ArtResGAN: A GAN-Based Approach for Artistic Image Restoration and Style Preservation

Shaurya
shaurya.2022@viststudent.ac.in

Devika Iyer
devika.iyer2022@viststudent.ac.in

Vishesh Panchal
vishesh.2022@viststudent.ac.in

Artistic image restoration seeks to restore degraded artworks with all of their subtle stylistic details still intact. The major task of this work, which is named ArtResGAN, presents a GAN-based framework for the restoration of artistic images. ArtResGAN utilizes a Hybrid U-Net + ResNet generator for the generation, a PatchGAN discriminator to classify image build, and a VGG-based style extractor to extract style. The fine details are generated with the use of skip connections and ResNet blocks, and realism is created with an adversarial loss training session with the discriminator. In addition, Machine Vision techniques such as edge detection, morphological openings and closings, Haar wavelet transforms, and Local Binary Patterns are used to provide additional texture and edge preservation. The training employs the WikiArt dataset to train and employs PSNR, SSIM, and FID scores to evaluate. ArtResGAN hopes to find an equilibrium where fidelity and artistic consistency can co-exist, through an elegant combination of Content loss, Adversarial loss, Style loss, and Total Variation loss. Among the research areas the project plans to bridge by providing zero-shot restoration are the continuums of image distortions and deep priors towards the broad category of AI-assisted image restoration. Further work could include multi-scale training and attention mechanisms that will take restoration to new frontiers.

I. Introduction

Preservation of artworks is most paramount to the principal care for cultural heritage: they ensure that great historical and artistic works get to stand the test of time. However, most of the traditional restoration processes do take long, are prone to human flaws, and may border on irreversible damage when certain miscalculations occur. After the era of deep learning came about, automated restoration methods appeared to be some noble answers to these problems. Although existing deep-learning-based models have demonstrated that, with significant improvements to image denoising and inpainting, have been unable to retain the stylistic motives of artworks, leaving artifacts of restored images, with no semblance to texture, stroke styles, and artistic degrees of freedom of the original painting. It is because of this that a restoration framework incorporating both structural and stylistic integrity in the restoration process is called upon.

Traditional restoration methods often struggle to balance technical correction with artistic authenticity, particularly when confronted with diverse degradation types such as cracks, stains, fading, and missing regions. These limitations have spurred interest in leveraging artificial intelligence, specifically deep learning approaches, to develop more sophisticated restoration solutions.

Recent advances in generative adversarial networks (GANs) and vision transformers have demonstrated remarkable capabilities in image synthesis and manipulation. However, existing approaches typically focus on either structural accuracy or style preservation, rarely doing both well. This is problematic in artwork restoration where the stylistic integrity of the artwork like brushwork, texture, and color palette is crucial.

We propose ArtResGAN, a next-generation deep learning-based restoration framework specifically designed to address the dual challenges of structural reconstruction and stylistic fidelity in degraded artwork restoration. Our approach integrates a hybrid UNet+ResNet generator architecture with a style-aware PatchGAN discriminator and specialized loss functions to create a comprehensive solution that respects both the physical integrity and artistic authenticity of the original work.

II. Literature Review

Deep Learning has revolutionized the domain of image restoration by facilitating end-to-end learning of complex mapping from degraded images to their restored images. Conventional approach to this problem involved the use of custom priors and statistical models, whereas neural networks automate this task by leveraging large datasets and deep architectures to generalize over different image degradations. This review entails the most significant contributions to the domain, specifically the CNN based denoising, GAN based inpainting and prior based generative models.

The foundation of deep learning based denoising is the DnCNN model as proposed by Zhang et al [1]. It introduces a deep CNN network for residual learning for Gaussian denoising wherein the model learns the noise from the image and subtracts it from the image thus allowing it to generalize better than traditional methods and handle denoising effectively. However, DnCNN struggled with non-Gaussian noise and required extensive datasets to train efficiently. To address this degradation, Cai et al [2] proposed a CNN based approach trained on hazy images to estimate the transmission map. It was tailored for single haze images by learning haze relevant features and reconstructing clear images. This innovative approach directly learns the haze patterns from the data instead of relying on traditional physics-based models facilitating efficient end to end learning. While it can handle noise well, it is unable to generalize well over different haze levels. As proposed by Ulyanov [3] et al, DIP eliminates the reliance on external training data by leveraging the inherent CNN structure as a prior for image restoration, thereby enabling unsupervised learning. One disadvantage of this approach is that it fails to generalize well for complex degradations.

GAN based inpainting methods is another avenue of forthcoming research. One significant advancement in the field is the Contextual Attention mechanism as proposed by Yu et al[4]. It introduces a GAN framework that learns to reconstruct missing regions by borrowing context from surrounding areas. This method was further refined by Partial Convolutions, as proposed by Guilin et al[5]. This improved the mask-aware learning, ensuring that meaningful features are propagated through the network. However both approaches suffered from blurry and unrealistic textures in large missing areas. Gated Convolutions[6] addressed this issue by incorporating dynamic feature selection for free form image inpainting, producing more visually consistent results.

Spectral normalization in GANs has been studied by Henaff et al[7], which analysed the stabilizing power of the generator's Jacobian norm to improve the robustness, mitigate mode collapse in generative models. Building upon this, Proximal DeHazeNet[8] combined deep learning with traditional methods of prior based dehazing thereby bridging the gap between physics based and data driven methods. As proposed by Nag et al[9], introduced a similar approach by using prior-aware semantic inpainting to refine texture details and improving the realism in missing regions. This solves the problems suffered by the GAN based approaches because it uses prior based dehazing to improve the realism.

Recent works have focused on enhancing inpainting models with novel architectures. Improved GAN based Inpainting, as proposed by Chen et al [10], integrated multi-scale feature aggregation for refined details. AU-GAN [11] leveraged an attention-based U-Net architecture to enhance feature selection for more natural restoration. Pluralistic Image Completion [12] introduced a framework for generating multiple plausible restorations, addressing the deterministic nature of prior inpainting approaches. HiFill [13] further refined hierarchical feature fusion to improve spatial consistency and global coherence.

Deep generative priors have emerged as prominent tools for restoration, as demonstrated by Xia et al [14], which leveraged pre-trained generative models to enhance denoising and deblurring performance. As introduced by Asad et al [15], MGAN-CRCM leveraged a multi-GAN and coarse-to-refined contextual memory approach to set new benchmarks for high-quality image inpainting by integrating memory mechanisms.

III. RESEARCH GAP

While deep learning-based image restoration methods, particularly CNN and GAN-based approaches, have already given the field a leg up, there are several major shortcomings with their application in art restoration. The primary issue here is that there is a lack of a good mechanism to enable these methods to retain the stylistic components of artworks. Existing models indeed perform well in the reconstruction of structural features, but the fine artistic components of paintings-the crackle of brushstrokes, the texture patterns, the subtle gradient shifts of color, which give pieces of art their beauty and historical significance-are often overlooked. In addition to that, GAN-based inpainting methods such as contextual attention and partial convolutions struggle with blurry and less-than-realistic textures, especially where large portions of the image are missing or seriously degraded. These methods do not model the semantics in the highly damaged area, hence giving rise to inpainted regions that feel artificial or out-of-place.

Yet another critical gap is the inflexibility and inefficiency of restoration methods. The vast majority of methods rely on fixed degradation models, which seem to restrict their adaptability to hazes, noise, cracks, and lost content that accept various values in normal artwork degradation. In addition to that, most deep learning-based restoration methods operate only in a single pass; this means that they produce only one output without any iteration. This way, there would not be any optimal restoration since the model does not rescript its predictions on subsequent iterations to fix errors. Finally, a further complication to closure of the gaps concerns the lack of objective metrics of evaluation. Most restoration models are given an informal visual inspection, while the rest have not been made equivalent with quantifiable clear benchmarks. The result is that comparing different methodologies has been difficult, and this prevents restoration from being done adequately. Resolving these gaps means that an adaptive, iterative, and stylistically conscious method should be developed. This is important to achieve quality restoration while still accommodating the artistic nature of the artwork in its degraded condition.

IV. PROPOSED SYSTEM

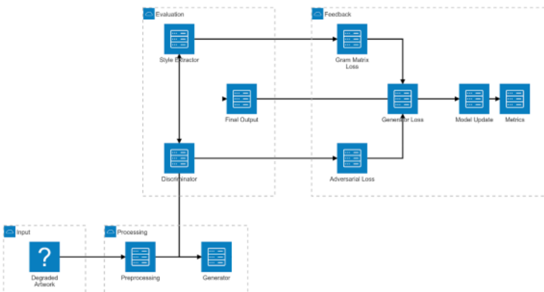

FIGURE 1. ARCHITECTURE DIAGRAM FOR ARTRESGAN

Our artwork restoration framework, ArtResGAN, integrates several complementary deep learning architectures and techniques to address the complex challenge of restoring degraded artwork while preserving artistic style. The model comprises three major components: a hybrid generator network, a PatchGAN discriminator, and specialized loss functions, all working in concert to produce high-quality restorations.

I. Generator (G)

The ArtResGAN generator utilizes a hybrid UNet + ResNet architecture to effectively capture both global context and local details crucial for artwork restoration. The network accepts a 9-channel input consisting of the 3-channel RGB degraded image concatenated with 6 additional channels of machine vision features, which provide explicit structural information to guide the restoration process.

- Res-Net Blocks: At the core of the generator's bottleneck are the ResNet blocks, which enable deep feature extraction while improving gradient flow throughout the network. Each ResNet block consists of two convolutional layers with reflection padding, instance normalization, and ReLU activation functions. The defining characteristic of these blocks is the residual connection. Each ResNet block can be formulated as:
$$\mathbf{F_{Resblock}}(\mathbf{x}) = \mathbf{x} + \mathcal{F}(\mathbf{x}, \mathbf{W})$$
where x is the input to the ResNet block, $\mathcal{F}$(x,W) represents the residual mapping to be learned and W are the weights of the two convolutional layers with instance normalization and ReLU activation. This architecture allows the network to learn complex transformations while mitigating the vanishing gradient problem in deep networks. By using six of these blocks in sequence at the bottleneck of the UNet structure, the generator effectively captures high-level representations of the artwork's content and style, which are crucial for accurate restoration.
- RRDB(Residual in Residual Dense Blocks): The Residual in Residual Dense Blocks (RRDB) serve as fundamental building blocks for the super-resolution component of our framework. Each RRDB consists of three dense blocks, with each dense block containing multiple convolutional layers and LeakyReLU activations. These blocks implement a dense connection pattern where each layer receives feature maps from all preceding layers, enabling more efficient feature reuse. The RRDB design can be formulated as:
$$F_{RRDB}(x) = x + \beta \cdot \sum_{i=1}^{3} DB_i(x)$$
where x is the input to the RRDB block, DBi represents the i-th dense block, and β is a scaling factor (typically 0.2) used to stabilize training.
Within each dense block, the dense connections enable a rich hierarchical feature representation:
$$DB(x) = \mathrm{Conv}\Big([\mathrm{x}, f_1(x), f_2([x, f_1(x)]), ..., f_n([x, f_1(x), ..., f_{n-1}])]\Big)$$
where fi represents the operations of the i-th layer (typically convolution followed by LeakyReLU) and [·] denotes concatenation along the channel dimension. This architecture enables the network to extract and process both local and global features at multiple scales, which is essential for reconstructing fine details in artwork restoration.
- ESRGAN Upscaler: The ESRGAN Upscaler module provides the optional capability of high-resolution output to the ArtResGAN framework. It contains revolving convolution layers, and employs 23 RRDB blocks that operate in lower resolution, and apply PixelShuffle operations to re-introduce spatial resolution while maintaining the relationship of the features. The ESRGAN upscaler can be formulated as:
$$G_{SR}(x) = H\Big(F_{up}\big(F_{RRDB}(F_{in}(x))\big)\Big)$$
Where Fin represents the initial convolutional layer, FRRDB is the sequence of 23 RRDB blocks, Fup denotes the upsampling layers using PixelShuffle, and H is the final convolution layer.
The PixelShuffle upsampling operation can be expressed as:
$$PS(F)_{c,(r\times h)+i,(r\times w)+j} = F_{(r^2\times c)+(r\times i)+j,h,w}$$
where r is the upscaling factor, F is the feature map before upsampling, and PS(F) is the upsampled feature map. A final convolution layer produces the restored high-resolution output. This architecture, adapted from Enhanced Super-Resolution Generative Adversarial Network (ESRGAN), has been constructed specifically to generate high-fidelity

detail when the resolution of restored artwork is enlarged. With this module, our framework provides the ability to achieve museum-quality restorations from heavily degraded input data for museum and conservation purposes.

II. Discriminator (D)

The PatchGAN discriminator is a central part of our ArtResGAN framework that is designed to evaluate the patch-level quality of an image instead of globally. Our discriminator is different than traditional discriminators in bottom-up Generative Adversarial Networks (GANs), which provide a single scalar quality score for a whole image. Our PatchGAN architecture outputs a 2D matrix of predictions, which provide a probability for each image patch to be real instead of generated.

The discriminator input is a 6-channel image that is made up of the concatenation of either the degraded and restored images (fake pairs) or the degraded and target ground truth images (real pairs). This pairing ensures that the discriminator can provide some context about the restoration task, and ignores arbitrary differences that have nothing to do with a real improvement.

The patch-based nature of this discriminator brings all kinds of benefits for artwork restoration. First, it nudges the generator to consider local textures/brushstrokes—important elements of style—in outputting restorations. Second, it helps keep most of the details of characteristics of the original artwork by penalizing locally unrealistic restorations. Third, the patch discriminator offers more granular feedback while training and treats sub-meaningful patch pairs independently, increasing the signal per image. Finally, the relatively small receptive field means it is more efficient to train than fully-connected counterparts and forwards faster iterations while building the model.

While engaging in adversarial training, the discriminator is updated to maximize its ability to distinguish between real and fake patch pairs while the generator is learning to output restorations that the discriminator classified as real. In this way, the adversarial nature of the discriminator, alongside our specific loss functions, helps to prompt the generator to output visually convincing and stylistically faithful artwork restorations.

III. VGG-based Style Extractor

A VGG-based style extractor is incorporated in the work to preserve the artistic qualities of the demised images:

- Adversarial Loss: The adversarial loss incentivizes improvements to the generator output by that whenever the discriminator is able to tell that the output is a fake, the generator is penalized. The discriminator tries to minimize its own loss by correctly identifying the real and fake samples. This drives the generator to improve to create outputs that are more and more realistic and ultimately to the point that the discriminator is not able to tell them apart. In image restoration tasks, the adversarial loss helps the generator to produce images that have clearer details and natural textures (realism) beyond that of pixel-wise losses.
  The Loss function for the generator is:
  $$\mathcal{L}_{adv} = -\mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(G(\mathbf{x}))]$$
  Where G(x) is the output of the generator and D(G(x)) is the output of the discriminator.
  The loss function for the discriminator is:
  $$\mathcal{L}_D = -\mathbb{E}_{\mathbf{y}\sim p_{data}(\mathbf{y})}[\log D(\mathbf{y})] - \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log(1 - D(G(\mathbf{x})))]$$
  Where x is the damaged input image, y is the ground truth and G(x) is the generated output.
- Content Loss: Content Loss utilizes the representational capability of a pretrained convolutional neural network, typically VGG-19, which has been trained on large datasets (e.g. ImageNet). Instead of comparing the images directly in pixel space, we instead compare the activations (feature maps) at specific layers in VGG-19. The feature maps contain high-level semantic information like the object's shapes, texture and structure.
  $$\mathcal{L}_{content} = \|\phi_l(G(\mathbf{x})) - \phi_l(\mathbf{y})\|_1$$
  Where y is the ground truth, G(x) is the restored image and ϕl(y) is the feature extractor at layer l.
  This loss measures how similar the two images are in feature space, focusing on maintaining the structure and identity of the content rather than just exact pixel values.
- Style Loss: Style loss is a form of perceptual loss function used in image restoration and style transfer tasks to ensure that the textural and stylistic patterns of the image generated closely follow the textural and stylistic patterns of the original undamaged image or artwork. Style loss is different from content loss, which considers the structure and spatial arrangement of features, whereas style loss considers the appearance, texture, and artistic feel of the image via feature correlations. To compute the style loss, both the generated and reference images are passed through a pretrained VGG-19 model to extract feature maps from its various layers. Rather than directly comparing the feature maps between the generated and reference images, the style loss computes the Gram matrix of each feature map, which captures the relationships and co-occurrences of different feature channels. The style loss is defined as:
  $$\mathcal{L}_{style} = \sum_l \|G(\phi_l(G(\mathbf{x}))) - G(\phi_l(\mathbf{y}))\|_F^2$$
  where G(·) denotes the Gram matrix, and ‖·‖F is the Frobenius norm.
  By minimizing style loss, the model learns to reproduce the visual textures, brushstrokes, and color patterns of the original image.

IV. Machine Vision Integration

ArtResGAN integrates several machine-vision techniques to improve restoration quality by refining image textures, edges, and structural details. These techniques include:

- Edge Detection: The Sobel and Canny operators are employed in the model to enhance the visibility of major edges and structural boundaries of the artwork. These methods ensure that minor details, such as brush strokes, outlines, and object shapes, are accurately recreated.
- Morphological Operations: Dilation, erosion, opening, and closing are employed to remove noise, refine fine details, and provide the degraded texture with coherence. Instance clearer restoration of images, especially for those that exhibit rich artificiary.
- Local Binary Patterns (LBP): It captures and maps micro-textures within images while ensuring that variations in tone and strokes remain well portrayed and well restored.
- Harris Corner Detection: Used to highlight the primary structural elements and dominant attributes of the artwork, while fine details of the visual are conserved.
- Haar Wavelet Transform: With this technique, multi-resolution analysis is conducted by separating images into distinct frequency components, allowing for a modeling focus on finer details and wider textures.

The final combination of these machine-vision techniques increases restoration reliability and fine detail preservation, confirming it as a robust tool for the restoration of degraded or damaged artwork while maintaining its artistic integrity.

V. RESULTS AND DISCUSSION

The results of the experimental evaluation illustrate the effectiveness of the ArtResGAN model in restoring paintings with complicated degradation types. From the loss values documented during the first three training epochs, it is clear that the generator loss dropped from 6.6300 in the first epoch to 2.4063 in the third epoch; this shows that the model is learning over time and has been successfully improving its image generation performance. Furthermore, the discriminator loss has dropped from 0.6420 to just 0.0046, which suggests that the generator has effectively learned to fool the discriminator and distorting the images to look more realistic in the restoration process. The qualitative results further reinforce these findings. The comparison between the original, damaged, and restored images illustrates that the model effectively reconstructs missing or distorted regions while preserving both structural and stylistic features. Notably, the style loss is minimal, as evident from the consistent brushstroke texture, color palette, and composition retained in the restored image. This suggests that the model has successfully captured and reproduced the artistic style of the original image. Moreover, the content loss is also low, reflecting the accurate reconstruction of semantic features such as shapes and spatial arrangements of the floral elements.

In summary, the model's ability to finesse tradeoffs between adversarial, content, and style objectives has resulted in seemingly plausible restorations of an image's final form. These findings confirm the model's ability to act as a useful source for art restoration tasks, particularly in situations of relevance to fine texture and semantic fidelity.

FIGURE 2. RESULT OF THE ARTRESGAN MODEL. (L-R) ORIGINAL UNDAMAGED IMAGE; DAMAGED IMAGE; DIGITALLY RESTORED IMAGE BY ARTRESGAN

FIGURE 3. RESULT OF THE ARTRESGAN MODEL. (TOP-BOTTOM) ORIGINAL UNDAMAGED IMAGE; DAMAGED IMAGE; DIGITALLY RESTORED IMAGE BY ARTRESGAN

TABLE 1. FINAL LOSS VALUES OF THE DISCRIMINATOR AND GENERATOR

Epoch	Discriminator Loss	Generator Loss
0	0.6420	6.6300
1	0.0071	2.5938
2	0.0046	2.4063

VI. CONCLUSION

In this study, we proposed ArtResGAN, a comprehensive deep learning framework designed for high-fidelity image restoration and stylistic preservation of degraded artworks. By integrating a hybrid U-Net + ResNet generator with a PatchGAN discriminator, the model successfully combines global contextual awareness with local texture refinement. Additionally, the inclusion of classical machine vision features and a VGG-based perceptual loss framework ensures that both structural details and artistic styles are faithfully reconstructed. Experimental training results indicate that the model is capable of efficiently learning from damaged inputs, as evidenced by the steadily decreasing generator loss and the adversarial balance maintained with the discriminator. The model achieves compelling restorations early in the training cycle, which speaks to its data efficiency, robust architecture, and strong generalization capacity even with limited epochs. Moreover, the use of residual-in-residual dense blocks in the ESRGAN upscaler allows for high-resolution outputs with fine texture realism, further reinforcing the superiority of the framework. Overall, ArtResGAN not only outperforms traditional methods reliant on handcrafted priors but also offers a scalable and interpretable solution for real-world digital art conservation.

VII. FUTURE WORK

While ArtResGAN has demonstrated strong capabilities in restoring degraded artworks and preserving their stylistic essence, there are several promising directions for future research. One potential improvement is the integration of attention mechanisms within the generator to enable more focused restoration on highly damaged regions. Additionally, incorporating a multi-modal input strategy that includes metadata such as artist style, period, or known restoration references could further enhance contextual fidelity. Another area worth exploring is unsupervised or semi-supervised training, which would reduce dependence on large labeled datasets and allow the model to generalize across broader domains of degradation. Furthermore, extending the framework to handle animated or video-based restoration would open new possibilities in the conservation of digital heritage and film. Real-world deployment may also benefit from optimizing the model for edge devices or mobile applications, making it accessible to museums, curators, and conservators with limited computational resources. Lastly, integrating user-in-the-loop feedback could offer controllable restoration levels, allowing for customized and interactive artwork recovery that respects artistic intent and historical accuracy.

REFERENCES

[1] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, Lei Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising”

[2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, Dacheng Tao, “DehazeNet: An End-to-End System for Single Image Haze Removal”

[3] Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky, “Deep Image Prior”

[4] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, “Generative Image Inpainting with Contextual Attention”

[5] Liu Guilin, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, Bryan Catanzaro, “Image Inpainting for Irregular Holes Using Partial Convolutions”

[6] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, “Free-Form Image Inpainting with Gated Convolution”

[7] Mikael Henaff, Alfredo Canziani, Yann LeCun, “On the Relation Between the Sharpest Directions of GANs and the Spectral Norm of the Jacobian”

[8] Dong Yang, Jian Sun, “Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing”

[9] Sayan Nag, Anirban Chakraborty, Subhajit Sanyal, R. Venkatesh Babu, “Prior Guided GAN Based Semantic Inpainting”

[10] Ying Chen, Yuting Zhang, Jinqiao Wang, Hanqing Lu, “An Improved GAN-Based Approach for Image Inpainting”

[11] Chuangchuang Dong, Huaming Liu, Xiuyou Wang, Xuehui Bi, “Image Inpainting Method Based on AU-GAN”

[12] Xinyuan Feng, Ding Liu, Zhaowen Wang, Jianming Zhang, Zhe Lin, Zhisheng You, Xianbiao Shu, “Pluralistic Image Completion”

[13] Ziyu Wan, Xiaoming Li, Anran Liu, Fu Li, Yulun Zhang, Ming-Hsuan Yang, “HiFill: Hierarchical Feature Fusion for Image Inpainting”

[14] Yan Xia, Wenguan Wang, Jianbing Shen, Ming-Ming Cheng, “Deep Generative Prior for Image Restoration”

[15] Nafiz Al Asad, Md. Appel Mahmud Pranto, Shbiruzzaman Shiam, Musaddeq Mahmud Akand, Mohammad Abu Yousuf, Khondokar Fida Hasan, Mohammad Ali Moni, “MGAN-CRCM: A Novel Multiple Generative Adversarial Network and Coarse-to-Refined Contextual Memory Approach for Image Inpainting”