LIIF-GAN: Learning Representations With Local Implicit Image Function and GAN for Realistic Images on a Continuous Scale (Supplementary Material)

Jun Seok Kang[1,2]
js.kang@ust.ac.kr
Sang Chul Ahn[1,2]
asc@kist.re.kr

[1] KIST school
University of Science and Technology
Seoul, Republic of Korea

[2] Artificial Intelligence and Robotics Institute
Korea Institute of Science and Technology
Seoul, Republic of Korea

1 Detailed explanation of the LIIF-GAN architecture

The detailed structures of the encoder and discriminator that are used in the LIIF-GAN are in Fig. 1. The encoder is based on the EDSR[2]. Sixteen residual blocks are used to construct the encoder. Each block contains two conv2d layers, a ReLU activation function, a scalar multiplication layer, and a skip connection. The output features of the 8th residual block are used as mid-layer features. The output features of the last Conv2d layer are used as the final layer features. The structure of the discriminator follows the ESRGAN[3]. The discriminator consists of four discriminator blocks and one conv2d block. Except for the first block, the discriminator blocks contain two conv2d layers, two batch normalization layers, and two leakyReLU activation functions. For the first block, we remove the first batch normalization. The kernel size of each conv2d layer is 3x3.

2 Additional Quantitative Evaluation: CelebA-HQ

As described in the main paper, we place the quantitative evaluation results for the CelebA-HQ dataset in Table 1. Like in the case of the DIV2K results, the results of EDSR, ESRGAN, and Real-ESRGAN, which target fixed-scale super-resolution, show promising results on the trained scale(x4) but huge performance degradation on the untrained scales. In contrast, the LIIF-GAN works better even on untrained scales. The LPIPS scores of the LIIF-GAN are better than those of the LIIF on all scales. Note that the LIIF-GAN has lower PSNR/SSIM scores than the LIIF because improving perceptual quality often means sacrificing PSNR/SSIM scores. It is well-known as a perception-distortion tradeoff[1]. Also, the

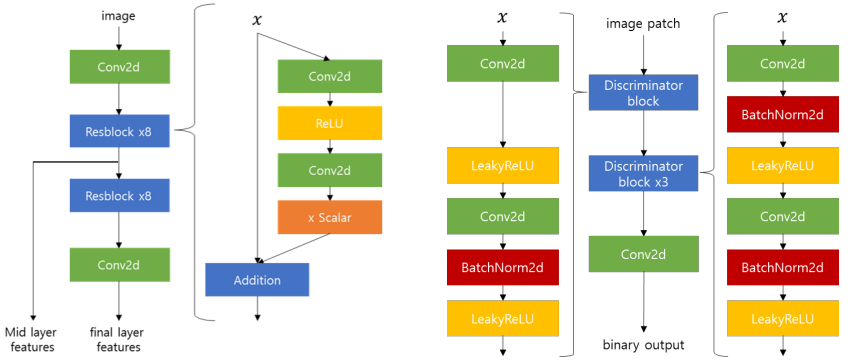

Figure 1: (a) The detailed architecture of the encoder in the LIIF-GAN. As a backbone, we use the EDSR network, which contains 16 residual blocks. (b) The detailed architecture of the discriminator in the LIIF-GAN. We use a simple feed-forward network that contains four discriminator blocks.

LIIF-GAN shows better PSNR and LPIPS scores than the ESRGAN and Real-ESRGAN on all scales except the x4 scale. While two ablation models (LIIF-GAN-S and LIIF-GAN-SF) have similar PSNR/SSIM and LPIPS scores to the LIIF-GAN, we have already checked that the output of the LIIF-GAN model can represent image structure better in qualitative comparison. In summary, the LIIF-GAN can reconstruct realistic images better on a continuous scale than conventional methods and ablation models. Note that the SSIM scores of the LIIF-GAN are lower than those of the ESRGAN. But we also see that the LIIF-GAN's and ESRGAN's scores on the x4 scale (in-distribution scale) are even below the bicubic's. So, we think that the SSIM score on the CelebA-HQ dataset is less reliable than others.

2.1 Additional Qualitative Results on the DIV2K dataset

Figs. 2 and 3 show several cropped result images for the DIV2K dataset. Each image is reconstructed from a four-times downscaled input image. The images in the first column are from the LIIF, and those in the second column are from the LIIF-GAN. We can see the details of textures have disappeared in the images of the LIIF. For example, the details of birds' wing feathers, the grass in the mountains, the textures of snow, the patterns in the butterfly's body, the penguin's feathers, the patterns on the helm, the owl's feathers, and the textures of the house roofs are not clearly visible. On the contrary, in the results of the LIIF-GAN, we can see that those are well represented. From the results, we can conclude that the LIIF-GAN method is more advantageous for obtaining realistic images than the previous SOTA (LIIF) method.

2.2 Additional Qualitative results on the CelebA-HQ dataset

Figs. 4 and 5 show the cropped result images on the CelebA-HQ dataset. Each image is four times up-scaled from a 64x64 resolution image. The images in the first and third columns are from the LIIF, and those in the second and fourth columns are from the LIIF-GAN. We can see the details of textures have disappeared in the images of the LIIF. For example, the

Table 1: The PSNR, SSIM, and LPIPS scores on the CelebA-HQ dataset. Each score is calculated by comparing it with the ground truth image. The up and down arrows mean that the higher and lower scores are better, respectively. N/A means not available. The bold text denotes the best scores. The texts in red indicate the best scores among the models adopting GAN. Note that there is a perception-distortion tradeoff[1], meaning that adoption of GAN can degrade the PSNR/SSIM scores.

Upscale		x2	x3	x4	X4.5	x5	X5.5	x6	X6.5	x7
Bicubic	PSNR↑	30.9714	29.2363	28.4716	28.2475	28.0842	27.9647	27.8762	27.8063	27.7518
	SSIM↑	0.9341	0.8774	0.8347	0.8193	0.8068	0.7973	0.7899	0.7846	0.7805
	LPIPS↓	0.0833	0.2030	0.2879	0.3136	0.3372	0.3556	0.3699	0.3830	0.3930
EDSR(x4)	PSNR↑	N/A	N/A	**31.5746**	28.5453	28.1175	27.9908	27.9038	27.8249	27.7682
	SSIM↑	N/A	N/A	**0.8941**	0.8282	0.8107	0.8000	0.7919	0.7857	0.7810
	LPIPS↓	N/A	N/A	0.1266	0.2110	0.2362	0.2752	0.3220	0.3568	0.3771
ESRGAN(x4)	PSNR↑	N/A	N/A	28.7142	27.8192	27.4741	27.489	27.5702	27.5575	27.5219
	SSIM↑	N/A	N/A	0.8343	0.8100	0.7944	0.7873	0.7829	0.7790	0.7753
	LPIPS↓	N/A	N/A	**0.0464**	0.1392	0.1757	0.2355	0.2988	0.3401	0.3643
Real-ESRGAN(x4)	PSNR↑	N/A	N/A	26.2777	26.1900	26.1801	26.1369	26.0222	25.8703	25.6724
	SSIM↑	N/A	N/A	0.7706	0.7576	0.7516	0.7451	0.7380	0.7317	0.7257
	LPIPS↓	N/A	N/A	0.0995	0.1099	0.1182	0.1289	0.1414	0.1547	0.1680
LIIF	PSNR↑	**35.5841**	**32.7435**	31.4977	**31.122**	**30.8457**	**30.6499**	**30.4998**	**30.3812**	**30.2843**
	SSIM↑	**0.9696**	**0.9284**	0.8927	**0.8785**	**0.8662**	**0.8561**	**0.8475**	**0.8406**	**0.8345**
	LPIPS↓	0.0152	0.0683	0.1271	0.1491	0.1671	0.1822	0.1942	0.2046	0.2133
LIIF-GAN-S	PSNR↑	32.9169	29.898	28.7579	28.3655	28.0506	27.9345	27.8234	27.7208	27.6255
	SSIM↑	0.9491	0.8807	0.8276	0.8052	0.7847	0.7727	0.7616	0.7519	0.7429
	LPIPS↓	0.0079	0.0249	0.0478	**0.0603**	**0.0740**	**0.0860**	**0.0970**	**0.1078**	**0.1177**
LIIF-GAN-SF	PSNR↑	32.9727	30.0689	28.6173	28.1539	27.8675	27.5539	27.3645	27.1901	27.1289
	SSIM↑	0.9491	0.8823	0.8219	0.7971	0.7752	0.7587	0.7440	0.7308	0.7169
	LPIPS↓	0.0080	**0.0245**	0.0479	0.0611	0.0767	0.0908	0.1040	0.1148	0.1258
LIIF-GAN	PSNR↑	32.6175	30.4208	28.7587	28.3877	28.4902	27.9698	27.8260	27.7487	27.8936
	SSIM↑	0.9146	0.8895	0.8289	0.8084	0.7981	0.7764	0.7644	0.7544	0.7497
	LPIPS↓	**0.0075**	0.0300	0.0484	0.0615	0.0852	0.0891	0.1014	0.1130	0.1370

details of skin, hair, eyebrows, mustache, and eyes are not visible in the results of the LIIF. On the contrary, the results of the LIIF-GAN contain those details. The results show that the LIIF-GAN is more advantageous for obtaining realistic images than the previous SOTA (LIIF) method.

References

[1] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[2] 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR) workshops*, pages 136–144, 2017.

[3] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) workshops*, page 0, 2018.

Figure 2: The results of LIIF and LIIF-GAN on the DIV2K dataset. The images in the first column are from the LIIF, and those in the second column are from the LIIF-GAN.

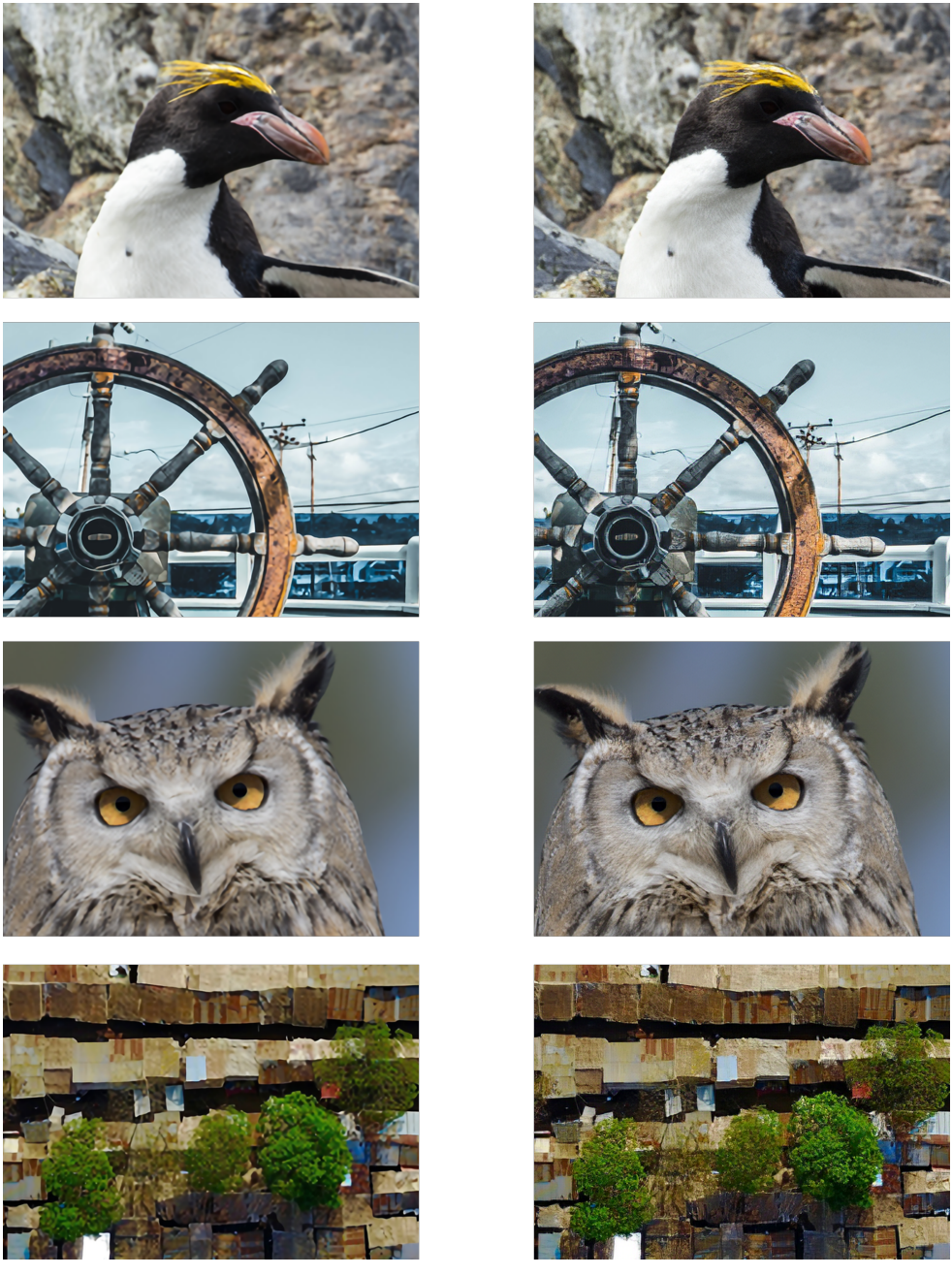

Figure 3: The results of LIIF and LIIF-GAN on the DIV2K dataset. The images in the first column are from the LIIF, and those in the second column are from the LIIF-GAN.

Figure 4: The results of LIIF and LIIF-GAN on the CelebA-HQ dataset. The images in the first and third columns are from the LIIF, and those in the second and fourth columns are from the LIIF-GAN.

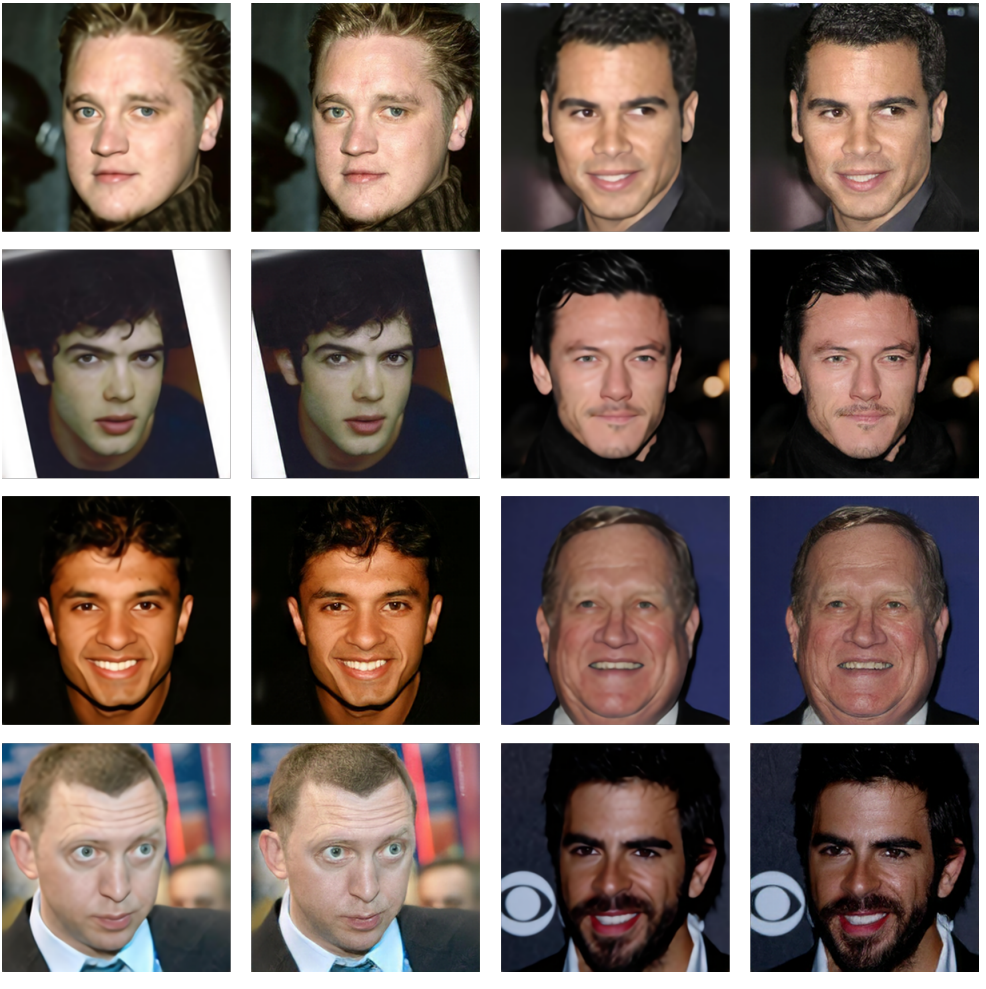

Figure 5: The results of LIIF and LIIF-GAN on the CelebA-HQ dataset. The images in the first and third columns are from the LIIF, and those in the second and fourth columns are from the LIIF-GAN.