.881	22.06	0.621	23.04	0.627
w/o Position Embedding $\mathcal{S}$	23.66	0.843	24.13	0.874	22.13	0.620	22.92	0.622
SNR Value as Mask	22.66	0.855	24.77	0.887	22.01	0.617	22.94	0.627
Use 1D Priors via Con.	23.01	0.853	23.83	0.878	22.07	0.622	22.93	0.628
Use 2D Priors via Con.	22.68	0.862	24.11	0.880	22.08	0.618	23.06	0.625
Full Setting	**24.70**	**0.878**	**25.50**	**0.892**	**22.34**	**0.623**	**23.34**	**0.630**

Outline

- Introduction
- Framework
- Method
- Experiment
- Conclusion

Conclusion

- Explore the utilization of **features from a pre-trained model** to enhance the performance of a restoration model.

- Introduce a novel refinement module **PTG-RM** that employs PTG-SVE and PTG-CSA mechanisms, which focus on formulating **optimal operation ranges** and **attention strategies** guided by the pre- trained features.

- Demonstrate the effectiveness and generalization ability of this approach.