Zero-Shot Image Super-Resolution with Depth Guided Internal Degradation Learning Supplimentary Material

Xi Cheng, Zhenyong Fu(✉), and Jian Yang(✉)

Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education,
Jiangsu Key Lab of Image and Video Understanding for Social Security,
PCA Lab, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
{chengx,z.fu,csjyang}@njust.edu.cn

1 Study on Detailed Model Desgin

1.1 Low-contrast constraint

Low-contrast constraint is an important aspect in our proposed method. Low-contrast patches contain little useful semantic and texture information and if we do not ignore these patches, they will be harmful for the convergence. In addition, more patches will also slow down the training of our network. Table 1 shows the performance and running speed comparison on image_86 with original 640×480 resolution from NYU depth v2 dataset [8] with and without the proposed low-contrast constraint. We use NIQE [6] and PI [3] as the comparison metrics.

1.2 Patch size

Patch size is also an important parameter that affects the performance. Larger patches will enlarge the receptive field and help the network achieve better results for images with repetitive contents or stronger self-similarities. However, too large patch size will increase the computational burden and reduce the number of available training patches, making the network train insufficiently. Table 2 shows the performance and running speed comparison on image_86 of NYU depth v2 dataset [8] with patch sizes from 64×64 to 256×256.

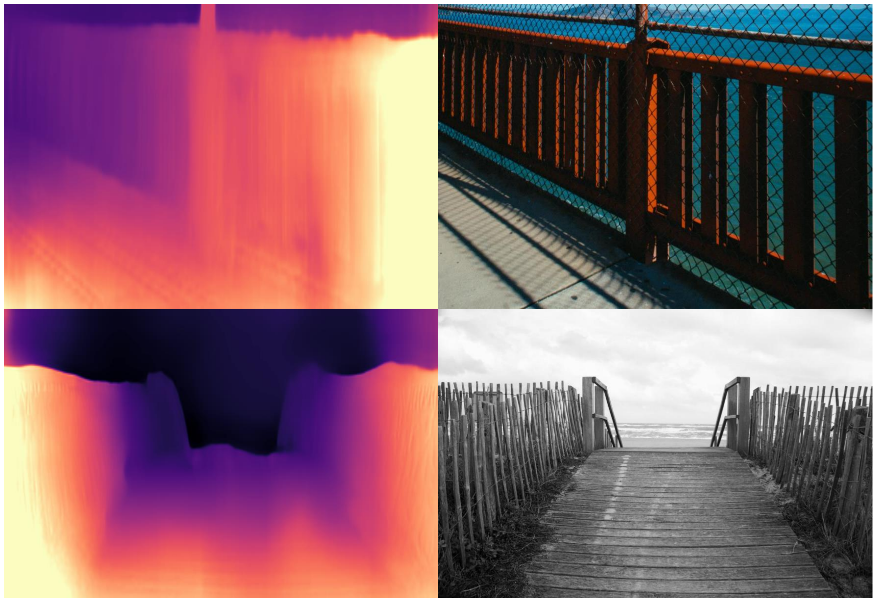

Fig. 1: Visulization of photos with different depth directions in the natural environment.The depth distribution of the first group of images is horizontal, and the depth on the left is greater than the depth on the right. The depth distribution of the second group is vertical, the upper depth is greater than the lower depth

1.3 Sampling mode

Sampling mode for the selection of useful HR/LR patches is an interesting topic. Empirically, when taking a photo, it often produces a photo with smaller depth values in the vicinity and larger depth values in the distance. This experience also holds for many datasets [4]. Thus sampling patches according to the vertical position is an empirically feasible sampling mode which does not need to calculate the depth map and is simpler to obtain the HR/LR patches. However,

Mode	NIQE (Lower is better)	PI (Lower is better)	Second/epoch
without low-contrast	**4.79**	**5.41**	**8.53**
with low-contrast	5.21	5.71	9.68

Table 1: Performance and speed comparison with and without the low-contrast constraint.

Patch-Size	NIQE (Lower is better)	PI (Lower is better)
64	5.85	6.05
96	5.32	5.72
128	**4.79**	**5.41**
192	6.04	6.12
256	5.86	6.04

Table 2: Performance and speed comparison between different patch sizes.

Mode	NIQE (Lower is better)	PI (Lower is better)
depth-guided sampling	**4.79**	**5.41**
vertical sampling	6.80	6.11

Table 3: Performance comparison with different sampling mode.

in reality, due to the different angles of the photographer, the perspective relationship between the objects captured in the image may not necessarily be vertical, i.e., it may also be left or right directions. Table 3 gives an example on image_86 from NYU dataset, super-resolved with vertical sampling and depth-guided sampling, respectively. Fig. 1 shows two sets of images taken with the camera at different angles. It can be easily found that the depth distribution on the image has a great relationship with the camera angle. Only sampling in the vertical direction will cause great instability. Therefore, using the depth prior should be a better choice.

1.4 Running time

Running time is very important for zero-shot image super-resolution. In this section we evaluate the running time of ZSSR, KernelGAN and our proposed method on images with size of 640x480 and report the average runtimes of these three methods in Table 4. KernelGAN and our DGDML-SR almost double the runtime of ZSSR. The reason is that ZSSR does not learn but adopts bicubic

Model	Running Time
ZSSR	4min46s
KernelGAN	8min23s
Ours	8min6s

Table 4: Running time comparison between ZSSR, KernelGAN and our proposed DGDML-SR.

as the degradation kernel, while KernelGAN and our method learn the image-specific degradation kernel, which causes the longer runtimes of KernelGAN and our method. Our method is a little faster than KernelGAN, although KernelGAN learns a linear degradation kernel while our method learns another image-specific degradation simulation network (DSN) as the degradation kernel.

2 Additional Visual Results

In this section, we provide more visual results on Urban100 [5] and DIV2k [1] datasets via our proposed method. Fig. 2, Fig. 3 and Fig. 4 show the full image and the estimated depth map on the left. On the right, we show the detailed visual comparison of our method with RCAN [9], ZSSR [7] and Kernel GAN [2]. The patches are taken from the red region of the images on the left. Among these methods, our proposed DGDML-SR can generate images with the most delicate textures with the least artifacts, leading to the best visual quality.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
2. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. arXiv preprint arXiv:1909.06581 (2019)
3. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 pirm challenge on perceptual image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops (September 2018)
4. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
5. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5197–5206 (2015)
6. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2012)
7. Shocher, A., Cohen, N., Irani, M.: "zero-shot" super-resolution using deep internal learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3118–3126 (2018)
8. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European Conference on Computer Vision. pp. 746–760. Springer (2012)
9. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301 (2018)

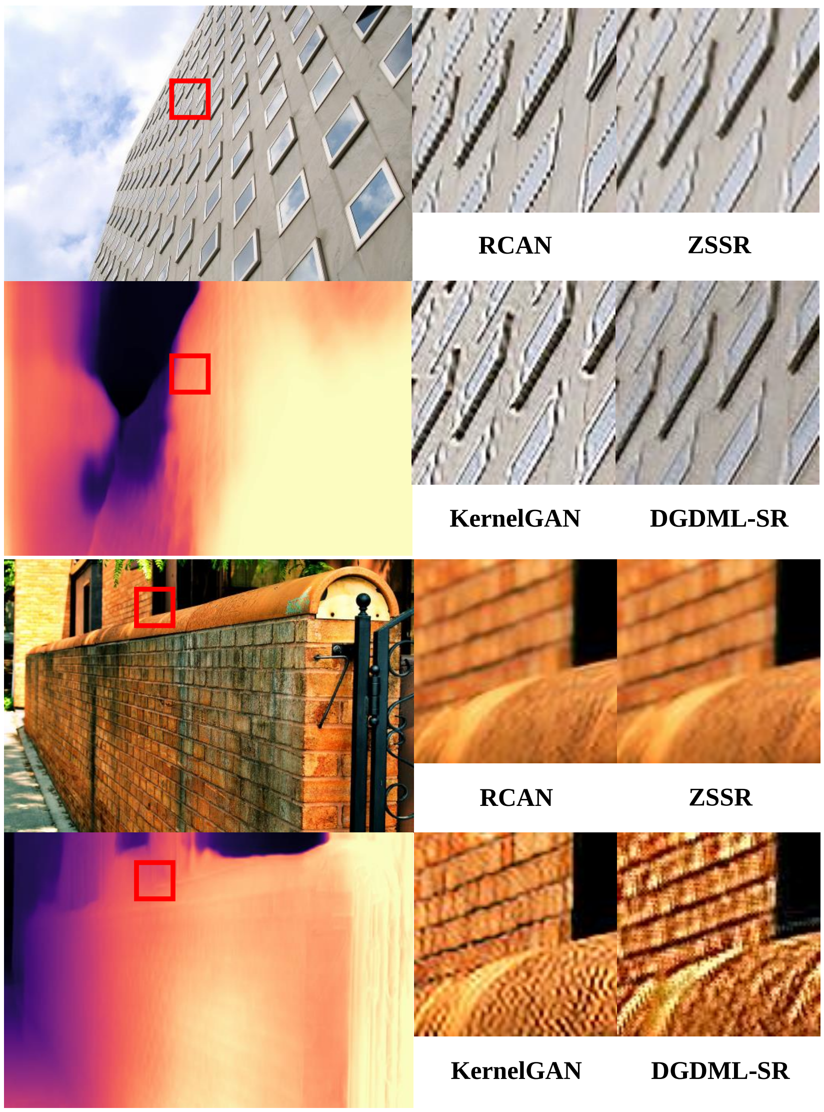

Fig. 2: Visual comparison of our proposed DGDML-SR with RCAN, ZSSR and Kernel GAN on Urban100 dataset with estimated depth.

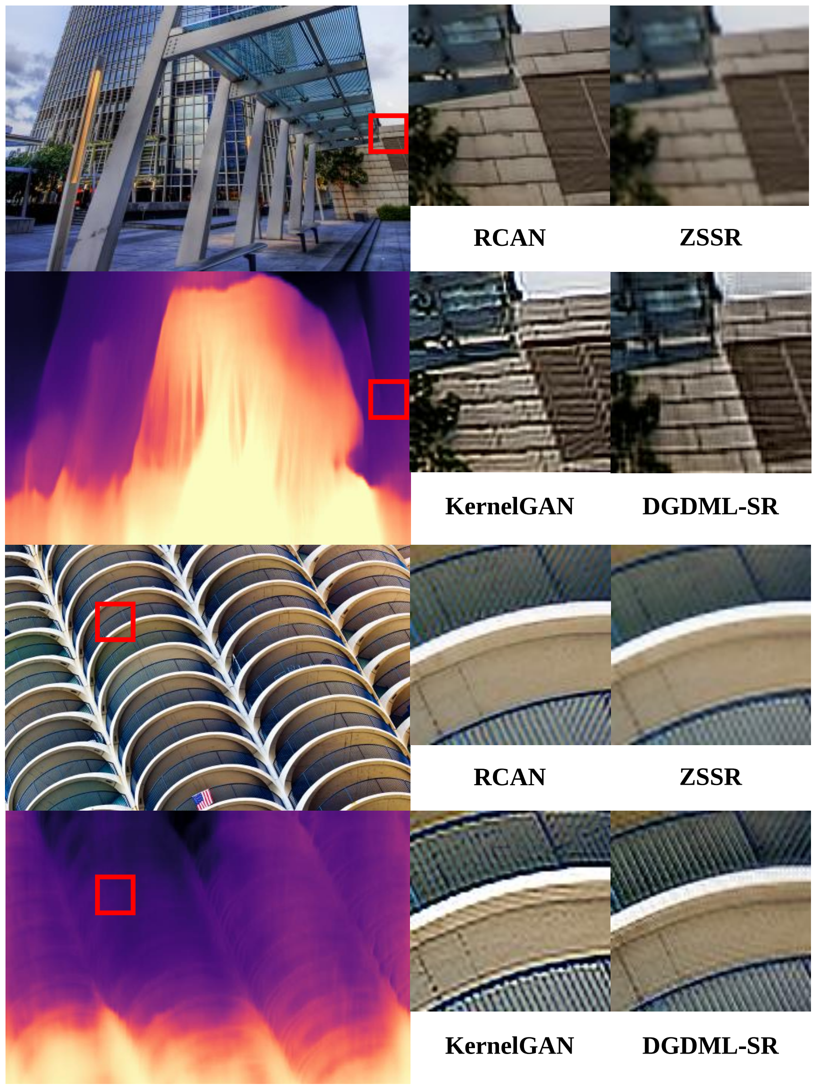

Fig. 3: Visual comparison of our proposed DGDML-SR with RCAN, ZSSR and Kernel GAN on Urban100 dataset with estimated depth.

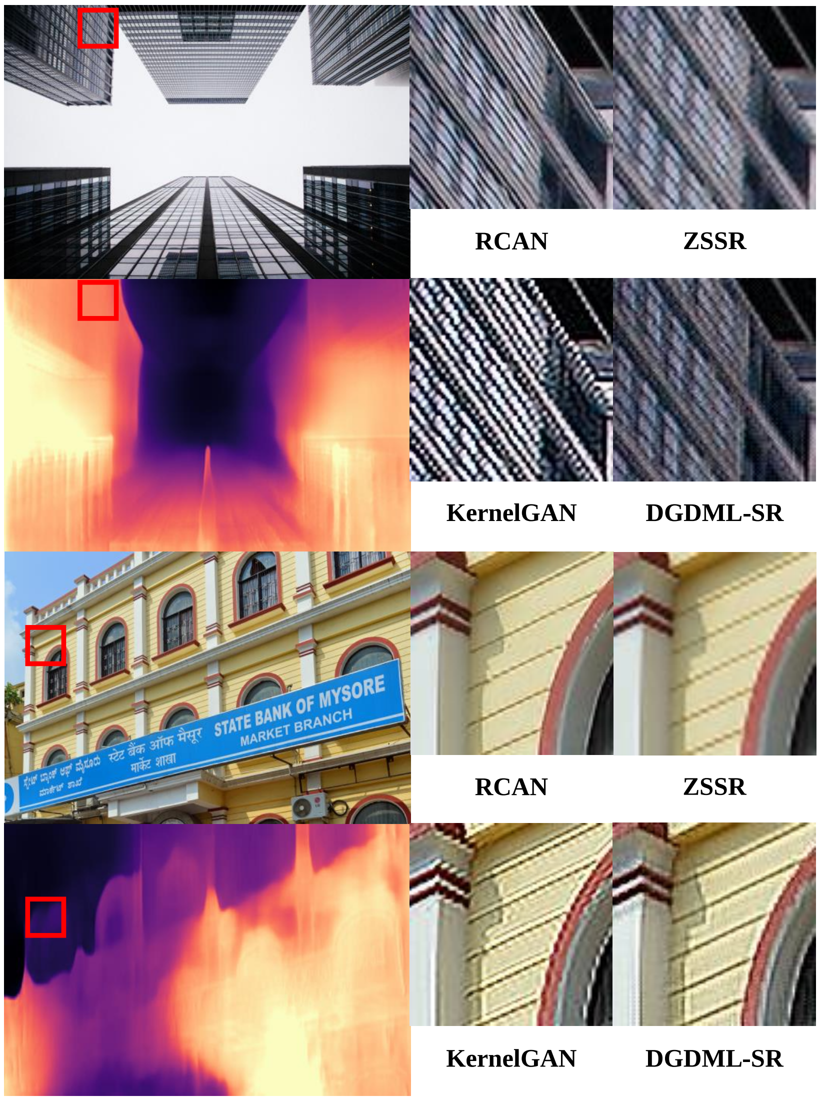

Fig. 4: Visual comparison of our proposed DGDML-SR with RCAN, ZSSR and KernelGAN on the DIV2K dataset with estimated depth.