AI Image Remastering

Guide: Meysam Effati

Group E

- Kuncheria Tom
- Prince Thomas
- Remya Kannan Kandari
- Shanmuga Priyan Jeevanandam
- Sravya Somala

Introduction

Problem Statement

Many images are captured or stored at low resolutions due to technical limitations, storage constraints, or legacy data. This project aims to address the challenge of restoring and enhancing these low-resolution images, transforming them into higher-quality versions with improved details, clarity, and visual appeal.

Significance

High-quality image restoration is crucial in fields like entertainment, medical imaging, and digital art, where preserving visual fidelity is paramount. This project focuses on developing an AI-based remastering technique to revive lost details and make low-resolution images suitable for modern applications.

Data Collection and Preparation

1

Dataset Selection

High-resolution images from Kaggle's Image Super-Resolution dataset were chosen. The dataset included a variety of image types, such as natural scenes, art, anime images.

2

Data Preprocessing

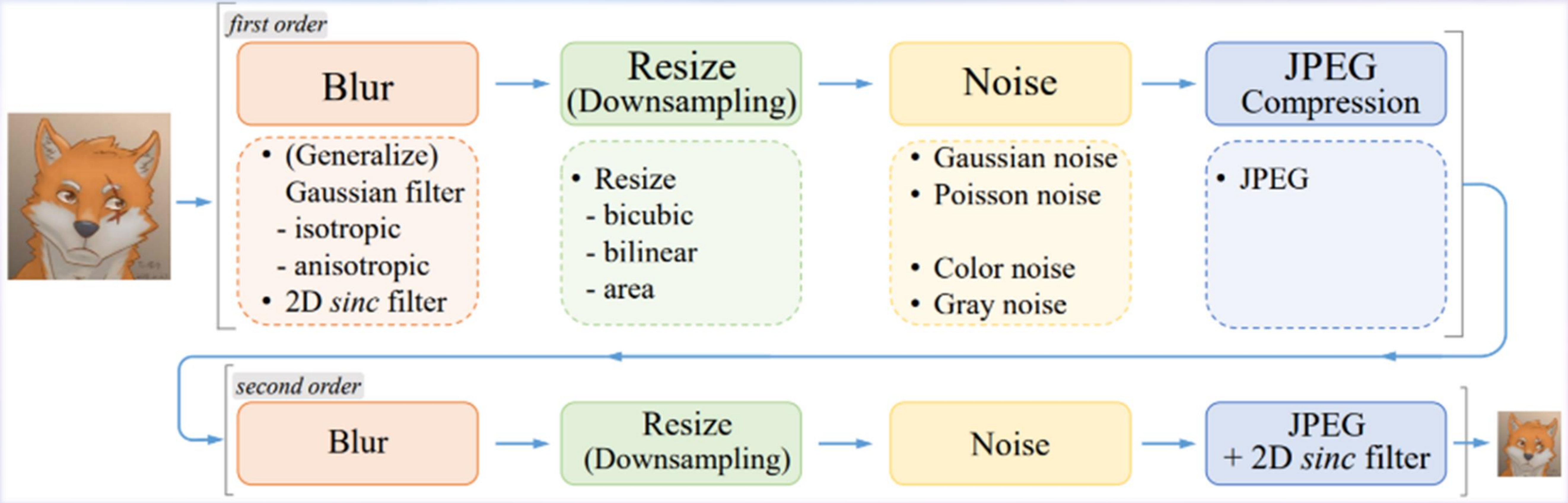
Low-resolution images were created from high-resolution counterparts by applying degradation techniques like downsampling, noise addition, and compression.

3

Challenges Addressed

Generating realistic low-resolution images posed challenges, which were overcome through careful calibration of degradation parameters.

Image Degradation Architecture



Simulate real-world image degradation to create challenging low-resolution images for training.

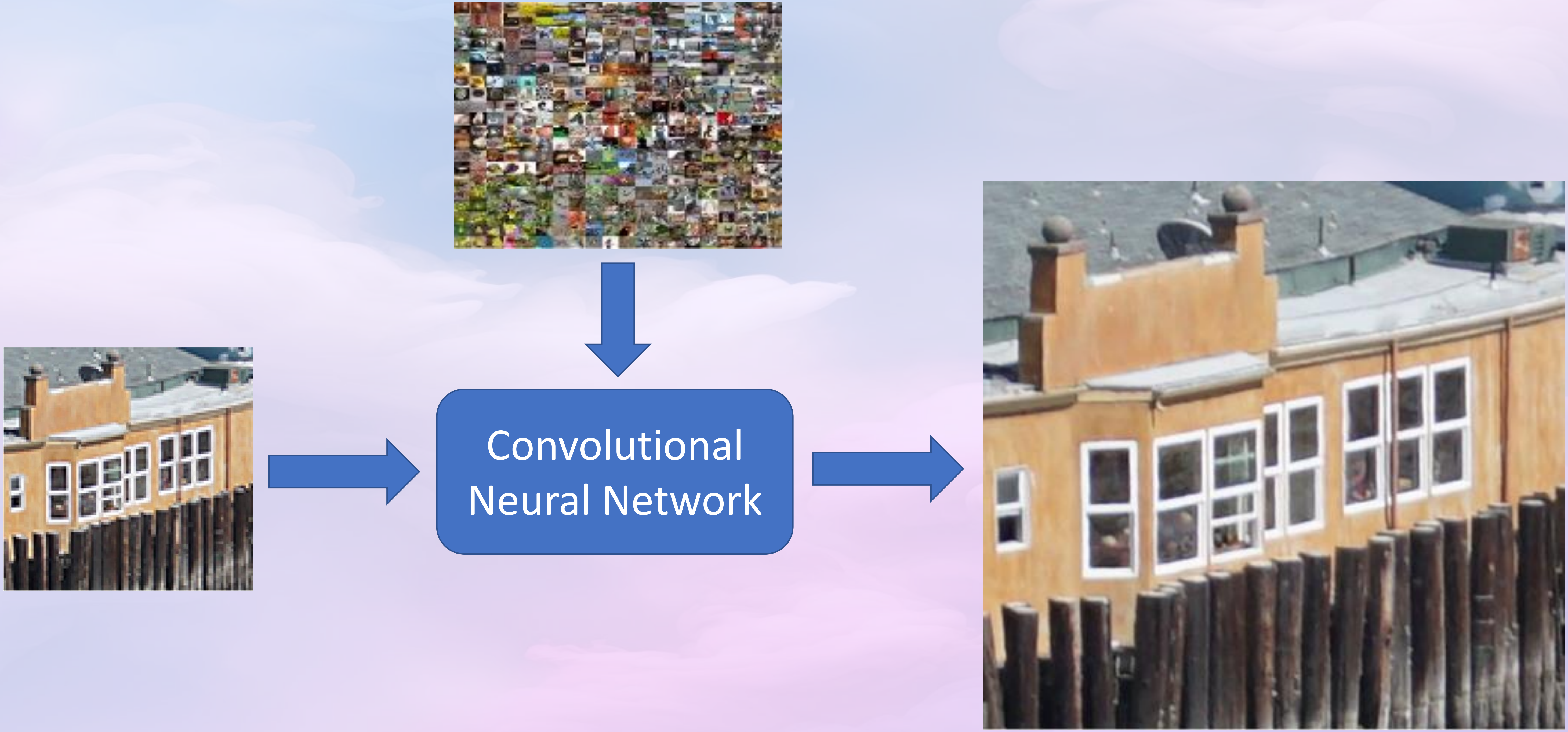
- Downsampling: Reduced image resolution using a scaling factor.
- Noise Addition: Introduced Gaussian noise to mimic common artifacts.
- Compression: Applied JPEG compression to emulate data loss in storage and transmission

Traditional Image Upscaling Techniques

LR Image

Nearest Neighbor Upscaling

Bicubic Interpolation


Convolutional
Neural Network

Methodology

Techniques and Algorithms

Real-ESRGAN: Leveraged as the primary architecture for image super-resolution, utilizing GANs to enhance image details and textures.

Transfer Learning: Implemented by loading pre-trained Real-ESRGAN weights and fine-tuning only specific layers.

Data Augmentation: Employed techniques like cropping, flipping, and rotation to diversify the dataset, enhancing the model's ability to generalize.

Justification

Real-ESRGAN: Chosen for its proven capability to produce high-quality, visually appealing images, excelling in detail preservation.

Transfer Learning: Selected to capitalize on the pre-trained model's existing knowledge, reducing training time and computational requirements while improving model performance on our dataset.

Layer Selection: The early layers, which capture general features like edges and textures, were kept constant. The later layers, responsible for more specific, high-level features, were fine-tuned to adapt to our unique image content.

Assumptions

The assumption was that the early layers of the Real-ESRGAN model already learned general image features effectively, so freezing these layers would prevent overfitting and expedite training. Fine-tuning the later layers allowed the model to adapt to the specific characteristics and degradation patterns of our dataset.

Architecture of Real ESRGAN

1

Generator Network

Utilizes a deep residual network with skip connections for feature extraction and upsampling module to scale low-resolution images to high-resolution outputs. Designed to handle real-world images with complex textures and details.

2

Discriminator Network

The discriminator uses a U-Net architecture integrated with a pre-trained VGG19 model. The U-Net structure enables the model to capture both global and local features effectively, while the VGG19 layers, pre-trained on a vast dataset, provide powerful feature extraction capabilities.

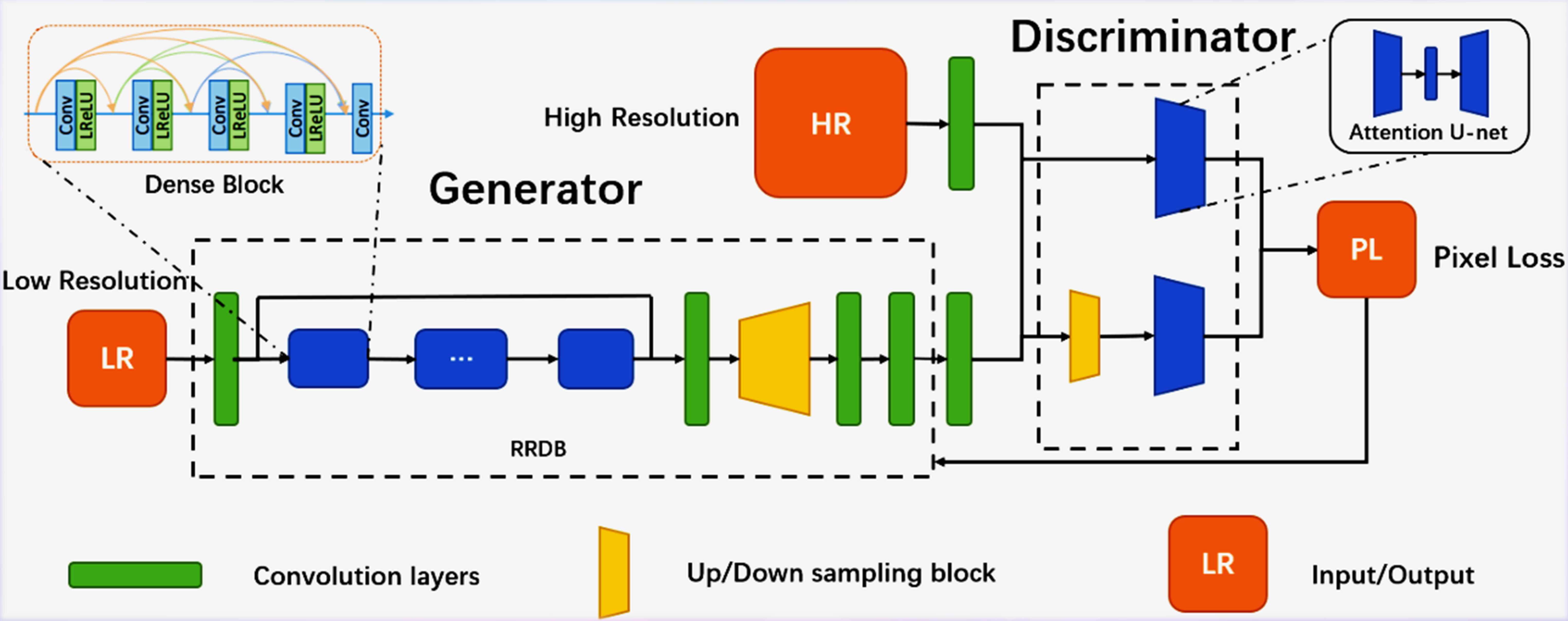
3

Loss Functions

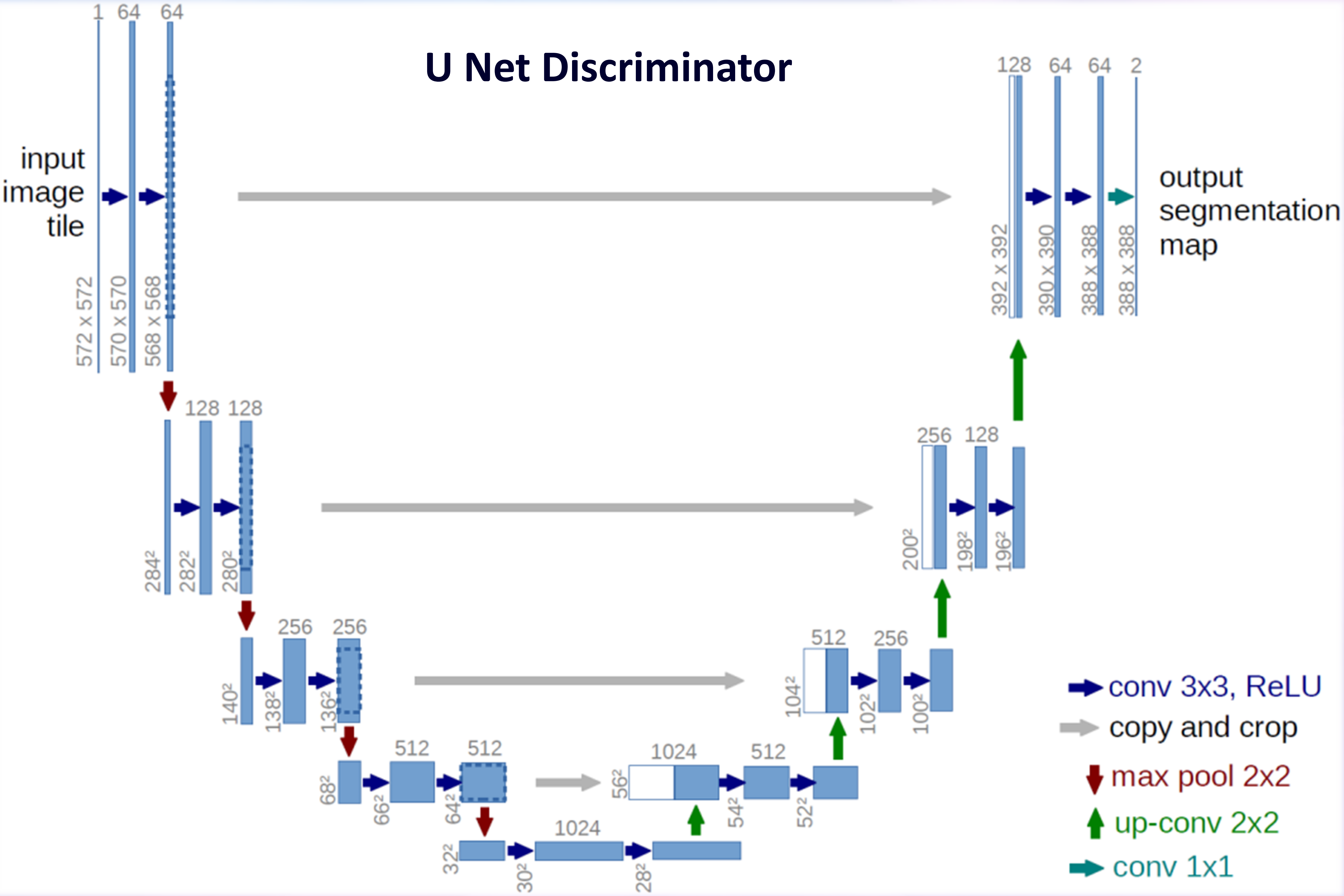
Pixel Loss: Measures the difference between the generated image and the ground truth at the pixel level.

Perceptual Loss: Compares high-level feature representations, ensuring the output preserves perceptual quality and texture.

Adversarial Loss: Encourages the generator to produce more realistic images by fooling the discriminator, leading to outputs that are visually indistinguishable from real images.


Discriminator
Attention U-net
High Resolution
HR
Conv
LReLU
Dense Block
Generator
PL
Pixel Loss
Low Resolution
LR
...
RRDB
Convolution layers
Up/Down sampling block
LR
Input/Output

U Net Discriminator



Model Training and Evaluation

1

Training Process

Model trained on high-resolution and degraded images, focusing on minimizing perceptual loss using VGG19, and optimizing for PSNR and SSIM metrics.

2

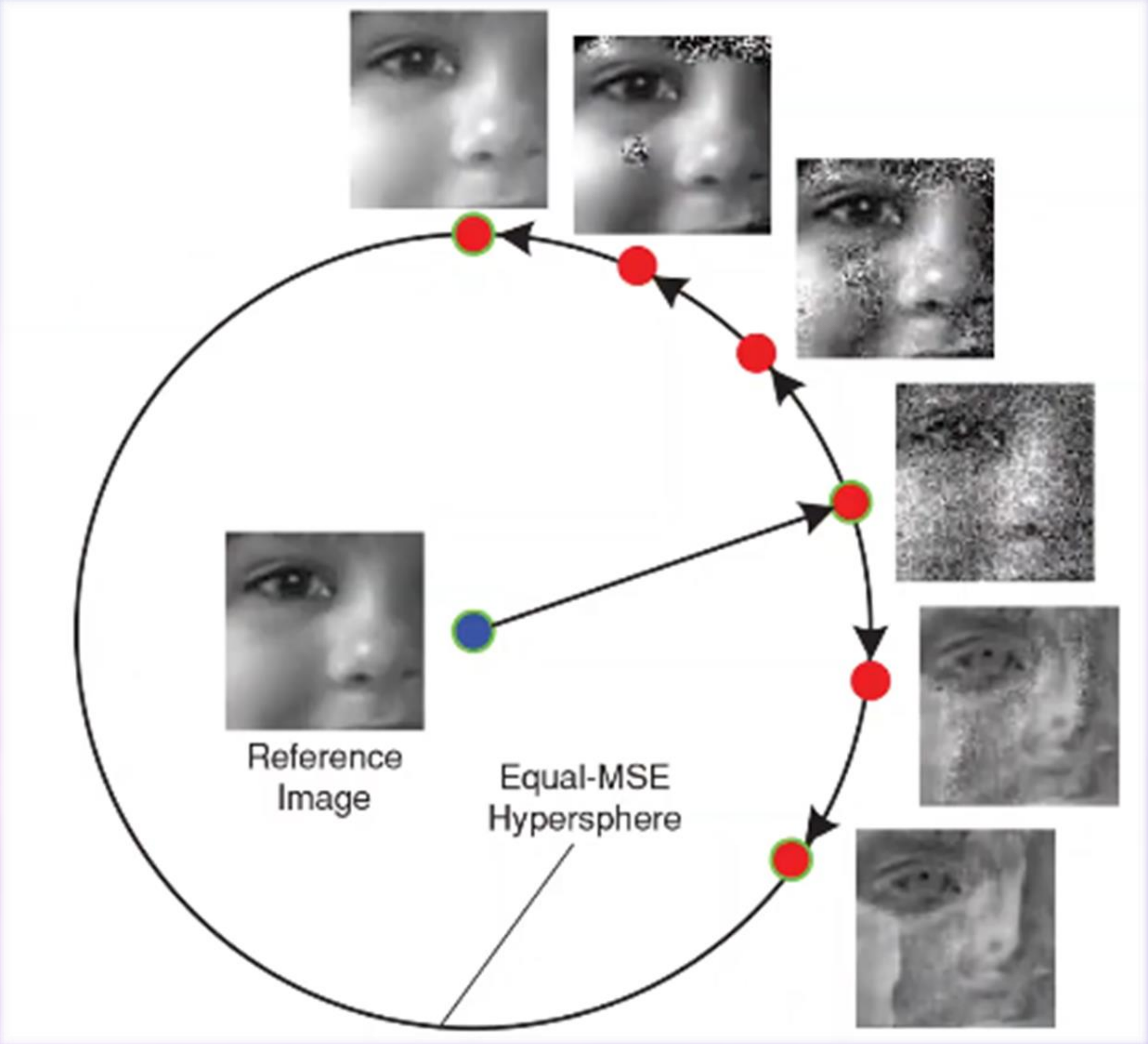
Evaluation Metrics

Model evaluated based on the quality of restored images using PSNR, SSIM, and visual assessments.

3

Challenges Addressed

Overfitting and convergence issues were mitigated through careful hyperparameter tuning and data augmentation techniques.



Analysis and Results

Model	PSNR	SSIM
Real-ESRGAN	32.5	0.92
ESRGAN	30.2	0.88
SRGAN	28.7	0.85

RealESRGAN significantly improved image quality, reducing artifacts and enhancing details compared to other models. PSNR and SSIM scores were higher for RealESRGAN, with visual comparisons showing superior restoration quality.

Restored Image (x2)

Input Image

800x534, 59KB

1600x1068, 140KB

Original Image

Restored Image

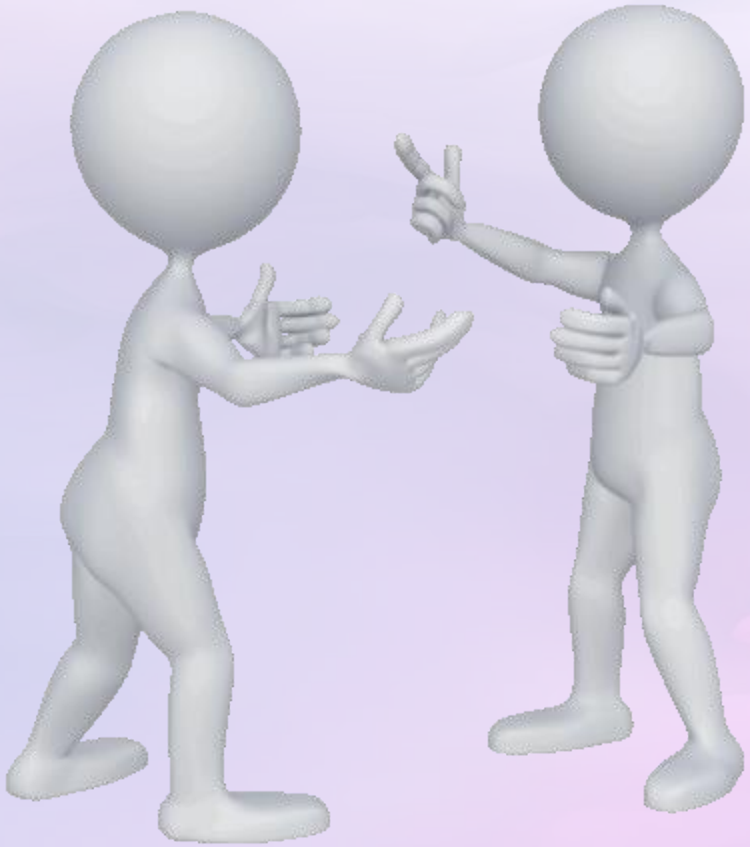
Competitor Model
PSNR: 29.4230, SSIM: 0.9427, VGG Loss: 2.8433

Our 8X Model
PSNR: 26.6906, SSIM: 0.9070, VGG Loss: 9.8749

Commercial Product (Remini)

Restored Image

Discussion



1

Interpretation

The model effectively enhanced image resolution, showing robust performance on general images. Results indicate strong detail preservation and artifact reduction, validating the model's efficacy for typical use cases.

2

Limitations

Higher computational requirements compared to simpler models. Struggled with very low-resolution images, particularly those featuring human faces, due to dataset limitations and the model's sensitivity to facial details.

3

Future Work

Explore integration of GFPGAN for improved face enhancement and to address the current model's limitations with human faces.

Conclusion

Key Findings and Contributions

- Demonstrated Real-ESRGAN's effectiveness in enhancing image quality across various types, including general images and anime-style visuals.
- Successfully addressed image degradation through downsampling, noise addition, and compression, improving restoration results.

Lessons Learned

- Gained insights into the complexities of image restoration and the importance of fine-tuning models for specific tasks.
- Importance of dataset diversity and quality in training models for specific applications.
- Need for specialized models like GFPGAN for tasks involving human faces, indicating future directions for improvement.

References

Key Citations:

•Wang, X., et al. (2021). "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data." Link to paper

•Ledig, C., et al. (2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." Link to paper

•He, K., Zhang, X., et al. (2016). "Deep Residual Learning for Image Recognition." Link to paper

Acknowledgments:

•We would like to thank the authors of Real-ESRGAN for providing the pre-trained model weights used in our project.

•Special thanks to the developers of TensorFlow, PyTorch, and other libraries that made this research possible.

•Gratitude to our guide and peers for their invaluable feedback and support throughout the project.

Thank You