Plug-and-Play Dynamic In-context Learning with Stochastic Regularization for Screen Content Image Super-Resolution

Supplementary Material

1. Overview

In this supplementary material, we primarily elaborate on the conventional local implicit functions in the SCI SR task. We also analyze the impact of hyperparameters in our DIL and StRe, their effectiveness on natural images, and present additional visualization results.

2. Up-sampling Implicit Functions

Previous work focused on designing up-sampling implicit functions, such as ITSRN [7], which directly map the coordinates and up-scale to latent features. Other methods include LTE [4], which uses Fourier transforms to represent dominant frequency information in the frequency domain

$$\begin{aligned}\hat{\mathbf{F}}_j &= \mathcal{U}_\gamma^{LTE}(\mathbf{F}, r, C) \\ &= g_a(\tilde{\mathbf{F}}_j) \odot \begin{bmatrix}\cos(\pi(g_f(\tilde{\mathbf{F}}_j)\boldsymbol{\delta} + g_p(\hat{\mathbf{c}})))\\ \sin(\pi(g_f(\tilde{\mathbf{F}}_j)\boldsymbol{\delta} + g_p(\hat{\mathbf{c}})))\end{bmatrix},\end{aligned}$$

where $g_a : \mathbb{R}^d \rightarrow \mathbb{R}^{2d'}$, $g_f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'\times 2}$, and $g_p : \mathbb{R}^2 \rightarrow \mathbb{R}^{d'}$ are learnable estimators for amplitude, frequency, and phase, respectively. $j \in C$ is the queried coordinate, $\tilde{\mathbf{F}} \in \mathbb{R}^{rH\times rW\times d}$ denotes the upsampled features obtained from $\mathbf{F}$ through nearest-neighbor interpolation. $\boldsymbol{\delta} = [\boldsymbol{\delta}^h, \boldsymbol{\delta}^w] \in \mathbb{R}^2$ is a local-grid coordinate that denotes the relative coordinate between the query j and its nearest neighbor in the LR input. $\hat{\mathbf{c}} \in \mathbb{R}^2$ is the cell size ($= 2/r$). And BTC [6] that utilize B-spline basis with positive constraints and compact support to preserve discontinuities in SCIs, which can be formulated as

$$\begin{aligned}\hat{\mathbf{F}}_j &= \mathcal{U}_\gamma^{BTC}(\mathbf{F}, r, C) \\ &= g_c(\tilde{\mathbf{F}}_j) \odot \left[\beta^n\left(\frac{\boldsymbol{\delta}^h - g_k^h(\tilde{\mathbf{F}}_j)}{g_d^h(\hat{\mathbf{c}})}\right) \otimes \beta^n\left(\frac{\boldsymbol{\delta}^w - g_k^w(\tilde{\mathbf{F}}_j)}{g_d^w(\hat{\mathbf{c}})}\right)\right],\end{aligned}$$

where β^n denotes n-th order B-spline functions. $g_c : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$, $g_k : \mathbb{R}^d \rightarrow \mathbb{R}^{2\sqrt{d'}}$, and $g_d : \mathbb{R}^2 \rightarrow \mathbb{R}^{\sqrt{d'}}$ are learnable and control the coefficient, knot and dilation of B-spline basis, respectively. For a more detailed explanation, please refer to their original text.

3. Performance on Nature Images

Although our method is tailored for screen content images, we have also tested it on natural images. We train the LTE-based [4] method (*i.e.*, LTE+Ours) on the DIV2K [1] (including 800 natural images) dataset and evaluated it on Set5 [2], Set14 [8], B100 [5], and Urban100 [3]. The results are presented in Tab. 1. Our results indicate that DIL is not suitable for natural images. This is likely because it is difficult to find similar elements in a natural image, making it challenging to leverage structural similarities for super-resolution. However, our StRe method still demonstrates good performance improvements on natural images, suggesting its potential in addressing frequency imbalance is-

Train set: DIV2K Test set	Method	In-scale ×2	 ×4	Out-of-scale ×6	 ×8
Set5 ($n = 5$)	LTE	38.23	32.61	29.32	27.26
	LTE+DIL	38.19	32.56	29.33	27.24
	LTE+StRe	38.44	32.73	29.38	27.32
	LTE+DIL+StRe	38.34	32.59	29.34	27.27
Set14 ($n = 14$)	LTE	34.09	28.88	26.71	25.16
	LTE+DIL	34.11	28.80	26.70	25.13
	LTE+StRe	34.25	28.89	26.83	25.27
	LTE+DIL+StRe	34.21	28.87	26.77	25.19
B100 ($n = 100$)	LTE	32.36	27.77	26.01	24.95
	LTE+DIL	32.33	27.76	25.95	24.89
	LTE+StRe	32.47	27.91	26.17	25.08
	LTE+DIL+StRe	32.41	27.80	26.08	25.00
Urban100 ($n = 100$)	LTE	33.04	26.81	24.28	22.88
	LTE+DIL	33.04	26.82	24.25	22.86
	LTE+StRe	33.21	26.98	24.35	22.94
	LTE+DIL+StRe	33.10	26.91	24.33	22.88

Table 1. Performance on natural image benchmarks. We utilize LTE as our base method and train on DIV2K.

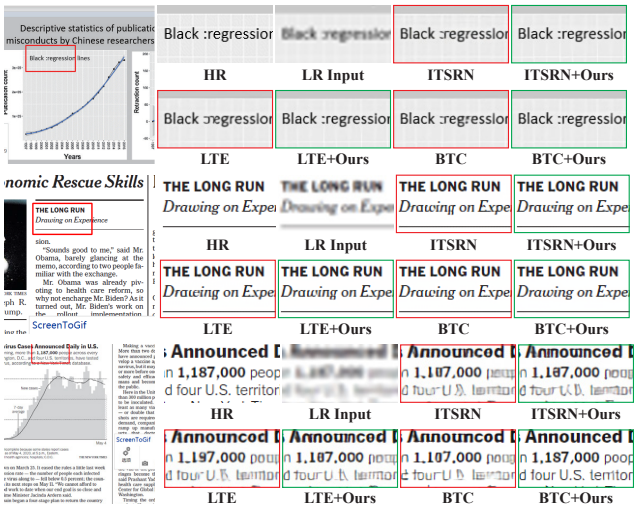

Figure 1. More visualization on SCI1K.

Figure 2. More visualization on SCID.

Figure 3. More visualization on SIQAD.

sues in these images to some extent.

4. Addtional Visualization

Fig. 1, 2 and 3 present additional visualization comparisons. These results demonstrate that our methods effectively address the limitations of existing methods that overemphasize local information, thereby significantly improving super-resolution performance.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR Workshops*, pages 126–135, 2017. 1

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *In Proceedings of the British Machine Vision Conference*, 2012. 1

[3] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pages 5197–5206, 2015. 1

[4] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *CVPR*, pages 1929–1938, 2022. 1

[5] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pages 416–423. IEEE, 2001. 1

[6] Byeonghyun Pak, Jaewon Lee, and Kyong Hwan Jin. B-spline texture coefficients estimator for screen content image super-resolution. In *CVPR*, pages 10062–10071, 2023. 1

[7] Jingyu Yang, Sheng Shen, Huanjing Yue, and Kun Li. Implicit transformer network for screen content image continuous super-resolution. *NeurIPS*, 34:13304–13315, 2021. 1

[8] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on Curves and Surfaces*, pages 711–730. Springer, 2010. 1