Stochastic Window Transformer for Image Restoration

Jie Xiao, Xueyang Fu, Feng Wu, Zheng-Jun Zha

University of Science and Technology of China

Source code: https://github.com/jiexiaou/Stoformer

Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022)

National Engineering Laboratory for Brain-Inspired Intelligence Technology and Application

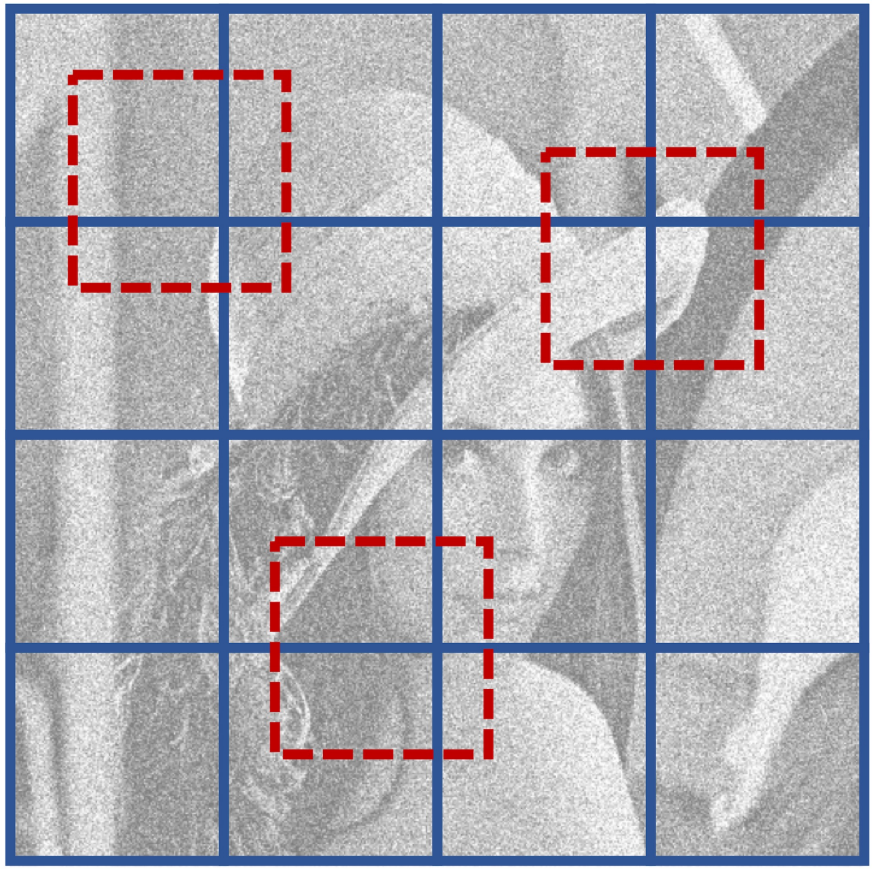
CONTENT

Fixed Window Partition

Translation equivalence ✗

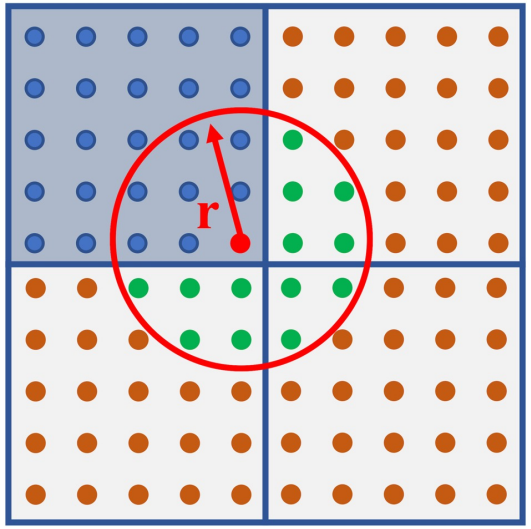
↓

Breaking the translation invariance

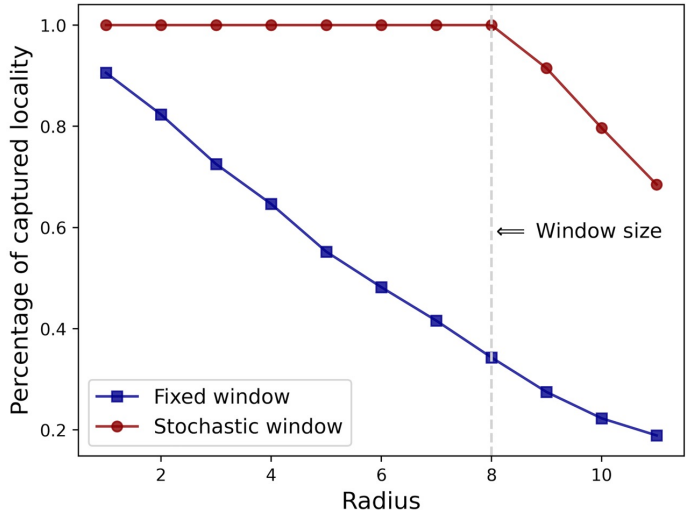


Fixed Window Partition

Intact Local Relationship? ✗

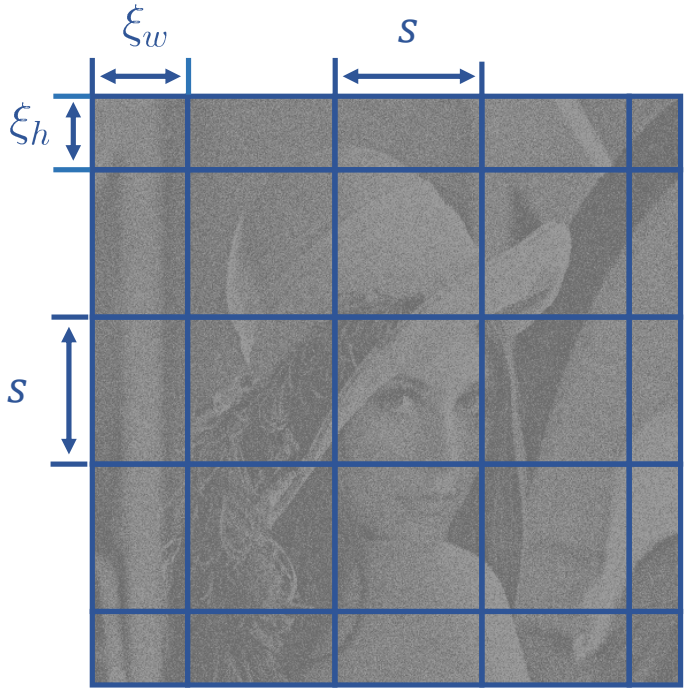


The fixed local window cannot faithfully capture local relationships.



Percentage of captured locality between the fixed and stochastic window strategy.

Refuse infinite favoritism



$$z_l = \mathrm{SA}(\mathrm{Par}(x_{l-1}; s, 0, 0)) + x_{l-1},$$
$$x_l = \mathrm{MLP}(z_l) + z_l,$$
$$z_{l+1} = \mathrm{SA}(\mathrm{Par}(x_l; s, \frac{s}{2}, \frac{s}{2})) + x_l,$$
$$x_{l+1} = \mathrm{MLP}(z_{l+1}) + z_{l+1},$$

目录 CONTENT

Stochastic Window Strategy

Treating all the local window fairly

$$z_l = \mathrm{SA}(\mathrm{Par}(x_{l-1}; s, \xi_h^l, \xi_w^l)) + x_{l-1}, \quad (\xi_h^l, \xi_w^l) \sim \mathbb{U}(\Re_s),$$
$$x_l = \mathrm{MLP}(z_l) + z_l,$$

Uniform distribution

$$\Re_s := [0, \ldots, s-1] \times [0, \ldots, s-1],$$

Layer Expectation Propagation

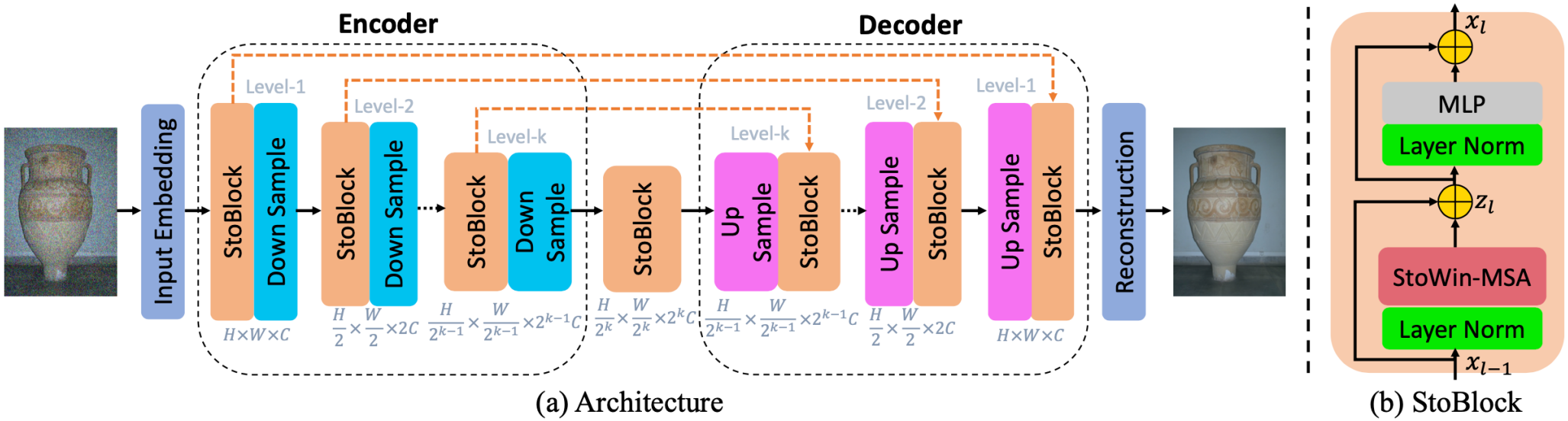
Aggregating expected feature per layer

$$
\begin{aligned}
F(x)^{\text{test}} &= \sum_{\xi_h^0,\xi_w^0,\ldots,\xi_h^{N-1},\xi_w^{N-1}} F(x;\xi_h^0,\xi_w^0,\ldots,\xi_h^{N-1},\xi_w^{N-1})\mathbb{U}(\xi_h^0,\xi_w^0,\ldots,\xi_h^{N-1},\xi_w^{N-1}) \\
&= \sum_{(\xi_h^{N-1},\xi_w^{N-1})} \cdots \sum_{(\xi_h^0,\xi_w^0)} F(x;\xi_h^0,\xi_w^0,\ldots,\xi_h^{N-1},\xi_w^{N-1})\mathbb{U}(\xi_h^0,\xi_w^0)\cdots\mathbb{U}(\xi_h^{N-1},\xi_w^{N-1}),
\end{aligned}
$$

$$
\begin{aligned}
z_l^{\text{test}} &= \sum_{(\xi_h^l,\xi_w^l)} \text{SA}(\text{Par}(x_{l-1}^{\text{test}};s,\xi_h^l,\xi_w^l))\mathbb{U}(\xi_h^l,\xi_w^l) + x_{l-1}^{\text{test}} \\
&= \mathop{\mathbb{E}}_{(\xi_h^l,\xi_w^l)\sim\mathbb{U}} \left[\text{SA}(\text{Par}(x_{l-1}^{\text{test}};s,\xi_h^l,\xi_w^l))\right] + x_{l-1}^{\text{test}}, \\
x_l^{\text{test}} &= \text{MLP}(z_l^{\text{test}}) + z_l^{\text{test}}.
\end{aligned}
$$

Layer-wise expectation

Network



(a) Architecture

(b) StoBlock

目录 CONTENT

❸ | RESULTS & DISCUSSION

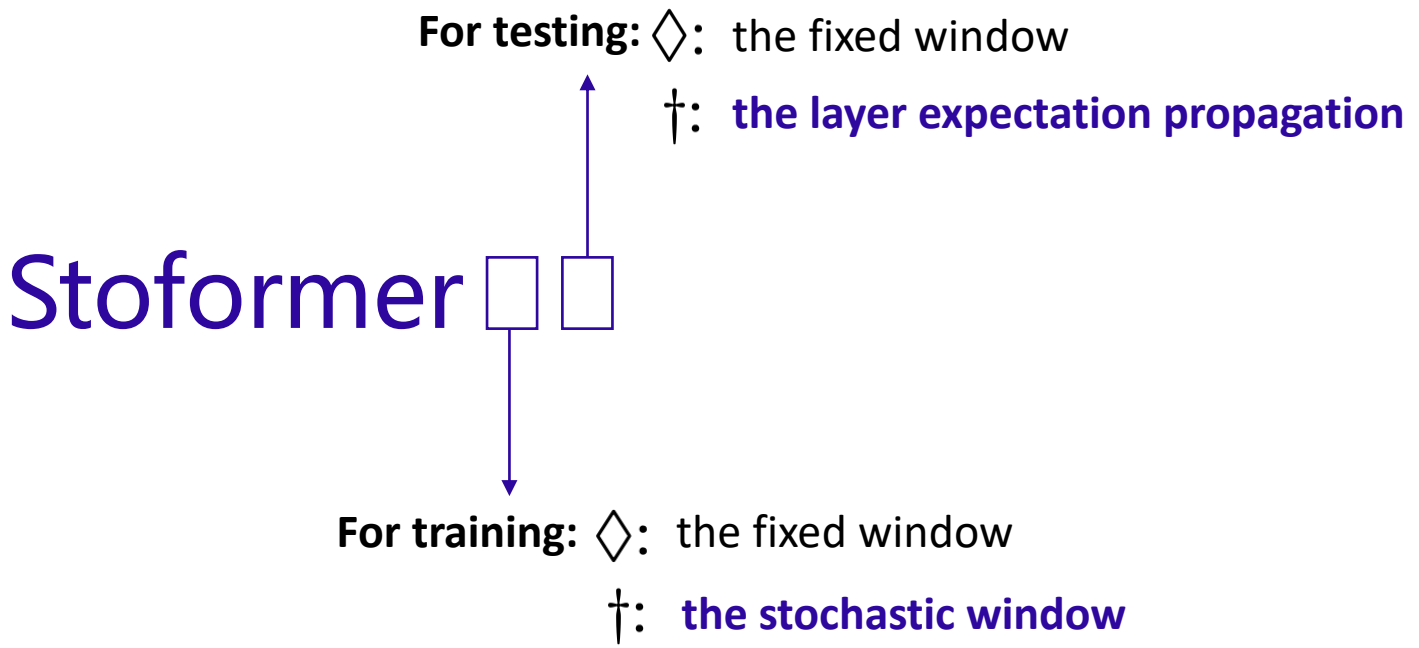


Image Deraining

Method	SPA-Data	
	PSNR	SSIM
GMM [30]	34.30	0.9428
DDN [13]	36.97	0.9604
SPANet [49]	40.24	0.9811
JORDER-E [55]	40.78	0.9811
RCDNet [48]	41.47	0.9834
SPAR [38]	44.10	0.9872
Uformer [52]	47.84	0.9925
Stoformer◊◊	47.80	0.9925
Stoformer◊†	46.95	0.9917
Stoformer†◊	48.85	0.9935
Stoformer††	48.97	0.9938

Image Denoising

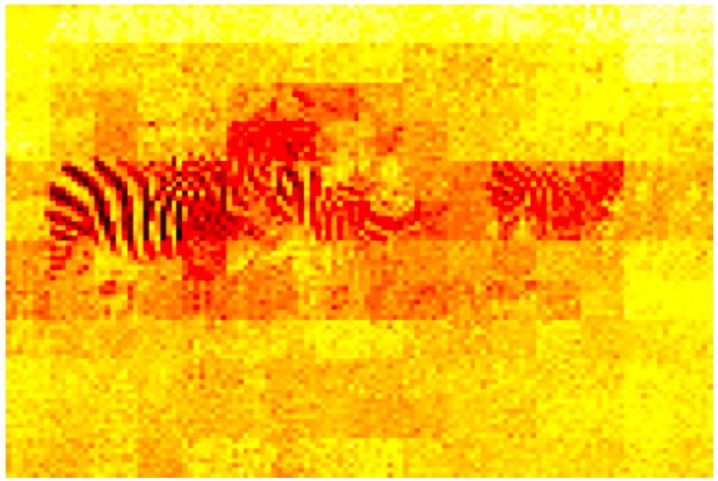
Method	CBSD68			Kodak24			McMaster			Urban100		
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
IRCNN [64]	33.86	31.16	27.86	34.69	32.18	28.93	34.58	32.18	28.91	33.78	31.20	27.70
FFDNet [65]	33.87	31.21	27.96	34.63	32.13	28.98	34.66	32.35	29.18	33.83	31.40	28.05
DnCNN [63]	33.90	31.24	27.95	34.60	32.14	28.95	33.45	31.52	28.62	32.98	30.81	27.59
VDN [57]	33.90	31.35	28.19	-	-	-	-	-	-	-	-	-
FuncNet [34]	34.28	-	-	35.25	-	-	-	-	-	-	-	-
DRUNet [67]	34.30	31.69	28.51	35.31	32.89	29.86	35.40	33.14	30.08	34.81	32.60	29.61
Restormer [59]	34.39	31.78	28.59	35.44	33.02	30.00	35.55	33.31	30.29	35.06	32.91	30.02
Stoformer◊◊	34.34	31.73	28.52	35.32	32.91	29.83	35.53	33.35	30.34	35.04	32.83	29.66
Stoformer◊†	34.30	31.73	28.50	35.22	32.90	29.80	35.40	33.22	30.18	35.00	32.78	29.61
Stoformer†◊	35.10	32.40	29.13	35.50	33.08	30.00	36.00	33.83	30.80	35.37	33.14	30.00
Stoformer††	35.13	32.47	29.16	35.53	33.12	30.03	36.03	33.86	30.84	35.42	33.19	30.06

Image Deblurring

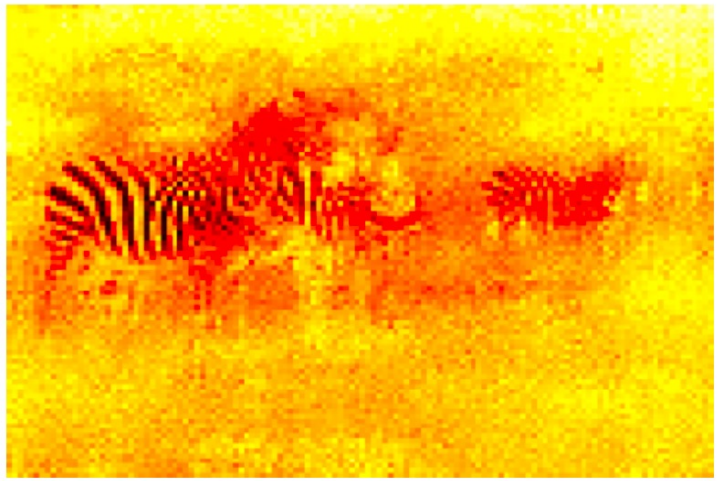
Method	GoPro		HIDE	
	PSNR	SSIM	PSNR	SSIM
Nah *et al.* [37]	29.08	0.914	25.73	0.874
DeblurGAN [22]	28.70	0.858	24.51	0.871
DeblurGAN-v2 [23]	29.55	0.934	26.61	0.875
DBGAN [66]	31.10	0.942	28.94	0.915
IPT [2]	32.52	-	-	-
MPRNet [60]	32.66	0.959	30.96	0.939
SPAIR [38]	32.06	0.953	30.29	0.931
Stoformer◊◊	32.80	0.959	30.73	0.937
Stoformer◊†	31.62	0.950	29.94	0.928
Stoformer†◊	33.17	0.963	30.91	0.940
Stoformer††	33.24	0.964	30.99	0.941

Eliminating Blocking Artifacts

(a) Input

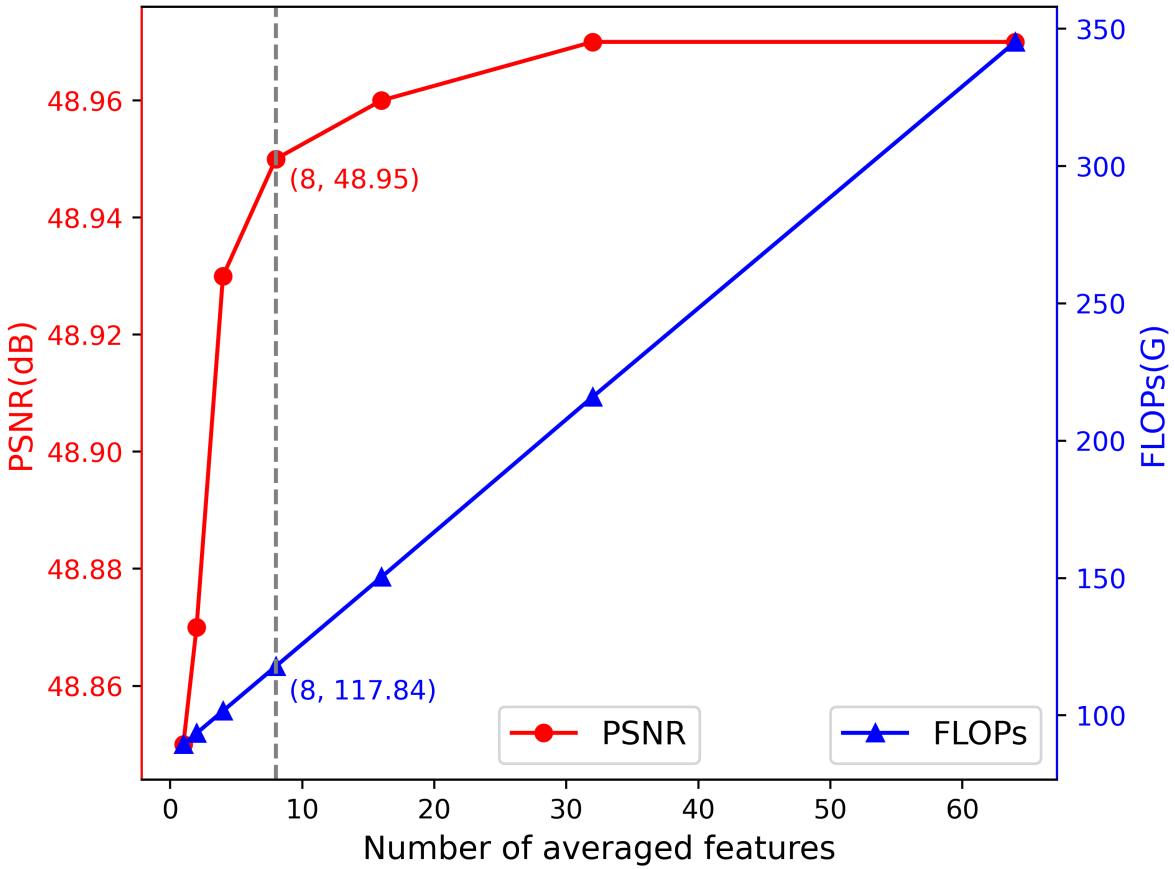


(b) Fixed window strategy

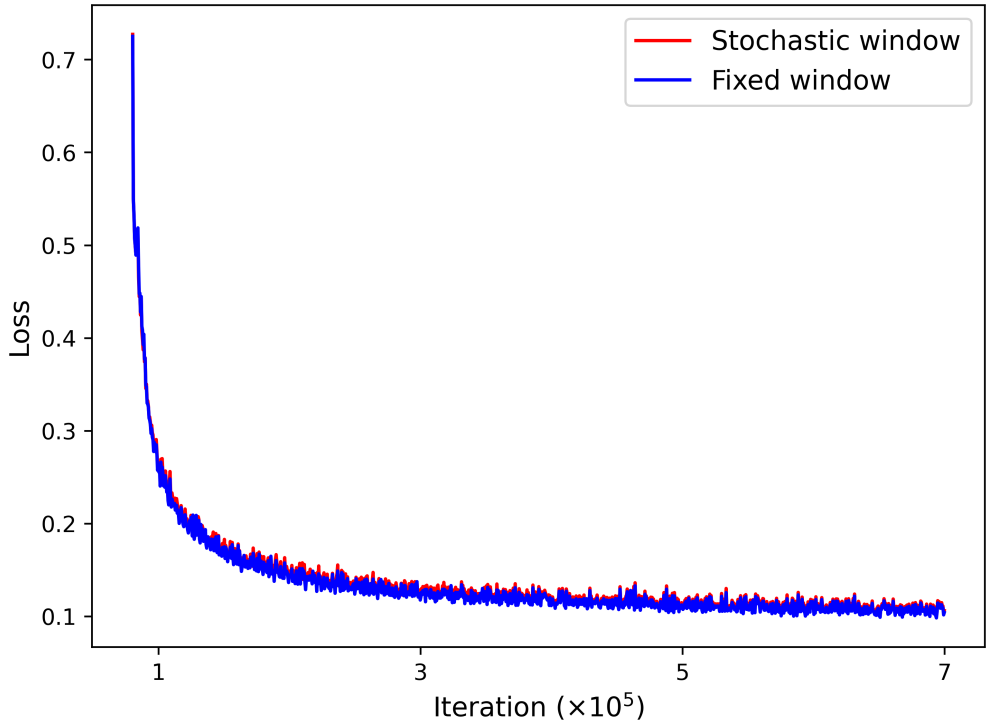


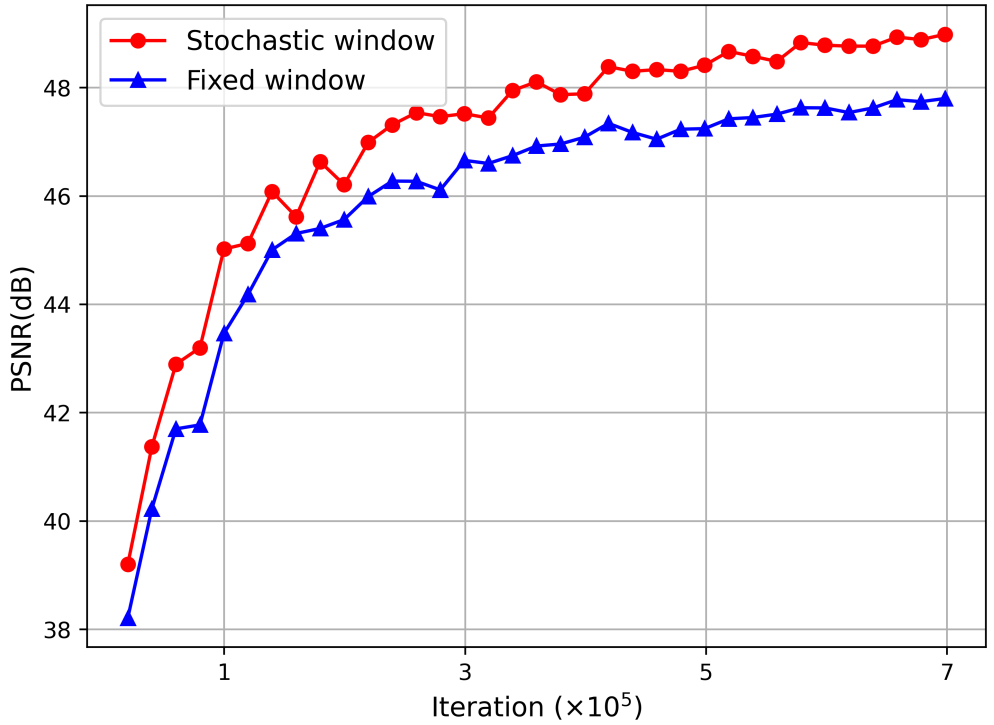
(c) Stochastic window strategy

Trade-off in Layer Expectation Propagation



Boosting Model Generalization





Extension?

More restoration tasks

High-level tasks where local attention is need

目录 CONTENT

4 | CONCLUSION

1. Translation invariance breaking and loss of local relationships in existing transformer-based image restoration approaches.

2. We propose a new stochastic window strategy.

3. Extensive experiments to validate the effectiveness.

Thank you for your listening

NEL-BITA
类脑智能技术及应用国家工程实验室