Cross-Scale Self-Supervised Blind Image Deblurring via Implicit Neural Representation

Tianjing Zhang[1], Yuhui Quan[2], Hui Ji[1]

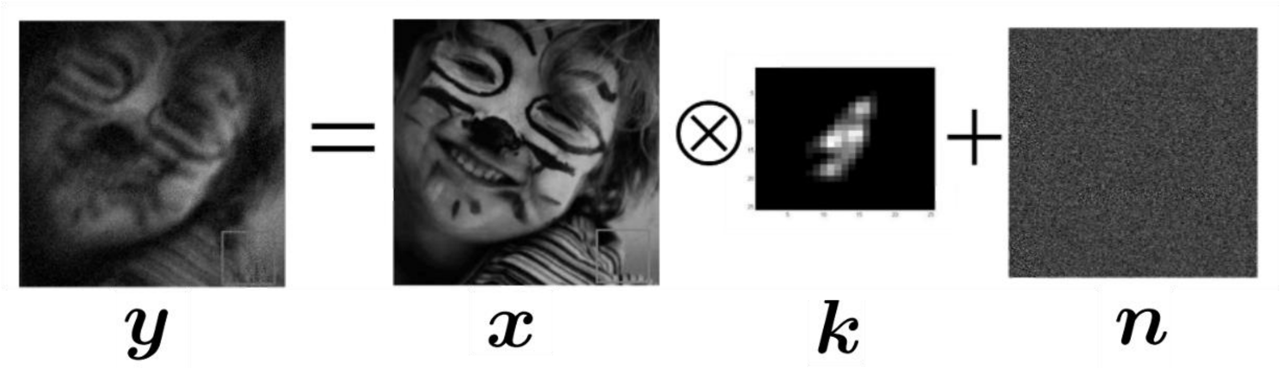
[1] Department of Mathematics, National University of Singapore

[2] School of Computer Science and Engineering, South China University of Technology

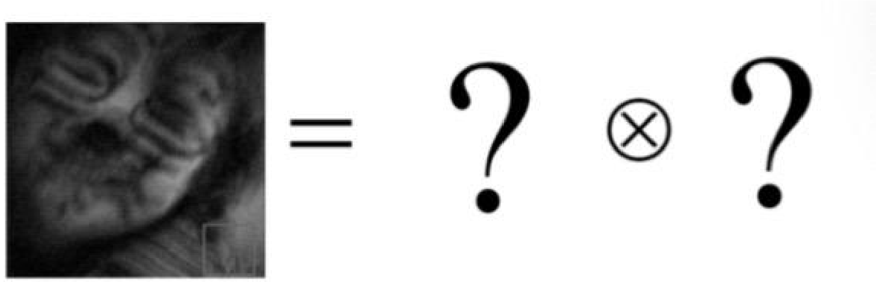
Background

- Uniform blurring, usually can be described as the convolution:

$$\boldsymbol{y} = \boldsymbol{k} \otimes \boldsymbol{x} + \boldsymbol{n}$$



- Blind image deblurring (BID): $\boldsymbol{y} \longrightarrow (\boldsymbol{k}, \boldsymbol{x})$



- Challenge: solution ambiguity: $\boldsymbol{y} = \boldsymbol{k} \otimes \boldsymbol{x} = \boldsymbol{\delta} \otimes \boldsymbol{y}$

Self-Supervised Deep Learning Methods

NEURAL INFORMATION PROCESSING SYSTEMS

- DNN-based re-parametrization of latent $\boldsymbol{x}/\boldsymbol{k}$:

$$\boldsymbol{x} := \mathcal{G}_{\boldsymbol{x}}(\cdot;\Theta_{\boldsymbol{x}}) \qquad \boldsymbol{k} := \mathcal{G}_{\boldsymbol{k}}(\cdot;\Theta_{\boldsymbol{k}})$$

- Standard self-supervised reconstruction loss:

$$\mathcal{L}_{sr}(\Theta_{\boldsymbol{k}},\Theta_{\boldsymbol{x}}) := ||\mathcal{G}_{\boldsymbol{k}}(\cdot;\Theta_{\boldsymbol{k}}) \otimes \mathcal{G}_{\boldsymbol{x}}(\cdot;\Theta_{\boldsymbol{x}}) - \boldsymbol{y}||_2^2$$

- Challenge: Overfitting due to the lack of ground truth (GT) data.

$$\mathcal{G}_{\boldsymbol{x}}(\cdot;\hat{\Theta}_{\boldsymbol{x}}) \to \boldsymbol{y} \qquad \mathcal{G}_{\boldsymbol{k}}(\cdot;\hat{\Theta}_{\boldsymbol{k}}) \to \boldsymbol{\delta}$$

Main Idea and Contributions

NEURAL INFORMATION PROCESSING SYSTEMS

Two key questions for self-supervised BID:

Q1: How to formulate a better self-supervised loss?

A1: **A cross-scale loss function**:

Leveraging the resolution-independent properties of Implicit Neural Representation (INR) for latent images/kernels.

Q2: How can we efficiently train the two NN generators to ensure accurate convergence to the latent images and kernels?

A2: **A progressive coarse-to-fine scheme**:

Enhancing training efficiency and ensuring the convergence to GT image/kernel.

Self-supervised cross-scale loss for BID

NEURAL INFORMATION PROCESSING SYSTEMS

- Without GT images, the only readily available loss function to train the generators is the fitting loss:

$$L_{\text{fit}}(\Theta_{\boldsymbol{k}}, \Theta_{\boldsymbol{x}}) = \mathcal{M}_f(\boldsymbol{y} - \boldsymbol{k} \otimes \boldsymbol{x}) = \mathcal{M}_{\text{fit}}\Big(\Phi_{\boldsymbol{k}}(\mathbb{I}_{\boldsymbol{k}}; \Theta_{\boldsymbol{k}}) \otimes \Phi_{\boldsymbol{x}}(\mathbb{I}_{\boldsymbol{x}}; \Theta_{\boldsymbol{x}}), \boldsymbol{y}\Big)$$

where $\mathcal{M}_f(\cdot)$ is some distance metric.

- **Such fitting loss clearly is not sufficient to resolve solution ambiguity!**

Self-supervised cross-scale loss for BID

NEURAL INFORMATION PROCESSING SYSTEMS

- To alleviate over-fitting, the down-sampled version of $\boldsymbol{y}$, denoted as $\boldsymbol{y}_{\downarrow_s}$ for scale s , has often been used to initiate the blur kernel estimate.

- However, $(\boldsymbol{x} \otimes \boldsymbol{k})\downarrow_2 \neq \boldsymbol{x}\downarrow_2 \otimes \boldsymbol{k}\downarrow_2$

- We present a cross-scale constraint that accurately characterizes the connection between $(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{k})$ at different scales:

Proposition 1. *For a kernel (filter)* $\boldsymbol{k}$, *let* $\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3$ *denote its associated QMF filters defined by*

$$\boldsymbol{g}_1[m,n] = (-1)^m \boldsymbol{k}[m,n],\ \boldsymbol{g}_2[m,n] = (-1)^n \boldsymbol{k}[m,n],\ \boldsymbol{g}_3[m,n] = (-1)^{m+n} \boldsymbol{k}[m,n],$$

for any $[m,n] \in \mathbb{I}_{\boldsymbol{k}}$. *Then, we have the following relation between consecutive two dyadic scales:*

$$(\boldsymbol{x}\downarrow_2) \otimes (\boldsymbol{k}\downarrow_2) = \frac{1}{4}\Big((\boldsymbol{x} \otimes \boldsymbol{k})\downarrow_2 + \sum_{d=1}^{3} (\boldsymbol{x} \otimes \boldsymbol{g}_d)\downarrow_2\Big).$$

Self-supervised cross-scale loss for BID

NEURAL INFORMATION PROCESSING SYSTEMS

- We introduce a scale consistency loss across two consecutive scales:
  For each scale s :

$$L_{\text{cross}}^{(s)}(\Theta_{\boldsymbol{k}}, \Theta_{\boldsymbol{x}}) = \mathcal{M}_c\Big(4(\boldsymbol{x}^{(s)}\downarrow_2) \otimes (\boldsymbol{k}^{(s)}\downarrow_2), (\boldsymbol{x}^{(s)} \otimes \boldsymbol{k}^{(s)})\downarrow_2 + \sum_{1\le d\le 3} (\boldsymbol{x}^{(s)} \otimes \boldsymbol{g}_d^{(s)})\downarrow_2\Big)$$

$$= \mathcal{M}_c\Big(4(\boldsymbol{x}^{(s+1)}) \otimes (\boldsymbol{k}^{(s+1)}), (\boldsymbol{x}^{(s)} \otimes \boldsymbol{k}^{(s)})\downarrow_2 + \sum_{1\le d\le 3} (\boldsymbol{x}^{(s)} \otimes \boldsymbol{g}_d^{(s)})\downarrow_2\Big)$$

- Ablation study on the $\mathcal{L}_{\text{cross}}$ in terms of of PSNR/SSIM.

Category	Manmade	Natural	People	Saturated	Text	Average
w/o $\mathcal{L}_{\text{cross}}$	21.19/0.778	25.84/0.887	30.74/0.918	17.69/0.682	26.75/0.917	24.44/0.836
Ours	**23.24/0.893**	**26.27/0.933**	**31.53/0.944**	**17.76/0.683**	**27.01/0.930**	**25.16/0.879**

The scale-consistency loss providing additional regularization for training two INR-based generators.

Resolution-free INR-based Generators

- The blur kernel and the latent image are re-parameterized by two INR models:

$$\begin{cases} \boldsymbol{k}[\mathbb{I}_{\boldsymbol{k}}] = \Phi_{\boldsymbol{k}}(\mathbb{I}_{\boldsymbol{k}}; \Theta_{\boldsymbol{k}}) & : \quad \boldsymbol{k}[i,j] = \Phi_{\boldsymbol{k}}\big([i,j]\big),\ [i,j] \in \mathbb{I}_{\boldsymbol{k}}; \\ \boldsymbol{x}[\mathbb{I}_{\boldsymbol{x}}] = \Phi_{\boldsymbol{x}}(\mathbb{I}_{\boldsymbol{x}}; \Theta_{\boldsymbol{x}}) & : \quad \boldsymbol{x}[i,j] = \Phi_{\boldsymbol{x}}\big([i,j]\big),\ [i,j] \in \mathbb{I}_{\boldsymbol{x}}, \end{cases}$$

- Let $\boldsymbol{k}^{(0)} = \boldsymbol{k}, \boldsymbol{x}^{(0)} = \boldsymbol{x}$ denotes the original scale, we can form both the kernel and the image in a dyadic pyramid:

$$\boldsymbol{k}^{(s)} = (\boldsymbol{k}^{(s-1)}) \downarrow_2 \quad \text{and} \quad \boldsymbol{x}^{(s)} = (\boldsymbol{x}^{(s-1)}) \downarrow_2, \quad \text{for } 1 \le s \le S_0.$$

- INR enables the model to generate the prediction with higher/lower resolutions from the same learned model, facilitating multi-scale processing and cross-scale interaction.

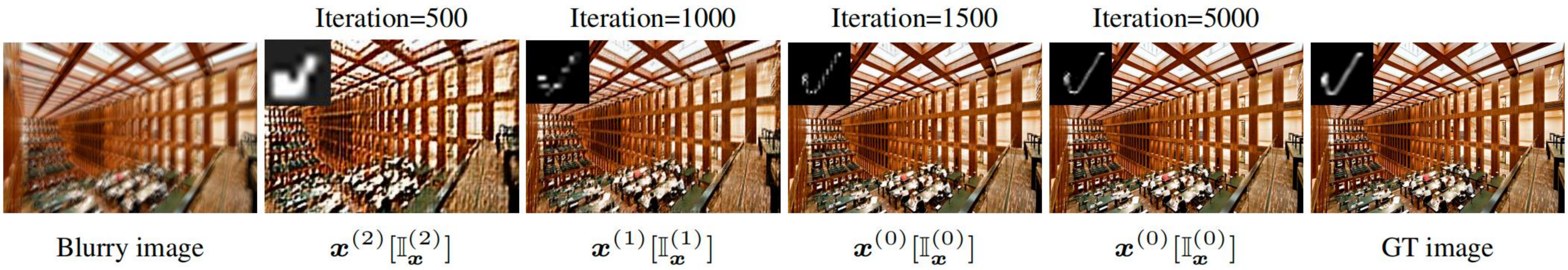
Progressively coarse-to-fine training

- **First stage** (Initialization): at the scale S_0

  Training the NNs with only the fitting loss $\mathcal{L}_{\text{fit}}^{(S_0)}$.
- **Second stage:** progressive refines the training from the scale S_0 to 0

  Training the NNs at the scale *s* with the loss $\mathcal{L}_{\text{fit}}^{(s)} + \lambda \mathcal{L}_{\text{cross}}^{(s)}$.
- **Final stage:** tuning ar the scale 0

  Training the NNs at scale 0 with only the fitting loss $\mathcal{L}_{\text{fit}}^{(0)}$.



Progressively coarse-to-fine training

NEURAL INFORMATION PROCESSING SYSTEMS

- **First stage** (Initialization): at the scale S_0

  Training the NNs with only the fitting loss $\mathcal{L}_{\text{fit}}^{(S_0)}$.
- **Second stage:** progressive refines the training from the scale S_0 to 0

  Training the NNs at the scale *s* with the loss $\mathcal{L}_{\text{fit}}^{(s)} + \lambda\mathcal{L}_{\text{cross}}^{(s)}$.
- **Final stage:** tuning ar the scale 0

  Training the NNs at scale 0 with only the fitting loss $\mathcal{L}_{\text{fit}}^{(0)}$.
- Ablation study on the croos scale and progressive training in terms of of PSNR/SSIM.

Category	Manmade	Natural	People	Saturated	Text	Average
Single-scale	22.04/0.803	25.93/0.890	30.33/0.933	17.68/0.688	24.76/0.886	24.14/0.840
w/o Progressive	20.36/0.742	23.91/0.829	26.35/0.821	17.22/0.675	22.88/0.857	22.14/0.790
Ours	**23.24/0.893**	**26.27/0.933**	**31.53/0.944**	**17.76/0.683**	**27.01/0.930**	**25.16/0.879**

PSNR results on blind uniform deblurring

NEURAL INFORMATION PROCESSING SYSTEMS

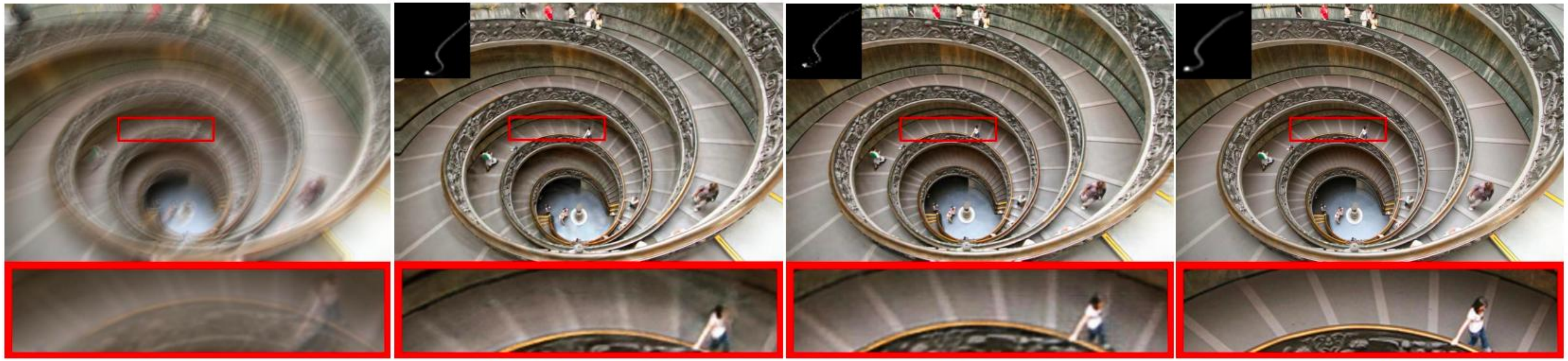
- [Lai et al.'s Dataset]: 100 images categorized into five groups, and covers 4 different kernels whose size ranges from 31×31 to 75×75.

	Non-learning		Supervised		Self-Supervised Method			
Methods	Yan et al. [CVPR'17]	Yang &Ji [CVPR'19]	MPRNet [CVPR'21]	Restormer [CVPR'22]	SelfDeblur [CVPR'20]	DEBID [TCSVT'23]	MCEM [CVPR'23]	Ours
Manmade	19.32	19.99	17.39	17.87	20.35	22.14	23.06	**23.24**
Natural	23.69	24.33	20.53	21.07	22.05	26.18	26.00	**26.27**
People	27.01	27.22	22.85	23.15	25.94	31.25	31.02	**31.53**
Saturated	16.46	17.04	15.35	15.58	16.35	**18.43**	17.21	17.76
Text	17.42	20.35	16.01	16.67	20.16	23.00	25.46	**27.01**
Average	19.89	21.79	18.42	18.89	20.97	24.29	24.55	**25.16**

Visual Results

NEURAL INFORMATION PROCESSING SYSTEMS

➢ Visual results on blind uniform deblurring [Lai Dataset].



Blurry Image | MCEM [CVPR’23] | Ours | GT

➢ Visual results on real blurry image.

Blurry Image | SelfDeblur[CVPR’20] | MCEM [CVPR’23] | Ours

Limitations and Future work

Limitation1: Computational cost for processing a large number of images as the method requires training the model for each individual sample.

Potential solution: **Meta-learning or Testing-time adaptation.**

Limitation2: Only applicable to handle uniform blurring, as it relies on the convolution model.

Future work: **Extending this approach to handle non-uniform blur.**

Thank you for your attention!