Image Resolution and Denoising Enhancement with Deep Learning

Tejas Murali and Gautham Varma
UHID: 1612032 and 2328782

Abstract

We developed and evaluated two complementary approaches for image processing: one focusing on image resolution enhancement and the other on image denoising. For resolution enhancement, GAN and ESRGAN architectures were employed to synthesize realistic images and improve resolution. While the GAN generated high-fidelity images, the ESRGAN successfully enhanced resolution but occasionally introduced artifacts due to the balance in loss function and discriminator feedback. For image denoising, we implemented U-Net-inspired and ResNet-inspired autoencoders to reconstruct clean images from noisy inputs. Both architectures were evaluated on Gaussian noise-augmented datasets, demonstrating notable improvements in Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE). Our findings highlight the trade-offs between realism and resolution in enhancement tasks, alongside the strengths of perceptual fidelity and pixel-level accuracy in denoising tasks, emphasizing the importance of tailored designs for high-quality image generation and restoration.

Introduction

The first section of the project is about image denoising, which is a critical preprocessing step in computer vision. Our work developed and evaluated two complementary—U-Net-inspired and ResNet-inspired autoencoders—designed to handle noisy images efficiently. We evaluate these models on their performance metrics, scalability, and qualitative visual results.

The next section is improving resolution in images. Resolution enhancement has widespread applications, including medical imaging, digital media, and underwater imaging. GAN and ESRGAN utilize deep learning to improve pixel quality and upscale images, with ESRGAN capable of increasing image resolution up to four times its original size. Results demonstrate that GAN provides a modest improvement in image quality, while ESRGAN significantly enhances resolution, albeit with minor artifacts. Final losses calculated are low during the initial epochs itself, indicating that the models are performing well.

Data and Preprocessing

The dataset of choice that was decided to use in this research project study is the MIT-Adobe FiveK dataset. It was chosen due to its richness and diversity in content. It consists of 5,000 high-resolution images of various genres like landscapes, portraits, and urban scenes. This dataset shall provide a good representation of real- world image enhancement challenges.

Next, we will discuss the different pre-processing techniques and stages that were used for both sections of the project. We utilized a shared dataset of 3000 high-resolution images, resized to 256×256 for computational efficiency. Gaussian noise with factors of 0.2 and 0.3 was applied to simulate real-world scenarios. The dataset was split into 80% for training and 20% for validation/testing, with normalization applied to scale pixel values between 0 and 1.

For the preprocessing stage in the image resolution enhancement part, the main steps involved were first downloading the images from MIT-Adobe FiveK dataset website, saving them in a google drive folder. Then, we converted the images from DNG to JPG, which is more convenient to use when performing operations on images in general. Finally, for the ESRGAN, we downscaled the images so we could see the effect of the model's up sampling ability while it produces better resolution pictures.

Methodology

Table 1: Architecture of U-Net and ResNet Autoencoders for image Denoising

Component	U-Net-Inspired Autoencoder	ResNet-Inspired Autoencoder
Encoder	Conv2D layers with filters: 64, 128, 256, 512.	Residual blocks with filters: 64, 128, 256, 512.
	ReLU activation layers followed by MaxPooling2D for downsampling	Shortcut connections for enhanced feature retention and gradient flow
	Skip connections to preserve spatial details for the decoder	MaxPooling2D layers for progressive spatial reduction.
Bottleneck	Two Conv2D layers (512 filters) with ReLU activation	Residual blocks with 512 filters for compact, high-level feature extraction.
	Maintains reduced spatial resolution for feature representation.	
Decoder	UpSampling2D layers to reconstruct spatial dimensions	UpSampling2D layers for spatial reconstruction
	Concatenated skip connections and Conv2D layers to refine features	Skip connections for incorporating encoder details.
		Residual blocks for refining reconstructed feature maps.
Output Layer	Final Conv2D layer with sigmoid activation to map pixel values to [0, 1].	Final Conv2D layer with sigmoid activation to align pixel values with input.

Table 2: Architecture of ESRGAN and GAN Models for image enchantment.

Aspect	GAN (Generative Adversarial Network)	ESRGAN (Enhanced Super-Resolution GAN)
Objective	Generate realistic images by synthesizing from down sampled inputs.	Enhance low-resolution images to high-resolution with improved details.
Generator Architecture	Convolutional layers with ReLU for down sampling and transpose convolution with Tanh for up sampling.	Residual Blocks with skip connections and PixelShuffle for efficient up sampling.
Discriminator Architecture	Convolutional layers with LeakyReLU and a final Sigmoid activation, classifying real vs. fake images.	Similar architecture as GAN with modifications made for improved feedback to the generator.
Loss Functions	Combines adversarial loss (Binary Cross-	MSE loss for pixel-wise accuracy, incorporating

	Entropy) for realism and L1 content loss for fidelity.	adversarial loss for perceptual quality.
Training Approach	Alternates between optimizing the discriminator to classify real vs. fake and the generator to produce realistic images.	Focuses on MSE loss for generator training, transitioning to adversarial training since the discriminator is updated.
Key Features	Balances adversarial competition to improve realism while retaining fidelity via L1 loss.	Efficient up sampling with PixelShuffle and superior image quality using residual connections.
Advantages	Simple and effective for general-purpose image generation tasks.	Tailored for super-resolution tasks, producing sharper and more detailed images.
Limitations	Limited performance in tasks requiring high pixel-level accuracy.	Requires tuning of discriminator and loss weights for optimal results.

Why This Approach?
GANs effectively synthesize realistic images using adversarial training, with L1 loss ensuring fidelity. ESRGANs enhance this by incorporating Residual Blocks and PixelShuffle for super-resolution, with discriminator modifications in progress to improve feedback for finer details.

Alternatives Considered:
SRGAN: Predecessor to ESRGAN, limited by batch normalization and less refined loss functions, reducing detail quality.

The chosen methods leverage adversarial training, advanced architectures, and targeted losses for high-quality image generation and enhancement, balancing practicality and effectiveness.

Experimentation:

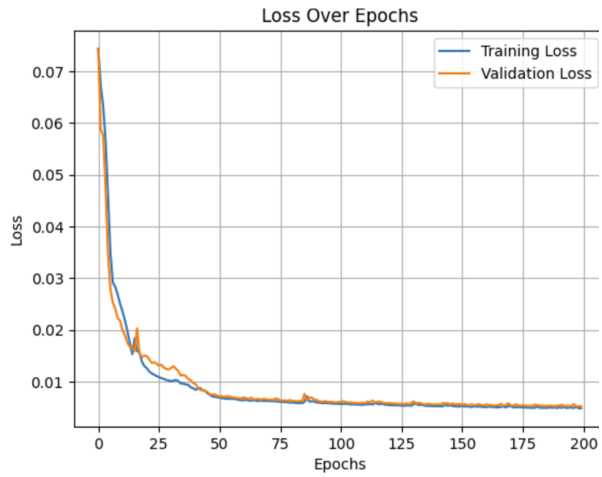
Experimentation for Image Denoising

Parameter	**U-Net Architecture**	**ResNet Architecture**
Loss Function	Mean Squared Error (MSE)	Perceptual Loss (VGG16 Features)
Optimizer	Adam (LR = 0.0001)	Adam (LR = 0.0001)
Early Stopping	Yes	Yes
Epochs	200	120
Batch Size	32	32

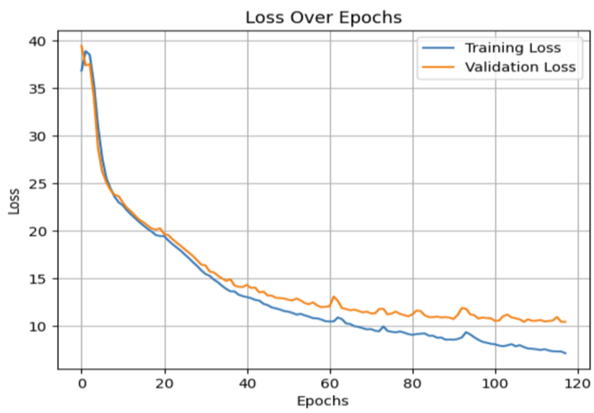
We experimented with various hyperparameters, including learning rates (0.001, 0.0005, 0.0001), batch sizes (16, 32, 64), and Gaussian noise factors (0.2 and 0.3). For the U-Net-inspired model, a noise factor of 0.3 and a learning rate of 0.0001 provided stable convergence. Similarly, the ResNet-inspired model employed a noise factor of 0.2 with the same learning rate to ensure consistent comparisons. The ResNet-inspired model utilized perceptual loss, leveraging pre-trained VGG16 features to focus on high-level perceptual fidelity rather than strict pixel-wise accuracy, ensuring visually coherent denoised outputs.

Early stopping callbacks were implemented for both models to prevent overfitting. For the ResNet-inspired model, training halted at 118 epochs, with the best weights restored from epoch 108. Training loss curves showed steady improvement, while validation loss plateaued, reflecting the models' ability to generalize effectively. The accompanying graphs highlight these trends, illustrating the stable training and convergence of both architectures.

U-Net Encoder Loss Graph:



ResNet Encoder Loss Graph:



Experimentation for Image Resolution Enhancement:

This section is the experimentation for the Image Resolution Enhancement. For this part of the project, 2 different models were used; namely, the ESRGAN and GAN. The first experiment conducted on ESRGAN shown below was after optimal hyperparameters (learning rate, batch size, number of layers, and number of epochs) were found. Other experiments before this can be found in the attached code file. Graphs show 50 epochs to view the overall trend, but the models themselves were trained up to 200 epochs for best results.

- Experiment 1: ESRGAN Model Resolution Enhancement

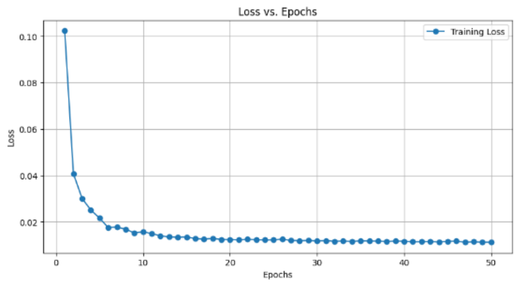


Figure 1: Loss vs. Epochs Graph for ESRGAN Training

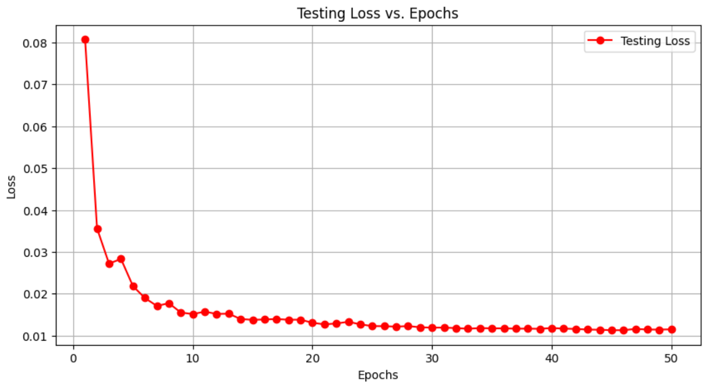


Figure 2: Loss vs. Epochs Graph for ESRGAN Testing

- Experiment 2: GAN Model Resolution Enhancement

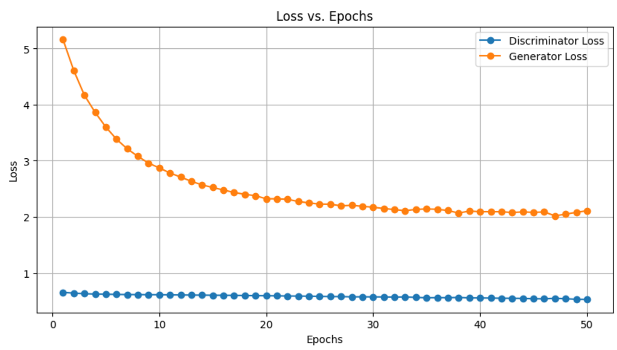


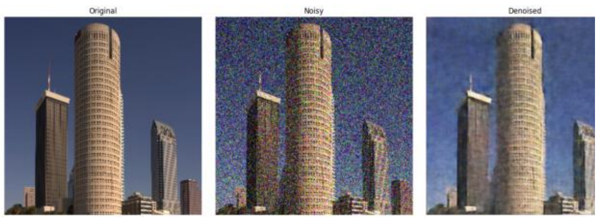
Figure 3: Loss vs. Epochs Graph for GAN Training

Figure 4: Loss vs. Epochs Graph for GAN Testing

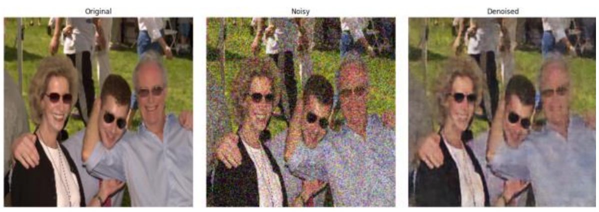
Results and Metrics for U-Net and ResNet-Inspired Autoencoders:

Metric	U-Net-Inspired Autoencoder	ResNet-Inspired Autoencoder
PSNR (dB)	23.75	21.85
MSE	0.0052	0.0069
Best Epoch	108	118

U_Net image:



ResNet image:

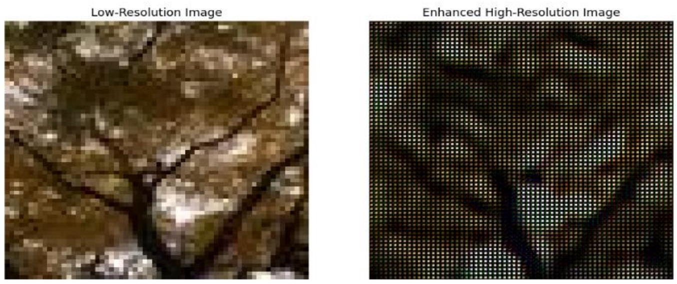


The U-Net-inspired autoencoder exhibited superior pixel-wise reconstruction, achieving a PSNR of 23.75 and an MSE of 0.0052. These metrics underscore its ability to effectively denoise images by minimizing pixel-level intensity differences. The architecture's skip connections and hierarchical feature extraction proved vital for preserving spatial details, leading to smooth convergence and minimal overfitting during training.

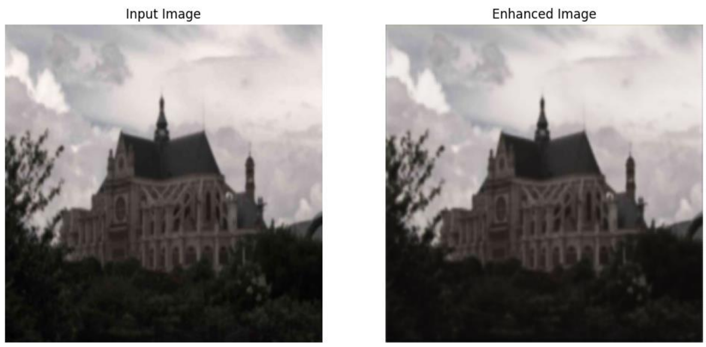
In contrast, the ResNet-inspired autoencoder achieved a PSNR of 21.85 and an MSE of 0.0069. While these values are slightly lower than those of the U-Net model, the ResNet-based approach demonstrated its strength in handling complex noise patterns due to its residual connections. These connections facilitated robust gradient flow, enhancing the model's stability and generalization, albeit with slightly higher validation loss due to the added complexity.

Results for Image Resolution Enhancement with ESRGAN and GAN:

ESRGAN Image:



GAN Image:



The ESRGAN and GAN models both overall did a decent job in enhancing the resolution of the original images by increasing the pixel quality and pixel count in the images. As you can see, with both models running up to 200 epochs, the ESRGAN enhanced picture comes with a few artifacts, but the picture details appear a bit sharper. As for the GAN model, the enhanced image is very slightly improved in resolution, and the generator does a very good job of generating a realistic image compared to the original. Losses discussed in the methodology section were used as a metric to compare how well the enhancement process through the deep learning models turned out.

Insights and Novel Contributions:

U-Net-Inspired Autoencoder

1. **Efficient Skip Connections:** The U-Net architecture leveraged skip connections to effectively bridge the encoder and decoder, preserving fine-grained spatial details often lost during downsampling. This feature was critical in retaining the structural integrity of denoised images.

2. **Balanced Reconstruction through MSE Loss:** The pixel-wise Mean Squared Error (MSE) loss function ensured precise reconstruction of denoised images without introducing artifacts. While this method excels in minimizing numerical differences, it revealed limitations in perceptual quality, such as smooth but less visually realistic outputs.

The U-Net model, while effective in denoising, struggled with capturing high-level contextual features due to its reliance on shallow feature representations. This limitation informed the need for an alternative approach to address more complex noise patterns and perceptual fidelity. we proposed the ResNet-inspired autoencoder particularly for handling high-level semantic information and achieving perceptual quality,. This design incorporated several innovative improvements to address these challenges.

ResNet-Inspired Autoencoder

1. **Novel Use of Residual Connections**: The architecture employed residual blocks within both the encoder and decoder, ensuring enhanced feature retention and mitigating vanishing gradients. This design facilitated efficient training and preserved both low-level and high-level features, crucial for reconstructing noisy inputs.

2. **Perceptual Loss Integration**: Unlike the U-Net model, which relied solely on pixel-wise accuracy, the ResNet-inspired model incorporated perceptual loss. By leveraging the pre-trained VGG16 model to extract high-level features, this approach emphasized visual quality over exact numerical reconstruction. The outputs appeared more natural and visually realistic, demonstrating significant improvements in perceptual fidelity.

3. **Architectural Innovation**: Combining residual blocks with skip connections uniquely balanced fine-grained detail retention and high-level feature extraction. This hybrid approach resolved the limitations observed in the U-Net model, making the ResNet-inspired autoencoder more suitable for perceptually driven tasks. Early stopping further optimized training by halting at validation loss plateau, restoring the best-performing model weights.

While the U-Net-inspired autoencoder served as an effective baseline leveraging established techniques, the ResNet-inspired architecture introduced a novel hybrid design with residual connections and perceptual loss, demonstrating enhanced performance for perceptually challenging denoising tasks.

GAN Insights

The GAN demonstrated its strength in producing realistic images closely resembling the original low-resolution inputs. Its generator, using transposed convolutions and L1 content loss, focused on preserving the overall structure and fidelity of the input data. The discriminator effectively guided the generator to synthesize images that were visually indistinguishable from real samples. However, while the images appeared realistic, the GAN lacked the capability to enhance resolution significantly, which is essential for applications requiring fine detail and texture reconstruction.

ESRGAN Insights

The ESRGAN, leveraging Residual Blocks and PixelShuffle, successfully enhanced the resolution of low-resolution images, capturing finer textures and intricate details. However, this came at the cost of introducing artifacts in some regions, likely due to the imbalance between pixel-wise MSE loss and adversarial loss in guiding the generator. The current discriminator's performance, modeled after the GAN's structure, provided insufficient feedback

for perceptual refinement. This underscores the need for more sophisticated perceptual loss functions or improved discriminator architectures to achieve both high resolution and artifact-free results.
These findings highlight the trade-off between realism and resolution enhancement, offering insights into the strengths and limitations of each approach for specific image synthesis objectives.

Conclusion:

In this project, we implemented and evaluated two deep learning architectures for image denoising: a U-Net-inspired autoencoder and a ResNet-inspired autoencoder. The U-Net model effectively retained spatial details due to its skip connections, making it highly effective for tasks requiring precise pixel-level accuracy. The ResNet-inspired model, with its residual connections and use of perceptual loss, demonstrated robustness in handling complex noise patterns and emphasized perceptual quality. Together, these models showcase complementary strengths and provide a solid foundation for advanced image restoration tasks. For the image resolution enhancement section, we implemented the ESRGAN and GAN model architectures. The GAN, with its convolutional-based generator and discriminator, effectively produced realistic images closely resembling the originals, demonstrating the strength of its adversarial and L1 losses in balancing realism and fidelity. Similarly, the ESRGAN, featuring Residual Blocks and PixelShuffle for upsampling, successfully enhanced image resolution, showcasing its capability to add fine details, though some artifacts were observed. Both models highlight their unique strengths, with the GAN excelling in realism and the ESRGAN pushing the boundaries of resolution enhancement.

Future Work:

1. Diffusion Models:
   Explore state-of-the-art Denoising Diffusion Probabilistic Models for iterative noise removal and enhanced stability
2. Dataset Expansion: Train and validate the models on larger and more diverse datasets, including various real- world noise distributions, to strengthen their robustness and generalization capabilities.
3. Future work for image resolution enhancement could focus on refining the ESRGAN's discriminator to provide more precise feedback and reduce artifacts, improving overall image quality. For the GAN, incorporating additional perceptual loss functions or exploring advanced architectures could further enhance realism and fidelity. Additionally, experimenting with hybrid models combining elements of both approaches could unlock new possibilities for image synthesis and enhancement.

References

1. **Ronneberger, O., Fischer, P., & Brox, T.** (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. [Link: https://arxiv.org/abs/1505.04597]
2. **He, K., Zhang, X., Ren, S., & Sun, J.** (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [Link: https://arxiv.org/abs/1512.03385]
3. "A Comprehensive Review on Microgrid Control Strategies." *Electric Power Systems Research*, vol. 189, 2020, 106602. [Link: Diving deeper into underwater image enhancement: A survey - ScienceDirect]
4. "MIT-Adobe FiveK Dataset." *MIT CSAIL*, https://data.csail.mit.edu/graphics/fivek/. Accessed 9 Dec. 2024.