Application of Latent Diffusion Models (LDMs) in Data-Scarce Scenarios

Kefei Ding*

College of Science, Shanghai University, Shanghai, China

Abstract. The paper mainly focuses on the use of Latent Diffusion Models (LDMs) to address data scarcity, with an in-depth analysis of their practical performance in two important domains: medical imaging and modern industrial manufacturing. As a type of diffusion model, LDMs operate in low-dimensional latent spaces, avoiding the mode collapse issues of traditional Generative Adversarial Networks (GANs) while significantly reducing computational costs. In medical imaging, LDMs aid in generating high-quality, clinically relevant data while respecting privacy constraints; in industrial manufacturing, they support key tasks like enhancing defect detection by supplementing scarce defect samples. The paper further explores core challenges, including the lack of tailored evaluation criteria for LDM-generated images and risks to data privacy, like potential sensitive information leakage. Alongside this, the paper outlines future optimization directions, covering improvements to LDMs' generalization capabilities and the development of more suitable assessment metrics for data-scarce scenarios so as to drive practical applications better.

1 Introduction

1.1 Research Background

A multitude of artificial intelligence generated content (AIGC) models have emerged as a result of the great technological breakthroughs in the field of artificial intelligence in recent years. As those large-scale AI models continue to evolve, with improvements in accuracy, efficiency, and adaptability across diverse aspects, they are poised to exert a profound impact on human society's productivity [1].

Currently, the quantity and caliber of the data set used determine how well the majority of machine learning models function. However, in certain fields, data scarcity arises due to various reasons. Take the field of medical imaging technology as an example, medical data is often restricted by privacy concerns, ethical and moral constraints, and the rarity of diseases as well [2]. In the industrial field, the complexity of data collection procedures also leads to data scarcity.

Traditional methods to address data scarcity, including traditional data augmentation, transfer learning and active learning, have limited effectiveness in complex scenarios. For

* Corresponding author: coffeevol@shu.edu.cn

example, the image generated by data augmentation methods only undergoes geometric transformations but does not alter the core semantics of the image, which may distort the anatomical structure.

Generative models aim to synthesize new data that matches the distribution of real samples, which emerges as a promising alternative. Among these models, Generative Adversarial Networks (GANs) once dominated. However, they encounter issues like mode collapse and training instabilities when data is scarce [3].

Diffusion models (DMs), which are likelihood-based, have shown outstanding performance in image generation and do not exhibit training instabilities or mode-collapse. Latent Diffusion Models (LDMs), as a kind of diffusion model, overcome computational bottlenecks by operating in latent spaces, enabling high-fidelity data synthesis with lower GPU memory than pixel-space diffusion [4]. Therefore, applying LDMs to tackle data scarcity is crucial for boosting large models' rapid learning from datasets and advancing research across fields.

1.2 Research Objective

The paper focuses on the application of LDMs in alleviating data scarcity, mainly analyzing their effectiveness in medical imaging and industrial manufacturing, two fields with relatively severe data limitations. At the same time, the authors will elaborate on the challenges encountered by LDMs in data-scarce environments and their future prospects, aiming to promote the innovative solution of cross-domain data scarcity problems.

2 Basic Theories and Principles of LDM

2.1 Diffusion Model

DMs are probabilistic generative models that simulate the gradual denoising process of data to learn data distributions, consisting of two essential phases: forward diffusion and reverse denoising. With the objective function (1), the training objective of DMs is to reduce the discrepancy between the noise that was actually introduced and the noise that the model anticipated.

$$L_{DM}=E_{x,\ \varepsilon\sim N(0,1),t}[\|\epsilon-\epsilon_\theta(x_t,t)\|_2^2] \quad (1)$$

Where x represents the real input image, $\varepsilon\sim\mathcal{N}(0,1)$ is the added Gaussian noise, t is the step in the diffusion process, x_t is the noisy image at step t , $\epsilon_\theta(x_t,t)$ is the predicted noise, $\| \epsilon-\epsilon_\theta(x_t,t) \|_2^2$ is a specific L_2 loss that quantifies the gap between the predicted and true noise.

2.2 Latent Space Construction

In the latent space construction phase, a pretrained autoencoder transforms images inputted $x \in R^{H\times W\times 3}$ into latent representations $z = \varepsilon(x) \in R^{h\times w\times c}$ with down sampling factor $f = H/h = W/w$. Then the image can be reconstructed from the latent space by the decoder D, giving $\tilde{x} = \mathcal{D}(z)$.

To enable efficient denoising of $\epsilon_\theta(z_t,t)$ in the latent space, LDM adopts a time-conditional UNet built primarily with 2D convolutional layers and reweighted bound. Those designs allow the model to adjust denoising strategies based on diffusion steps, reducing computational costs while preserving spatial details and ignoring redundant information.

Since LDM modifies DM's pixel-space denoising to latent-space denoising, the objective function of LDM can now be stated as:

$$L_{LDM}=E_{\varepsilon(x),\epsilon\sim N(0,1),t}\left[\|\epsilon-\epsilon_\theta(z_t,t)\|_2^2\right] \quad (2)$$

Two kinds of regularization are also introduced to avoid the latent spaces with high arbitrariness. The first is KL-regularization, and the other is VQ-regularization.

2.3 Cross-Attention Mechanism

To enable efficient learning of attention models for multi-modal inputs, LDM integrates a cross-attention mechanism into the UNet backbone, making it a flexible conditional generator. A domain-specific encoder τ_θ converts inputs y into a unified intermediate representation $\tau_\theta(y) \in \mathbb{R}^{M\times d_\tau}$, which is then linked to UNet's latent features via a cross-attention layer. This layer identifies key information by calculating relevance between query (Q) and key (K) vectors, and transmits critical features through value (V) vectors via function (3):

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d}}\right)\cdot V \quad (3)$$

Then the joint objective becomes[4]:

$$L_{cLDM}=E_{\varepsilon(x),y,\epsilon,t}\left[\|\epsilon-\epsilon_\theta(z_t,t,\tau_\theta(y))\|_2^2\right] \quad (4)$$

3 STATE-OF-THE-ART Applications

3.1 Medical Data Scarcity

3.1.1 Medical Image Synthesis

As mentioned earlier, the field of medical imaging faces a problem of data scarcity, which stems primarily from the rarity of some diseases and strict medical privacy policies. Against this backdrop, data augmentation has become an effective solution to help models overcome the scarcity of training set data. However, due to the complex structure of medical images, images generated by simple traditional data augmentation methods may distort anatomical structures, such as rotation and cropping. Meanwhile, the medical imaging field has a clear demand for 3D image data, but the sole use of diffusion models presents issues of excessively high computational Costs and large memory requirements.

Currently, a solution called the slice-by-slice latent diffusion model (SBLDM) has been proposed in the medical imaging domain. Building on the LDM, SBLDM integrates a 2-dimensional variational autoencoder (2D VAE) with a positional embedder to process volumetric data in a slice-by-slice encoding fashion. This approach enables the training of 3D diffusion models while addressing the problems of high computational costs and memory requirements. Additionally, the SBLDM supports precise control over tumor features, allowing it to avoid anatomical structure distortion with high quality even under data-scarce conditions. Table 1 indicates that SBLDM outperforms 3D-LSGAN, a type of deep convolutional generative adversarial network, in two important metrics: structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), with higher scores in both. Although 3D-LSGAN boasts significant efficiency advantages in terms of architectural design and

sampling time, when trained using only a small sample dataset containing 100 images, the details of the generated images become blurred, and their quality degrades noticeably[5].

Table 1. Performance Comparison Table of Different Models.

Method	PSNR	SSIM	params	Sampling time
3D-LSGAN	20.091	0.601	71M	0.002s
3D-LDM	21.034	0.677	728M	13.324s
SBLDM	21.466	0.731	159M	30.900s
SBLDM(DDIM)	21.701	0.726	159M	1.965s

3.1.2 Clinical Data Generation

In the field of clinical medicine data scarcity, LDMs also play a significant role. For instance, in the task of imputing medical datasets, they can fill in missing values in the time series of certain vital sign variables. For instance, when a patient's heart rate or blood pressure data at a specific moment is missing due to missed inspections or equipment failures, the model generates reasonable values to supplement the dataset. Compared with the emphasis on physiological structures when generating medical images, clinical data tends to focus more on the medical logical associations between various physiological variables.

LSSDM (Latent Space Score-Based Diffusion Model) is also a model built on LDM. It introduces Graph Convolutional Networks (GCNs) to process multivariate time series, which can be understood as different vital sign variables, thereby establishing medical logical associations between various physiological variables. Furthermore, it combines Transformer and CNN to obtain coarsely reconstructed missing values, and then inputs these reconstructed values into the diffusion model to generate accurate imputed values[6].

3.2 Industrial Manufacturing

3.2.1 Defect Detection Enhancement

Surface Defect Detection (SDD) is a challenging task that involves identifying samples with defects. Due to the slowness and high cost of human inspection, automated defect detection systems have emerged, which can effectively address these problems by learning classifiers from nominal and defective training samples. However, defective samples are far rarer than normal ones, making data collection difficult. To tackle this, LDMs can be used to generate defect samples that conform to the real distribution, relying on anomaly descriptions provided by experts or a small number of actual defective samples.

4 Challenges and future prospects

In data-scarce applications, the lack of evaluation criteria for LDM-generated data and data abuse risks merit attention. Traditional evaluation criteria like FID and LPIPS measure distribution similarity between generated and real data, but fail to evaluate attribute relationships. While metrics such as SaD and PaD address this issue, they still have limitations, for instance, failing to take into account the problem of attribute importance ranking[7].

Regarding the risk of abuse of generated data, some generative models, for instance, are prone to patient data memorizing in the medical imaging domain. This leads to patient re-identification since the models produce copies of patient data rather than new synthetic samples[8].

Looking ahead, LDM models present substantial potential for optimization in data-scarce domains. First, efforts should be directed toward enhancing their comprehension and generalization capabilities. Some generative models over-memorize original samples, leading to the creation of replicas or even potential privacy violations. It might be a good choice to add noise to raw data to prevent accurate memorization of sample features, or to calculate the feature similarity between produced and original samples while implementing penalty mechanisms to prevent over-memorization. To fully evaluate the adaptation of LDMs in specialized fields, it is also beneficial to create and improve assessment indices that are adapted to data scarcity.

5 Conclusion

Latent Diffusion Models (LDMs) have demonstrated significant application value in data-scarce domains, especially in scenarios such as medical imaging, clinical data generation, and industrial manufacturing defect detection. Through innovative architectural designs, such as the slice-by-slice processing of SBLDM, the conditioning mechanism of BlastDiffusion, and the integration of Graph Convolutional Networks in LSSDM, they have effectively alleviated the training bottlenecks caused by data scarcity in various fields. Meanwhile, they perform better than conventional methods in terms of domain flexibility, computational economy, and generation quality. However, this field still faces challenges such as inadequate evaluation criteria for generated data and risks of data abuse. In the future, it is necessary to further optimize the understanding and generalization capabilities of LDMs, reduce the risk of sample replication through anti-memorization mechanisms, and establish a specialized evaluation system adapted to data-scarce scenarios, so as to promote their reliable application and sustainable development in more professional fields.

References

1. J.S. Hou, Research on few-shot image generation based on diffusion models, MA thesis, Nanjing University of Science and Technology (2024)
2. A. Kazerouni, E.K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, D. Merhof, Diffusion models in medical imaging: A comprehensive survey. Med. Image Anal. **88**, 102846 (2023)
3. A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
4. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10684-10695 (2022)
5. A. Kebaili, J. Lapuyade-Lahorgue, P. Vera, S. Ruan, 3D MRI synthesis with slice-based latent diffusion models: Improving tumor segmentation tasks in data-scarce regimes, in Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, 1-5 (2024)
6. G. Liang, N. Abiri, A.S. Hashemi, J. Lundström, S. Byttner, P. Tiwari, Latent Space Score-based Diffusion Model for Probabilistic Multivariate Time Series Imputation, in Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 1-5 (2025)
7. D. Kim, M. Kwon, Y. Uh, Attribute based interpretable evaluation metrics for generative models. arXiv preprint arXiv:2310.17261 (2023)

8. S.U.H. Dar, M. Seyfarth, I. Ayx, T. Papavassiliu, S.O. Schoenberg, R.M. Siepmann, S. Engelhardt, Unconditional latent diffusion models memorize patient imaging data. Nat. Biomed. Eng. (2025) 1-15