

时序证据双门控青花瓷风格迁移

徐子墨*, 李嘉辉

景德镇陶瓷大学信息工程学院, 江西 景德镇

收稿日期: 2026年4月27日; 录用日期: 2026年5月20日; 发布日期: 2026年5月27日

摘要

LoRA微调技术已广泛应用于稳定扩散(Stable Diffusion)文生图模型,但在青花瓷高精度图像生成中,难以同时保证纹饰风格学习与器型结构完整性,常出现形变。本文提出一种适用于Transformer架构、无需额外训练的LoRA风格迁移方法:将LoRA注入过程解耦为时间门控与提示词证据门控,并通过乘法融合动态调节风格强度——在结构敏感阶段抑制扰动,在细节生成阶段增强纹饰表现。实验证明,相比经开发集调参后的恒定强度对比,LPIPS相对下降52.1%,SSIM相对提升31.1%,MS-SSIM、PSNR与DISTS亦呈一致改善,有效实现了结构保持下的高质量风格迁移。

关键词

结构保护型风格迁移, 稳定扩散, 扩散模型, 推理时控制, 青花瓷生成

Time-Evidence Dual Gating for Blue-and-White Porcelain Style Transfer

Zimo Xu*, Jiahui Li

School of Information Engineering, Jingdezhen Ceramic University, Jingdezhen Jiangxi

Received: April 27, 2026; accepted: May 20, 2026; published: May 27, 2026

Abstract

LoRA (Low-Rank Adaptation) fine-tuning has been widely adopted in Stable Diffusion for text-to-image generation. However, when applied to high-fidelity image synthesis of blue-and-white porcelain, it often struggles to simultaneously preserve structural integrity of vessel forms and accurately capture intricate decorative styles, frequently resulting in geometric distortions. To address this challenge, we propose a novel LoRA-based style transfer method compatible with Transformer architec-

*通讯作者。

tures that requires no additional training. Our approach decouples the LoRA injection process into two dynamic gating mechanisms: a time-aware gate and a prompt-evidence gate. These gates are multiplicatively fused to adaptively modulate style intensity—suppressing stylistic perturbations during structure-sensitive denoising steps while enhancing decorative detail rendering in later stages. Experimental results demonstrate significant improvements over a baseline using constant, development-set-tuned LoRA strength: LPIPS decreases by 52.1%, SSIM increases by 31.1%, and consistent gains are observed across MS-SSIM, PSNR, and DISTS metrics. This confirms the effectiveness of our method in achieving high-quality style transfer while faithfully preserving structural fidelity.

Keywords

Structure-Preserving Style Transfer, Stable Diffusion, Diffusion Models, Inference-Time Control, Blue-and-White Porcelain Generation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

青花瓷图像生成服务于数字文博、陶瓷设计与创意生产等应用场景，其结果评价通常同时关注器型结构稳定性与纹饰风格还原度。对于此类任务，生成结果既需保持器口、腹部比例及整体轮廓等几何信息的完整性，又需呈现青花纹样的层次关系、笔触特征与装饰语言。因此，如何在结构保持与风格表达之间实现有效协调，已成为青花瓷高质量生成中的核心问题。

随着 Transformer 主干(如 Diffusion Transforme, DiT; Rectified Flow 及其 Transformer 变体)在文本到图像模型中的广泛应用[1]-[3]，基于 LoRA (Low-Rank Adaptation)的风格迁移逐渐成为稳定扩散模型轻量化适配的重要技术路径。与此同时，扩散采样过程具有显著的阶段性特征：在高噪声阶段，模型更新更侧重于低频几何与整体结构的稳定建立；随着信噪比提升，优化重心逐步转向纹理、边缘及局部细节的细化[4] [5]。围绕生成控制问题，已有方法如 ControlNet、T2I-Adapter 与 IP-Adapter [6]-[8]通过附加条件分支提供约束能力，但通常伴随额外训练与部署成本。

然而，现有 LoRA 风格迁移方法大多采用恒定强度注入策略，在青花瓷这类结构敏感任务中仍面临显著局限。当注入强度偏高时，口沿、腹部比例等关键结构更易发生漂移；当注入强度偏低时，青花纹样则容易出现表现不足、层次不清等问题。结合扩散过程的阶段性特征可知，恒定强度注入在前期可能对结构重建造成额外扰动，而在后期又难以针对纹饰细节提供充分增强。相比之下，现有附加控制方法虽具有一定效果，但并不适用于强调轻量化与即插即用的推理流程。因此，在无需额外训练的条件下实现面向扩散阶段特性的 LoRA 强度调度，仍是当前研究中的关键空白。

针对上述问题，本文提出 TE-LoRA (Time-Evidence LoRA)方法，在推理阶段对 LoRA 注入强度实施分阶段调度：以时间门控控制注入时机，以证据门控调节注入幅度，并通过二者乘积得到各去噪步的有效强度。该方法仅依赖前向 Hook 对标量进行动态更新，不修改基座参数，亦不引入额外训练过程。基于 Stable Diffusion 3.x 的青花瓷生成实验表明，与经过调参的恒定强度基线相比，TE-LoRA 在结构相关指标上表现出持续优势，同时基本保持语义一致性。由于该方法的核心建立在扩散阶段性机制而非特定题材语义之上，因此可为其他结构敏感风格迁移任务提供一种推理期强度调度思路。(图 1)

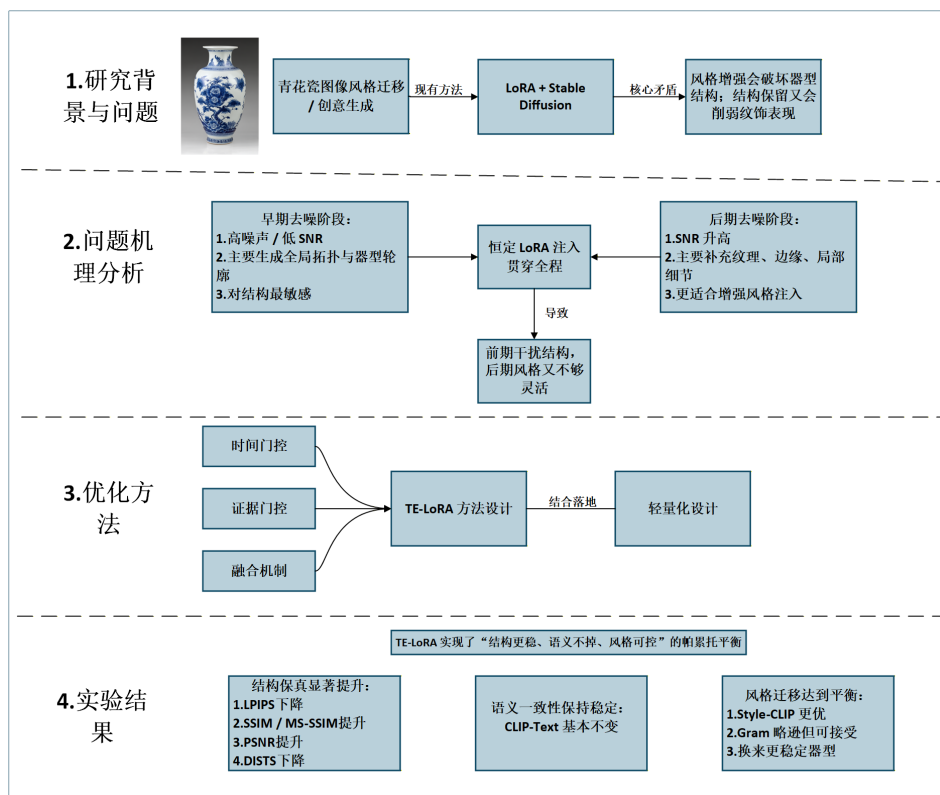


Figure 1. Overall flowchart

图 1. 总体流程图

2. 方法

2.1. 问题动机

现有基于 LoRA 的风格迁移方法通常采用恒定强度注入, 这在处理结构敏感型内容(如青花瓷器型)时, 容易引发“保结构”与“强风格”之间的根本性冲突。本文认为, 这一冲突源于恒定注入策略与扩散模型内在的去噪阶段性不匹配。

扩散模型的去噪过程并非均质。在信噪比很低的早期去噪阶段, 模型的主要任务是依据文本提示和先验知识, 从噪声中推断出图像的全局拓扑结构和主体轮廓[4]。此时, 潜变量的更新方向对最终生成结果的几何形态(如器型的口沿、腹部比例、外轮廓)具有决定性影响。随着 Transformer 主干在扩散模型中的普及, 其全局注意力机制使得特征扰动更容易传播。若在此结构敏感阶段高强度注入 LoRA 风格信息, 风格相关的注意力偏差会系统性干扰跨空间位置的几何关系建模, 导致轮廓漂移或比例失调。

随着去噪进行, 信噪比升高, 图像的主体结构已基本确定。后续阶段的更新重心逐渐转向填充和细化纹理、边缘与局部细节[4]。此时, 恰是增强风格注入以丰富纹样、色彩和质感细节的理想时机。恒定强度策略在此阶段可能无法提供足够的风格表达力, 导致纹样“发虚”或层次不足。

基于上述洞察, 提出核心假设: LoRA 风格注入的强度不应是恒定的, 而应是一个与去噪阶段(时间)和生成内容(提示词证据)动态适配的函数。通过抑制早期注入以保护结构, 并在后期释放强度以增强风格, 有望在无需重新训练模型的情况下, 实现更好的“结构 - 风格”平衡。为此, 本文提出时序 - 证据双门控机制, 其核心是通过时间门控实现刚性结构保护, 并通过证据门控实现柔性风格调制。TE-LoRA 正是基于这一机制而设计, 见图 2。

2.2. 双门控调度器

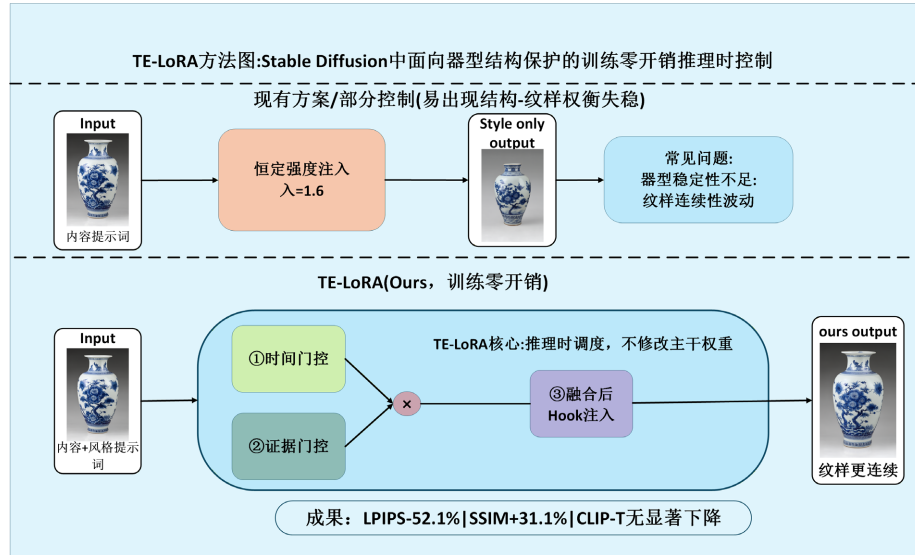


Figure 2. TE-LoRA method framework diagram
图 2. TE-LoRA 方法框架图

为了将上述动机转化为可操作的算法, 设计了 TE-LoRA (Time-Evidence LoRA)。其核心思想是将每一步的有效注入强度 λ_{eff}^t 解耦为两个可解释的门控因子: 时间门控 $g_T(t)$ 和证据门控 $g_E(t)$, 并通过乘法进行融合:

$$\lambda_{\text{eff}}^t = \lambda_{\text{cap}} \cdot g_T(t) \cdot g_E(t) \quad (1)$$

其中, λ_{cap} 是一个用户定义的强度上限, 用于控制风格影响的全局尺度。

2.2.1. 时间门控机制

时间门控是 TE-LoRA 实现分阶段调度的核心, 直接源于 3.1 节所述的扩散阶段性理论。在此采用一种“暖启动”策略来刚性地区分结构形成期与细节生成期:

$$g_T(t) = \begin{cases} 0, & \text{if } t < t_{\text{warmup}} \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

其中, $g_T(t)$ 表示第 t 个去噪步的时间门控输出, t 表示当前去噪步, t_{warmup} 表示结构保护窗口。在去噪步数 $t < t_{\text{warmup}}$ 时, 时间门控输出为 0, 完全切断 LoRA 风格的注入。这确保了模型在结构形成的关键期, 其自注意力机制不受任何外部风格偏置的干扰, 从而刚性保障几何轮廓的稳定收敛。一旦步数超过 t_{warmup} , 门控完全打开(输出为 1), 允许最大程度的风格注入以丰富细节。本文主实验采用阶跃函数形式, 以验证刚性保护的有效性。

2.2.2. 证据门控机制

时间门控提供了基础的时序调度。为进一步适应多样化的文本提示, 在此引入了证据门控, 在时间门控开启的窗口内, 进行基于提示词语义的细粒度强度调制。其计算基于一个预定义的主题词库 V_{content} 和风格词库 V_{style} 。词库构建遵循三项原则: 一是任务相关, 仅保留青花瓷器型与纹样生成中稳定出现且语义明确的词项; 二是中英文对应, 将同义词并列收录以降低语言形式差异带来的偏置; 三是低歧义可复现, 剔除多义或弱指示词并固定词表版本。以本文实验脚本的默认词库为例, V_{content} 共 17 项, 代

表词包括 vase、bottle、jar、花瓶、器形、口沿、腹、肩; V_{style} 共 15 项, 代表词包括 blue and white porcelain、qinghua、cobalt blue、glaze、青花瓷、青白釉、钴蓝、釉面、缠枝纹。需要说明的是, 本文实际生成提示词模板为 blue and white porcelain, 青花瓷, vase, [纹样词], 其中纹样词来自 11 类集合(如缠枝纹、海水纹、花卉纹、龙纹等); 词库中的其他词项用于证据估计的语义扩展, 不要求在每条提示词中同时出现。该构建方式可在提示词措辞变化或中英文混写时降低证据估计波动; 具体而言, 当提示词发生同义替换、语言切换或词序变化时, 风格证据比的命中统计仍保持在可控范围内, 从而减少证据门控的误触发并提升方法鲁棒性。通过词条匹配统计证据比 r_e :

$$r_e = \frac{r_{\text{style}}}{r_{\text{style}} + r_{\text{subj}}} \quad (3)$$

原始证据因子定义为 $g_E^{raw} = \text{clip}(r_e, 0.1, 0.9)^{\max(1, \alpha)}$ 其中 α 为证据增益系数, 用于调节模型对风格证据的响应灵敏度, r_e 表示风格证据比, r_{style} 表示提示词中与风格词库匹配的词条数, r_{subj} 表示提示词中与主题词库匹配的词条数。为提升鲁棒性引入双重约束: (1) 设置冻结步数 f , 在 $t < f$ 时强制 $g_E^{raw} = 1$, 确保在去噪最早期完全由时间门控主导, 避免初始噪声阶段的统计波动; (2) $t \geq f$ 时引入基于信息熵的正则化, 信息熵公式为:

$$H(r_e) = -r_e \log(r_e) - (1 - r_e) \log(1 - r_e) \quad (4)$$

当提示词中风格与主题词数量相近(熵值高, 指示语义模糊)时, 对证据因子进行约束, 防止因输入不确定性导致生成不稳定, 当 $H(r_e)$ 较高(即 r_e 接近 0.5, 语义更模糊)时, 触发证据回拉, 令 $g_E^{raw} = 0.5g_E^{raw} + 0.25$, 从而下调 $g_E(t)$; 否则保持 $g_E(t)$ 不变。最终经过平滑处理的证据因子记为 $g_E(t)$, 用于第 t 步的证据因子。

2.2.3. 融合机制与参数语义

有效强度 λ'_{eff} 由公式 ($\lambda'_{\text{eff}} = \lambda_{\text{cap}} \cdot g_T(t) \cdot g_E(t)$) 计算得出, 其中, λ'_{eff} 表示第 t 个去噪步的有效注入强度, λ_{cap} 表示全局强度上限, $g_T(t)$ 表示时间门控函数, $g_E(t)$ 表示证据门控函数, t 表示当前去噪步。乘法融合定义了一个层次化且条件依赖的控制逻辑。在此框架中, 时间门控 $g_T(t)$ 充当一级条件, 控制风格注入的全局启用时机。当注入被允许后, 证据门控 $g_E(t)$ 则作为二级调制器, 依据提示词语义对其进行细粒度调节。标量 λ_{cap} 作为全局系数, 定义了风格化影响的理论上限。这种级联的乘法关系确保了控制信号传递路径的清晰性, 从而提升了方法的解析性。

该控制律中的每个参数均具有明确的物理或功能语义, 其中, w 为时间门控结构保护步数, f 为证据冻结步数, α 为证据增益系数, $\text{entropy}_{\text{thresh}}$ 为熵门控阈值。共同服务于“在结构形成期抑制扰动, 在细节合成期增强风格”的核心目标。 λ_{cap} 作为乘法链的公共系数, 是平衡风格强度与内容保真度的首要参数, 直接决定了风格影响力的最大强度。关键参数 t_{warmup} 定义了结构保护时窗, 即去噪过程中应完全屏蔽风格干扰的初始步数范围。这一机制是保障主体几何结构稳定收敛的关键。证据增益系数 α 则线性调制模型对文本风格描述的响应。此外, 为确保系统在初始高噪声阶段及面对语义模糊提示词时的鲁棒性, 引入了两项辅助性约束: 冻结步数 f 在去噪最前期暂时固定证据门控输出, 以抑制初始统计波动引入的扰动; 阈值 $\text{entropy}_{\text{thresh}}$ 在提示词语义存在歧义时对证据因子进行约束, 防止因输入不确定性引发生成过程的不稳定。通过上述设计, TE-LoRA 将复杂的“结构-风格”权衡问题, 形式化表达为一个由语义清晰的参数构成的动态调度系统, 从而兼具良好的可控性与可调性。

2.3. 前向 Hook 实现

为实现上述动态调度并保持“训练零开销”的优势, 采用前向 Hook 机制在推理时动态调制 LoRA

分支的权重。对于 Transformer 主干[1]中任一注入 LoRA 的线性层(如 q, k, v, o 投影), 其前向传播过程被修改为:

$$y = W_{\text{base}}x + \lambda_{\text{eff}}^t \cdot (W_{\text{loa}}x) \quad (5)$$

其中 y 表示该线性层的输出特征, W_{base} 是原始预训练权重, W_{loa} 是训练好的 LoRA 适配器权重。在每一步去噪开始前, 调度器根据当前步数 t 和提示词计算出 λ_{eff}^t , 并将其作为一个全局标量传递给所有被 Hook 的模块。

这种基于前向 Hook 的实现, 使得 TE-LoRA 成为一个完全运行于推理期的独立模块。它无需修改或访问模型内部权重 ($W_{\text{base}}, W_{\text{loa}}$), 仅通过一个标量实现全局强度调度, 因此计算与显存开销极低。该方法与任何基于 Transformer 架构并支持 LoRA 注入的预训练扩散模型兼容, 实现了即插即用的结构保护型风格迁移。在我们的实现中, 该 Hook 被统一应用于模型内关键的线性投影层, 确保风格调制能全面生效。

3. 实验

3.1. 实验设置

任务与提示词。使用单一青花瓷 LoRA, 测试集覆盖 11 类典型纹样(如缠枝、海水、龙纹等)。提示词模板为 blue and white porcelain, 青花瓷, vase, [纹样词]。实验在 10 个随机种子上进行配对测试, 形成 $11 \times 10 = 110$ 个配对评估单元(用于表 1~3 的统计检验)。按三路结果图(Base/Style/Ours)计, 共 330 张图像($base = 110, style = 110, ours = 110$)。本文使用的青花瓷 LoRA 权重为自训练版本; 文中“纹样词/类别名”主要用于工程化提示与分组评测, 不严格等同于陶瓷史或考古学语境下的标准学术命名。

对比方法与参数选择。以“恒定强度注入(Style-only)”为强基线, 采用统一自动调参流水线, 在开发集进行公平选参。Style-only 仅搜索恒定强度 λ , 网格为 $\{0.8, 1.0, 1.2, 1.4, 1.6\}$ (5 组); TE-LoRA 搜索网格为 $\lambda_{\text{cap}} \in \{1.0, 1.2, 1.4\}$ 、 $w \in \{2, 3, 4\}$ 、 $\alpha \in \{1.5, 2.0, 2.5\}$ 、 $entropy_thresh \in \{1.5, 2.0\}$, 并固定 $f = 3$, 共 $3 \times 3 \times 3 \times 2 \times 1 = 54$ 组。开发集采用 3 轮随机子集评估: 每轮从 seed 池[17, 33, 42, 54, 69]抽取 3 个 seed, 并使用 4 个纹样提示词, 因此每参数组合每轮评估 12 个单元。总评估量为 $(5 + 54) \times 12 \times 3 = 2124$ 个评估单元。评分采用 Q^* ($w_{\text{ssim}} = 0.4, w_{(1-lpips)} = 0.4, w_{\text{clip}} = 0.2, \gamma = 0.5, \delta = 1e-3$), 先在全部参数组合上做 min-max 归一化, 再跨轮取均值选优; 最终 Style-only 选得 $\lambda = 1.6$, TE-LoRA 选得 $\lambda_{\text{cap}} = 1.0$ 、 $w = 2$ 、 $\alpha = 1.5$ 、 $entropy_thresh = 1.5$ 、 $f = 3$ 。该流程采用受限且可解释的参数网格, 用于检验机制在不同配置下的稳定性。结果显示, 改进方向在全量 110 个配对单元上保持一致, 表明收益来源于时序-证据耦合控制本身。

评测指标体系:

(1) 结构保真组: 以未注入 LoRA 的输出为几何参照。-SSIM (Structural Similarity Index Measure) [9]/MS-SSIM (Multi-Scale Structural Similarity) [10]: 分别从单尺度与多尺度衡量轮廓、边缘及层级结构一致性, 对应器口、外轮廓连续性是否被破坏。-PSNR (Peak Signal-to-Noise Ratio) [11]: 量化像素层面的保真度, 用于验证结构修复是否伴随整体失真回落(作为辅助证据而非单一决策依据)。

(2) 感知质量组: -LPIPS (Learned Perceptual Image Patch Similarity) [12]: 衡量感知域差异, 反映“看起来是否仍是同一器型”的整体漂移风险。-DISTS (Deep Image Structure and Texture Similarity) [13]: 联合结构与纹理统计, 反映纹样细节改善是否以结构扭曲为代价。

(3) 语义与风格组: -CLIP-Text (Contrastive Language-Image Pre-training text similarity) [14]: 作为语义护栏, 检验提示词语义是否被时序调度破坏。-Style-CLIP (基于 CLIP 的风格相似度指标)/Gram Dis-

tance: 分别衡量语义层风格贴合度与纹理统计贴合度。在结构保护型任务中, 本文以“结构组 + 感知组”作为主判据, 以“语义与风格组”作为约束与权衡判据: 若结构显著改善且 CLIP-Text 无显著下降, 则允许出现小幅纹理统计差异。

3.2. 定量结果

表 1 显示, TE-LoRA 在结构与感知指标上取得了显著优势: LPIPS 相对基线下降 52.1% (0.2593→0.1243), SSIM 提升 31.1% (0.6394→0.8385), 同时 MS-SSIM、PSNR、DISTS 同向改善。就任务含义而言, LPIPS/DISTS 的同步下降意味着“器型感知漂移 + 纹理失真”同时减弱, SSIM/MS-SSIM/PSNR 的同步上升意味着“全局轮廓 + 多尺度局部结构”更稳定。CLIP-Text 得分保持稳定(0.3668 vs 0.3661), 说明在保持风格迁移有效性的同时, 语义护栏未被破坏。

Table 1. Main results (mean \pm standard deviation, N = 110)

表 1. 主结果(均值 \pm 标准差, N = 110)

方法	LPIPS \downarrow	SSIM \uparrow	MS-SSIM \uparrow	PSNR \uparrow	DISTS \downarrow	CLIP-T \uparrow
恒定强度注入	0.2593 \pm 0.104	0.6394 \pm 0.208	0.7854 \pm 0.076	16.24 \pm 2.12	0.1579 \pm 0.034	0.3668 \pm 0.00
TE-LoRA	0.1243 \pm 0.074	0.8385 \pm 0.096	0.8882 \pm 0.068	20.53 \pm 2.97	0.0931 \pm 0.033	0.3661 \pm 0.008

图 3 展示整体性能分布对比(测试集 N = 110)。TE-LoRA 在保持风格迁移有效性并维持语义一致性的同时, 大幅压缩了 LPIPS 分布区间, 呈现出更低且收敛的 LPIPS 分布, 表明其对长尾样本的鲁棒性。

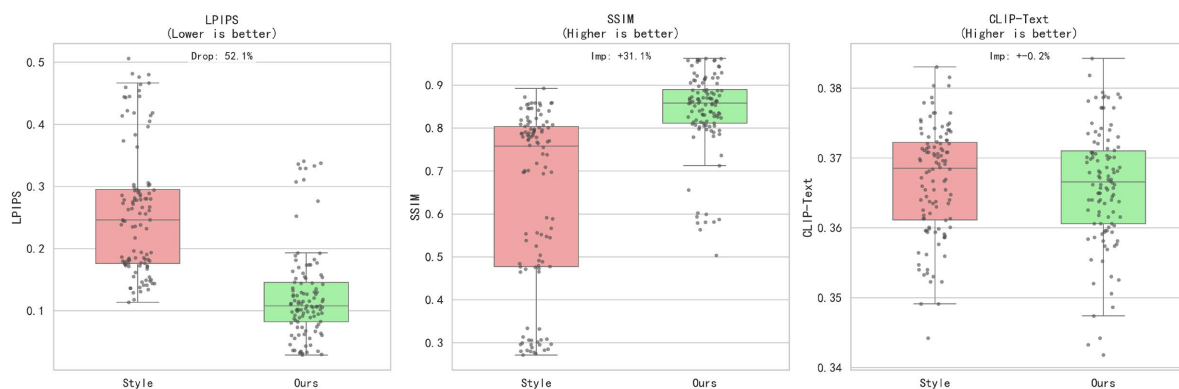


Figure 3. Boxplot comparison of performance distribution on the blue-and-white porcelain test set

图 3. 箱线图整体分布对比图

表 2 进一步展示了 TE-LoRA 在最差样本(Min/Q25)上的改善, 证明了其对结构崩坏风险(长尾失真)的有效抑制; 这也是工业批量生成中最关键的可用率增益来源

Table 2. Quantile distribution of structural fidelity index (N = 110)

表 2. 结构保真指标的分位数分布(N = 110)

方法	LPIPS \downarrow (min/Q25/med/Q75/max)	SSIM \uparrow (min/Q25/med/Q75/max)
恒定强度注入($\lambda = 1.6$)	0.1340/0.2460/0.2787/0.2960/0.4804	0.2747/0.3007/0.4785/0.8072/0.8587
TE-LoRA	0.0287/0.0811/0.1097/0.1810/0.3362	0.5038/0.7687/0.8105/0.8423/0.9588

Table 3. Style fidelity evaluation (Style Fidelity, Top-k = 3, N = 110)
表 3. 风格保真度评测 (Style Fidelity, Top-k = 3, N = 110)

对比维度	指标	Base 均值	Ours 均值	差异 Δ	P-value	结论
Ours vs Base	Style-CLIP \uparrow	0.9491	0.9505	+0.0015	0.0326	风格显著迁徙
	Gram Dist \downarrow	2.05e-5	1.91e-5	-0.14e-5	<0.001	纹理显著接近
Ours vs Style	Style-CLIP \uparrow	0.9464	0.9505	+0.0042	<0.001	语义风格更优
	Gram Dist \downarrow	1.50e-5	1.91e-5	+0.41e-5	<0.001	纹理统计略差

结果表明 TE-LoRA 在语义层风格相似度 (Style-CLIP) 上优于恒定强度, 而在纹理统计 (Gram) 上略逊。该现象可解释为: TE-LoRA 通过前期抑制强注入减少结构扰动, 部分高频纹理统计不再被“过量放大”, 从而换取更稳定的器型骨架。结合结构组指标的 (LPIPS、SSIM/MS-SSIM、PSNR、DISTS) 一致显著改善, TE-LoRA 实现了“语义风格更优、纹理略逊、结构更稳”的可解释帕累托平衡。TE-LoRA 仅需前向 Hook 标量更新, 无额外显存占用, 时延开销可忽略, 具备较高工程集成度。

3.3. 定性结果与补充图像

图 4 展示了不同注入策略在器型几何与纹样组织上的可视差异。总体上, 恒定强度注入能够提升风格显著性, 但在结构敏感阶段更容易将高频纹理扰动提前写入几何骨架, 表现为口沿厚薄不均、肩腹过渡变形、瓶体中轴稳定性下降等现象。相较之下, TE-LoRA 在前期抑制注入、后期释放强度的控制下, 器口 - 器肩 - 器腹的轮廓关系更接近 Base 参考, 外轮廓连续性与轴向对称性更稳定。

从纹样层面观察, 恒定强度组在高曲率区域更易出现笔触堆叠、纹带粘连或局部断裂; TE-LoRA 则在保持蓝白釉面风格强度的同时, 使纹样边界与器型曲面贴合更自然, 连续纹带的走向更一致, 局部细节的层次关系更清晰。该现象说明证据门控并未削弱风格迁移目标, 而是将风格注入约束在更合适的时序区间内完成。

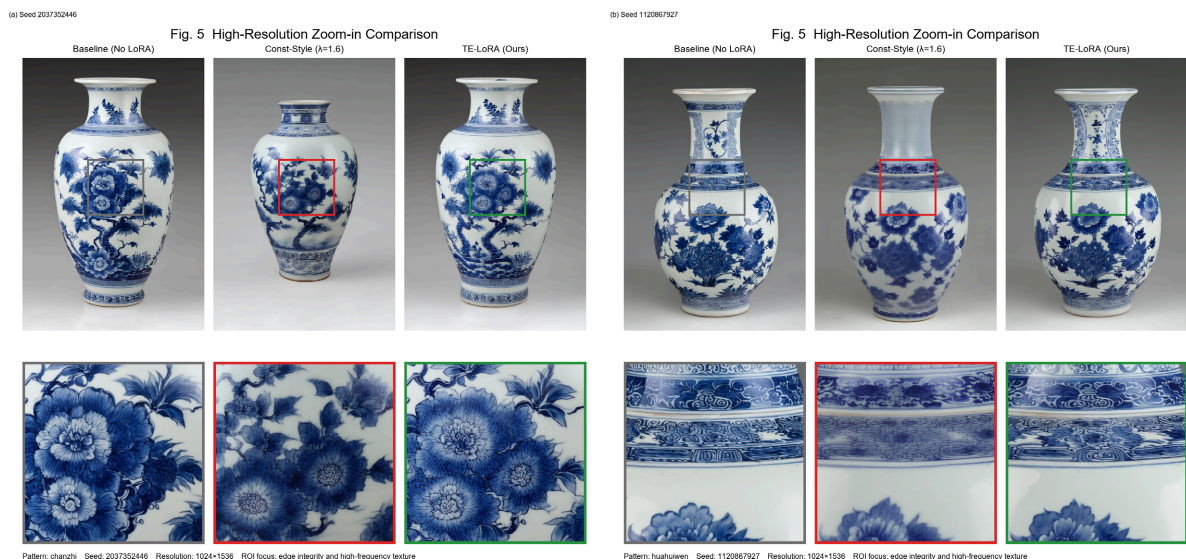


Figure 4. Comparison of local details

图 4. 局部细节对比图

图 4 的局部放大进一步显示了两种策略在“结构边界 - 纹样细节”耦合处的差异: 恒定强度组更易

在边界附近产生伪边与涂抹感, 而 TE-LoRA 的边缘过渡更干净、细节组织更连贯。该定性趋势与定量结果一致, 即结构相关指标(LPIPS、SSIM/MS-SSIM、PSNR、DISTS)同步改善而 CLIP-Text 基本稳定, 支持“结构优先且风格可控”的机制结论。

4. 讨论

4.1. 结果分析

总体结果显示, 在语义一致性保持条件下, TE-LoRA 相较恒定强度 LoRA 实现了更优的风格迁徙与结构保真。配对统计中, LPIPS、DISTS 下降, SSIM/MS-SSIM、PSNR 上升, CLIP-Text 基本稳定, 说明增益主要来自结构与感知质量改善。

可以看到, 当应用于以 Transformer 为主干的 Stable Diffusion 模型时, 它比普通 LoRA 更加适配其结构。TE-LoRA 的时序与证据双门控机制是与 Transformer 主干注意力通路和扩散模型动力学相匹配的机制性增益。同时观察到总体改进、尾部收缩、剂量边界可控三类信号, 表明该机制能够有效抑制 $q/k/v/o$ 通路中的早期过量注入扰动。结构与感知指标在 110 个配对样本中稳定胜出, 最差分位(Min/Q25)明显改善, 且当 λ_{cap} 从 1.0 提高至 1.4 时 LPIPS 由 0.205 上升至 0.356。

4.2. 局限与未来工作

在证明了 TE-LoRA 在青花瓷生成场景上的有效性后, 更希望进一步探索多场景多模型中更具挑战性的问题, 例如将 TE-LoRA 应用于山水画场景生成或者与其它 LoRA 方法的融合技术, 由于实际生产需求的增加, 风格迁徙在这里更为重要, 并且还理论认为, 如果没有类似的流程, 要达到同等性能需要更多算力和数据训练, 这同样是由于结构与风格难以平衡的问题。

4.3. 结论

在本文中, 提出了一种新的、面向结构保护的推理时风格迁移流程 TE-LoRA, 并将其应用于 Stable Diffusion 的青花瓷创意生成任务。证明了该方法在“结构保真 - 风格表达”两方面的有效性: 相较恒定强度注入基线, TE-LoRA 在 LPIPS、SSIM、MS-SSIM、PSNR 与 DISTS 等指标上取得一致且显著的改进。总体而言, 实验结果验证了该流程在性能上优于传统恒定注入策略。报告了其鲁棒性提升, 体现在不同纹样类别与随机种子下的稳定收益; 以及其灵活性提升, 体现在无需额外训练、仅通过单一推理期调度机制即可完成可控迁移。总体上, 这表明像 TE-LoRA 这样的时序与证据联合调度方法, 在结构敏感型风格迁移任务中具有更高实用价值, 并有望推广到更广泛的生成式内容创作场景。

基金项目

本研究得到江西省高等学校省级大学生创新创业训练计划项目“‘青花焕新’——AI 驱动的扩散模型青花瓷图案创新生成与个性化设计”(No.S202510408003)资助。

参考文献

- [1] Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., *et al.* (2024) Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. arXiv: 2403.03206. <https://arxiv.org/abs/2403.03206>
- [2] Peebles, W. and Xie, S. (2023) Scalable Diffusion Models with Transformers. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 4172-4181. <https://doi.org/10.1109/iccv51070.2023.00387>
- [3] Liu, X., Gong, C. and Liu, Q. (2023) Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. *International Conference on Learning Representations (ICLR) 2023*, Kigali, 1-5 May 2023, 1-33.
- [4] Lin, S., Liu, B., Li, J. and Yang, X. (2024) Common Diffusion Noise Schedules and Sample Steps Are Flawed. 2024

-
- IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2024, 5392-5399. <https://doi.org/10.1109/wacv57701.2024.00532>
- [5] Karras, T., Aittala, M., Aila, T. and Laine, S. (2022) Elucidating the Design Space of Diffusion-Based Generative Models. arXiv: 2206.00364.
- [6] Zhang, L., Rao, A. and Agrawala, M. (2023) Adding Conditional Control to Text-To-Image Diffusion Models. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3813-3824. <https://doi.org/10.1109/iccv51070.2023.00355>
- [7] Mou, C., Wang, X., Xie, L., Wu, Y., Zhang, J., Qi, Z., *et al.* (2024) T2I-Adapter: Learning Adapters to Dig Out More Controllable Ability for Text-To-Image Diffusion Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, **38**, 4296-4304. <https://doi.org/10.1609/aaai.v38i5.28226>
- [8] Ye, H., Zhang, J., Liu, S., Han, X. and Yang, W. (2023) IP-Adapter: Text Compatible Image Prompt Adapter for Text-To-Image Diffusion Models. arXiv: 2308.06721.
- [9] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/tip.2003.819861>
- [10] Wang, Z., Simoncelli, E.P. and Bovik, A.C. (2003) Multiscale Structural Similarity for Image Quality Assessment. *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, Pacific Grove, 9-12 November 2003, 1398-1402. <https://doi.org/10.1109/acssc.2003.1292216>
- [11] Zhou Wang, and Bovik, A.C. (2009) Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, **26**, 98-117. <https://doi.org/10.1109/msp.2008.930649>
- [12] Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O. (2018) The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 586-595. <https://doi.org/10.1109/cvpr.2018.00068>
- [13] Ding, K., Ma, K., Wang, S. and Simoncelli, E.P. (2020) Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 2567-2581. <https://doi.org/10.1109/tpami.2020.3045810>
- [14] Radford, A., Kim, J. W., Hallacy, C., *et al.* (2021) Learning Transferable Visual Models from Natural Language Supervision. *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 18-24 July 2021, 8748-8763.