THU-AM-172

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

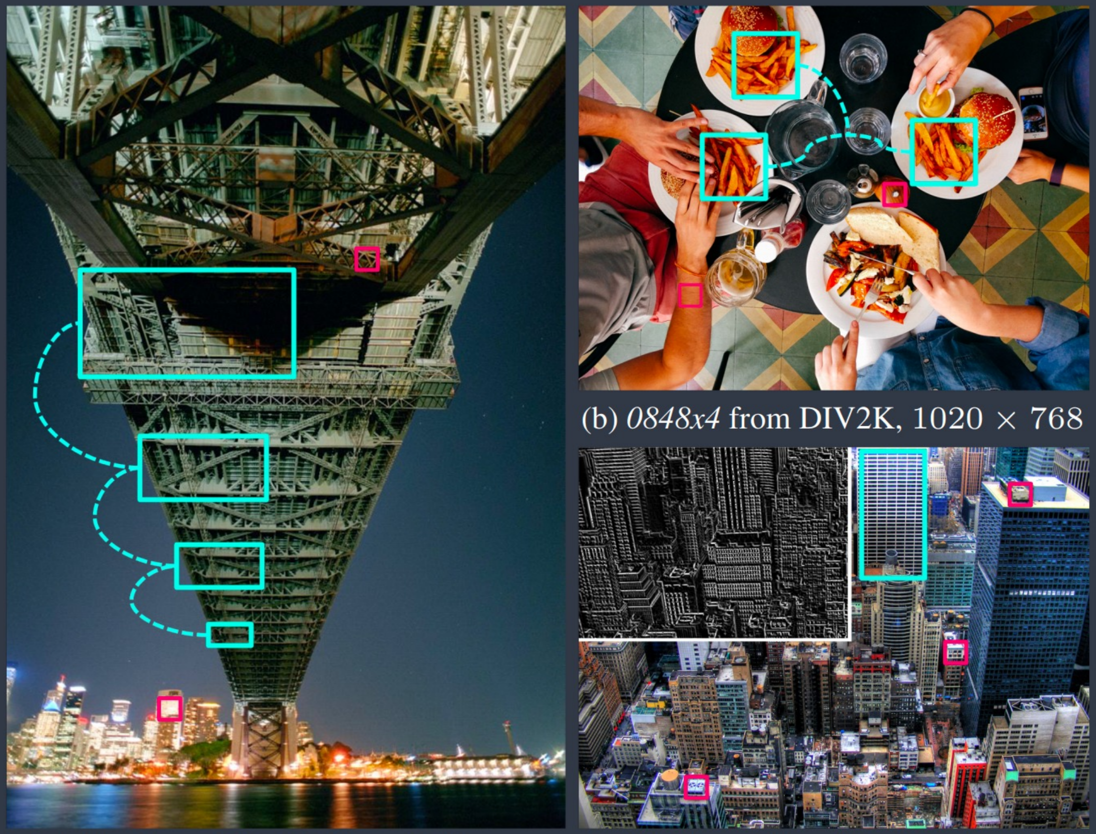
Efficient and Explicit Modelling of Image Hierarchies for Image Restoration

Yawei Li[1], Yuchen Fan[2], Xiaoyu Xiang[2], Denis Demandolx[2],
Rakesh Ranjan[2], Radu Timofte[1,3], Luc Van Gool[1,4]

ETH zürich Meta Julius-Maximilians-UNIVERSITÄT WÜRZBURG KU LEUVEN

Background

We propose to model image hierarchies efficiently and explicitly.



(a) *bridge* from ICB, 2749 × 4049

(b) *0848x4* from DIV2K, 1020 × 768

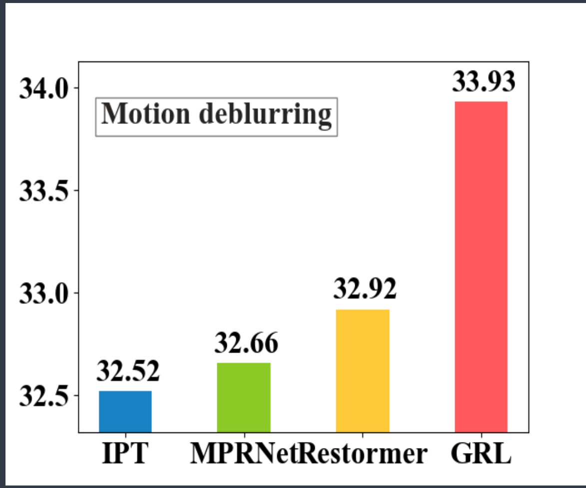
(c) *073* from Urban100, 1024 × 765

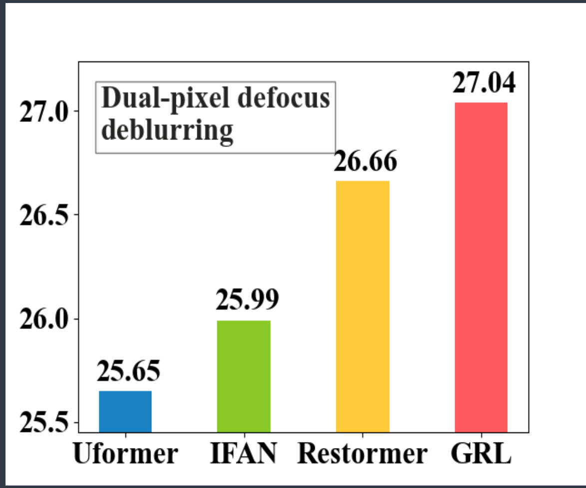
Global: multi-scale pattern repetition, same scale texture similarity

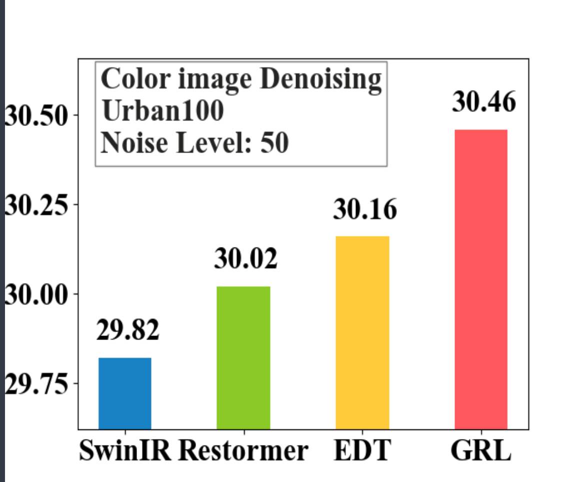
Regional: structure in a patch

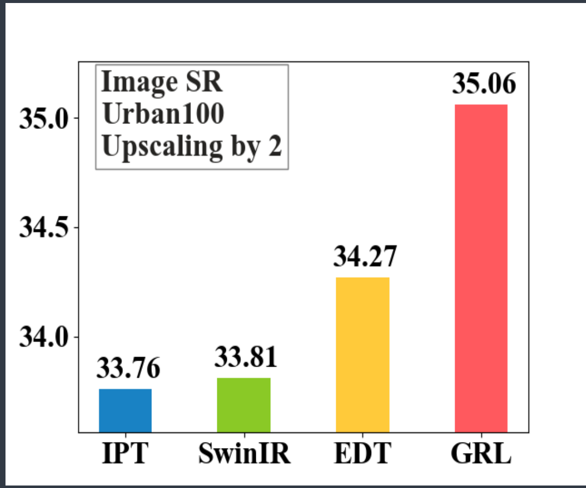
Local: edges

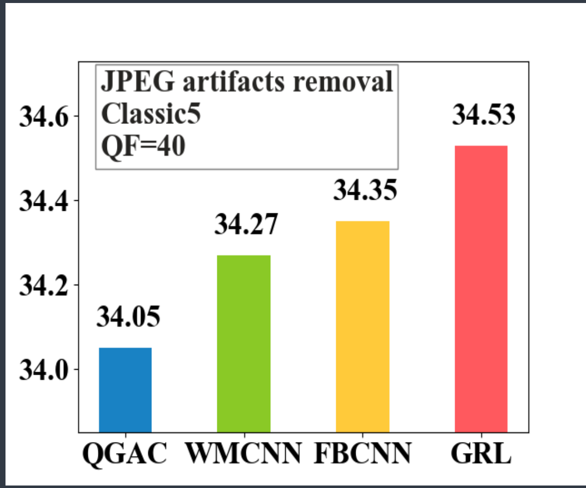
Performance improvement

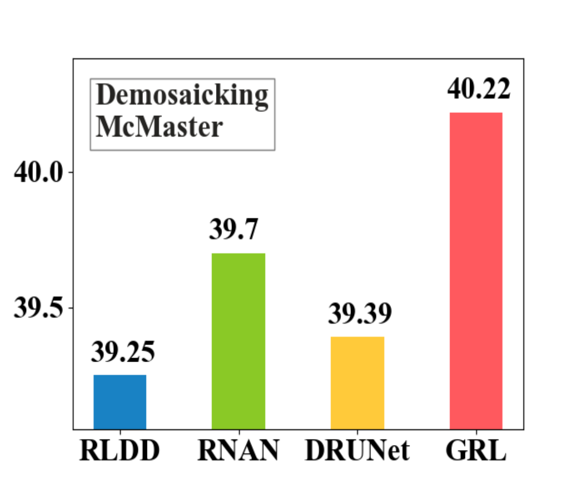












Challenges: Global range information modelling

1. Existing image restoration networks based on convolutions and window attention could not capture long-range dependencies explicitly by using a single computational module.

2. The increasing resolution of today's images poses a challenge for long-range dependency modelling

Research Questions

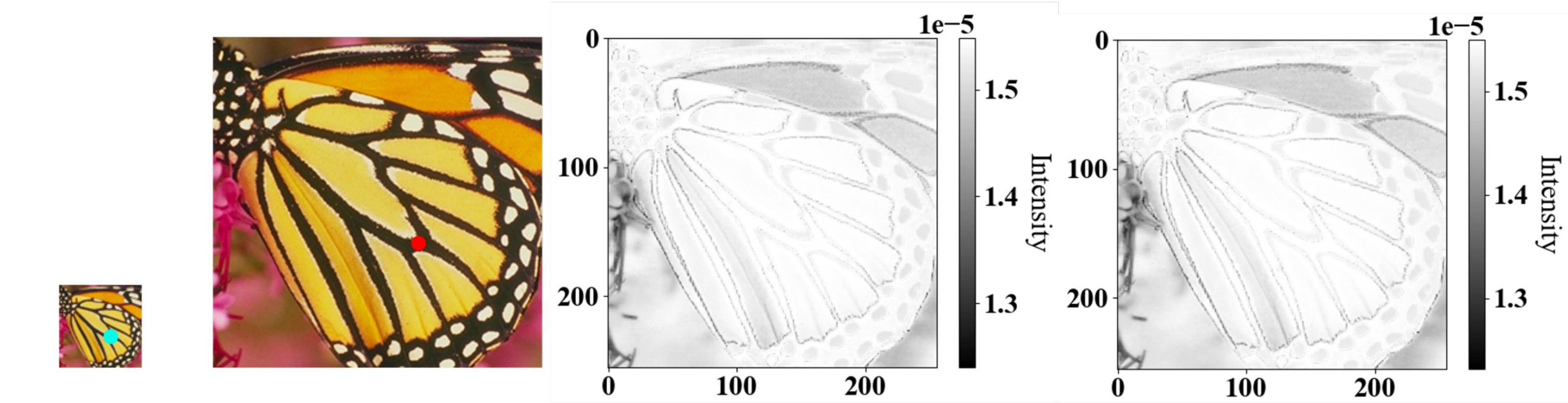
1. How to efficiently model global range features in high-dimensional images for image restoration;

2. How to model image hierarchies (local, regional, global) explicitly by a single computational module for high-dimensional image restoration;

3. How can this joint modelling lead to a uniform performance improvement for different image restoration tasks.

Motivation I: Cross-scale similarity

Images at different scales have similar structures



(a) LR image (b) HR image (c) Attention between the cyan and other pixels in (b). (d) Attention between the red pixel and other pixels in (b).

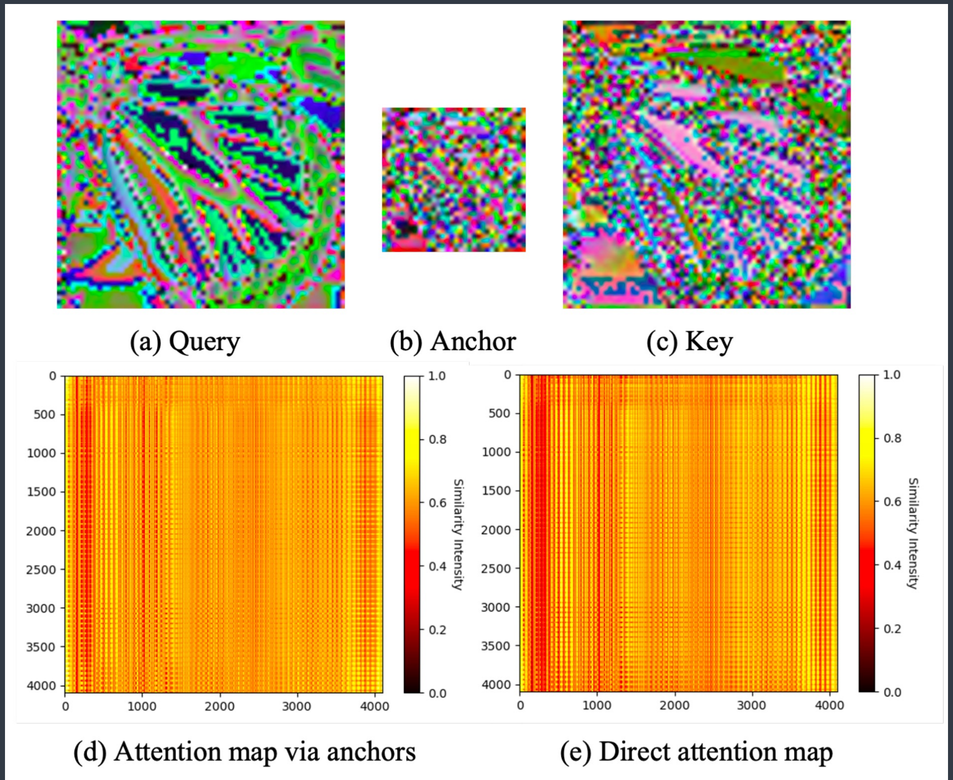
Motivation I: Cross-scale similarity

Images at different scales have similar structures

$$\mathbf{Y} = \mathbf{M}_e \cdot \mathbf{Z} = \mathbf{M}_e \cdot (\mathbf{M}_d \cdot \mathbf{V}),$$

$$\mathbf{M}_d = \mathrm{Softmax}(\mathbf{A} \cdot \mathbf{K}^T/\sqrt{d}),$$

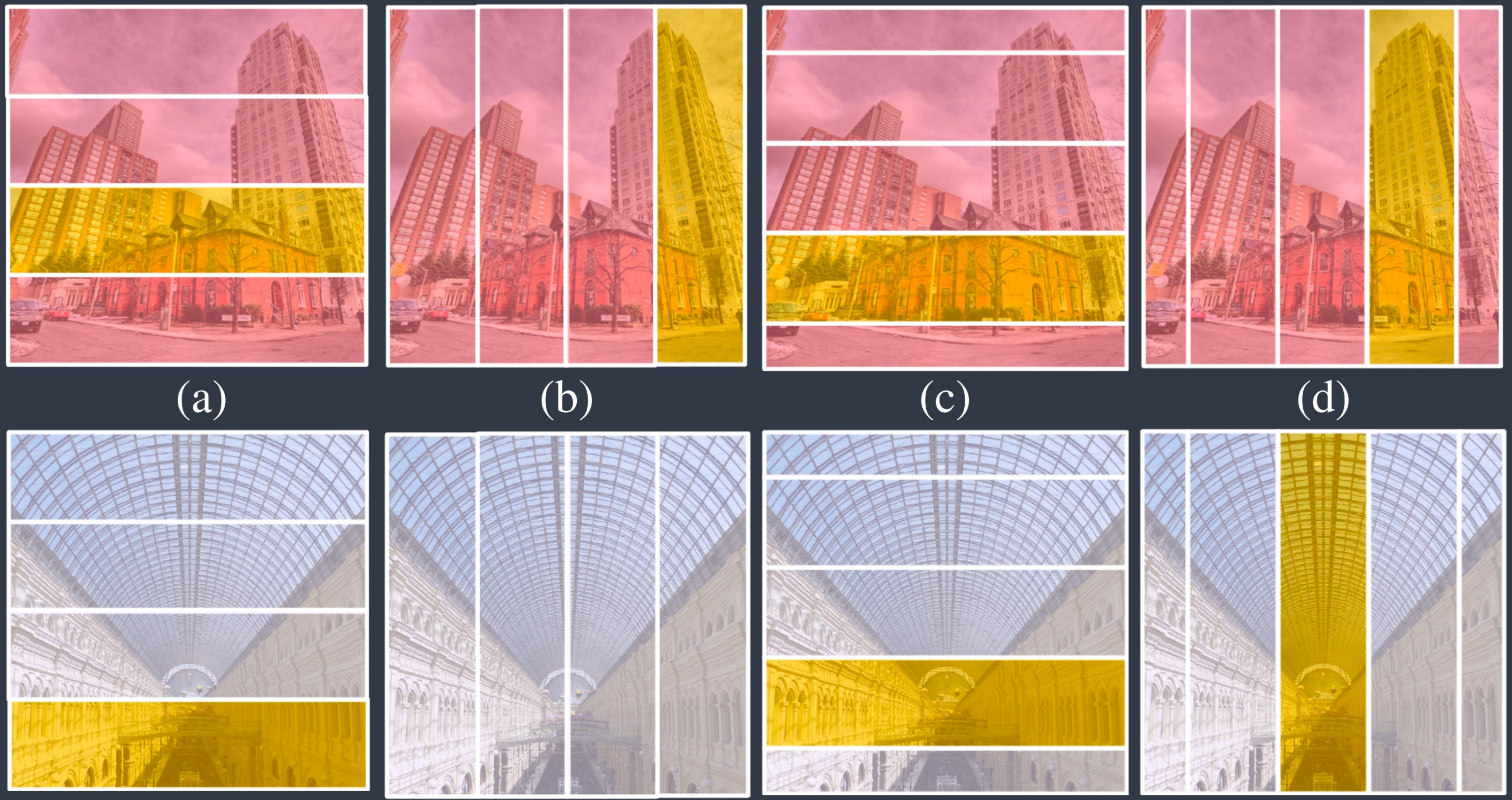
$$\mathbf{M}_e = \mathrm{Softmax}(\mathbf{Q} \cdot \mathbf{A}^T/\sqrt{d}),$$



Pearson correlation coefficients (0.9505)

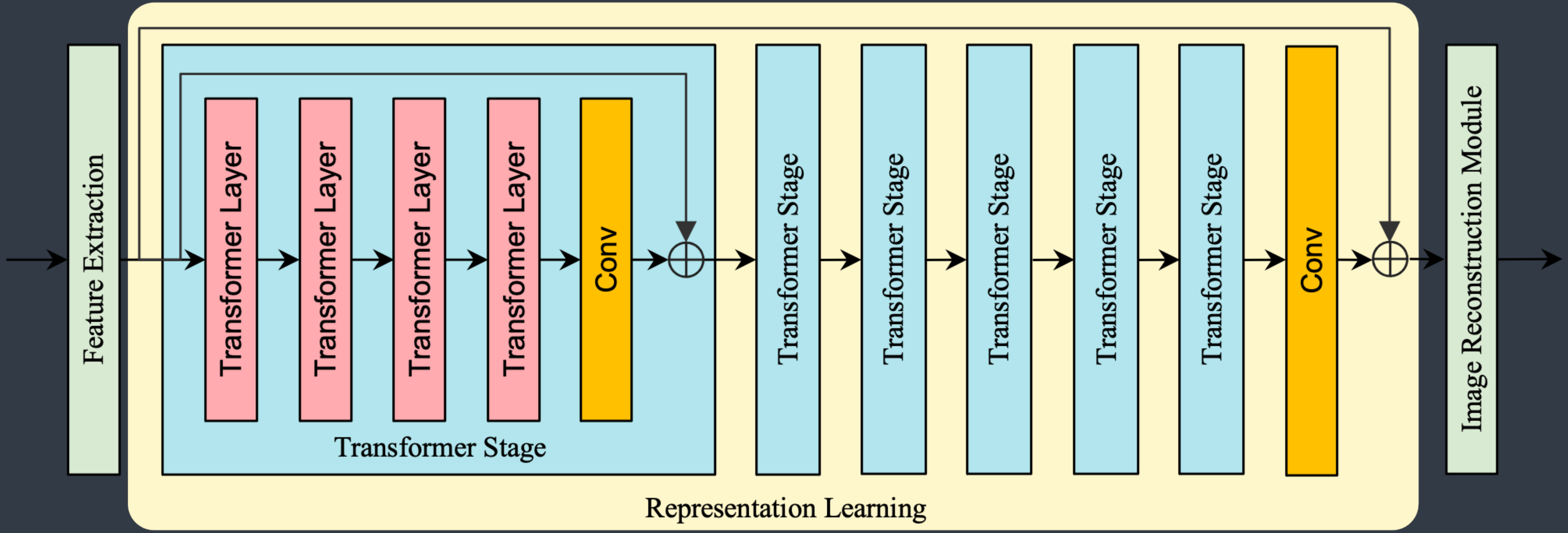
Motivation II: Anisotropic image features

The natural images contains anisotropic contents.


(a) (b) (c) (d)
(e) (f) (g) (h)

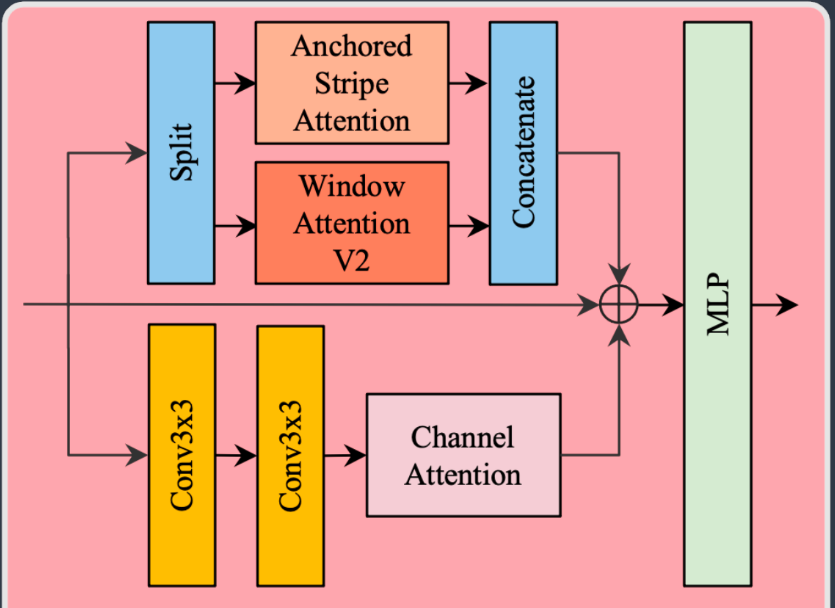
Network architecture

The representation learning module contains stages of transformer layers.

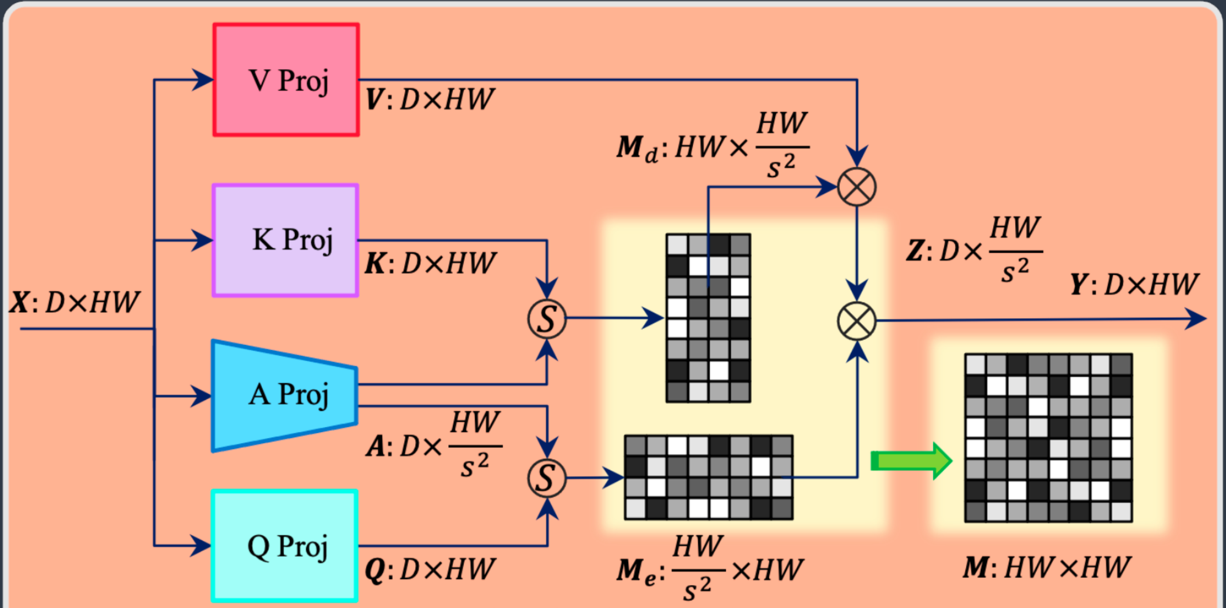


Network architecture

The transformer layer is equipped with global, regional, and local modelling blocks.



(b) Transformer Layer



(c) Anchored Stripe Attention

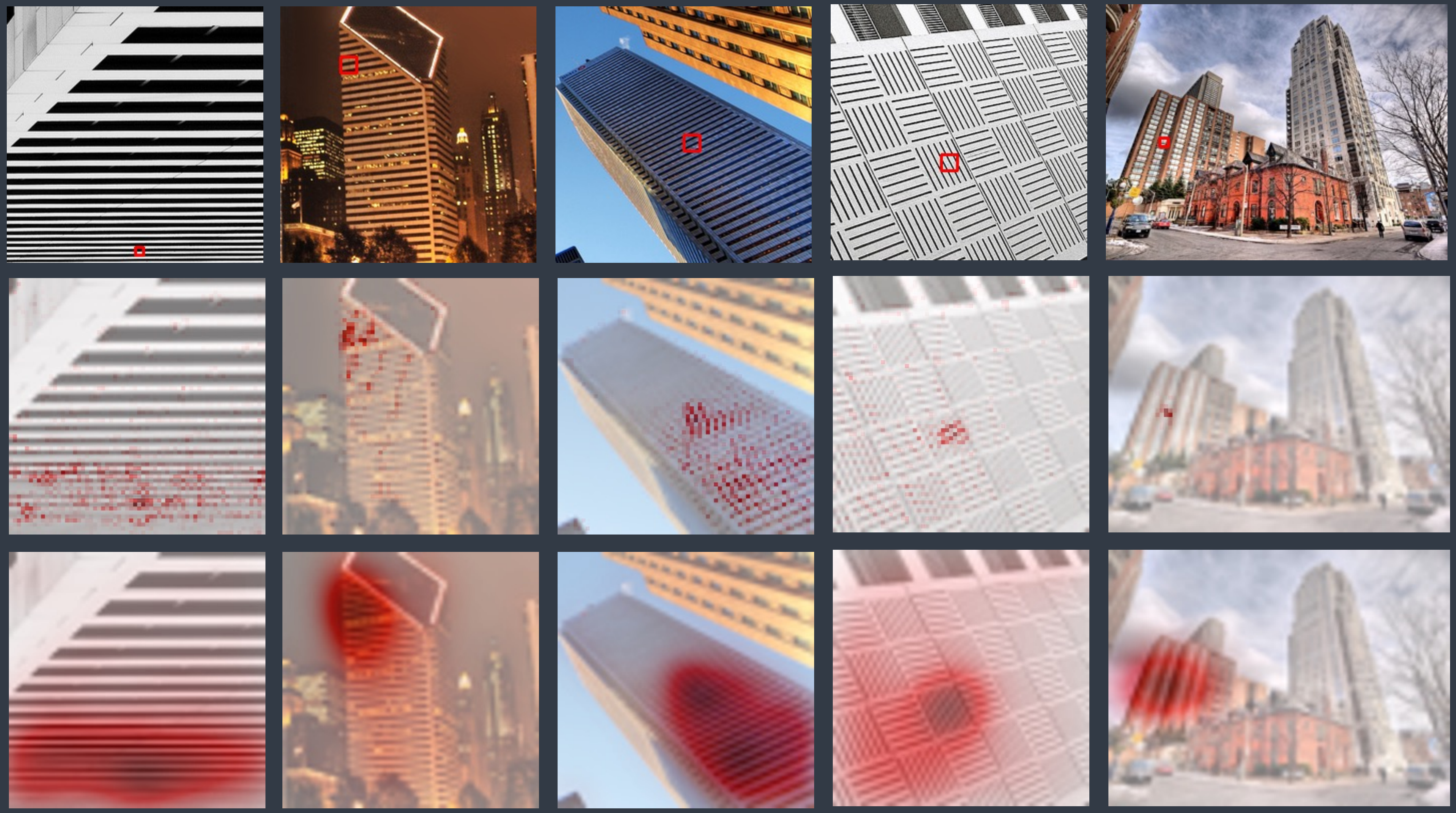
Analysis

The proposed method can utilize information beyond local range.

Input

LAM Attribution

Area of Contribution



Results

State-of-the-art performance.
More efficient.

Table 7. ***Color image denoising*** results. Model complexity and prediction accuracy are shown for better comparison.

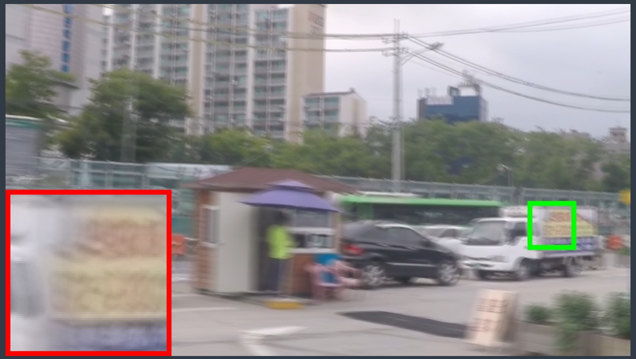
Method	**# Params** [M]	**Urban100** [28] σ=15	σ=25	σ=50
DRUNet [80]	32.64	34.81	32.60	29.61
SwinIR [43]	11.75	35.13	32.90	29.82
Restormer [76]	26.13	35.13	32.96	30.02
GRL-T	0.88	35.08	32.84	29.78
GRL-S	3.12	35.24	33.07	30.09
GRL-B	19.81	35.54	33.35	30.46

Results

1dB improvement over Restormer on GoPro.

Table 4. ***Single-image motion deblurring*** results. GoPro dataset [51] is used for training.

Method	GoPro [51] PSNR↑ / SSIM↑	HIDE [59] PSNR↑ / SSIM↑	Average PSNR↑ / SSIM↑
MIMO-UNet+ [8]	32.45 / 0.957	29.99 / 0.930	31.22 / 0.944
IPT [5]	32.52 / -	- / -	- / -
MPRNet [77]	32.66 / 0.959	30.96 / 0.939	31.81 / 0.949
Restormer [76]	32.92 / 0.961	31.22 / 0.942	32.07 / 0.952
GRL-B (ours)	33.93 / 0.968	31.65 / 0.947	32.79 / 0.958



Blurred

MPRNet

Restormer

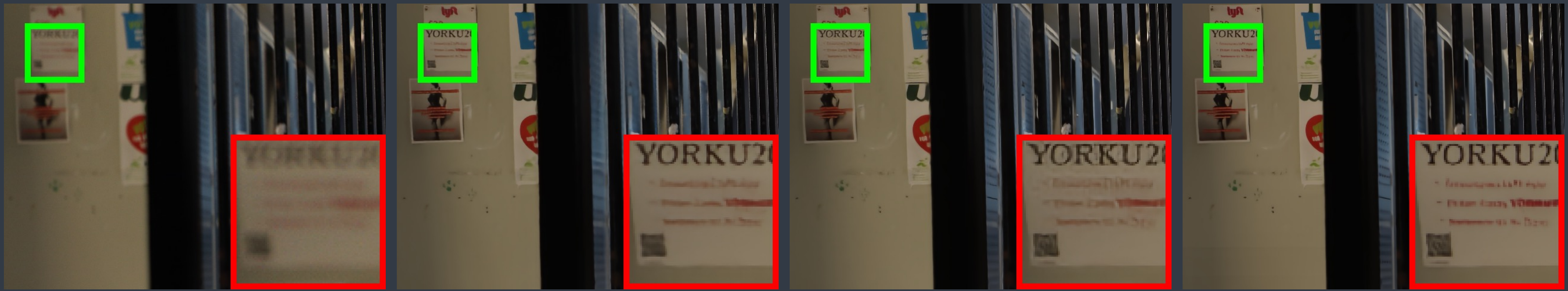
GRL (ours)

Results

Significant improvement for defocus deblurring

Table 2. ***Defocus deblurring*** results. **S:** single-image defocus deblurring. **D:** dual-pixel defocus deblurring.

Method	Indoor Scenes				Outdoor Scenes				Combined			
	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓	PSNR↑	SSIM↑	MAE↓	LPIPS↓
DPDNet$_D$ [1]	27.48	0.849	0.029	0.189	22.90	0.726	0.052	0.255	25.13	0.786	0.041	0.223
RDPD$_D$ [2]	28.10	0.843	0.027	0.210	22.82	0.704	0.053	0.298	25.39	0.772	0.040	0.255
Uformer$_D$ [74]	28.23	0.860	0.026	0.199	23.10	0.728	0.051	0.285	25.65	0.795	0.039	0.243
IFAN$_D$ [41]	28.66	0.868	0.025	0.172	23.46	0.743	0.049	0.240	25.99	0.804	0.037	0.207
Restormer$_D$ [76]	29.48	0.895	0.023	0.134	23.97	0.773	0.047	0.175	26.66	0.833	0.035	0.155
GRL$_D$-B	29.83	0.903	0.022	0.114	24.39	0.795	0.045	0.150	27.04	0.847	0.034	0.133



Blurred Uformer Restormer GRL (ours)

Results

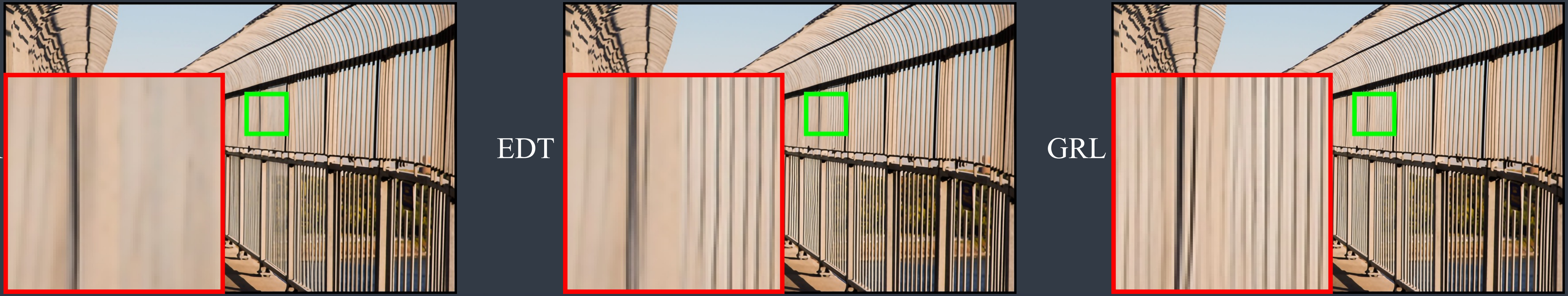


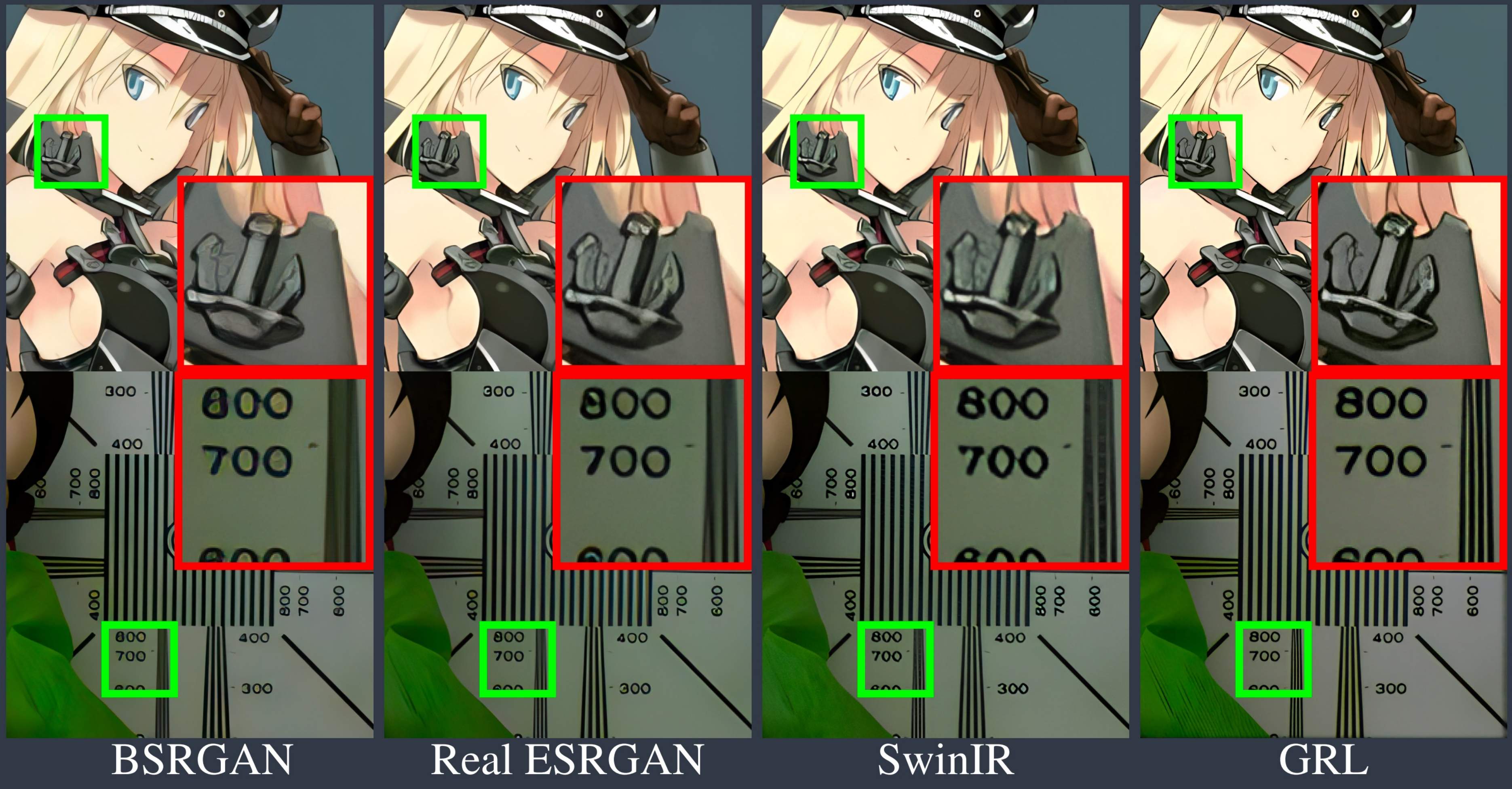
Table 10. ***Classical image SR*** results. Results of both lightweight models and accurate models are summarized.

Method	**Scale**	**# Params** [M]	**Set5** [3]		**Set14** [78]		**BSD100** [49]		**Urban100** [28]		**Manga109** [50]	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
RCAN [84]	×4	15.59	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
SAN [11]	×4	15.86	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169
HAN [52]	×4	64.20	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094	31.42	0.9177
SwinIR [43]	×4	0.90	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
EDT [42]	×4	0.92	32.53	0.8991	28.88	0.7882	27.76	0.7433	26.71	0.8051	31.35	0.9180
GRL-T (ours)	×4	0.91	32.56	0.9029	28.93	0.7961	27.77	0.7523	27.15	0.8185	31.57	0.9219
IPT [5]	×4	115.63	32.64	-	29.01	-	27.82	-	27.26	-	-	-
GRL-S (ours)	×4	3.49	32.76	0.9058	29.10	0.8007	27.90	0.7568	27.90	0.8357	32.11	0.9267
SwinIR [43]	×4	11.90	32.92	0.9044	29.09	0.7950	27.92	0.7489	27.45	0.8254	32.03	0.9260
EDT [42]	×4	11.63	33.06	0.9055	29.23	0.7971	27.99	0.7510	27.75	0.8317	32.39	0.9283
GRL-B (ours)	×4	20.20	33.10	0.9094	29.37	0.8058	28.01	0.7611	28.53	0.8504	32.77	0.9325

Results

Better visual quality.

Figure 7. Visual results for real-world image SR.



Conclusion

1. Inspired by two image properties (cross-scale similarity and anisotropic image features), anchored stripe self-attention module is proposed for efficient long-range dependency modelling.

2. Based on the new attention module, a network is proposed to efficiently and explicitly model image hierarchies in the global, regional, and local ranges.

3. Owing to the advanced computational mechanism, the proposed network architecture achieves state-of-the-art performances for various image restoration tasks.

Thanks for your attention!