NP-LoRA: Null Space Projection for Subject-Style LoRA Fusion

Chuheng Chen[1,2] chenchuheng@iie.ac.cn
Xiaofei Zhou[1,2]* zhouxiaofei@iie.ac.cn
Geyuan Zhang[1,2] zhanggeyuan@iie.ac.cn
Yong Huang[1,2] huangyong@iie.ac.cn
Gang Xiong[1,2] xionggang@iie.ac.cn

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Abstract

Low-Rank Adaptation (LoRA) fusion enables the composition of subject and style representations for controllable generation without retraining. However, existing approaches primarily operate through weight-level merging, without explicitly modeling how independently trained LoRAs interact in the shared parameter space. We adopt a geometric perspective on LoRA fusion, interpreting content and style LoRAs as occupying overlapping, non-orthogonal low-rank subspaces, where such overlap can lead to conflicting parameter updates that affect generation quality. This observation motivates us to reformulate LoRA fusion not merely as parameter combination, but as a problem of controlling how updates from overlapping subspaces are combined. Based on this insight, we propose Null Space Projection LoRA (NP-LoRA), a training-free framework that employs projection as a fusion operator to explicitly modulate cross-LoRA interactions. Specifically, NP-LoRA uses principal directions of the style LoRA to define a projection subspace and projects the content LoRA onto the complementary subspace (i.e., the null space of the style LoRA), suppressing interference along dominant style directions while preserving complementary information. To avoid the overly aggressive suppression of hard projection, we further formulate soft projection as a regularized optimization problem that balances content preservation against style-subspace suppression. This objective admits a closed-form solution, yielding a projection operator controlled by a single parameter that continuously interpolates between linear merging and hard projection. Extensive experiments across multiple pretrained LoRA pairs show that NP-LoRA achieves more balanced content-style composition compared to strong baselines, without requiring retraining. Code is available at https://github.com/blackzebra7/NP-LoRA.

1 introduction

Low-Rank Adaptation (LoRA) Hu et al. [2022] has become a standard approach for customizing diffusion models Ho et al. [2020], Song et al. [2021], Dhariwal and Nichol [2021], Rombach et al. [2022], enabling subject- or style-specific adaptation from only a few examples Ruiz et al. [2023a]. In practice, users typically train LoRAs independently for different factors, such as subject identity and artistic style, due to differences in data, prompts, and objectives. This naturally leads to a common requirement: composing multiple independently trained LoRAs for flexible and reusable generation. For example, users may wish to render a learned subject (e.g., a pet dog) in a separately

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

learned artistic style (e.g., Van Gogh's brushwork), as illustrated in Fig. 1(a). However, such fusion is often unreliable in practice: naive merging frequently produces degraded results, where either subject identity is distorted or the intended style is weakened (Fig. 1(b)). This behavior highlights a key challenge in personalized concept composition, where independently trained LoRAs must be combined while preserving their respective functionalities.

This difficulty arises because independently trained LoRAs share the same underlying model space, causing their parameter updates to compete along shared parameter directions during fusion. Existing methods primarily operate through weight-level merging in parameter space. Recent approaches, such as ZipLoRA Shah et al. [2024], DuoLoRA Roy et al. [2025], K-LoRA Ouyang et al. [2025], and LoRA.rar Shenaj et al. [2025], improve over naive addition through optimized fusion weights, timestep-dependent switching, or learned hypernetworks. While effective, these methods rely on heuristic combinations of LoRA updates and do not explicitly model how different LoRAs interact, often resulting in trade-offs between subject fidelity and stylistic consistency. We interpret these failure modes from a geometric perspective. Since LoRAs are trained within the same model space, content and style LoRAs may occupy correlated low-rank subspaces, leading to overlapping directions that may encode inconsistent semantics. Importantly, we do not assume that LoRAs are cleanly disentangled; instead, we consider realistic settings where semantics partially overlap. When updates from different LoRAs compete along these shared directions, the merged model may become biased toward one concept, degrading the balance between subject identity and style consistency. The goal is therefore to combine the content LoRA (ΔW_c) and the style LoRA (ΔW_s) into a single module that preserves both representations while mitigating such interference. We focus on pairwise subject–style composition, where the two LoRAs have distinct roles in preserving subject identity and transferring visual style.

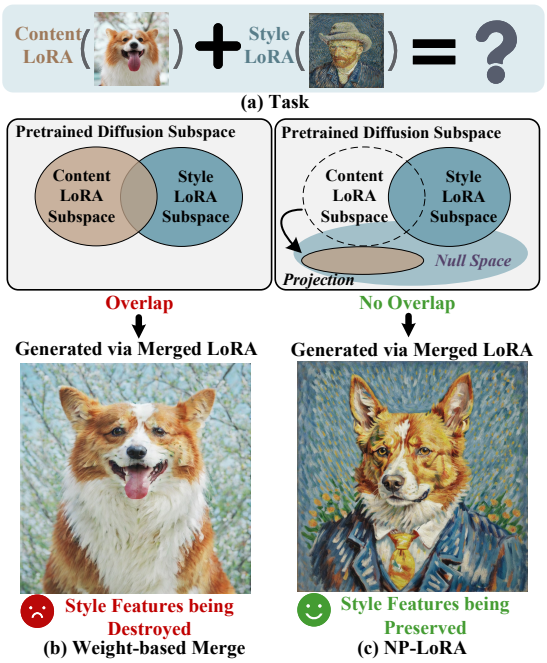


Figure 1: Illustration of our motivation. (a) The task is to compose a content LoRA (subject identity) with a style LoRA (artistic appearance) for controllable generation. (b) Weighted merging can degrade stylistic fidelity, potentially due to interference, as independently trained LoRAs may occupy correlated, non-orthogonal low-rank subspaces of the diffusion feature space. (c) NP-LoRA projects the content LoRA into the null space of the style LoRA, reducing overlap and better preserving stylistic characteristics.

Our empirical analysis (Sec. 4.1) reveals that LoRAs exhibit anisotropic low-rank structures, where a small number of dominant directions contribute disproportionately to generation and are particularly sensitive to cross-LoRA interference. This suggests that interference is concentrated along a few critical directions, making simple weight-level combinations insufficient. Consequently, effective fusion requires direction-wise control, selectively suppressing interfering components while preserving complementary ones. Motivated by this, we move beyond weight-level merging and instead explicitly control interactions in the shared subspace via projection.

To this end, we propose Null Space Projection LoRA (NP-LoRA), a training-free framework that redefines projection as a fusion operator for composing independently trained LoRAs. Unlike existing methods that rely on heuristic merging, NP-LoRA explicitly controls cross-LoRA interactions by modulating contributions along shared subspace directions. Specifically, we perform singular value decomposition (SVD) on the style LoRA ΔW_s to extract its dominant directions, and project the content LoRA ΔW_c onto the complementary subspace, suppressing components aligned with style-sensitive directions while preserving complementary information. To avoid the overly aggressive behavior of hard projection, we further formulate soft projection as a regularized optimization

problem that balances content preservation and style-subspace suppression. This objective admits a closed-form solution, yielding a projection operator controlled by a single parameter that continuously interpolates between linear merging and hard projection. NP-LoRA is training-free, requiring no additional data or optimization, and can be directly applied to pretrained LoRAs in a plug-and-play manner. This makes it particularly suitable for practical scenarios where users wish to compose LoRAs efficiently without retraining. Extensive experiments show that it achieves more balanced content–style composition than strong baselines across diverse settings.

Our main contributions are summarized as follows:

- We provide a geometric perspective on LoRA fusion, observing that independently trained LoRAs can exhibit substantial overlap in their low-rank subspaces, which may contribute to interference during fusion. This perspective offers an intuitive explanation for common failure patterns in existing merging methods.
- We propose Null Space Projection LoRA (NP-LoRA), a novel training-free framework that reformulates LoRA fusion as controlling contributions along shared subspace directions, and introduces projection as an explicit fusion operator for modulating cross-LoRA interactions.
- We formulate soft projection as a regularized optimization problem with a closed-form solution, yielding a parameterized operator that interpolates between linear merging and hard projection, enabling flexible control over the content–style trade-off.
- We conduct extensive experiments across multiple backbones, LoRA ranks, and pretrained LoRA pairs, demonstrating that NP-LoRA consistently achieves more balanced and robust content–style composition compared to strong baselines.

2 Related Work

Customization and LoRA composition in diffusion models. Diffusion models Ho et al. [2020], Song et al. [2021], Dhariwal and Nichol [2021], Rombach et al. [2022] have supported image generation and broader visual tasks, including image restoration and generative refinement for concealed visual perception He et al. [2025a,b]. In text-to-image customization, methods range from early few-shot personalization Alaluf et al. [2023], Voynov et al. [2023], Zhang et al. [2023b], Ruiz et al. [2023a] to efficient, reference-conditioned, and unified subject–style generation Kumari et al. [2023], Avrahami et al. [2023], Shi et al. [2024], Xiao et al. [2025], Xie et al. [2023], Mao et al. [2026], Wu et al. [2026], Xu et al. [2026]. Low-Rank Adaptation (LoRA) Hu et al. [2022] and its variants Hayou et al. [2024], Kopiczko et al. [2023], Ren et al. [2024], Zhang et al. [2023a], Zhou et al. [2024], Zi et al. [2023] provide efficient alternatives via rank-decomposed adapters, enabling modular customization. Prior work has explored LoRA composition by learning more disentangled representations (e.g., B-LoRA Frenkel et al. [2024], B4M Xu et al. [2025]) or by designing mechanisms for multi-concept generation (e.g., CSGO Xing et al. [2024]). In contrast, we focus on *post-hoc composition of independently trained LoRAs*, where no retraining is assumed. Existing LoRA composition methods for image generation can be roughly categorized into object composition Dong et al. [2024], Gu et al. [2023], Jiang et al. [2025], Liu et al. [2023], Yang et al. [2024], Zhong et al. [2024a], Zhang et al. [2025a], Zou et al. [2025b], Simsar et al. [2025] and content-style fusion Shah et al. [2024], Ouyang et al. [2025], Shenaj et al. [2025], Zhang and Xiong [2025]. Recent methods compose multiple LoRAs through contrastive test-time optimization, direct composition, intrinsic skill composition, cached composition, or a learned weight basis Meral et al. [2025], Zhong et al. [2024b], Foteinopoulou et al. [2025], Zou et al. [2025a], Manor et al. [2026]. Recent content-style methods, such as ZipLoRA Shah et al. [2024], DuoLoRA Roy et al. [2025], K-LoRA Ouyang et al. [2025], and LoRA.rar Shenaj et al. [2025], improve over naive addition through weighted fusion, adaptive switching, or hypernetworks. However, these approaches operate primarily through weight-level operations and do not explicitly model the geometry of cross-LoRA subspace interactions. Our work is complementary to these strategies: instead of designing new training procedures or learned fusion modules, we analyze post-hoc LoRA fusion through the geometry of low-rank subspaces and introduce a projection-based operator to control their interactions.

Orthogonality and Subspace Methods in Model Adaptation. Orthogonality and subspace methods are widely used in model adaptation to mitigate interference, often through singular value decomposition (SVD), null-space constraints, or orthogonal projections. Recent works Stoica et al., Borse et al. [2025], Farhadzadeh et al. [2025], Fang et al., Zhang et al. [2025b], Xiong and Xie [2026]

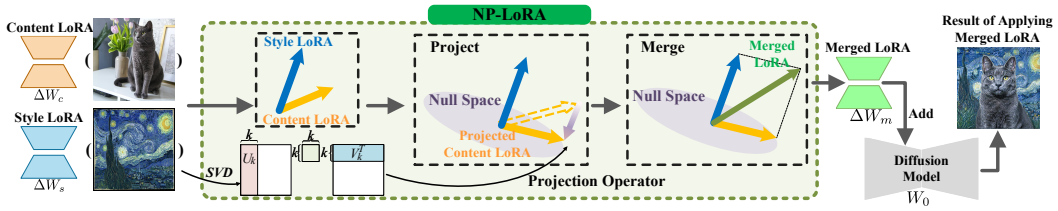


Figure 2: Overview of the proposed method. NP-LoRA takes pretrained content and style LoRAs as inputs. The style LoRA is decomposed via singular value decomposition (SVD) to construct a null space, onto which the content LoRA is projected. This design enables training-free fusion through a closed-form projection operator.

leverage these ideas in different settings, but target objectives distinct from post-hoc LoRA fusion. KnOTS Stoica et al. uses SVD to extract shared bases for LoRA alignment and aggregation, while SubZero Borse et al. [2025] reduces content–style leakage through orthogonal cross-attention at the feature level. LiON-LoRA Zhang et al. [2025b] focuses on coordinating spatial–temporal LoRAs in video diffusion. Closest to our work, ProLoRA Farhadzadeh et al. [2025] and LoRA-Null Tang et al. [2025] also exploit orthogonal or null-space properties across LoRAs, but primarily use them for re-adaptation, initialization, or interference reduction during sequential learning. In contrast, we study *training-free post-hoc composition of independently trained content and style LoRAs*. Rather than using projection as a training constraint or initialization strategy, we use it directly as a fusion operator and derive a closed-form soft projection that balances content preservation and style-subspace suppression.

3 Background

LoRA Fine-tuning. Low-Rank Adaptation (LoRA) Hu et al. [2022] efficiently adapts pretrained diffusion models Ho et al. [2020], Song et al. [2021], Dhariwal and Nichol [2021], Rombach et al. [2022] by inserting low-rank updates into existing weight matrices. Given a pretrained weight $W_0 \in \mathbb{R}^{m\times n}$, LoRA parameterizes the update as

$$\Delta W = BA, \quad B \in \mathbb{R}^{m\times r},\ A \in \mathbb{R}^{r\times n},\ r \ll \min(m,n). \tag{1}$$

During training, only A and B are optimized while W_0 is frozen. At inference, the adapted weight becomes $W = W_0 + \Delta W$.

Problem Setup. We consider the problem of composing two independently trained LoRAs: a content LoRA ΔW_c capturing subject identity, and a style LoRA ΔW_s capturing visual style. Given a pretrained diffusion model, our goal is to construct a merged LoRA ΔW_m:

$$\Delta W_m = \text{Merge}(\Delta W_c, \Delta W_s), \tag{2}$$

such that both content and style are preserved during generation. We omit layer indices for clarity, as the same operation is applied to each LoRA matrix Shah et al. [2024]. This setup focuses on post-hoc composition, where LoRAs are trained independently and must be combined without retraining.

4 Methodology

In this work, we propose NP-LoRA, a training-free framework for fusing independently trained LoRAs (Fig. 2). We view LoRA fusion as controlling interactions between overlapping subspaces. We first apply singular value decomposition (SVD) to identify dominant directions of the style LoRA, and project the content LoRA onto the complementary subspace (i.e., the null space of the style LoRA). This suppresses its contribution along style-critical directions while retaining complementary information. In the limiting case, a hard projection strongly reduces overlap but may discard useful components. We therefore introduce a soft projection with a tunable parameter, enabling continuous control over the trade-off between subject fidelity and style consistency.

4.1 Principal Direction Extraction via SVD

We analyze LoRA representations through singular value decomposition (SVD), decomposing the style LoRA as

$$\Delta W_s = U \Sigma V^\top, \tag{3}$$

where the columns of $V \in \mathbb{R}^{d_{in} \times r_s}$ are orthonormal right singular vectors. Each vector $v \in V$ defines a direction in parameter space, and its corresponding singular value in Σ reflects the strength of ΔW_s along that direction.

Following common practice Ouyang et al. [2025], Ruiz et al. [2023a], we treat the leading singular vectors as dominant directions in the LoRA update space. As shown in Fig. 3, perturbing dominant directions produces more noticeable stylistic changes than less dominant ones, suggesting that certain directions may contribute more strongly to generation and be more susceptible to interference during LoRA fusion. We emphasize that this analysis is intended as an illustrative geometric observation rather than a rigorous statistical characterization of LoRA interference.

Having identified these dominant style directions, we next examine how they behave under existing merging strategies. A common approach is linear fusion:

$$\Delta W_{\mathrm{m}} = a \Delta W_c + b \Delta W_s, \quad a, b \in \mathbb{R}. \tag{4}$$

However, such combinations do not explicitly account for interactions between LoRAs, and may introduce competing contributions along shared directions, potentially degrading stylistic consistency.

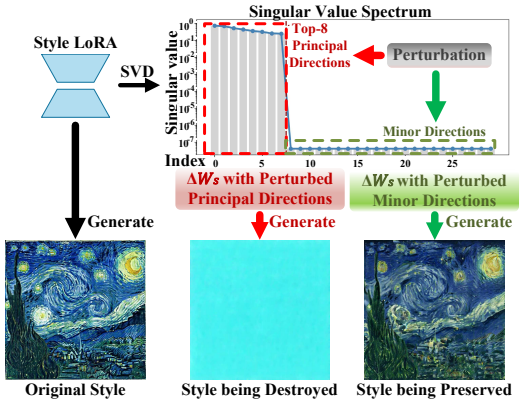


Figure 3: Singular value spectrum of a LoRA and perturbation effects. Perturbing dominant directions produces more noticeable stylistic changes than perturbing less principal directions.

Why weight-based merging fails. We analyze whether weighted combinations can preserve style-critical directions. Let the style subspace be spanned by the dominant singular vectors of ΔW_s, with projection operator $P = V_k V_k^\top$.

We decompose the content LoRA as

$$\Delta W_c = \Delta W_c (I - P) + \Delta W_c P, \tag{5}$$

where $\Delta W_c P$ lies in the style subspace.

For a weighted merge $\Delta W_m = a \Delta W_c + b \Delta W_s$, its projection onto the style subspace is

$$\Delta W_m P = a\, \Delta W_c P + b\, \Delta W_s. \tag{6}$$

For nonzero a, preserving the original style component would require $a\, \Delta W_c P \approx (1 - b)\, \Delta W_s$. Because the same coefficient a scales content contributions both within and outside the style subspace, weighted merging cannot adjust the two independently.

4.2 Hard Projection for Subspace Control

Having identified the dominant style directions and observed that weight-based merging does not explicitly account for their interactions, we introduce a projection-based formulation to modulate cross-LoRA contributions. We use the style LoRA as the reference because, in our evaluated subject–style pairs, direct merging tends to weaken stylistic characteristics more than subject structure.

We define the projection onto the dominant style subspace as $P = V_k V_k^\top$, where V_k spans the style subspace. In practice, k is set to the LoRA rank, so that the full low-rank subspace is considered without introducing additional selection hyperparameters. The null space projection is given by:

$$P_{\text{null}} = I - P = I - V_k V_k^\top, \tag{7}$$

which suppresses components aligned with style-critical directions while retaining complementary information.

Applying this projection to the content LoRA $\Delta W_c \in \mathbb{R}^{m\times n}$ yields:

$$\Delta W_c^{\text{proj}} = \Delta W_c P_{\text{null}} = \Delta W_c (I - V_k V_k^\top). \tag{8}$$

The merged LoRA is then defined as:

$$\begin{aligned} \Delta W_m &= \Delta W_s + \Delta W_c^{\text{proj}} \\ &= \Delta W_s + \Delta W_c (I - V_k V_k^\top). \end{aligned} \tag{9}$$

This formulation restricts the contribution of the content LoRA along dominant style directions, allowing the style LoRA ΔW_s to primarily govern these subspaces.

From a geometric perspective, this can be viewed as reducing the overlap between content and style subspaces along critical directions. In the limiting case, this projection minimizes the contribution of ΔW_c within the style subspace, which can be quantified by:

$$\|\Delta W_c^{\text{proj}} P\|_F^2, \tag{10}$$

representing the residual energy of the projected content LoRA within the style subspace.

4.3 Soft Projection for Balanced Fusion

Instead of treating projection as a binary operation, we reinterpret LoRA fusion as a continuous subspace interaction problem. In practice, content and style LoRAs often share partially aligned directions, and hard projection may overly suppress useful components. This motivates a flexible mechanism that modulates their interaction rather than strictly eliminating overlap.

We introduce a subspace-aware formulation:

$$\min_{\Delta W_c^{\text{proj}}} \|\Delta W_c^{\text{proj}} - \Delta W_c\|_F^2 + \mu \|\Delta W_c^{\text{proj}} P\|_F^2, \tag{11}$$

where the first term preserves content, and the second suppresses its energy within the style subspace. The parameter $\mu \geq 0$ controls the strength of this subspace interaction.

This optimization admits a closed-form solution, which leads directly to the merged LoRA:

$$\Delta W_m = \Delta W_s + \Delta W_c (I + \mu P)^{-1} = \Delta W_s + \Delta W_c \Big(I - \tfrac{\mu}{1+\mu} V_k V_k^\top \Big). \tag{12}$$

Equivalently, μ removes a fraction $\mu/(1+\mu)$ of the content component within the style subspace, retaining the remaining $1/(1+\mu)$, while leaving its complementary component unchanged; Tab. 1 shows the resulting trade-off.

We apply this modulation to the content LoRA while using the style subspace as the reference for controlling their interaction. As $\mu \to 0$, the formulation reduces to linear merging, while $\mu \to \infty$ approaches hard projection. Thus, μ provides continuous control over the trade-off between content fidelity and style preservation. Unlike strict orthogonal projection, our formulation continuously attenuates overlapping content and style components while preserving complementary ones. This allows effective fusion when the LoRA subspaces are partially aligned, without requiring complete separation.

5 Experiments

5.1 Experiment Setup

Datasets. Following common practice in customization Shah et al. [2024], Ouyang et al. [2025], we obtain LoRAs from publicly available images. For content, we adopt the DreamBooth Ruiz et al. [2023b] dataset, where each subject is represented by 4–5 reference images. For style, we use the StyleDrop Sohn et al. [2023] dataset, complemented with several representative artistic styles. For community-trained LoRAs, we use publicly available models from Hugging Face for evaluation as shown in Appendix.

Experimental details. The main experiments use SDXL v1.0 Podell et al. [2023]; the FLUX evaluation is reported separately. For training the LoRAs to be merged, including the training prompts

and ranks, we follow the settings in K-LoRA Ouyang et al. [2025], Ruiz et al. [2023a]. Importantly, NP-LoRA does not assume clean disentanglement or orthogonality between content and style LoRAs, but instead is designed to handle their interactions under potentially overlapping and entangled representations. We use a global $\mu = 0.5$ in the soft projection (Eq. 12) across the evaluated LoRA pairs and categories, including rank-16 and rank-64 experiments; the parameter controls the trade-off between subject preservation and style transfer (Sec. 5.3). In practice, the projection subspace is constructed using all singular directions of the style LoRA, i.e., we set k equal to the LoRA rank. Implementation details are provided in the Appendix. Experiments based on SDXL are performed on an NVIDIA RTX 3090 GPU.

Baselines. We compare our method with representative approaches for combining or learning multiple LoRAs. Among them, most methods merge two independently trained LoRAs, i.e., the model takes two trained LoRAs as input and produces merged generations. Specifically, we evaluate (i) the direct weight merging baseline $\Delta W_m = \Delta W_c + \Delta W_s$, (ii) ZipLoRA Shah et al. [2024] (ECCV2024), (iii) DuoLoRA Roy et al. [2025] (ICCV2025), (iv) K-LoRA Ouyang et al. [2025] (CVPR2025), and (v) LoRA.rar Shenaj et al. [2025] (ICCV2025).

5.2 Asymmetric Projection for Style Preservation

We first analyze which LoRA should be protected during projection. Direct merging consistently preserves subject structure better than style, suggesting that style representations are more susceptible to cross-LoRA interference. This motivates protecting the style LoRA during fusion. We compare two projection directions: projecting the style LoRA with respect to the content subspace, and projecting the content LoRA with respect to the dominant style subspace. As shown in Fig. 4, the former substantially weakens stylistic characteristics, while the latter achieves a better content–style balance. Based on this observation, NP-LoRA adopts an asymmetric projection strategy that preserves style-critical directions while suppressing conflicting content components. This observation is also consistent with Tab. 2 and Fig. 6, where direct merging exhibits a clear bias toward content over style.

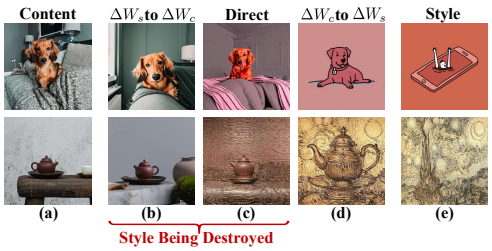


Figure 4: Comparison of LoRA projection. Projecting style into the content null space or direct merging degrades style, while projecting content into the style null space better balances content and style.

5.3 Ablation on Projection Strength

We analyze the effect of the projection strength controlled by μ, which adjusts the trade-off between subject fidelity and style consistency, as shown in Fig. 5. As μ increases from 0 to ∞, the suppression of content components aligned with the style subspace becomes stronger: $\mu \to 0$ corresponds to direct weight-based merging, while $\mu \to \infty$ approaches hard projection. We evaluate NP-LoRA with $\mu \in \{0, 0.1, 0.5, 1, 10, \infty\}$ to study this transition.

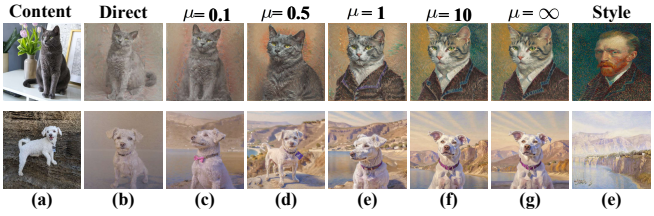


Figure 5: Effect of the projection strength μ in NP-LoRA. (a) and (e) are the content and style references, respectively. (c) corresponds to direct merging ($\mu = 0$), while (d)–(g) show increasing projection strengths from soft to hard projection. Smaller μ values tend to exhibit stronger content–style interference, whereas larger μ values better preserve style but may weaken content fidelity.

The results show a clear trade-off. Smaller μ values retain more content information but may weaken style consistency, whereas larger μ values better preserve style but can suppress content details. In particular, hard projection preserves the style subspace most aggressively, but often loses fine-grained subject information, highlighting the need for a softer projection strategy.

Empirically, $\mu = 0.5$ provides the best overall balance and is used as the default setting. The quantitative sweep in Tab. 1 shows how the metrics change across projection strengths.

The sweep shows a continuous content–style trade-off: style similarity increases as μ grows, while content similarity decreases. The default $\mu = 0.5$ attains the highest arithmetic mean and ties for the highest harmonic mean among the evaluated settings. For the evaluated pairs, $\mu \in [0.5, 1]$ provides a practical starting range for balancing content and style.

Table 1: Projection-strength sensitivity on the same 32 LoRA pairs and evaluation protocol as Tab. 2. Increasing μ attenuates the content component aligned with the style subspace.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
$\mu = 0$	0.75	0.55	0.54	0.23	0.51	0.42
$\mu = 0.1$	0.74	0.57	0.53	0.25	0.53	0.45
$\mu = 0.5$	0.73	0.59	0.52	0.33	**0.55**	**0.50**
$\mu = 1$	0.71	0.60	0.48	0.35	0.53	**0.50**
$\mu = 10$	0.69	0.60	0.40	0.37	0.52	0.48
$\mu \to \infty$	0.68	0.60	0.39	0.40	0.52	0.48

Table 2: Quantitative and preference-based evaluation of content–style LoRA fusion. (a) Automatic evaluation using CLIP/DINO similarities, with aggregated scores (S_{arith}, S_{harm}) summarizing the overall trade-off. (b) Preference evaluation by human users and GPT-5, reporting the proportion of cases where each method is selected as the best.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
Direct	0.75	0.55	0.54	0.23	0.51	0.42
ZipLoRA	0.75	0.56	0.55	0.24	0.52	0.44
DuoLoRA	0.75	0.55	0.56	0.26	0.53	0.46
K-LoRA	**0.76**	0.55	**0.57**	0.22	0.52	0.42
LoRA.rar	0.70	0.53	0.50	0.30	0.51	0.46
NP-LoRA	0.73	**0.59**	0.52	**0.33**	**0.55**	**0.50**

(a) Quantitative similarity metrics

Method	User ↑	GPT-5 ↑
Direct	8.44%	6.25%
ZipLoRA	9.22%	9.38%
DuoLoRA	11.09%	15.63%
K-LoRA	10.63%	9.38%
LoRA.rar	12.03%	12.5%
NP-LoRA	**48.59%**	**46.88%**

(b) Preference evaluation

5.4 Results

Quantitative comparisons. Unless otherwise stated, the experiments in this section use SDXL. Following prior works such as K-LoRA Ouyang et al. [2025] (18 pairs) and ZipLoRA Shah et al. [2024] (32 pairs), we evaluate NP-LoRA on 32 randomly selected content–style LoRA pairs, covering 5 animal subjects, 5 object-level categories, and 15 distinct artistic styles. For each pair, we generate 10 images with shared random seeds across all methods and report results averaged over pairs and seeds. Unless otherwise specified, all LoRAs are rank-8; results for other ranks are in the appendix. Unlike idealized settings, our evaluation includes pairs with substantial subspace overlap (often up to 98%, see Appendix), reflecting realistic entangled LoRA representations.

For evaluation, following prior work Shah et al. [2024], Ouyang et al. [2025], we adopt CLIP Radford et al. [2021] and DINO Zhang et al. [2022] similarity metrics to jointly assess subject preservation (content) and style alignment (style) Jiang et al. [2024], Shah et al. [2024], Ouyang et al. [2025]. Content scores compare generated images with subject references, while style scores compare them with style references; CLIP and DINO provide complementary measures of semantic alignment and visual-feature similarity. Stylization can change texture and local appearance even when subject identity remains recognizable; we therefore interpret the four similarities jointly with human preferences. Specifically, for CLIP we compute $S_{\text{content}}^{\text{CLIP}}$ and $S_{\text{style}}^{\text{CLIP}}$, and for DINO we compute $S_{\text{content}}^{\text{DINO}}$ and $S_{\text{style}}^{\text{DINO}}$. To obtain unified content-style similarity measures, we report both the arithmetic mean and the harmonic mean following the common F1-style aggregation scheme Van Rijsbergen [1979], Chinchor [1992], Ristani et al. [2016], Xian et al. [2017]. Given the four similarity scores $S_i \in \{S_{\text{content}}^{\text{CLIP}}, S_{\text{style}}^{\text{CLIP}}, S_{\text{content}}^{\text{DINO}}, S_{\text{style}}^{\text{DINO}}\}$, we report both the arithmetic mean $S_{\text{arith}} = \frac{1}{4}\sum_i S_i$ and the harmonic mean $S_{\text{harm}} = \frac{4}{\sum_i (1/S_i)}$, where S_{arith} reflects the overall trend, while the harmonic mean S_{harm} emphasizes joint fidelity to both content and style. As shown in Tab. 2 (a), our method achieves the highest S_{arith} and S_{harm}, showing its superior ability to balance subject fidelity and style consistency. Additional rank settings and comparisons with training-based methods are provided in the Appendix. In particular, we include results on real-world, community-trained LoRAs, which often exhibit entangled semantics and better reflect practical deployment scenarios.



Figure 6: Qualitative comparison with Direct Weight Merge, ZipLoRA, DuoLoRA, K-LoRA, LoRA.rar, and our proposed NP-LoRA, illustrating the trade-off between subject fidelity and style preservation.

Following K-LoRA Ouyang et al. [2025], we further evaluated perceptual quality through a user study involving 20 participants. The study was conducted on images generated from 32 merged LoRA pairs (Tab. 2 (b)), with 32 images sampled from each pair. Each case included outputs from SOTA baselines and our method, along with reference subject and style images. Participants selected the result best balancing subject fidelity and style consistency. As shown in Tab. 2 (b), our method was preferred in 48.59% of cases. We further adopt a GPT-5-based automatic evaluation protocol, which is more recent than the MLLM evaluator used in Shenaj et al. [2025]. To reduce potential positional and label bias, we randomize the order of candidate images and anonymize method identifiers during evaluation, restoring method names only after collecting the results. The results show a consistent 46.88% preference in favor of our method. Across the six evaluated methods, aggregate human and GPT-5 preference rates are strongly correlated (Pearson $r = 0.988$; Spearman $\rho = 0.928$).

Table 3: Efficiency comparison of LoRA merging strategies. Merge includes loading time; Gen./Img. is per-image generation time. LoRA.rar Shenaj et al. [2025] is excluded for its needs pretraining.

Method	Merge ($s^{\dagger}$)↓	Gen./Img. (s)↓
Direct	13.47	22.90
ZipLoRA	>10 min	29.93
DuoLoRA	>5 min	23.93
K-LoRA	12.25*	60.39*
NP-LoRA	13.44	23.21

* Only includes LoRA loading time, as K-LoRA performs on-the-fly merging during generation.

Efficiency Comparison. To evaluate efficiency, we compare both the merging and generation times in Tab. 3. LoRA.rar Shenaj et al. [2025] is excluded because it depends on pretrained models for parameter fusion. ZipLoRA Shah et al. [2024] is the slowest due to additional fusion optimization, whereas K-LoRA Ouyang et al. [2025] is efficient during initialization but slows down for dynamic fusion in the sampling stage. In contrast, NP-LoRA maintains merging and inference efficiency comparable to Direct Merge baseline, yet consistently delivers superior results.

Qualitative comparisons. Qualitative results are shown in Fig. 6. Direct merging and existing methods Shah et al. [2024], Roy et al. [2025], Ouyang et al. [2025], Shenaj et al. [2025] achieve only partial fusion, typically preserving either content or style but failing to balance both consistently. In contrast, NP-LoRA achieves the best results, achieving a superior balance between content fidelity and style preservation. A qualitative comparison with jointly training method in the Appendix shows that NP-LoRA attains superior results.

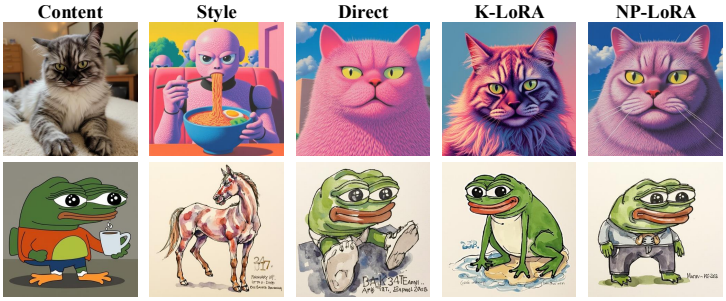


Figure 7: Qualitative comparison on FLUX. Each row shows a subject–style pair and results from Direct Merge, K-LoRA, and NP-LoRA.

Generalization to FLUX.

We further evaluate NP-LoRA on the FLUX backbone to examine its applicability beyond SDXL. We compare it with Direct Merge and K-LoRA Ouyang et al. [2025]; the latter provides a public FLUX implementation. As shown in Fig. 7, all three methods produce recognizable subject–style compositions, while NP-LoRA shows a slightly better balance between subject characteristics and the target style in these examples. The matched quantitative comparison in Tab. 4 supports this trend: NP-LoRA attains the highest aggregate scores ($S_{\text{arith}} = 0.56$, $S_{\text{harm}} = 0.54$) and the highest style similarities among the three methods.

Table 4: Matched quantitative comparison on FLUX using the same LoRA pairs, prompts, and random seeds.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
Direct	0.71	0.55	0.59	0.36	0.55	0.52
K-LoRA	0.73	0.54	0.62	0.32	0.55	0.50
NP-LoRA	0.70	0.57	0.58	0.40	**0.56**	**0.54**

6 Conclusions

We propose NP-LoRA, a training-free framework for post-hoc fusion of independently trained content and style LoRAs. We view LoRA fusion as a subspace interaction problem and use projection to modulate content contributions along dominant style directions. With a soft projection mechanism, NP-LoRA provides continuous control over the trade-off between subject fidelity and style consistency. Our analysis and experiments show that content and style LoRAs exhibit substantial subspace interaction, and that controlling these interactions leads to more balanced fusion than weight-based merging. Extensive evaluations, including quantitative metrics, user studies, and GPT-5-based assessments with randomized and anonymized protocols, show that NP-LoRA consistently improves content-style composition while remaining training-free and computationally efficient.

Limitations. Like prior methods Shah et al. [2024], Ouyang et al. [2025], Shenaj et al. [2025], NP-LoRA treats layers independently, which may leave cross-layer interference unresolved. We evaluate only pairwise subject–style fusion; handling multiple content LoRAs or a joint style subspace remains future work.

Acknowledgments and Disclosure of Funding

This study was supported by the National Natural Science Foundation of China (No. 62176252). The authors declare no competing interests.

References

Yuval Alaluf, Elad Richardson, Gal Metzer, and Daniel Cohen-Or. A neural space-time representation for text-to-image personalization. *ACM Transactions on Graphics (TOG)*, 42(6):1–10, 2023.

Omri Avrahami, Kfir Aberman, Ohad Fried, Daniel Cohen-Or, and Dani Lischinski. Break-a-scene: Extracting multiple concepts from a single image. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–12, 2023.

Shubhankar Borse, Kartikeya Bhardwaj, Mohammad Reza Karimi Dastjerdi, Hyojin Park, Shreya Kadambi, Shobitha Shivakumar, Prathamesh Mandke, Ankita Nayak, Harris Teague, Munawar Hayat, et al. Subzero: Composing subject, style, and action via zero-shot personalization. *arXiv preprint arXiv:2502.19673*, 2025.

Nancy Chinchor. Muc-4 evaluation metrics. In *Proceedings of the 4th Conference on Message Understanding*, MUC4 '92, page 22–29, USA, 1992. Association for Computational Linguistics. ISBN 1558602739. doi: 10.3115/1072064.1072067. URL https://doi.org/10.3115/1072064.1072067.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf.

Jiahua Dong, Wenqi Liang, Hongliu Li, Duzhen Zhang, Meng Cao, Henghui Ding, Salman H Khan, and Fahad Shahbaz Khan. How to continually adapt text-to-image diffusion models for flexible customization? *Advances in Neural Information Processing Systems*, 37:130057–130083, 2024.

Junfeng Fang, Houcheng Jiang, Kun Wang, Yunshan Ma, Jie Shi, Xiang Wang, Xiangnan He, and Tat-Seng Chua. Alphaedit: Null-space constrained knowledge editing for language models. In *The Thirteenth International Conference on Learning Representations*.

Farzad Farhadzadeh, Debasmit Das, Shubhankar Borse, and Fatih Porikli. Zero-shot adaptation of parameter-efficient fine-tuning in diffusion models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 16144–16160. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/farhadzadeh25a.html.

Niki Foteinopoulou, Ignas Budvytis, and Stephan Liwicki. Loratorio: An intrinsic approach to lora skill composition, 2025. URL https://arxiv.org/abs/2508.11624.

Yarden Frenkel, Yael Vinker, Ariel Shamir, and Daniel Cohen-Or. Implicit style-content separation using b-lora. In *European Conference on Computer Vision*, pages 181–198. Springer, 2024.

Yuchao Gu, Xintao Wang, Jay Zhangjie Wu, Yujun Shi, Yunpeng Chen, Zihan Fan, Wuyou Xiao, Rui Zhao, Shuning Chang, Weijia Wu, et al. Mix-of-show: Decentralized low-rank adaptation for multi-concept customization of diffusion models. *Advances in Neural Information Processing Systems*, 36:15890–15902, 2023.

Soufiane Hayou, Nikhil Ghosh, and Bin Yu. Lora+: Efficient low rank adaptation of large models. *arXiv preprint arXiv:2402.12354*, 2024.

Chunming He, Chengyu Fang, Yulun Zhang, Longxiang Tang, Jinfa Huang, Kai Li, zhenhua guo, Xiu Li, and Sina Farsiu. Reti-Diff: Illumination degradation image restoration with retinex-based latent diffusion model. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Learning Representations*, volume 2025, pages 43332–43352, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/6b9980876e7335e359df01911b5107d5-Paper-Conference.pdf.

Chunming He, Fengyang Xiao, Rihan Zhang, Chengyu Fang, Deng-Ping Fan, and Sina Farsiu. Reversible unfolding network for concealed visual perception with generative refinement, 2025b. URL https://arxiv.org/abs/2508.15027.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

Dongsheng Jiang, Yuchen Liu, Songlin Liu, Jin'e Zhao, Hao Zhang, Zhen Gao, Xiaopeng Zhang, Jin Li, and Hongkai Xiong. From clip to dino: Visual encoders shout in multi-modal large language models, 2024. URL https://arxiv.org/abs/2310.08825.

Jiaxiu Jiang, Yabo Zhang, Kailai Feng, Xiaohe Wu, Wenbo Li, Renjing Pei, Fan Li, and Wangmeng Zuo. Mcˆ 2: Multi-concept guidance for customized multi-concept generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2802–2812, 2025.

Dawid J Kopiczko, Tijmen Blankevoort, and Yuki M Asano. Vera: Vector-based random matrix adaptation. *arXiv preprint arXiv:2310.11454*, 2023.

Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1931–1941, 2023.

Zhiheng Liu, Ruili Feng, Kai Zhu, Yifei Zhang, Kecheng Zheng, Yu Liu, Deli Zhao, Jingren Zhou, and Yang Cao. Cones: Concept neurons in diffusion models for customized generation. *arXiv preprint arXiv:2303.05125*, 2023.

Hila Manor, Rinon Gal, Haggai Maron, Tomer Michaeli, and Gal Chechik. Spanning the visual analogy space with a weight basis of loras. In Paolo Favaro, Zuzana Kukelova, Atsuto Maki, Anna Rohrbach, Konrad Schindler, and Federico Tombari, editors, *Computer Vision – ECCV 2026*, pages 323–341, Cham, 2026. Springer Nature Switzerland. ISBN 978-3-032-37372-4.

Zhendong Mao, Mengqi Huang, Fei Ding, Mingcong Liu, Qian He, and Yongdong Zhang. RealCustom++: Representing images as real textual word for real-time customization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(2):2078–2095, 2026. doi: 10.1109/TPAMI.2025.3623025.

Tuna Han Salih Meral, Enis Simsar, Federico Tombari, and Pinar Yanardag. Contrastive test-time composition of multiple lora models for image generation. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 18090–18100, 2025. doi: 10.1109/ICCV51701.2025.01681.

Ziheng Ouyang, Zhen Li, and Qibin Hou. K-lora: Unlocking training-free fusion of any subject and style loras. In *CVPR*, 2025.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Pengjie Ren, Chengshun Shi, Shiguang Wu, Mengqi Zhang, Zhaochun Ren, Maarten de Rijke, Zhumin Chen, and Jiahuan Pei. Melora: Mini-ensemble low-rank adapters for parameter-efficient fine-tuning. *arXiv preprint arXiv:2402.17263*, 2024.

Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 17–35, 2016. Proposes IDF1, a harmonic mean of ID precision and recall for tracking evaluation.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

Aniket Roy, Shubhankar Borse, Shreya Kadambi, Debasmit Das, Shweta Mahajan, Risheek Garrepalli, Hyojin Park, Ankita Nayak, Rama Chellappa, Munawar Hayat, et al. Duolora: Cycle-consistent and rank-disentangled content-style personalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15395–15404, 2025.

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, 2023a.

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, 2023b.

Viraj Shah, Nataniel Ruiz, Forrester Cole, Erika Lu, Svetlana Lazebnik, Yuanzhen Li, and Varun Jampani. Ziplora: Any subject in any style by effectively merging loras. In *European Conference on Computer Vision*, pages 422–438. Springer, 2024.

Donald Shenaj, Ondrej Bohdal, Mete Ozay, Pietro Zanuttigh, and Umberto Michieli. Lora.rar: Learning to merge loras via hypernetworks for subject-style conditioned image generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2025.

Jing Shi, Wei Xiong, Zhe Lin, and Hyun Joon Jung. Instantbooth: Personalized text-to-image generation without test-time finetuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8543–8552, 2024.

Enis Simsar, Thomas Hofmann, Federico Tombari, and Pinar Yanardag. Loraclr: Contrastive adaptation for customization of diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 13189–13198, 2025.

Kihyuk Sohn, Nataniel Ruiz, Kimin Lee, Daniel Castro Chin, Irina Blok, Huiwen Chang, Jarred Barber, Lu Jiang, Glenn Entis, Yuanzhen Li, et al. Styledrop: Text-to-image generation in any style. *arXiv preprint arXiv:2306.00983*, 2023.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=St1giarCHLP`.

George Stoica, Pratik Ramesh, Boglarka Ecsedi, Leshem Choshen, and Judy Hoffman. Model merging with svd to tie the knots. In *The Thirteenth International Conference on Learning Representations*.

Pengwei Tang, Yong Liu, Dongjie Zhang, Xing Wu, and Debing Zhang. Lora-null: Low-rank adaptation via null space for large language models. *arXiv preprint arXiv:2503.02659*, 2025.

C.J. Van Rijsbergen. *Information Retrieval*. Butterworths, 1979. ISBN 9780408709293. URL `https://books.google.co.jp/books?id=t-pTAAAAMAAJ`.

Andrey Voynov, Qinghao Chu, Daniel Cohen-Or, and Kfir Aberman. p+: Extended textual conditioning in text-to-image generation. *arXiv preprint arXiv:2303.09522*, 2023.

Shaojin Wu, Mengqi Huang, Yufeng Cheng, Wenxu Wu, Jiahe Tian, Yiming Luo, Fei Ding, and Qian He. Unified customized generation by disentangled reward modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 34418–34427, 2026.

Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. Zero-shot learning – the good, the bad and the ugly. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4582–4591, 2017. Introduces the harmonic mean (H-score) of seen/unseen accuracies for balanced evaluation in generalized zero-shot learning.

Guangxuan Xiao, Tianwei Yin, William T Freeman, Frédo Durand, and Song Han. Fastcomposer: Tuning-free multi-subject image generation with localized attention. *International Journal of Computer Vision*, 133(3):1175–1194, 2025.

Shaoan Xie, Zhifei Zhang, Zhe Lin, Tobias Hinz, and Kun Zhang. Smartbrush: Text and shape guided object inpainting with diffusion model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22428–22437, 2023.

Peng Xing, Haofan Wang, Yanpeng Sun, Qixun Wang, Xu Bai, Hao Ai, Renyuan Huang, and Zechao Li. Csgo: Content-style composition in text-to-image generation. *arXiv preprint arXiv:2408.16766*, 2024.

Yifeng Xiong and Xiaohui Xie. Oplora: Orthogonal projection lora prevents catastrophic forgetting during parameter-efficient fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 34088–34096, 2026.

Yiyan Xu, Qiulin Wang, Wenjie Wang, Yunyao Mao, Xintao Wang, Pengfei Wan, Kun Gai, and Fuli Feng. UniCustom: Unified visual conditioning for multi-reference image generation. *arXiv preprint arXiv:2605.12088*, 2026. URL `https://arxiv.org/abs/2605.12088`.

Yu Xu, Fan Tang, Juan Cao, Yuxin Zhang, Oliver Deussen, Weiming Dong, Jintao Li, and Tong-Yee Lee. B4m: Breaking low-rank adapter for making content-style customization. *ACM Transactions on Graphics*, 44(2):1–17, 2025. doi: 10.1145/3728461.

Yang Yang, Wen Wang, Liang Peng, Chaotian Song, Yao Chen, Hengjia Li, Xiaolong Yang, Qinglin Lu, Deng Cai, Boxi Wu, et al. Lora-composer: Leveraging low-rank adaptation for multi-concept customization in training-free diffusion models. *arXiv preprint arXiv:2403.11627*, 2024.

Andi Zhang, Xuan Ding, Haofan Wang, Steven McDonagh, and Samuel Kaski. Rethinking inter-lora orthogonality in adapter merging: Insights from orthogonal monte carlo dropout. *arXiv preprint arXiv:2510.03262*, 2025a.

Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

Jia-Chen Zhang and Yu-Jie Xiong. Subject or style: Adaptive and training-free mixture of loras. *arXiv preprint arXiv:2508.02165*, 2025.

Longteng Zhang, Lin Zhang, Shaohuai Shi, Xiaowen Chu, and Bo Li. Lora-fa: Memory-efficient low-rank adaptation for large language models fine-tuning. *arXiv preprint arXiv:2308.03303*, 2023a.

Yisu Zhang, Chenjie Cao, Chaohui Yu, and Jianke Zhu. Lion-lora: Rethinking lora fusion to unify controllable spatial and temporal generation for video diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14569–14579, 2025b.

Yuxin Zhang, Nisha Huang, Fan Tang, Haibin Huang, Chongyang Ma, Weiming Dong, and Changsheng Xu. Inversion-based style transfer with diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10146–10156, 2023b.

Ming Zhong, Yelong Shen, Shuohang Wang, Yadong Lu, Yizhu Jiao, Siru Ouyang, Donghan Yu, Jiawei Han, and Weizhu Chen. Multi-lora composition for image generation. *Transactions on Machine Learning Research*, 2024, 2024a.

Ming Zhong, yelong shen, Shuohang Wang, Yadong Lu, Yizhu Jiao, Siru Ouyang, Donghan Yu, Jiawei Han, and Weizhu Chen. Multi-loRA composition for image generation. *Transactions on Machine Learning Research*, 2024b. ISSN 2835-8856. URL `https://openreview.net/forum?id=25FT0DqhVZ`.

Hongyun Zhou, Xiangyu Lu, Wang Xu, Conghui Zhu, Tiejun Zhao, and Muyun Yang. Lora-drop: Efficient lora parameter pruning based on output evaluation. *arXiv preprint arXiv:2402.07721*, 2024.

Bojia Zi, Xianbiao Qi, Lingzhi Wang, Jianan Wang, Kam-Fai Wong, and Lei Zhang. Delta-lora: Fine-tuning high-rank parameters with the delta of low-rank matrices. *arXiv preprint arXiv:2309.02411*, 2023.

Xiandong Zou, Mingzhu Shen, Christos-Savvas Bouganis, and Yiren Zhao. Cached multi-lora composition for multi-concept image generation. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu, editors, *International Conference on Learning Representations*, volume 2025, pages 75638–75666, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/bc8f76d9caadd48f77025b1c889d2e2d-Paper-Conference.pdf.

Xiandong Zou, Mingzhu Shen, Christos-Savvas Bouganis, and Yiren Zhao. Cached multi-lora composition for multi-concept image generation. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=4iFSBgxvIO.

Appendix Overview

This appendix provides additional technical details, derivations, and extended analyses to support the findings presented in the main paper. Specifically, we first derive the formulation of the proposed soft projection operator in Sec. A and present details for efficient implementation in Sec. C. Sec. D investigates which null-space construction (U-space or V-space) better preserves stylistic fidelity. Sec. G provides an additional comparison with joint training to further validate the effectiveness of NP-LoRA over training-based approaches. Sec. H demonstrates the controllability of our method under various prompt conditions, while Sec. I evaluates its robustness across different random seeds. Sec. J extends our validation to the Flux model to assess cross-model generalization. Finally, Sec. L details the GPT-5 evaluation protocol used for quantitative preference assessment.

Together, these sections offer deeper insights into NP-LoRA's formulation, implementation, and empirical robustness beyond the main paper.

A Derivation of the Soft Projection Operator

We start from the relaxed objective introduced in the main text:

$$\min_{\Delta W_c^{\text{proj}}} \|\Delta W_c^{\text{proj}} - \Delta W_c\|_F^2 + \mu \|\Delta W_c^{\text{proj}} P\|_F^2, \tag{13}$$

where the first term preserves proximity to the original content adapter, and the second penalizes its energy within the style subspace defined by $P = V_k V_k^\top$.

Closed-form solution. Taking the derivative of (13) with respect to ΔW_c^{proj} and setting it to zero yields

$$\Delta W_c^{\text{proj}} - \Delta W_c + \mu \, \Delta W_c^{\text{proj}} P = 0, \tag{14}$$

which can be rearranged as

$$\Delta W_c^{\text{proj}} (I + \mu P) = \Delta W_c. \tag{15}$$

Thus, the solution is given by

$$\Delta W_c^{\text{proj}} = \Delta W_c (I + \mu P)^{-1}. \tag{16}$$

Simplifying the inverse. Since $P = V_k V_k^\top$ is a rank-k projector with $V_k^\top V_k = I_k$, the inverse admits the closed form

$$(I + \mu P)^{-1} = I - \tfrac{\mu}{1+\mu} V_k V_k^\top, \tag{17}$$

which follows from spectral decomposition or the Woodbury identity.

Substituting (17) into (16), we obtain

$$\Delta W_c^{\text{proj}} = \Delta W_c \Big(I - \tfrac{\mu}{1+\mu} V_k V_k^\top \Big). \tag{18}$$

Final fusion rule. The merged LoRA is then defined as

$$\Delta W_m = \Delta W_s + \Delta W_c^{\text{proj}} = \Delta W_s + \Delta W_c\Big(I - \tfrac{\mu}{1+\mu} V_k V_k^\top\Big). \tag{19}$$

This formulation continuously interpolates between direct merging ($\mu = 0$) and hard projection ($\mu \to \infty$), providing a controllable trade-off between style preservation and content retention.

B Randomly Sampling LoRA Pairs.

Following prior works including K-LoRA Ouyang et al. [2025] (18 pairs) and ZipLoRA Shah et al. [2024] (32 pairs), we evaluate our NP-LoRA on **32 randomly sampled content–style LoRA pairs**, covering **5 animal subjects**, **5 object-level categories**, and **15 distinct artistic styles**.

The specific content–style combinations are listed below:

`dog`	in watercolor painting style3, in kid crayon drawing style, in oil painting style3, in watercolor painting style2
cat2	in watercolor painting style3, in watercolor painting style2, in oil painting style1, in oil painting style2
dog2	in sticker style, in watercolor painting style3, in oil painting style3
vase	in line drawing style, in oil painting style3, in sticker style
candle	in 3d rendering style2, in kid crayon drawing style, in watercolor painting style2
colorful_sneaker	in watercolor painting style1, in line drawing style, in wooden sculpture
bear_plushie	in oil painting style2, in wooden sculpture, in 3d rendering style
dog5	in cartoon line drawing style, in glowing 3d rendering style
wolf_plushie	in flat cartoon illustration style, in abstract rainbow colored flowing smoke wave design, in 3d rendering style
dog8	in cartoon line drawing style, in flat cartoon illustration style, in glowing 3d rendering style, in sticker style

Content names follow the DreamBooth conventions Ruiz et al. [2023a]; style captions follow StyleDrop Sohn et al. [2023], with numeric suffixes indicating occurrence order among repeated entries in its `data.json`.

To further analyze the geometric relationship between independently trained LoRAs, we measure the overlap between two LoRA updates as the proportion of content-LoRA energy lying in the style subspace:

$$\text{Overlap}(A, B) = \frac{\|P_B \Delta W_A\|_F^2}{\|\Delta W_A\|_F^2}, \qquad P_B = V_B V_B^\top, \tag{20}$$

where V_B denotes the orthonormal basis of the right singular subspace of the style LoRA. Across randomly sampled content–style LoRA pairs, we observe consistently high overlap values, typically around 98%, suggesting substantial shared subspace structure between independently trained LoRAs. These overlap values should be interpreted as an empirical indicator of subspace correlation rather than a direct measure of interference strength.

C Implementation Details

In our theoretical formulation, we employ the singular value decomposition (SVD) to obtain the right singular subspace of the LoRA update matrix $\Delta W_s = B_s A_s$, where $A_s \in \mathbb{R}^{r\times n}$ and $B_s \in \mathbb{R}^{m\times r}$. Specifically, the original projection of ΔW_c onto the orthogonal complement of ΔW_s is computed using the right singular vectors V_k of ΔW_s, i.e.,

$$\Delta W_c^{\perp} = \Delta W_c (I - V_k V_k^\top). \tag{21}$$

However, directly computing V_k via SVD or PCA on the large matrix $\Delta W_s \in \mathbb{R}^{m\times n}$ is computationally expensive, as its complexity scales with $\mathcal{O}(mnr)$.

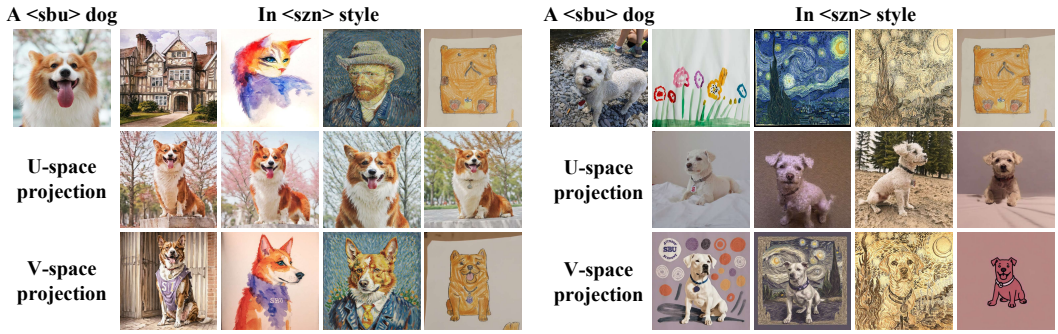


Figure 8: Comparison of output-Space (U) and parameter-space (V) projections for null-space construction. The U-space projection fails to remove style interference, leading to distorted stylistic features, whereas the V-space projection effectively eliminates such interference and preserves the intended style appearance.

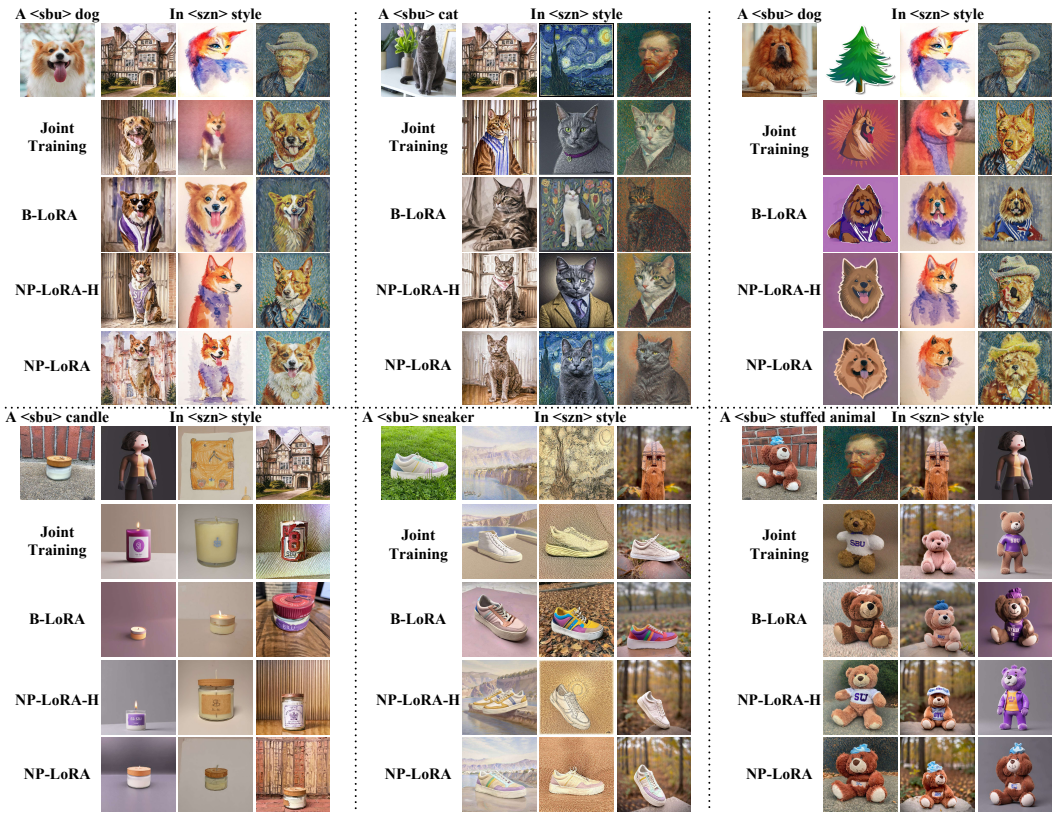


Figure 9: Qualitative comparison with joint training. Joint training exhibits unstable performance and often fails to merge content and style effectively, while N-LoRA achieves consistent and well-balanced fusion without retraining.

Efficient QR-based implementation. To improve efficiency, we replace the SVD with a QR-based projection that is mathematically equivalent but far more efficient. Observe that ΔW_s has rank at most r and its right singular subspace is equal to the column space of $A_s^\top$:

$$\text{span}(V) = \text{span}(A_s^\top). \tag{22}$$

Therefore, we compute a thin QR decomposition on $A_s^\top$:

$$A_s^\top = QR, \tag{23}$$

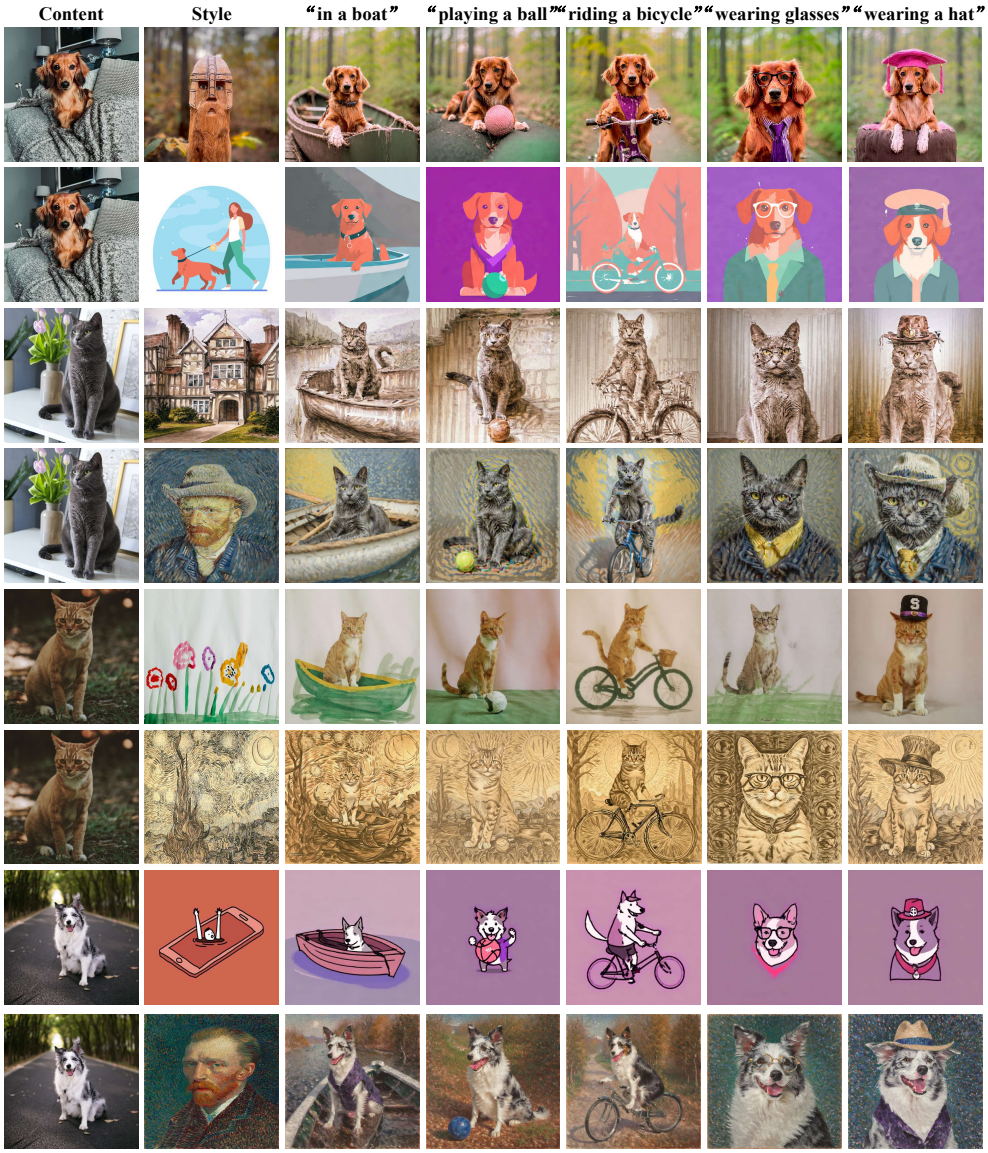


Figure 10: Our method effectively modifies the object's actions and environment while maintaining the original style.

where $Q \in \mathbb{R}^{n\times r}$ has orthonormal columns spanning $\mathrm{span}(A_s^\top)$, and $R \in \mathbb{R}^{r\times r}$ is an upper triangular matrix. We then replace V with Q and compute

$$\Delta W_c^{\text{proj}} = \Delta W_c - \Delta W_c Q Q^\top, \tag{24}$$

which requires only a QR decomposition of a small matrix of size $n \times r$ (where $r \ll n, m$), reducing complexity to $\mathcal{O}(nr^2)$.

Proof of equivalence. Let $\Delta W_s = B_s A_s$ denote the LoRA matrix. Since LoRA uses low-rank factors ($\mathrm{A}_s \in \mathbb{R}^{r\times n}$) and ($B_s \in \mathbb{R}^{m\times r}$) with full column and row rank (i.e., $rank = r$), ($B_s^\top B_s$) is symmetric positive definite. Then

$$\Delta W_s^\top \Delta W_s = A_s^\top (B_s^\top B_s) A_s. \tag{25}$$

Since $B_s^\top B_s$ is symmetric positive definite, it admits a Cholesky factorization $B_s^\top B_s = C^\top C$ with C invertible. Substituting gives

$$\Delta W_s^\top \Delta W_s = (C A_s)^\top (C A_s). \tag{26}$$

As C represents a full-rank linear transform, it does not alter the column space of $A_s^\top$. Hence, the right singular subspace of ΔW_s coincides with $\mathrm{span}(A_s^\top)$:

$$\mathrm{Col}(\Delta W_s^\top \Delta W_s) = \mathrm{span}(A_s^\top). \tag{27}$$

Therefore, the orthogonal projectors constructed via SVD and QR are identical:

$$P = V_k V_k^\top = QQ^\top. \tag{28}$$

In practice, this means that replacing the expensive low-rank SVD with a simple QR decomposition of $A_s^\top$ yields a mathematically equivalent projection while reducing computation time by an order of magnitude.

D Which Null-Space Construction Preserves Style: U-space or V-space?

In the main paper, we construct the null space using the right singular vectors V from the SVD decomposition $\Delta W_s = U\Sigma V^\top$, since interference in LoRA fusion occurs in the parameter space (i.e., the column space of ΔW_s) rather than in the activation space represented by U. Specifically, U spans the output (activation) domain of ΔW_s, describing how style LoRA influences feature activations, while V spans the input (parameter) domain, capturing how LoRA updates are distributed across the weight columns. Because LoRA fusion directly manipulates weight parameters, interference emerges along these column directions, making V the appropriate basis for constructing the null space. Projecting in the V-space thus directly eliminates content-induced perturbations along the style-critical directions within the weight domain. In contrast, constructing the null space based on U operates in the output (activation) space, which suppresses both content and style responses and weakens overall expressiveness. As shown in Fig. 8, projections built from V effectively preserve stylistic consistency, while those derived from U fail to disentangle content and style, resulting in noticeable style degradation.

E Additional Ablation on Soft Projection

We refer to the version without soft relaxation as the hard variant, denoted as NP-LoRA-H. The qualitative results are shown in Fig. 9, and the quantitative comparison is provided in Tab. 6. As illustrated, NP-LoRA-H effectively preserves the style but tends to lose content details. In contrast, our full NP-LoRA achieves a better balance between content and style, yielding the highest overall scores. This confirms the necessity of the proposed soft relaxation mechanism.

F Generalization to Different LoRA Ranks

We further conduct the same experiments on LoRAs with different ranks (rank-16 and rank-64). As shown in Tab. 5, the results exhibit consistent trends across both settings, leading to the same conclusion: NP-LoRA achieves a better balance between content preservation and style consistency compared to existing methods.

G Additional Comparison with Joint Training and B-LoRA

We further compare our method with a joint training baseline, where a new LoRA is trained from scratch using mixed content and style images. The qualitative results are shown in Fig. 9, and the quantitative comparison is provided in Tab. 6. As observed, the joint-training approach exhibits high instability: while it occasionally produces reasonable fusion, it often fails to effectively combine content and style. This instability may arise from catastrophic forgetting during multi-objective optimization, where learning new style features interferes with previously acquired content representations. We also observe that B-LoRA tends to preserve stylistic patterns but frequently loses content identity, indicating an imbalanced fusion behavior. In contrast, NP-LoRA achieves consistent and harmonious fusion without retraining, effectively preserving both content fidelity and stylistic characteristics. Quantitatively, as shown in Tab. 6, NP-LoRA outperforms joint training and B-LoRA across all metrics, achieving the highest overall score and demonstrating superior content–style preservation.

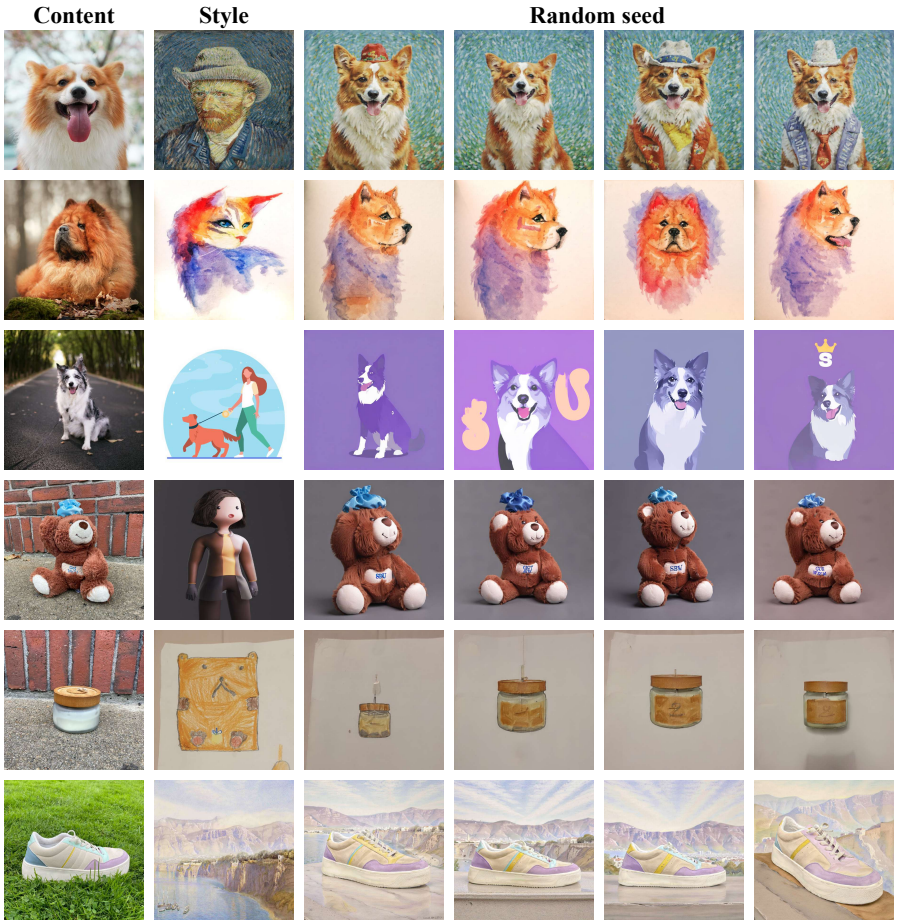


Figure 11: Results obtained with randomly selected seeds demonstrate the stability and robustness of our NP-LoRA.

H Prompt Control

We conduct experiments to assess whether our method can alter the object's actions, the surrounding environment, or introduce new elements by adjusting the prompts. As shown in 10, after modifying the prompts, our method effectively preserves the original object's characteristics and stylistic attributes, while seamlessly integrating new elements or scene details.

I Robustness Test

We evaluate the robustness of our method by running experiments with different random seeds to examine stability across stochastic conditions. As shown in Fig. 11, our approach produces consistently stable results under varying seeds, matching the qualitative trends observed in the main paper. Specifically, NP-LoRA achieves a balanced trade-off between content preservation and style consistency, yielding the most visually harmonious results.

J Further Generalization Validation on FLUX

To further validate the generality of NP-LoRA, we apply our method to the FLUX backbone using publicly available LoRAs covering diverse subjects and styles. As shown in Fig. 12 and Fig. 13, NP-LoRA consistently preserves a better balance between subject identity and stylistic fidelity across various domains. These results demonstrate that our projection-based fusion mechanism remains

Table 5: Comparison under different LoRA ranks.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
Direct	0.77	0.56	0.58	0.22	0.53	0.42
ZipLoRA	0.79	0.57	0.59	0.23	0.54	0.44
DuoLoRA	0.78	0.56	0.60	0.26	0.55	0.46
K-LoRA	0.80	0.55	0.60	0.22	0.54	0.43
LoRA.rar	0.73	0.53	0.53	0.31	0.52	0.47
NP-LoRA	0.78	0.60	0.58	0.30	0.56	0.50

(a) Rank 16 LoRAs

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
Direct	0.80	0.55	0.61	0.23	0.55	0.44
ZipLoRA	0.80	0.55	0.62	0.24	0.55	0.45
DuoLoRA	0.79	0.56	0.62	0.25	0.56	0.46
K-LoRA	0.82	0.54	0.63	0.23	0.55	0.44
LoRA.rar	0.73	0.54	0.55	0.31	0.53	0.48
NP-LoRA	0.79	0.60	0.61	0.29	0.57	0.50

(b) Rank 64 LoRAs

Table 6: Comparison with joint training methods under the same experimental setting as Tab. 2. Evaluation uses CLIP and DINO similarity scores for content preservation and style alignment. We emphasize the aggregated metrics (S_{arith} and S_{harm}), as they best reflect the overall trade-off between content and style.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$S_{\textbf{arith}} \uparrow$	$S_{\textbf{harm}} \uparrow$
Joint Training	0.67	0.61	0.49	0.25	0.50	0.43
B-LoRA	0.72	0.55	0.51	0.26	0.51	0.44
NP-LoRA-H	0.68	0.60	0.39	0.40	0.52	0.48
NP-LoRA	0.73	0.59	0.52	0.33	**0.55**	**0.50**

effective under different diffusion architectures, highlighting the robustness and general applicability of the proposed approach.

To further evaluate robustness under realistic conditions, we quantitatively compare NP-LoRA on community-trained FLUX LoRAs, where each content–style pair contains 10 generated images. These LoRAs are publicly available and often contain complex, entangled semantics, making fusion more challenging than cleanly trained LoRAs. Importantly, NP-LoRA does not require perfectly disentangled LoRAs; instead, it is designed to handle such entangled representations by modulating their interactions. As shown in Tab. 7, NP-LoRA outperforms K-LoRA, the strongest applicable baseline on FLUX, especially in style preservation and overall trade-off. This confirms that NP-LoRA remains effective for difficult and imperfectly disentangled LoRA pairs, and can robustly handle real-world LoRAs with mixed semantic and stylistic information.

K Human Preference Questionnaire

We provide the questionnaire interface used in the human preference study. The study involved 20 participants and was conducted on images generated from 32 merged LoRA pairs (Tab. 2(b)), with 32 generated samples evaluated for each pair. For each case, participants were presented with the outputs of NP-LoRA and competing state-of-the-art baselines, together with the corresponding reference subject and style images. Participants were asked to select the result that best balanced subject fidelity and style consistency. The retained records do not contain per-annotator responses, so inter-annotator agreement cannot be reliably computed.

The evaluation instructions shown to participants were as follows:

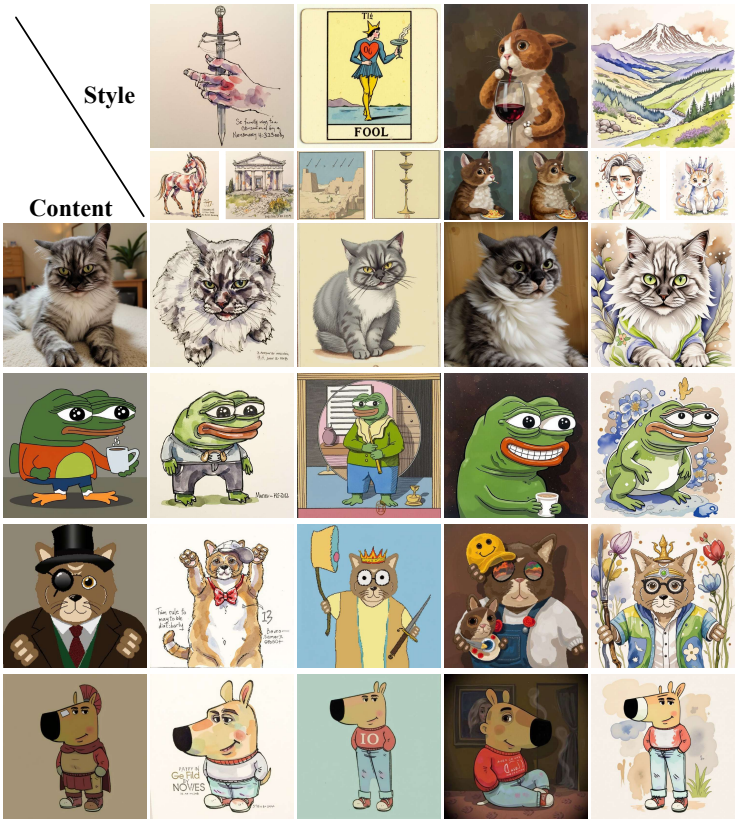


Figure 12: Qualitative results of NP-LoRA on the Flux backbone using diverse publicly available LoRAs. Each image corresponds to the combination of the content LoRA shown above and the style LoRA shown on the left, illustrating the results produced by our method.

Table 7: Quantitative comparison on community-trained FLUX LoRAs. NP-LoRA achieves a better overall content–style trade-off than K-LoRA, especially in terms of style preservation and harmonic aggregation.

Method	$S_{\text{content}}^{\text{CLIP}}$	$S_{\text{style}}^{\text{CLIP}}$	$S_{\text{content}}^{\text{DINO}}$	$S_{\text{style}}^{\text{DINO}}$	$\mathbf{S}_{\text{arith}}$	$\mathbf{S}_{\text{harm}}\uparrow$
K-LoRA	0.73	0.54	0.62	0.32	0.55	0.50
NP-LoRA	0.70	0.57	0.58	0.40	0.56	0.54

Evaluation Instructions. Each test group contains the original content image whose subject should be preserved, the original style image whose style should be fused, and fusion results from six anonymous methods. Please select the method that best preserves the content subject while successfully integrating the target style. After submission, your personal selection statistics will be displayed.

Prompt used for GPT-5 evaluation.

Task: Evaluate and select the candidate image that best balances Subject Fidelity and Style Consistency.

Inputs: - First image = A.jpg: Subject Fidelity reference. It shows the core subject/content that should be preserved. - Second image = B.jpg: Style reference. It shows the artistic style, colors, texture, and lighting that should be applied. - The remaining six images are candidate results. - Candidate images are anonymized and shown in randomized order. - Each candidate is labeled only by an anonymous ID: $Candidate_1$, $Candidate_2$, $Candidate_3$, $Candidate_4$, $Candidate_5$, $Candidate_6$.

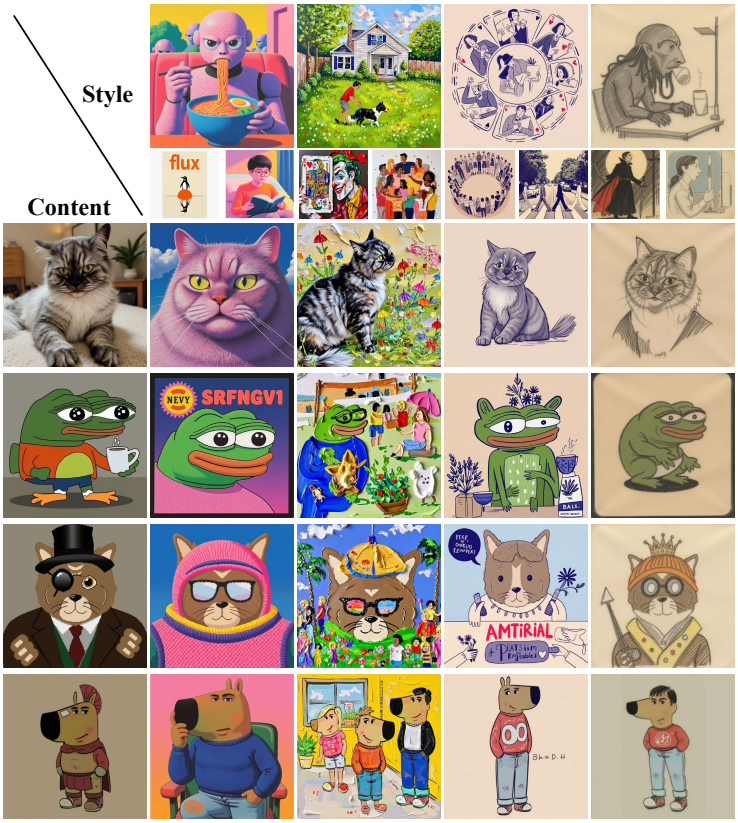


Figure 13: Qualitative results of NP-LoRA on the Flux backbone using diverse publicly available LoRAs. Each image corresponds to the combination of the conten t LoRA shown above and the style LoRA shown on the left, illustrating the results produced by our method.

```
Evaluation Criteria: 1) Subject Fidelity: retain the key
characteristics, structure, and recognizability of the subject in
A.jpg. 2) Style Consistency: apply the style, texture, colors,
and lighting of B.jpg. 3) Balance: prefer the candidate that best
preserves the subject while also matching the reference style.
Crucial Requirement: Do not choose an image that only preserves the
subject but fails to apply the style. Do not choose an image that
strongly mimics the style but distorts or loses the subject. Select
the candidate with the best overall balance between subject fidelity
and style consistency.
Output Rules: You MUST respond with ONLY ONE of the following exact
strings:
```
$Candidate_1$ $Candidate_2$ $Candidate_3$ $Candidate_4$ $Candidate_5$ $Candidate_6$
```
Do NOT output anything else. Do NOT include punctuation, brackets,
markdown, or explanation.
```

L Details for GPT-5 Evaluation Protocol

In the main paper, we employed GPT-5 to automatically evaluate and select the candidate image that best balances Subject Fidelity and Style Consistency. To reduce potential positional and label bias, we randomized the order of candidate images and anonymized all method identifiers during evaluation. The exact prompt used for GPT-5 is listed below to ensure full reproducibility.

For each evaluation case, GPT-5 received the reference subject image (`A.jpg`), the reference style image (`B.jpg`), and six anonymized candidate images labeled only as `Candidate_1`–`Candidate_6`. The mapping between anonymous candidate labels and actual methods was randomly generated

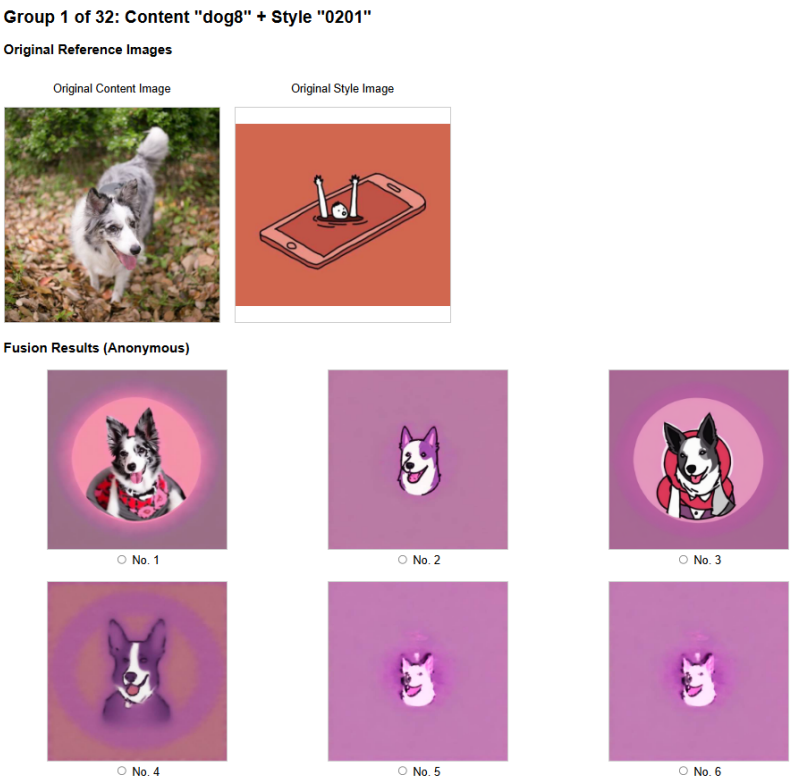


Figure 14: Questionnaire screenshot used in the human preference study. Participants selected the result that best balanced subject fidelity and style consistency based on the generated outputs and the reference subject/style images.

for each case and restored only after evaluation for result aggregation. Each case was judged independently to avoid cross-sample bias. The final GPT-5 preference score reports the proportion of cases in which each method was selected as the best-balanced result.

M Broader Impact

This work studies training-free LoRA fusion for controllable image generation, which may positively impact efficient personalization and creative content generation by enabling flexible composition of subject and style adapters without additional training.

At the same time, improved controllable generation techniques may also introduce potential misuse risks, including unauthorized identity/style composition, misleading synthetic media generation, or misuse of publicly shared LoRA adapters. Our work does not introduce a new generative backbone model or dataset, but instead focuses on improving the composition of existing LoRA adapters. We encourage responsible use consistent with the licenses and intended purposes of the underlying models and assets.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state that NP-LoRA is a training-free LoRA fusion method based on null-space projection, and the claims are supported by quantitative, qualitative, and preference-based evaluations. See Abstract, Sec. 1 and Sec. 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss limitations including the current focus on binary content–style LoRA fusion, the dependence on pretrained LoRA quality, and the need for further extension to multi-LoRA composition. See Limitations in Sec. 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides the assumptions and derivations for the hard and soft projection operators, including the closed-form solution of the regularized projection formulation. See Sec. 4 and Appendix A.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper reports the evaluated backbones, LoRA sources, baselines, prompts, metrics, and implementation details needed to reproduce the main experimental results. See Sec. 5 and Appendix B, C.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: See supplemental material.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies the evaluated datasets/LoRAs, baselines, backbones, hyperparameters, projection settings, metrics, and inference details. See Sec. 5 and Appendix B, C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: We report aggregate preference and automatic evaluation results over multiple LoRA pairs and prompts, but do not include formal confidence intervals or significance tests.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper reports the computational cost of merging and generation, and provides hardware and runtime details needed to estimate the compute required for reproduction. See Sec. 5.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research follows the NeurIPS Code of Ethics and uses existing models, LoRAs, and evaluation protocols in a responsible and anonymized manner.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper discusses positive impacts for efficient personalization and controllable generation, as well as potential misuse risks such as unauthorized style or identity composition. See Appendix M.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release a new high-risk generative model or scraped dataset; it proposes a training-free fusion method for existing LoRA adapters.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the existing diffusion backbones, LoRA adapters, datasets, and evaluation tools used in the experiments, and respects their corresponding licenses and terms of use. See Sec. 5.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A]

Justification: The paper does not introduce or release a new dataset, model checkpoint, or benchmark asset.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [Yes]

Justification: The paper includes a preference study; the evaluation protocol, instructions, and participant/task details are described in Appendix K.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The human evaluation only involves preference judgments over generated images and does not collect sensitive personal information or involve intervention with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: The paper uses GPT-5 only as an auxiliary preference evaluator, not as part of the proposed method; the evaluation protocol is described in the experimental section. See Sec. 5.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.