Improved B-LoRA for Style Preservation and Transformation

Wenrong Yin

School of Electrical Engineering, Computing and Mathematical Sciences

Curtin University

This report is submitted for < Computer Science Project 2 COMP 6016>

November 2024

DECLARATION

This document is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Signed:_____Wenrong Yin_____ _____

Date:_____07/11/24_____

Wenrong Yin（21194683)

Curtin University

ABSTRACT

This study enhances Block Low-Rank Adaptation (B-LoRA) for style preservation and transformation in text-to-image generation by focusing on specific transformer blocks (W2-W5) within the SDXL architecture. The improved configuration captures essential style features—such as color accuracy, texture fidelity, and multi-element consistency—across diverse prompts. While effective with simpler styles, challenges remain with complex backgrounds, where subjects may be overshadowed. Future recommendations include adaptive layer control and fine-tuning for specific style types. Improved B-LoRA offers a promising approach for consistent, adaptable style transfer in creative applications. Our codes will be available at https://github.com/RongWenYin/improved_blora

ACKNOWLEDGEMENTS

Thank you to Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template used to produce this document.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 Rationale

Text-to-image generation has emerged as a powerful application within generative AI, enabling the creation of high-quality images from simple text prompts. This technology holds great potential across fields such as digital art, advertising, and content creation. However, a major challenge within this domain is style preservation—the ability to maintain a consistent aesthetic across generated images while allowing for transformations to suit diverse prompts. Although diffusion models like SDXL have advanced text-to-image generation, they often struggle with preserving an intended style consistently across various content. This project addresses the need for enhanced style fidelity and flexible style transformation by improving the B-LoRA[1] method through insights from InstantStyle[2], offering a new approach to balancing style control and adaptability in generative models.

1.2 Statement of Problem

Figure 1.1. Overview of b-lora 's UNET

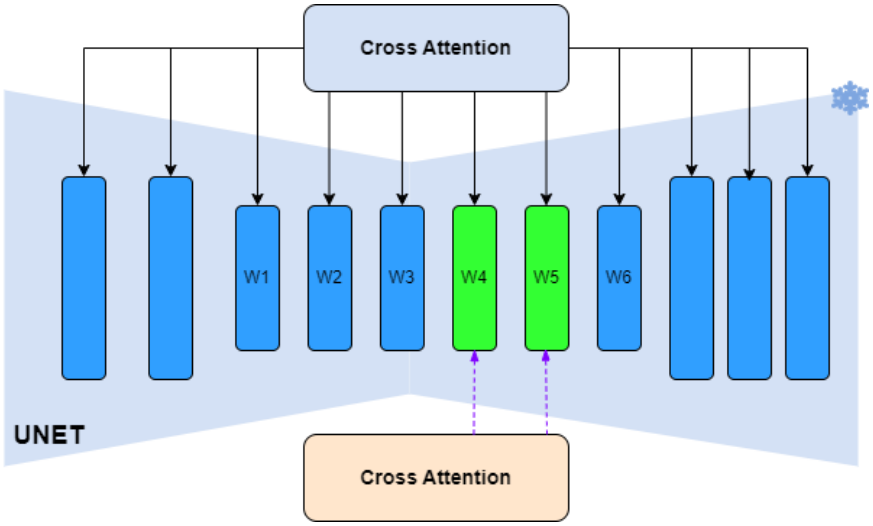
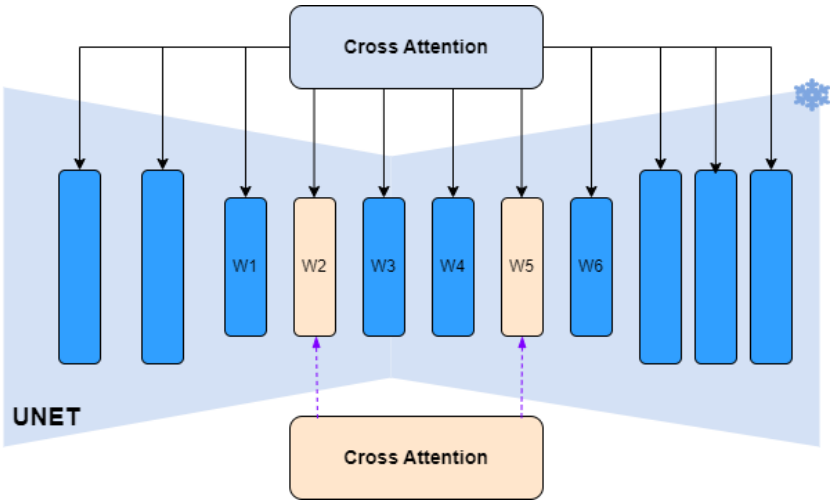


Figure 1.2 Overview of InstantStyle 's UNET



While existing methods for style control in diffusion models have achieved some success, each has its limitations. B-LoRA uses Low-Rank Adaptation to fine-tune layers 4 and 5 (as shown in Figure 1.1), effectively separating style from content, yet it struggles to maintain consistent style across different prompts. On the other hand, InstantStyle focuses on style preservation by injecting reference features into layers 2 and 5(as shown in Figure 1.2), achieving high style fidelity with minimal model adjustment but lacking flexibility in transforming or adapting style to new content. This project addresses the gap between these two methods, aiming to develop a technique

that combines the strengths of both. Specifically, we aim to leverage InstantStyle's layer-based focus within the B-LoRA framework to enhance style preservation while allowing for controlled transformations, thus meeting the dual requirements of consistency and adaptability.

1.3 Delimitation and Limitations

This research focuses on improving style preservation within the SDXL architecture (as shown in Figure 1.1), by targeting transformer layers W2 and W5. We limit our study to these layers, drawing from InstantStyle's findings, and use a single reference photo to test the model's ability to apply consistent style across diverse prompts. However, the study has certain limitations. The results may be specific to SDXL and may not directly translate to other generative models without further adaptation. Additionally, the outcomes may vary depending on the style reference chosen, meaning the findings may not generalize across all types of style references or domains.

1.4 Theoretical Framework

This research focuses on improving style preservation within the SDXL architecture by targeting transformer layers W2 and W5. We limit our study to these layers, drawing from InstantStyle's findings, and use a single reference photo to test the model's ability to apply consistent style across diverse prompts. However, the study has certain limitations. The results may be specific to SDXL and may not directly translate to other generative models without further adaptation. Additionally, the outcomes may vary depending on the style reference chosen, meaning the findings may not generalize across all types of style references or domains.

1.5 Definition of Terms

B-LoRA: Block Low-Rank Adaptation, a technique that fine-tunes specific layers within a model to enable control over style and content independently.

InstantStyle: A method for style preservation that injects style features directly into selected layers, particularly layers 2 and 5, allowing for minimal model adjustments.

SDXL: A powerful latent diffusion model used as the base architecture in this study, known for its ability to handle complex text-to-image transformations.

Transformer Blocks: In the SDXL architecture, transformer blocks play a critical role in controlling how style and content are represented in generated images. Each block is made up of multiple attention layers, which allow the model to focus on specific aspects of the image, such as fine details or broader stylistic elements.

1.6 Assumptions

This study operates under a few key assumptions. We assume that focusing on layers W2 and W5, as identified by InstantStyle, will improve B-LoRA's ability to consistently apply style across different prompts while preserving content. It is also assumed that a single photo reference will suffice to test the model's capability to maintain style fidelity and apply the style across varied content without further fine-tuning.

1.7 Research Approach

The research approach is experimental, combining B-LoRA's adaptability with InstantStyle's insights into selective layer focus within the SDXL model. We focus on testing transformer layers W2 and W5 for their ability to control style while maintaining content integrity. Using a single photo as the style reference, we generate images across a range of prompts to analyze the effectiveness of each configuration. This approach allows us to assess the degree to which the proposed method enhances style preservation and enables flexible transformations across diverse content.

1.8 Contributions

This project makes several contributions to the field of text-to-image generation and style transfer. It presents an improved methodology for style preservation by integrating insights from InstantStyle into the B-LoRA framework, demonstrating that focusing on W2 and W5 within the SDXL architecture enhances style consistency. This work can serve as a foundation for future studies that aim to refine style control in generative models, particularly in applications where consistent and adaptable style is critical.

1.9 Thesis Structure

The thesis is structured to provide a comprehensive view of the project. Chapter 1 (Introduction) outlines the project's rationale, objectives, and structure. Chapter 2 (Literature Review) explores existing work in style transfer, diffusion models, and

layer-focused techniques. Chapter 3 (Methodology) details the experimental approach, including model configuration and evaluation strategies. Chapter 4 (Results) presents the findings from the experiments, examining the effectiveness of different transformer block configurations. Finally, Chapter 5 (Conclusion) summarizes the main insights, discusses the implications of the findings, and suggests potential directions for future research.

1.10 Summary

This chapter has introduced the context, rationale, and objectives of the project, focusing on improving B-LoRA with insights from InstantStyle to achieve better style preservation and transformation within the SDXL model. By targeting layers W2 and W5, this project aims to enhance both style fidelity and flexibility, offering a refined approach to style application in text-to-image generation.

2 Review of Literature

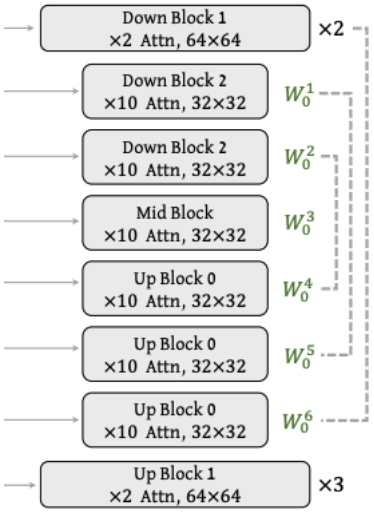
2.1 Background

Text-to-image generation is a rapidly advancing field in generative AI, where models like diffusion models have led to significant improvements in generating high-quality images from text prompts. A central challenge, however, is achieving consistent style preservation across varied prompts while allowing for adaptable style transformations. This challenge is particularly prominent in applications where personalized aesthetics or brand consistency are required. The focus of this study is to build upon existing style preservation methods, particularly B-LoRA and InstantStyle, to improve style fidelity and adaptability in the SDXL model.

2.2 Relevant Topics

SDXL Architecture[4]: In this study, we use the recently introduced Stable Diffusion[5] XL (SDXL), an enhanced version of the original Stable Diffusion model. Both SDXL and Stable Diffusion are latent diffusion models (LDMs), where the diffusion process occurs in the latent space of a pre-trained image autoencoder. SDXL improves upon its predecessor with a significantly larger UNet backbone—approximately three times the size of the original model. The SDXL architecture (as shown in Figure 2.1), consists of 70 attention layers, organized within a UNet structure, each comprising cross-attention and self-attention mechanisms to capture complex image interactions. For this study, these 70 attention layers are grouped into 11 transformer blocks, making SDXL highly effective for complex style and content transformations.

Figure 2.1 Illustration of SDXL architecture



LoRA (Low-Rank Adaptation)[3]: LoRA is a technique designed to fine-tune specific layers or blocks within a model by introducing low-rank adaptations to the model's parameters. Instead of modifying the entire network, LoRA makes adjustments only to a subset of parameters in targeted layers, reducing computational costs and memory usage. LoRA's ability to focus on specific layers makes it especially effective for tasks that require fine-grained control, like style and content separation in text-to-image models.

Implicit Style-Content Separation using B-LoRA[1]: B-LoRA, or Block Low-Rank Adaptation, aims to separate style and content within generative models. This approach utilizes low-rank adaptation (LoRA) to fine-tune specific blocks in the U-Net model (typically focusing on layers 4 and 5) to control high-level style features, such as color, texture, and other aesthetic aspects. B-LoRA's strength lies in its ability to adjust style and content independently, enabling flexible content modifications while maintaining a distinct style. However, B-LoRA's control over style consistency across prompts is limited, as it tends to vary in style application when prompt content changes significantly. This lack of consistency reveals a gap in B-LoRA's ability to ensure uniform style application across diverse inputs, particularly in cases where the same style must be applied repeatedly across different scenarios.

InstantStyle: Free Lunch towards Style-Preserving in Text-to-Image Generation[2]: InstantStyle adopts a unique "tuning-free" approach to style preservation. Instead of fine-tuning the model, InstantStyle directly injects style features into specific layers of

the U-Net, particularly layers 2 and 5, which capture foundational and structural style elements. Layer 2 primarily controls basic style features like color and texture, while layer 5 focuses on layout and spatial structure. By concentrating on these layers, InstantStyle achieves consistent style application across prompts with minimal adjustments to the model, ensuring high style fidelity without extensive computational overhead. However, this approach lacks flexibility in style transformation, as it doesn't provide fine-grained control over style variations in response to different content prompts. Thus, while InstantStyle is efficient in maintaining style fidelity, its application is limited when dynamic style adjustments are necessary.

2.3 Critique

Although both B-LoRA and InstantStyle offer valuable solutions for style control, they have notable limitations when applied individually. B-LoRA's focus on layers 4 and 5 enables effective separation of style and content, yet it often results in inconsistencies in style when generating images from different prompts. This inconsistency challenges its application in areas requiring uniform style application. Conversely, InstantStyle's tuning-free layer injection into layers 2 and 5 preserves style effectively but lacks adaptability, making it difficult to modify the style dynamically. These limitations point to the need for a combined approach that can provide both stability and flexibility, particularly by focusing on layers that balance foundational style features with structural adaptability.

2.4 Summary

This review highlights the advancements and limitations within existing methods for style preservation and transformation, specifically B-LoRA and InstantStyle. By combining B-LoRA's adaptable layer-specific adjustments with InstantStyle's selective layer injection, this study seeks to develop a method that enhances style fidelity while allowing controlled style transformation. Focusing on layers W2 and W5 within the SDXL model, the proposed approach aims to address the gaps in current style transfer methods, providing a balanced solution for applications requiring both consistent style and adaptable content.

3 METHODOLOGY

3.1 Methodology

The objective of this study is to develop an Improved B-LoRA approach that enhances style preservation and controlled transformation within text-to-image generation models. Building on the original B-LoRA framework, this research introduces modifications that leverage SDXL v1.0 as the base model while incorporating insights from InstantStyle to focus on specific transformer blocks (primarily W2 and W5) that are hypothesized to be crucial for style fidelity and adaptability. The modified B-LoRA implementation aims to achieve a refined balance between style consistency and the ability to transform style based on diverse prompts.

3.2 Approach

The approach centers on improving B-LoRA by focusing on strategic layer modifications within SDXL, specifically targeting transformer blocks W2 and W5 as identified in InstantStyle's approach. We utilize SDXL as the base model for implementing B-LoRA, with a focus on analyzing different combinations within SDXL's transformer blocks (W1 to W6) to explore optimal style preservation and transformation. This enhanced B-LoRA uses SDXL's UNet backbone with modifications that allow selective style control through pairwise combinations of these blocks. Key configurations, especially W2-W5 and W4-W5, are tested to assess their

impact on style consistency and transformation quality, as these layers are critical for handling foundational style features.

We generate images using pairwise combinations of these blocks to identify patterns in style preservation across various configurations. These modifications position Improved B-LoRA as a technique that combines style preservation with adaptability, making it more versatile than the original B-LoRA.

3.3 Context of Study

This study addresses the limitations of B-LoRA and InstantStyle individually in achieving style consistency and flexibility in diffusion models. The original B-LoRA method separates style from content by adapting layers 4 and 5, but it struggles with maintaining consistent style across prompts. Meanwhile, InstantStyle achieves high fidelity with its tuning-free layer focus on layers 2 and 5 but lacks transformation flexibility. This research develops an Improved B-LoRA that combines the strengths of both methods, focusing on W2 and W5 to enhance style consistency and transformation adaptability, with SDXL's architecture providing a robust foundation for these modifications.

3.4 Instrumentation

The primary model for this study is SDXL v1.0, selected for its expanded UNet architecture with 70 attention layers across 11 transformer blocks. This architecture enables fine-grained control over style elements, which is essential for the modifications introduced in Improved B-LoRA. During fine-tuning, both the model weights[6] and text encoders are kept frozen, with only B-LoRA weights being updated. The Adam optimizer is used with a learning rate of 5e-5, while center cropping is applied as the sole data augmentation method to maintain style integrity. LoRA weights are set to a rank of r = 64, and each prompt is used across 1,000 optimization steps to stabilize style output.

3.5 Data Collection

Data collection is structured around generating images using diverse prompts with a single style reference photo to establish a baseline for style consistency. By saving model checkpoints every 1,000 steps up to 5,000 steps, the study examines the impact of training duration on style retention and transformation. All analysis is conducted on

Kaggle, utilizing its GPU resources for efficient processing and enabling consistent benchmarking. To test Improved B-LoRA's adaptability, multiple style reference photos are selected, each varying in color, texture, and artistic style. This ensures that the model's response to different stylistic elements can be evaluated.

3.6 Treatment of Data

Style reference images used in the experiments are preprocessed to fit within a 1024x1024 pixel size, with accepted formats being PNG or JPG to ensure compatibility and consistency in style application.Generated images are analyzed for style fidelity and adaptability across various transformer block combinations and training steps. Improved B-LoRA's performance is compared against baseline configurations to assess its effectiveness in maintaining consistent style while adapting to new prompts.

3.7 Evaluation

The evaluation focuses on two main aspects: style preservation and controlled transformation. Improved B-LoRA's configurations—such as W2-W5, W4-W5, and W5 alone—are compared with InstantStyle configurations, specifically W2-W5 and W5 alone, to determine which setup best maintains the reference style while allowing for adaptable transformations. Key metrics include color fidelity, texture consistency, and transformation adaptability, all of which are crucial for achieving both uniformity and versatility in style application. These metrics provide a foundation for assessing whether Improved B-LoRA meets its objective of enhancing style fidelity and adaptability.

3.8 Summary

This methodology outlines the steps taken to develop and evaluate Improved B-LoRA within the SDXL model, focusing on achieving a balance between style preservation and transformation flexibility. Through strategic layer modifications, checkpointing, and detailed evaluation criteria, this study aims to position Improved B-LoRA as a robust solution for style control in text-to-image generation. All experiments were executed on Kaggle's free GPU resources, ensuring accessible and efficient computation.

4 RESULTS

4.1 Methods

The evaluation process involves testing different transformer block combinations within SDXL to assess their impact on style preservation and transformation. To determine the optimal number of training steps for achieving style fidelity and transformation quality, model checkpoints were saved at 1,000-step intervals, from 1,000 to 5,000 steps. At each checkpoint, images were generated to analyze style consistency and transformation quality over the training duration. This process specifically emphasized the W2-W5 and W4-W5 combinations within the SDXL model, as these layers are hypothesized to have the greatest influence on style preservation and adaptability.

We evaluated style preservation by generating images with different transformer block configurations for both B-LoRA and InstantStyle:

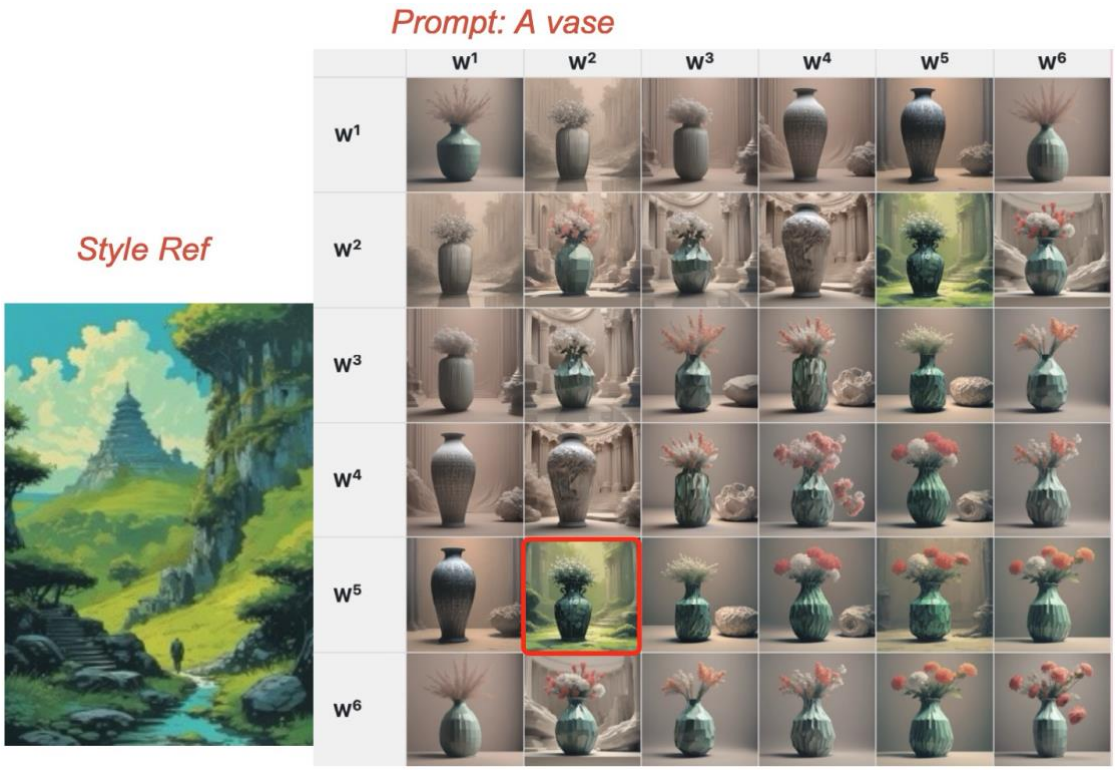
- For B-LoRA, configurations included W4-W5, W2-W5, and W5 alone to assess how each setup impacts style consistency and transformation control.
- For InstantStyle, configurations focused on W2-W5 and W5 alone, allowing direct comparison with B-LoRA's performance in similar configurations.

4.2 Results and analysis

4.2.1 Evaluate the effectiveness of Improved B-LoRA

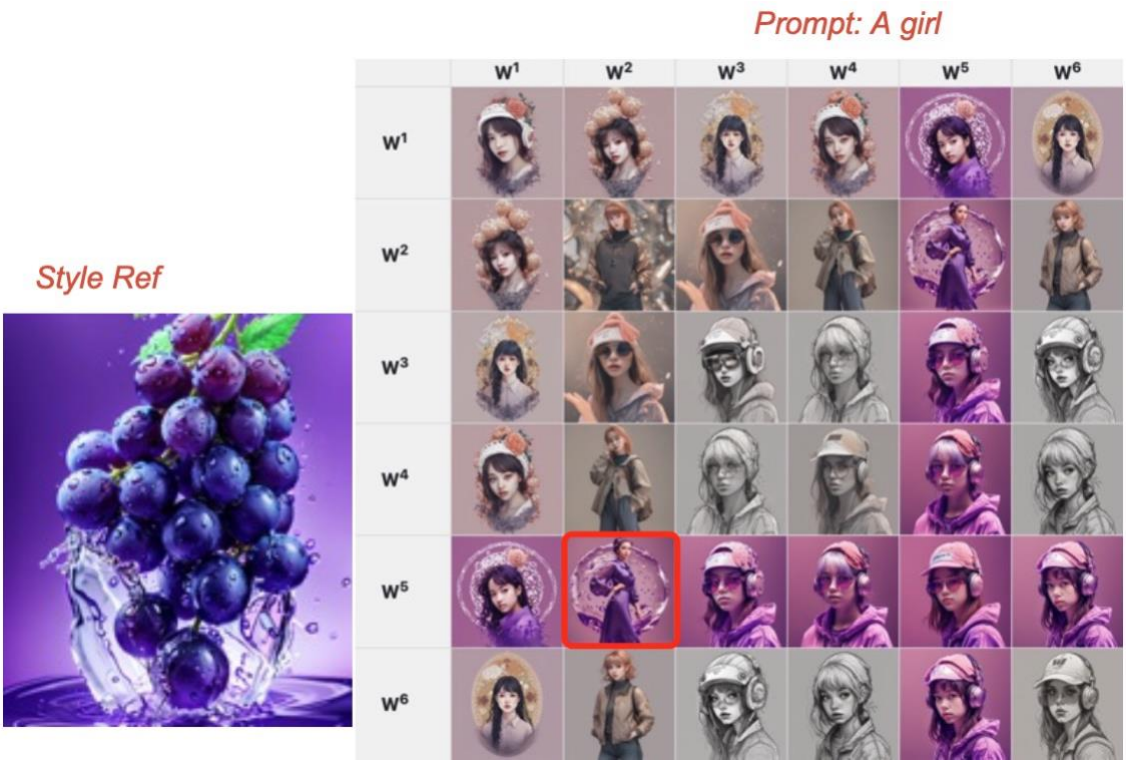
To evaluate the effectiveness of Improved B-LoRA for style preservation and transformation, I used two reference images with distinct styles. The first reference depicted a natural landscape with vibrant greens and soft textures, while the second was an image of purple grapes showcasing a glossy, reflective quality. The evaluation process focused on testing different transformer block combinations within SDXL, exploring pairwise combinations of layers from W1 to W6.

Figure 4.1 The performance of different combinations within SDXL's transformer blocks (w1 to w6) Example1



In Figure 4.1, which uses the natural landscape style with the prompt "A vase" different combinations demonstrated unique impacts on style preservation. The W2-W5 combination effectively captured the green tones and soft textures from the landscape, balancing style fidelity with content clarity. This setup maintained the vase structure while subtly integrating the landscape's naturalistic qualities, making it an ideal configuration for situations where both style and content are essential.

Figure 4.2 The performance of different combinations within SDXL's transformer blocks (w1 to w6) Example 2



In Figure 4.2, which uses the grape style reference with the prompt "A girl," an important observation emerges. The style reference image contains water droplets on the grapes, adding a glossy, reflective quality. Among the configurations tested, only the W2-W5 combination successfully captures this water element, creating a subtle glossy texture that resembles the reference. This suggests that W2-W5 is particularly effective at preserving intricate style details.

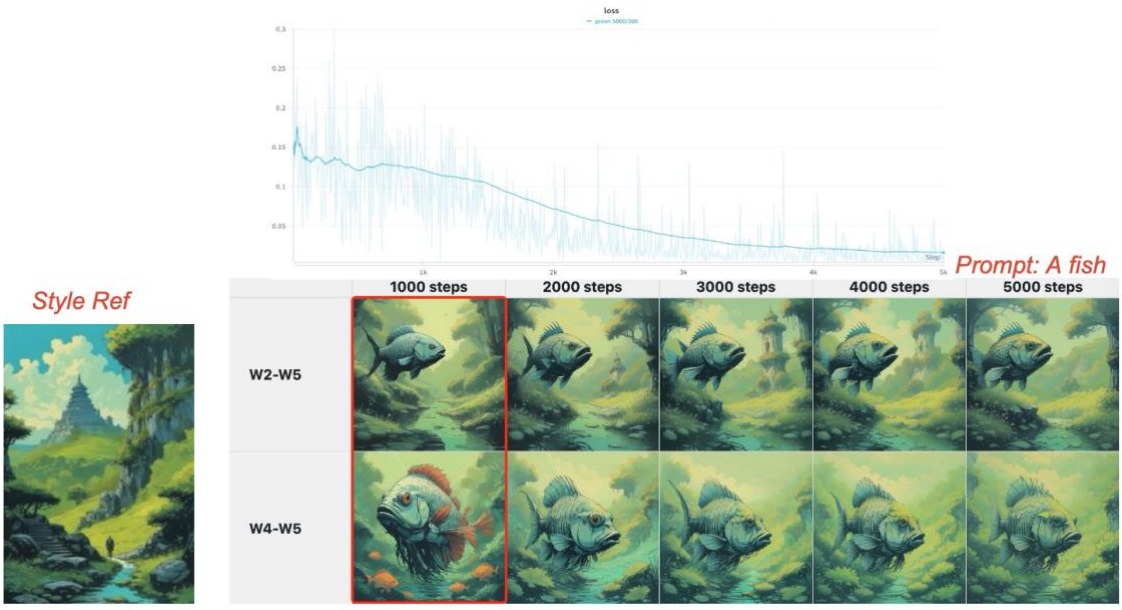
Additionally, configurations involving W5 consistently introduce the prominent purple hue from the grape reference, indicating that W5 plays a key role in transferring the core color palette. The presence of this vibrant purple across all W5 combinations highlights its importance in maintaining color fidelity, ensuring that primary stylistic elements（like color）—are consistently preserved.

4.2.2 Find the best number of training steps

In this experiment, a forest-themed image with green tones was used as the style reference, and "A fish" was set as the prompt to evaluate how well the model could adapt the forest style to the new content.

In Figure 4.3, the loss curve shows a steady decline, indicating effective convergence. In the generated images, by 1,000 steps, the model has already captured essential elements of the forest style, incorporating green tones and textures reminiscent of the reference image. While the loss continues to decrease with additional steps, the visual quality and style fidelity in the images do not show noticeable improvement beyond 1,000 steps. Therefore, we selected 1,000 steps as the optimal training duration, as it provides a satisfactory balance between style adaptation and training efficiency without unnecessary additional steps.

Figure 4.3 Style Fidelity and Transformation Quality Across Training Steps for W2-W5 and W4-W5 Configurations



4.2.3 Comparison

Figures 4.4, 4.5, and 4.6 provide a comparative analysis of Improved B-LoRA against B-LoRA (W4-W5) and InstantStyle (W2-W5 and W5 configurations). The purpose of this comparison is to evaluate the effectiveness of Improved B-LoRA in maintaining style fidelity and adaptability across different prompts while using distinct style references.

Figure 4.4 Comparison (Example 1)

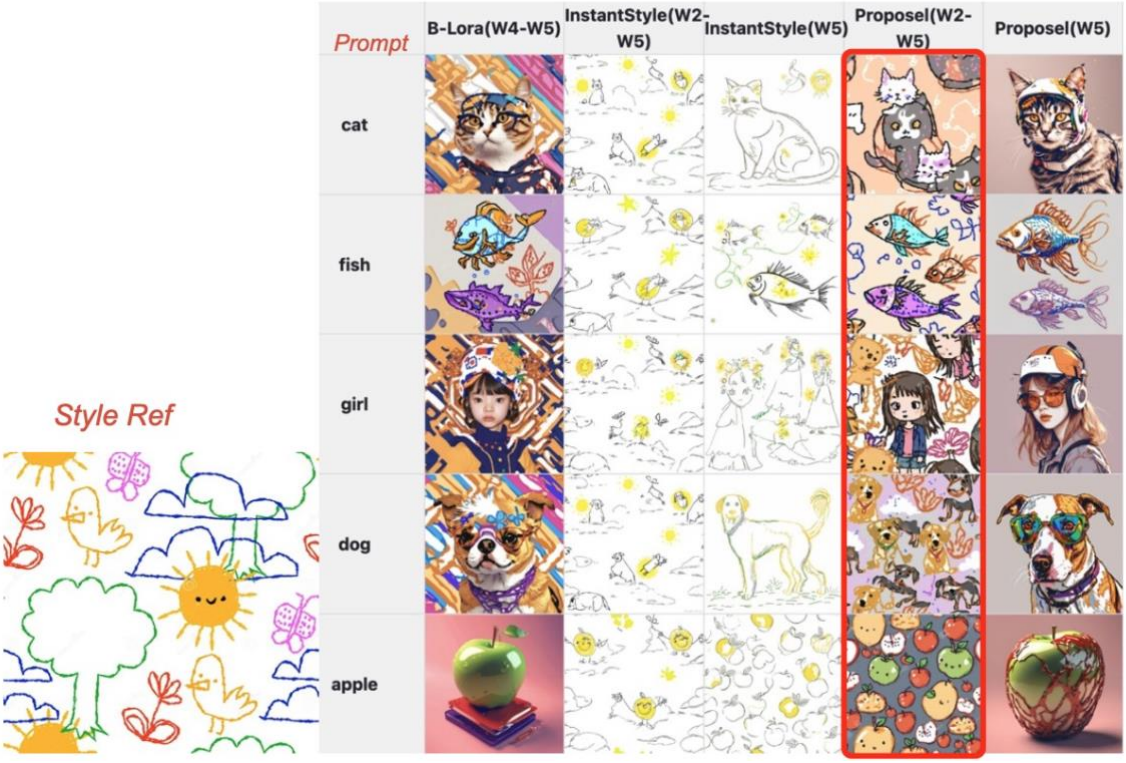


Figure 4.5 Comparison (Example 2)

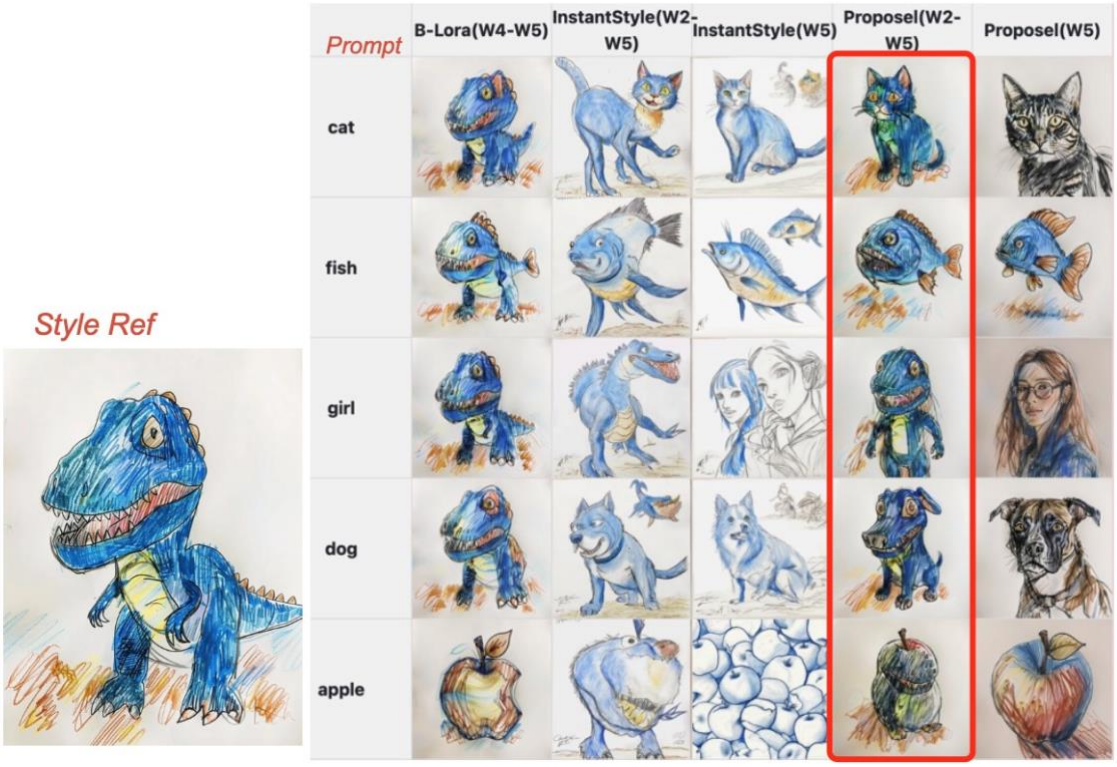
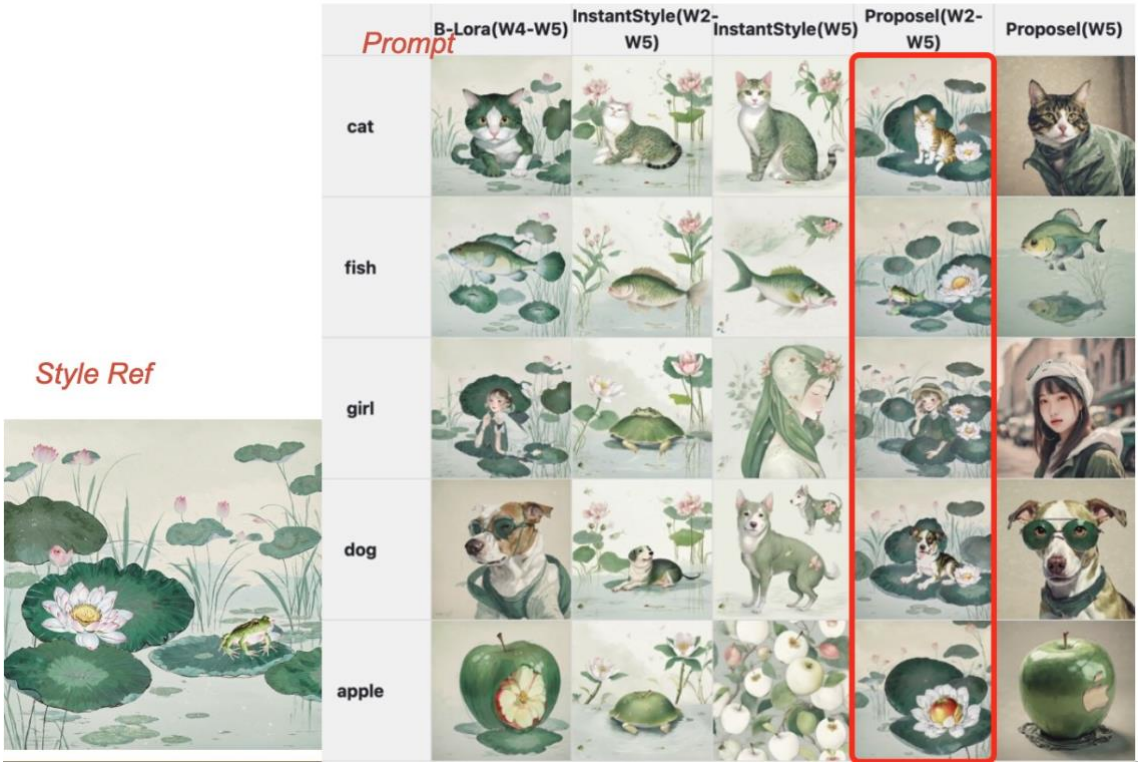


Figure 4.6 Comparison (Example 3)



i. **Texture and Detail Adaptation:** Across all three style references, Proposal (W2-W5) consistently captures and adapts the unique textures and fine details, whether it's the vibrant playfulness of the cartoon-like style in Figure 4.4, the rough, hand-drawn sketch quality in Figure 4.5, or the soft, serene tones of the landscape-inspired style in Figure 4.6. This demonstrates that Proposal (W2-W5) can adapt diverse textures effectively, bringing out the core characteristics of each reference.

ii. **Multi-Element and Object Consistency:** The Proposal (W2-W5) configuration excels in preserving and consistently applying multiple elements from the reference images across varied prompts. In Figure 4.4's colorful reference, it integrates various small elements such as clouds, sun shapes, and trees into the prompts without overcrowding or distorting the main subject. Similarly, in Figure 4.6's nature-inspired reference, Proposal (W2-W5) manages to retain essential background elements like lily pads and flowers, subtly present in prompts like "fish" and "girl." This multi-element consistency highlights the

Proposal's capability to incorporate complex backgrounds seamlessly into generated image

iii. **Background Integration:** Proposal (W2-W5) demonstrates effective background integration across different prompts and references, maintaining coherence without overpowering the subject. For example, in Figure 4.5's dinosaur-style reference, Proposal (W2-W5) subtly integrates rough, pencil-like backgrounds that enhance the sketchy quality, while in Figure 4.6, it brings in soft, natural scenery that complements the subjects. This balanced background integration is less evident in other configurations, where the backgrounds either become too dominant (e.g., in B-LoRA W4-W5) or too simplified (in InstantStyle). Proposal (W2-W5) strikes a harmonious balance, retaining the aesthetic value of the background without overshadowing the primary content.

iv. **Color Fidelity:** Color fidelity is one of the standout strengths of Proposal (W2-W5). In each example, it accurately captures the core color palettes of the references: the vibrant and varied colors in Figure 4.4, the earthy and muted tones in Figure 4.5, and the soft greens and pastels in Figure 4.6. By preserving these color characteristics, Proposal (W2-W5) enhances the authenticity and visual appeal of each generated image, maintaining alignment with the original styles.

v. **Line Quality and Structure:** Across different styles, Proposal (W2-W5) successfully maintains the line quality of the reference images. Whether it's the playful, uneven lines in Figure 4.4, the rough, sketch-like strokes in Figure 4.5, or the soft, delicate lines in Figure 4.6, Proposal (W2-W5) consistently captures and applies these qualities. This is especially evident in prompts such as "cat" and "apple," where the line structure aligns closely with the reference, enhancing the stylistic fidelity.

In summary, the Proposal (W2-W5) configuration in Improved B-LoRA consistently outperforms both B-LoRA (W4-W5) and InstantStyle (W2-W5 and W5) across multiple style aspects. It excels in capturing textures, preserving multi-element consistency, integrating backgrounds effectively, maintaining color fidelity, and aligning line quality with the style reference. This demonstrates that Proposal (W2-W5) is a powerful approach for style preservation and transformation,

successfully adapting varied style references across different prompts while retaining the core stylistic elements.

4.2.4 Qualitative Results

Figure 4.7 Qualitative Comparison of Style Adaptation Across Diverse Prompts

In this section, we analyze the visual outcomes of Improved B-LoRA using various style references and prompts, as shown in Figure 4.7. The model generally performs well in capturing the essence of each style reference, but certain areas still present challenges.

i. **Strengths in Style Fidelity and Texture Adaptation:** Improved B-LoRA consistently captures the unique characteristics of each style. The autumn landscape style maintains warm colors and layered textures, while the origami-inspired style accurately reproduces paper-like textures across different prompts. This texture adaptation enhances the realism and depth in each generated image.

ii. **Color Accuracy Across Styles:** Improved B-LoRA shows reliable color reproduction, preserving vibrant purples in the grape style and soft, muted tones in the origami-inspired style. This color fidelity improves the visual consistency and alignment with the reference images, contributing to the model's adaptability across various prompts.

iii. **Consistency in Multi-Element Styles**: For complex references, such as the autumn landscape, the model integrates multiple elements (like trees) without overcrowding the main subject. This balance allows the style to enhance, rather than overwhelm, the generated images, making it particularly effective in rich, multi-element styles.

However, Improved B-LoRA consistently captures the unique characteristics of each style. The autumn landscape style maintains warm colors and layered textures, while the origami-inspired style accurately reproduces paper-like textures across different prompts. This texture adaptation enhances the realism and depth in each generated image.

4.3 Reflection

Reflecting on the outcomes of Improved B-LoRA reveals its strong capabilities in style adaptation but also its limitations when dealing with complex or intricate styles. Improved B-LoRA effectively maintains essential stylistic elements, such as color accuracy, texture, and multi-element composition, which adds depth and authenticity to generated images across various prompts. However, as the style complexity increases — involving layered textures, detailed backgrounds, or multiple elements — there is a noticeable impact on subject prominence.

For instance, styles with dominant backgrounds, such as the autumn and green foliage themes, tend to overwhelm smaller subjects, making it difficult for key elements like "dog" or "fish" to stand out clearly. Similarly, in cases where intricate textures or fine details are central to the style, there is sometimes an unintended scaling effect, where primary subjects appear diminished or less defined within the complex backdrop. This suggests that, as style complexity rises, Improved B-LoRA faces challenges in balancing background integration with maintaining a clear focus on the main subject.

These reflections highlight a need for further refinement in handling complex styles, where controlling the prominence of background elements and adjusting subject scaling could improve adaptability. Future iterations of Improved B-LoRA could benefit from a more nuanced approach to background and subject prioritization, enabling better performance in styles with layered or intricate details.

4.4 Summary

In summary, Improved B-LoRA demonstrates effectiveness in capturing and preserving a range of stylistic qualities, with strengths in texture fidelity, color adaptation, and multi-element integration. However, as style complexity increases, challenges arise in balancing background details and subject clarity. These findings underscore the model's robustness in simpler styles and point to areas for improvement in managing intricate backgrounds and scaling issues within more complex stylistic references.

5 Conclusion

5.1 Conclusions

This project set out to improve B-LoRA's ability to preserve and adapt styles across diverse prompts, focusing on the W2-W5 configuration within the SDXL architecture. Improved B-LoRA successfully maintains core stylistic elements, such as color and texture fidelity, across a range of styles and subjects. While it excels with simpler styles, some challenges remain with complex backgrounds, where main subjects can sometimes be overshadowed. Overall, the results affirm that Improved B-LoRA offers enhanced style fidelity and flexibility compared to traditional methods.

5.2 Implications

Improved B-LoRA has significant potential in creative fields requiring consistent style adaptation across varied content, such as digital art and design. Its ability to handle intricate style details while adapting to different subjects makes it a versatile tool. However, further optimization could enhance its effectiveness with more complex styles, widening its application potential.

5.3 Recommendations

For future research, it is recommended to explore dynamic layer adjustment to balance style complexity, enhanced subject isolation techniques like selective masking for better subject prominence in intricate styles, and fine-tuning for specific style types to

optimize performance for minimalist versus detailed styles. These improvements could further refine B-LoRA's adaptability and effectiveness across a broader range of stylistic applications.

5.4 Conclusion

Improved B-LoRA effectively balances style fidelity and adaptability, meeting the project's objective of creating a model that adapts consistently across varied prompts. While further improvements are needed for complex styles, this work represents a valuable advancement in style-preserving image generation.

6 REFERENCES

[1] Frenkel, Y., Vinker, Y., Shamir, A., & Cohen-Or, D. (2024). Implicit Style-Content Separation using B-LoRA. *arXiv preprint arXiv:*2403.14572. https://doi.org/10.48550/arXiv.2403.14572

[2] Wang, H., Wang, Q., Bai, X., Qin, Z., & Chen, A. (2024). Instantstyle: Free lunch towards style-preserving in text-to-image generation. ar*Xiv preprint arXiv*:2404.02733.

[3] Hu, J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. *ArXiv,* abs/2106.09685.

[4] Dustin Podell, Zion English, Kyle Lacey, A. Blattmann, Tim Dockhorn, Jonas Muller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *ArXiv,* abs/2307.01952, 2023.sdxl

[5] Robin Rombach, A. Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis ¨ with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, 2021

[6] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* pages 22500– 22510, 2023