Supplementary for DINN360: Deformable Invertible Neural Network for Latitude-aware 360° Image Rescaling

Yichen Guo[1], Mai Xu[1]*, Lai Jiang[2]*, Leonid Sigal[2], Yunjin Chen

[1] School of Electronic and Information Engineering, Beihang University, Beijing, China
[2] Department of Computer Science, University of British Columbia, Vancouver, Canada

Figure 1. Examples of 360° images in F-360iSOD [14] and SUN360 [11] datasets. Some randomly selected patches and their HF components at low-/high-latitude regions are enlarged in the orange/blue boxes below.

In this supplementary document, we give more details about the data analysis, method and experiments.

1. Additional analyses of *Finding 1*

As introduced in Section 3 of the main text, our *Finding 1* validates that in 360° images, low-latitude regions tend to contain more textures, leading to larger HF components. Fig. 1 illustrates some examples of indoor and outdoor scenes in F-360iSOD [14] and SUN360 [11] datasets, respectively. Note that the HF components are calculated by Haar transformation [1]. It can be seen that, for both indoor and outdoor scenes, the image patches at low-latitude regions (in orange) contain more textures with larger HF components, compared with those at high-latitude regions

(in blue). This again verifies our *Finding 1*.

2. Additional formulations of invertible blocks

As introduced in Section 4.1 of the main text, the structure of invertible deformable and projection blocks endow DINN360 the ability for reverse procedure through the same model and parameters [3]. Let $\mathbf{a}_1^l$ and $\mathbf{a}_2^l$ denote the inputs of the l-th invertible block, the corresponding outputs $\mathbf{a}_1^{l+1}$ and $\mathbf{a}_2^{l+1}$ can be obtained by

$$\begin{aligned}
\mathbf{a}_1^{l+1} &= \mathbf{a}_1^l \odot \exp(\rho(\mathbf{a}_2^l)) + \rho'(\mathbf{a}_2^l), \\
\mathbf{a}_2^{l+1} &= \mathbf{a}_2^l \odot \exp(\rho(\mathbf{a}_1^{l+1})) + \rho'(\mathbf{a}_1^{l+1}),
\end{aligned} \tag{1}$$

Table 1. Results of WS-PSNR (dB) and WS-SSIM ($\times 10^{-2}$) of different sample seeds for $4\times$ rescaling on ODISR dataset.

| Sample seed	1	2	3	4	5	6	7	8	9	10
WS-PSNR	31.92406	31.92394	31.92366	31.92348	31.92324	31.92296	31.92262	31.92418	31.92416	31.92380
WS-SSIM	89.90252	89.90227	89.90163	89.90126	89.90081	89.90021	89.89954	89.90279	89.90271	89.90195
Sample seed	11	12	13	14	15	16	17	18	19	20
WS-PSNR	31.92262	31.92260	31.9227	31.92259	31.92270	31.92263	31.92273	31.92257	31.92266	31.92263
WS-SSIM	89.89954	89.89970	89.89976	89.89963	89.89970	89.89950	89.89979	89.89965	89.89969	89.90000
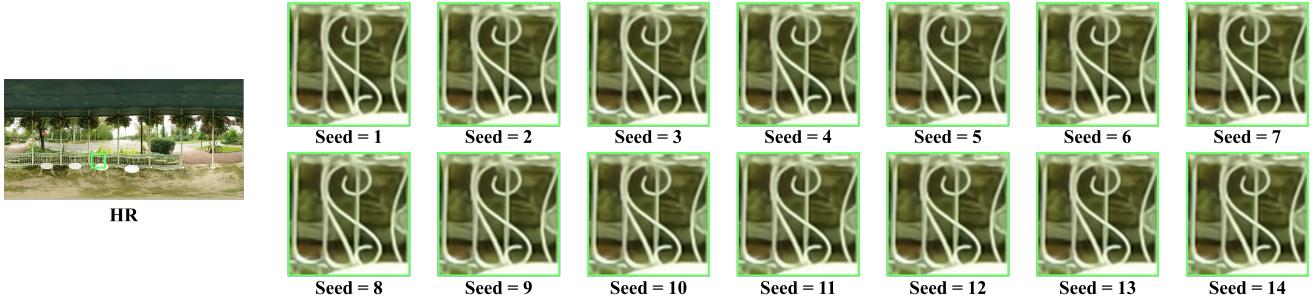
Figure 2. Examples of quantitative results with different sample seeds for $4\times$ rescaled HR images on ODISR dataset.

where $\rho(\cdot)$ and $\rho'(\cdot)$ denote the learnable scale and translation functions, and $\odot$ is the Hadamard product. Therefore, the corresponding reverse procedure can be formulated as follows,

$$
\begin{aligned}
\mathbf{a}_2^l &= (\mathbf{a}_2^{l+1} - \rho'(\mathbf{a}_1^{l+1})) \odot \exp(-\rho(\mathbf{a}_1^{l+1})), \\
\mathbf{a}_1^l &= (\mathbf{a}_1^{l+1} - \rho'(\mathbf{a}_2^l)) \odot \exp(-\rho(\mathbf{a}_2^l)).
\end{aligned}
\tag{2}
$$

3. Additional results

3.1. Ablation results on randomly sample

As introduced in Section 4.1, the latent variable $\widetilde{\mathbf{z}}$ in our DINN360 is randomly sampled from a normalized Gaussian distribution $\mathcal{N}(0, 1)$. Here, we pre-set 20 different random seeds for the sampling of $\mathcal{N}(0, 1)$, and then evaluate the performance of the rescaled HR images. The quantitative and qualitative results are shown in Tab. 1 and Fig. 2, respectively. It can be seen that DINN360 is able to generate the high-quality HR images with imperceptible visual differences, which is independent of the specific sampled latent variable. This validates that our DINN360 successfully learn to project the HF component into a prior distribution.

3.2. More subjective results of HR images

To validate the qualitative performance in generalization experiments, we randomly select some $4\times$ and $8\times$ rescaled images in SUN360 [11], F-360iSOD [14] and YouTube360 [9] datasets. Fig. 3 shows these subjective results for some randomly selected $360°$ images with the low-latitude and high-latitude regions being zoomed in. It can be seen that, compared with other rescaling methods, DINN360 is able to recover more realistic textures and better image details,

Table 2. Values of some key hyper-parameters.

Reference	Parameters	Values
Eq. (13)	Feedback proportion	$\alpha = 0.3$
Section 4.2	Number of IPBs	4
	Number of IDBs	4
Eq. (12)	Coefficient of HR loss	$\lambda_H = 1$
	Coefficient of LR loss	$\lambda_L = 5 \times 10^{-2}$
	Coefficient of Latent variable loss	$\lambda_{latent} = 1 \times 10^{-5}$
Section 5.1	Initial learning rate	2.5×10^{-4}
	Learning rate decay	0.7
	Weight decay	1×10^{-5}
	Training batch size	16
	Training patch size	160×160
	Training iteration	5×10^5

such as object boundaries and facial characteristics. This validates the generalization ability of our DINN360 method for generating high-quality HR images. To further measure the subjective improvement, we also evaluate 3 subjective metrics for ours and HCFlow, in terms of FID [4] ↓ (9.98 v.s.11.37), BRISQUE [8] ↓ (43.64 v.s.47.27), and VGCN [13] ↑ (59.59 v.s.56.08). The subjective results of all compared methods and more qualitative examples will be added in revision.

3.3. Qualitative results of downscaled LR images

To validate the qualitative performance of downscaling, we also randomly select some $2\times$, $4\times$ and $8\times$ downscaled LR images in ODISR [2] dataset. We compare our DINN360 method with (1) traditional interpolation methods, i.e., Bicubic, Bilinear and Lanczos; (2) 2D rescaling methods, i.e.,
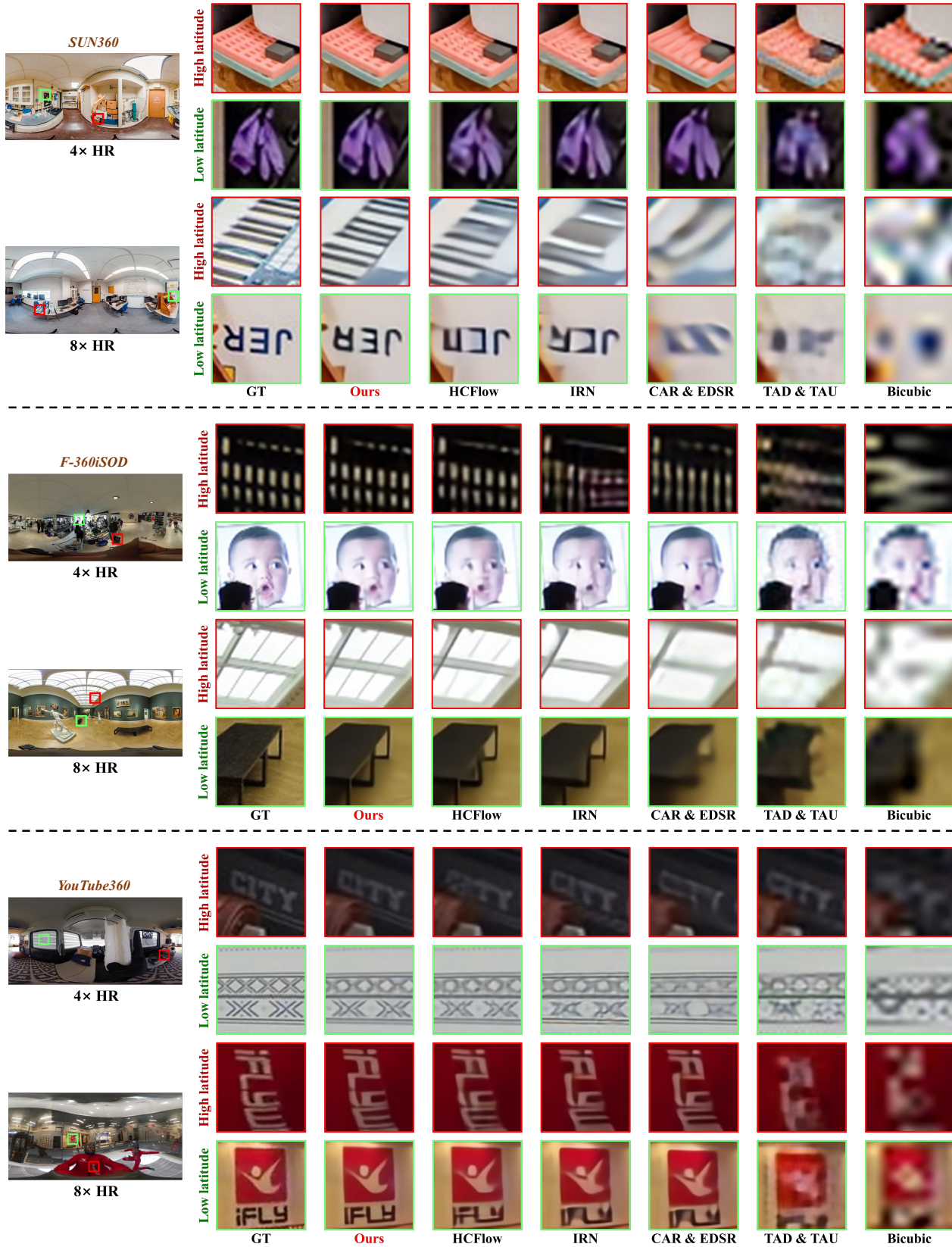
Figure 3. Quantitative results of $4\times$ and $8\times$ rescaled HR images on SUN360, F-360iSOD and YouTube datasets.
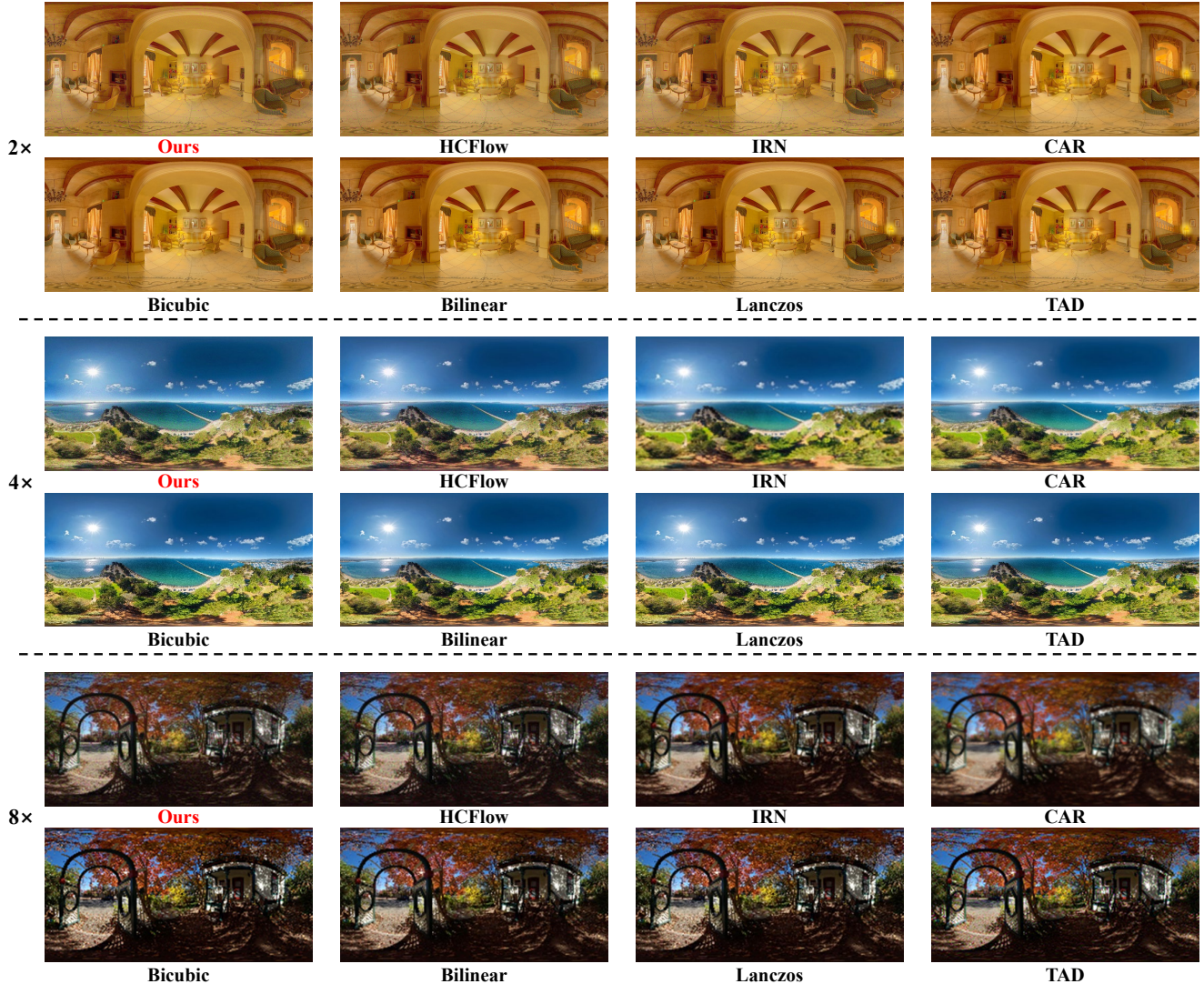
Figure 4. Quantitative results of $2\times$, $4\times$ and $8\times$ rescaled LR images on ODISR dataset.

TAD & TAU [5], CAR & EDSR [7,10], IRN [12] and HCFlow [6]. As can be seen in Fig. 4, the downscaled LR images of our and other compared methods have comparable performance, which are all visually valid for exhibition.

3.4. Application on 360° image SR

Here, we directly apply the reverse upscaling of DINN360 for $4\times$ 360° SR over the Bicubic downscaled LR images. The experimental results show that, compared with SOTA 360° SR method 360SISR [9], our method improves WS-PSNR by 0.12dB, 0.15dB, 0.13dB and 0.24dB on ODISR, SUN360, F-360iSOD and YouTube360 datasets. This validates the effectiveness of our DINN360 method on SR application.

4. Hyper-parameters of DINN360

Tab. 2 lists the key hyper-parameters of our DINN360, including the parameters about model structures, loss functions and model training, respectively. As introduced in Section 5.3 of the main text, the hyper-parameters are tuned to achieve the best final performance.

References

[1] Kenneth R Castleman. *Digital image processing*. Prentice Hall Press, 1996. 1

[2] Xin Deng, Hao Wang, Mai Xu, Yichen Guo, Yuhang Song, and Li Yang. Lau-net: Latitude adaptive upscaling network for omnidirectional image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9189–9198, 2021. 2

[3] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 1

[4] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 2

[5] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 399–414, 2018. 4

[6] Jingyun Liang, Andreas Lugmayr, Kai Zhang, Martin Danelljan, Luc Van Gool, and Radu Timofte. Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4076–4085, 2021. 4

[7] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 4

[8] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012. 2

[9] Akito Nishiyama, Satoshi Ikehata, and Kiyoharu Aizawa. 360 single image super resolution via distortion-aware network and distorted perspective images. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1829–1833. IEEE, 2021. 2, 4

[10] Wanjie Sun and Zhenzhong Chen. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Transactions on Image Processing*, 29:4027–4040, 2020. 4

[11] Jianxiong Xiao, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Recognizing scene viewpoint using panoramic place representation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2695–2702. IEEE, 2012. 1, 2

[12] Mingqing Xiao, Shuxin Zheng, Chang Liu, Zhouchen Lin, and Tie-Yan Liu. Invertible rescaling network and its extensions. *International Journal of Computer Vision*, pages 1–26, 2022. 4

[13] Jiahua Xu, Wei Zhou, and Zhibo Chen. Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5):1724–1737, 2020. 2

[14] Yi Zhang, Lu Zhang, Wassim Hamidouche, and Olivier Deforges. A fixation-based 360 benchmark dataset for salient object detection. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3458–3462. IEEE, 2020. 1, 2