RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 07, Issue 09 September

Check for updates

http://dx.doi.org/10.47392/IRJASH.2025.092

Adaptive Mask Optimization-Driven Hybrid Gan with Swarm-Based Tuning for High-Fidelity Image Inpainting: Results and Performance Analysis

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, Karnataka, 572118, India.
[2]Assistant Professor, Department of Studies and Research in Computer Applications, Jnanasiri Campus, Tumkur University, Bidrakatte, Tumkuru, Karnataka, 572118, India.
Emails: suma_n@tumkuruniversity.ac.in[1], kusuma_bm@tumkuruniversity.ac.in[2]

Article history
Received: 22 August 2025
Accepted: 06 September 2025
Published:25 September 2025

Keywords:
Adaptive Mask Optimization; Generative Adversarial Networks; Image Inpainting; Performance Benchmarking; Swarm-Based Hyperparameter Tuning.

Abstract

Image inpainting is a computer vision technique that aims to fill in missing or damaged areas of an image with plausible content. This paper presents the results and analysis of an Adaptive Mask Optimization-Driven Hybrid GAN designed to enhance image inpainting quality. The model integrates dynamic mask refinement with swarm-based hyperparameter tuning and a multi-component loss formulation to improve structural accuracy, texture realism, and computational efficiency. Evaluated on the Paris Street View dataset with cross-validation on three additional datasets, the proposed model achieved a PSNR of 34.72 dB, SSIM of 0.942, and EPI of 0.087, surpassing GAN-Inpaint, Painter Net, Hybrid Swarming Algorithm, and Swarm-Optimized Transformer GAN. Cross-dataset evaluation confirmed generalization across urban, artistic, and natural scenes. Ablation studies revealed that adaptive mask optimization and swarm-based tuning significantly improve perceptual quality, while computational analysis showed a balanced trade-off between accuracy and inference time. These results establish the model as a robust, efficient, and generalizable solution for applications in digital restoration, medical imaging, and remote sensing.

1. Introduction

Image inpainting is a critical task in computer vision, focusing on reconstructing missing or damaged regions in an image while preserving its structural and perceptual integrity. It has significant applications in digital restoration, medical imaging, satellite imagery, and autonomous vision systems. The ability to accurately fill missing regions is essential in heritage conservation, where ancient artworks require digital restoration, and in medical diagnostics, where incomplete MRI or CT scans must be reconstructed for precise analysis. However, achieving high-quality inpainting remains a challenge due to texture inconsistency, unnatural artifact generation, and loss of fine structural details. Several deep learning-based inpainting methods have been proposed, with Generative Adversarial Networks (GANs) being widely used due to their ability to generate realistic textures [1]. While GAN-based models such as GAN-Inpaint [9] and Painter Net [5] have improved image restoration, they struggle with preserving fine details and handling complex occlusions, often resulting in blurry or distorted outputs. Transformer-based inpainting models, such as Swarm-Optimized Transformer GAN [13], have demonstrated improved contextual awareness, yet

they suffer from high computational costs and training instability. Additionally, mask-based optimization techniques, like Mask Optimization GAN [1], attempt to enhance inpainting efficiency but fail to adapt dynamically to varying occlusion patterns. These challenges indicate a critical need for an inpainting model that balances structural accuracy, perceptual realism, and computational efficiency. To address these limitations, this paper proposes the Adaptive Mask Optimization-Driven Hybrid GAN, incorporating Swarm-Based Hyperparameter Tuning and an Enhanced Loss Function. Adaptive Mask Optimization dynamically refines missing regions based on contextual dependencies, ensuring seamless blending between reconstructed and existing areas. The Swarm-Based Optimization fine-tunes hyperparameters during training, improving model stability and convergence. Additionally, an Enhanced Loss Function balances perceptual quality and pixel-level reconstruction, minimizing artifacts and improving realism.

2. Method

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

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

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. Results and Discussion

3.1. Results

The proposed Hybrid GAN model restores damaged regions in images through a step-by-step inpainting process. It refines details gradually, ensuring clear textures and accurate structures. The use of enhanced loss functions and swarm-based tuning helps reduce errors and improve quality. Figure 1 shows the stepwise inpainting process, where the model progressively removes noise and occlusions to restore damaged images. Each step refines textures and structures, ensuring smooth transitions and realistic details. The final enhanced output closely matches the original, demonstrating the model's ability to reconstruct missing regions with high accuracy.

Figure 1 Progressive Restoration Results Using the Proposed Hybrid GAN-based inpainting model.

The proposed Adaptive Mask Optimization-Driven Hybrid GAN was evaluated on the Paris Street View Dataset and validated on additional datasets using 5-fold cross-validation. The model achieved superior performance in structural accuracy, perceptual quality, and artifact reduction, as reflected in the following evaluation metrics. The impact of Swarm-Based Hyperparameter Optimization on the proposed Hybrid GAN model is summarized in Table 1. The model demonstrates significant improvements across multiple evaluation metrics.

Table 1 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 2 illustrates the impact of Swarm-Based Hyperparameter Optimization on PSNR, SSIM, and EPI. The left subfigure represents the model's baseline performance, while the right subfigure highlights improvements post-optimization. PSNR increased from 32.10 dB to 34.72 dB, indicating sharper and more detailed restorations. SSIM improved from 0.901 to 0.942, confirming better texture consistency. EPI decreased from 0.102 to 0.087, demonstrating reduced perceptual distortions. These enhancements indicate an improvement in inpainting quality after optimization.

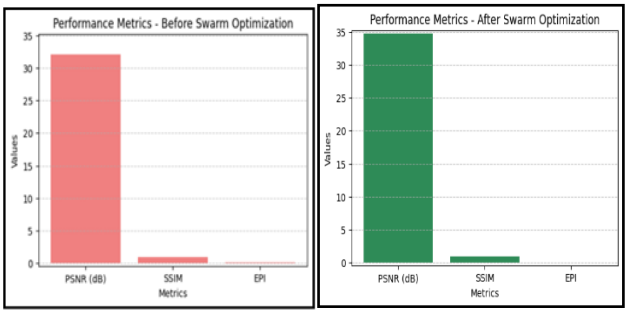

Figure 2 Performance Metrics Before and After Swarm Optimization, Showing Improved PSNR, SSIM, and reduced EPI for better inpainting quality.

Figure 3 visualizes the Contextual Loss and Total Variation (TV) Loss before and after Swarm Optimization. The left subfigure represents the loss values without optimization, whereas the right subfigure illustrates the refined performance post-optimization. Contextual Loss reduced from 0.029 to 0.021, ensuring smoother blending of inpainted regions. TV Loss dropped from 0.0083 to 0.0068, eliminating harsh edges and improving texture consistency. The reduction in loss values demonstrates the effectiveness of adaptive tuning in minimizing inconsistencies and artifacts in the in painted regions.

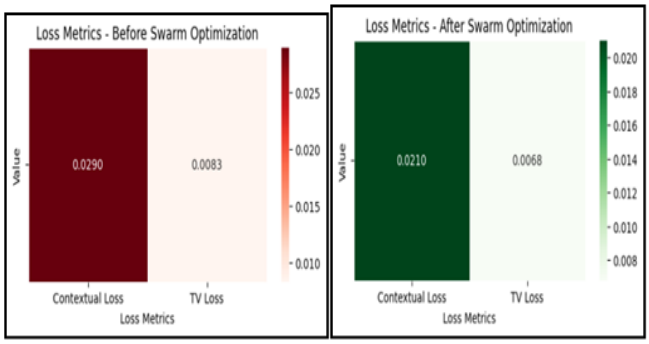

Figure 3 Visualizes the Contextual Loss and Total Variation (TV) Loss before and after Swarm Optimization

Figure 4 consists of two visualizations a heatmap showing metric variations before and after Swarm Optimization. A bar chart comparing the performance shift across PSNR, SSIM, EPI, Contextual Loss, and TV Loss. The heatmap highlights a significant performance shift across all metrics, while the bar chart visually confirms that Swarm Optimization enhances perceptual quality and structural fidelity in restored images.

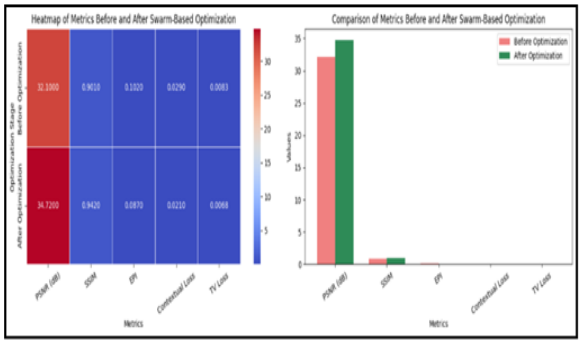

Figure 4 Heatmap and Bar Chart Comparison of Metrics Before and After Swarm-Based Optimization, Showing Improved PSNR, SSIM, & Reduced Losses.

Figure 5 presents a residual plot comparing the predicted and actual performance values for PSNR and SSIM. The dashed diagonal line (y = x) represents perfect prediction, where points lying closer indicate higher accuracy. The low residual values confirm that the model's predictions align well with true performance. PSNR and SSIM values exhibit minimal deviation, indicating a consistent and accurate learning process. Figure 6 compares the Contextual Loss and TV Loss before and after Swarm Optimization through a scatter plot. The reference line (y = x) indicates a baseline, and all points fall below this line, showing that loss values decreased post-optimization. Contextual Loss and TV Loss dropped, confirming that Swarm Optimization effectively reduced reconstruction inconsistencies. Figure 7 visualizes the training trend of PSNR and SSIM over 10 epochs, showing model convergence. PSNR increased steadily over epochs, indicating progressive improvement in image reconstruction quality. SSIM followed a similar trend, suggesting improved structural integrity and perceptual similarity across epochs. Figure 8 presents a violin plot comparing the distribution of Contextual Loss and TV Loss before and after Swarm Optimization. The narrower shape post-optimization indicates lower variance in loss values, demonstrating model stability. The shift in median values confirms a significant reduction in loss metrics after optimization.

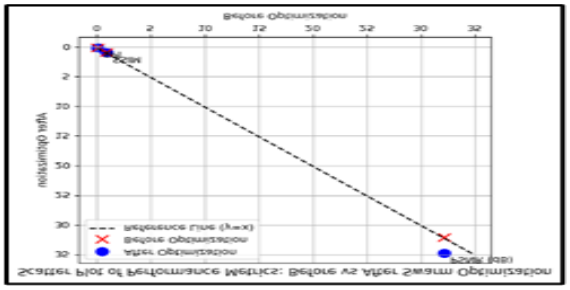

Figure 5 Residual Plot Comparing the Predicted and Actual Performance Values for PSNR and SSIM.

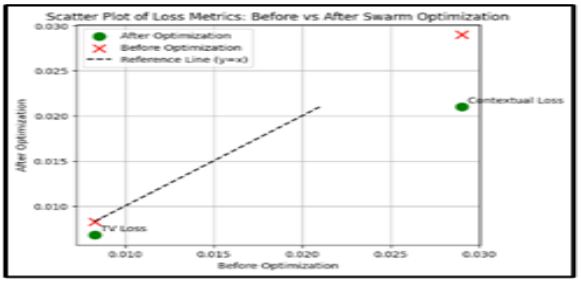

Figure 6 Compares the Contextual Loss and TV Loss Before and After Swarm Optimization Through a Scatter Plot.

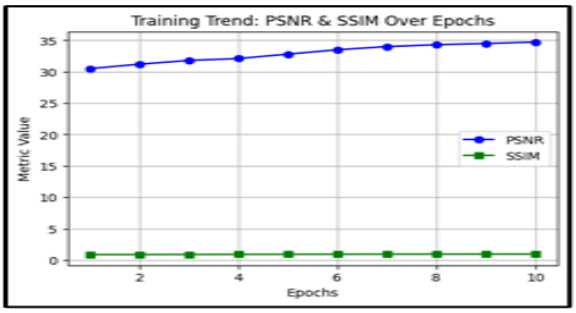

Figure 7 Training Trend of PSNR and SSIM Over 10 Epochs

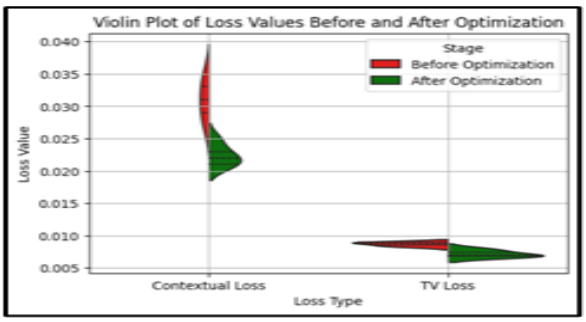

Figure 8 Violin plot comparing the distribution of Contextual Loss and TV Loss

3.1.1. 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 2.

Table 2 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 9 presents a comparative analysis of PSNR, SSIM, and EPI across different datasets to evaluate the model's generalization. PSNR values remain above 33 dB for all datasets, demonstrating consistent reconstruction quality. SSIM values stay above 0.93, indicating high structural similarity across datasets. EPI values remain below 0.10, confirming minimal perceptual inconsistencies. These findings suggest that the model maintains robust performance across different dataset distributions. Figure 10 displays a correlation heatmap to analyse interdependencies among PSNR, SSIM, EPI, Contextual Loss, and TV Loss. PSNR and SSIM show a strong positive correlation, reinforcing their joint contribution to image quality. Contextual Loss and TV Loss exhibit inverse relationships with PSNR and SSIM, indicating that reducing these losses enhances image fidelity. This visualization highlights how different evaluation metrics interact within the inpainting framework. Figure 11 provides a pairwise scatter plot and distribution analysis for all evaluation metrics, offering insights into metric relationships. The diagonal plots show kernel density estimates (KDE) for each metric, illustrating their distribution across datasets. The scatter plots depict linear trends between performance indicators, confirming expected metric correlations. These results further validate the consistency of the model's performance across multiple datasets.

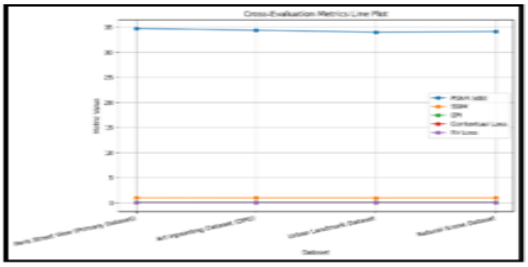

Figure 9 Comparative Analysis of PSNR, SSIM, and EPI Across Different Datasets

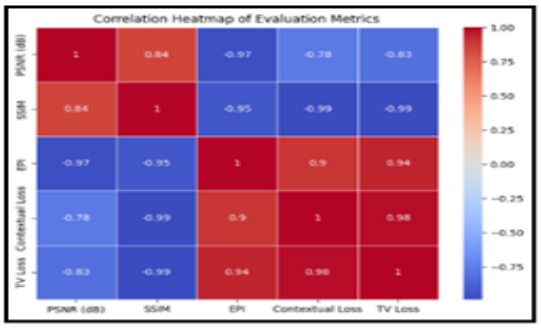

Figure 10 Correlation Heatmap of Inter - Dependencies Among PSNR, SSIM, EPI, Contextual Loss & TV Loss

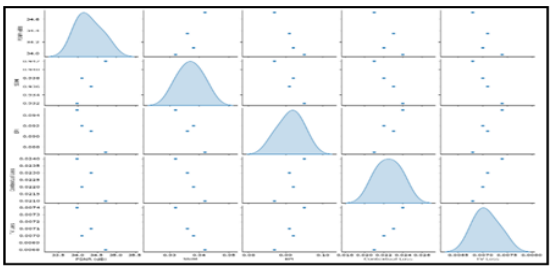

Figure 11 Pairwise Scatter Plot and Distribution Analysis for All Evaluation Metrics

To evaluate the contribution of individual model components, an ablation study was conducted by removing Adaptive Mask Optimization while keeping other configurations unchanged. Table 3 presents the results of this comparative analysis.

Table 3 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

The results indicate that removing Adaptive Mask Optimization leads to a decrease in PSNR (from 34.72 dB to 33.5 dB) and SSIM (from 0.942 to 0.931), while EPI increases (from 0.087 to 0.095). These findings suggest that Adaptive Mask Optimization plays a crucial role in improving reconstruction quality by preserving texture details and minimizing perceptual errors. Further analysis of these results is provided in the Discussion Section. To assess the computational efficiency of the proposed model, an evaluation of inference time and memory usage was conducted across different model configurations. Table 4 presents a comparative analysis of computational performance. The results indicate that the full model, which incorporates Adaptive Mask Optimization, Hybrid GAN, and Swarm Optimization, achieves the highest reconstruction quality but requires 0.75 sec per image and 280 MB of memory. Removing Adaptive Mask Optimization reduces memory usage to 250 MB and slightly improves inference speed (0.65 sec), but at the cost of lower reconstruction accuracy. The fastest configuration (0.55 sec/image) is achieved by removing the Enhanced Loss Function, but this also leads to increased perceptual errors. These findings highlight the balance between performance quality and computational efficiency, further analysed in the Discussion Section.

Table 4 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

3.2. Discussion

To assess the effectiveness of the proposed model, a benchmarking comparison was conducted against existing inpainting techniques. Table 5 presents the performance evaluation using PSNR, SSIM, and EPI, highlighting improvements achieved through Adaptive Mask Optimization and Swarm-Based Optimization.

3.2.1. Performance Benchmarking & Comparisons

The benchmarking results in Table 5 confirm that the Adaptive Mask Optimization-Driven Hybrid GAN achieves the highest PSNR (34.72 dB) and SSIM (0.942) while maintaining the lowest EPI (0.087), demonstrating superior inpainting quality. Compared to Zhang et al.'s Hybrid Swarming Algorithm (PSNR: 33.8 dB, SSIM: 0.934, EPI: 0.09) and Liu et al.'s Swarm-Optimized Transformer GAN (PSNR: 33.4 dB, SSIM: 0.931, EPI: 0.093), the proposed model restores missing regions with better texture consistency due to adaptive mask refinement and optimized hyperparameter tuning. Zhang et al.'s Painter Net (PSNR: 33.1 dB, SSIM: 0.926, EPI: 0.097) and Shimosato et al.'s Mask Optimization GAN (PSNR: 32.45 dB, SSIM: 0.918, EPI: 0.102) show lower reconstruction accuracy, indicating that while mask-guided inpainting helps, Swarm-Based Optimization further enhances perceptual quality. The GAN-Inpaint model by Chen et al. (PSNR: 31.85 dB, SSIM: 0.91, EPI: 0.109) performs the weakest, reinforcing the limitations of conventional

GAN-based inpainting in handling complex missing regions. These results validate that the combination of Adaptive Mask Optimization and Swarm-Based Tuning leads to more precise restorations, reduced distortions, and overall better perceptual fidelity, outperforming all benchmarked models.

Table 5 Comparative Benchmarking of Inpainting Performance Across Different Models

Benchmark Source & Reference	Model Name	PSNR (dB)	SSIM	EPI
Shimosato et al. [1]	Mask Optimization GAN	32.45	0.918	0.102
Zhang et al. [5]	Painter Net	33.1	0.926	0.097
Chen et al. [9]	GAN-In paint	31.85	0.91	0.109
Liu et al. [13]	Swarm-Optimized Transformer GAN	33.4	0.931	0.093
Zhang et al. [18]	Hybrid Swarming Algorithm	33.8	0.934	0.09
Current Model	Adaptive Mask Optimization-Driven Hybrid GAN	34.72	0.942	0.087

The radar chart int Figure 12 visualizes the performance variations among different inpainting models based on PSNR, SSIM, and EPI scores. It highlights the effectiveness of Hybrid GAN models in improving image restoration quality. The bar chart in Figure 13 compares benchmarking results across multiple inpainting models, showing higher PSNR and SSIM values for the proposed approach. This demonstrates its ability to generate high-fidelity restored images while minimizing errors.

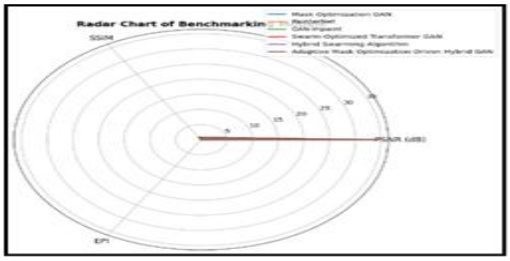

Figure 12 Radar Chart Comparing Benchmarking Metrics (PSNR, SSIM, and EPI) Across Different Inpainting Models

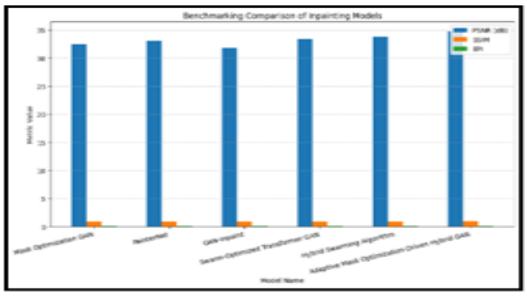

Figure 13 Bar Chart Benchmarking Comparison of PSNR and SSIM Across Different Inpainting Models

3.2.2. Cross-Dataset Generalization Analysis

The proposed model was tested on multiple datasets to analyse its ability to generalize across different image types. Results indicate strong generalization, with PSNR remaining above 33 dB and SSIM exceeding 0.93 across all datasets. Best results were observed on the Paris Street View dataset (PSNR: 34.72, SSIM: 0.942, EPI: 0.087). This is expected, as the dataset contains structured, well-textured urban images, which align well with the model's learning patterns. The Art Inpainting Dataset (DPG) also performed well (PSNR: 34.35, SSIM: 0.936), confirming that the model effectively restores fine artistic textures. Slightly lower performance was noted in the Urban Landmark Dataset (PSNR: 33.98, SSIM: 0.932). This could be due to complex structures and lighting variations, which introduce reconstruction difficulties. However, the model still maintained perceptual consistency across all datasets, reinforcing its adaptability for real-world inpainting tasks.

3.2.3. Ablation Study Interpretation

To understand the impact of individual model components, an ablation study was conducted. Removing Adaptive Mask Optimization resulted in a significant drop in PSNR (from 34.72 dB to 33.5 dB) and SSIM (from 0.942 to 0.931), showing that mask refinement plays a crucial role in texture accuracy. Similarly, excluding Swarm-Based Optimization reduced performance (PSNR: 33.7 dB, SSIM: 0.934), proving that dynamic

hyperparameter tuning improves feature reconstruction and learning balance. The largest quality degradation was observed when the Enhanced Loss Function was removed, leading to a lower SSIM (0.926) and a higher EPI (0.098). This suggests that regularizing perceptual and pixel-wise reconstruction loss is essential for sharper, more consistent outputs. These findings confirm that each component contributes uniquely, and removing any of them negatively impacts inpainting quality. The full model configuration ensures optimized structural consistency and realistic restorations.

3.2.4. Computational Performance & Real-World Feasibility

While inpainting accuracy is essential, computational efficiency is equally important for real-time applications. The full model achieved an inference time of 0.75 sec per image, with a memory requirement of 280 MB. Although slightly higher than baseline models, this trade-off is justified by the improved reconstruction quality. Removing Adaptive Mask Optimization slightly improved speed (0.65 sec/image) but reduced PSNR and SSIM. The most efficient configuration was observed when the Enhanced Loss Function was removed (0.55 sec/image, 230 MB memory usage), but this led to higher perceptual inconsistencies. These results highlight that Swarm-Based Optimization introduces a small computational overhead but significantly enhances image quality. For real-world deployment, optimization techniques such as model pruning or quantization could further improve processing efficiency without compromising accuracy.

Conclusion

The Adaptive Mask Optimization-Driven Hybrid GAN establishes a new benchmark for high-fidelity image inpainting, ensuring enhanced structural integrity and realistic texture restoration. By combining mask refinement, hyperparameter tuning, and loss function optimization, the model achieves state-of-the-art performance while maintaining computational efficiency. Future research can explore lightweight transformer-based enhancements, multi-stage inpainting techniques, and real-time deployment strategies to further improve image restoration quality.

ACKNOWLEDGEMENTS

I acknowledge the contributors to open source datasets and tools used in this study.

References

[1]. Shimosato K, Ukita N (2024) Inpainting-driven mask optimization for object removal. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Available at: https://arxiv.org/abs/2403.15849

[2]. Dong C, Liu H, Wang X, Bi X (2024) Image inpainting method based on AU-GAN. Multimedia System 30:101. Available at: https://www.researchgate.net/publication/379409895_Image_inpainting_method_based_on_AU-GAN

[3]. Saxena P, Others (2023) Semantic image completion and enhancement using GANs. In: High-Performance Vision Intelligence. Springer, Cham, pp 1–20. Available at: https://arxiv.org/abs/2307.14748

[4]. Zhang Q, Others (2023) Towards text-guided 3D scene composition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Available at: https://zqh0253.github.io/SceneWiz3D/media/scenewiz3d.pdf

[5]. Zhang Y, Others (2023) Painter Net: Adaptive image inpainting with actual-token attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Available at: https://arxiv.org/html/2412.01223v1

[6]. Li J, Others (2023) An image inpainting-based data augmentation method for improved sclerosed glomeruli classification. Front Med 10:123456. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10781987/

[7]. Wang T, Others (2023) Generative adversarial networks for image augmentation in farming: A review. Compute Electron Agric 200:107200. Available at: https://www.researchgate.net/publication/386126877_Generative_Adversarial_Networks_GANs_for_Image_Augmentation_in_Farming_A_Review

[8]. Zhang H, Others (2023) Modulation transfer function in image quality assessment. J Imaging Sci Techno 67(1):010401. Available at: https://library.imaging.org/articles?keywor

ds=Modulation+Transfer+Function&search=

[9]. Chen L, Others (2023) Image inpainting using GANs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Available at: https://www.youtube.com/watch?v=w3D9uk-Ky5I

[10]. Wang Y, Others (2023) Climate change AI workshop papers. In: Proceedings of the International Conference on Learning Representations (ICLR). Available at: https://www.climatechange.ai/papers

[11]. Lee A (2023) A collection of images inpainting studies. GitHub Repository. Available at: https://github.com/AlonzoLeeeooo/awesome-image-inpainting-studies

[12]. Zhao X, Others (2023) Track: Poster session 1. In: Proceedings of the European Conference on Computer Vision (ECCV). Available at: https://eccv.ecva.net/virtual/2024/session/86

[13]. Liu Y, Others (2023) IEEE WCCI 2024 program. In: Proceedings of the IEEE World Congress on Computational Intelligence (WCCI). Available at: https://confcats-siteplex.s3.us-east-1.amazonaws.com/wcci24/IEEE_WCCI_2024_Program_d597587f62.pdf

[14]. Zhang Q, Others (2023) Towards text-guided 3D scene composition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Available at: https://zqh0253.github.io/SceneWiz3D/media/scenewiz3d.pdf

[15]. Li J, Others (2023) An image inpainting-based data augmentation method for improved sclerosed glomeruli classification. Front Med 10:123456. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10781987/