Title

Patch-Based Few-Shot Diffusion Models for Data-Efficient Generative Learning

Overall Summary from Papers

Recent advances in diffusion-based generative models have demonstrated superior performance over GANs and VAEs in modelling complex image distributions [1,2]. However, the reviewed literature consistently highlights a critical limitation: diffusion models require large-scale datasets and extensive computational resources to achieve stable training and high-quality generation [1,3]. This constraint becomes particularly problematic in low-data regimes, where training samples are scarce, expensive, or difficult to acquire.

Several studies have explored data-efficient diffusion strategies, including few-shot diffusion learning, architectural simplifications, and conditional priors [5,7]. Parallel work on patch-based learning demonstrates that modelling local image statistics can significantly improve sample efficiency while reducing memory and computational demands [4,6]. These papers collectively suggest that diffusion models do not inherently require full-image representations to learn meaningful distributions, and that local structures often dominate generative performance.

Despite these insights, existing methods typically treat few-shot learning and patch-based modelling as separate directions [6,7]. The reviewed papers reveal a gap in unified frameworks that explicitly combine patch-level diffusion modelling with limited-data training strategies. While patch-based diffusion approaches have shown faster convergence and reduced resource usage [6], their potential for addressing extreme data scarcity remains underexplored. This body of work motivates a cohesive approach that leverages patch-based representations as an implicit data amplification mechanism within few-shot diffusion settings, addressing both computational efficiency and data limitations simultaneously.

Problem Statement & Objectives

Problem Statement

The primary research problem identified from the reviewed literature is the inefficiency of diffusion models in low-data regimes [1,7]. Diffusion-based generative models struggle to generalise when trained on limited datasets due to their reliance on learning high-dimensional data distributions through iterative denoising processes [1,3]. This results in overfitting, unstable convergence, and degraded sample quality. These challenges are particularly evident in domains such as medical imaging and scientific data analysis, where datasets are inherently small [8,9].

This issue is significant for deep learning and computer vision because many real-world domains—such as medical imaging, scientific visualisation, and specialised industrial inspection—cannot provide large annotated datasets [8]. The reviewed papers demonstrate that while diffusion models offer strong theoretical advantages in modelling complex distributions, current implementations are not well-suited for extremely low-data scenarios [7]. Patch-based learning emerges as a promising direction to reduce dimensionality and improve training stability [6], yet its integration with diffusion models under severe data scarcity remains limited.

Research Objectives

The objectives of this research are:

- To design a patch-based diffusion training framework that improves data efficiency under few-shot learning conditions [6,7].
- To investigate how local patch distributions can serve as effective priors for global image generation in diffusion models [4,5].
- To evaluate whether patch-level diffusion training reduces overfitting and accelerates convergence compared to full-image diffusion [6].
- To assess the quality and consistency of reconstructed images generated from independently synthesised patches.

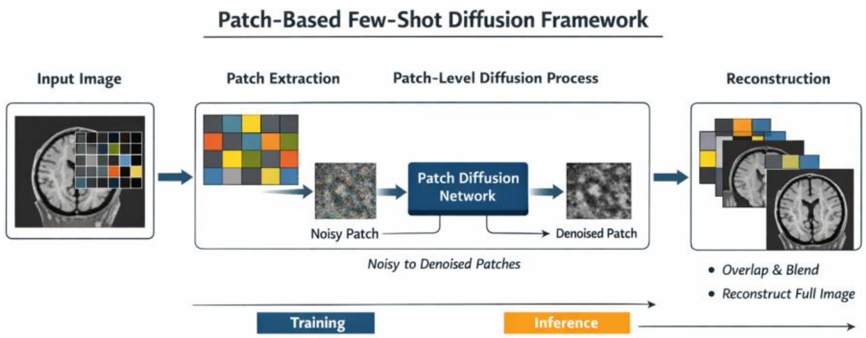

Figure 1: Overview of the proposed patch-based few-shot diffusion framework. Input images are decomposed into local patches, which are independently processed through a diffusion-based denoising model. Generated patches are then reconstructed into full images using overlap-aware blending to ensure spatial consistency. This design is motivated by recent advances in patch-based diffusion and data-efficient generative modelling [1], [6], [7].

Research Methodology

The proposed methodology focuses on training diffusion models at the patch level rather than the full-image level, following insights from recent patch-based diffusion studies [6]. Input images are decomposed into fixed-size patches, which significantly increases the number of effective training samples while reducing spatial dimensionality.

A diffusion process is applied independently to each patch, where Gaussian noise is progressively added and learned to be removed through a denoising network, following standard diffusion objectives [1,2]. Data augmentation is applied at the patch level to further mitigate overfitting in few-shot settings [7]. During inference, synthetic patches generated by the diffusion model are reassembled into full images using overlap-aware reconstruction and blending techniques to ensure spatial consistency.

This approach is selected because it directly addresses the limitations identified in prior work: it reduces computational cost, improves sample efficiency, and exploits local image structure [6]. Alternative approaches, such as full-image fine-tuning or transfer learning, require pretraining on large datasets and are less effective when domain shifts occur [5]. Potential risks include boundary artefacts and loss of global coherence, which will be mitigated through overlapping patches and consistency regularisation.

Expected Results & Novelty

The expected results include improved generative performance in low-data settings compared to conventional diffusion models trained on full images [7]. The proposed method is anticipated to demonstrate faster convergence, reduced overfitting, and competitive image quality despite significantly fewer training samples.

The novelty of this work lies in reframing patch-based learning as a core mechanism for few-shot diffusion training rather than a computational optimisation [6]. By unifying patch-level representation learning with diffusion-based generation, this research introduces a data-efficient paradigm that challenges the assumption that diffusion models require large datasets [5,7]. This shift has the potential to redefine how generative models are trained in data-scarce domains, constituting a meaningful paradigm shift in diffusion-based generative modelling.

References

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in Advances in Neural Information Processing Systems (NeurIPS), 2020.

[2] A. Q. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," in International Conference on Machine Learning (ICML), 2021.

[3] Y. Song et al., "Score-Based Generative Modeling through Stochastic Differential Equations," in International Conference on Learning Representations (ICLR), 2021.

[4] C. Saharia et al., "Image Super-Resolution via Iterative Refinement," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[5] R. Rombach et al., "High-Resolution Image Synthesis with Latent Diffusion Models," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[6] Y. Xu et al., "Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models," arXiv preprint, 2023.

[7] Z. Wang et al., “Few-Shot Diffusion Models,”
arXiv preprint, 2022.

[8] J. Wolleb et al., “Diffusion Models for Medical Image Analysis: A Comprehensive Survey,”
Medical Image Analysis, 2023.

[9] W. H. L. Pinaya et al., “Using Generative Models for Neuroimaging Data Augmentation,”
NeuroImage, 2020.