Identity-Preserving Portrait Stylization with LoRA-Based Diffusion Models

Oğuz Kağan Hitit [*1] **İdil Görgülü** [*1]

Abstract

This project addresses the challenge of identity preservation in exemplar-based portrait style transfer using LoRA-adapted diffusion models. While recent methods such as ConsisLoRA (Chen et al., 2025) improve style and content consistency, they often fail to retain the subject's facial identity in the stylized output. We propose an identity-aware extension to the existing training pipeline, where an identity regularization loss based on face embeddings (ArcFace, DINOv2, or CLIP) is introduced during LoRA fine-tuning. Our approach aims to maintain the subject's unique features while applying high-quality artistic styles such as Pixar-inspired cartoons. We will evaluate the method on facial image datasets and compare it against existing baselines using identity similarity and perceptual quality metrics. This project combines recent advances in parameter-efficient adaptation with facial representation learning to bridge the gap between stylization and semantic fidelity.

1. Introduction

Portrait style transfer is a fascinating task in the field of generative modeling, where the goal is to stylize a real-world face image into an artistic domain—such as a Pixar-style cartoon—while preserving the subject's identity and facial structure. With the rise of powerful generative models, high-quality stylization has become increasingly feasible. Recent LoRA-based approaches have enabled exemplar-based high-resolution stylization, allowing the transfer of complex styles from a single reference image to new content images (Shah et al., 2023; Frenkel et al., 2024; Chen et al., 2025).

However, a persistent challenge in stylized portraits is the loss of identity. In many cases, while the generated image successfully adopts the artistic characteristics of the style domain, it fails to retain key identity features such as facial shape, expression, and landmarks (Chen et al., 2025). This issue becomes particularly problematic in applications like avatar creation, personalized artistic filters, or identity-aware content generation, where the recognizability of the original face is critical.

This project investigates the problem of identity loss in exemplar-based portrait style transfer and aims to address it using LoRA-adapted diffusion models. We aim to enhance the content preservation capability of LoRA by incorporating an identity-aware loss during training. This will be achieved by introducing a regularization term based on face recognition embeddings, which ensures that the stylized output remains semantically close to the original in identity space. We find this topic particularly interesting because it touches on the intersection of feature preservation and creative stylization, two domains that are often in conflict in generative modeling. The project is challenging due to the need to balance artistic abstraction with structural fidelity, as well as the difficulty of quantifying "identity" in a generative context.

2. Related Work

The task of portrait style transfer has been extensively explored, with notable advancements in several architectures such as generative adversarial networks (GANs) and diffusion models. Our review focuses specifically on diffusion-based approaches, particularly those that use fine-tuning methods like Low-Rank Adaptation (LoRA).

2.1. Style Transfer on Diffusion Models

Diffusion models have recently become powerful generative tools for image synthesis, offering diverse, coherent and high-quality outputs by learning to reverse a denoising process of an underlying data distribution. Their flexibility yields widespread use in various image manipulation tasks such as style transfer. These models have been mostly used for generating a new image from a textual description of the style. However, this approach faces challenges due to the necessity of crafting text prompts to convey the desired style that are detailed and accurate enough. This requires a thorough understanding of the style features and great effort (Li, 2024). One commonly used approach to overcome this

[*]Equal contribution [1]Department of Computer Engineering. Correspondence to: Oğuz Kağan Hitit <ohitit20@ku.edu.tr>, İdil Görgülü <igorgulu21@ku.edu.tr>.

COMP547 Deep Unsupervised Learning, Spring 2025.

is learning style attributes from a single input image and transferring the learned attributes to a natural image. Recent approaches such as (Li, 2024), (Ukarapol, 2023) and (Chen et al., 2025) have leveraged LoRA modules to efficiently learn and apply such style attributes within diffusion models, enabling more flexible and lightweight style transfer workflows.

2.2. LoRA-Based Style Transfer

Recent approaches have leveraged Low-Rank Adaptation (LoRA) techniques to fine-tune diffusion models for style transfer tasks. For instance, Ukarapol (2023) demonstrated that integrating LoRA with diffusion models enables efficient style transfer with minimal computational resources. This project finetunes an existing diffusion model with images and captions of the style they aim to transfer to another image. The model they chose is fine-tuned using Monet paintings paired with descriptive captions, associating the words "A Monet painting" with the desired style. During training, LoRA weights are injected to adapt the model effectively. Despite the advancements this project introduced, challenges such as content inconsistency and style misalignment persist, leading to potential identity loss in the stylized outputs.

2.3. ConsisLoRA

To address the challenges in LoRA-based style transfer, Chen et al. (2025) propose ConsisLoRA, a method that enhances both content and style consistency by optimizing LoRA weights to predict the original image rather than noise predictions. This enables their model to better recognize structural features and reduce content leakage. Chen et al. (2025) also introduce a two-step training strategy in which a content-consistent LoRA is trained first and a separate style LoRA is trained from scratch using style-specific prompts afterwards. The first step balances detail preservation whereas the second step performs the style transfer. By separating content and style training into two distinct phases, their approach aims to minimize content leakage. ConsisLoRA also incorporates inference guidance mechanisms which allow for separate adjustment of content and style intensities, providing flexible control for image generation.Despite these improvements, ConsisLoRA acknowledges that preserving individual identities in facial imagery still remains as a significant shortcoming, attributing this limitation to the constrained capacity of LoRA modules.

3. The Approach

This project builds upon the ConsisLoRA framework (Chen et al., 2025). Our method is based on Stable Diffusion v1.5 and uses Low-Rank Adaptation (LoRA) modules. We adopt ConsisLoRA's two-phase training strategy:

- **Content LoRA**: Trained first to preserve structure, layout, and semantic fidelity using x_0-prediction loss.
- **Style LoRA**: Trained after content LoRA is frozen, to learn visual texture, color, and local style attributes from the exemplar image.

The key difference in our approach is the addition of an identity-preserving loss term during the training of the content LoRA. This loss is not present in ConsisLoRA.

Specifically:

- During content LoRA training, we extract identity embeddings of both the original (input) and stylized (generated) images using a pre-trained identity representation model (ArcFace, DINOv2, or CLIP).
- We compute the cosine similarity between these embeddings and minimize the difference by adding it as a regularization term in the total training loss.
- The full loss becomes:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{x_0} + \lambda \cdot \mathcal{L}_{\text{identity}},$$

  where $\mathcal{L}_{x_0}$ is the x_0-prediction loss used by ConsisLoRA and $\mathcal{L}_{\text{identity}}$ enforces similarity in identity features.

ConsisLoRA relies on perceptual and structural alignment losses, which are insufficient for preserving identity-defining facial features. By explicitly regularizing identity in a dedicated feature space, our method proposes that key features like facial structure, pose, and expression remain close to the original.

4. Experimental Evaluation

4.1. Datasets

To train and evaluate our model, we will use the following datasets:

- **Flickr-Faces-HQ (FFHQ)**: This dataset includes 70,000, 1024x1024 images of human faces, offering a diverse range of ages, ethnicities, and backgrounds (Karras et al., 2019).

We also plan to construct a custom dataset containing images for stylistic references. We will gather various cartoonistic and artistic styles from various datasets such as as the Anime Face Dataset (Splcher, 2019), the Toonify dataset (Pinkney & Adler, 2020), and the WebCaricature benchmark (Huo et al., 2018)

4.2. Evaluation Metrics

To assess the performance of our approach, we will rely on the following three metrics, consistent with our reference work, ConsisLoRA (Chen et al., 2025):

- **DINO Similarity**: We will use a pre-trained DINOv2 model (Oquab et al., 2023) to extract semantic features from both the original (content) and stylized images.
- **CLIP Similarity**: To evaluate both content and style alignment, we will compute cosine similarity between CLIP embeddings (Radford et al., 2021) of the stylized image and the reference style image.
- **DreamSim Distance**: We will report the DreamSim distance (Mokady et al., 2023), a learned perceptual similarity metric designed to reflect human visual preferences.

4.3. Baseline Comparisons

Our approach will be evaluated against ConsisLoRA (Chen et al., 2025). Also, if time permits, some of the baseline models reported in the ConsisLoRA study such as ZipLoRA (Shah et al., 2023) and B-LoRA (Frenkel et al., 2024) will be incorporated into our study.

5. Work Plan

Activity	Deadline
Complete the literature search	April 19
Reproduce results of ConsisLoRA baseline	April 26
Prepare progress report	May 3
Implement identity-aware loss integration	May 10
Conduct experiments and evaluate performance	May 31
Prepare final report and presentation	June 10

References

Chen, B., Zhao, B., Xie, H., et al. Consislora: Enhancing content and style consistency for lora-based style transfer. *arXiv preprint arXiv:2503.10614*, 2025.

Frenkel, Y., Vinker, Y., Shamir, A., and Cohen-Or, D. Implicit style-content separation using b-lora, 2024. URL https://arxiv.org/abs/2403.14572.

Huo, Q., Chen, J., Li, R., Mei, S., Zhao, Y., Zeng, G., and Qiao, Y. Webcaricature: A benchmark for caricature recognition. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, 2018. URL http://www.cs.nju.edu.cn/rl/WebCaricature.htm. Dataset available at http://www.cs.nju.edu.cn/rl/WebCaricature.htm.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019.

Li, S. Diffstyler: Diffusion-based localized image style transfer. *arXiv preprint arXiv:2403.18461*, 2024.

Mokady, R., Alaluf, Y., Patashnik, O., Pritch, Y., and Dekel, T. Dreamsim: Learning new perceptual metrics with dreamlike training data. *arXiv preprint arXiv:2311.18713*, 2023.

Oquab, M., Darcet, T., Moutakanni, T., Assran, M., Neverova, N., Li, G., Goyal, P., Bojanowski, P., Misra, I., Joulin, A., et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Pinkney, J. and Adler, D. Toonify yourself. https://github.com/justinpinkney/toonify, 2020. Accessed: 2024-04-05.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Shah, V., Ruiz, N., Cole, F., Lu, E., Lazebnik, S., Li, Y., and Jampani, V. Ziplora: Any subject in any style by effectively merging loras, 2023. URL https://arxiv.org/abs/2311.13600.

Splcher. Anime face dataset. https://www.kaggle.com/datasets/splcher/animefacedataset, 2019. Accessed: 2024-04-05.

Ukarapol, T. Diffusion lora style transfer. https://github.com/trapoom555/Diffusion-LoRA-Style-Transfer, 2023.