GLEAN: Generative Latent Bank for Large-Factor Image Super-Resolution Supplementary Material

Kelvin C.K. Chan[1] Xintao Wang[2] Xiangyu Xu[1] Jinwei Gu[3,4] Chen Change Loy[1]
[1]S-Lab, Nanyang Technological University
[2]Applied Research Center, Tencent PCG [3]Tetras.AI. [4]Shanghai AI Laboratory
{chan0899, xiangyu.xu, ccloy}@ntu.edu.sg xintao.wang@outlook.com gujinwei@tetras.ai

We first provide the implementation details of GLEAN in Sec. 1. We then provide additional qualitative results on various categories and scale factors in Sec. 2.1. Finally, we demonstrate the application of GELAN to the task of image retouching in Sec. 2.2.

1. Training Details of GLEAN

We adopt pre-trained StyleGAN[1] [4] or StyleGAN2[2] [5] as our generative latent bank. In this section, we assume the latent bank is pre-trained and present the training details of GLEAN (*i.e*. the encoder-bank-decoder network). Note that the weights of the latent bank are fixed when training GLEAN to better employ the generative prior and to avoid biasing to the training distribution.

We train GLEAN on five categories including human faces, cats, cars, towers, and bedrooms. The training and test datasets used in our experiments are summarized in Table 1. Since StyleGAN produces images with fixed size, we resize the images in the datasets for our experiments.

Table 1: **Datasets used in our experiments.**

	Train	**Test**
Human faces	FFHQ [4]	CelebA-HQ [3]
Cats	LSUN-train [13]	CAT [14]
Cars	LSUN-train [13]	Cars [7]
Bedrooms	LSUN-train [13]	LSUN-validate [13]
Towers	LSUN-train [13]	LSUN-validate [13]

Following previous works [11, 12], the objective function for GLEAN consists of three terms. MSE loss is used to guide the fidelity of the output images:

$$\mathcal{L}_{mse} = \frac{1}{N}||\hat{y} - y||_2^2, \tag{1}$$

where N, $\hat{y}$, and y denote the number of pixels, the output image, and the ground-truth image, respectively. We further incorporate perceptual loss [2] and adversarial loss [1] to improve the perceptual quality:

$$\mathcal{L}_{percep} = \frac{1}{N}||f(\hat{y}) - f(y)||_2^2, \tag{2}$$

$$\mathcal{L}_{gen} = \log\left(1 - D(\hat{y})\right), \tag{3}$$

where $f(\cdot)$ denotes the feature embedding space of the VGG16 [10] network, and D corresponds to the StyleGAN discriminator. The resulting objective function is a weighted mean of the three losses:

$$\mathcal{L}_g = \mathcal{L}_{mse} + \alpha_{percep}\cdot\mathcal{L}_{percep} + \alpha_{gen}\cdot\mathcal{L}_{gen}. \tag{4}$$

In all our experiments, we set α_{percep}=α_{gen}=10^{-2}. For the discriminator, we maximize

$$\mathcal{L}_d = \log\left(1 - D(\hat{y})\right) + \log D(y). \tag{5}$$

We adopt Cosine Annealing Scheme [8] and Adam optimizer [6] in training. The number of iterations is 300K and the initial learning rate is 10^{-4}. The batch size is 8 for human faces and 16 for other categories. We train our models using two Nvidia V100 GPUs.

2. Qualitative Results

2.1. Super-Resolution

Randomly-Selected Examples. In Fig. 1, we show the results of randomly-selected examples from CelebA-HQ [3]. By optimizing only the latent codes, PULSE [9] produces outputs with low-fidelity. In contrast, guided by the encoder features and our generative latent bank, GLEAN achieves remarkable quality and fidelity, demonstrating the effectiveness of our designs.

Scale Factors and Categories. GLEAN is extensible to various scale factors (from 8× to 64×) and categories (*e.g*. faces, cats, cars, bedrooms, towers). From Fig. 2 to Fig. 7, we see that GLEAN outperforms DGP and ESRGAN^+ in both fidelity and quality. It is noteworthy that the performance of DGP and ESRGAN^+ are less promising on categories other than human faces.

[1]GenForce: https://github.com/genforce/genforce
[2]BasicSR: https://github.com/xinntao/BasicSR

2.2. Image Retouching

In interactive image retouching, users can manually edit the images based on their preference. For instance, users can change the facial expression of an object and perform geometric transformations for enlarging eyes. However, a perfect output requires tedious and precise retouching. As a result, artifacts are common in the outputs from amateur retouching.

GLEAN allows the possibility of performing realistic refinement of imperfect retouching. More specifically, given a retouched image, we can first downsample the image to a smaller resolution, where the artifacts vanished. We can then upsample it back to the original resolution. With GLEAN as a powerful super-resolver, we can obtain an output with unnatural artifacts suppressed.

As shown in Fig. 8, GLEAN is able to correct the unnatural artifacts introduced by amateur retouching while being similar to the retouched images, realistic, and coherent with the unaltered regions. In addition, since GLEAN requires only a single forward pass, it can be used in interactive image editing software to allow a more flexible retouching.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 1

[2] Justin Johnson, Alexandre Alahi, and Fei-Fei Li. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2017. 1

[3] Tero Karras, Timo Ailo, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018. 1, 3

[4] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 1

[5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. *arXiv preprint arXiv:1912.04958*, 2019. 1

[6] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 1

[7] Jonathan Krause, Michael Stark, Jia Deng, and Fei-Fei Li. 3D object representations for fine-grained categorization. In *ICCV*, 2013. 1

[8] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 1

[9] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. PULSE: Self-supervised photo upsampling via latent space exploration of generative models. In *CVPR*, 2020. 1, 3

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

[11] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, 2018. 1

[12] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. ESRGAN: Enhanced super-resolution generative adversarial networks. In *ECCVW*, 2018. 1

[13] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. 1

[14] Weiwei Zhang, Jian Sun, and Xiaoou Tang. Cat head detection - how to effectively exploit shape and texture features. In *ECCV*, 2008. 1

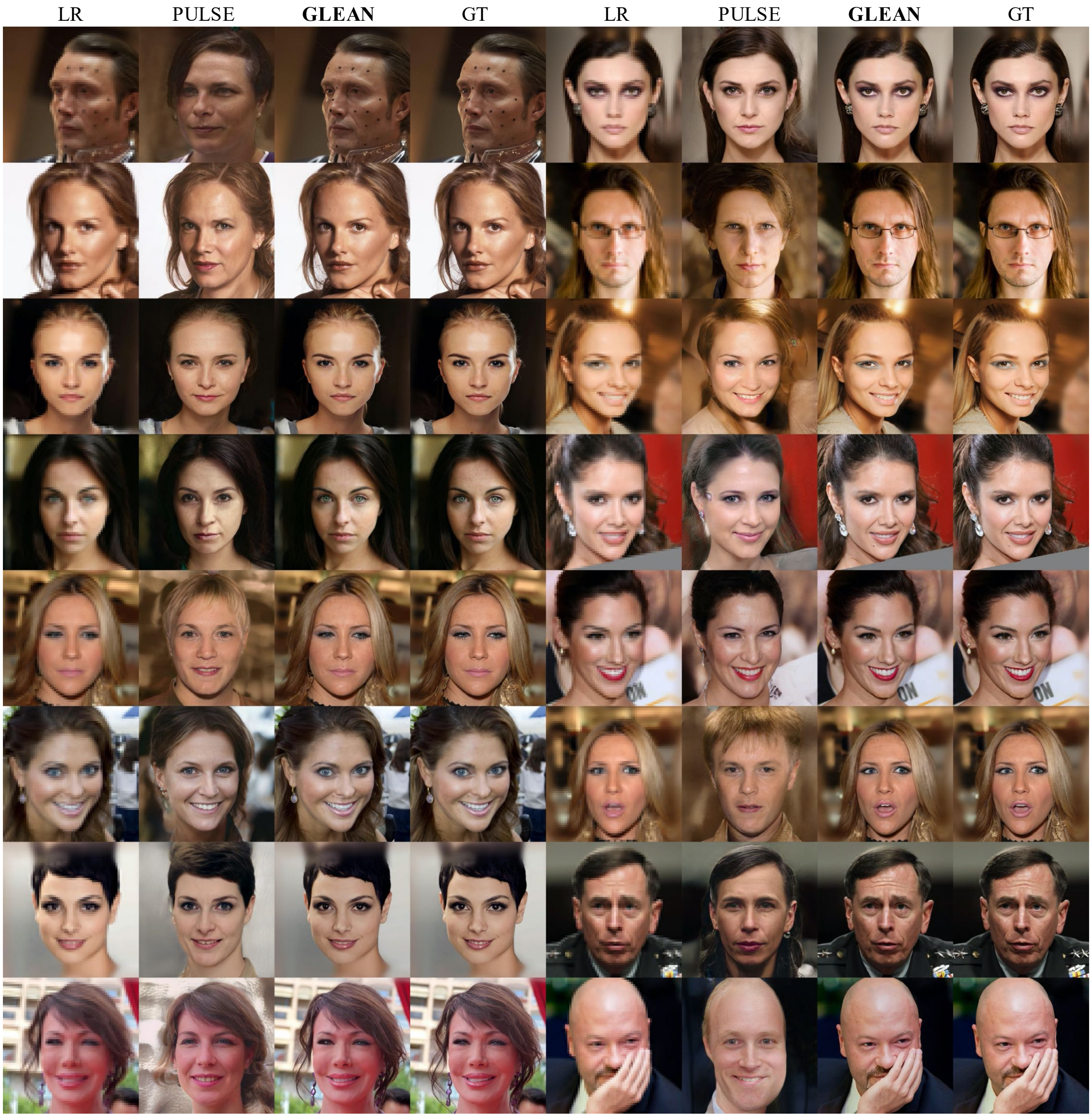

Figure 1: **Comparison to PULSE on randomly-selected examples from CelebA-HQ [3].** By optimizing only the latent vectors, the outputs of PULSE [9] differ significantly from the ground-truths. With our novel designs, GLEAN produces outputs highly similar to the ground-truths.

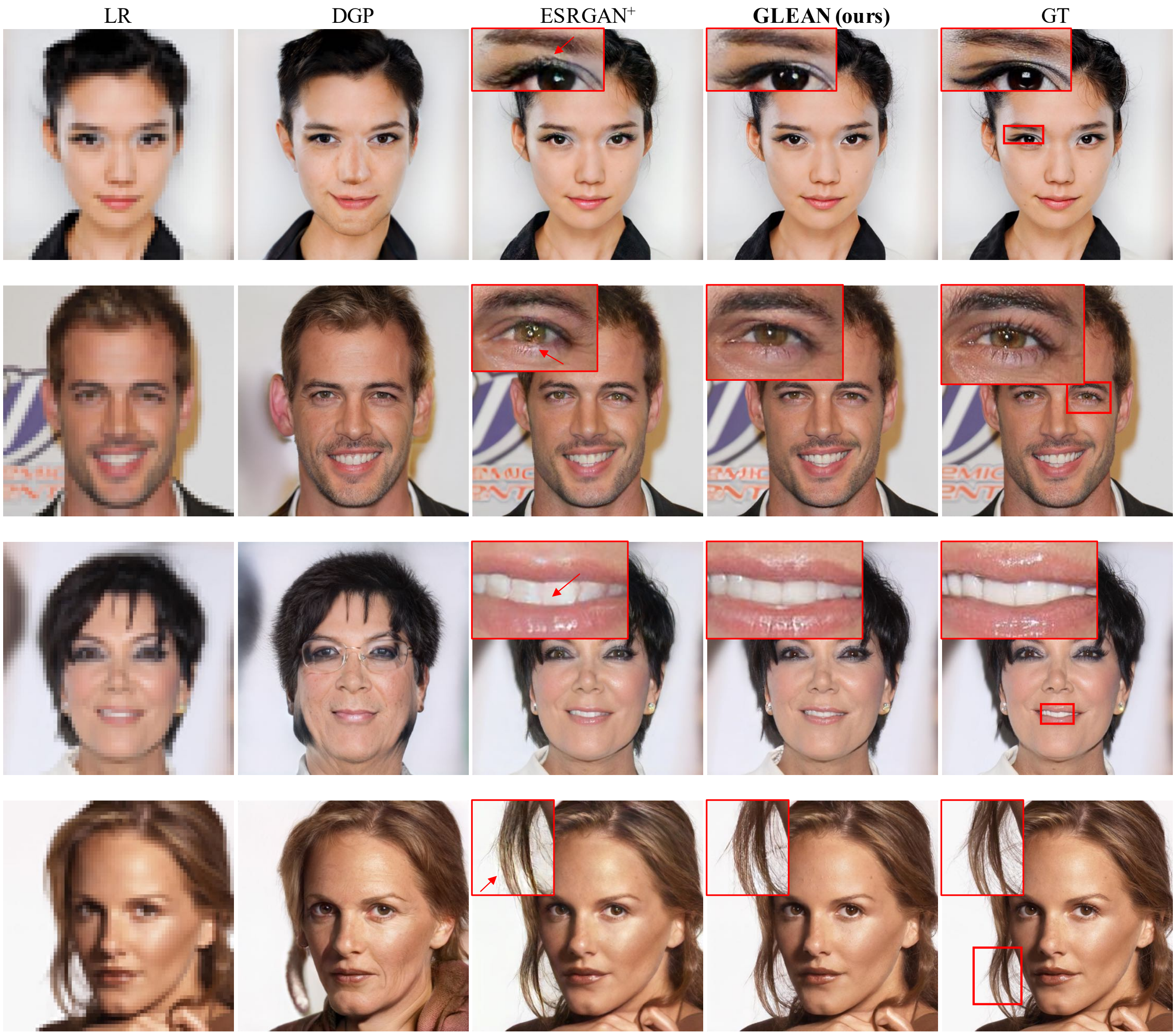

Figure 2: **Comparison with DGP and ESRGAN$^+$.** The outputs of DGP show noticeable identity differences to the ground-truths. ESRGAN$^+$ shows unpleasant artifacts for the fine details. **(Zoom-in for best view)**

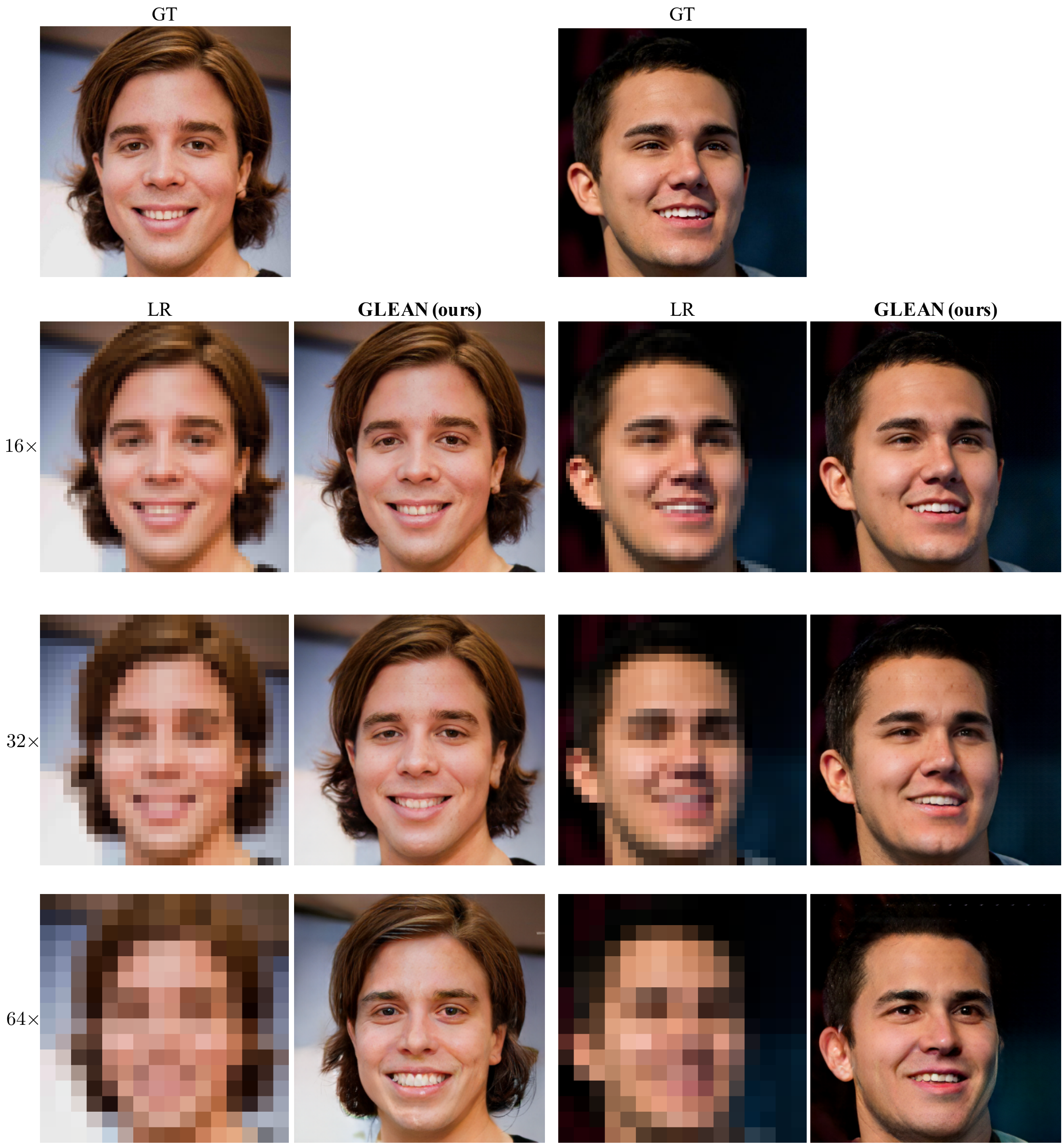

Figure 3: **Performance of GLEAN on 16×, 32×, and 64× SR.** GLEAN is able to synthesize images well resembling the ground-truths for up to 64× upsampling.

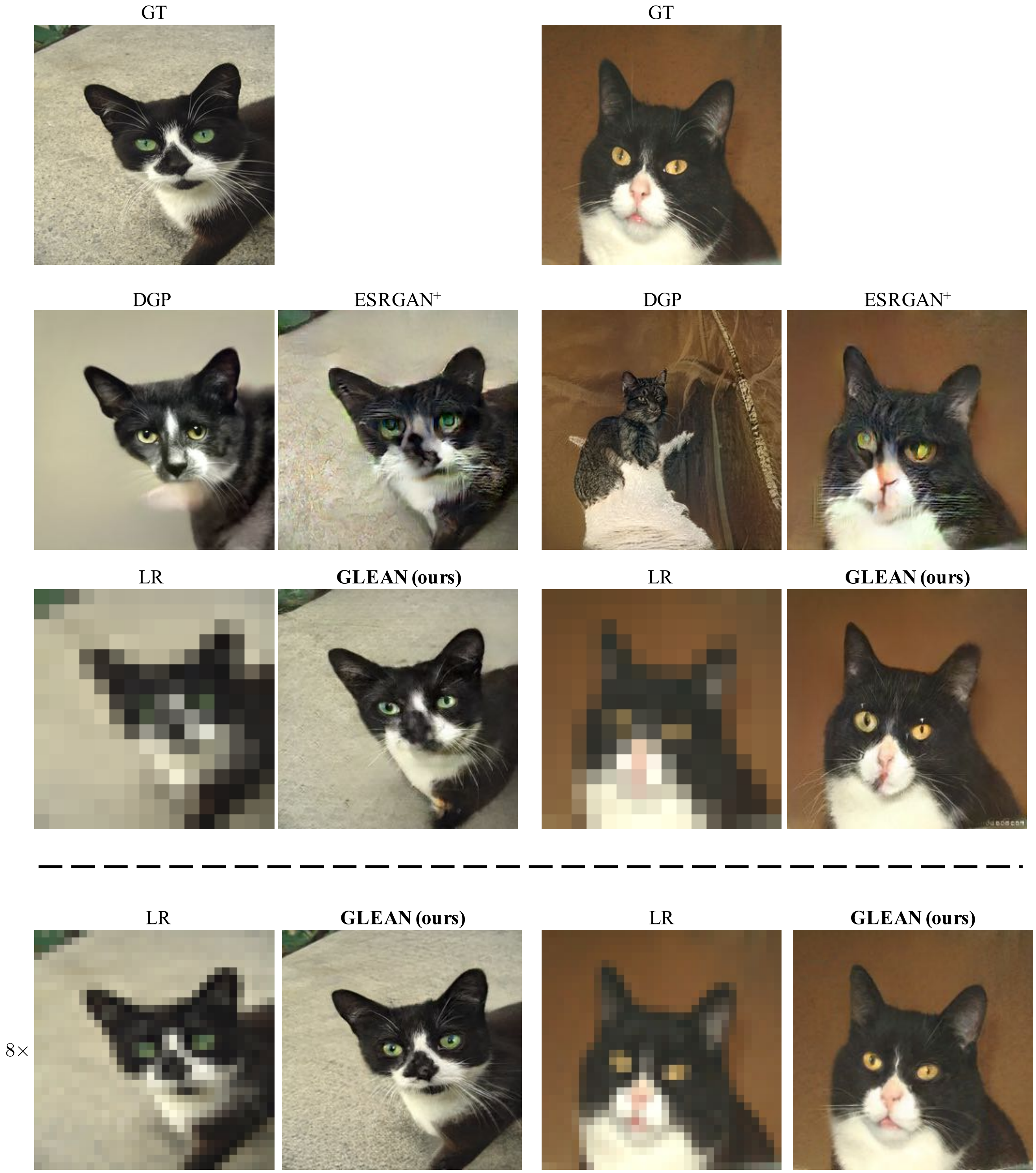

Figure 4: **(Top) Comparison with DGP and ESRGAN$^+$ on *Cats*.** DGP produces outputs with low fidelity; ESRGAN$^+$ fails to synthesize realistic textures. **(Bottom) Performance of GLEAN on 8$\times$ SR.** GLEAN produces realistic outputs that are highly similar to the ground-truths. **(Zoom-in for best view)**

Figure 5: **(Top) Comparison with DGP and ESRGAN$^+$ on *Cars*.** DGP produces outputs with low fidelity; ESRGAN$^+$ fails to synthesize realistic textures. **(Bottom) Performance of GLEAN on 8$\times$ SR.** GLEAN produces realistic outputs that are highly similar to the ground-truths. **(Zoom-in for best view)**

Figure 6: **(Top) Comparison with DGP and ESRGAN$^+$ on *Bedrooms*.** DGP produces outputs with low fidelity; ESRGAN$^+$ fails to synthesize realistic textures. **(Bottom) Performance of GLEAN on 8× SR.** GLEAN produces realistic outputs that are highly similar to the ground-truths. **(Zoom-in for best view)**

Figure 7: **(Top) Comparison with DGP and ESRGAN$^+$ on *Towers*.** DGP produces outputs with low fidelity; ESRGAN$^+$ fails to synthesize realistic textures. **(Bottom) Performance of GLEAN on 8× SR.** GLEAN produces realistic outputs that are highly similar to the ground-truths. **(Zoom-in for best view)**

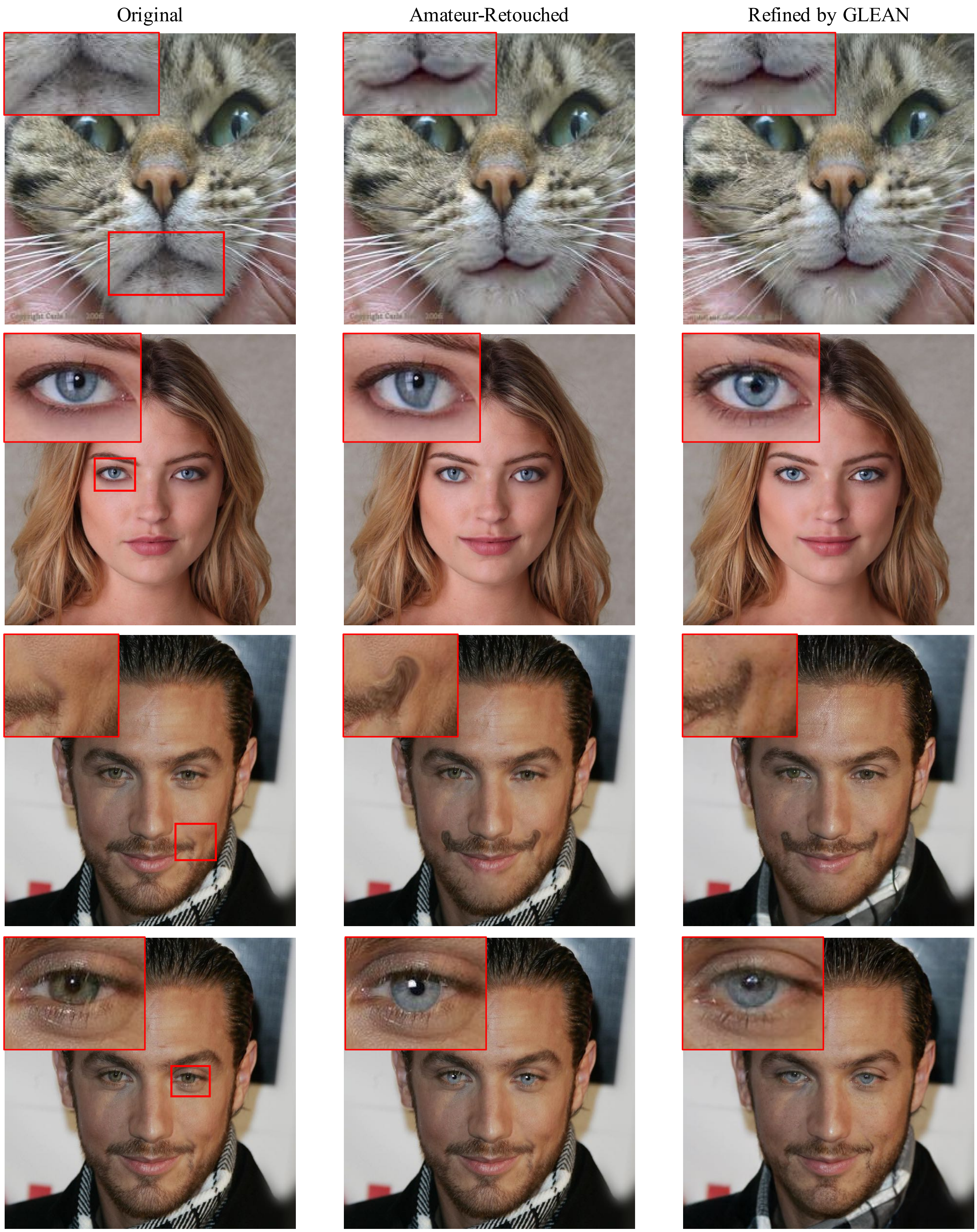

Figure 8: **Results on image retouching.** GLEAN can be used to correct unpleasant artifacts introduced by amateur retouching. **(Zoom-in for best view)**