D2C-SR: A Divergence to Convergence Approach for Real-World Image Super-Resolution

Youwei Li [1] Haibin Huang [2] Lanpeng Jia [3] Haoqiang Fan [1]
Shuaicheng Liu [4,1,†]

[1] Megvii Technology, Beijing, China [2] Kuaishou Technology, Beijing, China
[3] Great Wall Motor Company Limited, Baoding, China
[4] University of Electronic Science and Technology of China, Chengdu, China
{liyouwei, fhq}@megvii.com jialanpeng@gwm.cn jackiehuanghaibin@gmail.com
liushuaicheng@uestc.edu.cn
†Corresponding Author

1 Overview

In this supplementary material, we first introduce more details of our D2CRealSR dataset in Sec. 2, including the collection and processing method. In Sec. 3, we discuss more about simulated SISR and real-world SISR. Then, in Sec. 4, we demonstrate more experimental results and details. Finally, we show more visual comparison results and divergence results in Sec. 5, including the results on the RealSR and D2CR RealSR datasets.

2 D2CRealSR Dataset

Existing real-world SISR datasets generally include images pairs on x2, x3 and x4 scaling factors, such as RealSR dataset [2] and DRealSR dataset [13]. Our proposed D2CRealSR dataset have images pairs on x8 scaling factor, which larger than the existing real-world datasets. The ill-posed problem is more obvious on D2CRealSR dataset.

Images pairs in our dataset are taken by zooming Sony α-7. The number of training pairs in our dataset is 100, and testing set has 15 images pairs selected randomly. The D2CRealSR provides, to the best of our knowledge, the first general purpose benchmark on x8 scaling factor for real-world SISR. It is difficult to obtain images pairs with good alignment effects in natural scenes. These difficulties in the alignment process include depth-of-field misalignment, noticeable perspective misalignment, lens distortion misalignment and resolution alignment ambiguity [15]. Therefore, using the DSLR camera to directly acquire natural scene images pairs generally cannot be well aligned. We have found misalignment in DRealSR dataset, mostly due to depth-of-field misalignment. To solve these misalignment problems, we collect some high-resolution images, firstly. These images are mostly taken by DSLR from existing datasets or our own shooting dataset. Different from City100 [3], which only have city scenes, we have selected the images with different scenes including indoor and outdoor. It contains a variety of target objects with rich texture. Then, we print these scenes on high-quality postcards just like City100, and we get images pairs by taking photos of postcards. For each scaling factors, we measured a safety distance under a controlled environment in professional laboratory such that the textures and reflections of postcards have no effect on the captured data. These postcards are on a fixed plane, which can reduce the influence of depth-of-field misalignment. It also weakens noticeable perspective misalignment. Further, we blur the HR images of pairs in during calculating matrix M in SIFT algorithm and apply the M on the source HR images to align source HR images with LR images. In this way, it can weaken the influence of resolution alignment ambiguity. We use multiple iterative alignment methods to ensure alignment results. Cropping the center area after each alignment can reduce misalignment caused by lens distortion. Several examples of our D2CRealSR dataset are shown in Fig. 1, and it will be made publicly available later.

3 Simulated SISR and Real-World SISR

We focus primarily on real-world SR for two reasons. Firstly, more works have focused on real-world SR because the great gap between simulated and real-world degradation hinders practical SR applications [2,13]. Secondly, as shown in Fig. 2, real-world degradation loses more information than bicubic compared with the HR, and ill-posed problem becomes more apparent in real-world datasets. The motivation of our approach is also to alleviate the ill-posed problem. Therefore, the real-world datasets can better reflect the effectiveness of our method. Nevertheless, as shown in Table 1, we still achieve good performance on bicubic benchmarks.

4 More Experimental Results and Details

4.1 Details in Divergence Loss

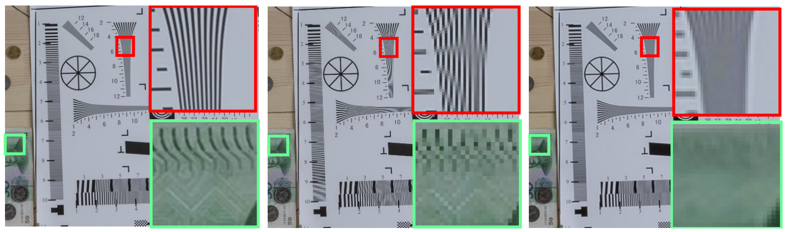
The traditional SR methods cannot be further optimized due to the ill-posed nature and it always exists a minimum distance between I_{SR} and I_{HR}. In our experiments, too large margin cause performance degradation. Therefore, in Fig. 3, we count the cumulative histogram of the average $L2$ distance based on pre-trained model with only $L2$ loss. And we set the margin within a suitable range. Theoretically, as the convergence of the network progresses, the α in the divergence loss should be reduced accordingly. Otherwise the loss of divergence can affect the impact of content consistency loss leading to degradation of results, as shown in Table 2. Therefore, in our experiments, we will decay the weight of the divergence loss after the set training step. On the other hand, the divergence loss mainly helps to maintain the stability of the divergence process, so if the divergence loss is completely removed it may also lead to degradation of the performance due to incomplete divergence or unstable divergence stage in some situations.

(a) City scenes

(b) Natural scenes

(c) Indoor scenes

Fig. 1: Examples of D2CRealSR dataset.



(a) HR Ground Truth PSNR (b) Bicubic Downsampling (x4) 24.01 dB (c) Real-world Degradation (x4) 23.09 dB

Fig. 2: Comparison of bicubic degradation and real-world degradation (x4). Real-world degradation loses more information than bicubic degradation.

Table 1: Comparison on bicubic datasets (Set5 [1] and B100 [10]).

	Method	x2		x3		x4		x8	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	Bicubic	33.66	0.9299	30.39	0.8682	28.42	0.8104	24.40	0.6580
	VDSR [5]	37.53	0.9590	33.67	0.9210	31.35	0.8830	25.93	0.7240
	LapSRN [6]	37.52	0.9591	33.82	0.9227	31.54	0.8850	26.15	0.7380
	MemNet [11]	37.78	0.9597	34.09	0.9248	31.74	0.8893	26.16	0.7414
	EDSR [8]	38.11	0.9602	34.65	0.9280	32.46	0.8968	26.96	0.7762
	RCAN [16]	**38.27**	0.9614	**34.74**	0.9299	**32.63**	**0.9002**	27.31	**0.7878**
	SRMDNF [14]	37.79	0.9601	34.12	0.9254	31.96	0.8925	-	-
	D-DBPN [4]	38.09	0.9600	-	-	32.47	0.8980	27.21	0.7840
	RDN [17]	38.24	0.9614	34.71	0.9296	32.47	0.8990	-	-
	D2C-SR(Ours)	38.26	**0.9619**	34.72	**0.9302**	32.62	0.8999	**27.46**	0.7814
B100	Bicubic	29.56	0.8431	27.21	0.7385	25.96	0.6675	23.67	0.5480
	VDSR [5]	31.90	0.8960	28.83	0.7990	27.29	0.7260	24.49	0.5830
	LapSRN [6]	31.08	0.8950	28.82	0.7980	27.32	0.7270	24.54	0.5860
	MemNet [11]	32.08	0.8978	28.96	0.8001	27.40	0.7281	24.58	0.5842
	EDSR [8]	32.32	0.9013	29.25	0.8093	27.71	0.7420	24.81	0.5985
	RCAN [16]	32.41	0.9027	**29.32**	0.8111	27.77	0.7436	24.98	0.6058
	SRMDNF [14]	32.05	0.8985	28.97	0.8025	27.49	0.7337	-	-
	D-DBPN [4]	32.27	0.9000	-	-	27.72	0.7400	24.88	0.6010
	RDN [17]	32.34	0.9017	29.26	0.8093	27.72	0.7419	-	-
	D2C-SR(Ours)	**32.41**	**0.9041**	29.28	**0.8125**	**27.77**	**0.7466**	**24.98**	**0.6140**

Table 2: Effect of the margin and α in divergence loss (RealSR x4).

α	*margin*	PSNR	SSIM	PSNR(I_D^1)	PSNR(I_D^2)	PSNR(I_D^3)	PSNR(I_D^4)
1e-3	1e-4	29.63	0.830	29.132	29.134	29.142	29.180
1e-2	1e-4	29.35	0.824	28.252	28.288	28.800	28.290
1e-2	1e-3	29.26	0.824	28.234	28.200	28.587	28.203
1e-1	1e-3	28.66	0.813	26.853	26.894	27.537	27.285

Table 3: Performance of different branches.

Scale	PSNR	SSIM	PSNR(I_D^1)	PSNR(I_D^2)	PSNR(I_D^3)	PSNR(I_D^4)
2x	34.40	0.926	34.340	34.311	34.314	34.323
3x	31.33	0.871	31.205	31.176	31.203	31.162
4x	29.72	0.831	29.547	29.559	29.539	29.567
8x	30.55	0.871	30.414	30.416	30.389	30.385

4.2 Performance of Different Branches

In Table 3, we list the performance on the different branches after divergence and after convergence. It can be seen that the results of the different branches have a divergent distribution effect, producing a variation in the PSNR. This is due to the independence of the different branches of our divergence stage and the stabilizing effect of the divergence loss. And the performance after convergence significantly exceeded the performance of each branch individually.

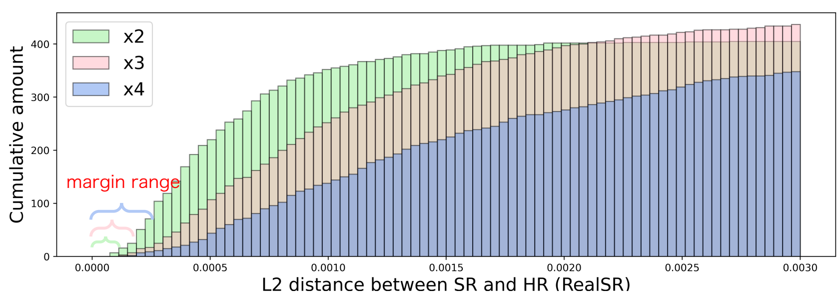


Fig. 3: Histogram of the average $L2$ distance.

4.3 Residual Domain in Divergence Loss

In Fig. 4, we show the visualization of the results after computing the residuals between HR and SR prediction in divergence loss. Residual results have a higher response in the high-frequency region. Therefore, adding residual processing to the divergence loss allows the divergence predictions to focus more on high-frequency texture regions, which is exactly what we want to achieve.

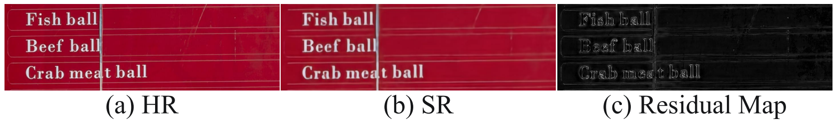


Fig. 4: The visualization result of residuals in divergence loss on x4 scaling factor.

5 More Visual Results

5.1 More Visual Comparison Results on RealSR Dataset

We show more visual comparison results in Fig. 5. Our method is compared with other methods, SRResNet [7], EDSR [8], RCAN [16], CDC [13] and LP-KPN [2]. As shown in the figure, our proposed D2C-SR framework can restore more reasonable prediction results. Other classic super-resolution methods produce unnatural results due to the ill-posed problem.

Fig. 5: More Visual comparison for x4 SR on RealSR [2] dataset. We compare Bicubic, SRResNet [7], EDSR [8], ESRGAN [12], SRFlow [9], RCAN [16], LP-KPN [2], CDC [13] and our D2C-SR.

5.2 More Visual Results on D2CRealSR Dataset

In Fig. 6, we show more visual results from our proposed D2C-SR method on D2CRealSR dataset. We crop the images pairs and show the patch to facilitate the presentation of image details. The LR images patchs are upsampled using bicubic method.

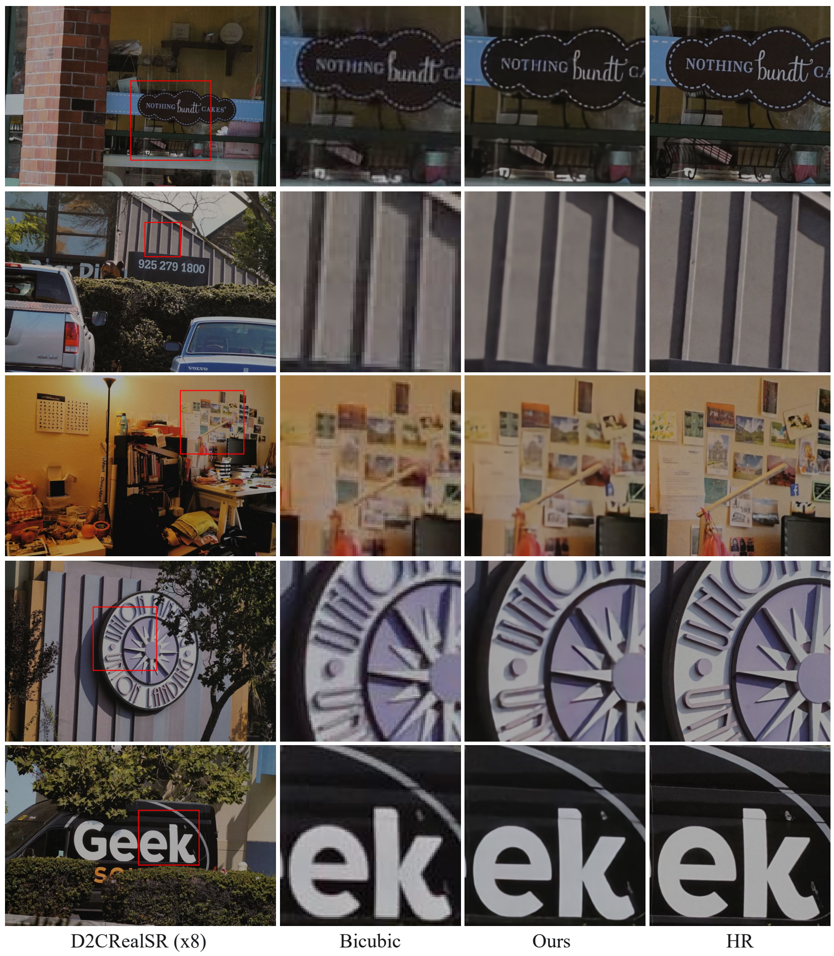


Fig. 6: More visual results on D2CRealSR dataset.

References

1. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC (2012) 4
2. 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: CVPR. pp. 3086–3095 (2019) 1, 2, 6
3. Chen, C., Xiong, Z., Tian, X., Zha, Z.J., Wu, F.: Camera lens super-resolution. In: CVPR. pp. 1652–1660 (2019) 1
4. Haris, M., Shakhnarovich, G., Ukita, N.: Deep back-projection networks for super-resolution. In: CVPR. pp. 1664–1673 (2018) 4
5. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR. pp. 1646–1654 (2016) 4
6. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: CVPR. pp. 624–632 (2017) 4
7. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR. pp. 4681–4690 (2017) 6
8. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPRW. pp. 136–144 (2017) 4, 6
9. Lugmayr, A., Danelljan, M., Van Gool, L., Timofte, R.: Srflow: Learning the super-resolution space with normalizing flow. In: ECCV. pp. 715–732 (2020) 6
10. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV. vol. 2, pp. 416–423 (2001) 4
11. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: ICCV. pp. 4539–4547 (2017) 4
12. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: ECCV. pp. 0–0 (2018) 6
13. 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) 1, 2, 6
14. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: CVPR. pp. 3262–3271 (2018) 4
15. Zhang, X., Chen, Q., Ng, R., Koltun, V.: Zoom to learn, learn to zoom. In: CVPR. pp. 3762–3770 (2019) 1
16. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV. pp. 286–301 (2018) 4, 6
17. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR. pp. 2472–2481 (2018) 4