AI6126: REPORT For Project 2

Names: Joseph UWINEZA
Matric No: G2303477F
Program: MSAI

NTU, April 2024

Abstract

This report elucidates the development and assessment of a machine learning model crafted for the purpose of blind face super-resolution. The task entailed elevating the quality of low-resolution, degraded face images to high-quality versions, employing a dataset of 5000 high-quality images for training and 400 low-quality to high-quality image pairs for validation from the FFHQ dataset [3]. The challenge demanded the formulation of the low-quality images on-the-fly during training, utilizing a specified second-order degradation pipeline. The principal metric for performance evaluation is the Peak Signal-to-Noise Ratio (PSNR)[2]. The model's architecture was constrained to less than 2,276,356 trainable parameters, and the development was steered clear of external data and pre-trained models, ensuring the authenticity of learning. This document presents the model architecture, loss functions, training curves, and and discuss the results..

1 Introduction

Face super-resolution[1], commonly termed as face hallucination, is an intricate task within the domain of image processing, aiming to reconstruct a high-resolution face image from a low-resolution counterpart[5]. The intricacies of human facial features demand an approach that is both delicate and precise, further compounded when dealing with blind super-resolution where the degradation process is unknown. This report documents the journey of crafting such a model as part of AI6126's second project, aligning with the stipulations laid out in the challenge description.

The endeavor embarked on with the training of the network using a mini-dataset provided, which was followed by meticulous tuning of hyper-parameters upon validation. A key aspect of the challenge was the imperative on-the-fly generation of low-quality images using the degradation pipeline provided during training. Moreover, to align with real-world applications, the model's architecture was constrained to a parameter count reflective of practical deployment considerations.

A salient feature of the project was the dual emphasis on quantitative PSNR scores and qualitative perceptual quality. The latter was particularly underscored for solutions that ventured to implement adversarial training within a Generative Adversarial Network (GAN) framework, aiming for results that resonate with the natural human perception of faces.

Subsequent sections of this report will delve into the specifics of the model used, elucidating on the chosen loss functions and the rationale behind them, the training strategy employed, and the experimental results obtained. Furthermore, the report will also touch upon the additional enhancements and data processing techniques that were harnessed to attain the final results. This introduction sets the stage for a detailed narrative of the methodology and insights gleaned from this intriguing task of blind face super-resolution.

2 Data preprocessing

The dataset for this project consists of images from the FFHQ dataset, including a training set of 5000 high-quality images and a validation set with 400 pairs of images. An initial exploratory analysis assessed the dataset's distribution to confirm its fitness for the intended training. A key aspect of dataset preparation was the synthetic degradation process. Since low-quality training images were not provided, we generated them by introducing various artifacts such as Gaussian blur, downsampling, noise, and compression to high-quality images. This method was designed to closely mimic real-world image degradation scenarios.

To ensure diversity in degradation and enhance the model's ability to generalize, low-quality images were created dynamically during training. The model encountered a variety of imperfections, bolstering its adaptability. The validation set was subjected to the same degradation pipeline to maintain consistent evaluation standards. All preparation procedures, including this degradation pipeline, were automated to guarantee reproducibility and allow for straightforward replication of the dataset. Throughout the data preparation stage, no external data was used to comply with challenge guidelines and to test the model's effectiveness solely on the given dataset. This approach guaranteed a well-rounded and realistic training environment for the model to learn super-resolution on degraded face images.

Two degradation steps were implemented in data preprocessing. Initially, images could be upscaled, downscaled, or remain unchanged based on specific probabilities, with a range of alterations to mimic natural image degradation. The second step introduced another layer of degradation, enhancing the robustness of the model. The final touch might include a sinc filter, recreating the sinc blur effect often observed in real-life images.

3 Model Architecture

The model employed is a Modified SRResNet (Super-Resolution Residual Neural Network) without Batch Normalization. It is part of the RealESRNet model type [4], specifically for a scale of 4x super-resolution. The architecture consists of: 3 input channels and 3 output channels (for RGB images). 64 feature maps. 16 residual blocks. An upscaling factor of 4. This model is designed to learn an end-to-end mapping from low-resolution to high-resolution images. This model architecture sum up to **1,517,571** parameters.

4 Loss Function

The L1 Loss function is central to the training of the super-resolution model, prioritizing high-fidelity reconstructions by penalizing the absolute differences between the super-resolved output and high-resolution ground truth. With a loss weight of 1.0, it places a premium on pixel-level accuracy, essential for capturing the fine details necessary for realistic images. Consequently, the model is trained to produce super-resolved images that are visually compelling and true to the original high-resolution source.

5 Evaluation Metrics

The Modified SRResNet model utilizes key evaluation metrics to assess its performance in generating high-resolution images from low-quality inputs on the FFHQ dataset. Validation is conducted every 5000 iterations to monitor progress and detect issues, incorporating both visual inspections and quantitative measures.

Key Metrics: Peak Signal-to-Noise Ratio (PSNR) [2]: This metric evaluates the quality of super-resolved images across the entire image and all color channels, with higher PSNR values indicating better image quality. Natural Image Quality Evaluator (NIQE): NIQE measures the perceptual quality of images without a reference, aiming for lower scores to reflect a closer resemblance to natural images.

6 Experiment

Our experiments targeted facial image super-resolution, utilizing a Modified SRResNet (without Batch Normalization) within the RealESRNetModel framework with a 4x upscaling factor. Using a single GPU, the setup emphasized consistency with a fixed manual seed. The model underwent training across 300,000 iterations with an Adam optimizer and a cosine annealing scheduler for learning rate adjustments. The goal was to minimize L1 Loss for pixel accuracy, using a weight of 1.0 to achieve high-fidelity reconstructions.

Validation with 400 paired images allowed for performance tuning and employed PSNR and NIQE metrics without border influence. Progress was diligently logged every 100 iterations, with checkpoints every 5,000 iterations. Real-time tracking was supported by TensorBoard and Weights & Biases for thorough training analysis.

I have used the **SCSEGPU_M** provided by the School of Computer Science and Engineering.

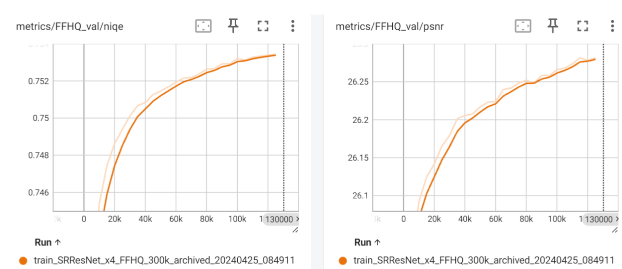
The Table 1, shows hyperparameters which play crucial roles in the behavior and performance of the training process for our task. They are typically chosen through empirical testing and adjusted for optimal performance on the validation set.

Table 1: Hyperparameters Table.

Hyperp.	Values	Hyperp.	Values
Learn. Rate	2e-4	Total Iter	300000
Scale	4	Sec. Gray Noise Pr.	0.4
Num_GPUs	**1 SCSE_M**	Sec. JPEG Range	[70, 95]
Num_Blocks	20	Beta Values	[0.9, 0.99]
num_Parameters	**1,517,571**	Queue Size	176
Resize Prob	[0.2, 0.7, 0.1]	Blur Kernel Size	21
Resize Range	[0.2, 1.5]	Kernel Prob	[0.45, 0.25, ...]
Gauss N. Prob.	0.5	Sinc Prob	0.1
Noise Range	[1, 20]	Blur Sigma	[0.2, 3]
Poi. Scale Range	[0.05, 2]	BetaP Range	[1, 2]
Gray Noise Prob	0.4	BetaG Range	[0.5, 4]
JPEG Range	[50, 95]	2nd Res Range	[0.3, 1.2]
2nd G. N. Prob	0.5	Upscale Factor	4
2nd N. Range	[1, 15]	Optimizer	Adam

7 Results and Conclusion

(a) Pixel loss

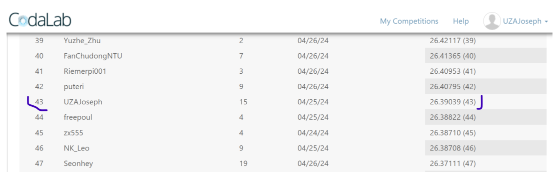


(b) NIQE (left) and PSNR (right)

Figure 1: (a) presents the learning curve, depicting a downward trend in pixel loss, indicating the model's improvement in accuracy. The network's performance stabilizes around 80,000 iterations, hinting at convergence

(b) The descent in NIQE and ascent in PSNR throughout the training suggest that the network is progressively generating sharper and more accurate high-resolution images, essential for assessing the model's super-resolution capabilities

Figure 2: presents a comparative analysis of super resolution techniques applied to low-quality images. The top row displays the original low-resolution images used as input, each characterized by pixelation and blurriness that obscure details and features. The bottom row showcases the output of our super-resolution process, which has noticeably enhanced the clarity, sharpness, and definition of each subject. The transformation is evident in the restoration of facial features, textures, and overall image fidelity. These results demonstrate the efficacy of our super-resolution algorithm in extracting and reconstructing high-resolution details from degraded inputs, offering promising applications in fields such as digital forensics, archival restoration, and image enhancement technologies.



39	Yuzhe_Zhu	2	04/26/24	26.42117 (39)
40	FanChudongNTU	7	04/26/24	26.41365 (40)
41	Riemerpi001	3	04/26/24	26.40953 (41)
42	puteri	9	04/26/24	26.40795 (42)
43	UZAJoseph	15	04/25/24	26.39039 (43)
44	freepoul	4	04/25/24	26.38822 (44)
45	zx555	4	04/24/24	26.38710 (45)
46	NK_Leo	9	04/25/24	26.38708 (46)
47	Seonhey	19	04/26/24	26.37111 (47)

Figure 3: Shows that i have got **26.39** accuracy , under the username of **UZAJoseph** at **43th** position

References

[1] Kaiming He Xiaoou Tang Chao Dong, Chen Change Loy. Image super-resolution using deep convolutional networks, 2015. Accessed: 2024-04-10.

[2] Francisco C. de Oliveira Paulo S. Rodrigues Fernando A. Fardo, Victor H. Conforto. A formal evaluation of psnr as quality measurement parameter for image segmentation algorithms, 2016. Accessed: 2024-04-13.

[3] NVlabs. Ffhq-dataset / flickr-faces-hq dataset (ffhq). Accessed: 2024-04-14.

[4] Chao Dong Ying Shan Xintao Wang, Liangbin Xie. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Aug 2021. Accessed: 2024-04-11.

[5] Steven C.H. Hoi Zhihao Wang, Jian Chen. Deep learning for image super-resolution: A survey, 2020. Accessed: 2024-04-10.

A Pixel loss

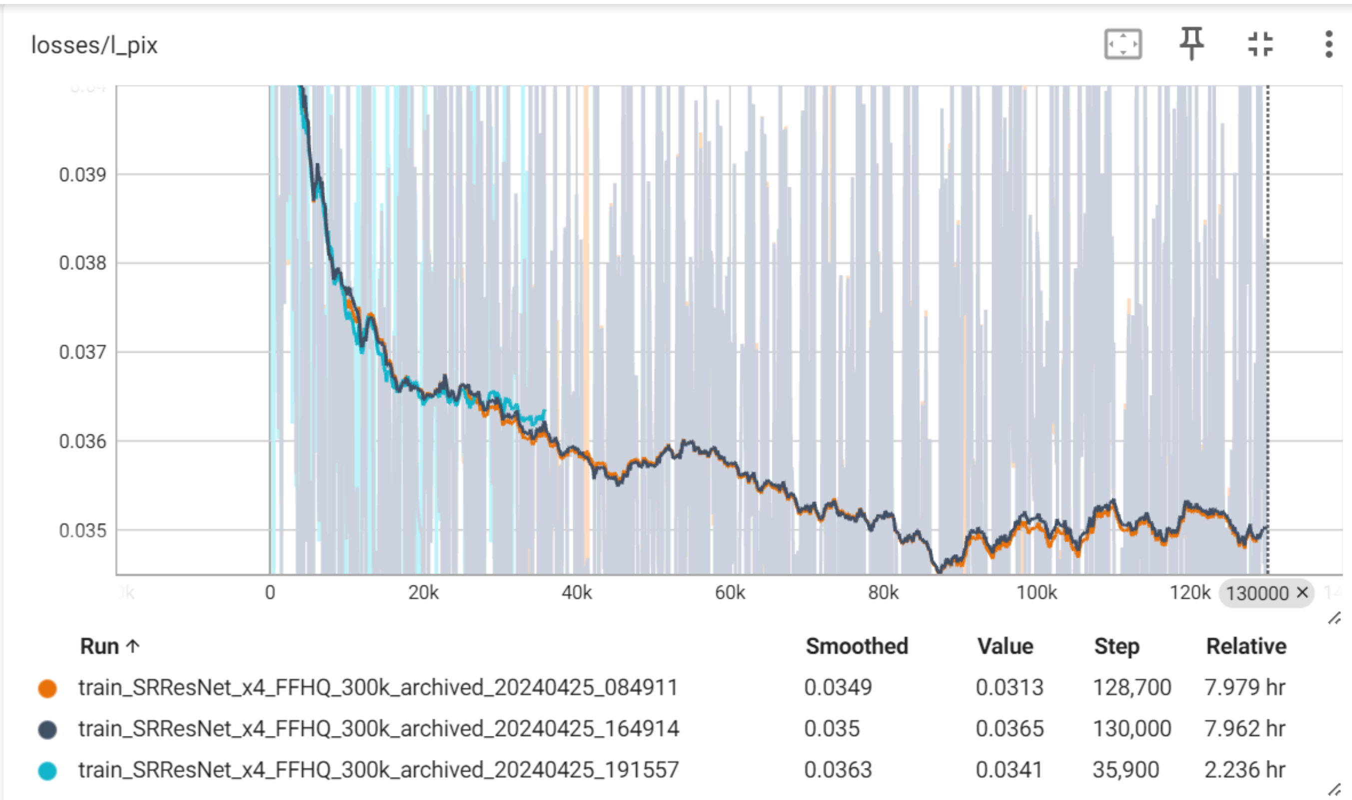


Figure 4: Shows pixel loss on all possible checkpoints during the training

B Metrics

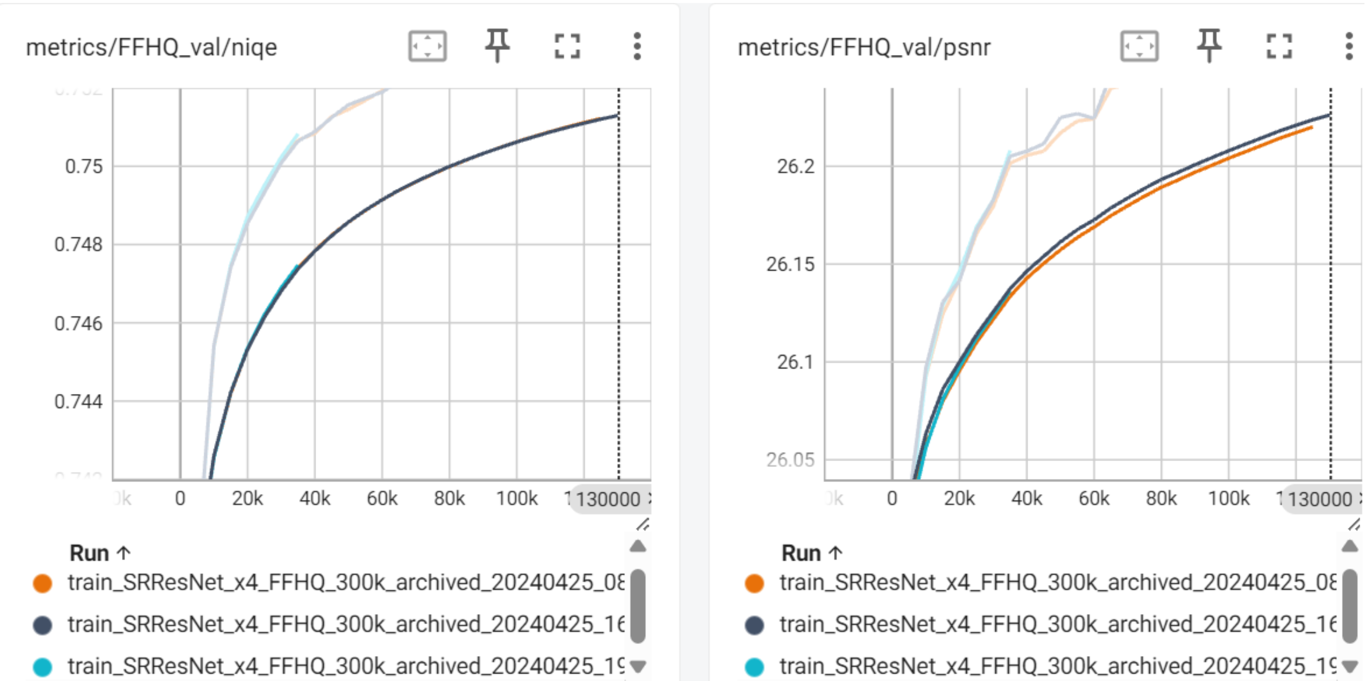


Figure 5: Shows both NIQE (left) and PSNR (right) on all possible checkpoints during the training