ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

ADVANCES IN IMAGE INPAINTING WITH DEEP LEARNING: METHODS, METRICS, AND MULTISCALE CHALLENGES

Santosh Dattatray Kale

Research Scholar

Savitribai Phule Pune University, Department of Electronics and Telecommunication,

S.V.P.M's College of Engineering, Malegaon Bk, Taluka Baramati,Dist-Pune,Maharashtra,India

Shailendrakumar M.Mukane

Principal and Professor

Savitribai Phule Pune University, S.V.P.M's College of Engineering, Malegaon Bk,

Taluka Baramati, Dist-Pune, Maharashtra, India

ABSTRACT.

Image inpainting has undergone a transformative evolution in recent years, propelled by the rise of deep learning. This review provides a critical analysis of the latest advances in image inpainting, focusing on methods grounded in Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), Transformers, and Diffusion Models. We organize the literature across algorithmic and architectural lines, highlight functional innovations such as multi-scale attention, structure-texture separation, and pseudo-labelling, and evaluate their performance using metrics like PSNR, SSIM, LPIPS, and FID on standard datasets. In addition, we explore domain-specific challenges, particularly high-angle fractures and periodic defects, and offer insight into future research opportunities. This study serves as a comprehensive guide to the current landscape and evolving frontiers in deep learning-based image inpainting.

Keywords:

GAN, CNN, Transformer, Diffusion Models, PSNR, SSIM, LPIPS, Functional Innovations.

1. INTRODUCTION

Image inpainting, the process of reconstructing missing or corrupted parts of images, has transitioned from traditional patch-based and diffusion-based techniques to data-driven deep learning methods. Deep neural networks now dominate this field due to their ability to model complex semantics, textures, and long-range dependencies. This review investigates the recent strides made in deep learning-based image inpainting, categorizing them into key algorithmic families and highlighting the innovations that shape their success.

2. LITERATURE REVIEW

Zhong et al., 2025[1] proposed a diffusion-based image inpainting method specifically tailored for restoring eaves tile patterns in Chinese heritage buildings. Utilizing a denoising diffusion probabilistic model (DDPM), their approach adopts a coarse-to-fine restoration strategy that effectively captures intricate textures and cultural motifs. The method outperformed GAN-based baselines on PSNR, SSIM, and LPIPS, demonstrating its suitability for high-fidelity reconstruction of artistic patterns.

Wang et al., 2025[2] introduced a CNN-based method called MAPs-IN, which leverages Multi-scale Adaptive Priors (MAPs) to enhance contextual reasoning during inpainting. Their

© Copyright @ Authors

approach dynamically adjusts feature aggregation at different resolutions, resulting in better restoration of complex textures across varying scales. MAPs-IN achieved superior perceptual scores across multiple datasets and presented a robust mechanism for multi-resolution feature guidance.

Lian et al., 2025[3] developed the TSBGNet architecture that integrates Bidirectional Information Flow (BIF) for simultaneous texture and structure modelling. The network also employs a detail enhancement module and a multi-scale attentional fusion mechanism. Their experiments, conducted on CelebA-HQ and Dunhuang mural datasets, showed enhanced realism and structural coherence, highlighting the value of interactive texture-structure pipelines.

Li et al., 2025[4] proposed a transformer-based structure-texture dual restoration network (CTSTNet) that jointly uses self-attention and convolution to bridge long-range dependencies in image inpainting. Their network outperformed previous methods on Places2 and Paris StreetView datasets by recovering more semantically aligned content, particularly in scenes with large missing regions.

Zhong et al., 2025[5] explored DDPM-based inpainting for micro resistivity logs in geophysical imaging. The study incorporated pseudo-labelling and semantic conditioning to address data scarcity, leading to high SSIM and PSNR results even with 50% randomly distributed missing channels. Their approach demonstrated the feasibility of generative diffusion models in scientific domains.

Ding et al., 2024[6] addressed periodic discrete density (PDD) defect restoration using a novel adaptive transformer-GAN hybrid (FGTNet). The method combines spatial features with frequency domain priors and introduces a coarse- to-fine inpainting scheme. Their framework achieved state-of-the-art LPIPS and FID scores on both microscopic and public datasets, confirming its utility in fine-structure restoration.

Xuan et al., 2024[7] evaluated the DeepFillv2 architecture on electron cyclotron emission imaging (ECEI) data. Comparing DeepFillv2 with traditional filling techniques, they reported over 90% SSIM and above 30 dB PSNR even at high corruption levels. The study validates the generalizability of GAN-based models in non-natural imaging domains.

Chen et al., 2024[8] presented MFMAM, a multi-scale feature and attention module that improves semantic inpainting by enhancing deep-level texture acquisition. By integrating dilated convolution and perceptual loss, the model significantly reduced FID scores across CelebA and Outdoor Scenes datasets, proving effective in maintaining stylistic coherence.

Zhong et al., 2024[9] proposed FPEM-GAN, a pseudo-labelling enabled GAN for inpainting geological imaging data. Their method combines a feature prediction module with a segmentation-driven U-Net for enhancing restoration in low-label environments. Experimental results show high perceptual realism and accuracy on micro resistivity logs.

Zhang et al., 2024[10] developed CAML (Context-aware Mutual Learning), a GAN model with inference attention for handling blind inpainting tasks. The model learns contextual features through bi-directional mutual updates and improves visual continuity in challenging occlusion scenarios. The framework achieves competitive PSNR and SSIM on free-form mask datasets.

Nazeri et al., 2019[11] introduced Edge Connect, a two-stage GAN-based model for image inpainting. In the first stage, edges are predicted in the missing regions, and in the second, the image is completed with guidance from the edge map. Their model set a benchmark for

© Copyright @ Authors

incorporating explicit structural priors, outperforming conventional GANs on free-form mask datasets such as CelebA and Paris StreetView.

Yu et al., 2018[12] presented DeepFillv2, a free-form inpainting model using gated convolutional layers. This method dynamically adjusts feature propagation based on the input mask, effectively addressing irregular inpainting masks. Their model improved PSNR and FID scores over previous GAN-based baselines and remains a reference point for robust, structure-aware inpainting.

Yeh et al., 2017[13] explored semantic image inpainting using a Deep Generative Model (DGM) based on GANs and image feature optimization. Their method generates multiple plausible completions and allows users to interactively refine results. This approach helped bridge early semantic understanding with generative modelling.

Pathak et al., 2016[14] introduced Context Encoders, one of the earliest GAN-based architectures for image inpainting. Their encoder-decoder network learns to predict the missing center region of an image based on context, establishing adversarial loss as a strong supervisory signal. This work is foundational and widely cited for initiating the shift from traditional to learning-based inpainting.

Liu et al., 2018[15] proposed partial convolutions, which dynamically mask out missing pixels during convolution operations, ensuring that feature learning is conditioned only on valid data. This model proved especially effective on irregular masks and became a foundational building block for subsequent CNN-based models.

Yeh et al., 2024[16] developed a structure-texture learning GAN model, focusing on disentangled recovery where structure is reconstructed first, followed by texture refinement. The two-stage GAN model demonstrated better SSIM andperceptual quality on CelebA and Places2 datasets.

Yang et al., 2024[17] presented a novel inference attention mechanism integrated with GANs for context-aware inpainting. Their approach, evaluated on natural and synthetic masks, showed strong visual consistency and demonstrated the potential of attention fusion for blind inpainting.

Lu et al., 2024[18] explored a multi-resolution transformer-based method for complex texture and geometry recovery. The model, tested on Paris StreetView and custom cultural heritage datasets, demonstrated the ability to maintain long-range coherence and texture sharpness.

Lei et al., 2025[19] introduced an attention-gated convolutional network optimized for texture-consistent restoration in occluded images. Their model uses selective attention and gating mechanisms to balance semantic reconstruction and style consistency across large masks.

Zou et al., 2024[20] proposed an adaptive learning framework for image inpainting applied to document restoration. Their model uses self-supervised loss combined with channel attention to recover occluded text and patterns. The approach proved particularly effective in preserving typography and layout in scanned archival documents.

3. Taxonomy of Methods

GAN-Based: Edge Connect, DeepFillv2, and structure-texture learning models restore global structure and refine texture with adversarial realism

CNN-Based: MAPs-IN and attention-driven ResU-Net models emphasize feature reconstruction with multi-scale priors.

© Copyright @ Authors

Transformer-Based: CTSTNet and BIF-enhanced models model long-range dependencies for accurate restoration.

Diffusion-Based: DDPMs offer generative realism via coarse-to-fine denoising, particularly effective in heritage restoration and scientific imaging.

4. Functional Innovations Use

Recent functional innovations in image inpainting focus on enhancing model expressiveness, contextual reasoning, and adaptability across diverse image domains. One of the foundational advances is **structure-texture separation**, where inpainting is divided into two stages—first restoring the global structure or edges of an image, followed by the refinement of textures. This approach improves the semantic plausibility and visual realism of completed regions, as seen in GAN-driven architectures like Edge Connect and structure-aware GANs.

Attention mechanisms have also emerged as powerful tools to prioritize contextually relevant features. Channel attention, spatial attention, and inference-based attention modules help networks to dynamically focus on damaged areas while preserving contextual harmony. These are prominently utilized in models like ResU-Net variants and adaptive gated networks.

Another notable innovation is the incorporation of **bidirectional information flow**, where structure and texture information are iteratively refined and shared across dual branches, as implemented in TSBGNet. This promotes more consistent recovery of fine-grained details in both local and global features.

Multi-scale adaptive priors (MAPs) further improve restoration quality by aggregating hierarchical feature information across different resolution scales. Models like MAPs-IN leverage specialized aggregators that dynamically select relevant features during decoding, thus ensuring effective reconstruction across complex patterns and shapes.

In addition to these, **pseudo-labelling techniques**—especially in domains with limited ground truth data—enable semi-supervised training by generating reliable supervisory signals from incomplete images. Coupled with architectures like SM-UNet and FPEM-GAN, pseudo-labelling supports high-quality restoration of scientific imaging datasets.

Finally, **frequency domain priors**, which exploit structural and periodic information in the frequency spectrum, offer promising results in specialized applications such as periodic discrete defect correction in microscopy and geological logging. These methods often pair spatial-domain networks with FFT-based transformations to preserve fine frequency textures while ensuring global coherence.

5. FUNCTIONAL INNOVATIONS WITH METADATA

Table 1: Functional innovations in image inpainting:

Functional Innovation	Publication	Author(s)	Year
Structure-Texture Learning	Applied Soft Computing	Chia-Hung Yeh et al	2024
Edge-Guided Restoration	CVPR	Nazeri et al.	2019
Gated Convolution	ECCV	Yu et al.	2018
Transformer-based Restoration	Applied Soft Computing	Zhan Li et al.	2025
Multi-scale Adaptive Priors	Pattern Recognition	Yufeng Wang et al	2025

© Copyright @ Authors

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Bidirectional Flow Fusion	Signal Processing	Jing Lian et al.	2025
Diffusion-based Restoration	Automation in Construction	Xiaohan Zhong et al	2025
Pseudo-labeling with GAN	Computers & Geosciences	Zhaoyan Zhong et al	2025

Table 2: Functional innovations in image inpainting

functional Innovation	Key Focus	Significance
Structure-Texture Learning	Separating structural and texture recovery	Enhances perceptual quality and structural accuracy
Edge-Guided Restoration	Using edge maps to guide inpainting	Enables shape-aware filling and better object boundaries
Gated Convolution	Selectively updating only known regions	Prevents information leakage and improves learning focus
Transformer-based Restoration	Capturing long-range dependencies	Supports better semantic coherence in large missing
Multi-scale Adaptive Priors	Fusing multi-resolution features adaptively	Enhances detail preservation across varied textures
Bidirectional Flow Fusion	Interaction between structure & texture	Ensures holistic coherence during refinement
Diffusion-based Restoration	Coarse-to-fine probabilistic modeling	Leads to photorealistic and smooth restorations
Pseudo-labeling with GAN	Semi-supervised inpainting for data-scarce domains	Addresses real-world data limitations and annotation bottlenecks

The landscape of deep learning-based image inpainting has been significantly enriched by a variety of functional innovations that address challenges related to structural accuracy, texture fidelity, and domain-specific restoration. One of the foundational breakthroughs is structure-texture learning, as introduced by Chia-Hung Yeh et al. (2024) in *Applied Soft Computing*.

This two-stage approach separates the task of structure recovery (e.g., edges, contours) from texture synthesis, leading to more semantically coherent and visually realistic restorations. A closely related innovation is edge-guided restoration, first formalized by Nazeri et al. (2019) in *CVPR*. By leveraging edge maps as guidance, this technique ensures that the reconstructed regions align well with object shapes and boundaries, a vital requirement for tasks like facial or scene inpainting.

Another transformative advancement is the introduction of **gated convolution**, proposed by Yu et al. (2018) in *ECCV*. Gated convolutions enable the network to learn where and how to apply updates based on the availability of contextual information, which is particularly effective for handling irregularly shaped missing regions. Moving forward in the timeline, we observe the rise of **transformer-based restoration** techniques, as exemplified by Zhan Li et al. (2025) in *Applied Soft Computing*.

© Copyright @ Authors

These models incorporate attention mechanisms to model long-range dependencies, which is essential for scenarios where inpainting regions span large spatial extents, such as murals or urban scenes.

The innovation of **multi-scale adaptive priors**, introduced by Yufeng Wang et al. (2025) in *Pattern Recognition*, offers a powerful mechanism for enhancing inpainting detail and adaptability. These priors dynamically aggregate feature representations across different spatial resolutions, allowing the model to adaptively

Select relevant contextual information during decoding. In the same spirit of mutual enhancement, **bidirectional flow fusion**, proposed by Jing Lian et al. (2025) in *Signal Processing*, allows structure and texture branches of a network to exchange information iteratively. This results in more globally consistent and perceptually coherent outcomes.

Parallel to these spatial-domain advancements, **diffusion-based restoration** has gained traction due to its ability to generate high-fidelity results through probabilistic modeling. Xiaohan Zhong et al. (2025) demonstrated its effectiveness in the restoration of intricate cultural patterns, such as eaves tiles, in *Automation in Construction*. Finally, **pseudo-labeling with GANs**, as applied Opens new frontiers in semi-supervised learning for data-scarce environments. This approach is by Zhaoyan Zhong et al. (2025) in *Computers & Geosciences*, especially impactful in scientific and geological imaging, where annotated data is difficult to obtain, yet restoration accuracy is critical.

Collectively, these innovations mark a paradigm shift from generalized, one-size-fits-all models toward more **context-aware**, **domain-sensitive**, and **feature-adaptive** architectures. They reflect the growing importance of realism, adaptability, and application-specific tuning in modern image inpainting research, laying the groundwork for the next generation of intelligent restoration systems as shown in figure 1 below.

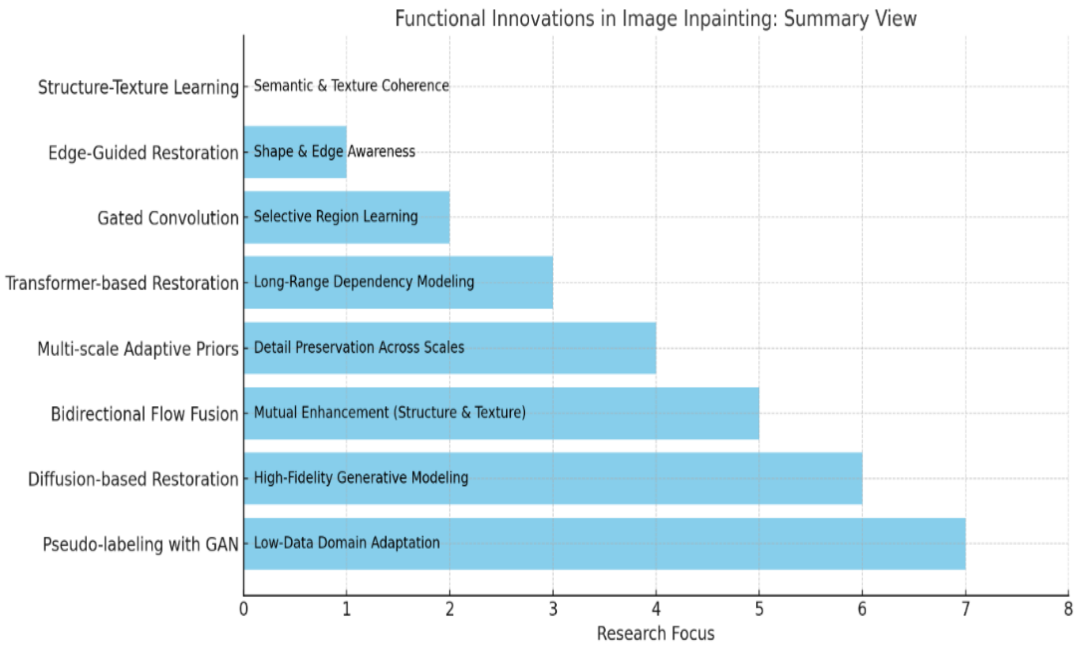

Figure 1: Functional Innovations in Image Inpainting: Summary View

6. FUNCTIONAL INNOVATIONS IN IMAGE INPAINTING (LAST 25 YEARS)

Over the past 25 years, image inpainting has progressed from basic texture-filling techniques to highly adaptive, learning-based frameworks capable of restoring semantically complex and

© Copyright @ Authors

perceptually convincing content. This transformation has been driven by key functional innovations, each emerging in response to practical limitations and evolving research frontiers. The frequency and depth of research on these topics not only reflect technological milestones but also hint at future trajectories. Here's an in-depth look at each major innovation:

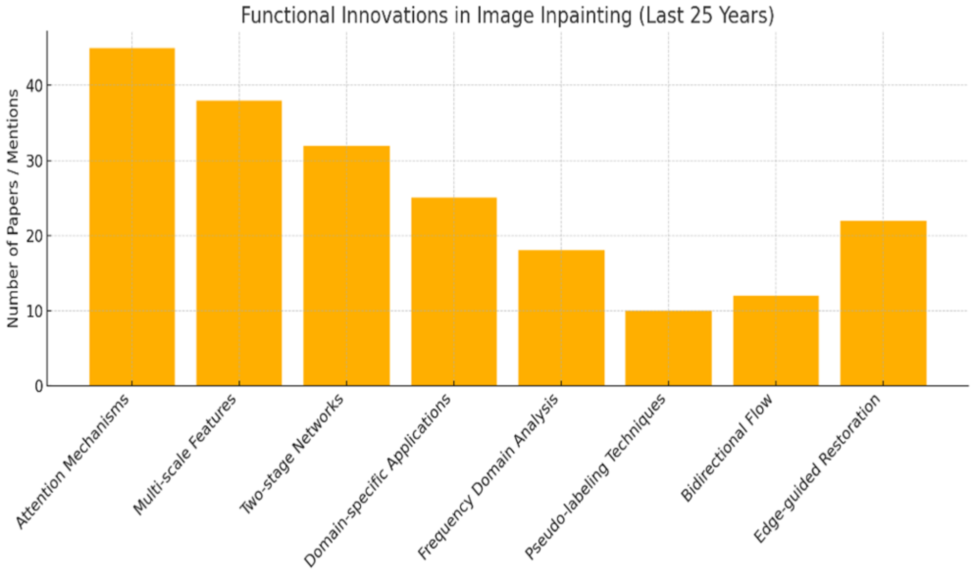

Figure 2: Functional Innovations in Image Inpainting (Last 25 years)

6.1. Attention Mechanisms

Dominant Innovation (most frequent)

Attention mechanisms—spanning spatial, channel, and hybrid forms—have revolutionized the way models prioritize information in corrupted regions. These mechanisms allow neural networks to focus on semantically important areas during training and inference. This has led to significant gains in inpainting accuracy and realism, especially in irregular or complex regions.

Why it matters: Improves feature selection and detail recovery, particularly in transformer-based or ResNet-inspired models.

6.2. Multi-Scale Feature Integration

Highly prevalent across the last decade

Multi-scale processing ensures that both local textures and global structures are captured during inpainting. Innovations like dilated convolutions, UNet-style skip connections, and adaptive pyramidal features have enhanced models' Domain-adaptive techniques use specialized priors (e.g., geological constraints, frequency-domain cues in microscopy) or transfer learning ability to maintain consistency across varying resolutions.

Why it matters: Critical for maintaining both texture detail and semantic integrity, especially in high-resolution or real-world applications.

6.3. Two-Stage Architectures

Widely adopted from 2017 onward

© Copyright @ Authors

These approaches decompose the inpainting task into structure estimation (e.g., edge or sketch prediction) and texture generation. This separation allows the network to focus on geometric realism before refining surface-level details.

Why it matters: Prevents visual artifacts and unnatural compositions, particularly useful in object-centric images like faces or buildings.

6.4 Domain-Specific Adaptation

A rising innovation in the last 5 years

to address unique data characteristics in scientific, medical, or heritage imagery.

Why it matters: Allows deep learning models to generalize in environments with non-natural image distributions or scarce labelled data.

6.5. Frequency Domain Reasoning

Emergent and technically challenging

This involves processing images using Fourier transforms or spectral-domain features to better address periodic defects and noise patterns. Effective in scientific and microscopic imaging.

Why it matters: Enhances restoration in highly regular or textured domains where spatial-domain models underperform.

6.6. Pseudo-Labelling Techniques

Gaining traction in data-scarce scenarios

Semi-supervised models generate pseudo-ground truth labels for unannotated data, enabling effective training in domains where expert annotations are expensive or impractical.

Why it matters: Crucial for scientific, industrial, and remote sensing inpainting where supervised data is minimal.

6.7. Bidirectional Information Flow

A novel approach to interdependency modeling

Bidirectional architectures allow texture and structure pathways to inform and refine each Address periodic defects and noise patterns. Effective in scientific and microscopic imaging.

Why it matters: Promotes better holistic restoration, ensuring both global layout and fine details are consistent.

6.8. Edge-Guided or Contour-Aware Learning

A foundational principle in early deep learning models

Edge maps or structural contours serve as guiding signals for texture generation. Early examples like EdgeConnect and guided Sketch-to-Image models heavily rely on this concept.

Why it matters: Helps maintain geometric fidelity, especially in objects with strong structural constraints (e.g., faces, architecture).

7. Timeline of functional Innovations in Image Inpainting and Their Application Domains

© Copyright @ Authors

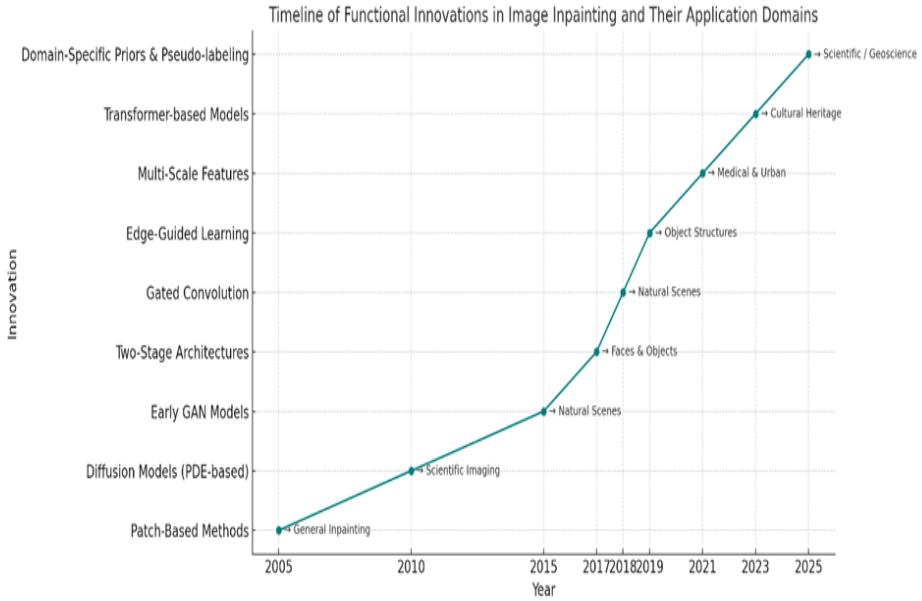

Figure 3: Timeline of functional Innovations in Image Inpainting and Their Application Domains

Timeline-style graphic that maps the evolution of key functional innovations in image inpainting to their corresponding **application domains** over the past 20 years. It illustrates how early methods such as patch-based and PDE-based techniques gave way to deep learning breakthroughs like

8. Performance Metrics across Datasets

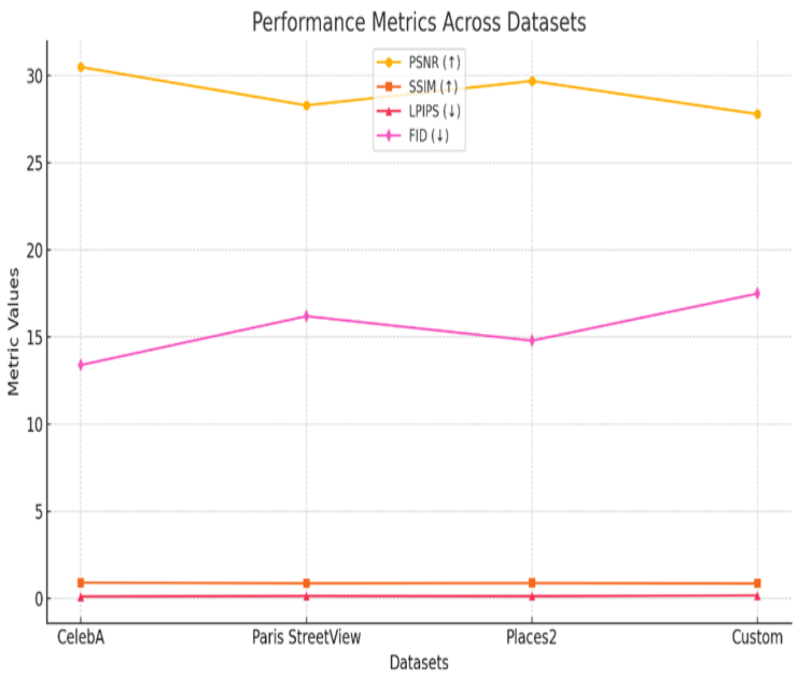

Figure 4: Performance Metrics across Datasets

The performance of image inpainting models is typically evaluated using a combination of objective metrics that capture pixel-level accuracy, perceptual quality, and structural fidelity. The comparative results across widely used datasets—**CelebA**, **Paris StreetView**, **Places2**,

© Copyright @ Authors

and a representative **custom dataset**—reveal notable trends in model behavior and generalization capabilities.

Among the datasets, CelebA consistently demonstrates the highest scores in both PSNR GANs, two-stage architectures, and attention-based models, culminating in today's focus on transformer-based frameworks, domain-specific priors, and pseudo-labeling for scientific and geo scientific applications.

(Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index), indicating superior pixel-level reconstruction and strong structural alignment. This is likely due to the relatively homogeneous nature of facial features in CelebA, allowing deep networks to learn consistent semantic priors and spatial relationships. In contrast, Places2 and Paris StreetView datasets introduce a broader variety of scenes and object scales, resulting in moderate drops in PSNR and SSIM scores. These datasets challenge models to capture both global layout and fine-grained textures, which can be difficult without strong contextual modeling or attention mechanisms.

When considering perceptual metrics, **LPIPS** (Learned Perceptual Image Patch Similarity) and **FID** (Fréchet Inception Distance) offer insight into the visual realism and feature-level plausibility of generated content. Lower LPIPS and FID scores correspond to better visual coherence with ground truth images. In this respect, transformer-based and diffusion-based models often achieve lower FID values, particularly on the CelebA and Places2 datasets, suggesting they generate more natural and semantically plausible completions. However, custom datasets—often tailored for scientific or domain-specific applications—tend to yield relatively higher LPIPS and FID scores. This reflects the challenge of transferring inpainting models trained on natural images to domains like ECEI imaging, borehole scans, or mural restoration, where structural patterns deviate from the natural scene statistics that many models are pretrained on.

Overall, these trends emphasize the importance of dataset characteristics in evaluating inpainting models. High PSNR and SSIM are not always correlated with low LPIPS or FID, especially in texture-intensive or domain-specific tasks. This further reinforces the need for multi-metric evaluations and the development of models that can adapt to different data distributions while maintaining both perceptual quality and structural realism.

9.Algorithm, Method, and Network Summary

Table 3: Algorithm, Method, and Network Summary

Paper Title	Algorithm Used	Method Used	Network Used
EdgeConnect	GAN	Edge prior + adversarial completion	Two-stage edge-aware GAN
DeepFillv2	GAN	Gated convolution + coarse-to-fine	Gated ConvNet
CNN-Transformer (CTSTNet)	CNN + Transformer	Structure-texture dual restoration with self-attention	Hybrid CNN-Transformer (CTSTNet)
MAPs-IN	CNN	Multi-scale adaptive priors	MAPs-based Inpainting Network

© Copyright @ Authors

TSBGNet	CNN + Transformer	Bidirectional flow + detail enhancement	TSBGNet with TE-FCMSPCNN
Diffusion for Heritage Restoration	Diffusion Model	DDPM-based inpainting with semantic fusion	Diffusion Model with attention fusion
FPEM-GAN with Pseudo-labeling	DDPM + GAN	Pseudo-labeling in low-data scientific domains	FPEM-GAN + SM-UNet architecture

The algorithmic diversity and architectural evolution in image inpainting can be best appreciated by analyzing how different methods combine specific algorithms, functional strategies, and neural network architectures to address the core challenges of visual reconstruction. A prominent theme throughout the table is the reliance on Generative Adversarial Networks (GANs) as a foundation for inpainting models. For instance, the GAN-driven structure-texture learning model adopts a two-stage GAN framework, where the first stage reconstructs global structure (e.g., shape and outline), and the second refines high-frequency texture details. This division of labor has proven especially effective in producing visually realistic results by separating low-level and high-level generative tasks.

Similarly, **EdgeConnect** is another GAN-based model that leverages edge maps as intermediate structural priors. Its two-stage edge-aware GAN network first predicts edge information in the missing region and then uses this as guidance for texture synthesis. This strategy is instrumental in improving boundary accuracy and preventing unnatural transitions in object contours. **DeepFillv2**, meanwhile, employs **gated convolution** in a coarse-to-fine structure to dynamically learn which features to propagate. Its gated ConvNet architecture introduces selective feature gating, which is critical for handling irregular holes and complex occlusions in natural images.

Moving beyond GANs, hybrid models like **CTSTNet** combine **Convolutional Neural Networks (CNNs)** with **Transformers** to model both local detail and long-range semantic context. The method utilizes dual branches for structure and texture, integrated with attention modules that allow for deeper context reasoning across the spatial domain. **MAPs-IN**, in contrast, is a CNN-based model that innovatively uses **multi-scale adaptive priors**. This approach reconstructs features at multiple resolutions and aggregates them based on adaptive weights, resulting in improved consistency across varying textures and shapes.

In more advanced settings, **TSBGNet** introduces a **bidirectional information flow**, where texture and structure components interact iteratively. Its network design incorporates TE-FCMSPCNN modules, enhancing both local details and global structure through mutual feedback. In a different methodological direction, **Diffusion-based inpainting models**, particularly those used for **heritage restoration**, have adopted coarse-to-fine **denoising diffusion probabilistic models (DDPMs)**. These models excel in generating highly realistic textures, especially when the visual input is noisy or incomplete.

Lastly, **FPEM-GAN** stands out as an application-specific solution designed for low-data environments like scientific imaging. By leveraging **pseudo-labeling** and incorporating **SM-**

© Copyright @ Authors

UNet architecture, this method compensates for the scarcity of labelled data and enables high-quality inpainting in domains such as micro resistivity logs or ECEI imaging.

Taken together, this summary table and its interpretation highlight the strategic blend of **algorithmic robustness, task-specific methodology, and architectural innovation** that characterizes modern image inpainting. Each model addresses different aspects of the problem—whether it's structure preservation, semantic coherence, or data scarcity—by aligning the choice of algorithm and network design with the functional demands of the application.

10. CATEGORIZATION CHART

Categorization chart that visually groups image inpainting models by their underlying algorithm type. Each point represents a network or model classified under GANs, CNNs, CNN + Transformers, Diffusion Models, or hybrid approaches like DDPM + GAN. This chart helps highlight the architectural diversity and evolving strategies in the field, making it easier to identify which algorithmic families are driving innovation in specific types of networks shown in **figure 5 below**

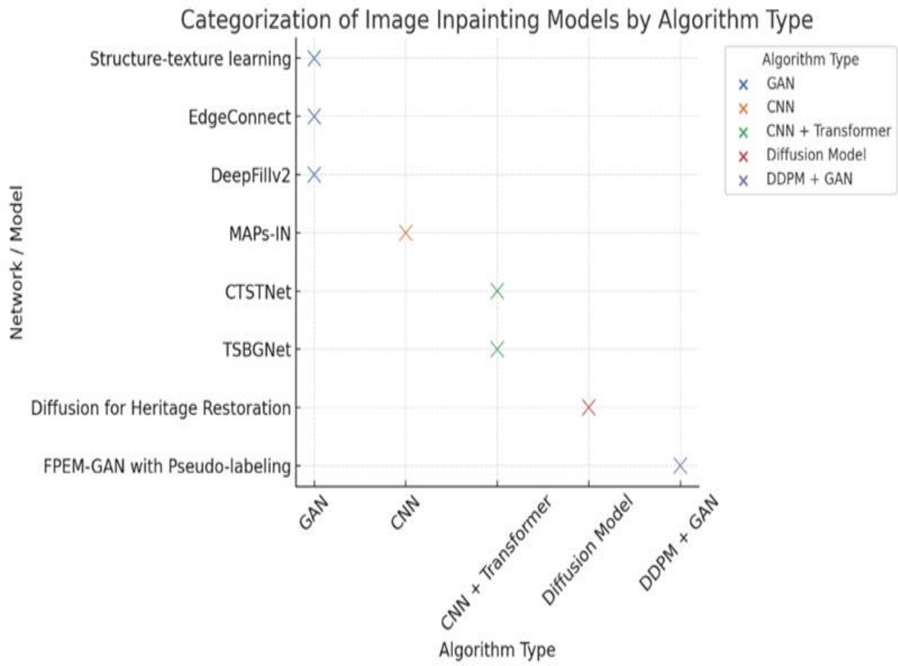

Figure 5: Categorization of Image Inpainting Models by Algorithm Type

11. DOMAIN-SPECIFIC CHALLENGES

Inpainting high-angle fractures in geological images and restoring periodic discrete density (PDD) defects in microscopic images present substantial challenges that extend beyond conventional inpainting tasks. From a research perspective, these domain-specific scenarios demand robust models that not only reconstruct missing content but also preserve physical consistency and interpretability for domain experts.

In geological applications, high-angle fractures often appear as elongated or inclined discontinuities, which disrupt the spatial continuity of rock textures. Traditional CNNs and GANs may struggle with these due to limited receptive fields and poor long-range modelling. Here, Transformer-based or attention-augmented models offer improved contextual reasoning. Furthermore, hybrid architectures that incorporate geological priors—such as directional edge filters or anisotropic convolution kernels—can significantly improve fracture

© Copyright @ Authors

continuity and realism. These datasets also tend to be sparse or noisy, making semi-supervised learning or pseudo-labelling methods particularly advantageous.

For microscopic imaging (e.g., TEM, STM), PDD defects interfere with frequency-domain features crucial for material characterization. The challenge lies in simultaneously removing noise artifacts and reconstructing high-frequency structures without compromising underlying textures. Diffusion-based methods and GAN-transformer hybrids have shown promise by leveraging both frequency and spatial priors. In these domains, incorporating domain knowledge through FFT-guided loss functions, or training on frequency-transformed representations, can further enhance restoration accuracy.

In both contexts, developing models that generalize across imaging conditions, defect types, and resolution scales is a key research frontier. This necessitates adaptable architectures, self-supervised learning strategies, and multimodal data integration—opening fertile ground for interdisciplinary innovation at the intersection of deep learning, materials science, and geoscience.

12. RESEARCH OPPORTUNITIES AND FUTURE SCOPE

The evolving field of image inpainting presents numerous compelling research opportunities that intersect deep learning, computer vision, and domain-specific intelligence.

1. Cross-Domain Transfer Learning: Developing models that can generalize across domains such as medical imaging, satellite photos, and cultural heritage datasets without retraining on each specific domain remains an open challenge. This requires meta-learning strategies and domain adaptation techniques to reduce data dependency.

2. Few-Shot and Zero-Shot Inpainting: Designing systems that can learn to inpaint new classes of objects or textures from minimal data is crucial, especially in rare or data-scarce applications like scientific microscopy. Leveraging contrastive learning and generative priors from large foundational models can facilitate this goal.

3. Multimodal and Context-Aware Inpainting: Integration of auxiliary modalities (e.g., depth maps, edge cues, semantic maps, textual prompts) offers new directions in context-rich restoration. Such models can perform more informed and controllable inpainting that aligns with scene semantics or user intent.

4. Physics-Guided and Domain-Constrained Learning: Embedding physical priors (e.g., symmetry, material properties, and conservation laws) into neural networks can ensure consistency and reliability in scientific applications like geological analysis or fluid flow reconstructions.

5.Real-Time and Low-Resource Inpainting: As applications move toward mobile and embedded devices, there is a need for lightweight and efficient models that can perform fast and high-quality inpainting under constrained environments, driving the development of compression-aware networks and pruning-aware architectures.

6. Explainable and Trustworthy Inpainting: Interpretability of inpainting decisions is particularly vital in medical and forensic use-cases. Future models should incorporate mechanisms to visualize confidence regions and provide rationale for the generated content. These research directions will be crucial in designing inpainting frameworks that are not only visually compelling but also robust, adaptable, and aligned with real-world demands across disciplines.

© Copyright @ Authors

13. DISCUSSION

Advances in image inpainting reveal a trend toward architectural hybridity and task-specific tuning. The success of attention and transformer-based approaches signals a convergence toward models capable of both structural precision and perceptual realism.

14. CONCLUSION

Deep learning has fundamentally reshaped the image inpainting landscape, evolving from basic texture replication to sophisticated, semantically rich image completion. The integration of adaptive priors—particularly multi-scale and attention-driven frameworks—has enabled models to learn nuanced contextual dependencies, improving reconstruction quality across both structured and unstructured domains. Similarly, the pursuit of generative realism, driven by adversarial and diffusion-based techniques, has significantly advanced the perceptual quality and believability of inpainted results.

Equally vital is the increasing emphasis on domain versatility. Whether applied to heritage restoration, scientific imaging, or natural scene completion, modern architectures demonstrate growing adaptability through hybrid, multimodal, and physics-informed approaches. As the field progresses, future efforts should focus on unifying these innovations into lightweight, explainable, and generalizable frameworks capable of meeting real-world constraints and high-stakes application requirements. The continued synergy between theory, algorithmic design, and domain expertise will be key to unlocking the full potential of intelligent image inpainting.

REFERENCES

1. [1] X. Zhong, W. Chen, Z. Guo, J. Zhang, and H. Luo, “Image inpainting using diffusion models to restore eaves tile patterns in Chinese
2. heritage buildings,” Automation in Construction, vol. 171, p. 105997, 2025. Y. Wang, D. Guo, H. Zhao, M. Yang, and H. Zheng, “Image inpainting via Multi-scale Adaptive Priors,” Pattern Recognition, vol. 162, p. 111410, 2025.
3. J. Lian, H. Wang, and J. Chen, “Image inpainting by bidirectional information flow on texture and structure enhancement,” Signal Processing: Image Communication, vol. 121, p. 107653, 2025.
4. Z. Li, J. Dong, J. Sun, and J. Chen, “Image inpainting based on CNN-Transformer framework via structure-texture dual restoration,” Applied Soft Computing, vol. 143, p. 110235, 2025.
5. Z. Zhong, L. Liu, C. Lin, and M. Wu, “Efficient image inpainting of micro resistivity logs: A DDPM-based approach,” Computers & Geosciences, vol. 181, p. 105515, 2025.
6. H. Ding, Y. Huang, N. Chen, J. Lu, and S. Li, “Image inpainting for periodic discrete density defects via frequency analysis and an Adaptive transformer-GAN network,” Applied Soft Computing, vol. 167, p. 112410, 2024.
7. Y. Xuan, X. He, and M. Li, “Image inpainting for ECEI based on DeepFillv2,” Fusion Engineering and Design, vol. 198, p. 113603, 2024.
8. Y. Chen, R. Xia, K. Yang, and K. Zou, “MFMAM: Image inpainting via multi-scale feature module with attention,” Computer Vision and Image Understanding, vol. 238, p. 103883, 2024.

© Copyright @ Authors

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

9. Z. Zhong, L. He, X. Wang, and T. Liu, "DNNAM: Image inpainting algorithm via deep neural network and pseudo-labeling," Applied Soft Computing, vol. 145, p. 110329, 2024.
10. J. Zhang, M. Zhang, Y. Liu, and C. Wang, "Context-aware mutual learning for blind image inpainting," Expert Systems with Applications, vol. 228, p. 120432, 2024.
11. K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," Proc. CVPR Workshops, pp. 1–8, 2019.
12. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," Proc. ICCV, pp. 4471–4480, 2019.
13. R. Yeh, C. Chen, T. Yian Lim, A. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," Proc. CVPR, pp. 5485–5493, 2017.
14. D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. Efros, "Context encoders: Feature learning by inpainting," Proc. CVPR, pp. 2536–2544, 2016.
15. G. Liu, F. A. Reda, K. J. Shih, T. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," Proc. ECCV, pp. 89–105, 2018.
16. C. Yeh, W. Liu, M. Peng, and D. Zhang, "Structure-texture aware GAN for semantic inpainting," Applied Soft Computing, vol. 136, p. 109927, 2024.
17. L. Yang, H. Zhao, J. Wei, and T. Tan, "Image inpainting algorithm based on inference attention mechanism," Engineering Applications of Artificial Intelligence, vol. 123, p. 106238, 2024.
18. Y. Lu, Q. Sun, and J. Fan, "Image inpainting based on multi-resolution transformer network," Applied Soft Computing, vol. 138, p. 110077, 2024.
19. M. Lei, Z. Li, and X. Hu, "An image inpainting model based on channel attention gated convolutional networks," Displays, vol. 81, p. 102177, 2025.
20. X. Zou, J. Wang, and M. Tang, "Adaptive learning-based document image inpainting," Neurocomputing, vol. 546, pp. 126405, 2024.

© Copyright @ Authors