

Supplementary Material: Learning Dual-Level Deformable Implicit Representation for Real-World Scale Arbitrary Super-Resolution

Zhiheng Li¹, Muheng Li¹, Jixuan Fan², Lei Chen^{1*}, Yansong Tang²,
Jiwen Lu¹, and Jie Zhou¹

¹ Department of Automation, Tsinghua University, China

² Shenzhen International Graduate School, Tsinghua University, China

{lizhihan21, li-mh20, fjx23}@mails.tsinghua.edu.cn,
leichen@sz.tsinghua.edu.cn, tang.yansong@sz.tsinghua.edu.cn,
{lujiwen, jzhou}@tsinghua.edu.cn

A Further Details about RealArbiSR Dataset

Camera Calibration and Camera Setting. We use two checkerboards for the calibration of the focal lengths in RealArbiSR dataset. The first checkerboard is designed for the scale factors of $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, $\times 3.5$, and $\times 4.0$. In this case, the checkerboard is annotated with seven concentric rectangles, which are labeled by ‘GT’, ‘ $\times 1.5$ ’, ‘ $\times 2.0$ ’, ‘ $\times 2.5$ ’, ‘ $\times 3.0$ ’, ‘ $\times 3.5$ ’, and ‘ $\times 4.0$ ’ at the top-right corners of the corresponding rectangles, as illustrated in Figure S1(a). The second checkerboard is designed for the rest scale factors, annotated with six rectangles with the labels of ‘GT’, ‘ $\times 1.7$ ’, ‘ $\times 2.3$ ’, ‘ $\times 2.7$ ’, ‘ $\times 3.3$ ’, and ‘ $\times 3.7$ ’, demonstrated in Figure S1(b). The camera is set to aperture priority mode. The focus, exposure, white balance, and ISO are set to automatic. We prefer to capture images in bright light conditions, because captured images tend to be noisy in the dark environment. We make sure there is no inappropriate blur due to depth-of-field by manual check. While collecting images, we gradually zoom out the camera to collect all LR-HR image pairs. For each scene, we take the images captured at the longest focal length as the ground truths, and the low-resolution versions are cropped from the red-dotted regions at shorter focal lengths, as shown in Figure S2.

Image Alignment. We first use the image registration algorithm with luminance adjustment [1] to coarsely align the low-resolution images with their high-resolution ground truths, and then crop the corresponding central regions of all images. In this way, the aberration effect can be minimized in the LR-HR image pairs. Next, we finer align the cropped LR-HR image pairs by adopting the image registration algorithm with luminance adjustment again. The borders of the aligned images are shaved to remove the distorted regions caused by the registration algorithm. We set 5 iterations for the optimization process of both

* indicates the corresponding author.

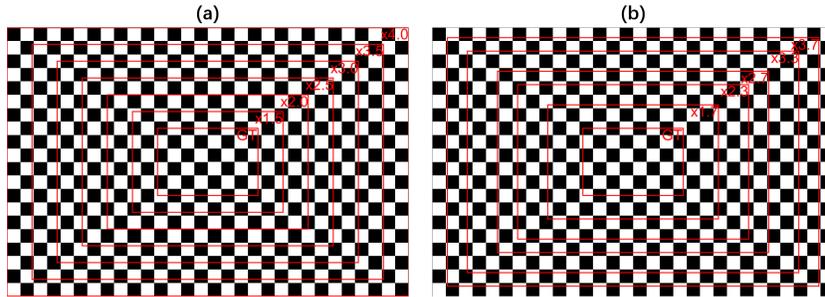


Fig. S1: The checkerboards for the calibration of the focal lengths with the scale factors of (a) $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, $\times 3.5$, $\times 4.0$; and (b) $\times 1.7$, $\times 2.3$, $\times 2.7$, $\times 3.3$, $\times 3.7$.

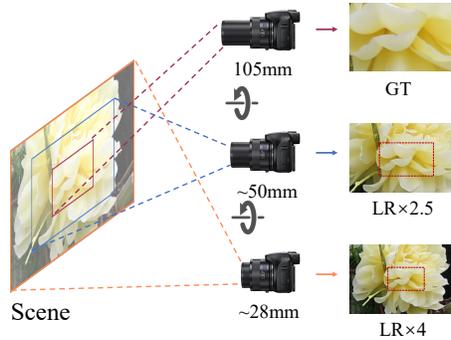


Fig. S2: The illustration of dataset collection. The images taken at the longest focal length (105mm) are used as the ground truths, and the low-resolution versions are cropped from the red-dotted regions from the images taken at shorter focal lengths.

coarse and fine alignments. After all these image pre-processing, we conduct a careful manual check for all images. Image pairs with inappropriate blur, moving objects, inappropriate exposure, etc., are all discarded.

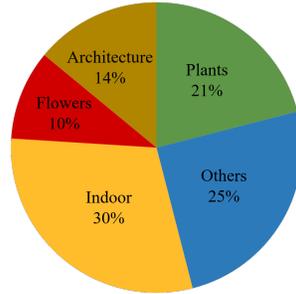
Dataset Statistics. The resolutions of LR and HR images for different scale factors are listed in Table S1. The RealArbiSR dataset covers diverse scenes in indoor and outdoor environments. We present the content distribution of our RealArbiSR dataset in Figure S3. Some ground-truth examples of the RealArbiSR dataset are illustrated in Figure S4.

B Further Analysis of Training Scaling Factors in RealArbiSR Dataset

To further analyze the effect of non-integer scale factors in the training of real-world scale arbitrary SR, we present the experimental results with the RDN [7] backbone at all scale factors in Table S2 and S3. As shown in Table S2 and S3,

Table S1: The resolutions of LR and HR images for different scale factors in RealArbiSR dataset.

Scale factor	$\times 1.5$	$\times 2.0$	$\times 2.5$	$\times 3.0$	$\times 3.5$	$\times 4.0$
HR	1212×792	1196×776	1180×760	1164×744	1148×728	1132×712
LR	808×528	598×388	472×304	388×248	328×208	283×178
Scale factor	$\times 1.7$	$\times 2.3$	$\times 2.7$	$\times 3.3$	$\times 3.7$	
HR	1241×816	1219×782	1215×783	1188×792	1184×777	
LR	730×480	530×340	450×290	360×240	320×210	

**Fig. S3:** The content distribution of our RealArbiSR dataset.**Fig. S4:** Image examples of the RealArbiSR dataset.

the models which are trained at all scale factors (including $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, $\times 3.5$, and $\times 4.0$, indicated as ‘All’ in Table S2 and S3) perform better than the ones trained only at integer scale factors (including $\times 2.0$, $\times 3.0$, and $\times 4.0$, indicated as ‘ $\times 2 \times 3 \times 4$ ’ in Table S2 and S3).

Table S2: Quantitative Analysis of training scale factors in RealArbiSR dataset. The highest PSNR at each scale factor on each method is bolded. ‘ $\times 2 \times 3 \times 4$ ’ represents the models are trained at the scale factors of $\times 2.0$, $\times 3.0$, and $\times 4.0$. ‘All’ represents the models are trained at the scale factors of $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, $\times 3.5$, and $\times 4.0$. The models are tested at the scale factors from $\times 1.5$ to $\times 4.0$ with a step of $\times 0.5$ in RealArbiSR dataset.

Method	Training Scale	$\times 1.5$	$\times 2.0$	$\times 2.5$	$\times 3.0$	$\times 3.5$	$\times 4.0$
RDN-LIIF [3]	$\times 2 \times 3 \times 4$	36.77	34.38	32.53	31.31	30.32	29.65
	All	37.14	34.41	32.60	31.40	30.34	29.70
RDN-LTE [4]	$\times 2 \times 3 \times 4$	36.84	34.44	32.63	31.43	30.44	29.76
	All	37.24	34.52	32.76	31.53	30.54	29.84
RDN-CiaoSR [2]	$\times 2 \times 3 \times 4$	36.84	34.68	32.91	31.65	30.67	29.95
	All	37.38	34.70	32.96	31.68	30.77	30.07
RDN-DDIR	$\times 2 \times 3 \times 4$	37.22	34.81	32.99	31.76	30.77	30.05
	All	37.63	35.02	33.20	31.91	30.94	30.21

Table S3: Quantitative Analysis of training scale factors in RealArbiSR dataset. The highest PSNR at each scale factor on each method is bolded. ‘ $\times 2 \times 3 \times 4$ ’ represents the models are trained at the scale factors of $\times 2.0$, $\times 3.0$, and $\times 4.0$. ‘All’ represents the models are trained at the scale factors of $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, $\times 3.5$, and $\times 4.0$. The models are tested at the scale factors of $\times 1.7$, $\times 2.3$, $\times 2.7$, $\times 3.3$, and $\times 3.7$ in RealArbiSR dataset.

Method	Training Scale	$\times 1.7$	$\times 2.3$	$\times 2.7$	$\times 3.3$	$\times 3.7$
RDN-LIIF [3]	$\times 2 \times 3 \times 4$	34.63	32.24	31.31	30.18	29.63
	All	34.66	32.40	31.45	30.28	29.71
RDN-LTE [4]	$\times 2 \times 3 \times 4$	34.71	32.40	31.47	30.32	29.74
	All	34.74	32.44	31.55	30.39	29.81
RDN-CiaoSR [2]	$\times 2 \times 3 \times 4$	34.45	32.46	31.66	30.55	29.94
	All	34.54	32.50	31.67	30.56	29.96
RDN-DDIR	$\times 2 \times 3 \times 4$	35.06	32.75	31.84	30.70	30.12
	All	35.07	32.88	31.96	30.75	30.15

C Out-of-distribution Testing in RealArbiSR Dataset

We conduct out-of-distribution testing in RealArbiSR dataset. To do this, we train one model at the scale factors of $\times 1.5 \times 2.0 \times 2.5 \times 3.0$, and $\times 3.5$, and test it at the scale factors of $\times 3.7$ and $\times 4.0$ in RealArbiSR dataset. As shown in Table S4, our DDIR model achieves the best results in the out-of-distribution testing, compared to other baselines.

Table S4: Quantitative comparison of out-of-distribution testing on RealArbiSR dataset in PSNR(dB). The highest PSNR at each scale factor is bolded. One model is trained at the scale factors of $\times 1.5$, $\times 2.0$, $\times 2.5$, $\times 3.0$, and $\times 3.5$, and tested at the scale factors of $\times 3.7$ and $\times 4.0$ in RealArbiSR dataset.

Method	EDSR Backbone		RDN Backbone	
	$\times 3.7$	$\times 4.0$	$\times 3.7$	$\times 4.0$
LIIF [3]	29.59	29.49	29.73	29.72
LTE [4]	29.55	29.64	29.83	29.92
CiaoSR [2]	29.84	29.86	30.03	29.99
DDIR	29.88	29.99	30.11	30.11

D Analysis on Simulated and Real SR Experiments in RealArbiSR Dataset

We compare the bicubic and real-world degradation in RealArbiSR dataset. It demonstrates the advantage of our RealArbiSR dataset compared to synthetic scale arbitrary methods with bicubic degradation. For our DDIR model with bicubic degradation, we remove the deformation field and deformation branch because they are specifically designed for real-world degradation and do not work for bicubic degradation. As shown in Table S5, the performance of models with bicubic degradation all drops by a large margin. It proves bicubic degradation in synthetic scale arbitrary super-resolution fails to generalize in real-world degradation for all models at all scale factors. Further qualitative comparison is shown in Figure S5 and S6.

E Quantitative Comparison between DDIR and Real-World SR Methods

We compare the quantitative results between our DDIR model and other real-world SR methods [1, 5, 6] on RealSR dataset in PSNR(dB). For our DDIR method, one model is trained and tested at the scale factors of $\times 2.0$, $\times 3.0$, and $\times 4.0$. For other real-world SR methods, different models are trained and tested

Table S5: Quantitative Analysis of bicubic and real-world degradations with the RDN [7] backbone in RealArbiSR dataset. The highest PSNR at each scale factor on each method is bolded. The models are trained and tested at the scale factors from $\times 1.5$ to $\times 4.0$ with a step of $\times 0.5$ in RealArbiSR dataset.

Method	Degradation	$\times 1.5$	$\times 2.0$	$\times 2.5$	$\times 3.0$	$\times 3.5$	$\times 4.0$
RDN-LIIF [3]	Bicubic	35.70	32.69	30.91	29.63	28.71	28.03
	Real	37.14	34.41	32.60	31.40	30.34	29.70
RDN-LTE [4]	Bicubic	35.68	32.69	30.89	29.61	28.69	28.01
	Real	37.24	34.52	32.76	31.53	30.54	29.84
RDN-CiaoSR [2]	Bicubic	35.67	32.69	30.89	29.62	28.71	28.03
	Real	37.38	34.70	32.96	31.68	30.77	30.07
RDN-DDIR	Bicubic	35.70	32.70	30.91	29.64	28.72	28.04
	Real	37.63	35.02	33.20	31.91	30.94	30.21

Table S6: Quantitative comparison between DDIR and real-world SR methods on RealSR dataset in PSNR(dB). The best metric result at each scale factor is bolded. For DDIR, one model is trained and tested at the scale factors of $\times 2.0$, $\times 3.0$, and $\times 4.0$. LP-KPN [1], CDC [6], and D2C-SR [5] train and test different models at different scales.

Method	RealSR		
	$\times 2.0$	$\times 3.0$	$\times 4.0$
LP-KPN [1]	-	30.60	28.65
CDC [6]	33.96	30.99	29.24
RDN-D2C [5]	34.03	30.93	29.32
RDN-DDIR	34.35	31.15	29.48

for different scales. As shown in Table S6, even with only one model, our DDIR model still outperforms existing real-world SR methods, at which one model is trained at each scale factor.

F More Visual Results

We show more visual results on the RealArbiSR dataset and the RealSR dataset with real-world and bicubic degradations in Figure S5 and S6. For DDIR model with bicubic degradation, we also remove the deformation branch and keep the appearance embedding. As shown in Figure S5 and S6, our DDIR model reconstructs better image details and sharper edges compared to other methods. By comparing models with bicubic and real-world degradation, we can see synthetic scale arbitrary super-resolution methods with bicubic degradation fail to generalize in the real-world case.

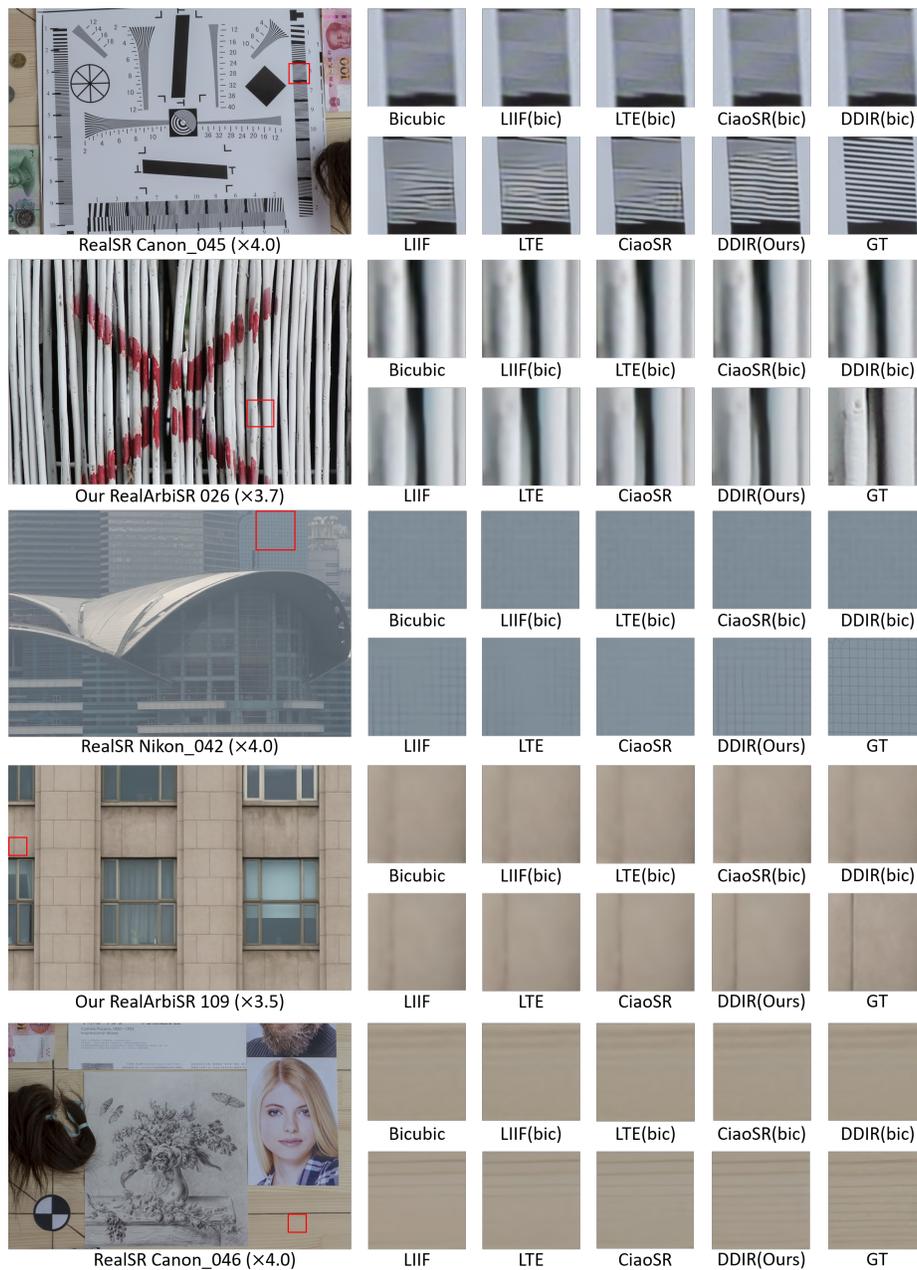


Fig. S5: Qualitative comparisons between different methods on benchmarks. Zoom in to have better views.

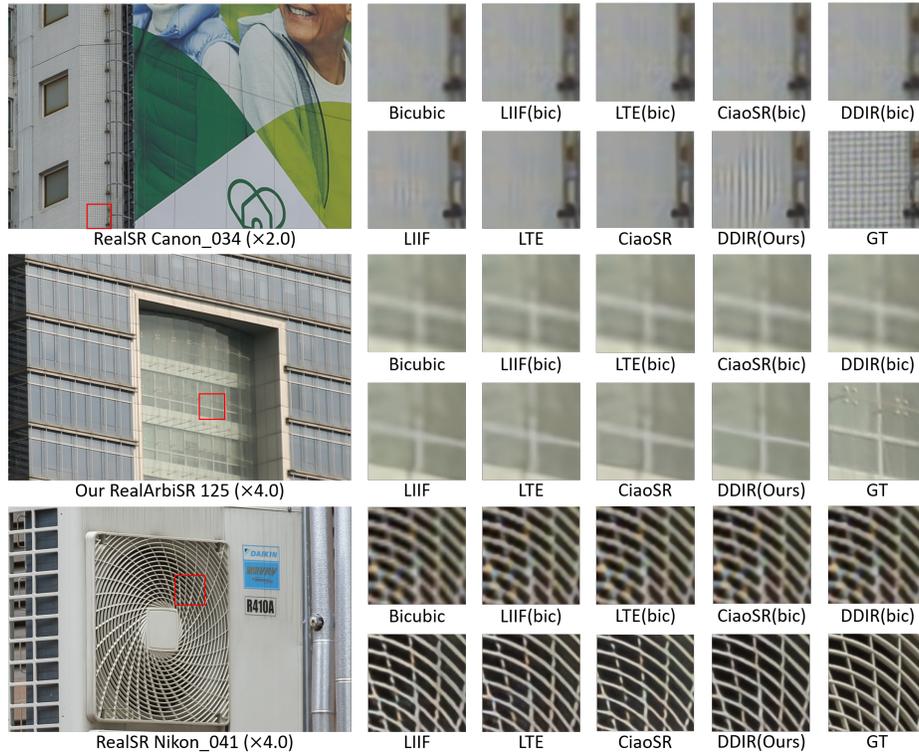


Fig. S6: Qualitative comparisons on benchmarks. Zoom in to have better views.

References

1. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: ICCV. pp. 3086–3095 (2019)
2. Cao, J., Wang, Q., Xian, Y., Li, Y., Ni, B., Pi, Z., Zhang, K., Zhang, Y., Timofte, R., Van Gool, L.: Cioasr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution. arXiv preprint arXiv:2212.04362 (2022)
3. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: CVPR. pp. 8628–8638 (2021)
4. Lee, J., Jin, K.H.: Local texture estimator for implicit representation function. In: CVPR. pp. 1929–1938 (2022)
5. Li, Y., Huang, H., Jia, L., Fan, H., Liu, S.: D2c-sr: A divergence to convergence approach for real-world image super-resolution. In: ECCV. pp. 379–394 (2022)
6. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: ECCV. pp. 101–117 (2020)
7. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR. pp. 2472–2481 (2018)