International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 10 | Oct 2025 www.irjet.net p-ISSN: 2395-0072

Image Inpainting and Outpainting using Deep Learning: A Survey

Kindipsingh Mallhi[1], Aditi Chhajed[2], Ojas Binakye[3], Priyansh Katariya[4] and Prof. Pramila M. Chawan[5]

[1,2,3,4] *B.Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India.*
[5] *Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India.*

Abstract - *Image inpainting and outpainting are crucial problems in computer vision, with applications ranging from restoration of historical artworks to modern-day photo editing, privacy preservation, and medical imaging. Traditional approaches such as diffusion and patch-based methods often fail to preserve semantic consistency for large missing regions. Recent advancements in deep learning, especially Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Partial Convolutional Networks (PConv), have revolutionized this domain. This paper surveys existing approaches, discusses their merits and limitations, and positions our proposed work of applying PConv for image inpainting and outpainting as a robust and semantically consistent solution.*

Key Words: **Image Inpainting, Image Outpainting, Deep Learning, CNN, GAN, Partial Convolution, Image Restoration**

1. INTRODUCTION

1.1 Brief overview of image inpainting and outpainting

Image inpainting refers to the process of reconstructing lost or deteriorated parts of an image so that the restored regions are visually plausible and semantically coherent. Outpainting extends an image beyond its original boundaries while maintaining contextual consistency. Both problems are of immense significance in fields like digital heritage restoration, entertainment, e-commerce, and medical imaging.

Early traditional methods relied on pixel diffusion and patch-based texture synthesis. While these approaches were widely used, they were computationally expensive and lacked semantic awareness, often producing unrealistic or repetitive patterns. The advent of deep learning introduced powerful alternatives such as Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Partial Convolutional Networks (PConv), which leverage large-scale datasets and semantic feature learning to produce realistic reconstructions.

This survey consolidates the progress across these methods, evaluates their strengths and limitations, and establishes the research gap for our project focused on PConv for inpainting and outpainting.

1.2 Problem

Image inpainting and outpainting deal with the reconstruction of missing or corrupted regions and the extension of image boundaries while ensuring that the generated content is visually plausible and semantically consistent. Traditional techniques, such as diffusion-based interpolation and patch-based texture synthesis, struggle with large missing regions and lack contextual awareness.

Although deep learning has significantly improved results, challenges remain unresolved. Maintaining semantic consistency across diverse scenes, handling irregularly shaped holes, and preserving fine details without blurriness are still difficult. Moreover, models must generalize across different image types—from natural landscapes to medical scans—while balancing reconstruction quality with computational efficiency. These challenges highlight the need for advanced frameworks such as PConv that can dynamically adapt to missing regions and improve both accuracy and stability.

1.3 Motivation

The motivation for this study arises from the growing demand for reliable image completion across multiple domains. In cultural heritage, it is vital for restoring damaged artworks, wall paintings, and manuscripts. In healthcare, medical imaging often requires reconstructing incomplete scans for accurate diagnosis. In everyday applications such as photo editing, privacy preservation, and digital content creation, users expect seamless object removal, background completion, and boundary extension.

Similarly, industries like entertainment, e-commerce, and gaming rely on visually appealing and contextually accurate image modifications. While CNN and GAN-based models have shown remarkable progress, they still face limitations in stability, handling irregular regions, and computational cost. This necessitates exploration of more efficient and robust alternatives such as PConv.

1.4 Applications

The applications of inpainting and outpainting are widespread and impactful:

Cultural Heritage: Digital restoration of deteriorated wall paintings and manuscripts.

Medical Imaging: Reconstruction of incomplete MRI and CT scans for better diagnosis.

Digital Photography and Media: Object removal, background filling, and image expansion.

Privacy Preservation: Removal of sensitive content without visible artifacts.

E-commerce and Entertainment: Enhancement of product images and generation of immersive environments for gaming and virtual reality.

2. LITERATURE SURVEY

Image inpainting and outpainting are crucial techniques in computer vision for reconstructing or extrapolating missing regions of an image. Traditional exemplar-based and diffusion approaches struggled with structural consistency and semantic accuracy. With the advent of deep learning, new architectures such as CNNs, GANs, and attention-based models have significantly advanced the field.

2.1 Partial Convolution Approaches

2.1.1 Liu et al. (2018) [1]: Proposed the Partial Convolution (PConv) framework, where invalid pixels are masked during convolution. Tested on ImageNet, Places2, and CelebA-HQ datasets, it achieved better performance on irregular holes compared to standard CNNs. However, it struggled with very large missing regions.

2.1.2 Patel et al. (2020) [2]: Developed a modified U-Net with PConv, improving restoration accuracy and reducing loss rates. Evaluated on CelebA-HQ, it demonstrated superior results for face inpainting tasks. Limitation: performance degraded when handling complex background textures.

2.1.3 Yan et al. (2021) [4]: Designed PCNet (PConv with Attention), which combined partial convolution and attention mechanisms. The model showed improved accuracy and faster convergence on Places2 and Paris Street View datasets. Limitation: struggled with fine facial features.

2.1.4 Chen et al. (2021) [6]: Applied PConv with a Sliding Window Strategy to restore Dunhuang murals. This approach yielded faster restoration with good accuracy. However, certain restored regions remained blurred with visible traces.

2.2 Weighted and Gated Convolutions

2.2.1 Kang et al. (2022) [5]: Proposed Weighted Convolutions (WConv) with normalized masks to handle invalid pixels. Demonstrated efficiency on ImageNet and CelebA, solving instability issues in PConv. However, higher loss values were observed in comparison to state-of-the-art models.

2.2.2 Yu et al. (2020) [7]: Introduced Gated Convolutions (GConv) for free-form inpainting. The model dynamically learned masks and improved boundary transitions, tested on Places2 and CelebA-HQ datasets. The drawback was poor performance with very large irregular masks.

2.3 Attention and GAN-Based Approaches

2.3.1 Yu et al. (2019) [3]: Introduced a Contextual Attention + GAN framework. The model exploited distant features through attention layers to fill missing regions. This approach produced visually realistic results on Places and CelebA datasets, but required high computational resources for training.

2.3.2 Chang et al. (2021) [8]: Extended inpainting to videos using 3D-Gated CNN + Temporal GAN loss. Applied on Free-form Video Inpainting datasets, it generated temporally consistent results. However, the approach failed with highly complex masks and thick occlusions.

2.4 Summary

This survey highlights the progression from mask-based CNN approaches to GAN and attention-driven models, addressing structural realism and contextual coherence. While CNN and GAN-based methods provide strong performance, challenges remain in handling large missing regions, complex semantic structures, and computational cost.

Table - 1: Summary of Literature on Image Inpainting and Outpainting using DL

Author(s) & Year	Method / Model	Dataset(s) Used	Metrics	Merits	Demerits
Liu et al. (2018)	Partial Convolution (PConv)	ImageNet, Places2, CelebA-HQ	L1, PSNR, SSIM, Style Loss, TV Loss	Handles irregular holes, reduces artifacts	Quality deteriorates for large missing regions

Patel et al. (2020)	Modified U-Net with PConv	CelebA-HQ	L1, MSE, PSNR, SSIM	Improved accuracy and reduced loss vs. classical PConv	Less effective on large missing structures
Yu et al. (2019)	Contextual Attention + GAN	Places, CelebA	PSNR, Perceptual Loss	Learns long-range dependencies, realistic results	Training is computationally expensive
Yan et al. (2021)	PCNet (PConv + Attention)	Places2, Paris Street View	PSNR, SSIM, Perceptual & Style Loss	Effective for large missing regions, faster convergence	Struggles with complex objects like faces
Kang et al. (2022)	Weighted Convolution (WConv)	ImageNet, CelebA	L1, L2, PSNR, SSIM	Solves instability due to invalid pixels	Higher loss values compared to others
Chen et al. (2021)	PConv + Sliding Window	Dunhuang Mural Dataset	Random Mask, Accuracy	Faster processing,promising accuracy	Some regions blurred with visible traces
Yu et al. (2020)	Gated Convolutions (GConv)	Places2, CelebA-HQ	L1, L2, GAN Loss	Soft mask learning, seamless boundary transitions	Struggles with large masks on faces
Chang et al. (2021)	3D-Gated CNN + Temporal GAN Loss	Free-form Video Inpainting (FVI)	FDI, MSE, LPI	High-quality video inpainting, efficient	Fails on thick/complex masks

3. PROPOSED SYSTEM

A. Problem Statement

The objective of the proposed system is to address the shortcomings of standard image inpainting techniques when dealing with irregular or large missing regions. Classical methods and standard deep learning models often fail to preserve semantic consistency and structural details, especially when the mask geometry is complex.

We propose a system based solely on Partial Convolutional Networks (PConv) for inpainting, which conditions each convolution operation on valid (unmasked) pixels and dynamically updates the mask throughout the network. This approach is designed to robustly reconstruct missing regions while minimizing artifacts and improving adaptation to diverse mask shapes.

B. Architecture Overview

The proposed architecture uses a U-Net-like encoder–decoder framework with skip connections, optimized for processing masked images.

- *Encoder Module:* Each encoder block consists of: Partial Convolutional Layer (PConv2D), which takes both the image and mask as input.
- *ReLU activation:* Dynamic mask update after each partial convolution, expanding the set of valid pixels. The encoder extracts hierarchical features, learning both local textures and global contextual information.
- *Latent Feature Representation:* The bottleneck layer integrates semantic features from valid image areas, enabling the network to learn contextual cues for filling large holes.
- *Decoder Module:* Each decoder block includes: 1. Upsampling (e.g., bilinear interpolation). Skip connections from corresponding encoder layers (for both image and mask).

  2. Partial Convolutional Layer and ReLU activation.
  The decoder reconstructs missing regions, aligning textures and structure with the surrounding context.
- *Skip Connections:* As in U-Net, skip connections allow low-level features to bypass the bottleneck, preserving edge and color details and improving visual sharpness.

The overall design of the proposed system is based on a U-Net-like encoder–decoder architecture with skip connections, tailored for processing masked images. The high-level structure of the model is illustrated in Fig. 1.

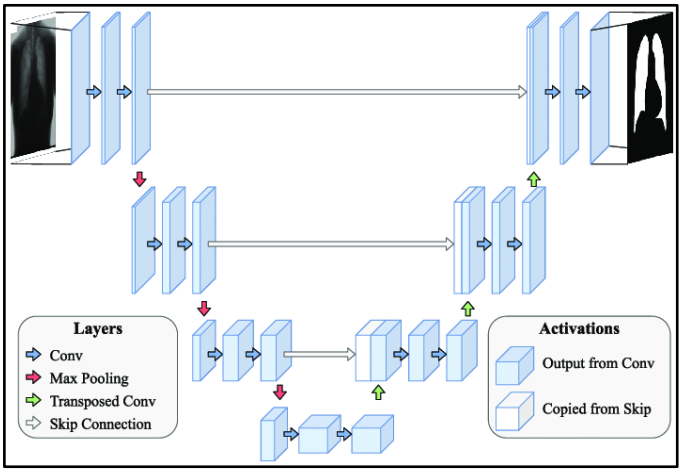

Fig - 1: U-Net Architecture

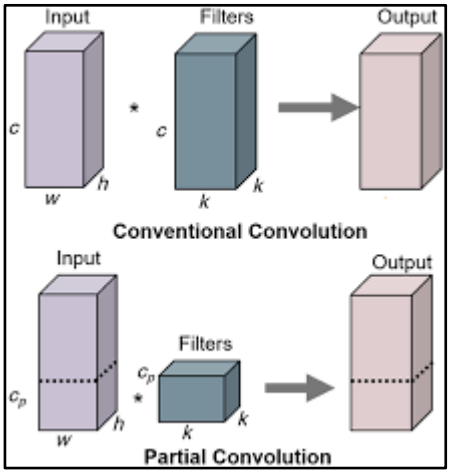

Fig - 2: Partial Convolutions

C. Mask Update Mechanism

The central innovation is the mask-aware partial convolution:

1. Each convolution operates only on valid pixels within the receptive field.
2. After each convolution, the mask is updated: if any valid pixels exist in the receptive field, the output pixel becomes valid; otherwise, it remains masked.
3. This dynamic mask update lets valid pixel regions grow deeper in the network, progressively shrinking the masked area and enabling semantic information to propagate.
4. The core principle of Partial Convolution (PConv) is that each convolution operation is conditioned only on valid pixels, with the mask being dynamically updated at every layer.
5. Fig. 2 illustrates how partial convolution differs from standard convolution and how the mask propagates through the network.

D. Loss Functions

The system is trained with a composite loss function tailored for inpainting:

- *Reconstruction Loss (L1):* Measures pixel-wise similarity between the output and ground-truth image, guiding recovery of fine details.
- *Total Variation Loss:* Reduces high-frequency artifacts and encourages smoothness.

The final loss is a weighted sum to balance local accuracy and global realism. In a basic implementation, pixel-wise L1 loss is sufficient.

E. Training Procedure

- *Dataset*: Training images are corrupted with randomly generated irregular masks, simulating diverse missing regions.
- *Input*: Each training sample consists of a masked image and its binary mask.
- *Training Loop:*
  1. Inputs are normalized and resized.
  2. The network receives the image and mask and outputs the inpainted image.
  3. Loss is computed over masked regions, and the model is updated using Adam optimizer.
- *Augmentation:* Standard image augmentation (flipping, rotation, cropping) can be used for robustness.
- *Batching*: Mini-batch training leverages GPU acceleration.

F. Evaluation Metrics

Performance is assessed using:

- *Peak Signal-to-Noise Ratio (PSNR):* Quantifies pixel-level reconstruction fidelity.
- *Structural Similarity Index (SSIM):* Evaluates perceptual similarity in luminance, contrast, and structure.
- *Qualitative Visual Inspection:* Human assessment of realism, artifact presence, and semantic coherence.

G. System Block Diagram

Proposed Pipeline

1. *Input Module:* Accepts a masked image and binary mask.
2. *Mask-Guided Encoder (PConv Layers):* Extracts features conditioned on valid pixels, dynamically updating the mask.
3. *Latent Feature Representation:* Encodes global context and semantics.
4. *Decoder with Skip Connections:* Reconstructs missing regions using encoder features and mask information.
5. *Loss Computation Unit:* Computes training loss (reconstruction and optionally perceptual/style/TV).
6. *Output Module:* Produces the completed image.

4. CONCLUSION

Image inpainting and outpainting have progressed from traditional diffusion and patch-based methods to advanced deep learning architectures that achieve semantically consistent and visually realistic reconstructions. Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and attention-based models have each enhanced structural coherence, contextual awareness, and perceptual quality. Among these, Partial Convolutional Networks (PConv) stand out for their ability to handle irregular masks and large missing regions by dynamically updating valid pixels during training. Despite these advancements, challenges such as preserving fine details, reducing computational complexity, and achieving generalization across diverse image domains remain unresolved. Outpainting further amplifies these challenges, demanding models that can extrapolate beyond original boundaries while maintaining semantic realism.

To address these limitations, our proposed approach leverages a PConv-enabled U-Net framework that integrates mask-aware feature extraction, skip connections, and tailored loss functions. This design enables sharper, contextually aligned, and artifact-free reconstructions suitable for real-world applications, including cultural heritage restoration, medical imaging, and digital media editing. Looking ahead, research may explore hybrid architectures that combine PConv with transformer-based models, multimodal training with text–image priors, and domain-specific fine-tuning for specialized use cases. Such advancements will push inpainting and outpainting closer to achieving high perceptual fidelity with computational efficiency, thereby enhancing their scalability and adoption in large-scale deployments.

REFERENCES

[1] T. Yu, C. Lin, S. Zhang, S. You, X. Ding, J. Wu, and J. Zhang, “End-to-End Partial Convolutions Neural Networks for Dunhuang Grottoes Wall-painting Restoration,” Dunhuang Academy; Jinan University; Tianjin Medical University; CSIRO; Tianjin University.

[2] J. Susan and P. Subashini, “Deep Learning Inpainting Model on Digital and Medical Images—A Review,” Avinashilingam Institute for Home Science and Higher Education for Women, India.

[3] S. Navasardyan and M. Ohanyan, “Image Inpainting with Onion Convolutions,” Picsart Inc., Yerevan, Armenia.

[4] A. Dash, G. Wang, and T. Han, “Attentive Partial Convolution for RGBD Image Inpainting,” New Jersey Institute of Technology, Newark, USA.

[5] S. S. Singh, A. N. Singh, B. R. Yadav, and K. Jayamalini, “Deep Learning Approach to Inpainting and Outpainting System,” Shree L.R. Tiwari College of Engineering, Maharashtra, India.

[6] C. Jo, W. Im, and S. E. Yoon, “In-N-Out: Towards Good Initialization for Inpainting and Outpainting,” Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea.

[7] C. Li, D. Xu, and H. Zhang, “Multi-stage Image Inpainting using Improved Partial Convolutions.”

[8] G. Liu, F. A. Reda, A. Tao, K. J. Shih, T. C. Wang, and B. Catanzaro, “Image Inpainting for Irregular Holes using Partial Convolutions,” NVIDIA Corporation.

BIOGRAPHIES

Kindipsingh Mallhi, B. Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

Aditi Chhajed, B. Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India.

Ojas Binayke, B. Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India.

Priyansh Katariya, B. Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India.

Prof. Pramila M. Chawan, is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B.E.(Computer Engineering) and M.E.(Computer Engineering) from VJTI College of Engineering, Mumbai University. She has 30 years of teaching experience and has guided 85+ M. Tech. projects and 130+ B. Tech. projects. She has published 148 papers in the International Journals, 20 papers in the National/International Conferences/ Symposiums. She has worked as an Organizing Committee member for 25 International Conferences and 5 AICTE/MHRD sponsored Workshops/STTPs/FDPs. She has participated in 17 National/International Conferences. Worked as Consulting Editor on JEECER, JETR,JETMS, Technology Today, JAM&AER Engg. Today, The Tech. World Editor - Journals of ADR Reviewer -IJEF, Inderscience. She has worked as NBA Coordinator of the Computer Engineering Department of VJTI for 5 years. She had written a proposal under TEQIP-I in June 2004 for 'Creating Central Computing Facility at VJTI'. Rs. Eight Crore were sanctioned by the World Bank under TEQIP-I on this proposal.

Central Computing Facility was set up at VJTI through this fund which has played a key role in improving the teaching learning process at VJTI. Awarded by SIESRP with Innovative & Dedicated Educationalist Award Specialization: Computer Engineering & I.T. in 2020 AD Scientific Index Ranking (World Scientist and University Ranking 2022) - 2nd Rank- Best Scientist, VJTI Computer Science domain 1138th Rank- Best Scientist, Computer Science, India.