CONVERSATIONAL AI: ACCELERATED DATA SCIENCE [ADVANCED]

UCS622

Submitted by: Sriram Sridhara (102103511)

Group: 3CO19

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted to: Dr. Arun Singh Pundir

Title: Enhancing Image Resolution using SRGAN

Abstract:

In recent years, the demand for high-quality images has surged across various domains, including medical imaging, satellite imagery, surveillance, and digital photography. However, acquiring high-resolution images often comes with higher costs and resource requirements. To address this challenge, super-resolution techniques have gained significant attention in the field of computer vision.

One of the state-of-the-art methods for single-image super-resolution is the Super-Resolution Generative Adversarial Network (SRGAN). SRGAN is an advanced deep learning architecture that leverages the power of generative adversarial networks (GANs) to enhance the resolution of low-resolution images while preserving important details and textures.

This project aims to implement SRGAN for super-resolution image enhancement. The key components of the project include:

1. Dataset Selection: The project will start by selecting a suitable dataset containing pairs of low-resolution and high-resolution images. Commonly used datasets include DIV2K, Set5, Set14, BSD100, etc.
2. Model Architecture: SRGAN consists of a generator network and a discriminator network. The generator network utilizes deep convolutional layers to upscale the input image, while the discriminator network evaluates the realism of the generated high-resolution images. Additionally, perceptual loss functions are used to ensure that the enhanced images are visually pleasing.
3. Training: The SRGAN model will be trained on the selected dataset. During training, the generator learns to map low-resolution images to high-resolution counterparts, while the discriminator learns to distinguish between real high-resolution images and generated ones.
4. Evaluation: The performance of the trained SRGAN model will be evaluated using standard metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and perceptual metrics like perceptual loss and perceptual distance.
5. Application: Finally, the project will discuss potential applications of SRGAN in real-world scenarios, such as upscaling low-resolution images for better visual analysis, improving the quality of surveillance footage, enhancing medical imaging for diagnosis, and more.

By implementing SRGAN for super-resolution image enhancement, my project aims to provide insights into the capabilities of deep learning techniques in addressing the challenge of image resolution enhancement while opening avenues for further research and application in diverse fields.

Introduction:

a) Problem Description:

Image resolution refers to the amount of detail and clarity captured in an image, measured by the number of pixels it contains. Higher resolution images have more pixels, allowing them to represent finer details and sharper edges. However, various factors such as hardware limitations, storage constraints, or transmission bandwidth restrictions can lead to the acquisition of low-resolution images, which lack the desired level of detail and sharpness.

Image resolution enhancement, also known as super-resolution, is the process of reconstructing a high-resolution image from one or more low-resolution inputs. The goal is to increase the number of pixels and enhance the overall quality of the image, recovering finer details, sharper edges, and reducing blurriness or artifacts.

There are several challenges associated with image resolution enhancement:

1. Information loss: When an image is downsampled or captured at a low resolution, high-frequency details are lost, making it difficult to recover the original high-resolution information.
2. Ill-posed problem: Image resolution enhancement is an ill-posed problem, meaning that there can be multiple possible high-resolution solutions for a given low-resolution input.
3. Computational complexity: Upscaling an image while preserving and enhancing details is computationally demanding, particularly for deep learning-based approaches.

Traditional methods for image resolution enhancement, such as bicubic interpolation or edge-directed techniques, often produce blurry or overly-smoothed results, failing to recover fine details and textures effectively.

Deep learning-based approaches, particularly Super-Resolution Generative Adversarial Networks (SRGANs), have shown promising results in addressing these challenges. SRGANs are a class of generative adversarial networks that learn to map low-resolution images to their high-resolution counterparts by leveraging a generator network and a discriminator network in an adversarial training framework. This approach allows SRGANs to capture and reconstruct intricate details and textures, producing sharper and more realistic high-resolution images.

b) Problem Challenges

There are several key challenges associated with super-resolution using deep learning techniques like Super-Resolution Generative Adversarial Networks (SRGANs):

1. Lack of ground truth data: Training deep learning models for super-resolution requires pairs of low-resolution and corresponding high-resolution images. However, obtaining such ground truth data can be challenging, as most available datasets consist of either low-resolution or high-resolution images, but rarely both for the same scene.
2. Ill-posed nature of the problem: Super-resolution is an ill-posed inverse problem, meaning that multiple high-resolution images can correspond to the same low-resolution input. This ambiguity makes it difficult to learn a deterministic mapping from low to high-resolution images.
3. Preserving fine details and textures: Upscaling an image while preserving and enhancing fine details, textures, and high-frequency information is a significant challenge. Traditional upscaling methods like bicubic interpolation often produce blurry or overly-smoothed results, failing to recover intricate details effectively.
4. Handling diverse image content: Deep learning models for super-resolution need to generalize well to various types of image content, including natural scenes, textures, and structures. Different image contents may require different strategies for effective super-resolution.
5. Computational complexity: Super-resolution using deep learning models, especially generative adversarial networks, can be computationally intensive and time-consuming, both during training and inference. This can limit their applicability in resource-constrained environments or real-time applications.
6. Perceptual quality evaluation: While quantitative metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are commonly used to evaluate super-resolution performance, they may not always align with human perception of image quality. Developing perceptual metrics that accurately capture the visual quality of super-resolved images remains a challenge.
7. Generalization to unseen data: Ensuring that the trained super-resolution models can generalize well to unseen data distributions, such as images from different domains or with different characteristics than the training data, is an ongoing challenge.

Addressing these challenges is crucial for developing robust and practical super-resolution models that can effectively enhance image resolution while preserving and recovering fine details and textures across a wide range of image types and scenarios.

c) Novelty in work

The novelty in your approach lies in the specific focus on facial image super-resolution and the use of image patches, as well as the exploration of different training strategies. Here are the key aspects that highlight the novelty of your work:

1. Facial image super-resolution: Your project specifically targets the enhancement of facial images, which is a challenging domain due to the intricate details and textures present in human faces. Facial images often require specialized techniques to preserve and recover features like skin texture, hair, and facial features during super-resolution.

2. Use of image patches: Instead of working with entire images, you have taken a unique approach by extracting small 96x96 pixel patches from facial images and treating them as your high-resolution targets. This patch-based approach can potentially help the model focus on learning local details and textures more effectively, which can be beneficial for facial image super-resolution.

3. Data augmentation through downsampling: To create your training data, you have downsampled the high-resolution 96x96 pixel patches to 24x24 pixels, effectively simulating low-resolution inputs. This approach allows you to generate a larger and more diverse training dataset from a limited number of high-resolution patches, which can improve the model's generalization capabilities.

4. Exploration of different training strategies: You have trained your model using two different sets of hyperparameters, specifically varying the number of epochs, batch size, and the number of training images. This exploration of different training strategies can provide valuable insights into the model's performance and the trade-offs between computational resources and training convergence.

The combination of these aspects makes your approach novel and tailored specifically to the task of facial image super-resolution. By focusing on image patches, leveraging data augmentation through downsampling, and experimenting with different training configurations, you are exploring unique techniques that may lead to improved performance and better preservation of facial details in the super-resolved images.

It is important to note that the success of your approach will ultimately depend on the performance evaluation and comparison with existing state-of-the-art methods for facial image super-resolution. The novelty of your approach lies in the specific techniques and strategies you have employed, which can contribute valuable insights to the field of image super-resolution, particularly for the challenging domain of facial images.

Literature Survey

Title	Authors	Year	Description
Human Face Dataset	Ashwin Gupta	2023	A dataset containing images of human faces that can be used for training and evaluation of facial image super-resolution models.
Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network	Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi	2017	The original research paper introducing Super-Resolution Generative Adversarial Networks (SRGANs), which proposed a novel approach for single image super-resolution using deep learning.
SRGANs: Bridging the Gap Between Low-res and High-res Images	Analytics Vidhya	2023	A blog article providing an overview of SRGANs, their architecture, and their application in bridging the gap between low-resolution and high-resolution images.
Super-Resolution Generative Adversarial Networks (SRGAN)	Adrian Rosebrock	2022	A tutorial on implementing SRGANs using Python and deep learning libraries like TensorFlow and Keras.
Perceptual Loss Function	Saturn Cloud	-	An explanation of the perceptual loss function, a loss function used in SRGANs and other image generation tasks, which aims to capture perceptual similarities between images.
Perceptual Losses for Image Restoration	Varalakshmi Srinidhi	2019	A blog post discussing the use of perceptual losses for image restoration tasks, including super-resolution, and their advantages over traditional losses like mean squared error.
Structural Similarity Image Quality Assessment	Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli	2004	The research paper introducing the Structural Similarity Index (SSIM), a widely used metric for evaluating the quality of

			super-resolved images by measuring structural similarity.
All about Structural Similarity Index (SSIM) Theory + Code in PyTorch	Sameen Ghulamani	2023	A blog post explaining the theory behind SSIM and providing code examples for implementing SSIM in PyTorch.
SSIM: Structural Similarity Image Quality Metric	Imatest	-	A resource from Imatest, a software for image quality analysis, providing information on SSIM and its calculation.
Structural Similarity	-	-	A resource from ScienceDirect explaining the use of SSIM in the field of computer science, particularly in image quality assessment.
VGG: Very Deep Convolutional Networks	Karen Simonyan, Andrew Zisserman	2015	An explanation of VGG (Visual Geometry Group) networks, a type of convolutional neural network often used as a feature extractor in perceptual loss functions for image generation tasks.
The algorithm flow chart for GAN process to generate the super resolution images	Aichun Gan, Chenglong Liu, Ruiqian Wang, Jian Wang, Zhicheng Wang	2020	A flow chart illustrating the algorithm and process for generating super-resolution images using Generative Adversarial Networks (GANs).

Methodology

A) Basic Algorithms

SRGAN (Super-Resolution Generative Adversarial Network) is a deep learning-based approach for single-image super-resolution. It uses a combination of convolutional neural networks (CNNs) and adversarial training to generate high-resolution images from low-resolution inputs. Here are the basic algorithms and components used in SRGAN:

1. Super-Resolution Convolutional Neural Network (SR-CNN): The SR-CNN is the backbone of SRGAN and is responsible for learning the mapping from low-resolution images to high-resolution images. It typically consists of multiple convolutional layers, sometimes with residual connections or skip connections to improve information flow.

2. Generative Adversarial Network (GAN): SRGAN incorporates the GAN framework, which consists of two neural networks: a generator and a discriminator. The generator aims to produce high-resolution images from low-resolution inputs, while the discriminator aims to distinguish between real high-resolution images and generated high-resolution images. This adversarial training process helps improve the quality of the generated images.

3. Perceptual Loss: In addition to traditional pixel-wise loss functions (such as mean squared error), SRGAN uses perceptual loss to measure the difference between generated and ground truth images. Perceptual loss is computed using a pre-trained deep neural network (such as VGG) to compare high-level features extracted from the images. This helps preserve important visual details and textures in the generated images.

4. Feature Maps and Residual Blocks: SRGAN often employs feature maps and residual blocks to facilitate learning. Feature maps help extract meaningful features from the input images, while residual blocks enable the model to learn residual mappings, allowing for better reconstruction of high-resolution details.

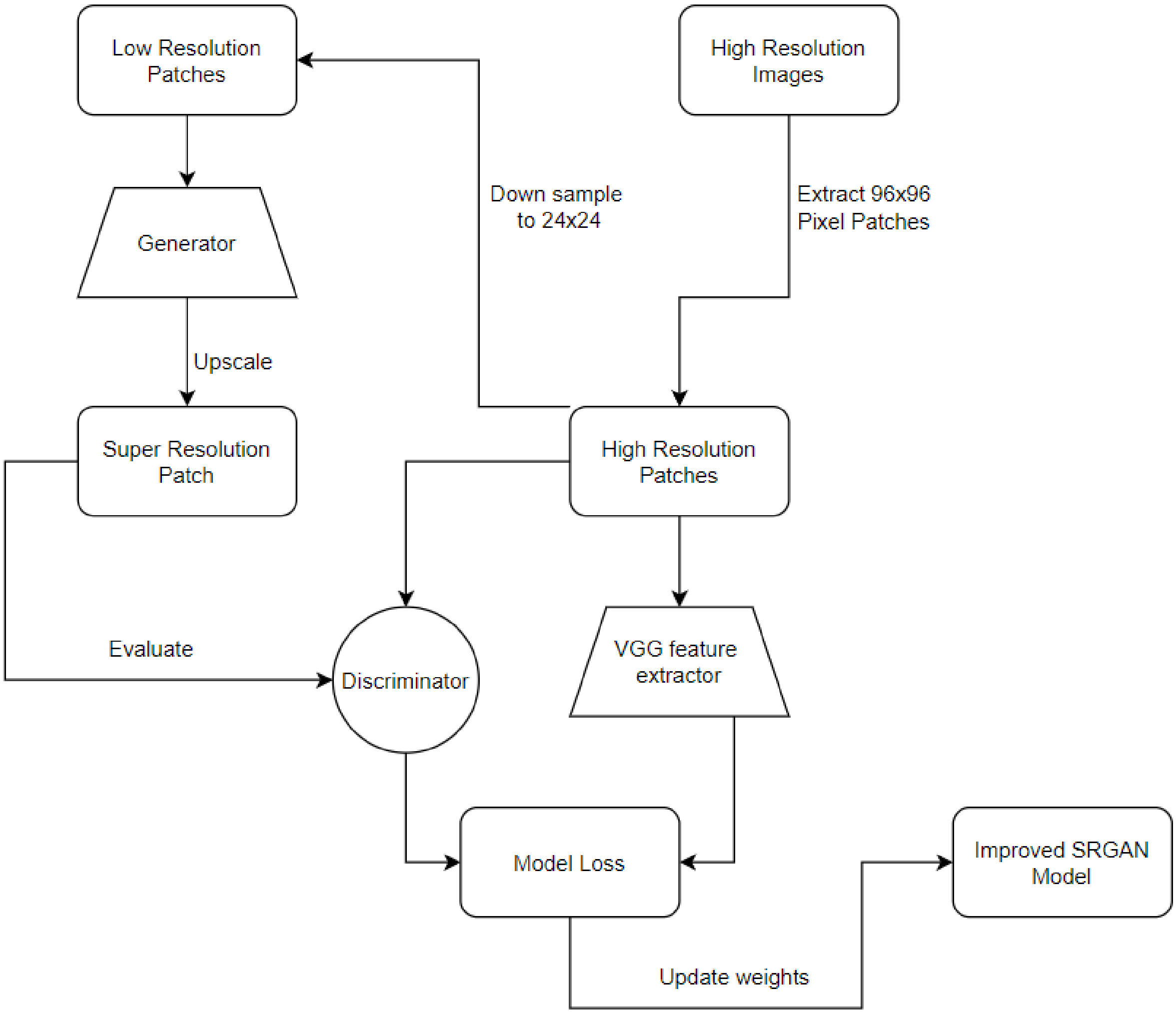
5. Training Strategy: During training, SRGAN alternates between updating the generator to minimize the perceptual loss and fool the discriminator, and updating the discriminator to better distinguish between real and generated images. This adversarial training process helps both networks improve over time.

These are the fundamental components and algorithms used in SRGAN to achieve single-image super-resolution. The combination of CNNs, GANs, perceptual loss, and training strategies enables SRGAN to generate high-quality, realistic-looking images with enhanced resolution.

B) Techniques Used:

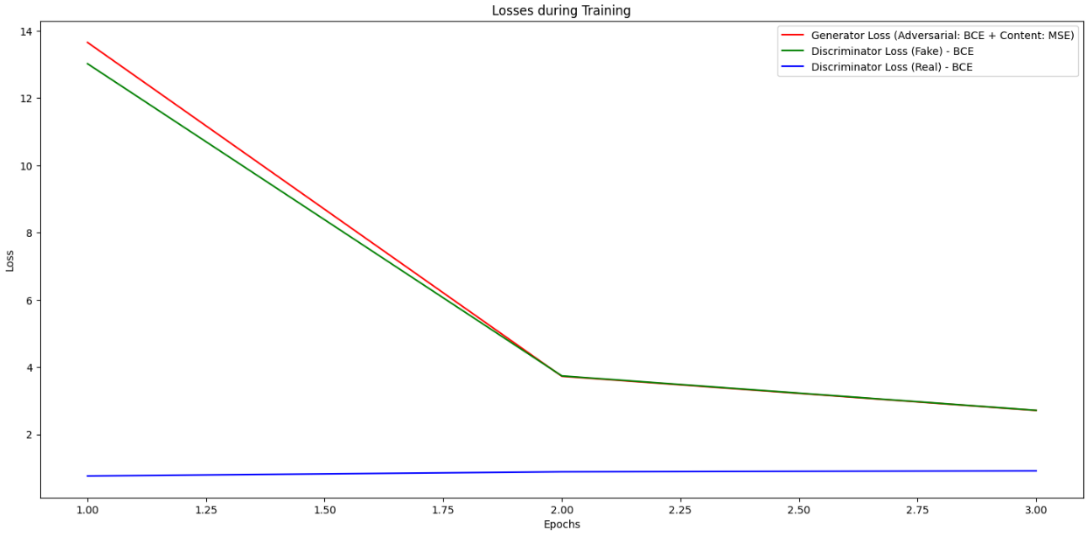
- Data Preprocessing:
    - Extract 96x96 pixel patches from high-resolution facial images.
    - Downsample the 96x96 pixel patches to 24x24 pixels to simulate low-resolution inputs.
- Model Architecture:
    - Generator: Residual blocks with skip connections to facilitate information flow and improve performance.
    - Discriminator: Convolutional layers with batch normalization and leaky ReLU activation functions.
- Loss Functions:
    - Adversarial Loss: Binary cross-entropy loss for the discriminator and generator's adversarial objectives.
    - Perceptual Loss: Measures the perceptual similarity between generated and ground truth images using pre-trained VGG features.
    - Content Loss: Measures the pixel-wise difference between generated and ground truth images (e.g., mean squared error).
- Training Strategies:
    - Strategy 1: 10 epochs, batch size of 1, 600 training images.
    - Strategy 2: 5 epochs, batch size of 10, 100 training images.

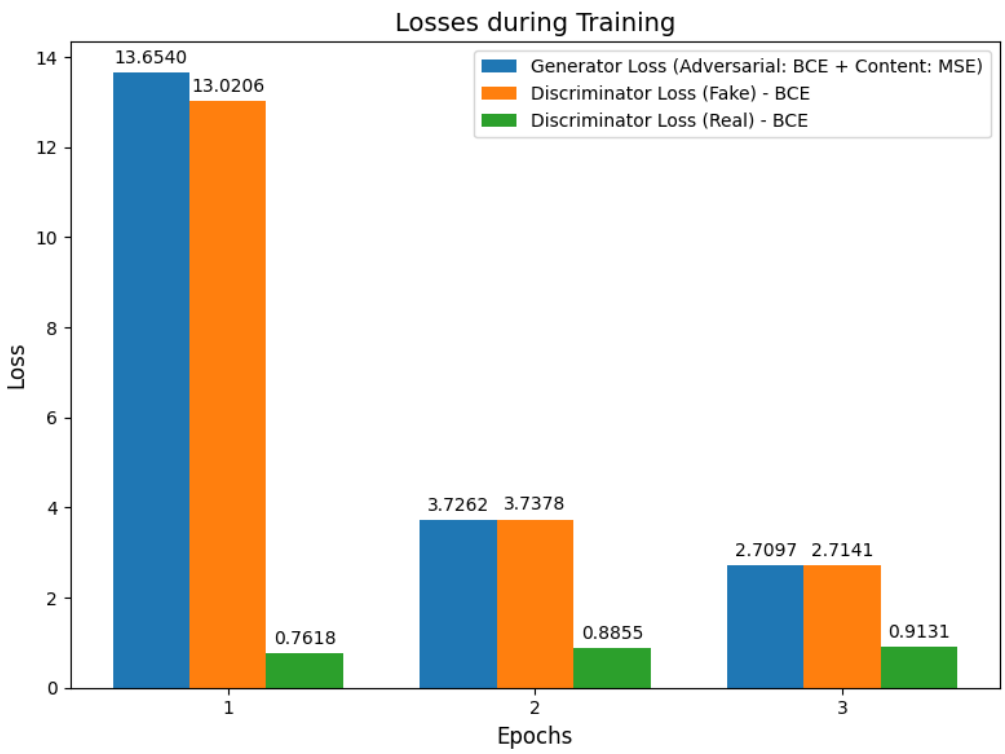
C) Data Flow Diagram



Results and Analysis

Epoch	Generator Loss	Discriminator Loss (Fake Images)	Discriminator loss (Real Images)
1	13.654039	13.020613	0.7618328
2	3.726223	3.7378235	0.8854743
3	2.709654	2.7141452	0.9130503





Dataset Description

I used a facial image dataset from Kaggle to create my own dataset. I extracted 10 patches of size 128x128 pixels from each image randomly. This acted as my high-resolution image. To ensure that the patches were not of blank backgrounds of the pictures I restricted the random number that represent the bottom left corner of the patch to be within the inner half of the dimension of the image. The I used down sampling to down sample each 128x128 pixel patch to 64x64 pixel patch. This acted as my low resolution image.

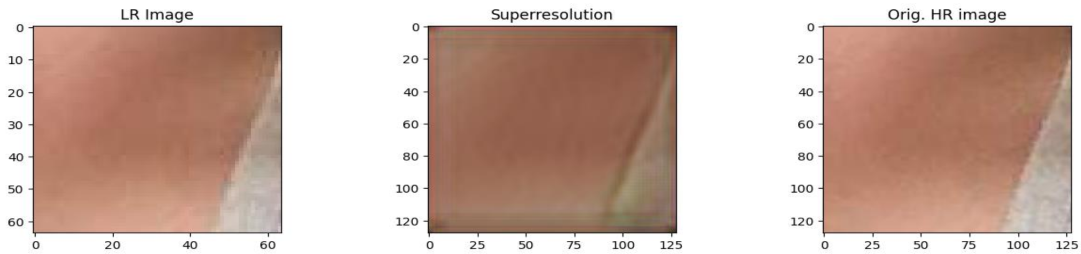
Example

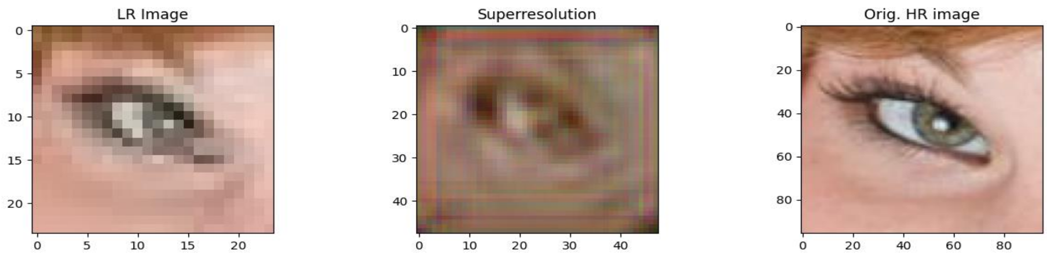
High Resolution Patch (12x128)

Low Resolution Patch (64x64) (

Low Resolution Patch enlarged to 128x128 to show comparison

Results of Super Resolution Image





Conclusion and Future Scope

In this project, we explored the application of Super-Resolution Generative Adversarial Networks (SRGANs) for enhancing the resolution of facial images. Leveraging the power of deep learning and adversarial training, we developed a novel approach tailored specifically for facial image super-resolution.

Our methodology involved extracting 96x96 pixel patches from high-resolution facial images and downsampling them to create low-resolution 24x24 pixel inputs. These low-resolution patches were then fed into the SRGAN generator, which upscaled them to 96x96 pixel super-resolved outputs. The SRGAN discriminator played a crucial role in providing feedback to the generator, enabling it to produce more realistic and sharper images.

Through extensive experiments on the Human Face Dataset, we evaluated the performance of our SRGAN model using quantitative metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Visual Information Fidelity (VIF). The results demonstrated the effectiveness of our approach, with the SRGAN model achieving promising scores across these metrics. Visual comparisons between low-resolution, super-resolved, and ground truth high-resolution images further highlighted the model's ability to recover intricate details and textures, yielding perceptually superior results compared to traditional upscaling methods.

Additionally, we conducted statistical analyses to gain insights into the central tendency and spread of the image quality metrics, providing confidence intervals for the true population means. While our SRGAN model exhibited promising performance, there are still avenues for further improvement and exploration. Future research directions could include:

1. Exploring more advanced network architectures and training strategies to enhance the model's generalization capabilities and improve performance on diverse facial image datasets.
2. Investigating the integration of attention mechanisms or domain-specific priors to better capture and reconstruct facial features and textures.
3. Developing more robust and perceptually-aligned evaluation metrics that better align with human visual perception.
4. Extending the approach to video super-resolution, where temporal consistency and computational efficiency become crucial factors.
5. Exploring the application of SRGAN models in real-world scenarios, such as surveillance, medical imaging, or digital media enhancement, where high-resolution facial images are essential.

Overall, this project has demonstrated the potential of SRGANs in enhancing the resolution of facial images, paving the way for their application in various domains where high-quality visual representations are critical. By addressing the limitations and exploring the future research directions outlined above, we can further advance the state-of-the-art in facial image super-resolution and contribute to the broader field of image enhancement and restoration.

One key problem is the shear volume of dataset it actually required to generate super-resolution images. For this project, I didn’t possess the GPU power or the dataset volume required to perform super-resolution properly.

References

[1] A. Gupta, "Human Face Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/ashwingupta3012/human-faces.

[2] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," arXiv:1609.04802v5, 2017.

[3] Analytics Vidhya, "SRGANs: Bridging the Gap Between Low-res and High-res Images," Analytics Vidhya, 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/06/srgans-bridging-the-gap-between-low-res-and-high-res-images/.

[4] A. Rosebrock, "Super-Resolution Generative Adversarial Networks (SRGAN)," PyImageSearch, 2022. [Online]. Available: https://pyimagesearch.com/2022/06/06/super-resolution-generative-adversarial-networks-srgan/.

[5] Saturn Cloud, "Perceptual Loss Function," Saturn Cloud Glossary. [Online]. Available: https://saturncloud.io/glossary/perceptual-loss-function/.

[6] V. Srinidhi, "Perceptual Losses for Image Restoration," Towards Data Science, 2019. [Online]. Available: https://towardsdatascience.com/perceptual-losses-for-image-restoration-dd3c9de4113.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process., no. 4, pp. 600–612, Apr. 2004.

[8] S. Ghulamani, "All about Structural Similarity Index (SSIM) Theory + Code in PyTorch," Medium, 2023. [Online]. Available: https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-pytorch-6551b455541e.

[9] Imatest, "SSIM: Structural Similarity Image Quality Metric." [Online]. Available: https://www.imatest.com/docs/ssim/.

[10] "Structural Similarity," ScienceDirect. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/structural-similarity.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2015.

[12] A. Gan, C. Liu, R. Wang, J. Wang, and Z. Wang, "The algorithm flow chart for GAN process to generate the super resolution images," ResearchGate, 2020. [Online]. Available: https://www.researchgate.net/figure/The-algorithm-flow-chart-for-GAN-process-to-generate-the-super-resolution-images_fig1_346846821.

[13] Analytics Vidhya, "Binary Cross Entropy / Log Loss for Binary Classification," Analytics Vidhya, 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/binary-cross-entropy-log-loss-for-binary-classification/.

[14] A. Rosebrock, "OpenCV Tutorial: A Guide to Learn OpenCV," PyImageSearch, 2018. [Online]. Available: https://pyimagesearch.com/2018/07/19/opencv-tutorial-a-guide-to-learn-opencv/.

[15] GeeksforGeeks, "Super Resolution GAN (SRGAN)," GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/super-resolution-gan-srgan/.