# Filling the Gaps: Context-Aware Image Inpainting via Diffusion


Shashank G. Sharma
M.Tech, Computer Science and Engineering (AI)
IIIT Delhi
Roll No: MT23088
Email: shashank23088@iiitd.ac.in

Manoj Telrandhe
M.Tech, Electronics and Communication Engineering (ML)
IIIT Delhi
Roll No: MT23176
Email: manoj23176@iiitd.ac.in



***Abstract*—This project aims to develop a model for image inpainting, which involves filling missing or corrupted parts of an image using diffusion models while preserving structural and contextual coherence. The proposed approach ensures high-quality image reconstruction and realistic restoration. The model will be evaluated on different datasets and compared against existing inpainting models using both quantitative and qualitative evaluations.**



***Index Terms*—Image Inpainting, Diffusion Models, Denoising, Image Restoration, Generative Models, Deep Learning**


## I. Introduction

Image inpainting is a critical task in computer vision, focusing on filling missing or corrupted parts of an image. While traditional methods often struggle with maintaining visual quality, diffusion models offer a promising solution by iteratively refining noisy inputs back to clear images. This project leverages diffusion models for inpainting, ensuring realistic reconstruction by conditioning on uncorrupted regions of the image. The model will be evaluated using metrics like PSNR and SSIM, comparing its performance with existing inpainting techniques.


Fig. 1. Illustration of the inpainting task: Incomplete images (top) and reconstructed images (bottom). Image taken from [1].

### A. Project Objectives

The main objectives of this project include:

- Developing a diffusion-based model for image inpainting.
- Ensuring seamless blending of inpainted regions with the surrounding areas.
- Experimenting with different noise schedules to improve performance.
- Evaluating the model's performance using PSNR, SSIM, and visual assessments.
- Benchmarking the model against existing inpainting methods.
- Optimizing computational efficiency and model convergence time.

## II. Related Work

Diffusion models have gained significant attention for image inpainting due to their ability to iteratively reconstruct missing content. Lugmayr et al. [2] proposed RePaint, which leverages denoising diffusion models to fill in corrupted regions by conditioning on visible areas, achieving superior results compared to traditional methods. Saharia et al. [3] introduced Palette, a versatile image-to-image diffusion model that performs well across various tasks, including inpainting. Anciukevičius et al. [4] extended diffusion models to 3D reconstruction and inpainting in their work, RenderDiffusion, demonstrating the model's ability to handle large occlusions. Additionally, Lugmayr et al. [1] showcased the use of diffusion models for reconstructing complex textures in large missing areas, reinforcing their applicability to inpainting tasks.

## III. Proposed Methodology

The proposed methodology focuses on leveraging diffusion models for image inpainting, involving the following key steps:

### A. Data Preparation

We will use publicly available datasets like CelebA [5] or Places365 [6], which are suitable for inpainting tasks. CelebA consists of high-quality images of faces, making it ideal for face inpainting, while Places365 includes natural scenes and is useful for more general inpainting tasks. Corruptions, such as random masking or occlusions, will be applied to a portion of the images to simulate missing regions. Preprocessing will include normalization and data augmentation techniques.

### B. Model Architecture

The inpainting model is based on the Denoising Diffusion Probabilistic Model (DDPM) proposed by Lugmayr et al. [1].

The forward diffusion process gradually corrupts an image, $x_0$, into a Gaussian noise, $x_T$, over $T$ steps:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)I),$$

where $\alpha_t$ controls the noise level at each step. The reverse process aims to iteratively denoise the corrupted image by estimating the original image through a neural network, $\epsilon_\theta(x_t, t)$, which predicts the added noise at each step:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)).$$

The model is trained to minimize the difference between the predicted and actual noise using a Mean Squared Error (MSE) loss function:

$$\mathcal{L} = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right].$$

### C. Training Procedure

The model will be trained by corrupting the input image and learning to reverse the process. The loss function will primarily use MSE to reduce the discrepancy between the generated and original images. Additionally, to improve perceptual quality, a loss based on VGG feature space (perceptual loss) may be added:

$$\mathcal{L}_{\text{perceptual}} = \sum_i \|\phi_i(x) - \phi_i(\hat{x})\|,$$

where $\phi_i$ denotes the feature map extracted from the $i$-th layer of a pre-trained VGG network.

### D. Evaluation Metrics

We will evaluate the model using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). These metrics are defined as:

$$\text{PSNR}(x, \hat{x}) = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}(x, \hat{x})}\right),$$

where MAX is the maximum possible pixel value, and

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

where $\mu_x$, $\mu_{\hat{x}}$, $\sigma_x$, and $\sigma_{\hat{x}}$ denote the means and standard deviations of the original and inpainted images, respectively.

### E. Comparative Analysis

Our diffusion-based approach will be benchmarked against GAN-based inpainting models and other state-of-the-art diffusion methods such as RePaint [1]. RePaint leverages a denoising diffusion model for image inpainting, showing superior performance in terms of realistic restorations and context-aware completion, and we will compare our model's performance against it using both quantitative and qualitative metrics.

### F. Optimization and Fine-Tuning

We will explore optimization techniques such as mixed-precision training and noise schedule adjustments to balance quality and computational efficiency. This will involve adjusting the variance schedule $\beta_t$ and minimizing the loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda\mathcal{L}_{\text{perceptual}}.$$

## IV. Implementation Plan and Timeline

The project will follow this tentative timeline:

- **Week 1:** Literature review and dataset preparation.
- **Week 2:** Model design and initial implementation.
- **Week 3:** Set up training pipeline and implement loss functions.
- **Week 4:** Train the model and perform early experiments.
- **Week 5:** Hyperparameter tuning and testing noise schedules.
- **Week 6:** Final evaluation and report writing.

**Shashank's Contribution:** Responsible for dataset preparation, model implementation, training pipeline setup, and hyperparameter tuning.

**Manoj's Contribution:** Focus on literature review, model design, loss function experimentation, and report writing.

## V. Expected Outcomes

We expect to achieve high-quality inpainting results with diffusion models, ensuring seamless blending of missing regions. The model should outperform traditional methods in terms of visual fidelity, evaluated through PSNR, SSIM, and qualitative assessments. Additionally, optimizing noise schedules and hyperparameters will enhance model efficiency and performance.

## VI. Conclusion

This proposal aims to develop a diffusion-based model for image inpainting, focusing on seamless restoration of missing regions with high visual fidelity. By experimenting with noise schedules and optimizing hyperparameters, we seek to improve both the quality and efficiency of the inpainting process. The results will contribute to advancing diffusion models in practical applications of image restoration.

## References


[1] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Image inpainting with denoising diffusion models," *OpenReview*, 2021. [Online]. Available: https://openreview.net/pdf?id=17YbAlc1tW

[2] ——, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 461–11 471. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Lugmayr_RePaint_Inpainting_Using_Denoising_Diffusion_Probabilistic_Models_CVPR_2022_paper.html

[3] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," *arXiv preprint arXiv:2111.05826*, 2021. [Online]. Available: https://arxiv.org/pdf/2111.05826

[4] T. Anciukevičius, Z. Xu, M. Fisher, P. Henderson, H. Bilen, N. J. Mitra, and P. Guerrero, "Renderdiffusion: Image diffusion for 3d reconstruction, inpainting and generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 12 608–12 618. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2023/html/Anciukevicius_RenderDiffusion_Image_Diffusion_for_3D_Reconstruction_Inpainting_and_Generation_CVPR_2023_paper.html

[5] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3730–3738.

[6] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2017.