

LCRO-GAN: A Multi-stage GAN Framework for Image-Based Reconstruction of Historical Monuments

Pranav Jigalur(✉), Likhith Gowda, Kripa S Rai, Praveen Koppal, and Sindhu R Pai

Department of Computer Science and Engineering, PES University, Bengaluru, India
pranav.jigalur@gmail.com, sindhurpai@pes.edu

Abstract. The virtual reconstruction and preservation of historical monuments is a critical problem in digital archaeology, especially when cultural heritage images contain damaged or missing regions. Existing image inpainting methods often struggle to restore such regions with both photorealistic quality and structural fidelity, limiting their usefulness for heritage reconstruction. This paper presents an efficient AI-based approach for achieving structural consistency in the reconstruction of damaged monument images. Specifically, a sequential multi-stage generative adversarial network (GAN) framework is proposed to restore missing regions while maintaining architectural structure. Furthermore, a novel masking strategy, called Largest Contour-based Regional Occlusion (LCRO), is introduced to generate architecturally realistic damage patterns for training and evaluation. Accordingly, the proposed framework, termed LCRO-GAN, is designed as a robust three-stage GAN pipeline for structurally consistent monument image reconstruction. The pipeline performs coarse reconstruction in the first stage, followed by high-resolution inpainting to recover fine architectural details, and a final refinement stage to enhance visual consistency and generate seamless outputs. Experimental results show that LCRO-GAN achieves a PSNR of 28.7 dB, an SSIM of 0.827, and an FID score of 11.5, indicating the effectiveness of the multistage approach for photorealistic image restoration. The findings further demonstrate that the combination of structured masking and multi-stage adversarial learning improves both structural consistency and reconstruction accuracy. Overall, the proposed framework offers a scalable approach for digital heritage restoration and serves as a useful tool for monument preservation projects.

Keywords: Generative Adversarial Networks (GAN) · LCRO Masking (Largest Contour-based Regional Occlusion) · Multi-stage GAN Framework · Peak Signal-to-Noise Ratio (PSNR) · Fréchet Inception Distance (FID) · Structural Similarity Index (SSIM)

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 428–444, 2027.
https://doi.org/10.1007/978-3-032-30491-9_27

1 Introduction

Historic monuments showcase concrete connections to our cultural heritage and they represent centuries of architectural achievements, artistry, and historical significance [1,2]. Consequently, digital reconstruction and conservation have become increasingly significant in India, as many ancient temples, forts, and archaeological monuments continue to deteriorate due to natural disasters, environmental pollutants, and human interference [3,4]. In normal images, local texture patterns are adequate for realistic image inpainting; however, for architectural reconstruction, it is necessary to follow strict patterns of structure, symmetry, and cultural design motifs [5]. Conventional image inpainting methods often fail to deliver photorealistic reconstruction and preserve structural consistency.

Significant improvements in the performance of deep learning techniques have been reported in the field of image synthesis and reconstruction with the help of Generative Adversarial Networks (GANs) [6]. Single-image restoration solutions that comprise multi-scale attention mechanisms and deep image priors have also shown promising results [7,8]. However, the performance of single-stage GAN models seems to suffer from difficulties in preserving architectural details in the reconstructed images. Single-stage GAN models typically struggle to preserve fine-grained architectural details and often introduce artifacts when applied to complex structural imagery [9]. In addition to this, existing masking strategies are simplistic, thus failing to generate realistic deterioration patterns observed in damaged monuments, which can limit their effectiveness in real-world restoration approaches [10]. Masking plays a crucial role in image reconstruction and inpainting tasks, as it defines the regions of missing or corrupted information that the model must restore. These techniques lack structural and semantic awareness, removing regions without considering underlying spatial structure or architectural importance of the scene. Such indiscriminate masking strategies can obscure essential structural elements, thereby producing reconstructions that are visually inconsistent and architecturally unrealistic. Existing studies on occlusion handling in deep models also reveal that these simple masking strategies have several limitations on theory and real-world applications [11]. These limitations highlight the need for more structure-aware masking techniques, since the final restoration depends largely on the masking approach.

This research attempts to bridge the gap in the digital reconstruction of ancient monuments and sculptures by proposing *LCRO-GAN*, a novel multi-stage Generative Adversarial Network framework designed specifically for Dravidian-style monument reconstruction. In contrast to conventional inpainting methods that use generic free-form masking, the proposed framework incorporates a *Largest Contour-based Regional Occlusion (LCRO)* masking technique to generate realistic, context-aware damage patterns that more closely resemble natural architectural deterioration. The LCRO-GAN pipeline follows a progressive restoration strategy. The initial stage creates a coarse structural foundation, followed by intermediate texture refinement for structural consistency, and a final stage that enhances high-resolution details by coordinating the use of multiple generative models.

In order to test and validate the proposed framework, a dataset of more than 20,000 high-resolution images of Indian temples and monuments were collected. The dataset was compiled using images obtained from multiple open-access, royalty-free image repositories. The images were further refined through systematic preprocessing and augmentation. Experimental evaluation shows that the proposed multi-stage framework achieves competitive PSNR, SSIM, and FID scores, highlighting its capability to produce photorealistic and structurally consistent restorations. Overall, the proposed approach provides a robust and scalable solution for the digital preservation of historical monuments and establishes an effective data-driven framework for cultural heritage reconstruction.

The remainder of this paper is structured as follows: Sect. 2 reviews related work on image inpainting and GAN-based reconstruction approaches. Section 3 describes the process of dataset creation. The detailed description of the proposed LCRO masking strategy is discussed in Sect. 4. The LCRO-GAN architecture and training methodology are presented in Sect. 5. Experimental results and comparative analysis are covered in Sect. 6. Finally, Sect. 7 concludes the paper and outlines directions for future research.

2 Related Work

The researches related to the domain of image processing and inpainting using generative AI has seen significant shift from standalone conditional GANs towards advanced multi-stage pipelines that decomposes the image reconstruction process into manageable pieces [6]. This improvement serves the requirements of architectural restoration, where the global structure, local texture, and semantic consistency needs to be maintained simultaneously. Techniques that combines pre-processing or multimodal strategies can assist in enhancing restoration and analysis [12].

Pix2Pix conditional GAN model has been used widely for image-to-image reconstruction tasks, especially for creating binary damage masks and guided inpainting. Rakhimol and Uma Maheswari demonstrated that through the application of conditional GANs for automatic mask generation and partial convolution networks for temple murals restoration, improving PSNR, MSE, and SSIM values can be made by greater extents [13].

The variants of DCGAN are simple models for texture synthesis in image restoration tasks. Recently, the application of attention and structural priors to DCGAN architecture has been done to enhance the restoration model's capacity to handle geometric complexity of architectural structures.

CycleGAN has also been utilized for style transfer and colorization in heritage restoration, enabling unpaired learning scenarios where the ground truth of the restored images is not available [14]. This is beneficial for the Indian temple restoration task because of the scarcity of information about the structures before damage.

Devalla and Arya proposed a GAN architecture using gated convolution and contextual attention layers that were trained on Indian temple and sculpture

images. Their method obtained an L1 loss of 0.29 after prolonged training, showing improved free-form missing region handling [15].

For temple murals of Tamil Nadu, Sumathi and Uma Devi proposed an edge- and line-guided diffusion patch GAN (WSFN + SCN + DIMD) that could restore large degraded regions. Their approach recorded stunning metrics: SSIM 0.8853, PSNR ∼29.88, MSE 0.0021, and LPIPS 0.0426 on proprietary degraded mural datasets.

Multiple works have blended modified U-Net damage detection with style transfer for priors in texture and PartialConv inpainting to produce perceptually coherent completions on Rajasthani painting images [16].

Previous research was dedicated to automatic mask generation blended with exemplar-based region-filling algorithms for both structure and texture reconstruction in Indian temple murals, with assessment using PSNR, MSE, and SSIM compared to conventional techniques [10].

Triple GAN frameworks have more recently been suggested whereby there are distinct networks dealing with varying aspects of restoration (structure, texture, refinement), and each step utilizes the output of the preceding step. Several works employed coarse-to-refinement pipelines where a coarse network produces crude completions, which are then fine-tuned by expert networks dedicated to texture and detail refinement. Sophisticated pipelines distinguish structural reconstruction from content creation, applying structure-specific networks to ensure architectural consistency while content networks address texture synthesis. The researchers have employed multiple discriminators operating at different scales (global + local/patch), effectively addressing both the macro-structural and micro-textural features in detail, to make training more stable and realistic [1].

The literature details significant advancements in GAN-based in-painting of Indian temples using techniques like structure-content decomposition, multi-discriminator architectures, and gated convolutions. However, there are still significant issues with standardization, expert verification, practical implementation, and cross-cultural transfer. To ascertain historical authenticity, future research will need expert-guided evaluation standards and high-quality labeled datasets from Indian temples.

3 Dataset Creation

This section describes the methodology followed for building the dataset used in this research, including data collection, preprocessing, deduplication, and measures taken to enhance data quality and diversity.

3.1 Data Acquisition

Data Sourcing and Retrieval. The dataset employed in this paper was assembled by the authors using web scraping and API-based retrieval, followed by manual curation from publicly available, copyright-free image repositories, including Pexels, Unsplash, and Pixabay, that support royalty-free usage for academic

research [17,18]. Libraries like `requests` and `BeautifulSoup` has been used to navigate web pages and identify image URLs. To obtain a broad range of images, keywords for specific temple names, common terms like "Dravidian style", "Indian temple architecture" and certain kingdom names like "The Vijayanagara Kingdom", "The Rastrakutas" has been used. This method gathered a diverse set of images exhibiting the diversity of Indian cultural heritage.

Initial Data Preprocessing. The image URLs were gathered and stored in a traceable JSON format. The subsequent phase included downloading and verification of each of the images. A python script was created for downloading the images, in which multithreading was implemented for network latency management as well as increased efficiency. Each image that was downloaded went through two validation steps:

- *Automated Checks:* The integrity of the files was checked by scripts to ensure that all the images were correct and maintained in their original files. This also optimized the image format in order to be successfully integrated into the training process.
- *Manual Filtering:* The manual screening was done to eliminate unnecessary or poor images. The images of modern buildings, irrelevant objects, low-resolution images that did not contribute to the learning process in the model were eliminated. This step was necessary to make the dataset of quality and integrity.

3.2 Dataset Refinement and Augmentation

Deduplication. To reduce the risk of model overfitting and potential dataset bias, a thorough deduplication process has been carried out before training. This has been developed through a hash-based file comparison technique to detect and eliminate similar images, regardless of file name or origin. Automated methods may not always capture near-duplicates or slight variations. Therefore, the dataset was also reviewed manually to identify and remove such cases. After completing this cleaning process, the dataset consisted of 19,998 distinct images.

Data Augmentation. Even though the amount of original images was also considerable, the diversity of the architecture styles was in the nature of the dataset which made it difficult to train an efficient model. This issue was addressed by significantly expanding the dataset through data augmentation techniques, which not only increased the dataset size but also enhanced the model's ability to generalize. These augmentations created modified versions of the existing images, artificially increasing the dataset size six-fold to a total of 119,936 images. The types of augmentations applied are:

- *Geometric Transformations:* Horizontal and vertical flips and rotations of 90°, 180°, and 270° has been performed. This brought in different perspectives and orientation to the model in such a way that it would be more resistant to different orientations.

– *Photometric Adjustments:* Images' brightness and contrast were changed slightly (within a ±10% range). This mimicked a range of illumination conditions and ensured the model that was trained was more robust against variations in environmental conditions.

Prior to feeding the images into our model, one final preprocessing step had to be performed. All images were resized to a standard resolution of 500×500 pixels. This was important for computational efficiency and consistent model training. Free-form masks were applied to the images subsequently. This is simulating damage and forming the "broken" images which our model has been trained to fix. The free-form topology of the masks avoids overfitting the model to the shape of damage and requires it to learn to complete missing information from context. All images used in this work were obtained from open-access platforms that allow free use for research and academic purposes. No copyrighted or restricted images were included in the dataset. The final curated dataset was generated solely for experimental evaluation of the proposed LCRO-GAN framework.

4 Data Preprocessing

The performance of a deep learning model, in this case, a Generative Adversarial Network (GAN) for an image-to-image translation task such as architectural reconstruction, rests to a large extent on the realism and quality of its training dataset. Towards this end, having a paired dataset of images of undamaged, original monuments and their synthetically damaged counterparts is crucial. Although there are many ways to create image masks, generic approaches are not adequate for domain specific architectural reconstruction.

4.1 Rationale for a Custom Masking Strategy: Limitations of Conventional Methods

Throughout the image inpainting domain in general, there are a number of masking techniques that are widely utilized. They can be generally classified as:

– *Geometric Masks:* These are simple, regular shapes like rectangles or circles used to mask out portions of an image. Although easy to produce, they are structure-naive and fail to model realistic patterns of damage on intricate structures such as monuments (Fig. 1a).
– *Irregular (Free-form) Masks:* These techniques, typically involving random walks or "brush stroke" generation, form masks with irregular edges. While they produce more detail than geometric masks, they are nonetheless image content agnostic. The occlusions they produce are random and most unlikely to be a whole architectural element (e.g., a dome, a spire, or the whole facade) (Fig. 1b).
– *Noise-Based Masks:* Methods such as introducing random pixel noise (e.g., salt-and-pepper) emulate corruption to the signal instead of the global physical deterioration typical of deteriorating monuments (Fig. 1c).

The major drawback of such traditional methods is that they are semantically oblivious. The structural deterioration of monuments is seldom random; it is usually in large, connected areas that relate to specific architectural features. A fallen roof, a ruined wall, or a weathered facade represents a region-based and structurally meaningful loss of information.

A model learned from images of easily dissectible rectangular holes might be very good at inpainting textures and lines but would probably not convincingly reconstruct complicated, non-rectilinear features having distinctive architectural grammar. The mask itself needs to tell the model what kind of damage it is supposed to fix. Hence, to learn a model for authentic architectural restoration, a damage simulation strategy is needed that emulates damage in semantically and structurally meaningful ways. A new unsupervised machine learning solution is specifically designed for this task.



(a) Geometric Mask (b) Free-form Mask (c) Noise-based Mask

Fig. 1. Comparison of conventional masking techniques.

4.2 The Proposed Technique: Largest Contour-Based Regional Occlusion (LCRO)

To alleviate the above-mentioned drawbacks, the Largest Contour-based Regional Occlusion (LCRO) technique has been introduced. The main goal of LCRO is to detect and completely mask the most salient object in an image—presumed to be the monument—thus mimicking extensive, realistic structure damage instead of superficial, haphazard holes. This unsupervised method utilizes a pipeline of conventional computer vision operations. The result of this process is shown in Fig. 2.

Edge Detection for Structural Boundary Identification. The process begins by converting the input BGR image (I) into a single-channel grayscale representation, I_{gray}. This reduces computational complexity and narrows the analysis to luminance gradients, which are the major indicators of structural form. After this, the Canny edge detection algorithm is applied. This is a critical step for distinguishing the monument's silhouette and primary structural

lines from the background and internal textures. The thresholds of the algorithm, $\theta_{\text{low}} = 50$ and $\theta_{\text{high}} = 150$, were selected empirically to strongly map these boundaries over our dataset. The outcome is a binary edge map, E.

Dominant Object Segmentation via Contour Analysis. With the use of the edge map E, a set C containing all unique outer boundaries will be obtained, with every boundary c forming the boundary of a shape detected.

$$C = \{c_1, c_2, ..., c_n\} = \text{findContours}(E) \tag{1}$$

The LCRO algorithm uses the central heuristics for segmentation. It is supposed that, in a well-composed photograph of a monument, the visual salient subject is the monument. Thus, it will be framed by the largest area with the profile. This maximum contour, $c_{\max}$, was found by enumerating the set C and choosing the one with largest area.

$$c_{\max} = \underset{c \in C}{\text{argmax}}(\text{Area}(c)) \tag{2}$$

This simple intuitive method allows us to do sound, unsupervised foreground segmentation that is application-specific without the need to use computationally expensive deep learning segmentation models.

Mask Generation and Synthetic Occlusion. The main contour $c_{\max}$ has been split off to form a binary mask M. This is accomplished by initializing a zero-(black) image and drawing in white (pixel value 255) the filled contour over it. Thus a neat clean mask will be obtained that covers the principal architectural building. And lastly, This mask will be used to produce the "damaged" image, I_{damaged}. The mask M will be utilized as an index, replace every pixel in the original color image I which is in the region of the mask with a block white color $[255, 255, 255]$. This is to mimic a total and absolute occlusion of the monument.

$$I_{\text{damaged}}(p) = \begin{cases} [255, 255, 255] & \text{if } M(p) = 255 \\ I(p) & \text{if } M(p) = 0 \end{cases} \tag{3}$$

4.3 Automated Pipeline for Dataset Generation

To generalise this process to our entire data of monument images, an automated batch processing pipeline was used. This script manipulates the entire pipeline: it loads raw images, executes the LCRO method on each of them, and saves outputs neatly. The pipeline produces:

1. The damaged image (I_{damaged}) with the white occlusion.
2. The binary mask (M) employed to generate the occlusion.

In addition, the pipeline constructs a necessary manifest file (`image_pairs.txt`), monitoring the file paths for every triplet (original, degraded, mask). Our data loader directly consumes this manifest file during GAN training, which guarantees a clean and reproducible pipeline from data preparation to training the model.



Fig. 2. LCRO masking.

5 Proposed Methodology

This section provides an overview of the LCRO-GAN framework, including the architecture of each component, the associated loss functions, the training methodology, and the evaluation metrics.

5.1 The LCRO-GAN Framework

The proposed methodology, LCRO-GAN, introduces a multi-stage generative adversarial network architecture designed to achieve high-fidelity image reconstruction of damaged historical structures. The overall architecture of the framework is illustrated in Fig. 3. The core of the architecture is a sequential, three-stage pipeline designed to address the inherent limitations of image inpainting. Image reconstruction, particularly for objects with intricate geometric structures and complex textures such as historical structures, involves addressing multiple interrelated sub-problems such as creating global structural consistency, reconstructing realistic and consistent textures, and restoring fine-grained, high-frequency features. To effectively address these problems, the framework decomposes the reconstruction process into multiple stages, where each stage addresses distinct challenges associated with restoration.

The pipeline begins with the first generator, denoted as G_1, which utilizes an efficient U-Net architecture based on Deep Convolutional GAN design. This stage generates a coarse but structurally coherent completion of the missing regions. The output from this stage is subsequently passed to a second generator in the sequence, G_2, which uses a CycleGAN-based architecture with an efficient ResNet architecture. This stage focuses on refining the coarse reconstruction by generating realistic textures and ensuring domain consistency. The final stage uses a third generator, G_3, based on the Pix2Pix conditional GAN architecture with an efficient ResNet design to enhance the refined image with sharper edges and high-frequency details, producing the final, photorealistic output. This progressive refinement procedure, from coarse global shapes to fine local details, can

be interpreted as a form of implicit curriculum learning. Rather than attempting to tackle the extremely challenging and ill-posed high-resolution inpainting problem in a single end-to-end operation, a process fraught with training instability and mode collapse, the framework instead steps through the learning process in multiple stages.

5.2 Stage 1: Coarse Structural Inpainting (G_1)

The first stage of the LCRO-GAN architecture is dedicated to create the foundational structure of the inpainted region. This is achieved by the generator G_1, which receives the masked input image I_{damaged} and produces a coarse reconstruction I_{coarse}. The primary objective of this stage is to provide a structurally realistic foundation for subsequent refinement steps. To achieve this, the design of G_1 is optimized for learning global, low-frequency information efficiently.

Generator Architecture (G_1). The generator G_1 is based on an encoder-decoder architecture inspired by Deep Convolutional GANs (DCGANs). The encoder decimates the input image with strided convolutions to obtain a low-dimensional latent vector without pooling. Leaky ReLU activation functions are applied to promote stable gradients and avoid the "dying ReLU" risk. The decoder then upsamples the latent vector using transposed convolutions to generate the inpainted image as output. It uses ReLU activations, with a final Tanh layer scaling pixel values to the range $[-1, 1]$. To improve training stability, Batch Normalization is applied following each convolution and transposed convolution layer (excluding the generator output and discriminator input). This normalization minimizes internal covariate shift and prevents issues such as vanishing or exploding gradients.

Loss Function (L_{G_1}). The generator G_1 is trained using a composite loss function with competing adversarial and reconstruction terms:

$$L_{G_1} = L_{\text{adv}}^{(1)} + \lambda_1 L_{L1}^{(1)} \tag{4}$$

where:

- L_{G_1}: Total loss of the Stage-1 generator (G_1).
- $L_{\text{adv}}^{(1)}$: Adversarial loss corresponding to the discriminator supervision for G_1.
- $L_{L1}^{(1)}$: L_1 reconstruction loss measuring the mean absolute error between generated and ground-truth images.
- λ_1: Weighting coefficient that controls the contribution of the L_1 loss term.

The loss function used to train the first-stage generator is defined in Eq. (4), which combines adversarial supervision with pixel-wise reconstruction loss. In Eq. (4), λ_1 balances the weight of the L1 reconstruction loss. The adversarial loss ($L_{\text{adv}}^{(1)}$) encourages G_1 to generate images that are indistinguishable from real ones, creating realism. The L1 loss ($L_{L1}^{(1)}$) computes the mean absolute pixel difference between the generated and corresponding ground-truth images. L1

loss is preferred over L2 loss because it produces sharper, less blurry outputs. At this point, the aim is to ensure structural accuracy rather than complete photorealism. Therefore, intricate losses such as perceptual loss are not included to ensure that the optimization process focuses on coarse geometric reconstruction.

5.3 Stage 2: Texture and Consistency Refinement (G_2)

The second stage of the framework focuses on texture synthesis and structural improvement. The generator G_2 takes the coarse structural map I_{coarse} from the previous stage as input and generates a refined image, I_{refined}, by generating contextually consistent textures within the reconstructed regions.

Generator Architecture (G_2). The generator G_2 adopts a CycleGAN-style architecture, which is optimized for image-to-image translation tasks. The core building block of this network is Residual Blocks (ResBlocks), each consisting of convolutional layers with skip connections that add the input of the block to its output. This residual learning framework preserves the low-level features learned from the encoder, supports deep network architectures, and facilitates efficient gradient propagation. Since Instance Normalization operates independently across each sample in the batch, G_2 uses it instead of Batch Normalization. This allows the model to preserve image-specific textures and styles. U-Net skip connections link the encoder and decoder layers to preserve global structural information and fine details from I_{coarse}, while reflection padding is used to minimize edge artifacts.

Loss Function (L_{G_2}). Optimization of G_2 is driven by an aggregate loss:

$$L_{G_2} = L_{\text{adv}}^{(2)} + \lambda_2 L_{L1}^{(2)} + \lambda_3 L_{\text{cycle}} \tag{5}$$

where:

- L_{G_2}: Total loss of the Stage-2 generator (G_2).
- $L_{\text{adv}}^{(2)}$: Adversarial loss for the second-stage generator.
- $L_{L1}^{(2)}$: Pixel-wise L_1 reconstruction loss.
- L_{cycle}: Cycle-consistency loss enforcing structural consistency between forward and inverse mappings.
- λ_2, λ_3: Weighting parameters controlling the influence of reconstruction and cycle-consistency losses, respectively.

The optimization objective for the second-stage generator is defined in Eq. (5), which incorporates adversarial, reconstruction, and cycle-consistency constraints. Here, λ_2 and λ_3 regulate the term balance. The adversarial and L1 reconstruction losses enforce realism and pixel-wise accuracy, similar to Stage 1. In addition, cycle-consistency loss (L_{cycle}) is introduced to ensure that an image translated from one domain to another and back remains consistent with the original image. This regularizes G_2, enforcing structural consistency and preventing the generation of unrealistic or inconsistent textures.

5.4 Stage 3: High-Fidelity Detail Enhancement (G_3)

The final stage of the LCRO-GAN pipeline includes the final refinement of the reconstructed image, aimed at the recovery of high-frequency details, sharp structural boundaries, and overall photorealism. In this stage, a more sophisticated generator, G_3, is utilized, based on the Pix2Pix conditional GAN architecture which takes as input the refined image I_{refined} and generates the final output I_{final}.

Generator Architecture (G_3). The generator G_3 uses a coarse-to-fine multi-scale framework to generate high-resolution images. The global generator initially produces a low-resolution output to create semantic structure, while local enhancer networks refine high-resolution feature. The outputs from all scales are combined to generate the final reconstruction. This design captures both global coherence and fine textures effectively. Hierarchical residual blocks further refine feature representations at each level, enabling stable training of deep, high-capacity models.

Loss Function (L_{G_3}). The overall objective for G_3 emphasizes perceptual quality and is defined as:

$$L_{G_3} = L_{\text{adv}}^{(3)} + \lambda_4 L_{L1}^{(3)} + \lambda_5 L_{\text{perceptual}} + \lambda_6 L_{\text{feature_matching}} \tag{6}$$

where:

- L_{G_3}: Total loss of the Stage-3 generator (G_3).
- $L_{\text{adv}}^{(3)}$: Adversarial loss for the final-stage generator.
- $L_{L1}^{(3)}$: Pixel-wise L_1 reconstruction loss.
- $L_{\text{perceptual}}$: Perceptual loss computed using feature representations extracted from a pretrained VGG-19 network.
- L_{FM}: Feature matching loss based on intermediate discriminator feature maps.
- $\lambda_4, \lambda_5, \lambda_6$: Weighting coefficients for reconstruction, perceptual, and feature matching losses, respectively.

Here in Eq. (6), λ_4, λ_5 and λ_6 balance the contributions of each loss component. The adversarial and $L1$ losses enforce realism and pixel fidelity, respectively, while the perceptual loss computed using a pre-trained VGG-19 network enforces semantic and textural accuracy. The feature matching loss aligns intermediate discriminator features for real and generated images, thereby stabilizing adversarial training and improving the visual realism of the generated images. This set of losses steers G_3 to produce images that are structurally accurate and, more importantly, indistinguishable from real ones.

5.5 Adversarial Supervision and Training Protocol

The effective coordination of the three separate generator modules is achieved through a unified discriminator and a multi-stage training protocol. This design

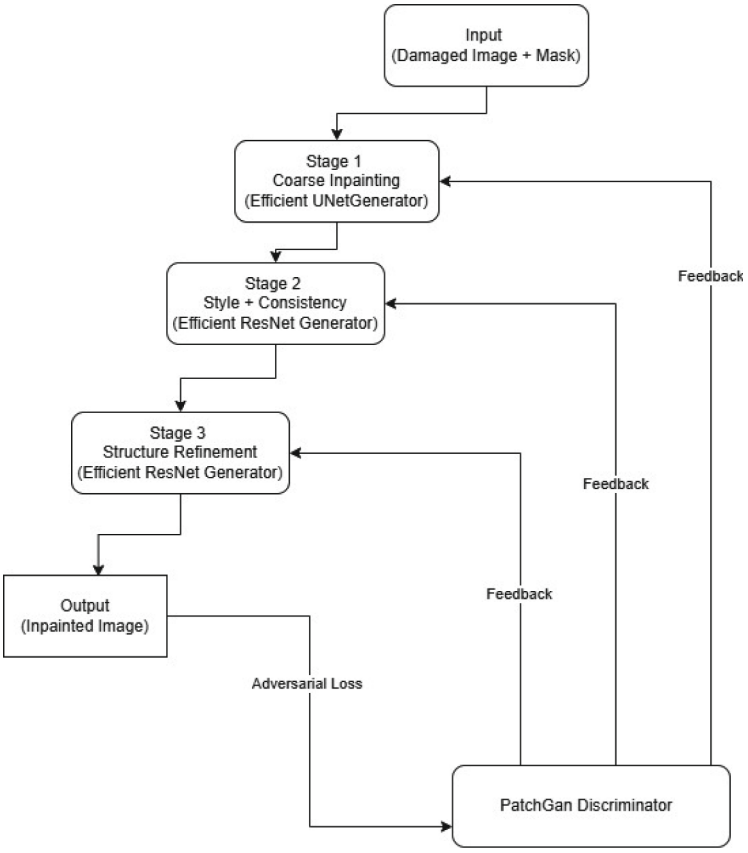

Fig. 3. High Level View of Proposed Methodology.

enables each stage to build upon the outputs of the preceding stage, resulting in a coherent, high-performance inpainting pipeline.

Unified Discriminator Design (D). A single discriminator network D provides adversarial supervision for all three generators. This discriminator uses the architecture of PatchGAN, which has been proven to be effective in various tasks such as image-to-image translation or image synthesis. Unlike a classic discriminator, which would return a single scalar value that classifies an entire image as real or fake, a PatchGAN is a fully convolutional network returning a 2D map of probabilities. Each value corresponds to the "realism" of an overlapping N $\times$ N patch of the input image (e.g., 70 $\times$ 70 patches). This patch-level assessment encourages the generator to produce realistic high-frequency details throughout the image. Moreover, this is more efficient than a full-image discriminator because it requires fewer parameters. A key characteristic of the discriminator is its ability to accept inputs at multiple scales, enabling it to effectively supervise

the outputs of all three generators, which may exhibit varying levels of detail or resolution.

Comprehensive Training Strategy. The model is optimized using a staged and joint optimization approach. Considering that the Adam optimizer assumes a learning rate that decreases in time, the training proceeds over four stages: (1) G_1 and the discriminator are trained for 100 epochs to produce coarse structural completions; (2) with G_1 frozen, G_2 is trained for 100 epochs to add realistic textures; (3) freezing both G_1 and G_2, G_3 is trained for 100 epochs to enhance fine details and high-resolution quality; and (4) Joint Fine-Tuning: all generators are unfrozen for an end-to-end training with a reduced learning rate. This hybrid training combines the stability of progressive learning and the comprehensive improvement of joint optimization. It enables earlier generators to modify their outputs to facilitate more seamless stage transitions, ultimately creating a final coherent and high-fidelity reconstruction.

5.6 Evaluation Metrics

Quantitative performance evaluation is conducted by aggregating a set of metrics capable of capturing pixel-level accuracy, perceptual realism, and distributional similarity in particular. PSNR measures pixel-wise reconstruction fidelity but often does not correlate well with human perception; SSIM tackles this aspect by comparing luminance, contrast, and structure between the generated and real images for a more perceptually correct measure. To further assess the perceptual aspect, LPIPS computes feature-space distances between deep network-processed generated and real images, leveraging VGG or AlexNet, demonstrating



(a) Masked Input (b) Restored Output

Fig. 4. Qualitative comparison of masked input images and corresponding reconstructions generated by the proposed LCRO-GAN framework.

strong correlation with human judgment. Finally, FID quantifies both the quality and diversity of generated images by comparing feature distributions of real and generated images using an InceptionV3 network, where lower scores indicate higher realism. Collectively, these metrics provide a comprehensive and rigorous evaluation of the model's reconstruction fidelity, perceptual quality, and generative performance.

6 Results and Discussion

The model was tested and trained on a high-performance computing environment. All tests were run in an NVIDIA Quadro RTX 5000 with 16 GB of its own VRAM as well as CUDA version 12.6. Every generator stage of the LCRO-GAN pipeline was trained with 100 epochs, G1 took about 15 minutes per epoch, G2 and G3 took about 30 minutes per epoch, as they are more complex to architecture. After the three-stage training, further 100 epochs of joint end to end fine-tuning were done again at a rate of 30 minutes per epoch to allow the training to become stable and the high-frequency photorealistic details and low-frequency structural priors to be co-learned on the entire model. The reconstruction results of the pipeline is illustrated in Fig. 4. This demonstrates the progression from coarse structural completion to high-fidelity refinement.

Table 1. Reconstruction Metrics Across LCRO-GAN Stages.

Stage	PSNR (dB)	SSIM	LPIPS	FID
Stage 1 – Coarse	24.8	0.712	0.073	48.2
Stage 2 – Refined	26.5	0.764	0.058	31.7
Stage 3 – Details	27.9	0.803	0.045	18.4
Final (Joint Fine-Tune)	**28.7**	**0.827**	**0.038**	**11.5**

Table 1 illustrates the reconstruction quality of the proposed multi-stage GAN model. The framework had been shown to be pixel-level accurate, with the final Joint Fine-Tune achieving a PSNR of 28.7 dB and an SSIM of 0.827. More importantly, the gradual training approach leads to extensive perceptual realism improvements. FID, which is a measure of the quality and distribution of images, dropped significantly by 76.1 percent since the starting coarse stage score of 48.2 scored lower in the last stage of 11.5. On that note, the LPIPS improved immensely and reduced to its current final score of 0.038, which is outstanding. These findings as a whole therefore suggest that the progressive optimization achieved the task of moving the framework out of producing a plausible structural completion into producing a high-fidelity, photorealistic reconstruction, thereby justifying the design of the framework.

7 Conclusion

The proposed LCRO-GAN model exhibits strong reconstruction ability and structural consistency of the dataset, which contains the images of various Dravidian-style temple architectures and sculptures. The model was able to reconstruct structurally coherent and realistic results from degraded inputs. Furthermore, the combination of a multi-stage GAN framework with the novel LCRO masking technique enables the model's ability to identify the overall architectural layout as well as the minute fine-grained details of temples and sculptures, thus demonstrating their effectiveness in the digital preservation and restoration of historical monuments.

Future work will focus on expanding the dataset to include wide variety of architectural styles across different regions and cultures, improving the model's generalization ability. Additionally, exploring context-aware image reconstruction techniques would eliminate the need of explicit damage masks. We aim to develop an end-to-end application with automated preprocessing, inference, visualization, and export tools, making the model usable by researchers, historians, and conservation experts for heritage restoration and digital preservation. These future advancements together position LCRO-GAN as a scalable foundation for next-generation digital heritage restoration systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Deng, X., Yu, Y.: Ancient mural inpainting via structure information guided two-branch model. Heritage Sci. **11**(1), 1–16 (2023). https://doi.org/10.1186/s40494-023-00972-x
2. Basu, A., et al.: Digital restoration of cultural heritage with data-driven computing: a survey. IEEE Access (2023). https://doi.org/10.1109/ACCESS.2023.3280639
3. Altaweel, M., Khelifi, A., Zafar, M.H.: Using generative AI for reconstructing cultural artifacts: examples using Roman coins. J. Comput. Appl. Archaeol. **7**(1), 301–315 (2024). https://doi.org/10.5334/jcaa.146
4. Gaber, J.A., Youssef, S.M., Fathalla, K.M.: The role of artificial intelligence and machine learning in preserving cultural heritage and art works via virtual restoration, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. X-1/W1-2023, pp. 185–192 (2023). https://doi.org/10.5194/isprs-annals-X-1-W1-2023-185-2023
5. Nogales, A., et al.: ARQGAN: an evaluation of generative adversarial network approaches for automatic virtual inpainting restoration of greek temples. Expert Syst. Appl. **180**, 115092 (2021). https://doi.org/10.1016/j.eswa.2021.115092
6. Guan, J., et al.: Progressive generative mural image restoration based on adversarial structure learning, NPJ Heritage Sci. **13**, 309 (2025). https://doi.org/10.1038/s40494-025-01883-9
7. Arumugam, C.D., Banothu, B.: Integrating multi-scale attention network with deep image prior for single image restoration. Int. J. Inf. Technol. **17**, 555–565 (2024)

8. Malagi, V.P., Naresh, E., Mithra, C., Suresh, B.V.N.V.K.: Adaptive residual convolutional neural network for moiré image restoration. Int. J. Inf. Technol. **17**(2), 755–765 (2024/2025)
9. RSINet: Inpainting remotely sensed images using triple GAN framework. In: Proceedings of the IEEE IGARSS (2022). https://doi.org/10.1109/igarss46834.2022.9884330
10. Mol, V.R., Maheswari, P.U.: The digital reconstruction of degraded ancient temple murals using dynamic mask generation and an extended exemplar-based region-filling algorithm. Heritage Sci. **9**(1), 1–18 (2021). https://doi.org/10.1186/S40494-021-00604-2
11. Chavan, S.A., Chaudhari, N.M., Pawar, U.B.: Mathematical analysis behind occlusion handling in image with deep learning. Int. J. Inf. Technol. **15**(7), 3727–3739 (2023)
12. Wajid, M.A.: A deep learning approach for image and text classification using neutrosophic fuzzy sets. Int. J. Inf. Technol. (2024)
13. Rakhimol, V., Maheswari, P.U.: Restoration of ancient temple murals using CGAN and PCONV networks. Comput. Graph. **109**, 83–92 (2022). https://doi.org/10.1016/j.cag.2022.11.001
14. Image restoration technology of Tang Dynasty tomb murals based on adversarial edge learning. Research Square (2023). https://doi.org/10.21203/rs.3.rs-2765366/v1
15. Devalla, V., Arya, A.: Temples restoration using gated convolution and contextual attention in generative adversarial networks. In: Proceedings of the IEEE International Conference on Smart Computing and Communication (ICSCCC) (2023). https://doi.org/10.1109/ICSCCC58608.2023.10176941
16. Singh, U., Maiti, S., Saini, A.K.: Ancient Indian murals digital restoration through image inpainting. IEEE SPIN (2023). https://doi.org/10.1109/SPIN57001.2023.10116111
17. Pexels, Free stock photos (2024). https://www.pexels.com
18. Pixabay, Royalty-free images (2025). https://pixabay.com