Supplementary File: Learning the Non-differentiable Optimization for Blind Super-Resolution

Zheng Hui[1] Jie Li[1] Xiumei Wang[1] Xinbo Gao[1,2,*]
[1] Visual Information Processing Lab, Xidian University, China
[2] The Chongqing Key Laboratory of Image Cognition,
Chongqing University of Posts and Telecommunications, China
zheng_hui@aliyun.com, leejie@mail.xidian.edu.cn, wangxm@xidian.edu.cn, xbgao@mail.xidian.edu.cn

Abstract

In this supplementary file, we first provide the details of our proposed adaptive modulation generative adversarial network (AMGAN) and then show the training procedure of AMGAN-RL. Finally, we show more visual results.

1. AMGAN

VGG perceptual loss. The VGG perceptual loss $\mathcal{L}_{\text{vgg}}$ compares the activation maps in the intermediate layers of well-trained VGG19 [5] network, which can be formulated as

$$\mathcal{L}_{\text{vgg}} = \sum_{l=1}^{5} w^l \frac{\left\| \Psi^l_{\mathbf{I}^{\text{HR}}} - \Psi^l_{\mathbf{I}^{\text{SR}}} \right\|_1}{N_{\Psi^l_{\mathbf{I}^{\text{HR}}}}},$$

where $\Psi^l_{\mathbf{I}^*}$ is the activation map of the relu$l$_1 layer given original input $\mathbf{I}^*$, $N_{\Psi^l_{\mathbf{I}^{\text{HR}}}}$ is the number of elements in $\Psi^l_{\mathbf{I}^{\text{HR}}}$. Following [2], we set $w^l = \frac{1e3}{\left(C_{\Psi^l_{\mathbf{I}^{\text{HR}}}}\right)^2}$. Here, C is the channel size of feature map $\Psi^l_{\mathbf{I}^{\text{HR}}}$.

Realness adversarial loss. Different from the standard GAN (including relativistic GAN [3]), RealnessGAN [6]'s discriminator outputs a distribution as the measure of realness.

$$\begin{aligned}\mathcal{L}_{\text{realness}} = &\ \mathbb{E}_{\mathbf{I}^{\text{LR}} \sim p_{\mathbf{I}^{\text{LR}}}} \left[\mathcal{D}_{\text{KL}} \left(\mathcal{A}_1 \| \mathcal{D} \left(G \left(\mathbf{I}^{\text{LR}} \right) \right) \right) \right] \\ &- \mathbb{E}_{\mathbf{I}^{\text{LR}} \sim p_{\mathbf{I}^{\text{LR}}}} \left[\mathcal{D}_{\text{KL}} \left(\mathcal{A}_0 \| \mathcal{D} \left(G \left(\mathbf{I}^{\text{LR}} \right) \right) \right) \right],\end{aligned}$$

where $\mathcal{D}_{\text{KL}}(\cdot|\cdot)$ denotes Kullback-Leibler (KL) divergence. Two virtual ground-truth distributions are required to stand for the realness distributions of real and fake images. We refer to these two distributions as $\mathcal{A}_1$ (real) and $\mathcal{A}_0$ (fake). Specifically, $\mathcal{A}_1$ and $\mathcal{A}_0$ are chosen to resemble the shapes of a Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and a Uniform distribution $\mathcal{U}(\mathbf{0}, \mathbf{I})$, respectively. The architecture of our discriminator is illustrated in Figure 1.

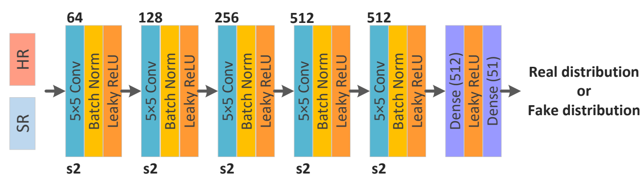

Figure 1. The structure of our discriminator. "s2" denotes the stride os 2.

Total loss. With L1 loss, VGG perceptual loss, realness adversarial loss, and DISTS loss [1], our final loss function is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_1 + 15\mathcal{L}_{\text{vgg}} + 0.01\mathcal{L}_{\text{realness}} + 10\mathcal{L}_{DISTS}.$$

To validate the effectiveness of the introduced *Realness GAN*, we conduct the same model trained with *Relativistic GAN*, denoted by AMGAN*. As shown in Table 1, AMGAN perform superior performance (PSNR & SSIM) than AMGAN*.

Table 1. Average LPIPS/PSNR results. The comparison is conducted using kernel widths 0.2, 1.3 and 2.6. Here, AMGAN* is trained with *Relativistic GAN*.

Dataset	Scores	AMGAN	AMGAN*
Set5	LPIPS↓	**0.0707**	0.1078
	PSNR↑	**31.02**	29.29
Set14	LPIPS↓	**0.1254**	0.2048
	PSNR↑	**27.14**	26.38
BSD100	LPIPS↓	**0.1698**	0.2862
	PSNR↑	**26.36**	25.63
PIRM_Val	LPIPS↓	**0.1197**	0.2291
	PSNR↑	**26.50**	25.91

*Corresponding author

2. The training procedure of AMGAN-RL

The algorithm of the detailed training process is summarized in Algorithm 1.

Algorithm 1 Training AMGAN-RL

Require: Critic network $Q\left(s, a\left|\theta^{Q}\right.\right)$ (random initialization); Pretrained actor $\mu\left(s\left|\theta^{\mu}\right.\right)$; Initialize replay buffer R; Initialize **Environment** with pre-trained renderer AMGAN.

1: **for** $t_{steps} < maxsteps$ **do**
2: Receive initial observation state s_t
3: Select action $a_t = \mu\left(s_t\left|\theta^{\mu}\right.\right)$ according to the current policy
4: Execute action a_t and observe reward r_t and observe rendered result s_t'
5: Store transition $(s_t, a_t, r_t, {s'}_t)$ in R
6: Sample a random minibatch of N transitions $(s_t, a_t, r_t, {s'}_t)$ from R
7: Update critic by minimizing the loss: $L\left(\theta^{Q}\right) = \frac{1}{N}\sum_i \max\left(0, -\gamma * Q\left(s_i, a_i\left|\theta^{Q}\right.\right)\right) + \left(r_i - Q\left(s_i, a_i\left|\theta^{Q}\right.\right)\right)$
8: Update the actor policy using the policy gradient: $\nabla_{\theta^{\mu}} J = \frac{1}{N}\sum_i \nabla_a Q\left(s, a\left|\theta^{Q}\right.\right)\Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu\left(s\left|\theta^{\mu}\right.\right)\Big|_{s=s_i}$
9: **return** $\mathbf{I}^{\mathrm{SR}} = \mathrm{AMGAN}\left(\mathbf{I}^{\mathrm{LR}}, \mu\left(\mathbf{I}^{\mathrm{LR}}\left|\theta^{\mu}\right.\right)\right)$ (Output the final SR result)

3. Visual results

We give more comparisons with methods of DAN [4] (blind) and USRGAN [7] (non-blind) on PIRM_Val (Figure 2), BSD100, and Set5 (Figure 3). It can be found that our AMNet-RL and AMNet_L-RL shows a little bit sharper than DAN. AMGAN produces texture naturally images, and AMGAN-RL generates sharper details.

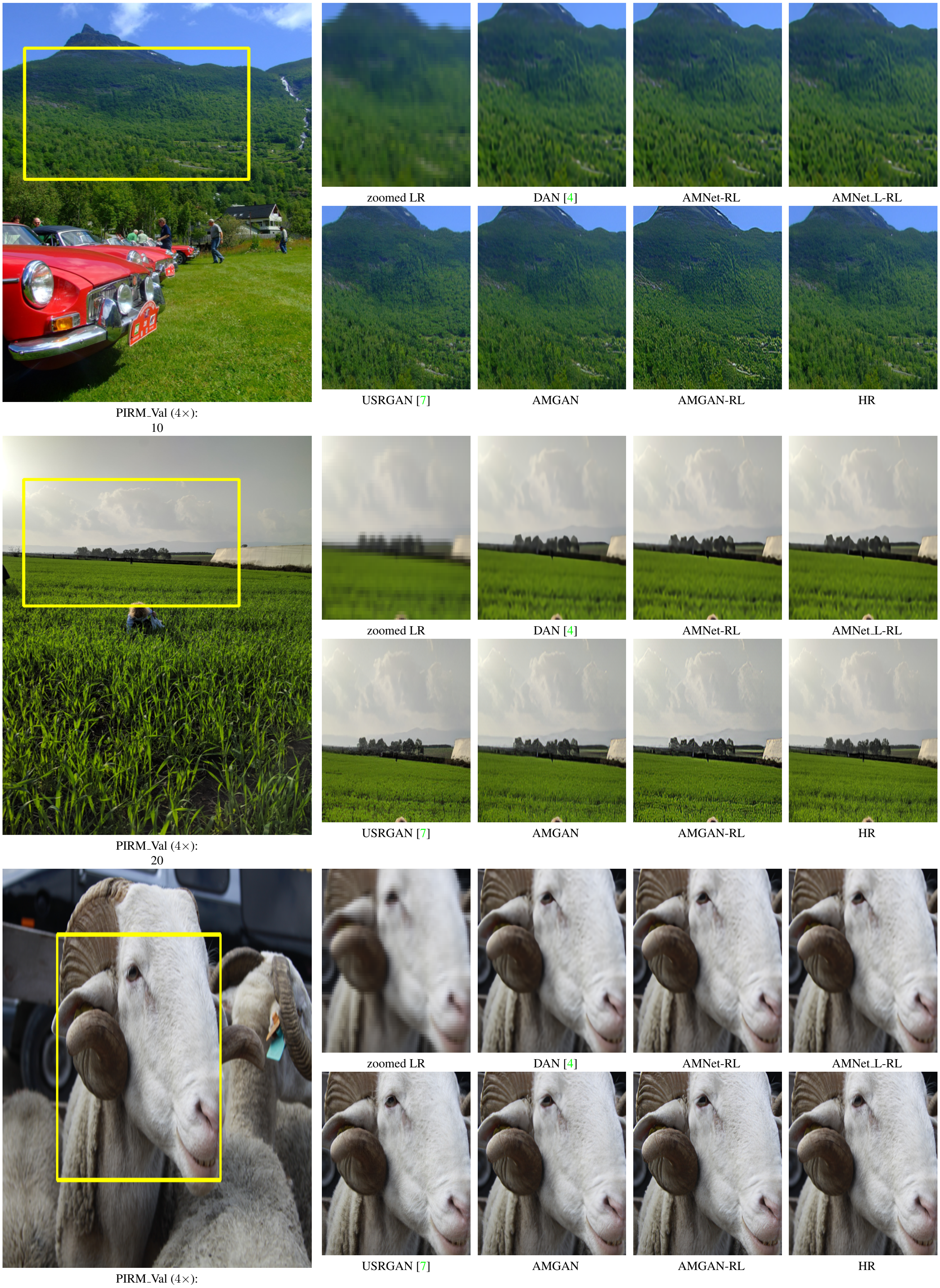

Figure 2. Visual results of various methods at scaling factor of 4. Note that USRGAN [7] adopts direct downsampling instead of *bicubic* downsampling, which is more simple than DAN [4] and our proposed method.

Figure 3. Visual results of various methods at scaling factor of 4. Note that USRGAN [7] adopts direct downsampling instead of *bicubic* downsampling, which is more simple than DAN [4] and our proposed method.

References

[1] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *arXiv preprint arXiv:2004.07728*, 2020. 1

[2] Zheng Hui, Jie Li, Xiumei Wang, and Xinbo Gao. Image fine-grained inpainting. *arXiv preprint arXiv:2002.02609*, 2020. 1

[3] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. In *ICLR*, 2019. 1

[4] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. Unfolding the alternation optimization for blind super resolution. In *NeurIPS*, 2020. 2, 3, 4

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 1

[6] Yuanbo Xiangli, Yubin Deng, Bo Dai, Chen Change Loy, and Dahua Lin. Real or not real, that is the question. In *ICLR*, 2020. 1

[7] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network fro image super-resolution. In *CVPR*, pages 3217–3226, 2020. 2, 3, 4