Blind Face Super-Resolution

1 Introduction

This challenge tasks participants with creating high-quality (HQ) face images from low-quality (LQ) ones using a dataset from FFHQ. The dataset contains 5000 HQ training images and 400 LQ-HQ image pairs for validation. During training, LQ images are generated from HQ images using the random second-order degradation pipeline that includes Gaussian blur, downsampling, noise, and compression.

2 Specification

Model	***RealESRNetModel [3]** HiFaceGAN [2] EDSR (Enhanced Deep Residual Networks for Single Image Super-Resolution) [1]
Loss Function	L1 Loss
Optimizer	Adam (lr=2e-4, momentum=(0.9,0.999), wd=0)
Scheduler	CosineAnnealingRestartLR period=[150000,150000]
Epoch	80 (5000 iter/epoch) – RealESRNetModel 25 – HiFaceGAN 100 -- EDSR
Number of Parameters	1,517,571 – RealESRNetModel >3M – HiFaceGAN 1,517,595 -- EDSR
Machine	GPU RTX 4090(24GB) * 1 GPU RTX 4090(24GB) * 1 GPU NVIDIA A40 CPU12 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz Ram 90GB

3 Results

RealESRNetModel:

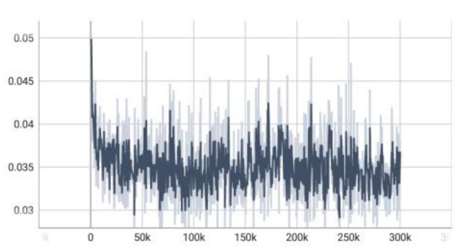

Figure 1: RealESRNetModel NIQE

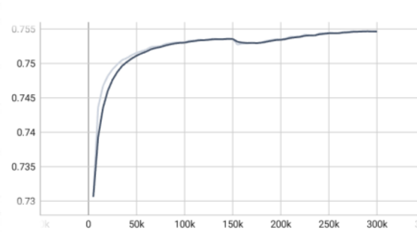

Figure 2: RealESRNetModel loss

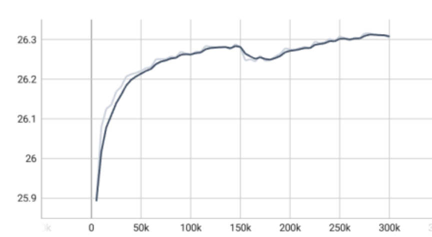

Figure 3: RealESRNetModel PSNR

Best Epoch (5000/epoch)	Training Loss	Validation NIQE	Validation PSNR	Best Test PSNR
73	3.1275e-02	0.7551	26.3272 dB	**26.51798 (14)**

RealESRNetModel with heavier degradation:

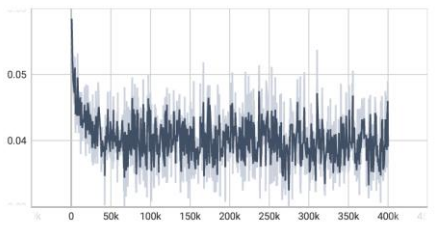

Figure 4: RealESRNetModel NIQE

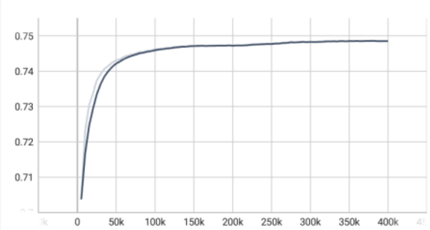

Figure 5: RealESRNetModel loss

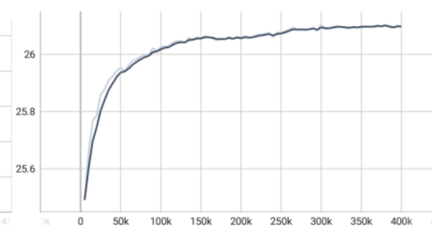

Figure 6: RealESRNetModel PSNR

Best Epoch (5000/epoch)	Training Loss L1	Validation NIQE	Validation PSNR	Test PSNR
56	3.843e-02	0.7483	26.1 dB	26.188

HiFaceGAN:

Best Epoch (5000/epoch)	Training Loss L1	Validation NIQE	Validation PSNR	Best Test PSNR
56		-	22.9 dB	-

EDSR:

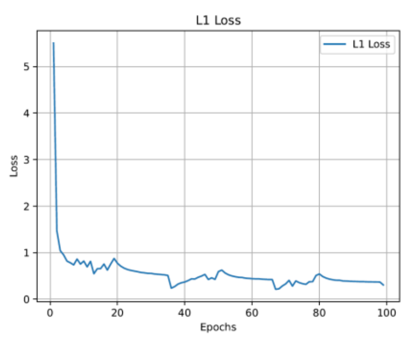

Figure 7: EDSR loss

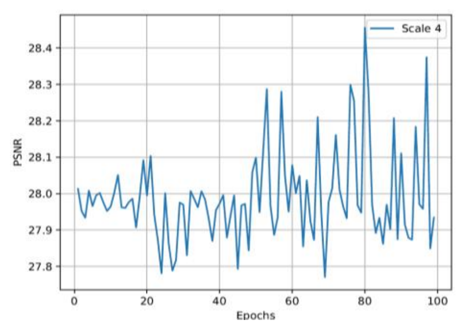

Figure 8: EDSR PSNR

Best Epoch (5000/epoch)	Training Loss	Validation NIQE	Validation PSNR	Best Test PSNR
56	0.399	-	7.16 dB	-

4 Dataset analysis and Preprocessing

Augmentation

In image super-resolution, we require pairs of low-resolution (LR) images and high-resolution (HR) images. As stated in the paper[3], the LR images are automatically generated on the fly during training through the random second-order degradation pipeline. In which Blur, Downsampling, Noise, and JPEG Compression are applied in random order twice. It is implemented in the basicsr repository, however, HiFaceGAN and EDSR do not follow this pipeline, we have implemented a degradation pipeline for them following the guidelines.

Figure 9: RealESRNetModel output

Figure 10: RealESRNetModel (heavier degradation) output

To observe the effect of image augmentation on super-resolution, we have increased the strength of degradation by widening the resize range, noise range, passion scale range, and jpeg range, while training with RealESRNetModel. The changed hyperparameters are shown in Appendix A. Part 3 shows that the model trained with heavier degradation has 0.22 dB and 0.28 dB less on validation and test set compared to the original setting, which proves heavier degradation does not necessarily improve the model's performance. Figure 9 and 10 depict that heavy augmentation would cause the model to generate more blurry results and less less able to learn pixel level details.

5 Model Selection

In this experiment, we have discovered three different methods for solving blind-face super-resolution. They are RealESRNetModel, HiFaceGAN, and EDSR. The HiFaceGAN is a GAN model which have UNet as the generator and VGG-19 as the discriminator.

HiFaceGAN is a face renovation model, it's architecture leverages a collaborative suppression and replenishment (CSR) approach. Within HiFaceGAN, multiple nested CSR units progressively replenish facial details based on hierarchical semantic guidance.

The EDSR architecture builds upon the SRResNet framework. It comprises multiple residual blocks and avoids batch normalization layers, opting for constant scaling layers to ensure consistent results.

Figure 11: Example output of the EDSR

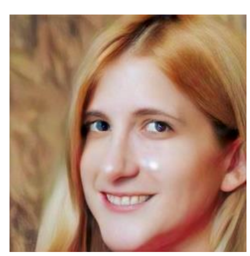

Figure 12: Example output of the HiFaceGAN

According to part 3, RealESRNetModel performs the best in validation PSNR with 26.3dB, while the HiFaceGAN and EDSR have only 22.9dB and 7.16. The low PSNR in EDSR is likely due to the loss of color information as shown in Figure 11. Meanwhile, Figure 12 shows HiFaceGAN has the potential to compare with RealESRNetModel with the help of suppression and replenishment module, as the HiFaceGAN has not converged, but we stopped the experiment due to the limitation of time and resources, as HiFaceGAN has a large model size which does not fit the objective of this experiment.

Summary

To conclude, RealESRNetModel outperformed HiFaceGAN and EDSR, achieving a validation PSNR of 26.3dB. Heavy degradation did not consistently improve performance. HiFaceGAN showed potential to achieve better result. EDSR struggled due to color information loss, achieving a validation PSNR of only 7.16dB.

Reference

[1] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image Super-Resolution," arXiv.org, https://arxiv.org/abs/1707.02921 (accessed Apr. 26, 2024).

[2] L. Yang et al., "Hifacegan: Face renovation via collaborative suppression and replenishment," arXiv.org, https://arxiv.org/abs/2005.05005 (accessed Apr. 26, 2024).

[3] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," arXiv.org, https://arxiv.org/abs/2107.10833 (accessed Apr. 26, 2024).

Appendix A

# the first degradation process	
resize_prob	[0.2, 0.7, 0.1] # up, down, keep
resize_range	[0.15, 1.5]
gaussian_noise_prob	0.5
noise_range	[1, 30]
poisson_scale_range	[0.05, 3]
gray_noise_prob	0.4
jpeg_range	[30, 95]
# the second degradation process	
second_blur_prob	0.8
resize_prob2	[0.3, 0.4, 0.3] # up, down, keep
resize_range2	[0.3, 1.5]
gaussian_noise_prob2	0.5
noise_range2	[1, 25]
poisson_scale_range2	[0.05, 2.5]
gray_noise_prob2	0.4

jpeg_range2	[30, 95]