[PROJECT 2]

[AI6126 : Advanced Computer Vision]

Riemer van der Vliet
G2304212K

[26 April 2023]

1 introduction

The objective of this mini-challenge is to generate high-quality (HQ) face images from corrupted low-quality (LQ) ones. The data for this task is sourced from FFHQ. For this challenge, we have provided a mini dataset comprising of 5000 HQ images for training and 400 LQ-HQ image pairs for validation. Please note that the LQ images are not included in the training set.

As commonly done within SR training, the data is purely synthetic, and constructed ”on the fly” by perfomring Gaussian blur, Downsampling, Noise, and Compression on the HQ images to make the LQ images. The LQ images are then used to train the model to generate the HQ images. In figure 1 some of the kernels used to transform the HQ images to LQ images are shown. Note that these are different for each of the samples.

1.1 Synthetic Data Generation

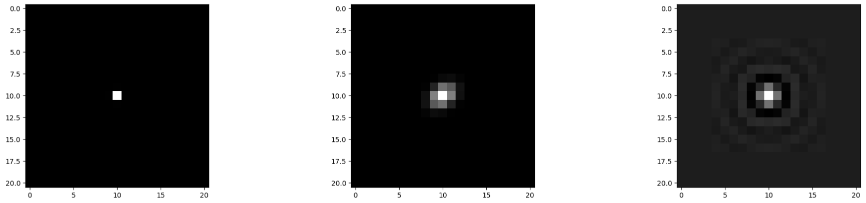

Figure 1: Example kernels in LQ synthesis. Note that each of these kernels is uniquely generated for each low-quality image.

This synthetic data generation process follows the methodology described in ”Training Real-World Blind Super-Resolution with Pure Synthetic Data” by Wang et al. The process involves the use of various kernels, as illustrated in Figure 1, to degrade HQ images into LQ images. Each kernel is uniquely generated for each image. The steps in the pipeline include:

1. **Initialization**
   - Load configuration options and paths to ground-truth (GT) images.
   - Initialize the file client based on the configuration.
2. **Get Item**
   - Retrieve and read the GT image for a given index.
   - Augment the GT image (horizontal flip, rotation).
3. **Generate Kernels**
   - Select kernel size and type (sinc or mixed) based on probability.
   - Generate and pad the kernel accordingly.
4. **Apply Final Sinc Filter**
   - Optionally apply a final sinc filter.
5. **Prepare Output**

- Convert and format the image and kernels into tensors.
- Return tensors and the image path.

1.2 Evaluation

The final test set consists of 400 LQ images to evaluate model performance. The evaluation metric is the Peak Signal-to-Noise Ratio (PSNR), a common measure in image processing to assess image quality. Higher PSNR values indicate better image quality, as shown in the following equation:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{1}$$

2 Other Models

This section presents an overview of other influential models and techniques in the field of super-resolution, particularly focusing on the innovations brought by BasicSR and Real-ESRGAN. These techniques offer alternative methods to achieve high-quality image super-resolution.

2.1 Techniques from BasicSR

BasicSR introduces several key techniques that enhance the quality of super-resolution outputs:

- **SRResNet:** This baseline architecture utilizes residual blocks to achieve effective super-resolution without extensive computational costs.
- **Perceptual Loss:** Unlike traditional loss functions like MSE, BasicSR employs perceptual loss that uses pretrained neural networks (e.g., VGG) to assess perceptual similarity between images, which often leads to more visually pleasing results.
- **Feature Fusion:** This technique allows the model to integrate and leverage information from multiple scales or network branches, enhancing the detail and quality of the upsampled images.

2.2 Techniques from Real-ESRGAN

Real-ESRGAN advances super-resolution through several sophisticated architectural improvements:

- **Enhanced Residual Dense Network (ERDN):** This architecture combines the strengths of residual and dense networks to improve image detail and texture in super-resolution tasks.
- **Adaptive Instance Normalization (AdaIN):** Used within the network, AdaIN adjusts the style of features dynamically, contributing to the flexibility and effectiveness of the super-resolution process.
- **Attention Mechanisms:** By incorporating attention mechanisms, Real-ESRGAN can focus more on significant regions of the image, thus prioritizing areas that most impact perceptual quality.

It is crucial to recognize that in super-resolution, the highest scores are often achieved not by merely producing visually appealing images but by optimizing for metrics such as the Peak Signal-to-Noise Ratio (PSNR). This metric critically influences model performance evaluation in this domain.

3 Results

In this mini-challenge, our objective was to generate high-quality (HQ) face images from corrupted low-quality (LQ) ones using the Real-ESRGAN model. Below we present some example outputs along with the loss curves observed during training.

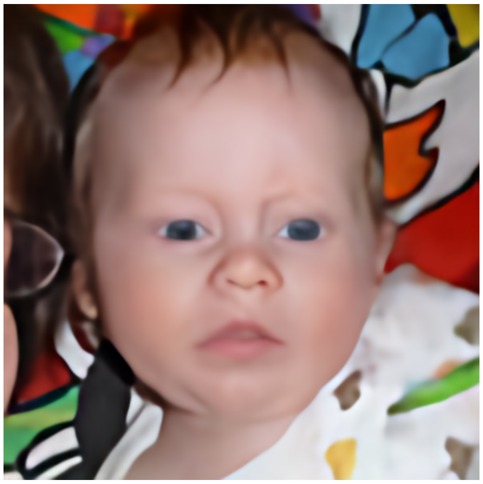

Figure 2: Example of HQ image generated from LQ input.

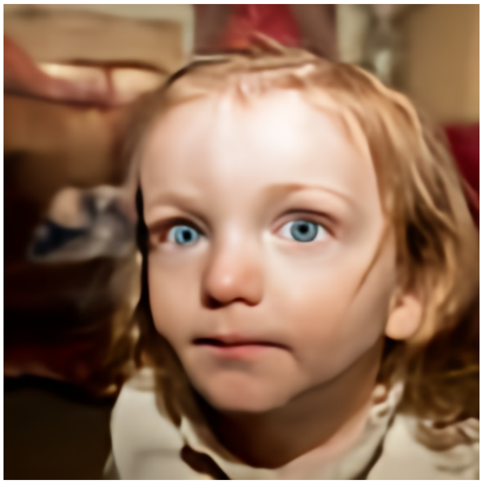

Figure 3: Another example of enhanced HQ output.

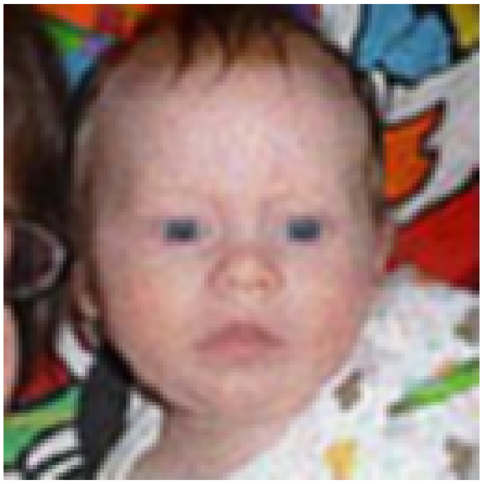

Figure 4: Further illustration of HQ output from Real-ESRGAN.

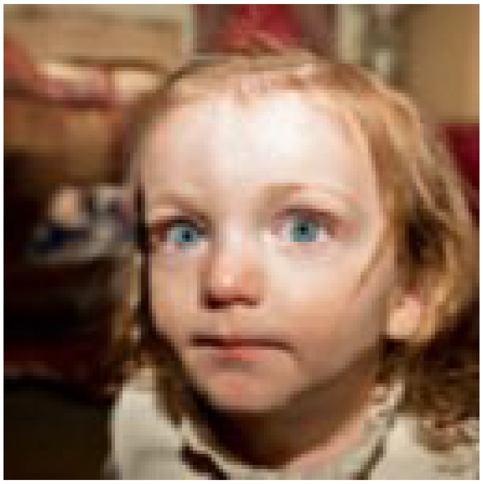

Figure 5: Additional HQ image showcasing model capabilities.

These results were achieved after training the model on the NTU GPU cluster using an A40 GPU for 115,000 iterations—the maximum allowed training duration. The model reached a Peak Signal-to-Noise Ratio (PSNR) score of 26.40953 in the blind test, demonstrating the effectiveness of the Real-ESRGAN model in enhancing image quality from LQ inputs. Detailed settings used during the training are provided in the appendix.

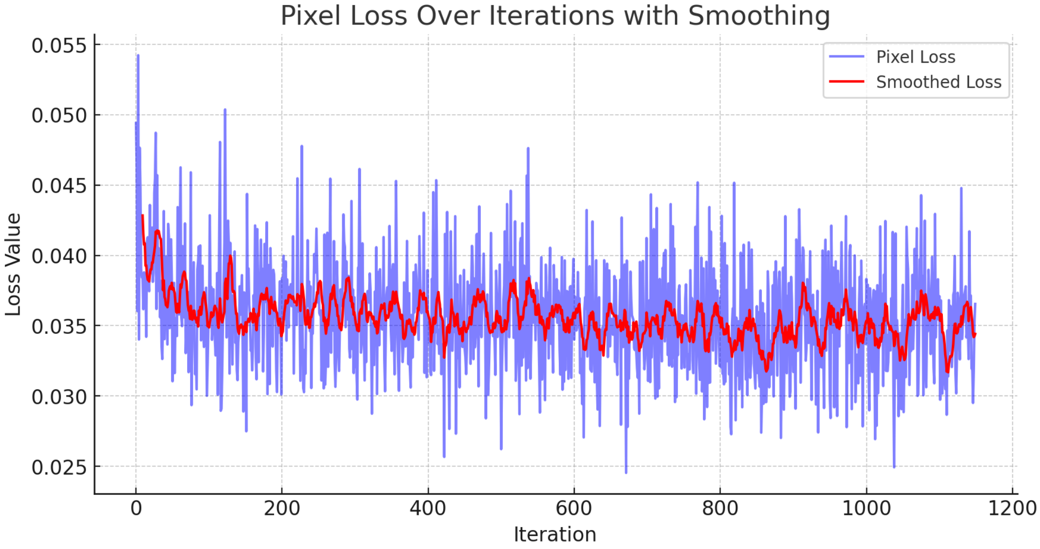

Figure 6: Training loss curves over 115,000 iterations.

4 Discussion

This project brings to light the delicate balance required in super-resolution between optimizing for human visual preferences and computational metrics. While human observers prioritize perceptual and semantic details, computational models tend to focus on precise pixel-based evaluations.

The use of Generative Adversarial Networks (GANs) equipped with perceptual loss functions is a promising approach to reconcile these perspectives. Perceptual loss allows for the generation of images that are more visually pleasing to humans by mimicking the way human vision processes images. However, this often results in lower Peak Signal-to-Noise Ratio (PSNR) scores, as the focus shifts from exact pixel accuracy to more qualitative aspects of image quality.

Despite the potential benefits of perceptual loss, it was not utilized in this experiment due to constraints on using external data. Understanding that the highest PSNR scores do not always correlate with the most visually appealing images is crucial.

5 References

6 Appendix