Single Image Defocus Deblurring via Implicit Neural Inverse Kernels (Supplemental Material)

1. Proof of Proposition 1

Consider an image space $\mathbb{R}^{H_0\times W_0}$. For a PSF $\boldsymbol{k} \in \mathbb{R}^{H\times W}$ with $H \ll H_0$ and $W \ll W_0$, its inverse kernel $\boldsymbol{k}^\dagger$ is defined in the frequency domain as follows: for $0 \le \omega_x < H_0, 0 \le \omega_y < W_0$,

$$\mathcal{F}(\boldsymbol{k}^\dagger)[\omega_x,\omega_y] = \begin{cases} \frac{1}{\mathcal{F}(\boldsymbol{k})[\omega_x,\omega_y]}, & \text{if } |\mathcal{F}(\boldsymbol{k})[\omega_x,\omega_y]| \neq 0; \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Consider the standard expansion-based dyadic upsampling operator $\mathcal{U}_s := \uparrow_s$, where the operator $\uparrow_s$: $\mathbb{R}^{H\times W} \to \mathbb{R}^{sH\times sW}$ can be defined in the Fourier domain as follows:

$$(\boldsymbol{k}\uparrow_s)[\omega_x,\omega_y] = \begin{cases} \boldsymbol{k}[\frac{\omega_x}{s},\frac{\omega_y}{s}], & \text{if } \frac{\omega_x}{s},\frac{\omega_y}{s} \in \mathbb{Z}; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In the remaining discussion, any index $[\omega_1,\omega_2]$ outside the range $[0,H_0-1]\times[0,W_0-1]$ is defined as $[\omega_1 \bmod H_0, \omega_2 \bmod W_0]$. Then, we have

$$\mathcal{F}(\boldsymbol{k}\uparrow_s)[\omega_x,\omega_y] = \mathcal{F}(\boldsymbol{k})[s\omega_x, s\omega_y]. \tag{3}$$

By the definition of inverse kernel, we have then

$$\begin{aligned} &\mathcal{F}((\boldsymbol{k}\uparrow_s)^\dagger)[\omega_x,\omega_y] \\ &= \begin{cases} \frac{1}{\mathcal{F}(\boldsymbol{k})[s\omega_x,s\omega_y]}, & \text{if } |\mathcal{F}(\boldsymbol{k})[s\omega_x,s\omega_y]| \neq 0; \\ 0, & \text{otherwise} \end{cases} \\ &= \mathcal{F}(\boldsymbol{k}^\dagger)[s\omega_x,s\omega_y] \\ &= \mathcal{F}((\boldsymbol{k}^\dagger)\uparrow_s)[\omega_x,\omega_y]. \end{aligned} \tag{4}$$

Thus, we have

$$(\boldsymbol{k}\uparrow_s)^\dagger = (\boldsymbol{k}^\dagger)\uparrow_s . \tag{5}$$

That is, the inverse kernel and the upsampling operator are commutative. Suppose that $\boldsymbol{k}^\dagger$ can be expressed as a linear combination over a set of atoms $\boldsymbol{V} = \{\boldsymbol{v}_n\}_{n=1}^N$:

$$\boldsymbol{k}^\dagger = \sum_{n=1}^N w_n \cdot \boldsymbol{v}_n, \tag{6}$$

Then, we have

$$(\boldsymbol{k}\uparrow_s)^\dagger = (\boldsymbol{k}^\dagger)\uparrow_s = (\sum_{n=1}^N w_n\cdot\boldsymbol{v}_n)\uparrow_s = \sum_{n=1}^N w_n\cdot(\boldsymbol{v}_n\uparrow_s). \tag{7}$$

Remark 1. *For interpolation-based upsampling operators, they usually can be expressed as*

$$\mathcal{U}_s(\boldsymbol{k}) = (\boldsymbol{k}\uparrow_s) * \boldsymbol{h}_s, \tag{8}$$

for some low-pass filter $\boldsymbol{h}_s$ whose size is related to s. For instance, $\boldsymbol{h}_s = [\frac{1}{2},1,\frac{1}{2}]$ for 1D linear interpolation on $s=2$. Based on (5) and that $(\boldsymbol{a}*\boldsymbol{b})^\dagger = \boldsymbol{a}^\dagger * \boldsymbol{b}^\dagger$, we have a generalized commutativity between inverse kernel and upsampling:

$$\big(\mathcal{U}_s(\boldsymbol{k})\big)^\dagger = (\boldsymbol{k}\uparrow_s)^\dagger * \boldsymbol{h}_s^\dagger = \big((\boldsymbol{k}^\dagger)\uparrow_s\big) * \boldsymbol{h}_s^\dagger. \tag{9}$$

As a result, in analog to (7), we have a generalized span:

$$(\mathcal{U}_s(\boldsymbol{k}))^\dagger = \sum_{n=1}^N w_n\cdot(\boldsymbol{v}_n\uparrow_s * \boldsymbol{h}_s^\dagger). \tag{10}$$

Such a generalized form is implicitly implemented in our INR-based model.

2. Additional Analysis

2.1. Influence of Atom Size and Atom Number

We investigate the impact of two main hyper-parameters, maximum atom size and atom number, in our INR-based model, by fixing one and varying the other. Concretely, we vary the maximum atom size to 7×7, 9×9, 11×11, 13×13, 15×15, 17×17, 19×19, respectively, with a fixed atom number of 10, and we vary the atom number to 2, 4, 6, 8, 10, 12, 14, 16, respectively, with a fixed maximum atom size of 15×15. The results plotted in Fig. 1 show that (a) when the atom size or atom number is too small, the performance of our INIKNet decreases (noticeable on RealDOF), probably due to under-fitting; (b) when setting the two hyper-parameters to larger values than the ones used in our experiments, the performance of INIKNet even has certain increase in some cases; (c) when the atom number is set too large, INIKNet performs slightly worse, probably due to overfitting; and (d) within a reasonable range of the hyper-parameters, INIKNet performs stably well.

2.2. Deblurring via INR Atoms vs. Non-INR Atoms

See Fig. 2 for some deblurred images on the CUHK dataset using INIKNet and its non-INR version respectively.

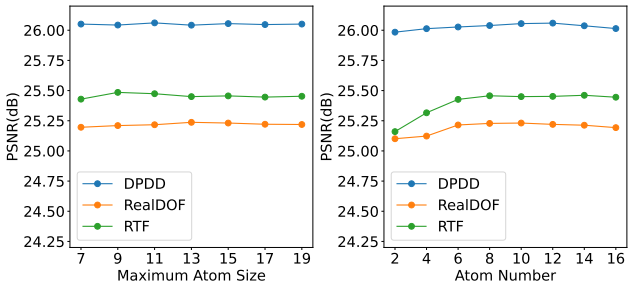

Figure 1: PSNR w.r.t. max atom size and atom number.

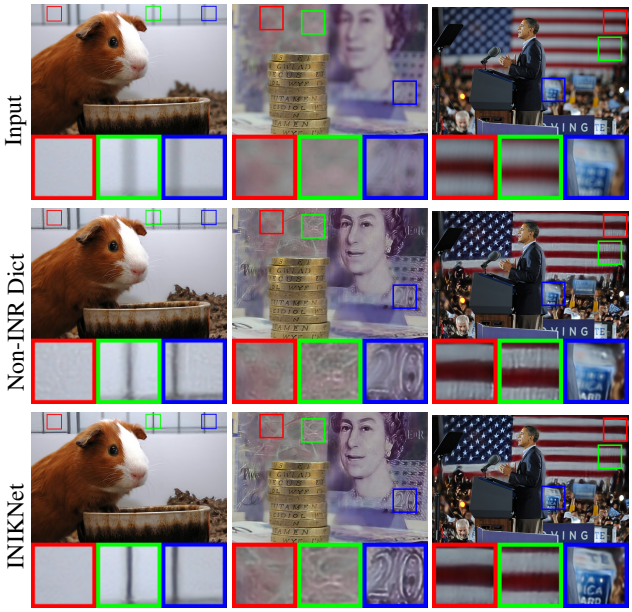

Figure 2: Deblurred image using INR-based model vs. non-INR model on CUHK dataset (without ground-truths). Zoom in for better views.

Leveraging INR-based inverse kernel atoms, INIKNet can recover cleaner textures with fewer artifacts.

2.3. Polar vs. Grid Coordinates in INR

Different from many existing works, our approach utilizes polar coordinates rather than grid coordinates as the input of INR. This does lead to some performance improvement (though not our main contribution), as demonstrated in Table 1.

Input	DPDD			RealDOF			RTF		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Grid	26.021	0.801	0.191	25.173	0.758	0.292	25.393	0.825	0.223
Polar	26.055	0.803	0.185	25.231	0.765	0.287	25.450	0.834	0.215

Table 1: Results of grid/polar coordinates used for INR-based atoms in INIKNet.

2.4. Performance Comparison to MIMOUNet

MIMOUNet [1] is a popular milt-scale NN for dynamic scene deblurring which employs asymmetric feature fusion to directly aggregate features of different scales. Table 2 presents the results of MIMOUNet trained on DPDD under the same training setting as ours. The MIMOUNet shows close performance to our INIKNet, but using a much larger model. In addition, its generalization performance on RealDOF and RTF is not as good as that of INIKNet, particularly that the PSNR gap is more than 0.8dB on RTF. These results again demonstrate the effectiveness of our approach.

Method	DPDD			RealDOF			RTF			#Par.
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	(M)
MIMOUNet	25.951	0.798	0.187	24.851	0.746	0.348	24.520	0.795	0.297	16.11
INIKNet	26.055	0.803	0.185	25.231	0.765	0.287	25.450	0.834	0.215	1.98

Table 2: Comparison with MIMOUNet trained on DPDD.

3. Additional Qualitative Results

In addition to those in the main paper, we provide extensive visual comparisons in Fig. 3, 4, 5, 6, 7, 8. When our INIKNet model is trained on DPDD, it not only performs well on the corresponding test split (see Fig. 3), but also shows superior generalization performance on RealDOF (see Fig. 4) and CUHK (see Fig. 5) both of which have a large domain gap against DPDD. Same thing happens when training on LFDOF (see Fig. 6, 7, 8). Under some extreme situations with severe blur, most methods fail while ours still works. Note that although the DRBNet trained on LFDOF achieves comparable performance to our INIKNet on LFDOF, its generalization performance to unseen patterns on CUHK is not as good as that of INIKNet (see Fig. 7, 8). These additional results further demonstrate the superiority of our proposed approach.

References

[1] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4641–4650, 2021. 2

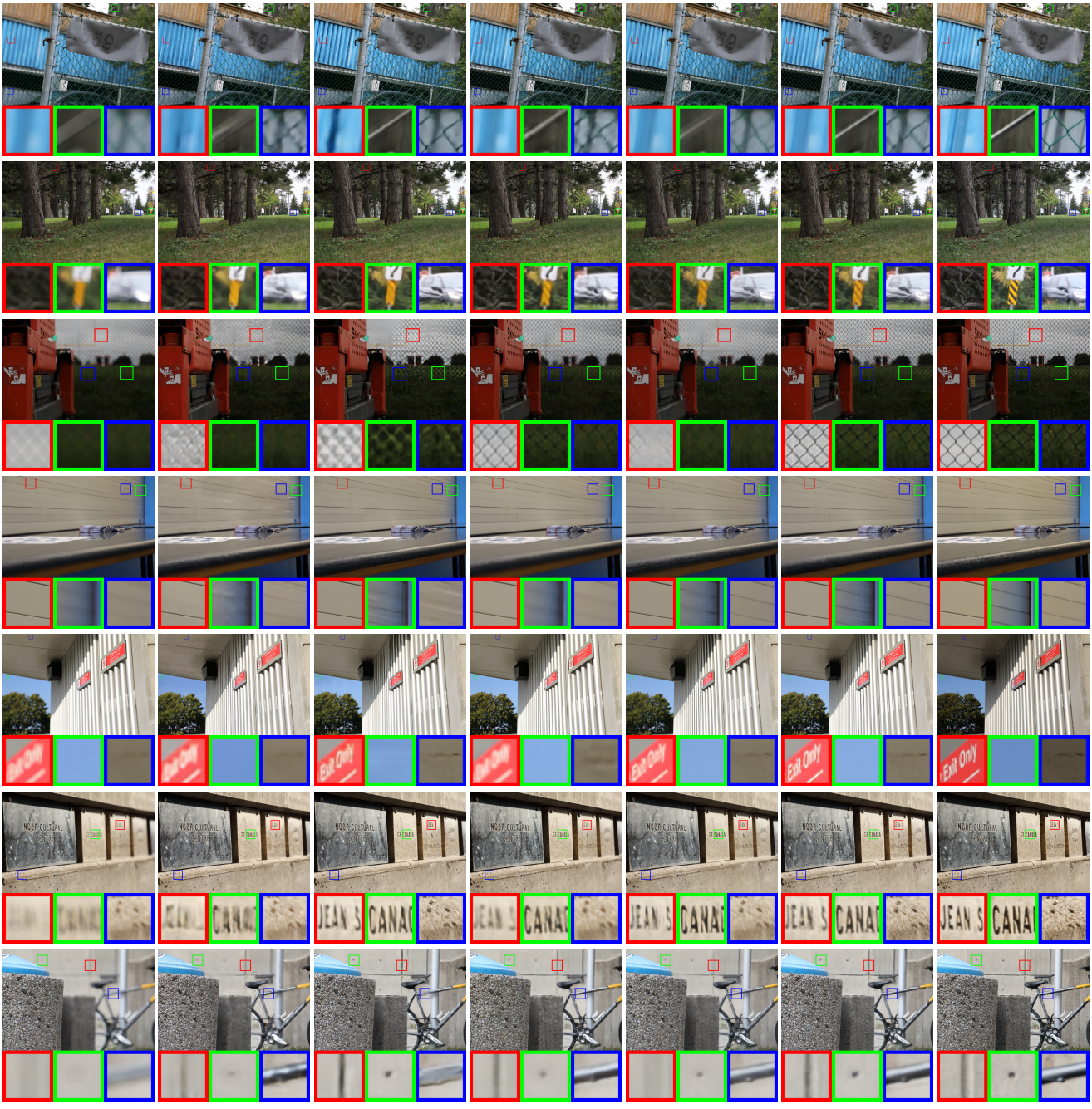

Figure 3: Results on DPDD using models trained on DPDD. Zoom in for better views.

Input IFANet KPAC GKMNet DRBNet INIKNet Ground Truth

Figure 4: Results on RealDOF using models trained on DPDD. Zoom in for better views.

Figure 5: Results on CUHK dataset using models trained on DPDD. Zoom in for better views.

Figure 6: Results on LFDOF using models trained on LFDOF. Zoom in for better views.

Figure 7: Results on CUHK dataset (without ground truths) using models trained on LFDOF. Zoom in for better views.

Figure 8: Results on CUHK dataset (without ground truths) using models trained on LFDOF. Zoom in for better views.