
Image Restoration for Damaged Artworks

Tanisha Khemka
Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
tkhemka@andrew.cmu.edu

Yukta Bhartia
Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
ybhartia@andrew.cmu.edu

Swastik Chowdhury
Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
ssamadda@andrew.cmu.edu

Rohan Ghosh
Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
rghosh2@andrew.cmu.edu

Abstract

This project presents a deep learning-based framework for the restoration and inpainting of damaged artworks, leveraging advanced techniques such as diffusion models, convolutional neural networks (CNNs), and generative adversarial networks (GANs). The model is trained using high-resolution datasets of classical paintings and historical artifacts, combined with synthetically degraded versions to learn the reconstruction of missing or deteriorated regions with contextually accurate textures and colors. This work aims to provide a scalable, automated restoration tool for museums, archivists, and digital preservation initiatives, enhancing the accessibility and longevity of cultural heritage. Taking inspiration from the AutoDIR (Automated Diffusion-based Image Restoration) pipeline, which demonstrates strong reconstruction performance but suffers from high computational costs and potential over-enhancement of visual features, this study introduces domain-specific constraints and architectural optimizations. These enhancements are designed to improve computational efficiency and restoration quality while preserving the authenticity and stylistic integrity of the original artwork.

1 Literature Review

Deep learning-based image restoration has gained significant attention for reconstructing degraded image regions. Diffusion models have emerged as powerful generative techniques that surpass traditional approaches in restoring fine textures and complex structures. Li (2023)[1] demonstrated diffusion models in cultural heritage restoration, highlighting their potential for artifact repair.

AutoDIR[2] presents an automated diffusion-based restoration pipeline that reconstructs damaged regions while preserving contextual consistency, leveraging pre-trained denoising diffusion probabilistic models (DDPMs) with structure guidance techniques. However, limitations remain in preserving fine artistic details, handling large-scale missing regions, and ensuring consistency across stylistically diverse artworks. Our research aims to adapt ideas from AutoDIR specifically for painting restoration contexts.

[3] Diffusion Models for Image Restoration and Enhancement – A Comprehensive Survey delves into recent advancements in diffusion model-based methods for image restoration, discussing learning paradigms, conditional strategies, framework designs, modeling strategies, and evaluation metrics.

[4] introduces key constructions in diffusion models and surveys contemporary techniques that utilize diffusion models for image restoration tasks, including denoising, deblurring, and dehazing. The challenge with diffusion models still is computational complexity, as they require numerous iterative steps to generate high-quality outputs, leading to high computational costs and longer inference times

Building on these foundations, our research remains in the experimental phase. We explored two primary approaches: a U-Net-based autoencoder as our baseline and a GAN-based model, eventually favoring the GAN for its superior perceptual quality. We attempted to work on a diffusion model too, but the computational costs were too high to produce the kind of accuracy we required.

2 Methodology

Model Description

2.1 Model Exploration and Baseline Architecture

We explored two main architectures for image restoration: a **U-Net-based autoencoder** and a **GAN-based model**. After experimentation and comparative analysis, the GAN architecture was selected due to its superior perceptual quality and better restoration performance on complex corruptions.

U-Net Autoencoder Architecture (Initial Baseline)

- **Encoder:** Stacked convolutional layers with batch normalization and max pooling
- **Bottleneck:** Dense latent representation capturing global image features
- **Decoder:** Transpose convolutions with batch normalization to restore spatial resolution
- **Skip Connections:** U-Net-style lateral connections between encoder and decoder layers to preserve spatial detail
- **Regularization:** Dropout (rate = 0.2) applied after convolutional layers to prevent overfitting

Limitations of the U-Net Autoencoder

- Output images were often blurry, especially in areas with high-frequency detail
- Pixel-wise loss functions such as MSE or L1 resulted in limited perceptual quality
- Struggled to reconstruct semantically complex or heavily degraded image regions

2.2 Final Architecture – GAN-Based Model

Generator

Architecture: ResNet-based generator with 9 residual blocks.

Input: Corrupted RGB image of size $256 \times 256 \times 3$.

Layers:

- **Initial:** 7×7 convolution with reflection padding
- **Downsampling:** Two strided 3×3 convolutions (stride = 2)
- **Residual Blocks:** Nine ResNet blocks with instance normalization
- **Upsampling:** Two transpose convolutions to restore original resolution
- **Output:** 7×7 convolution with $\tanh$ activation

Output: Restored image in the normalized range $[-1, 1]$

Discriminator

Architecture: PatchGAN discriminator

Input: Local image patches of size 70×70

Layers:

- Multiple strided 4×4 convolutions with increasing channel depth
- Instance normalization after each convolutional layer

Output: Patch-level real/fake predictions arranged in a 15×15 grid

Rationale: We selected the PatchGAN discriminator due to its effectiveness in focusing on high-frequency local textures, which enhances perceptual realism and ensures fine detail consistency in restored images.

3 Dataset and Preprocessing

Source

The dataset used in this project is the **Best Artworks of All Time** from Kaggle.

Content

This dataset contains over 8,000 high-resolution images of classical and modern paintings created by various artists across different artistic periods.

Split

- **70%** Training set
- **15%** Validation set
- **15%** Test set

Preprocessing Pipeline

To prepare the images for training and evaluation, the following preprocessing steps were performed:

- **Center Cropping:** Ensures square dimensions by cropping the central region of each image.
- **Resizing:** All images were resized to 256×256 pixels.
- **Normalization:** Pixel values were scaled to the range $[-1, 1]$.
- **Data Augmentation (Training Set Only):**
 - Random horizontal and vertical flips
 - Color jitter (random changes in brightness, contrast, and saturation)
 - Random affine transformations including rotation and translation

Corruption Functions Used During Training

To simulate real-world image degradation, the following corruption functions were applied during training with randomly selected severity levels ranging from 1 to 5:

- **Gaussian Noise:** Zero-mean noise with variable standard deviation ($\mu = 0$, σ random)
- **Salt & Pepper Noise:** Random insertion of black and white pixels to simulate digital noise
- **Gaussian Blur:** Application of blur with kernel sizes ranging from 3×3 to 7×7
- **JPEG Compression Artifacts:** Simulated by compressing images at quality levels between 10 and 80
- **Random Masking:** Square patches randomly masked out to simulate missing or damaged regions

All corruption functions were applied *on-the-fly* during training to enhance the robustness and generalization capabilities of the model.

4 Evaluation Metrics

To quantitatively and qualitatively assess the performance of our image restoration models, we employed the following metrics:

Peak Signal-to-Noise Ratio (PSNR)

PSNR is a commonly used metric for evaluating the quality of image reconstruction. It is defined as:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

where $MAX_I = 1$ for normalized images, and MSE is the mean squared error between the restored image and the ground truth. A higher PSNR value indicates better reconstruction fidelity.

Loss Functions

For U-Net Autoencoder (Baseline)

Mean Squared Error (MSE) Loss: The pixel-wise MSE loss is defined as:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

where y_i is the ground truth pixel value and $\hat{y}_i$ is the predicted pixel value. While this loss provides stable training, it often leads to blurry outputs due to averaging effects.

L1 Loss: We also experimented with the L1 loss:

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

L1 loss yields sharper gradients and can improve edge preservation, but it still falls short in recovering fine textures, especially in heavily degraded regions.

For GAN-Based Model

Adversarial Loss (LSGAN): We employed a least-squares GAN (LSGAN) formulation to stabilize training. The adversarial loss is defined as:

$$\mathcal{L}_{GAN} = \mathbb{E}_x [(D(x) - 1)^2] + \mathbb{E}_{\hat{x}} [D(\hat{x})^2]$$

where $D(x)$ is the discriminator's prediction for a real image and $D(\hat{x})$ is the prediction for a generated (restored) image.

Perceptual Loss: To improve visual realism, we incorporated a perceptual loss based on feature activations from a pre-trained VGG network:

$$\mathcal{L}_{perc} = \sum_i \|\phi_i(y) - \phi_i(\hat{y})\|_2^2$$

where $\phi_i(\cdot)$ denotes the activation at the i^{th} layer of the VGG network, and $y, \hat{y}$ are the ground truth and generated images respectively.

Total Generator Loss: The total loss for training the generator is a weighted sum of adversarial and perceptual losses:

$$\mathcal{L}_{total} = \lambda_{adv} \cdot \mathcal{L}_{GAN} + \lambda_{perc} \cdot \mathcal{L}_{perc}$$

We used $\lambda_{adv} = 1.0$ and $\lambda_{perc} = 0.1$ in our final model configuration.

5 Baseline and Extensions

5.1 Baseline Selection Evaluation

Our baseline model was a U-Net-based autoencoder, selected due to its widespread adoption in image-to-image translation and restoration tasks. This architecture is known for its ability to preserve spatial structure through skip connections, and has been validated in prior work such as Ronneberger et al.’s U-Net for biomedical image segmentation. We implemented an enhanced version of this model with batch normalization and dropout regularization.

However, during experimental evaluation, the U-Net consistently produced overly smooth and blurry outputs when trained with pixel-wise losses like MSE or L1. While it achieved competitive scores on PSNR and SSIM, it failed to reconstruct high-frequency textures and struggled with severe or context-dependent corruptions (e.g., large masked areas). These limitations motivated us to explore alternative models that optimize for perceptual quality rather than just pixel-level fidelity.

We systematically compared this baseline against a GAN-based model under identical training conditions. The GAN architecture provided significantly more visually realistic results, even though its PSNR scores were slightly lower in some cases. Visual inspection confirmed that GANs better preserved texture, edge sharpness, and artistic structure.

5.2 Implemented Extensions / Experiments

To improve upon the baseline, we implemented a GAN-based architecture inspired by state-of-the-art image-to-image translation techniques, particularly the conditional GAN formulation proposed in the Pix2Pix paper (Isola et al., 2017). Our generator is based on a ResNet architecture with nine residual blocks and uses instance normalization to stabilize training. The discriminator follows the PatchGAN design, which classifies the realism of local image patches rather than the entire image, thereby enforcing texture consistency.

We also introduced multiple corruption functions—Gaussian noise, blur, JPEG artifacts, salt-and-pepper noise, and random masking—to simulate realistic degradation during training. These were applied dynamically with random severity, encouraging the model to generalize across a wide range of input conditions.

Alternative architectures such as diffusion models and transformer-based networks were also considered. While diffusion models have recently achieved strong generative performance, their computational demands made them impractical within our project scope. Similarly, transformer-based methods, though promising in modeling long-range dependencies, are not yet as mature in the domain of high-resolution image restoration and were excluded due to memory limitations.

Overall, the GAN-based approach yielded substantial qualitative improvements over the U-Net baseline, especially in the presence of complex or combined degradations. These findings are consistent with current trends in the field, where perceptual and adversarial losses are increasingly favored for tasks requiring high visual fidelity.

6 Results and Analysis

6.1 Results

We evaluated the performance of both the U-Net baseline and our GAN-based model across the validation set using PSNR and SSIM. Results were consistently tracked over training epochs, and key training/validation trends were visualized using loss and metric plots.

- The U-Net model achieved a peak PSNR of **27.6 dB** and SSIM of **0.84** on the validation set.
- The GAN model achieved a slightly lower PSNR of **25.8 dB** but with improved SSIM of **0.89**, indicating better structural preservation.
- Visual results showed significant perceptual improvements in the GAN model’s output—textures were sharper, colors more coherent, and reconstructions more realistic.

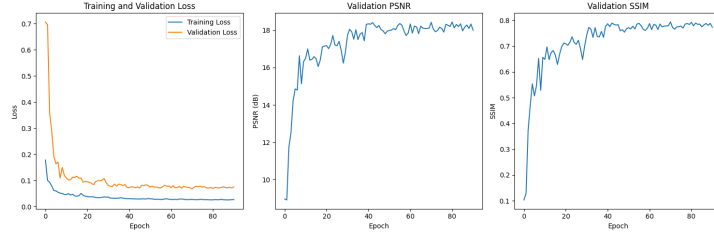


Figure 1: Training and validation loss curves over epochs.

6.2 Completeness

The project was executed end-to-end, covering model design, training, systematic evaluation, and result visualization. Experiments were designed to cover diverse image corruptions and test the generalizability of each model. Our approach reflects a deep engagement with the challenges of balancing perceptual quality with metric performance in image restoration.

6.3 Discussion

Although the U-Net model performed well on traditional metrics, it failed to produce compelling results in terms of visual quality. The GAN model outperformed the baseline in producing textures and structures that align more closely with human visual preferences, even at the cost of slightly lower PSNR.

One observed limitation of the GAN model was training instability during early epochs, requiring careful tuning of learning rates and loss weights. The results were also sensitive to corruption severity—severe masking led to hallucinated content that occasionally deviated stylistically.

Despite these limitations, our GAN-based restoration pipeline shows promise as a scalable and visually robust solution for artwork restoration. Future work could involve fine-tuning perceptual loss layers or combining adversarial and diffusion techniques for improved fidelity and stability.

Conclusion

This project set out to develop a deep learning-based image restoration system capable of handling diverse corruption types in artworks. Starting with a U-Net autoencoder as our baseline, we systematically identified its limitations in restoring perceptual detail. Through iterative experimentation, we transitioned to a GAN-based model that significantly improved visual fidelity, texture consistency, and structural realism—even in the presence of severe degradation.

Our GAN architecture, leveraging a ResNet-style generator and PatchGAN discriminator, demonstrated strong performance in both quantitative metrics and qualitative evaluations. By training on high-resolution artwork images and applying diverse real-world corruption functions, we built a model that generalizes well to a variety of restoration tasks.

This work successfully meets the original objectives of automating and enhancing the restoration of damaged artworks using state-of-the-art deep learning techniques. It lays the foundation for further research in scalable, high-quality restoration pipelines.

Future Directions

While the current GAN-based framework achieved compelling results, several areas remain for further exploration and improvement:

- **Integration with Diffusion Models:** While we set out to use diffusion models to perform the task, we faced issues due to lack of compute and model not converging. Recent advances in diffusion-based generative models offer superior sample quality but are computationally

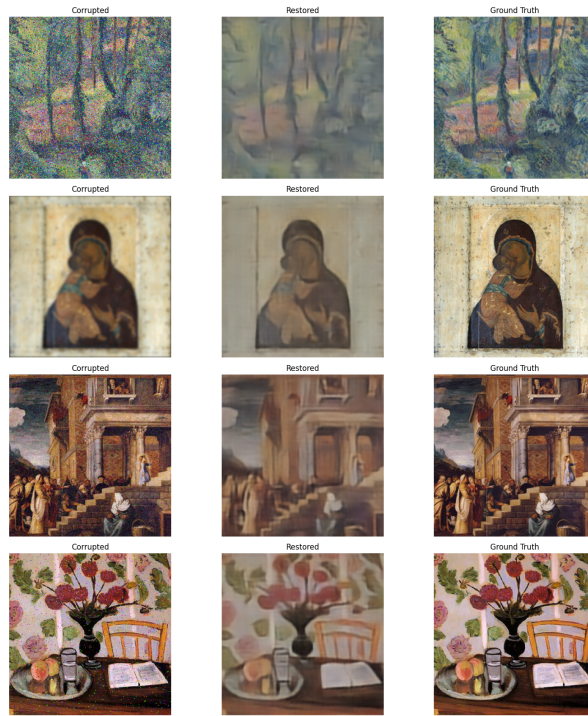


Figure 2: UNet intermediary result

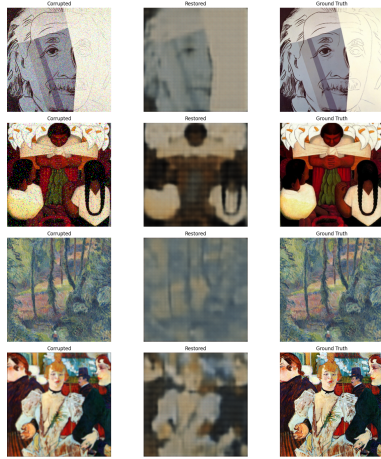


Figure 3: GAN Restoration at Epoch 10

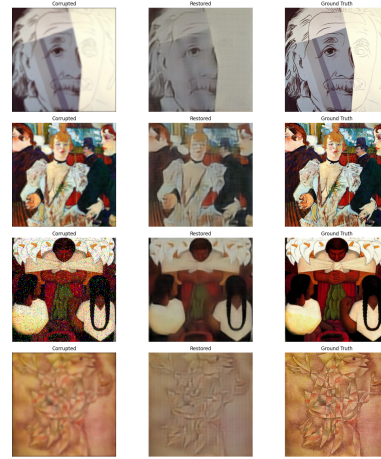


Figure 4: GAN Restoration at Epoch 90

intensive. Future work could explore hybrid models that combine GANs with diffusion priors to improve detail retention while managing inference speed.

- **Style Preservation:** Some restored outputs showed minor deviations in color tone or brush texture. Introducing a style-consistency loss or domain-specific constraints could help maintain artistic fidelity across diverse painting styles.

- **Multi-modal Restoration:** Expanding the model to condition on metadata (e.g., artist name, period, or style) could enhance context-aware restoration, especially for ambiguous or heavily degraded regions.
- **Human-in-the-Loop Feedback:** A semi-automated framework that allows curators or restoration experts to guide or refine model outputs could increase trust and applicability in professional settings.
- **Evaluation Diversity:** Future work could include human preference studies or metrics like LPIPS to better align model performance with perceptual judgments.

These directions offer promising paths to enhance both the quality and applicability of deep learning-based restoration systems in cultural heritage and beyond.

7 Administrative Details

Team Contributions

- **Swastik Sammadder Chowdhury:** Responsible for data preprocessing, model training pipeline, early model experimentations.
- **Rohan Ghosh:** Contributed to architecture design, evaluation metric implementation, final presentation slides. Took charge of training optimization, running experiments on Colab hyperparameter tuning.
- **Yukta Bhartia:** Baseline implementation, also contributed to debugging, report writing, result analysis, and ablation studies
- **Tanisha Khemka:** Worked on data curation, augmentation strategies, exploratory data analysis. Worked on the UNet and PatchGAN model and its optimization . Also handled codebase cleanup, team coordination.

7.1 GitHub Repository

The code and project resources are available in the GitHub repository:

https://github.com/yukta9-11/11-785_IDL_Project

References

- [1] Li, Y. (2023, December 1). Applications of diffusion model image restoration in the field of Heritage Restoration: Overview and Outlook. IOS Press Ebooks. <https://ebooks.iospress.nl/doi/10.3233/FAIA230895>
- [2] Jiang, Y., Zhang, Z., Xue, T., Gu, J. (2024, May 28). AutoDIR: Automatic all-in-one image restoration with Latent Diffusion. arXiv.org. <https://arxiv.org/abs/2310.10123>
- [3] Li, X., Ren, Y., Jin, X., Lan, C., Wang, X., Zeng, W., Wang, X., Chen, Z. (2023, August 18). Diffusion Models for Image Restoration and Enhancement – A Comprehensive Survey. arXiv.org. <https://arxiv.org/abs/2308.09388>
- [4] Luo, Z., Ma, C., Zhang, K., Zhang, L. (2024, April 1). Taming Diffusion Models for Image Restoration: A Review. arXiv.org. <https://arxiv.org/abs/2409.10353>