Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.20260354360

Low-Rank Adaptation (LoRA) for Visual Art Style Learning and Feature Fusion: A Narrative Review

Jialang Liu*

School of Information Technology, Shanghai Jianqiao University, China

**Corresponding author: Jialang Liu.*

Abstract

Recent developments in diffusion-based generative models, particularly Denoising Diffusion Probabilistic Models (DDPMs), have significantly advanced the field of Artificial Intelligence Generated Content (AIGC). Despite their strong generalization ability, base models such as Stable Diffusion often lack the precision required for capturing specific artistic styles or niche visual aesthetics. In this context, Low-Rank Adaptation (LoRA) has emerged as an efficient and practical solution for style-specific fine-tuning within Latent Diffusion Models (LDMs). This paper reviews the underlying mechanism of LoRA and examines how low-rank updates interact with attention layers to encode stylistic features. Compared with alternative approaches such as DreamBooth and Textual Inversion, LoRA offers a more balanced trade-off between computational efficiency and stylistic fidelity. In addition, the paper discusses the challenges of multi-style synthesis, with particular attention to the limitations of linear fusion methods and the potential of Dynamic Layer-wise Fusion (DLF). Finally, several future research directions are outlined, including interpretability and automated optimization, which are essential for improving controllability in generative systems.

Keywords

generative AI, Latent Diffusion Models, Parameter-Efficient Fine-Tuning (PEFT), Low-Rank Adaptation (LoRA), artistic style fusion

1. Introduction

The development of image generation models has undergone a clear transition in recent years, moving from Generative Adversarial Networks (GANs) [1] toward diffusion-based approaches such as Denoising Diffusion Probabilistic Models (DDPMs) [2]. Compared with GAN-based models, which are often difficult to train and may suffer from instability or mode collapse [1], diffusion models provide a more reliable framework by gradually learning to reconstruct data from noise [2]. This iterative denoising process has proven to be highly effective in generating high-quality images with strong consistency.

A major advancement in this area was the introduction of Latent Diffusion Models (LDMs) [3], which perform the diffusion process in a compressed latent space instead of directly operating on pixel-level representations. This significantly reduces computational cost while maintaining visual quality, making it feasible to deploy such models on consumer-level hardware [3]. As a result, systems like Stable Diffusion have

rapidly gained popularity in both academic research and creative industries [3]. In addition, recent work such as ControlNet further improves controllability by introducing structural guidance into diffusion models [4].

With the increasing adoption of AIGC technologies, the demand for controllability and personalization has grown. In practical scenarios such as game asset design, digital illustration, and marketing content generation, users often require outputs that follow a specific artistic style rather than generic visuals. However, pre-trained diffusion models are designed to be general-purpose and therefore struggle to capture fine-grained stylistic details, such as brushstroke patterns, lighting characteristics, or culturally specific aesthetics. Early neural style transfer methods [5] laid the foundation for this research direction, but failed to scale to the flexible generation capabilities of modern diffusion models.

Full-model fine-tuning is one approach to address this limitation, but it introduces several challenges. Updating all parameters in a large diffusion model is computationally expensive and time-consuming. More importantly, it can lead to catastrophic forgetting, where the model loses its ability to generate diverse outputs beyond the fine-tuned domain. This trade-off between specialization and generalization remains a key issue in generative modeling.

To overcome these challenges, Parameter-Efficient Fine-Tuning (PEFT) methods have been proposed. These methods aim to adapt large models to specific tasks while modifying only a small subset of parameters. Among them, Low-Rank Adaptation (LoRA) [6] has attracted significant attention due to its efficiency and flexibility. Instead of updating the entire model, LoRA introduces low-rank matrices that capture the essential changes required for a new task [6].

In this review, three main aspects are examined. First, the interaction between LoRA and the Query, Key, and Value matrices in attention layers is analyzed to better understand how stylistic features are encoded, building on prior work on attention control in diffusion models [7]. Second, LoRA is compared with other personalization approaches, with a focus on practical performance and limitations. Finally, the issue of multi-style fusion is discussed, particularly why simple weight combination often fails and how hierarchical strategies such as Dynamic Layer-wise Fusion (DLF) can provide more stable and interpretable results.

2. Mechanism of LoRA in Stable Diffusion

2.1 The Philosophy of Low-rank Adaptation

The core idea behind LoRA is that adapting a model to a new task does not necessarily require modifying all its parameters [6]. In many cases, the essential changes can be represented within a much lower-dimensional subspace. This observation forms the foundation of low-rank adaptation [6].

From a practical perspective, this means that instead of retraining a large model with millions or billions of parameters, one can focus only on a small set of additional parameters that capture the difference between the original task and the new task. This significantly reduces both computational cost and storage requirements.

2.2 Mathematical Framework

In LoRA, the weight update can be approximated as the product of two smaller matrices [6]:

$$\Delta W = B \times A \tag{1}$$

where B has dimensions d × r and A has dimensions r × k. The rank r is much smaller than the original dimensions, which reduces the number of trainable parameters while preserving essential information.

During the forward pass, the output can be written as:

$$h = W_0 \times x + \Delta W \times x \tag{2}$$

This formulation allows the original model weights to remain fixed, while only the low-rank matrices are trained. As a result, LoRA maintains the general knowledge of the base model while introducing task-specific adaptations [6].

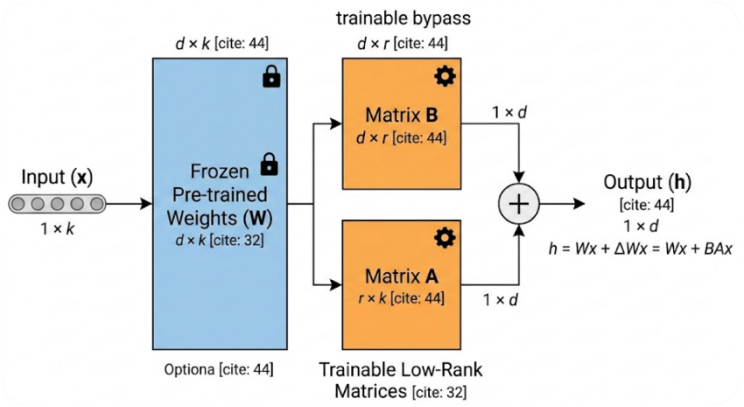
2.3 Role in Attention Layers (Q/K/V)

Attention mechanisms are central to modern diffusion models. In these layers, Query (Q), Key (K), and Value (V) matrices determine how different parts of the input interact with each other [7]. By applying LoRA to these matrices, it becomes possible to adjust how features are selected and combined. In practice, modifying the Query and Value matrices is often sufficient to influence stylistic representation. The Query matrix controls how attention is distributed, while the Value matrix determines the content being emphasized [7]. Through low-rank updates, LoRA effectively reweights internal representations, allowing the model to highlight specific textures, colors, and visual patterns associated with a given style.

2.4 Why Low-rank Works for Style Representation

Although images are high-dimensional, the defining characteristics of a style are relatively compact. Features such as color distribution, stroke patterns, and lighting conditions can often be described using a limited set of underlying rules [5]. This suggests that stylistic variations lie on a low-dimensional manifold within the model's parameter space. LoRA takes advantage of this property by focusing on these essential variations, rather than attempting to modify the entire model. This explains why relatively small rank values can still produce high-quality results.

Figure 1: Structural overview of the Low-Rank Adaptation (LoRA) mechanism in diffusion models



3. Comparative Analysis of Personalization Techniques

3.1 Textual Inversion

Textual inversion is one of the simplest personalization techniques. It works by learning a new embedding that represents a specific concept within the model's existing vocabulary space [8]. This approach is lightweight and easy to implement, making it accessible to a wide range of users. However, its expressive power is limited. Because it relies entirely on the model's existing knowledge, it struggles to capture complex styles or highly specific visual features [8]. In practice, it is more suitable for simple concepts rather than detailed stylistic transformations.

3.2 Dreambooth

DreamBooth takes a more comprehensive approach by fine-tuning the entire model. This allows it to achieve very high fidelity, especially when learning specific subjects such as faces or objects. Despite its effectiveness, DreamBooth has several drawbacks. It requires significant computational resources, including high-end GPUs and large amounts of memory. The training process is also relatively slow. Additionally, there is a risk of overfitting, which may reduce the model's ability to generalize.

3.3 LoRA

LoRA provides a balance between these two approaches. It offers strong performance while maintaining efficiency and flexibility [6]. Because LoRA modules are small and modular, they can be easily combined or reused across different tasks. In practical applications, this modularity is particularly valuable. Artists and developers can build libraries of LoRA models and mix them to achieve diverse visual effects. This makes LoRA a practical tool for real-world creative workflows.

4. Analysis of Multi-LoRA Fusion Strategies

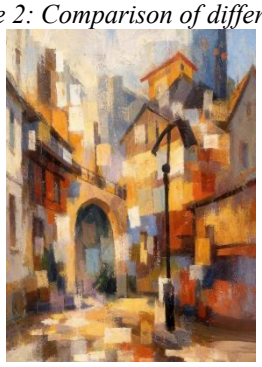
4.1 The Problem of Semantic Blurring

Combining multiple LoRA models is a common technique for generating complex styles. However, simple linear fusion often leads to unsatisfactory results. This issue is commonly referred to as semantic blurring. The main reason for this phenomenon is that neural network parameters interact in non-linear ways. When multiple weight updates are combined directly, they may interfere with each other, leading to inconsistencies in the generated output. As a result, images may appear visually confusing or lack clear structure. This problem becomes more pronounced when combining styles that differ significantly, such as geometric and painterly styles. In such cases, the model struggles to reconcile conflicting representations, a challenge that has been documented in prior style fusion research [5].

4.2 Dynamic Layer-wise Fusion (DLF)

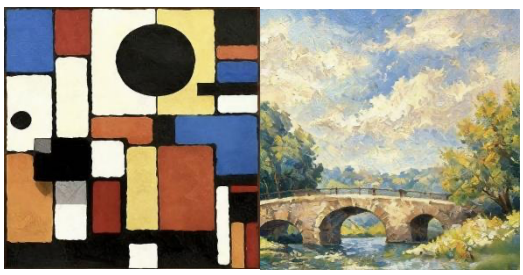
Dynamic Layer-wise Fusion addresses this limitation by introducing a more structured approach to combining LoRA models. Instead of applying the same weights across all layers, DLF assigns different weights to different parts of the network. In diffusion models, lower layers typically control global structure, while higher layers are responsible for fine details. By adjusting these layers independently, DLF allows for better separation between structure and style, leveraging the hierarchical feature learning properties of diffusion models [3]. For example, one LoRA model can be used to define the composition of an image, while another controls texture or color. This results in more coherent outputs and reduces the risk of semantic conflicts.

Figure 2: Comparison of different artistic style fusion strategies and their visual outcomes

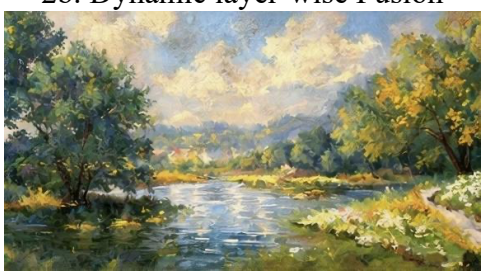


2a. Linear Fusion

2b. Dynamic layer-wise Fusion



Abstract Geometric, Impressionist



2c. Original Style

5. Future Work and Research Directions

5.1 Interpretability

Improving interpretability remains a key challenge. Although LoRA performs well in practice, it is still not entirely clear how individual parameters relate to specific visual outcomes. Future work could focus on clarifying how different components correspond to observable stylistic features.

One possible direction is to combine attention visualization methods to trace how LoRA matrices activate in relation to elements such as brushstroke patterns, color usage, and spatial composition. In addition, cross-modal alignment approaches can be used to measure how parameter updates correspond to semantic descriptions of style, often relying on cross-modal representations such as CLIP [9]. Establishing a clearer link between parameter space and visual output would help reduce the "black box" nature of LoRA-based methods. At the same time, better interpretability may reveal redundant parameters, offering a basis for further model compression without sacrificing core stylistic capabilities. Ultimately, this line of research could also support more precise control, allowing users to directly adjust specific stylistic attributes instead of relying on trial-and-error tuning.

5.2 Automated Optimization

Another promising direction lies in automating the selection of fusion parameters. At present, combining multiple LoRA models often depends on manual experimentation, which can be inefficient and inconsistent. Machine learning-based approaches, such as reinforcement learning, could help identify better configurations in a more systematic way.

For example, Bayesian optimization could be used to construct a surrogate relationship between fusion weights and generation quality, allowing the system to quickly narrow down effective parameter ranges. In multi-style scenarios, multi-objective optimization methods could balance the contribution of different styles and prevent important features from being lost due to poor weight allocation. In addition, lightweight predictive models could be trained to estimate the outcome of different LoRA combinations in advance, providing real-time suggestions to users. Incorporating user feedback into this process would further refine the optimization objectives, aligning results with individual aesthetic preferences. Overall, such an automated pipeline would lower the barrier for non-expert users and make high-quality style fusion more accessible.

5.3 Cross-model Adaptation

As generative models continue to evolve, transferring LoRA modules across different architectures is becoming increasingly important. Improving cross-model adaptability would enhance the reuse of trained modules and reduce redundant training efforts.

Currently, many LoRA modules are closely tied to the parameter structure and attention mechanisms of specific base models, which limits their portability. Future research could explore feature-space alignment techniques, mapping the stylistic representations encoded in LoRA matrices to the feature distributions of new models through linear or nonlinear transformations. Insights from transfer learning in other domains may also help in designing a more general style representation that is less dependent on a particular architecture. At the same time, compatibility frameworks could be developed to automatically detect structural differences between models and adjust parameter dimensions or insertion points accordingly. In the long term, such efforts could support the creation of a shared and standardized LoRA ecosystem, reducing repeated work caused by rapid model iteration.

5.4 Human-ai Collaborative Design

Another worthwhile direction is the integration of LoRA into interactive design systems. Rather than treating generative models as isolated tools, future systems could allow users to guide the generation process step by step. This human-in-the-loop approach would improve both usability and creative control.

For instance, interactive interfaces could provide modular controls for adjusting different stylistic components in real time, with immediate visual feedback during the generation process. When combined with structural guidance tools, users could define the layout or composition while using LoRA to refine stylistic details at a local level. Such systems could be tailored to professional workflows, including game design, animation, and commercial illustration, by offering predefined module combinations and parameter presets. In

addition, recording user interactions and preferences could enable the system to continuously refine its recommendations, forming a feedback-driven creative loop. This would help integrate AI capabilities with human creativity more naturally, positioning LoRA as a flexible assistant rather than a one-click solution. Overall, these directions indicate that future research will not only aim to improve output quality, but also to make generative models more interpretable, controllable, and better suited for real-world creative practice.

6. Conclusion

This paper has examined LoRA as an efficient approach for style adaptation in diffusion models [6]. By focusing on low-rank updates, it provides a practical alternative to full-model fine-tuning while preserving the general capabilities of the base model [3]. In addition, the discussion of fusion strategies highlights the importance of structured approaches for combining multiple styles. Methods such as Dynamic Layer-wise Fusion demonstrate that it is possible to achieve both flexibility and coherence in generative systems, addressing the core limitations of traditional linear fusion methods. As generative models continue to evolve, approaches that balance efficiency, flexibility, and interpretability will play an increasingly important role. LoRA represents a meaningful step in this direction and is likely to remain a key component in future AIGC workflows.

References

[1] I. Goodfellow et al., “Generative adversarial nets,” in Advances in Neural Information Processing Systems, vol. 27, 2014.

[2] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, 2020.

[3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695, 2022.

[4] S. Zhang et al., “Adding conditional control to text-to-image diffusion models,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3836–3847, 2023.

[5] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423, 2016.

[6] E. J. Hu et al., “LoRA: Low-rank adaptation of large language models,” in International Conference on Learning Representations, 2022.

[7] A. Hertz et al., “Prompt-to-prompt image editing with cross attention control,” arXiv preprint arXiv:2208.01626, 2022.

[8] R. Gal et al., “An image is worth one word: Personalizing text-to-image generation using textual inversion,” arXiv preprint arXiv:2208.01618, 2022.

[9] A. Radford et al., “Learning transferable visual models from natural language supervision,” in International Conference on Machine Learning, pp. 8748–8763, 2021.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access