Unveiling the Canvas: a Comprehensive Survey of Image Inpainting Techniques

Jasmeet Singh [1], Meenakshi Arora [2], Rohini Sharma [3]

[1] P.G. Student, Department of CSE, Sat Kabir Institute of Technology and Management, Ladrawan, Haryana, India

[2] Assistant Professor, of CSE, Sat Kabir Institute of Technology and Management, Ladrawan, Haryana, India

[3] Assistant Professor, GPGCW, Rohtak, Haryana, India

Abstract:

Image inpainting is a crucial field in computer vision and image processing that focuses on the reconstruction of missing or damaged parts of images in a visually plausible way. This survey provides an in-depth overview of the various techniques and methodologies developed over the years for image inpainting. We categorize the approaches into three primary classes: traditional methods, deep learning-based methods, and hybrid methods. Traditional techniques include diffusion-based, exemplar-based, and patch-based methods. Deep learning approaches encompass convolutional neural networks (CNNs), generative adversarial networks (GANs), and attention mechanisms. Hybrid methods combine traditional and deep learning techniques to leverage the strengths of both. We evaluate these methods based on their performance, computational complexity, and applicability to different types of image inpainting problems. This survey aims to provide a comprehensive understanding of the advancements in image inpainting and identify potential future research directions.

Keywords: Image Inpainting, Computer Vision, Image Processing.

INTRODUCTION: Image inpainting, also known as image completion or hole filling, is the process of restoring missing parts of an image to produce a complete, natural-looking picture. The demand for image inpainting techniques has surged due to their applications in areas such as photo restoration, object removal, and virtual reality[1]. Image inpainting is the process of reconstructing lost or deteriorated parts of an image in a visually plausible way. It involves filling in missing regions or removing unwanted objects while ensuring that the in painted areas blend seamlessly

with the surrounding pixels. This technique is essential in various applications such as photo restoration, object removal, and image-based editing [2]. Image inpainting has its roots in traditional art restoration, where artists would manually restore damaged paintings. The digital counterpart of this process began to develop with the advent of image processing and computer vision technologies. Over the years, image inpainting techniques have evolved from basic diffusion methods to advanced deep learning approaches.

Traditional Methods

1. **Diffusion-Based Methods**: These techniques propagate information from the known parts of the image to the missing regions using partial differential equations (PDEs). Examples include methods proposed by [3] that use image gradients to guide the inpainting process .

2. **Exemplar-Based Methods**: These methods fill in missing areas by copying similar patches from other parts of the image.[4] Developed a well-known exemplar-based inpainting algorithm that effectively handles large missing regions by prioritizing the filling order based on the image structure.

3. **Patch-Based Methods**: These techniques reconstruct missing regions by stitching together patches from the surrounding area. The method by [5] is a notable example that synthesizes textures by sampling patches from the known image regions.

Deep Learning-Based Methods

1. **Convolutional Neural Networks (CNNs)**: CNNs have significantly advanced the field of image inpainting by learning hierarchical features from large datasets. The U-Net architecture, introduced by [6], and encoder-decoder models are widely used for their ability to capture contextual information and generate realistic outputs .

2. **Generative Adversarial Networks (GANs)**: GANs improve inpainting quality by using a discriminator network to differentiate between real and in painted images, thereby pushing the generator to produce more realistic results. The Deep Fill model by [7]is a prominent example that uses GANs for high-quality image inpainting.

3. **Attention Mechanisms**: Attention mechanisms focus on relevant parts of the image, improving the handling of complex structures and fine details. Zhang et al. [7] developed a model that leverages contextual attention for accurate inpainting by attending to the most relevant areas of the image.

Hybrid Methods

Hybrid methods combine traditional and deep learning techniques to leverage the strengths of both. For example, combining patch-based methods with CNNs can improve the coherence of inpainted regions, while integrating exemplar-based techniques with GANs can enhance texture synthesis.

Evaluation Metrics and Datasets

The performance of image inpainting methods is typically evaluated using metrics such as:

➢ **Peak Signal-to-Noise Ratio (PSNR)**: Measures the ratio between the maximum possible power of a signal and the power of corrupting noise.

➢ **Structural Similarity Index (SSIM)**: Assesses the similarity between two images based on luminance, contrast, and structure.

➢ **Frechet Inception Distance (FID)**: Evaluates the quality of generated images by comparing the distribution of generated images to the distribution of real images.

Commonly used datasets for benchmarking inpainting algorithms include:

- ➤ **Paris StreetView**: Urban street scenes.
- ➤ **CelebA**: Celebrity faces dataset.
- ➤ **Places2**: A large-scale dataset with diverse scenes.

Challenges

Despite significant advancements, challenges in image inpainting remain. These include handling high-resolution images, preserving semantic consistency, and reducing computational costs. Future research may focus on improving the robustness of inpainting models, developing more efficient algorithms, and exploring new applications in emerging fields like augmented reality and medical imaging.

RESEARCH BACKGROUND

As computerized digital imaging technology advanced, some earliest conventional algorithm such as partial differential equations [3], sample-based image inpainting models [4], variational inpainting based on geometric image models [8], texture synthesis [9], and data-driven [10] methods were used to successfully repair minor damage such as scratches. A number of deep-learning-based image inpainting techniques [11] have highlighted their benefits in image transformation [12], creating images [13], and image semantic extraction of features [14], thereby encouraging the growth of image inpainting that utilize deep neural networks. These techniques have been made possible by the wide range of computing power and the development of artificial intelligence.

COMPARISON OF TRADITIONAL IMAGE INPAINTING APPROACHES

Table 1: A comparison of Traditional Image Inpainting Approaches

Aspect	Bertalmio et al. [3]	Criminisi et al. [4]	Shen and Chan [8]	Grossauer [9]
Method	Diffusion-Based Inpainting	Exemplar-Based Inpainting	Mathematical Models for Non texture Inpainting	Combined PDE and Texture Synthesis
Approach	Uses partial differential equations (PDEs) to propagate information from the known regions to the missing regions.	Copies similar patches from the known parts of the image to fill in the missing regions.	Uses variational models to inpaint local non texture regions, primarily focusing on edge-preserving and smoothing techniques.	Combines PDE-based diffusion with texture synthesis to handle both structure and texture information.
Strengths	- Effective for small, narrow missing areas. Preserves edge continuity.	- Handles large missing regions well. Good at maintaining texture consistency.	- Suitable for non-texture regions. Preserves edges and smooth transitions.	- Balances between structure reconstruction and texture synthesis. Can handle complex textures and structures.
Weaknesses	- Struggles with large missing areas. May produce blurry results for textured regions.	- Computationally intensive. May produce artifacts in highly structured regions.	- Limited to non-texture regions. Not suitable for complex textured areas.	- Computationally more demanding. Integration of PDE and texture synthesis can be complex.
Applications	- Small-scale image restoration. Edge-preserving smoothing.	- Object removal. Large-scale image inpainting.	- Restoration of non-texture areas. Edge-preserving image smoothing.	- Complex image restoration. Inpainting of images with both structure and texture.
Key Techniques	- PDE-based diffusion.	- Patch-based exemplar method. Priority-based filling.	- Variational models. Edge-preserving techniques.	- Combination of PDE and texture synthesis.
Performance	- High for small areas. Low for large areas with complex textures.	- High for large, textured areas. Can be slow.	- High for edge-preserving non-texture inpainting. Limited for textured areas.	- High for complex scenarios. Computationally intensive.

COMPARISON OF DEEP LEARNING BASED IMAGE INPAINTING APPROACHES

Encoding and decoding are steps in the machine learning method of image restoration. Through the use of an encoder, the damaged image is compressed into a latent space encoding. A recovered image is then created by expanding or decoding this compressed form.

Table 2: A comparison of Deep Learning Based Image Inpainting Approaches

Aspect	Qiang et al., [15]	Elharrouss et al., [16]	Qin et al., [17]	Ul Hassan, [18]
Focus	Survey on various deep learning-based image inpainting methods	Comprehensive review of image inpainting techniques	Review of deep learning-based image inpainting techniques	Classification of images using deep learning
Deep Learning Models Discussed	Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs)	GANs, CNNs, Autoencoders	GANs, CNNs, Attention Mechanisms, Transformer-based Models	AlexNet, CNNs
Key Contributions	Analysis of various models and their performance on different datasets	Detailed review of inpainting techniques with taxonomy and evaluation	Detailed evaluation of state-of-the-art models and future research directions	Demonstrated use of AlexNet for ImageNet classification
Performance Metrics	PSNR, SSIM, L1 loss, L2 loss	PSNR, SSIM, visual quality, computational complexity	PSNR, SSIM, MAE (Mean Absolute Error)	Accuracy, Top-1 and Top-5 error rates
Datasets Used/Discussed	CelebA, Places2, Paris StreetView	CelebA, ImageNet, Paris StreetView, Places2	CelebA, Paris StreetView, Places2, ImageNet	ImageNet
Challenges Addressed	Handling large missing regions, maintaining texture consistency	Filling large holes, edge preservation, semantic consistency	Large missing areas, texture generation, contextual understanding	Classification accuracy, generalization capability
Future Directions	Improved loss functions, hybrid models, unsupervised learning	Enhanced GAN architectures, context understanding, real-time applications	Transformer models, improved context understanding, multimodal learning	Improved network architectures, better training techniques
Strengths	Comprehensive analysis of model performances	Extensive review and categorization of techniques	In-depth review with focus on recent advancements	Demonstrates practical application and effectiveness of AlexNet
Limitations	Mostly focused on GANs and CNNs	Broad review may lack depth in specific methods	Primarily focused on recent techniques, less historical perspective	Limited to classification task, not specific to inpainting

This comparison highlights the various deep learning-based image inpainting approaches and their key characteristics as reviewed in the mentioned papers.

Table 3: A comparison of convolution-based Image Inpainting Approaches

References	[19]	[20]	[21]	[22]	[23]	[6]	[24]	[25]
Method	Dilated Convolutions	Globally and Locally Consistent Image Completion	Generative Image Inpainting with Auxiliary Contextual Reconstruction	Partial Convolutions	Free-Form Image Inpainting with Gated Convolution	U-Net	Inception Architecture	Aggregated Contextual Transformations
Key Contributions	Multi-scale context aggregati	Combined global and local context for	Contextual reconstruction to enhance	Using partial convolutions to	Gated convolutions for improve	U-shaped network architecture for	Improved inception modules for better	High-resolution image inpainting

	on using dilated convolutions	consistent inpainting	inpainting quality	handle irregular holes	d free-form inpainting	biomedical image segmentation	feature extraction	with aggregated contextual transformations
Main Technique	Dilated convolutions for expanding receptive field	Two-stage network: global and local discriminators	Contextual Reconstruction Networks	Mask-aware convolutions allowing irregular hole inpainting	Gated convolution layers controlling the update of features	Encoder-decoder architecture with skip connections	Inception modules with factorized convolutions	Transformer-based network aggregating context
Performance Metrics	PSNR, SSIM, qualitative visual results	PSNR, SSIM, qualitative visual results	PSNR, SSIM, FID, qualitative visual results	PSNR, SSIM, qualitative visual results	PSNR, SSIM, qualitative visual results	Dice coefficient, Jaccard index	Top-1 accuracy, Top-5 accuracy, computational efficiency	PSNR, SSIM, FID, qualitative visual results
Datasets Used/Discussed	Cityscapes, ADE20K, Pascal VOC	Places2, CelebA	Places2, CelebA, ImageNet	Places2, CelebA, Paris StreetView, ImageNet	Places2, CelebA, Paris StreetView	ISBI challenge dataset, LIDC-IDRI dataset	ImageNet, CIFAR-10, CIFAR-100	Places2, CelebA, ImageNet
Strengths	Effective receptive field expansion without increasing parameter count	Maintains global and local consistency	High-quality inpainting with contextual reconstruction	Handles irregular holes effectively	Improves free-form inpainting with gated convolutions	Effective in biomedical image segmentation with precise boundary delineation	High classification accuracy with efficient architecture	High-resolution inpainting with efficient context aggregation
Limitations	Limited to specific tasks, may not handle complex semantics well	Requires two-stage training, can be computationally expensive	May require significant computational resources	Relies heavily on mask quality, may not generalize well to all irregularities	Requires careful design of gated convolutions	Primarily designed for biomedical segmentation, not general-purpose inpainting	Primarily focused on classification, not specifically designed for inpainting	Computationally intensive, may require large datasets
Future Directions	Enhancing context aggregation techniques	Integration of more diverse datasets for training and evaluation	Improving efficiency and reducing computational complexity	Enhancing generalization to various types of image corruption	Combining gated convolutions with other advanced techniques	Adapting U-Net for more general image processing tasks	Expanding inception architecture for more general tasks	Enhancing model efficiency and robustness for real-world applications
Use Cases	Semantic segmentation, context aggregation	Image inpainting in natural images and textures	High-fidelity image completion for various	Image inpainting for natural images with	Free-form image inpainting for artistic	Biomedical image segmentation, medical image	Image classification, feature extraction	High-resolution image inpainting for professional

			application s	irregular masks	and practical applicati ons	analysis		image editing

This table provides a detailed comparison of the various convolution-based image inpainting approaches and their key characteristics.

CONCLUSION

The evolution of image inpainting techniques from traditional methods to sophisticated deep learning-based and convolutional approaches has significantly improved the quality and applicability of inpainting solutions. Traditional methods laid the groundwork but struggled with complex textures and semantic coherence. Deep learning-based approaches, especially those using GANs and attention mechanisms, have shown remarkable improvements in generating realistic inpainting results. Convolution-based methods, particularly dilated, partial, and gated convolutions, represent the latest advancements, addressing specific challenges such as handling irregular holes and maintaining context consistency. These methods combine the strengths of convolutional networks with innovative techniques to enhance the performance and flexibility of image inpainting tasks. Future research directions may focus on further improving the efficiency and generalization capabilities of these models, integrating multimodal learning, and exploring hybrid approaches that combine the strengths of different techniques to push the boundaries of what is achievable in image inpainting.

REFERENCES

1. R. S. Yogita, Shalini Bhadola, Kirti Bhatia, "A Deep Overview of Image Inpainting Approaches," Int. J. Innov. Res. Sci. Eng. Technol., vol. 11, no. 6, pp. 8744–8749, 2022.

2. R. S. Yogita, Shalini Bhadola, Kirti Bhatia, "Development and Analysis of Enhanced Image Inpainting Approach," Int. J. Trend Sci. Res. Dev., vol. 6, no. 4, pp. 1731–1737, 2022.

3. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000, pp. 417–424.

4. A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Trans. image Process., vol. 13, no. 9, pp. 1200–1212, 2004.

5. Efros, Berg, Mori, and Malik, "Recognizing action at a distance," in Proceedings Ninth IEEE International Conference on Computer Vision, 2003, pp. 726–733.

6. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, 2015, pp. 234–241.

7. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5505–5514.

8. J. Shen and T. F. Chan, "Mathematical models for local nontexture inpaintings," SIAM J. Appl. Math., vol. 62, no. 3, pp. 1019–1043, 2002.

9. H. Grossauer, "A combined PDE and texture synthesis approach to inpainting," in Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part II 8, 2004, pp. 214–224.

10. J. Hays and A. A. Efros, "Scene completion using millions of photographs," ACM Trans. Graph., vol. 26, no. 3, pp. 4-es, 2007.

11. J. Li, F. He, L. Zhang, B. Du, and D. Tao, "Progressive reconstruction of visual structure for image inpainting," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 5962–5971.

12. C. Meng et al., "Sdedit: Guided image synthesis and editing with stochastic differential equations," arXiv Prepr. arXiv2108.01073, 2021.

13. A. Nichol et al., "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," arXiv Prepr. arXiv2112.10741, 2021.

14. L. Liao, J. Xiao, Z. Wang, C.-W. Lin, and S. Satoh, "Guidance and evaluation: Semantic-aware image inpainting for mixed scenes," in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16, 2020, pp. 683–700.

15. Z. P. Qiang, L. B. He, X. Chen, and D. Xu, "Survey on deep learning image inpainting methods," J. Image Graph., vol. 24, no. 3, pp. 447–463, 2019.

16. O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," Neural Process. Lett. vol. 51, pp. 2007–2028, 2020.

17. Z. Qin, Q. Zeng, Y. Zong, and F. Xu, "Image inpainting based on deep learning: A review," Displays, vol. 69, p. 102028, 2021.

18. M. ul Hassan, "Alexnet imagenet classification with deep convolutional neural networks." 2018.

19. F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv Prepr. arXiv1511.07122, 2015.

20. S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," ACM Trans. Graph., vol. 36, no. 4, pp. 1–14, 2017.

21. Y. Zeng, Z. Lin, H. Lu, and V. M. Patel, "Cr-fill: Generative image inpainting with auxiliary contextual reconstruction," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 14164–14173.

22. G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 85–100.

23. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 4471–4480.

24. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

25. Y. Zeng, J. Fu, H. Chao, and B. Guo, "Aggregated contextual transformations for high-resolution image inpainting," IEEE Trans. Vis. Comput. Graph. 2022.