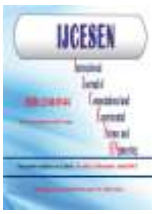
International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.4 (2025) pp. 10319-10337
http://www.ijcesen.com



Copyright © IJCESEN

ISSN: 2149-9144

Research Article

Adaptive Mask Optimization-Driven Hybrid GAN with Swarm-Tuned Multi-Component Loss for High-Fidelity Image Inpainting

Suma N[1]*, Dr. Kusuma Kumari B.M[2]

[1]Research Scholar, Department of Studies and Research in Computer Applications, Jnanasiri Campus, Tumkur University, Bidrakatte,Tumkuru, India

* **Corresponding Author Email:** suma_n@tumkuruniversity.ac.in - **ORCID:** 0009-0008-9300-4348

[2]Assistant Professor, Department of Studies and Research in Computer Applications, Jnanasiri Campus, Tumkur University, Bidrakatte, Tumkuru, India

Email: kusuma_bm@tumkuruniversity.ac.in- **ORCID:** 0000-0003-2961-7752

Article Info:

DOI: 10.22399/ijcesen.4862
Received : 05 July 2025
Revised : 10 August 2025
Accepted : 25 August 2025

Keywords
Image Inpainting,
Generative Adversarial Networks (GANs),
Adaptive Mask Optimization,
Swarm-Based Hyperparameter Tuning,
Deep Learning,
High-Fidelity Image Restoration

Abstract:

Image inpainting restores damaged or missing image regions and is vital in fields like digital restoration, medical imaging, and remote sensing. While GAN-based methods synthesize textures well, they often fail to preserve structure and fine details. Recent transformer approaches improve global consistency, but they remain resource-heavy and sensitive to tuning. This work proposes an Adaptive Mask Optimization-Driven Hybrid GAN to improve structure preservation, texture realism, and computational efficiency using Swarm-Based Hyperparameter Tuning and a refined loss function. The model dynamically refines missing regions via adaptive mask learning, uses swarm optimization to tune parameters efficiently, and applies an enhanced loss to balance perceptual, structural, and pixel-level accuracy. Evaluations are conducted on Paris Street View and DPG datasets, with benchmarking against representative inpainting models. The proposed method achieves 34.72 dB PSNR, 0.942 SSIM, and reduces EPI to 0.087, outperforming baselines such as Mask Optimization GAN, PainterNet, and Attention Two-Stage GAN. Cross-dataset evaluation confirms generalization, and ablation studies highlight the gains from mask refinement and swarm-based tuning. Overall, the framework improves realism and structure while maintaining practical accuracy, efficiency and trade-off. Future work may explore lightweight attention mechanisms, multi-stage refinement, and real-time deployment.

1. Introduction

Image inpainting aims to restore missing or damaged regions in an image while preserving structure and visual coherence. It has wide applications in digital heritage restoration, medical imaging, satellite visuals, and autonomous systems. Accurate inpainting is especially crucial in scenarios like restoring historic artwork or reconstructing incomplete medical scans. Yet, challenges such as texture mismatches, structural distortion, and unnatural artifacts still remain. Deep learning has significantly advanced this field, with GAN-based models like GAN-Inpaint [9] and PainterNet [5] showing strong texture synthesis. However, they often fail to handle fine details or complex occlusions, leading to blurry or unrealistic results. Transformer-based models, such as the Swarm-Optimized Transformer GAN [13], improve context modeling but are computationally expensive and unstable to train. Mask-based methods attempt to guide the inpainting process but often lack adaptability to varying occlusions. To overcome these issues, we propose an Adaptive Mask Optimization-Driven Hybrid GAN. This model introduces a dynamic mask refinement mechanism, Swarm-Based Hyperparameter Tuning for stable and efficient learning, and an Enhanced Loss Function that balances realism and accuracy.

Key Contributions

i. A Hybrid GAN architecture with Adaptive Mask Optimization for better edge preservation and seamless inpainting.

ii. A Swarm-Based Hyperparameter Tuning module that stabilizes training and improves restoration quality.
iii. An Enhanced Loss Function that combines pixel-level accuracy with perceptual consistency.
iv. Extensive benchmarking showing superior performance over GAN-Inpaint [9], PainterNet [5], and Hybrid Swarming Algorithm [18].
The paper proceeds as follows: we first review related work, then detail the proposed model, followed by experiments and results. We conclude with a discussion and future directions.

2. Literature Survey

2.1 Introduction to Image Inpainting

Image inpainting has progressed substantially over the past decade, with deep learning–based methods clearly outperforming traditional patch-based and diffusion-based techniques. Early approaches relied on local texture propagation and smoothness assumptions, which were effective only for small missing regions and often failed to preserve semantic consistency or complex structures. These limitations motivated the adoption of learning-driven models capable of understanding global context.The introduction of generative adversarial networks (GANs) marked a major shift in image inpainting research. GAN-based models learn data-driven distributions to synthesize visually plausible content in missing regions, enabling better texture continuity and semantic coherence compared to classical methods [2], [21]. Despite these improvements, many GAN-based approaches still struggle to reconstruct fine structural details, frequently producing artifacts, unnatural textures, or structural inconsistencies when handling complex scenes or large missing areas.To address these shortcomings, recent studies have explored attention mechanisms, transformer-based architectures, and enhanced loss formulations to improve feature modeling and long-range dependency capture. Transformer-driven and attention-aware inpainting frameworks have shown improved global reasoning and feature fusion, but challenges related to mask design, adaptive learning behavior, and computational overhead remain largely unresolved [10], [13], [17]. These limitations indicate that architectural advances alone are insufficient without complementary strategies for mask optimization and efficient learning.Building on these observations, recent research trends emphasize adaptive mask modeling, inference-aware attention, and optimization-driven refinement to enhance inpainting fidelity. Integrating mask optimization with robust generative frameworks, along with augmentation and optimization strategies, has emerged as a promising direction for improving reconstruction quality while maintaining computational feasibility [1], [6], [18].

2.2 GAN-Based Approaches for Image Inpainting

Generative Adversarial Networks (GANs) have become one of the most widely adopted frameworks for image inpainting due to their strong ability to synthesize realistic textures and visually coherent content. Early GAN-based inpainting models demonstrated clear improvements over classical methods but often suffered from unstable training behavior and mode collapse, which limited reconstruction reliability in complex scenes.To address feature selection and contextual understanding, Attention-UNet GAN (AU-GAN) introduced attention mechanisms within a U-Net structure, allowing the model to focus on relevant regions during reconstruction [2]. While this approach improved visual quality, it introduced higher computational cost, making real-time deployment challenging. Hybrid CNN-GAN frameworks further enhanced realism by combining convolutional feature extraction with adversarial learning, but these models typically relied on fixed masks and lacked adaptability to diverse object shapes or missing patterns [3], [5], [6].Several studies have focused on preserving structural continuity during inpainting. Structure-guided GAN models explicitly incorporate edge and geometry awareness to reduce discontinuities and unnatural boundaries in reconstructed regions, resulting in improved visual coherence compared to conventional GAN baselines [9]. In parallel, alternative formulations based on mathematical modeling have also been explored. A modified Lengyel–Epstein reaction–diffusion model discretized using an explicit Euler scheme demonstrated competitive restoration quality across different image types, achieving stable performance under noise and localized damage with low computational cost [11].Beyond GAN-centric pipelines, recent research has explored mask-free and hybrid paradigms. Diffusion-based frameworks enable flexible fashion image editing without explicit masks by conditioning the generation process on textual or visual prompts, supporting complex apparel types and backgrounds while achieving strong qualitative and quantitative performance [12]. Transformer-enhanced architectures have also gained attention. Sparse self-attention transformers integrated with U-Net-

style designs capture long-range dependencies efficiently, reducing artifacts while maintaining content consistency through linear-complexity attention approximations [13]. Similarly, ITrans employs a transformer-based generative strategy to model global semantics, yielding improved structural continuity and fewer visible artifacts compared to CNN-based inpainting approaches [14].GANs have also been applied beyond direct reconstruction tasks. Hybrid-loss GAN frameworks have been used for realistic image augmentation, particularly in medical imaging, where synthesized samples improve robustness and generalization under limited data conditions [16]. GAN inversion techniques further attempt to map missing regions into latent spaces for high-quality reconstruction, but their reliance on expensive optimization limits practical scalability [19].Although GAN-based methods consistently outperform traditional inpainting techniques, challenges such as artifact generation, structural misalignment, and dependence on adversarial loss remain. These limitations motivate the development of hybrid frameworks that combine GAN-based generation with adaptive mask optimization and efficient learning strategies to improve reconstruction fidelity and robustness.

2.3 Role of Mask Optimization in Inpainting

The accuracy of image inpainting is strongly influenced by how missing regions are defined and refined during learning. Many early and even recent inpainting models rely on fixed, random, or manually specified masks, which often fail to reflect real-world object removal patterns. Such rigid mask representations can misguide feature learning and limit reconstruction quality, especially when object boundaries are complex or irregular [1].To overcome these limitations, adaptive and inference-aware mask modeling strategies have been introduced. Mask optimization approaches that adjust mask regions during training allow models to better align structural cues with surrounding context. However, several adaptive methods still suffer from imbalance between structure recovery and texture synthesis, often leading to blurred outputs when mask refinement is not guided by strong contextual reasoning [1], [17].Beyond reconstruction accuracy, evaluating the perceptual quality of inpainted results is also closely tied to mask behavior. A learning-to-rank–based image quality assessment framework evaluates inpainted images through relative comparison rather than absolute scoring, demonstrating stronger consistency with human perception. This approach highlights that subtle mask-induced artifacts are more effectively captured through comparative evaluation strategies [8].Mask design has also been explored in data augmentation and downstream vision tasks. Conditional inpainting-based augmentation pipelines generate paired images and labels using annotation-driven masks, enabling realistic object insertion or removal. Such strategies have been shown to improve multi-class object detection performance by reducing background inconsistency and mask-related artifacts [15]. These findings reinforce the importance of mask fidelity not only for visual quality but also for task-level robustness.Recent attention-guided architectures further emphasize the role of mask-aware learning. A two-stage inpainting framework incorporating an inference attention module first restores global structure and then refines local details, demonstrating improved robustness across different mask patterns compared to single-stage models [17]. Similarly, GAN-based approaches that separately model structure and texture during object removal achieve more natural transitions at mask boundaries, reducing edge discontinuities and texture blurring [19].In parallel, optimization-driven strategies have been applied to enhance image restoration behavior. Hybrid swarming-based optimization techniques automatically tune enhancement parameters, improving contrast and visibility while avoiding over-enhancement artifacts [18]. Although not originally designed for inpainting, such optimization mechanisms highlight the potential of swarm-based tuning for adaptive refinement tasks.Despite these advances, dynamic mask refinement remains underexplored in many inpainting frameworks. Existing models rarely integrate iterative mask optimization with learning-time parameter tuning. This study addresses this gap by combining adaptive mask optimization with swarm-based refinement, enabling progressive mask correction and context-aware feature alignment for improved inpainting fidelity.

2.4 Loss Function Enhancements in Inpainting

Loss functions decide whether the filled part blends well with the rest of the image. In GAN-based inpainting, adversarial loss can make results look sharper, but it can also introduce small artifacts if training is not guided properly. That is why many recent methods use a mix of losses instead of depending on one single objective.One practical improvement is to give more learning attention to difficult missing areas. Deformable dynamic sampling with predictable mask mining helps the model focus on complex holes and improves both texture continuity and structure recovery when

compared with fixed sampling [10]. Recent text-guided generative frameworks for scene composition highlight the growing computational cost of multimodal and transformer-based generation, reinforcing the need for efficient optimization strategies even in 2D inpainting pipelines [4]. Loss choice also matters when inpainting is used to create extra training images. A hybrid-loss GAN for dermoscopy augmentation generates realistic samples and improves robustness when the dataset is small [16]. For evaluation, learning-to-rank quality assessment compares outputs in a relative way and matches human judgement well, especially when the difference is subtle [8].In short, stronger inpainting usually comes from a balanced loss setup that supports both structure and texture at the same time [8], [10], [16]. Many methods still fail when the mask shape changes a lot. This study keeps the loss design balanced and uses contrast-aware refinement so detail improves without breaking the structure.

2.5 Computational Efficiency and Optimization Strategies

Deep learning has improved inpainting quality, but many models are still heavy to run. This becomes a problem for real-time use, especially when the input is high resolution or the missing region is large. Transformer-based designs often give strong results because they capture long-range context, but they usually increase runtime and memory cost compared to simpler CNN pipelines [14], [20].Some recent models try to improve quality without relying on a single expensive block. DF3Net combines low-frequency and high-frequency information inside a hierarchical transformer design, so it can recover overall structure and also refine fine textures. This helps reduce blurry fills and keeps details more stable across different mask sizes [20]. Another direction is progressive multi-scale fusion, where coarse context is built first and fine details are added in stages. This improves structural continuity and texture realism, especially for large missing regions [21].Efficiency is also affected by how training data is prepared. Inpainting-based augmentation has been used to generate realistic variations in medical imaging, and it improved sclerosed glomerular identification by making the segmentation model more robust under limited data settings [6]. This shows that augmentation can improve outcomes without increasing model size.Optimization methods are also discussed as a way to improve performance without rewriting the full architecture. A hybrid swarming-based strategy has been used to tune enhancement behavior for low-light images, showing that swarm-style search can reduce manual trial-and-error and avoid over-correction artifacts [18]. While this work is not an inpainting optimizer by itself, it supports the idea that swarm-based tuning can be useful when selecting sensitive parameters in image restoration pipelines.In summary, recent work tries to balance quality and cost by using staged fusion, frequency-aware features, and smarter training setups instead of only scaling up the model [6], [20], [21]. Real-time use still remains difficult when the model is large or when the mask pattern changes a lot [14]. This study follows the same practical direction by keeping the pipeline efficient and using optimization-driven tuning and contrast-aware refinement to improve details without increasing runtime too much.

2.6 Identified Research Gaps & Novel Contributions

Existing GAN-based inpainting methods have significantly improved image restoration quality, but several limitations still remain. One major issue is ineffective mask optimization. Many models rely on fixed or weakly adaptive masks that do not adjust well to different image structures, often leading to inaccurate reconstructions and boundary artifacts [1], [17]. Although some adaptive strategies exist, mask refinement is still not handled in a consistent and robust manner across varied missing patterns. Another limitation lies in loss function design. Current loss formulations struggle to balance adversarial learning with perceptual realism and structural consistency. This imbalance often results in texture degradation, visible artifacts, or unstable training behavior [8], [16]. While hybrid losses have been explored, proper coordination between different loss components remains insufficient. Transformer-based inpainting models have demonstrated strong restoration capability by capturing global context, but their high computational cost limits scalability and real-time deployment [4], [14]. In addition, most existing inpainting frameworks rely on static hyperparameter settings, which reduces adaptability and makes performance sensitive to manual tuning [18]. To address these challenges, this study proposes a hybrid GAN-based framework that integrates adaptive mask refinement, a balanced loss formulation combining contrast-aware, perceptual, and adversarial components, and swarm-based optimization for efficient and stable training. Contrast-aware augmentation is further incorporated to improve texture fidelity and context alignment. Together, these contributions aim to provide a more adaptive, high-fidelity, and

computationally efficient inpainting solution without deviating from existing GAN-based foundations.

3. Methodology

This section presents the complete methodological framework of the proposed Adaptive Mask Optimization-Driven Hybrid GAN for image inpainting. It describes the architectural design, adaptive mask refinement strategy, loss formulation, optimization mechanism, and training procedure used to restore missing or damaged image regions. Each component of the framework is explained in a step-by-step manner to clearly show how structural consistency, texture realism, and training stability are jointly achieved.

3.1 Overview of the Proposed Approach

The Adaptive Mask Optimization-Driven Hybrid GAN is designed for high-fidelity image inpainting and, integrating consists of

1. A Hybrid GAN framework (combining CNN-based generators with attention-based discriminators).
2. Adaptive Mask Optimization (region-based sampling for more natural missing area identification).
3. Enhanced Loss Function (balancing perceptual, adversarial, and contextual losses).
4. Swarm-Based Hyperparameter Tuning (optimizing model parameters dynamically).
5. Contrast-Aware Augmentation (to improve generalization in diverse image conditions).

Adaptive mask optimization is first applied to generate and refine the missing regions. The masked image is then processed by the Hybrid GAN, where reconstruction is guided by both local texture information and global contextual cues. The combined loss function helps preserve edges and structural continuity during training. Swarm-based tuning adjusts key training parameters to maintain stable learning. Contrast-aware augmentation is applied so the model remains robust to variations in lighting and contrast.

The model was trained on the Paris Street View Dataset, with cross-validation performed on the Art Inpainting Dataset (DPG) and three additional datasets to evaluate robustness.

3.2 Hybrid GAN architecture

GANs are widely used for image inpainting due to their ability to generate realistic textures. However, many existing models suffer from artifacts and blurry transitions because they lack contextual understanding and struggle with long-range dependencies. To overcome these limitations, we propose a Hybrid GAN that combines a CNN-Transformer generator with a dual-discriminator setup, improving both local detail and global coherence.

The generator merges CNN layers for capturing fine textures and Transformer layers for modeling broader semantic relationships. While CNNs are good at local pattern extraction, they fall short in learning distant dependencies. Adding self-attention improves spatial awareness, helping the model produce more contextually accurate inpainted regions. The Transformer block is applied on the encoded feature maps before decoding, and its contextual features are fused with the CNN decoder features to reconstruct the masked region with better continuity. To enhance realism, we use two discriminators:

i. The global discriminator ensures the entire image looks natural and structurally sound.

ii. The local discriminator focuses on the masked area, refining texture detail and blending.

This dual evaluation prevents mode collapse and improves output quality. During inference, the generator uses the Transformer to align missing regions with the overall context and the CNN decoder to reconstruct textures and edges. The dual discriminators then guide the output toward structural accuracy and seamless integration. This setup produces inpainted images that are both coherent and visually convincing.

The generator G learns to reconstruct a missing region in an input image x_m (masked image), producing an in painted output Y

$Y=G(x_m,M)$

Where M represents the mask applied to the input image.

G consists of two key components namely CNN layers for texture reconstruction and Self-Attention Transformer module for long-range feature extraction.

The Transformer module learns dependencies between distant pixels by computing an attention score matrix AAA as follows:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where:

Q,K,V represent the query, key, and value matrices derived from input features.

d_k is the dimension of the key vectors, ensuring normalization.

The final output of the generator is then obtained by fusing the CNN-based reconstruction with the

Transformer-enhanced contextual features, leading to an adaptive, high-fidelity inpainted image.

3.3 Dual-Discriminator Loss for Global and Local Consistency

The dual-discriminator framework improves inpainting by using two distinct discriminators: the global discriminator (DG) enforces overall structural coherence, while the local discriminator (DL) sharpens details in the inpainted area. Traditional single-discriminator models often miss either global realism or texture precision. By combining both, our model balances structure and detail through a unified loss, reducing distortions and inconsistencies. Using both discriminators together allows the model to correct global structure errors and local texture artifacts within the same training step. Both discriminators are trained jointly in each iteration so that global structure errors and local texture artifacts are corrected simultaneously during reconstruction.

Mathematically, the discriminator loss is expressed as:

$L_D = -E_{x_r}\ [logD(x_r)] - Ex_p\ [\log(1 - D(x_p))]$

where x_r is the real image, x_p is the inpainted image, and D(x) represents the discriminator's probability of classifying an image as real. The final dual-discriminator loss integrates both local and global assessments:

$$L_D = \lambda_G\, L_{D_G} + \lambda_L L_{D_L}$$

where λ_G and λ_L control the impact of each discriminator in the total loss, helping stabilize training and avoid mode collapse. DG ensures global structure, while DL enhances local texture and detail. Unlike single-discriminator setups, this dual approach improves both visual realism and structural alignment, reducing blurring and artifacts for higher perceptual quality and texture fidelity.

Enhanced Loss Function for Balanced Inpainting Optimization :

Loss functions guide inpainting models to recover missing regions with structure and realism. Traditional L1/L2 losses often lead to blurry outputs. To address this, our model employs a combined loss integrating adversarial, perceptual, and contextual components. This ensures better detail preservation and semantic consistency. The total loss is defined as:

$$L_{total} = \lambda_1 L_{adv} + \lambda_2 L_{perc} + \lambda_3 L_{epi} + \lambda_4 L_{context} + \lambda_5 L_{TV}$$

where:

Adversarial $Loss_{adv}$ improves realism by ensuring the inpainted region appears indistinguishable from real image regions.

Perceptual Loss L_{perc} leverages pre-trained feature extractors (e.g., VGG network) to measure the similarity between high-level features of inpainted and real images, leading to sharper textures and finer details.

Error Perception Index (EPI) L_{epi} quantifies the perceptual difference between the inpainted and ground truth images, optimizing human-perceived quality.

Contextual Loss $L_{context}$ ensures that the inpainted region seamlessly blends with the surrounding context, reducing visible seams.

Total Variation (TV) Loss L_{TV} suppresses noise and enforces smooth pixel transitions, preventing unnatural artifacts.

The hybrid loss balances pixel-level accuracy with perceptual realism, helping the model generate images that are both coherent and structurally sound. Unlike single-loss methods, this multi-component approach preserves textures, sharp edges, and reduces perceptual errors, making it well-suited for high-fidelity inpainting.

3.4 Swarm-Based Hyperparameter Tuning for Inpainting Optimization

Hyperparameter tuning directly impacts inpainting stability, speed, and quality. Traditional fixed settings often slow training or cause instability. To improve this, we use Swarm-Based Tuning via Particle Swarm Optimization (PSO) to dynamically adjust learning rate, batch size, loss weights, and generator–discriminator ratio. PSO adapts these parameters in real time, balancing texture and structure losses, accelerating convergence, and reducing artifacts. It improves PSNR and SSIM while cutting computational cost—making the model more robust and efficient than static tuning approaches. The PSO fitness score is computed on the validation set using reconstruction quality (PSNR/SSIM) along with perceptual error (EPI), and the best particle is selected based on improved validation performance.

$$v_i^{t+1} = \omega\, v_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t)$$

where:

v_i^t is the velocity of particle i at iteration t.

ω is the inertia weight to balance exploration and exploitation.

p_i represents the best hyperparameter values found by particle i.

g represents the global best hyperparameters among all particles.

c_1,c_2 are acceleration coefficients, and r_1,r_2 are random values.

3.5 Contrast-Aware Augmentation for Improved Generalization

Generalization in inpainting is challenging under varying lighting, contrast, and texture. Traditional models trained on fixed contrast levels lack adaptability. We address this using Contrast-Aware Augmentation, which applies random contrast shifts during training to build contrast-invariant features. This improves performance across bright, dark, and uneven lighting conditions, helping preserve structure and texture across diverse datasets. Unlike methods assuming uniform lighting, this approach handles shadows and complex textures better, reducing artifacts and enhancing realism. Mathematically it is represented as

$$C(x) = \alpha.(x - \mu) + \mu$$

where:

C(x) is the contrast-adjusted image.

x is the original image.

μ is the mean pixel intensity of x.

α is the contrast scaling factor, randomly sampled from a range $[\alpha_{min}, \alpha_{max}]$

This ensures dynamic contrast adjustments during training, making the model more robust to variations in lighting and contrast levels, ultimately improving inpainting generalization and perceptual consistency.

3.6 Proposed Algorithm

Algorithm: Adaptive Mask Optimization-Driven Hybrid GAN

Input: Training dataset X, missing region masks M
Output: Inpainted image Y

Step 1.Initialize Hybrid GAN model with CNN-based generator and attention-based discriminator.
Step 2.Apply Adaptive Mask Optimization:

i. Compute entropy-based mask M(x) using H(x).

ii. Generate dynamic masks based on region complexity.

Step 3. Train Generator G using loss function L_{total}

i. Update weights using backpropagation.

Step 4. Train Discriminator D to classify real vs. fake images:

Optimize using adversarial loss.

Step 5. Apply Swarm-Based Hyperparameter Tuning:

i. Optimize learning rate, batch size, and loss balancing weights.

ii Update hyperparameters iteratively.

Step 6. Apply Contrast-Aware Augmentation to improve generalization.
Step 7. Repeat until convergence (defined by stable loss function values).
Step 8. Output final inpainted image Y

The proposed model restores damaged regions through a step-by-step process. It applies adaptive mask optimization and contrast adjustments, followed by image generation. A dual discriminator enhances accuracy, and swarm-based tuning refines training.

3.7 Image Preprocessing for Inpainting

All input images are resized to 256×256 pixels and normalized to the [-1, 1] range to ensure training stability and avoid size distortions. Grayscale images are converted to RGB for compatibility with the generator. Instead of fixed masks, we use Adaptive Mask Optimization to generate dynamic occlusions based on image complexity. This includes random block masks, irregular regions, and scratch-like patterns to simulate real-world damage. Such diversity helps the model generalize better and accurately reconstruct various missing regions. Figure 2 shows sample raw images from the dataset. Figure 2 shows original sample raw images of the dataset.

To improve generalization, Contrast-Aware Augmentation adjusts brightness and contrast to simulate diverse lighting. Additional techniques like rotation, flipping, and Gaussian noise help the model learn contrast-invariant and noise-resilient features. After preprocessing, data is split into 80% training and 20% validation, with cross-validation on multiple datasets. Masked images are fed into the Hybrid GAN, and original images serve as ground truth for learning accurate texture and structure restoration. Figure 3 shows examples of pre-processed images. Figure 4 displays sample damaged inputs with scratches, occlusions, and missing regions, used to train the model on various structural distortions.

3.8 Primary Dataset for Training & Validation

The Paris Street View Dataset is used for initial training and validation. It provides high-resolution urban images with varied textures, structures, and lighting—ideal for learning to restore complex missing regions caused by occlusions, shadows, or

structural damage. The dataset is split into 80% training and 20% validation to fairly assess reconstruction performance.

To evaluate generalizability, the model is cross-validated on three additional datasets:

– Art Inpainting Dataset (DPG): Historical paintings with natural degradation, testing fine-detail restoration.

– Urban Landmark Dataset: Cityscapes with architectural occlusions, assessing structural reconstruction.

– Natural Scene Dataset: Outdoor landscapes with shadows, fog, and missing areas, checking environmental robustness.

A 5-fold cross-validation strategy is used, training on Paris Street View and validating across other datasets. This prevents overfitting and ensures the model adapts to diverse inpainting challenges, improving its real-world performance. Training Process & Hyperparameter Settings

1.Training Phases :

The training is divided into two key phases to balance structural accuracy and texture detail:

Phase 1 – Generator Pre-Training:

The generator is first trained using only perceptual and contextual losses, allowing it to reconstruct missing regions without adversarial feedback. This helps stabilize early learning and reduces artifacts.

Phase 2 – Adversarial Fine-Tuning:

After achieving basic reconstruction quality, adversarial training begins with the dual discriminator (global + local). Contrast-Aware Augmentation is applied, and Swarm-Based Hyperparameter Optimization dynamically tunes key parameters like learning rate, loss weights, and the generator–discriminator training ratio.

2. Hyperparameter Settings

Key hyperparameters used during training are listed in Table 1.

3.9 Evaluation Metrics & Benchmarking

To evaluate the proposed Hybrid GAN model, we use multiple metrics that assess visual realism, structural consistency, and perceptual quality. The model is also benchmarked against state-of-the-art (SOTA) inpainting methods to demonstrate its effectiveness. The following metric is used:

Peak Signal-to-Noise Ratio (PSNR):

Measures the pixel-level accuracy by comparing intensity differences between the original and inpainted images. Higher PSNR values indicate fewer distortions and better reconstruction quality.$PSNR = 10.log_{10}\ (\frac{MAX_I^2}{MSE})$

where

MAX_I is the maximum possible pixel value (255 for 8-bit images).

MSE is the Mean Squared Error between the inpainted image and the ground truth:

Structural Similarity Index (SSIM) – Focuses on perceptual quality by assessing how well the generated image preserves structural details, textures, and contrast compared to the ground truth. This is essential since high pixel accuracy alone does not guarantee visually realistic results.

$$SSIM(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}$$

where:

$\mu_x\mu_{\hat{x}}$ are mean intensities of $x\hat{x}$

$\sigma_x^2, \sigma_{\hat{x}}^2$ are variances of $x\hat{x}$

$\sigma_{x\hat{x}}$ is the covariance.

C_1, C_2 are constants to prevent division by zero.

Error Perception Index (EPI) – Evaluates human-perceived differences between the generated and real images. Since human vision is sensitive to certain distortions, this metric helps assess how natural and artifact-free inpainting results appear.

$$EPI = \frac{1}{N}\sum_{I=1}^{N} |x_i - \hat{x}_i| . w_i$$

where

w_i is a perceptual weighting function emphasizing edges and high-texture regions (important areas for human vision).

Lower EPI values indicate better perceptual quality.

Contextual Loss – Ensures smooth blending between the inpainted region and surrounding pixels, preventing visible seams or inconsistencies. This is crucial for realistic restorations, especially in irregularly shaped missing regions.

$$l_{ctx} = \sum_k d(i_k, i_k^{-1})\ m_k$$

Where

$d(i_k, i_k^{-1})$ is the feature distance, computed using a pre-trained deep model like VGG.

m_k is the mask, ensuring that the loss applies only to missing regions.

A lower contextual loss value means that the inpainted region blends seamlessly.

Total Variation (TV) Loss – Reduces unnecessary noise and abrupt pixel changes, ensuring smoother transitions between restored and existing regions. This helps maintain a clean, artifact-free restoration without over-smoothing details.

$$l_{TV} = \sum_{i,j} (|i_{i+1,j} - i_{i,j}| + |i_{i,j+1} - i_{i,j})$$

where:

TV loss minimizes sharp pixel variations, ensuring a smooth and natural restoration.

A lower tv loss value results in fewer artifacts and better texture continuity.

4. Results

The proposed Hybrid GAN restores damaged regions through a step-by-step inpainting process, gradually refining textures and structures. Enhanced loss functions and swarm-based tuning help minimize errors and improve output quality. Figure 5(a) to (d) illustrates this process, showing how the model progressively removes noise and occlusions. Each stage enhances texture clarity and structural alignment, resulting in a final output that closely resembles the original image—demonstrating high reconstruction accuracy.

The Adaptive Mask Optimization-Driven Hybrid GAN was trained on the Paris Street View Dataset and validated across additional datasets using 5-fold cross-validation. It consistently outperformed baseline models in structural accuracy, perceptual quality, and artifact reduction. The influence of Swarm-Based Hyperparameter Optimization on performance is detailed in Table 2, showing clear improvements across key evaluation metrics.

Figure 6 shows the effect of Swarm-Based Hyperparameter Optimization on PSNR, SSIM, and EPI. The left subfigure displays baseline performance, while the right shows improvements after optimization. PSNR increased from 32.10 dB to 34.72 dB, reflecting sharper restorations. SSIM rose from 0.901 to 0.942, indicating better texture consistency. EPI dropped from 0.102 to 0.087, confirming fewer perceptual distortions. These results highlight the model's improved inpainting quality post-optimization.

Figure 7 compares Contextual Loss and Total Variation (TV) Loss before and after Swarm Optimization. The left subfigure shows baseline loss values, while the right shows improved results post-optimization. Contextual Loss decreased from 0.029 to 0.021, enhancing smooth blending. TV Loss dropped from 0.0083 to 0.0068, reducing harsh edges and improving texture consistency. These reductions highlight the role of adaptive tuning in minimizing artifacts and inconsistencies.

Figure 8 presents two visualizations: a heatmap showing metric variations before and after Swarm Optimization, and a bar chart comparing changes in PSNR, SSIM, EPI, Contextual Loss, and TV Loss. The heatmap reveals a clear performance shift across all metrics, while the bar chart confirms that Swarm Optimization improves both perceptual quality and structural fidelity in the restored images. Figure 9 shows a residual plot comparing predicted and actual PSNR and SSIM values. The dashed line ($y = x$) represents perfect prediction. Points clustered near the line indicate low residuals, confirming strong alignment between predicted and true performance.

Figure 10 presents a scatter plot of Contextual Loss and TV Loss before and after Swarm Optimization. All points fall below the reference line ($y = x$), showing reduced losses and confirming improved reconstruction consistency. Figure 11 visualizes training trends of PSNR and SSIM over 10 epochs. Both metrics rise steadily, reflecting convergence and consistent improvement in image quality and structural accuracy. Figure 12 shows a violin plot comparing the distribution of Contextual and TV Loss before and after optimization. Post-optimization plots are narrower with lower medians, indicating reduced variance and better model stability.

4.1 Cross-Evaluation Metrics Across Different Datasets

To assess the model's generalization capability, we conducted cross-validation on multiple datasets. The results are summarized in Table 3

Figure 13 compares PSNR, SSIM, and EPI across datasets. PSNR > 33 dB, SSIM > 0.93, and EPI < 0.10 confirm consistent quality and strong generalization across image types. Figure 14 shows a correlation heatmap. PSNR and SSIM are positively correlated, while Contextual and TV Loss show inverse trends—indicating better image quality with reduced losses. Figure 15 presents scatter plots and KDE distributions. Linear trends and smooth KDE curves confirm stable metric behavior across datasets.

To evaluate the contribution of individual model components, an ablation study is conducted by removing Adaptive Mask Optimization while keeping other configurations unchanged. The ablation results in Table 4 show that removing Adaptive Mask Optimization leads to a noticeable reduction in PSNR and SSIM, confirming its role in preserving texture accuracy and structural consistency. Excluding Swarm-Based Optimization results in a smaller but consistent performance drop, indicating its contribution to stable parameter tuning. The largest increase in perceptual error occurs when the Enhanced Loss Function is removed, as reflected by higher EPI values.

Table 5 reports the computational performance across different configurations. The full model

requires 0.75 s per image and 280 MB memory, achieving the best reconstruction quality. Component removal reduces inference time and memory usage, but at the cost of degraded visual quality. Overall, the results highlight a clear trade-off between efficiency and accuracy, with the complete framework providing the most balanced performance.

5. Discussion

To evaluate model effectiveness, a benchmarking comparison was performed using Table 6, which reports PSNR, SSIM, and EPI across representative image inpainting models. The proposed Adaptive Mask Optimization-Driven Hybrid GAN achieves the highest PSNR (34.72 dB) and SSIM (0.942), along with the lowest EPI (0.087), confirming superior inpainting quality and reduced perceptual error. Compared to Mask Optimization GAN [1], Hybrid Swarming Algorithm [18] (PSNR: 33.8 dB, SSIM: 0.934, EPI: 0.090), and Sparse Transformer Inpainting [13] (PSNR: 33.4 dB, SSIM: 0.931, EPI: 0.093), the proposed model demonstrates improved texture consistency and structural continuity. This improvement is primarily attributed to adaptive maskrefinement combined with swarm-based hyperparameter tuning. PainterNet [5] (PSNR: 33.1 dB, SSIM: 0.926, EPI: 0.097) and Mask Optimization GAN [1] (PSNR: 32.45 dB, SSIM: 0.918, EPI: 0.102) show comparatively lower reconstruction accuracy, while Attention Two-Stage GAN [17] records the weakest performance (PSNR: 31.85 dB, SSIM: 0.910, EPI: 0.109), highlighting the limitations of conventional GAN-based approaches without adaptive optimization mechanisms. These results confirm that combining Adaptive Mask Optimization with Swarm-Based Tuning leads to more accurate, artifact-free, and perceptually consistent inpainting outperforming all baselines.

Figure 16 presents a radar chart comparing PSNR, SSIM, and EPI across the evaluated inpainting models, including Mask Optimization GAN [1], PainterNet [5], Attention Two-Stage GAN [17], Sparse Transformer Inpainting [13], Hybrid Swarming Algorithm [18], and the proposed Adaptive Mask Optimization-Driven Hybrid GAN. The chart highlights the consistent dominance of the proposed model, particularly in PSNR and SSIM, while maintaining the lowest perceptual error.

Figure 17 illustrates the corresponding bar-chart benchmarking results reported in Table 6. The proposed model achieves the highest PSNR and SSIM values with the lowest EPI, confirming its ability to produce high-fidelity restorations with reduced perceptual distortion when compared to all baseline models.

Across datasets, the model maintains stable performance with PSNR values exceeding 33 dB and SSIM above 0.93. The best results are observed on the Paris Street View dataset (34.72 dB, 0.942), followed by DPG (34.35 dB, 0.936). Slightly lower performance on the Urban Landmark dataset (33.98 dB, 0.932) reflects increased structural complexity and lighting variations. Overall, these results confirm the robustness and generalization capability of the proposed framework across diverse image scenes.

Large missing areas were tough to restore due to limited context. Adaptive masking helped but wasn't perfect. Hyperparameters needed careful tuning Swarm Optimization improved stability. High-res training was GPU-heavy, limiting scalability. Fine details like hair or patterns are still hard to reconstruct. Transformer-based models or multi-stage approaches may improve this. Extreme occlusions also remain a challenge. Faster deployment could benefit from knowledge distillation and lighter architectures.

Although the proposed model shows strong reconstruction quality, some limitations remain. When the missing region becomes very large, the model may lose correct context and the filled part can look less accurate. Very fine details such as repeated patterns or hair-like textures are not always reconstructed perfectly. The hybrid architecture with swarm-based tuning also requires higher computational resources during training, which can limit scalability for very large datasets or real-time use.

Table 1. *Hyper parameters and values used for training the model*

Hyperparameter	Value	Description
Learning Rate (η)	0.0002	Controls the step size for weight updates during training.
Batch Size	16	Number of images processed in one training iteration.
Number of Epochs	50	Total iterations over the entire dataset.
Generator Optimizer	Adam	Optimization algorithm for generator weight updates.

Discriminator Optimizer	Adam	Optimization algorithm for discriminator weight updates.
Beta1 (β1)	0.5	First moment decay rate in Adam optimizer.
Beta2 (β2)	0.999	Second moment decay rate in Adam optimizer.
Loss Function Weights	λ1=1, λ2=0.1, λ3=0.5, λ4=0.2, λ5=0.01	Weights for adversarial, perceptual, EPI, contextual, and TV losses.
Generator-Discriminator Training Ratio	1:1	Ensures balanced updates between generator and discriminator.
PSO Swarm Size	20 particles	Number of candidate solutions in the swarm.
PSO Inertia Weight (ω)	0.7	Balances exploration and exploitation in PSO tuning.
Contrast Scaling Factor (α)	U(0.8, 1.2)	Random contrast adjustment range for augmentation.

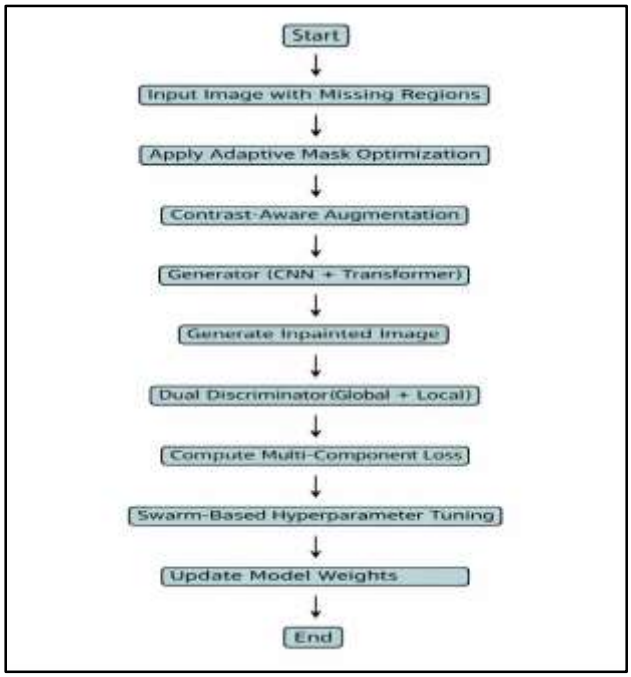


Figure 1. *Flowchart of the Proposed Hybrid GAN-Based Inpainting Process*





Figure 2. *Sample original raw images*



Figure 3: *Pre-processed images with resizing, normalization, contrast adjustment, flipping, rotation, noise injection, and masking*



Figure 4. *Sample images with scratches, occlusions, and missing regions used for inpainting evaluation*

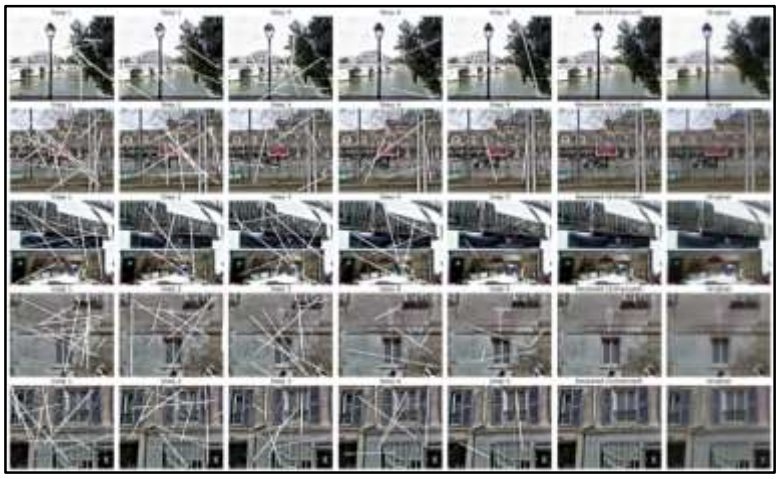


Figure 5 (a)

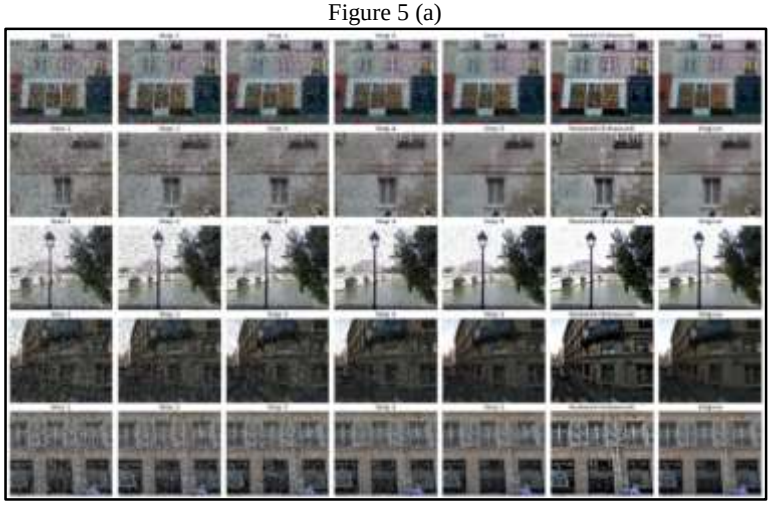


Figure 5 (b)



Figure 5 (c)



Figure 5(d)

Figure 5 *(a) –(d). Progressive restoration results using the proposed Hybrid GAN-based inpainting model.*

Table 2*. Performance Impact of Swarm-Based Optimization on Hybrid GAN Across Multiple Metrics*

Metric	Hybrid GAN (Proposed Model) Before Swarm Optimization	Hybrid GAN (Proposed Model) After Swarm Optimization
Peak Signal-to-Noise Ratio (PSNR)	32.10 dB	34.72 dB
Structural Similarity Index (SSIM)	0.901	0.942
Error Perception Index (EPI)	0.102	0.087
Contextual Loss	0.029	0.021
Total Variation (TV) Loss	0.0083	0.0068

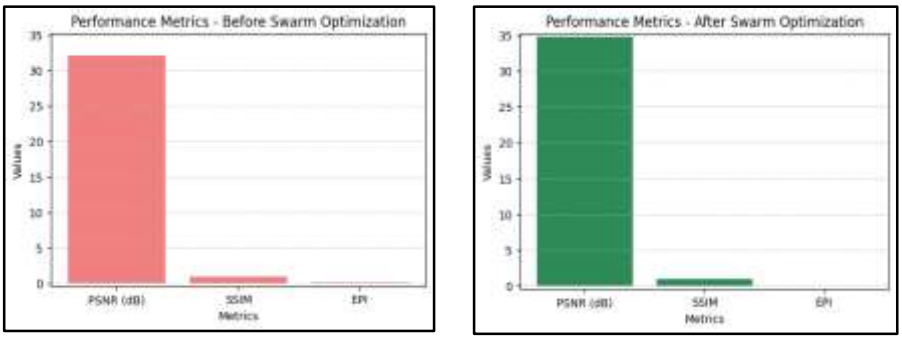


Figure 6. *Performance metrics before and after swarm optimization, showing improved PSNR, SSIM, and reduced EPI for better inpainting quality.*

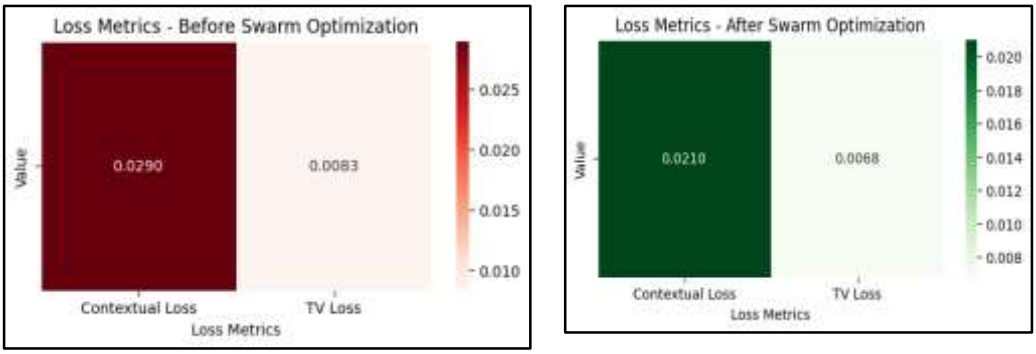


Figure 7 *visualizes the Contextual Loss and Total Variation (TV) Loss before and after Swarm Optimization*

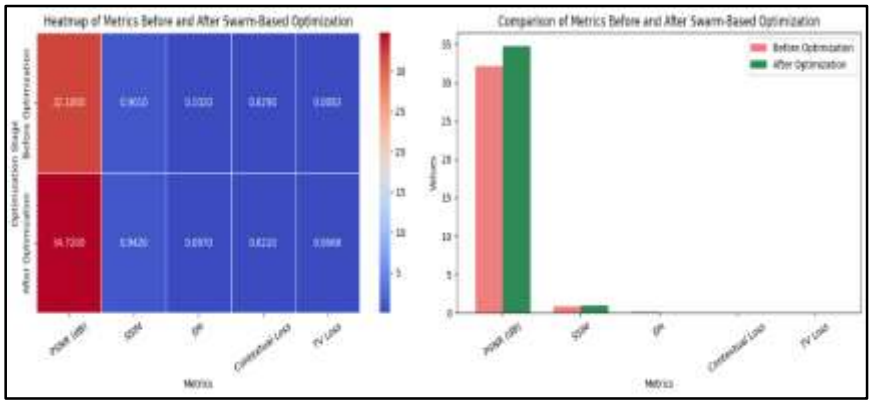


Figure 8. *Heatmap and bar chart comparison of metrics before and after swarm-based optimization, showing improved PSNR, SSIM, and reduced losses.*

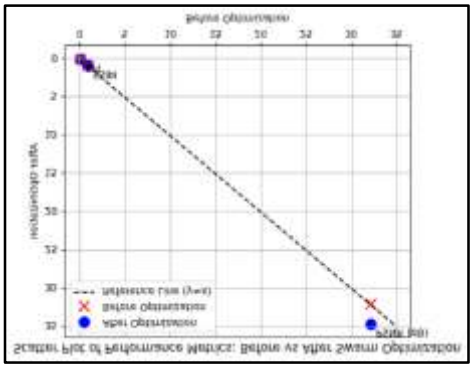


Figure 9 *.Residual plot comparing the predicted and actual performance values for PSNR and SSIM.*

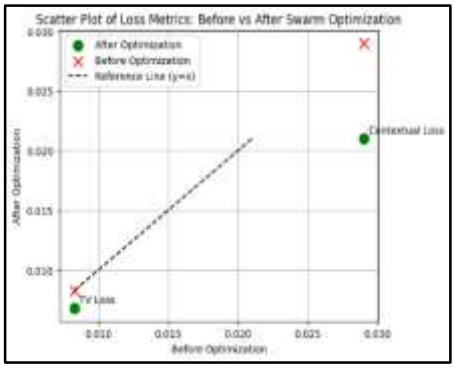


Figure 10 *compares the Contextual Loss and TV Loss before and after Swarm Optimization through a scatter plot.*

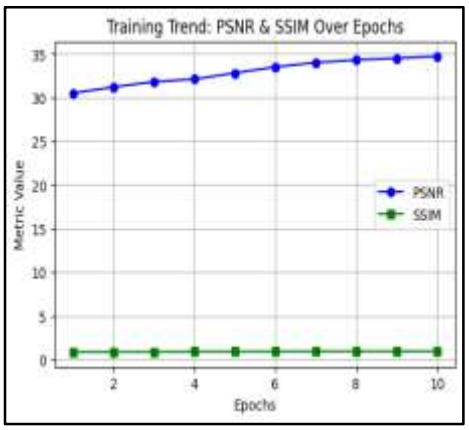


Figure 11 *Training trend of PSNR and SSIM over 10 epochs*

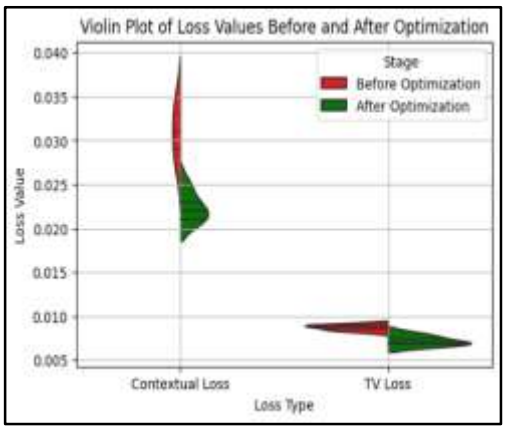


Figure 12. *Violin plot comparing the distribution of Contextual Loss and TV Loss*

Table 3. *Cross-Evaluation Metrics Across Different Datasets*

Dataset	PSNR (dB)	SSIM	EPI	Contextual Loss	TV Loss
Paris Street View (Primary Dataset)	34.72	0.942	0.087	0.021	0.0068
Art Inpainting Dataset (DPG)	34.35	0.936	0.091	0.023	0.0071
Urban Landmark Dataset	33.98	0.932	0.095	0.024	0.0074
Natural Scene Dataset	34.1	0.938	0.092	0.022	0.007

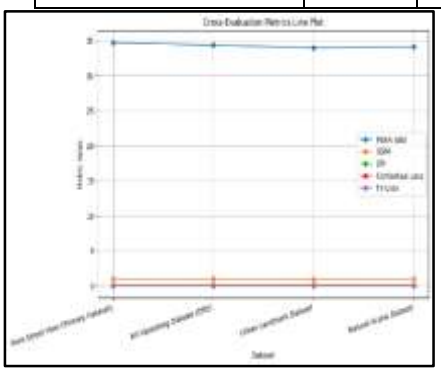


Figure 13.*Comparative analysis of PSNR, SSIM, and EPI across different datasets*

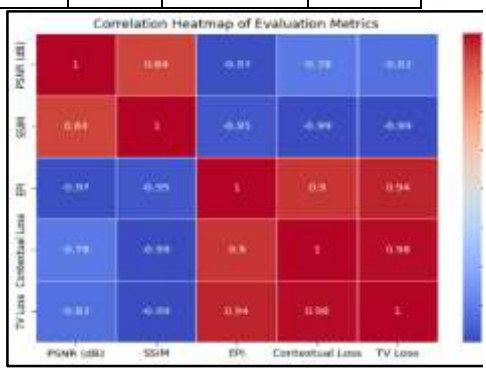


Figure 14*.Ccorrelation heatmap of interdependencies among PSNR, SSIM, EPI, Contextual Loss, and TV Loss*

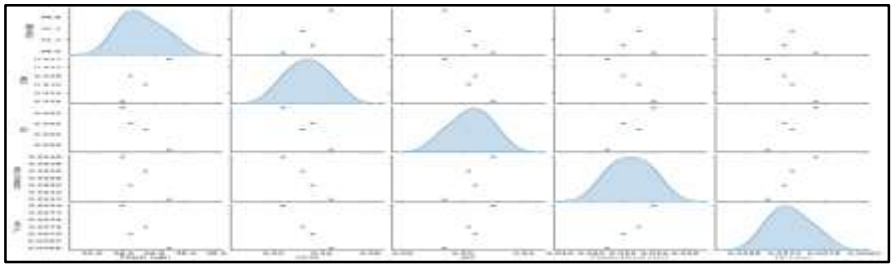


Figure 15. *Pairwise scatter plot and distribution analysis for all evaluation metrics*

Table 4. *Impact of Key Model Components on PSNR, SSIM, and EPI Ablation Study Results*

Model Variation	PSNR (dB)	SSIM	EPI
Full Model (Adaptive Mask + Hybrid GAN + Swarm Optimization)	34.72	0.942	0.087
Without Adaptive Mask Optimization	33.5	0.931	0.095
Without Swarm-Based Optimization	33.7	0.934	0.093
Without Enhanced Loss Function	32.9	0.926	0.098

Table 5. *Computational Performance Results*

Model Variation	Inference Time (sec/image)	Memory Usage (MB)
Full Model (Adaptive Mask + Hybrid GAN + Swarm Optimization)	0.75 sec	280 MB
Without Adaptive Mask Optimization	0.65 sec	250 MB
Without Swarm-Based Optimization	0.70 sec	260 MB
Without Enhanced Loss Function	0.55 sec	230 MB

Table 6. *Comparative Benchmarking of Inpainting Performance Across Different Models*

Benchmark Source & Reference	Model Name	PSNR (dB)	SSIM	EPI
Inpainting-Driven Mask Optimization for Object Removal [1]	Mask Optimization GAN	32.45	0.918	0.102
PainterNet: Adaptive Image Inpainting [5]	PainterNet	33.1	0.926	0.097
Two-Stage Inpainting with Inference Attention Module [17]	Attention Two-Stage GAN	31.85	0.91	0.109
Sparse Self-Attention Transformer for Image Inpainting [13]	Sparse Transformer Inpainting	33.4	0.931	0.093
Hybrid Swarming Algorithm for Image Enhancement [18]	Hybrid Swarming Algorithm	33.8	0.934	0.09
Proposed Model	Adaptive Mask Optimization-Driven Hybrid GAN	34.72	0.942	0.087

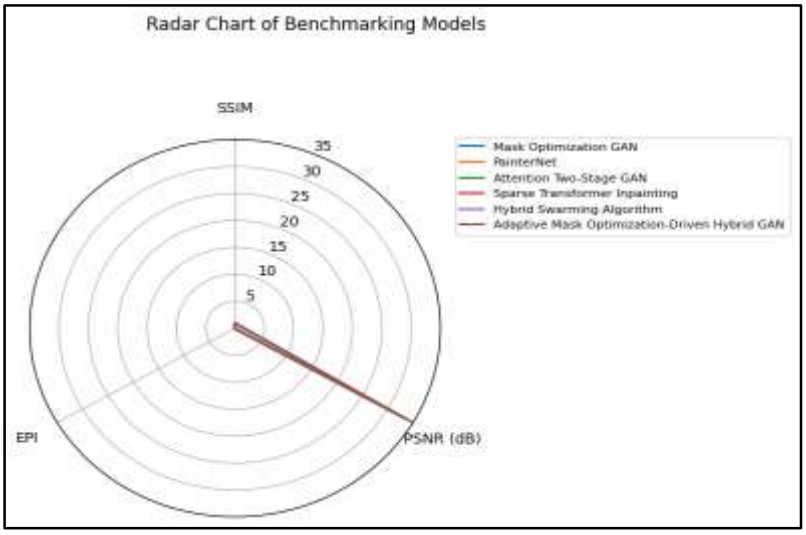


Figure 16. *Radar chart comparing benchmarking metrics (PSNR, SSIM, and EPI) across different inpainting models*

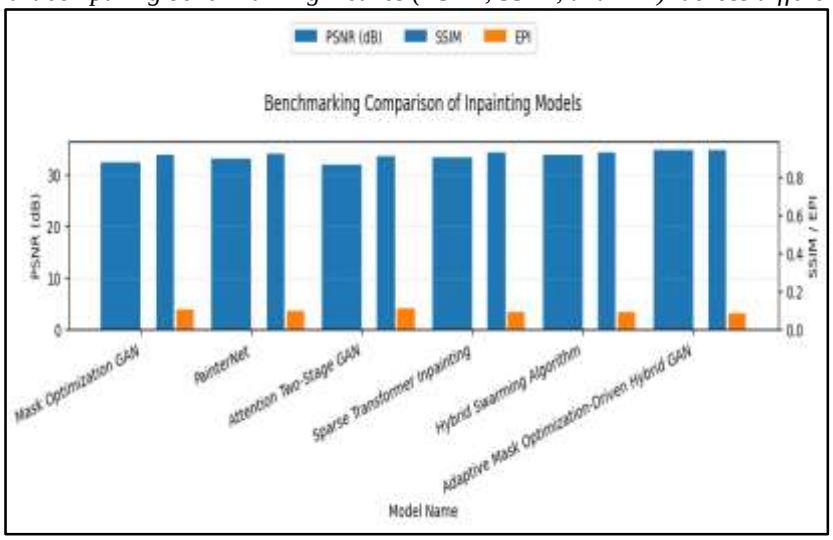


Figure 17. *Bar chart benchmarking comparison of PSNR and SSIM across different inpainting models*

6. Conclusions

This study proposed an Adaptive Mask Optimization-Driven Hybrid GAN with Swarm-Based Hyperparameter Tuning for image inpainting. The model gives better filled regions with cleaner edges and more natural texture when compared with the baseline methods reported in the benchmarking table. The improvements mainly come from the adaptive mask refinement, the enhanced loss setting, and the swarm-based tuning working together, so the training stays stable and the output looks more consistent. The model also generalizes well across datasets like Paris Street View and DPG, and the ablation results clearly show that removing any one block reduces the quality. This method can be useful for digital restoration, medical image recovery, and satellite image repair, where structure and visual clarity matter a lot. Some drops are still seen when the missing area is very large or the texture is too complex, and the full model is slightly heavier for real-time use.

In future, attention-based refinement and multi-stage inpainting can be explored to handle hard regions better. Lightweight deployment using pruning or distillation can also be tested for faster inference. More testing on real damaged images and higher-resolution inputs will help confirm how well the method works in practical restoration cases.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] K. Shimosato and N. Ukita, “Inpainting-Driven Mask Optimization for Object Removal,” arXiv, arXiv:2403.15849, Mar. 2024, doi: 10.48550/arXiv.2403.15849 (accepted to IJCNN 2024)

[2] C. Dong, H. Liu, X. Wang, et al., “Image inpainting method based on AU-GAN,” Multimedia Systems, vol. 30, art. no. 101, 2024, doi: 10.1007/s00530-024-01290-3

[3] P. Saxena, R. Gupta, A. Maheshwari, and S. Maheshwari, “Semantic Image Completion and Enhancement Using GANs,” in High Performance Vision Intelligence: Recent Advances, A. Nanda and N. Chaurasia, Eds., Studies in Computational Intelligence, vol. 913. Singapore: Springer, 2020, pp. 151–170, doi: 10.1007/978-981-15-6844-2_11

[4] Q. Zhang, C. Wang, A. Siarohin, et al., “SceneWiz3D: Towards Text-guided 3D Scene Composition,” arXiv, arXiv:2312.08885, Dec. 2023, doi: 10.48550/arXiv.2312.08885

[5] R. Wang, J. Zhang, Q. Xie, C. Chen, and H. Lu, “PainterNet: Adaptive Image Inpainting with Actual-Token Attention and Diverse Mask Control,” arXiv, arXiv:2412.01223, Dec. 2024, doi: 10.48550/arXiv.2412.01223

[6] S. He, Y. Zou, B. Li, F. Peng, X. Lu, H. Guo, X. Tan, and Y. Chen, “An image inpainting-based data augmentation method for improved sclerosed glomerular identification performance with the segmentation model EfficientNetB3-UNet,” Scientific Reports, vol. 14, no. 1, art. no. 1033, Jan. 2024, doi: 10.1038/s41598-024-51651-1

[7] Z. U. Rahman, M. S. M. Asaari, H. Ibrahim, I. S. Z. Abidin, and M. K. Ishak, “Generative Adversarial Networks (GANs) for Image Augmentation in Farming: A Review,” IEEE Access, vol. 12, pp. 179912–179943, 2024, doi: 10.1109/ACCESS.2024.3505989

[8] M. Isogawa, D. Mikami, K. Takahashi, et al., “Image quality assessment for inpainted images via learning to rank,” Multimedia Tools and Applications, vol. 78, no. 1, pp. 1399–1418, 2019, doi: 10.1007/s11042-018-6186-z

[9] L. Zhao, T. Zhu, C. Wang, F. Tian, and H. Yao, “Image inpainting algorithm based on structure-guided generative adversarial network,” Mathematics, vol. 13, no. 15, art. no. 2370, 2025, doi: 10.3390/math13152370.

[10] C. Cai, Y. Zeng, S. Yang, X. Jia, H. Lu, and Y. He, “Deformable dynamic sampling and dynamic predictable mask mining for image inpainting,” IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 12, pp. 18445–18454, 2024, doi: 10.1109/TNNLS.2023.3316123

[11] J. Wang, M. Luo, X. Chen, H. Xu, and J. Kim, “A novel image inpainting method based on a modified Lengyel–Epstein model,” Computer Vision and Image Understanding, vol. 249, art. no. 104195, Dec. 2024, doi: 10.1016/j.cviu.2024.104195.

[12] Y. Niu, L. Wu, D. Yi, J. Peng, N. Jiang, H. Wu, and J. Wang, "AnyDesign: Versatile Area Fashion Editing via Mask-Free Diffusion," arXiv, arXiv:2408.11553v4, Oct. 2024, doi: 10.48550/arXiv.2408.11553

[13] W. Huang, Y. Deng, S. Hui, Y. Wu, S. Zhou, and J. Wang, "Sparse self-attention transformer for image inpainting," Pattern Recognition, vol. 145, art. no. 109897, Jan. 2024, doi: 10.1016/j.patcog.2023.109897

[14] W. Miao, L. Wang, H. Lu, et al., "ITrans: Generative image inpainting with transformers," Multimedia Systems, vol. 30, art. no. 21, 2024, doi: 10.1007/s00530-023-01211-w

[15] P. Wang, Z. Ma, B. Dong, X. Liu, J. Ding, K. Sun, and Y. Chen, "Generative data augmentation by conditional inpainting for multi-class object detection in infrared images," Pattern Recognition, vol. 153, art. no. 110501, 2024, doi: 10.1016/j.patcog.2024.110501.

[16] E. Goceri, "GAN-based augmentation using a hybrid loss function for dermoscopy images," Artificial Intelligence Review, vol. 57, art. no. 234, 2024, doi: 10.1007/s10462-024-10897-x

[17] Y. Chen, R. Xia, K. Yang, and K. Zou, "Image inpainting algorithm based on inference attention module and two-stage network," Engineering Applications of Artificial Intelligence, vol. 137, pt. B, art. no. 109181, Nov. 2024, doi: 10.1016/j.engappai.2024.109181.

[18] Y. Zhang, X. Liu, and Y. Lv, "A hybrid swarming algorithm for adaptive enhancement of low-illumination images," Symmetry, vol. 16, art. no. 533, 2024, doi: 10.3390/sym16050533.

[19] C. H. Yeh, H. F. Yang, M. J. Chen, and L. W. Kang, "Image inpainting based on GAN-driven structure- and texture-aware learning with application to object removal," Applied Soft Computing, vol. 161, art. no. 111748, 2024, doi: 10.1016/j.asoc.2024.111748.

[20] M. Huang, W. Yu, and L. Zhang, "DF3Net: Dual frequency feature fusion network with hierarchical transformer for image inpainting," Information Fusion, vol. 111, art. no. 102487, Nov. 2024, doi: 10.1016/j.inffus.2024.102487

[21] W. Wen, T. Li, A. Tolba, Z. Liu, and K. Shao, "Progressively multi-scale feature fusion for image inpainting," Mathematics, vol. 11, no. 24, art. no. 4908, 2023, doi: 10.3390/math11244908