MixLoRA: Parameter-Efficient Style Mixing in Diffusion Models

Consulting & Analytics Club

Abstract

Large diffusion models such as Stable Diffusion are capable of producing high-quality images but are expensive to fine-tune for new styles or concepts. Low-Rank Adaptation (LoRA) enables parameter-efficient fine-tuning by introducing small trainable adapters while freezing the base model. However, standard LoRA is limited to representing a single task or style per adapter.

This project implements **MixLoRA**, a technique that enables the *linear composition of multiple LoRA adapters* at inference time, allowing controllable mixing of artistic styles without retraining the base model. We adapt the MixLoRA research idea to Stable Diffusion and demonstrate multi-style image generation using combinations of Monet, Van Gogh, and Studio Ghibli styles.

1 Introduction

Diffusion-based generative models have become the dominant paradigm for high-quality image synthesis. While powerful, these models are costly to retrain or fine-tune for each new artistic style or domain.

Low-Rank Adaptation (LoRA) addresses this limitation by introducing lightweight rank-decomposed matrices into attention layers, enabling efficient fine-tuning. However, traditional LoRA approaches suffer from a key limitation: **each adapter represents a single isolated capability**. Combining multiple styles typically requires either retraining or destructive merging of weights.

The MixLoRA approach overcomes this by enabling *composable LoRA adapters*, allowing multiple learned adaptations to be mixed using scalar weights at inference time.

2 Related Work

2.1 Diffusion Models

Diffusion models generate data by reversing a noise corruption process. Stable Diffusion combines a latent diffusion process with a frozen text encoder and U-Net backbone, making it a popular foundation for customization.

2.2 Low-Rank Adaptation (LoRA)

LoRA introduces low-rank matrices $A \in \mathbb{R}^{r\times d}$ and $B \in \mathbb{R}^{d\times r}$ such that:

$$W' = W + \alpha BA$$

where W is a frozen weight matrix and only A and B are trained.

2.3 MixLoRA

MixLoRA extends LoRA by allowing multiple adapters $\{(A_i, B_i)\}$ to be combined:

$$W' = W + \sum_i \lambda_i B_i A_i$$

where λ_i controls the contribution of each adapter.

3 System Overview

3.1 Objective

The goal of this project is to:

- Train separate LoRA adapters for different artistic styles
- Load multiple adapters simultaneously into Stable Diffusion
- Perform weighted mixing at inference time
- Generate images that reflect blended stylistic characteristics

3.2 High-Level Pipeline

1. Load pretrained Stable Diffusion model
2. Attach LoRA layers to attention modules
3. Train one LoRA adapter per style
4. Save adapters independently
5. Load multiple adapters during inference
6. Apply weighted composition (MixLoRA)

4 Architecture

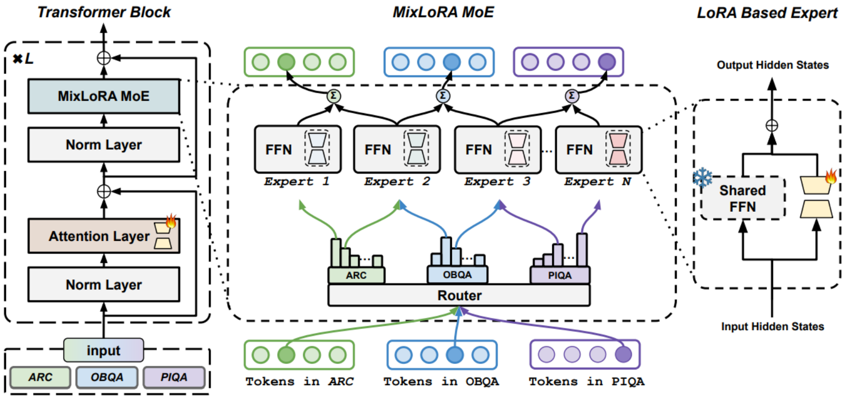

Figure 1: The architecture of MixLoRA transformer block. MixLoRA consists of n experts formed by an original FFN sublayer combined with different LoRAs, where the weights of the FFN sublayer are shared among all the experts.

5 Implementation Details

5.1 LoRA Layer Design

Each LoRA layer is implemented as a low-rank residual update:

1. Base weights remain frozen
2. Trainable matrices A and B are initialized
3. Scaling factor α/r stabilizes training

5.2 Adapter Training

Each style dataset (Monet, Van Gogh, Ghibli) is used to train a separate adapter:

1. Rank $r = 4$
2. Attention layers only
3. Text encoder frozen
4. Small batch size with gradient accumulation

5.3 Mixing Mechanism

During inference, multiple adapters are loaded and combined as:

$$\Delta W = \lambda_1 \Delta W_1 + \lambda_2 \Delta W_2 + \lambda_3 \Delta W_3$$

This allows smooth interpolation between styles without retraining.

6 Experiments

6.1 Experimental Setup

1. Base model: Stable Diffusion v1.5
2. Adapters: Monet, Van Gogh, Ghibli
3. Inference steps: 30
4. Guidance scale: 2.0

7 Discussion

The results demonstrate that MixLoRA enables:

1. Smooth stylistic interpolation
2. Non-destructive adapter composition
3. Reuse of independently trained adapters

However, extreme mixing weights can sometimes introduce artifacts, suggesting that normalization or adaptive weighting may further improve stability.

8 Limitations and Future Work

1. Mixing is linear and does not model nonlinear style interactions
2. No automated method for optimal weight selection

3. Limited evaluation beyond qualitative inspection

Future work could explore:

1. Learned mixing coefficients
2. Dynamic prompt-conditioned weighting
3. Extension to concept + style adapters

9 Conclusion

This project successfully implements MixLoRA for Stable Diffusion, demonstrating that multiple LoRA adapters can be composed at inference time to produce rich, controllable hybrid styles. MixLoRA significantly improves the flexibility and reusability of parameter-efficient fine-tuning methods in generative models.

9.1 Qualitative Results

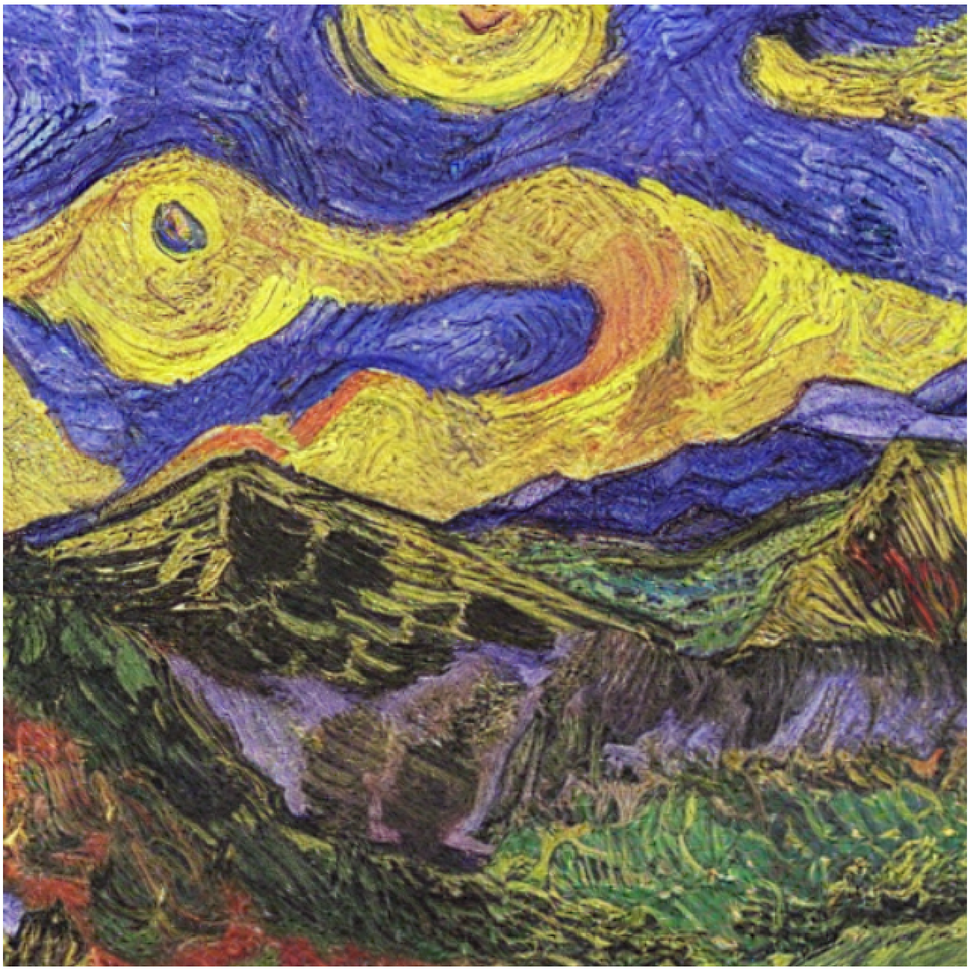

Figure 2: Prompt: "a mountain landscape at sunset, van gogh style"

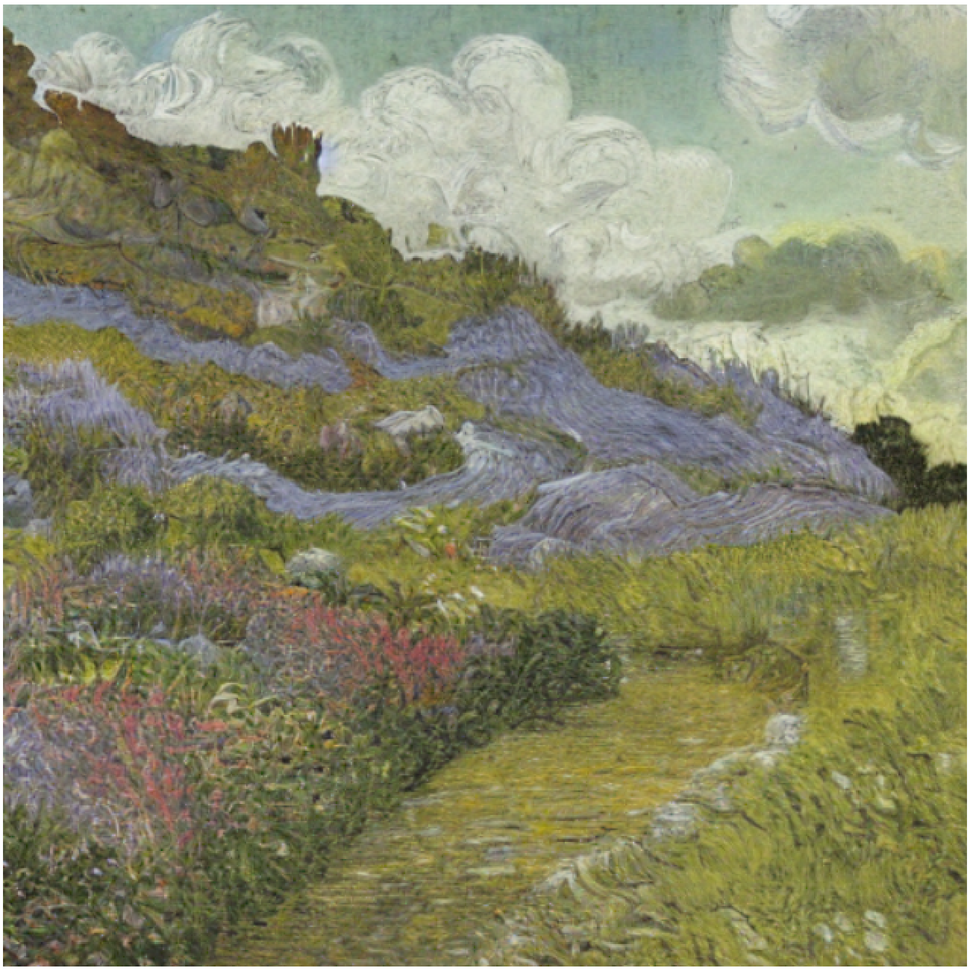

Figure 3: Prompt: ”a fantasy landscape with hills, flowers and sky, equal mix of van gogh texture, monet lighting and studio ghibli animation style”

References

1. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models”
2. MixLoRA: Efficient Adapter Mixing for Diffusion Models
3. Notebook: ”kaggle.com/code/manyadhamija/mixlora”