arXiv:2603.01103v1 [cs.CV] 1 Mar 2026

用于油画的扩散模型数据高效笔触生成

DANTONG QIN, 代尔夫特理工大学，荷兰ALESSANDRO BOZZON, 代尔夫特理工大学，荷兰XIAN YANG, 曼彻斯特大学，英国XUN ZHANG, 代尔夫特理工大学，荷兰YIKE GUO, 香港科技大学，香港特别行政区PAN WANG*, 代尔夫特理工大学，荷兰

许多创意多媒体系统建立在笔触或纹理等视觉基元之上，这些基元难以大规模采集，且与自然图像数据存在根本差异。这种数据稀缺使现代生成模型难以学习富有表现力且可控的基元，从而限制了其在过程感知内容创作中的应用。在本工作中，我们研究如何从少量手绘样本（$n = 470$）中学习类人笔触生成，并提出StrokeDiff，一种带有平滑正则化（SmR）的基于扩散的框架。SmR在训练过程中注入随机视觉先验，提供一种简单机制，可在稀疏监督下稳定扩散模型，且不改变推理过程。我们进一步展示如何通过基于贝塞尔曲线的条件模块使所学基元变得可控，并将其整合到完整的基于笔触的绘画流程中，包括预测、生成、排序和合成。这表明数据高效的基元建模能够支持富有表现力且结构化的多媒体内容创作。实验表明，所提方法能够生成多样且结构连贯的笔触，并实现纹理与分层更为丰富的绘画，这一点已由自动化指标和人工评估共同验证。

CCS 概念：• **应用计算** → **美术**；• **计算方法** → **计算机视觉**。

附加关键词与短语：笔触合成、基于笔触的渲染、扩散模型、艺术家绘画、少样本

ACM 引用格式：

DANTONG QIN、ALESSANDRO BOZZON、XIAN YANG、XUN ZHANG、YIKE GUO 和 PAN WANG。2018年。用于油画的扩散模型数据高效笔触生成。*ACM* 测量、分析与计算系统会议录37，4，第111号文章（2018年8月），23页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

绘画是一种分层、迭代的过程，艺术家通过连续的笔触构建视觉场景，每一笔都承载着空间、纹理和风格意图[3, 28, 55]。在创意多媒体系统中，基于笔触的渲染（SBR）是一种代表性范式，旨在通过将图像合成为一系列

*本文的通讯作者。

作者联系方式：Dantong Qin，D.Q.Dantong@tudelft.nl，代尔夫特理工大学，代尔夫特，荷兰；Alessandro Bozzon，A.Bozzon@tudelft.nl，代尔夫特理工大学，代尔夫特，荷兰；Xian Yang，xian.yang@manchester.ac.uk，曼彻斯特大学，曼彻斯特，英国；Xun Zhang，x.zhang-16@tudelft.nl，代尔夫特理工大学，代尔夫特，荷兰；Yike Guo，yikeguo@ ust.hk，香港科技大学，九龙，香港特别行政区；Pan Wang，P.Wang-2@tudelft.nl，代尔夫特理工大学，代尔夫特，荷兰。

特此授权免费制作本作品全部或部分内容的数字或纸质副本，供个人或课堂使用，但不得出于营利或商业目的制作或分发副本，且副本首页须保留本声明和完整引文。对于本作品中由作者以外的其他方拥有版权的组件，必须尊重其版权。注明出处后可以摘录。若要以其他方式复制、转载、发布到服务器或分发给他人，必须事先获得具体许可和/或支付费用。请联系permissions@acm.org 申请许可。©2018年版权归所有者/作者所有。出版权授权给 ACM。ACM 2476-1249/2018/8-ART111https://doi.org/XXXXXXX.XXXXXXX

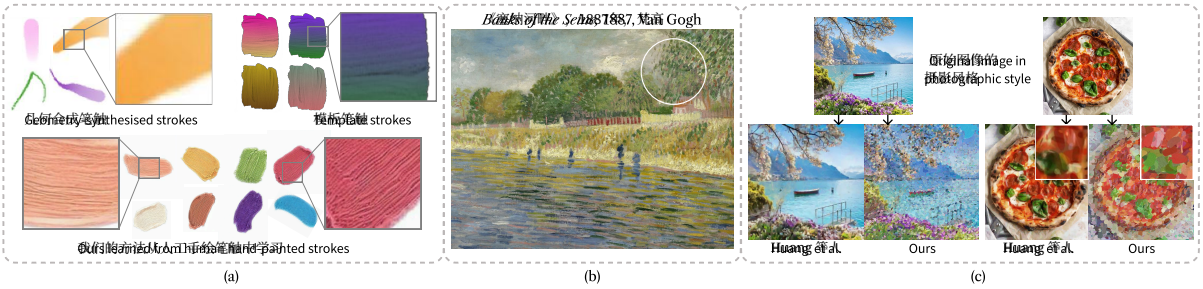


图 1. 笔触质量与绘画效果对比。(a) 以往的基于笔触的渲染方法采用几何笔触或模板笔触，表现力有限。我们的笔触从人工绘制的数据中学习而来，呈现出更丰富的结构与纹理。(b) 目标油画风格，展现出超越写实主义的抽象性。(c) 与 Huang 等人相比 [22], 我们的方法带来了更强的艺术领域迁移。

离散笔触。与照片级真实感合成不同，基于笔触的渲染并不旨在复现像素，而是通过绘画式表征传达语义内容。近期方法在生成与内容对齐的渲染结果方面取得了进展，但大多数方法依赖预定义的笔触模板 [19, 33, 42, 48, 50] 或几何合成基元[22, 36, 45, 53]。尽管这些方法具有高效和结构一致性等优点，但它们往往难以捕捉人类笔触的多样性、不规则性和纹理，而这些特征对于风格化抽象至关重要。因此，在数据驱动绘画流程中，类人笔触的表现空间仍未得到充分探索（对比见图1 ）。

为弥补这一空白，我们研究了如何利用扩散模型（DM）生成类人笔触这一问题。与大规模数据集广泛可用的自然图像领域相比，真实笔触数据仍然稀缺，在现有文献中也未得到充分探索。与基于 GAN 的笔触合成等早期方法相比， [25, 36, 53], 扩散模型在使用大量数据训练时表现出色，但其泛化到设计时基元的能力尚未得到充分了解；直接使用数百个样本对预训练模型进行微调，往往会导致结构退化。这凸显了在数据有限的情况下稳定扩散训练，同时保持艺术笔触的抽象性与风格特质所面临的挑战。

以往应对低资源扩散问题的工作，要么修改扩散路径（例如噪声调度[5, 9, 13, 30, 32, 58], 分类器（或无分类器）引导 [6, 7, 12]），要么注入提示词[8, 37] 或示例[16, 26, 40, 56]等外部条件。尽管这些策略在某些领域有效，但它们往往需要大规模调优，或引入测试时依赖，从而限制了可扩展性。在这项工作中，我们提出了 **StrokeDiff**，这是一种基于扩散的框架，旨在数据极其有限的情况下学习富有表现力的视觉基元。其核心是**平滑正则化（SmR）**，这是一种训练时策略，通过向前向扩散过程中注入随机视觉先验，弥补结构监督的不足。在训练期间注入微弱但多样的视觉线索，SmR 能够促使模型在低信号区间保留全局结构和语义连贯性。重要的是，SmR 仅在训练时发挥作用，不会修改推理流程，也不需要额外条件。因此，它是一种轻量且通用的正则化器，可在稀疏数据下稳定扩散模型，且不会引入测试时依赖或计算开销。

除了单个笔触之外，为了证明所学习的基元能够支持结构化内容创作，我们进一步引入了一个贝塞尔参数化条件模块，以实现可控笔触合成并促进其集成到绘画流程中。这一设计能够显式控制形状和位置等笔画属性，并支持将其排列成完整图像。为进一步改善构图，我们加入了排序损失来规范笔画顺序，从而减少重叠伪影并增强风格分层。

我们的实验表明，StrokeDiff 能够从 470 个手绘样本中生成高保真且多样化的笔触；在数据稀疏的情况下，其泛化能力优于噪声调度和基于 LoRA 的适配。在完整画作层面，它生成的结果更连贯，也更具风格表现力；我们通过 CLIP 对齐等语义指标以及人工偏好研究进行评估，后者考察风格、纹理和整体美感质量。我们的主要贡献如下：

- 平滑正则化（SmR）：一种基于随机视觉先验注入的训练时正则化策略，可在稀疏数据下稳定扩散训练，而不改变推理流程。
- 可控笔触合成与流程集成：我们设计了一个贝塞尔参数化条件模块，用于可控笔触生成，并将其接入完整的绘画流程，从而生成更能体现油画分层效果与纹理特质的笔触渲染结果。
- 评估：我们在笔触和绘画两个层面进行多维度评估，使用 CRD、FID、LPIPS 和人工研究，对基于模板、矢量、生成对抗网络及扩散模型的方法进行比较。

2 相关工作

2.1 基于笔触的渲染中的笔触

尽管生成模型在图像风格化方面也取得了进展 [21, 38, 54], ，基于笔触的渲染通过依次绘制笔触来构建图像，模拟人类艺术家层层绘制的方式 [22, 23, 34,44, 45]。现有方法大致可分为基于模板的方法、神经笔触生成方法和基于矢量图形的方法。基于模板的方法 [18, 19, 33, 42, 49, 50] 通过调整参数（如颜色、长度和角度），改变有限数量的笔触模板。尽管这类方法能保留绘画纹理的外观，但由于依赖固定模板，往往会产生重复的笔触模式。神经笔触生成方法 [36, 53, 61] 旨在通过从噪声中合成笔触来提高笔触多样性。然而，其中许多模型是在合成数据集（如 MyPaint[1] 或 FluidPaint[2]）上训练的，因此生成的结果过于平滑且具有几何感。尽管 Bidgoli 等人 [1]在真实笔触上训练了变分自编码器，但其方法并未整合到完整的绘画流程中。另一类方法采用基于矢量图形的方法[4, 17, 22, 29] ，将笔触拟合为参数曲线，例如贝塞尔样条。尽管这类方法具有很强的结构保真度，却往往牺牲了自由手绘所特有的随性和自然变化。尽管基于笔触的渲染取得了进展，实现多样、逼真且富有表现力的笔触仍是一项核心挑战。

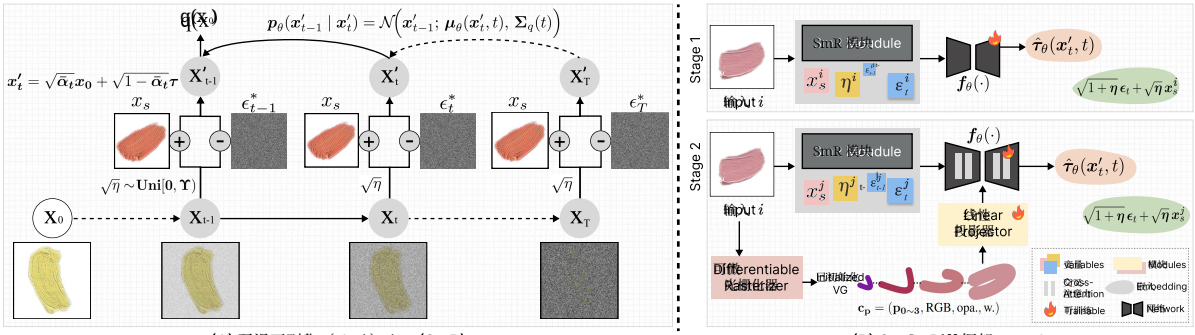
2.2 低数据量扩散方法

近期研究通过引入外部先验或提示词来引导生成，使扩散模型适用于低数据量场景，其中包括基于示例的个性化方法，如 DreamBooth [40] 和 Custom Diffusion [26], ，以及基于反演的方法 [35],和注意力操控技术 [10, 46]。这些方法在实现高保真度或可控生成方面效果显著，但通常依赖测试时输入或提示工程，而且往往局限于特定实例的合成。另一类研究则着重于在训练受限的情况下提升样本多样性，方法是通过噪声调度 [5, 9, 13, 30]、后验引导 [12],或多阶段细化框架（如 DoD [59],）来调整扩散路径；这些方法会利用早期阶段的视觉先验来增强语义结构和保真度。我们的方法在训练时对扩散路径进行调整，同时不依赖任何测试时条件。

3 方法

我们的目标是使用一小组手绘笔触训练扩散模型，使其能够生成在纹理和结构上都近似人类创作的笔触。我们提出的框架 *StrokeDiff*，通过视觉先验注入和笔画级可控性增强标准潜在

[1]https://github.com/mypaint/mypaint [2]
https://david.li/paint/



(A) Smooth Regularization (SmR) (B) StrokeDiff Framework

图 2. 我们方法的概述。(A) 平滑正则化在训练期间注入先验，以缓解有限数据下的模式崩溃。在每个时间步，从数据集中采样一个笔触 x_s，并与新采样的噪声项 $\epsilon^* \sim \mathrm{N}(0, I)$ 一起注入前向过程；两项均按因子 $\eta \sim \mathrm{Uni}[0, \Upsilon]$ 进行缩放。(B) StrokeDiff 框架分两个阶段进行训练。在阶段 1 中，模型学习在使用 SmR 但不使用条件输入的情况下生成笔触。在阶段 2 中，我们通过栅格转矢量模块引入参数条件化，该模块预测矢量化的笔画参数（例如控制点、不透明度、RGB、宽度）。这些参数随后用于控制特定笔触属性的生成。

扩散模型（见图2 了解整体概览）。我们首先训练模型，不使用推理时条件控制，并引入 平滑正则化（*SmR*），在训练过程中注入较弱的视觉先验，以缓解通过修改扩散过程对潜扩散模型（LDM）进行微调时观察到的模式崩溃。SmR 仅在训练期间应用，并在推理时停用，因此可以直接进行标准噪声采样，无需额外准备。为使输出兼容下游任务，我们还对笔触进行参数化，并支持基于笔画属性的可控生成。这些组件共同使 StrokeDiff 既能作为独立的笔触生成器，也能作为更广泛的绘画流程的基础；我们将在本节后文详细介绍该流程。

3.1 预备知识

扩散模型通过对初始高斯噪声输入反复去噪来生成样本。在前向过程中，干净图像 x_0 在马尔可夫链中通过逐步添加高斯噪声而被破坏。其一种表达方式为：

$$q(x_t \mid x_{t-1}) = \mathcal{N}\Big(x_t;\ \sqrt{\alpha_t}\, x_{t-1},\ (1-\alpha_t)I\Big), \tag{1}$$

其中，α_t控制从 x_{t-1}中保留的信息量。通过从 $t=1$到 T反复应用这一过程，我们可以写出以 x_0为条件的闭式分布：

$$q(x_t \mid x_0) = \mathcal{N}\Big(x_t;\ \sqrt{\bar{\alpha}_t}\, x_0,\ (1-\bar{\alpha}_t)I\Big), \tag{2}$$

其中，$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$。

为了逆转这一过程，我们训练神经网络 $\epsilon_\theta(x_t, t)$来预测添加到 x_0中的噪声 $\epsilon \sim \mathrm{N}(0, I)$，并最小化

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0,\epsilon,t}\Big[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\Big]. \tag{3}$$

利用这些预测，反向转移为：

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}\Big(x_{t-1};\ \mu_\theta(x_t, t),\ \Sigma_q(t)\Big), \tag{4}$$

其中，$\mu_\theta(x_t, t)$和 $\Sigma_q(t)$引导对 x_0的重建。

载于《ACM 测量、分析与计算系统会议录》，第37卷，第4期，第111号文章。出版日期：2018年8月。



图 3. 使用 Stable Diffusion v1-5 [39] 在缺乏足够多样性的典型笔触数据上进行微调的笔触生成。由于训练早期出现严重的模式崩溃，模型只能生成变化极少、类似纹理且缺乏真实笔触特征的图像。

3.2 通过视觉引导实现高保真笔触生成

3.2.1 问题设定。我们的工作聚焦于学习生成从单一艺术家处收集的笔触基元，其中包含470个扫描笔触。这些基元代表过程导向的视觉元素，在公共数据集中很少被单独观察到；对于开放域潜在扩散模型而言，它们在根本上属于分布外数据，因此无法通过提示调优或条件引导恢复。在该数据集上直接微调 Stable Diffusion 1.5 会导致严重的模式崩溃（见图3）。尽管这些笔触展现出家族级风格一致性，但其语义稀疏且结构多样，导致信息密度较低。因此，经过微调的模型无法在早期去噪步骤中推断全局结构。鉴于学习低频结构可能需要更长的时间范围 [13]，我们假设在标准扩散调度下，微弱的语义信号会迅速被噪声淹没。为此，我们提出一种正则化策略，以促进低信号区间中的结构保持。

3.2.2 随机视觉先验。我们在每个扩散步骤中引入视觉先验 x_s。我们不再仅依赖目标笔触的含噪重建，而是从训练集中采样 x_s 一个随机笔触，利用笔触间一致性提供结构感知引导。修改后的前向过程定义如下：

$$x'_t = \boxed{x_t} + \boxed{\sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\,x_s} - \boxed{\sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\,\epsilon^*_t}, \tag{5}$$

其中，η 是一个缩放因子（从 $\mathrm{Uni}[0, \Upsilon,\ 0 < \Upsilon < 1$ 中采样），用于调节先验注入的强度。每个训练实例仅采样一次，并在所有时间步间共享。值得注意的是，x'_t 并非由 x'_{t-1} 推导而来；x_t 在每个时间步与 x_s 进行插值，而 $\epsilon^* \sim \mathrm{N}(0,\ I)$ 是新采样的噪声，用于增加整体方差以维持多样性。否则，额外的结构会降低噪声幅度，并可能使训练产生偏差。所得分布 $q(x'_t \mid x_0)$ 同时包含数据项 $\sqrt{\bar{\alpha}_t}\,x_0$ 和视觉先验项 $\sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\,x_s$，如下所示：

$$= \mathcal{N}\Big(x'_t;\ \underbrace{\sqrt{\bar{\alpha}_t}\,x_0}_{\text{Data Term}} + \underbrace{\sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\,x_s}_{\text{Visual Prior}},\ \underbrace{(1+\eta)\,(1-\bar{\alpha}_t)\,\mathbf{I}}_{\text{Variance Term}}\Big). \tag{6}$$

视觉先验 x_s 不会在时间步之间传播。因此，注入的结构仅在时间上局部存在，不会沿着轨迹不断累积。与确定性调度方案不同，在后者中，噪声轨迹会在各个时间步和样本之间预先设定，这种不传播的特性能够防止模型记忆固定的结构模式。

我们推导出一个修改后的后验分布，该分布 $q(x'_{t-1}|x'_t, x_0)$

$$\propto \mathcal{N}\Bigg(x'_{t-1};\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})(1+\eta)x'_t + \sqrt{\bar{\alpha}_{t-1}}((1-\alpha_t-2\bar{\alpha}_t)\eta + 1-\alpha_t)x_0 + f(\eta)}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta},$$
$$\frac{((1+\alpha_t-2\bar{\alpha}_t)\eta + 1-\alpha_t)\,(1+\eta)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}\mathbf{I}\Bigg). \tag{7}$$

$f(\eta)$ 表示依赖于 η的独立项，表示为：

$$f(\eta) = \sum_{k=1}^{3} A_k \sqrt{\eta}^k, \tag{8}$$

系数 A_k由 α_t、 α_{t-1}、 $\bar{\alpha}_t$、 $\bar{\alpha}_{t-1}$ 和 x_s 决定。

3.2.3 平滑正则化。基于这一表述，我们将 τ 定义为噪声过程的一种替代表示：

$$\tau = \sqrt{1+\eta}\epsilon + \sqrt{\eta} x_s, \tag{9}$$

它同时包含学习到的噪声和注入先验。因此，我们可以将真实去噪转移均值 $\mu_q(x'_t,\ x_0)$ 参数化如下：

$$\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_t)x'_t + C_1 x'_t - \sqrt{\bar{\alpha}_{t-1}}\sqrt{1-\bar{\alpha}_t}(1-\alpha_t)\tau + C_2\tau + f(\eta)}{\sqrt{\bar{\alpha}_t}\,(1-\bar{\alpha}_t + (1+2\alpha_t - 3\bar{\alpha}_t)\eta)}, \tag{10}$$

其中， C_1 和 C_2 为系数，且每一项都乘以 η。由于 τ 引入了额外的随机性，模型必须同时学习**底层噪声及先验的影响**。因此，我们通过潜在扩散模型的神经网络参数化 $\hat{\tau}_\theta(x'_t,\ t)$，以此改进重建过程。于是，优化问题可简化为：

$$\mathbf{L}_\tau(\theta) = \mathbb{E}\left[||\tau - \hat{\tau}_\theta(x'_t, t)||^2\right]. \tag{11}$$

在推理过程中，将 $\eta = 0$ 设为 null 即可消除视觉先验，从而得到：

$$\mu_q(x'_t, t) = \mu_q(x_t, t) \quad \text{in DDPM}, \tag{12}$$

$$\Sigma_q(t) = \Sigma_q(t) \quad \text{in DDPM}. \tag{13}$$

这种设计仍兼容现有的扩散采样策略，使模型无需任何额外先验，仅凭高斯噪声即可生成图像。由于这种方法通过修改训练目标来“平滑”噪声，我们将其称为平滑正则化。在实践中，实现起来十分简单。此外，由于 x_s 是随机采样的，每个 x_0 都可以与多个不同的先验配对，从而有效扩大训练集，进一步提升笔触多样性。

3.3 面向下游应用的可控笔触合成

由于我们的笔触源自扫描图像，因此本身并不具备参数化形式。为实现参数化控制，我们采用可微光栅化器 [29]，先初始化一个简单的参数化基元，再通过迭代优化使其匹配目标笔触。具体而言，我们将路径和线段数量均设为 1，用单条三次贝塞尔曲线近似每个笔触：

$$c_p \ = \ (p_0,\ p_1,\ p_2,\ p_3,\ \mathrm{R},\ \mathrm{G},\ \mathrm{B},\ \mathrm{opacity},\ \mathrm{width}), \tag{14}$$

其中，每个 p_i 包含一个控制点的（x，y）坐标，从而得到 $c_p \in \mathbb{R}^{1\times 13}$。通过将笔触映射为这种矢量图形表示，我们无需手动记录笔刷轨迹 [1]，因此该方法适用于任何栅格化的笔触数据。

我们将此参数信息整合到模型中，方法是将c_p 与上下文特征 c（例如颜色或风格嵌入）拼接，并投影组合后的向量：

$$z \ = \ W\,[\,c_p;\,c\,], \tag{15}$$

其中， $[\cdot;\cdot]$ 表示向量拼接，W 是一个可学习的投影矩阵。随后，我们将 z 输入 U-Net 的交叉注意力机制，并对投影器和交叉注意力层进行微调。这种设置能够实现遵循贝塞尔曲线定义的形状的条件笔触生成，同时保持高保真度的光栅化效果。

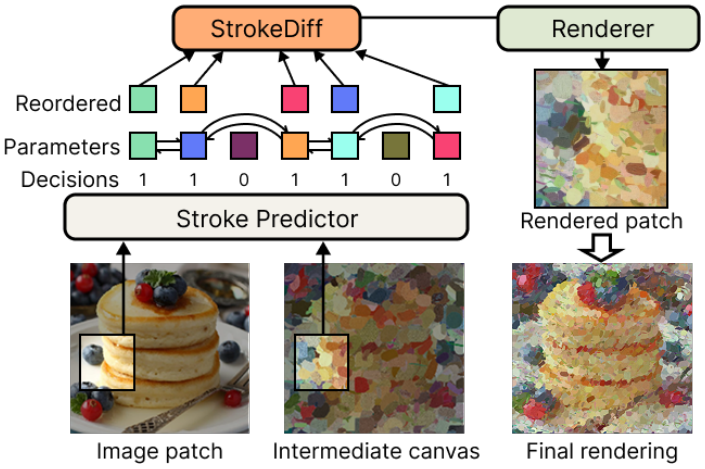


图 4. 绘画流程示意图。根据预测的排序分数 src_r 重新排列笔画参数，以确保渲染顺序连贯。

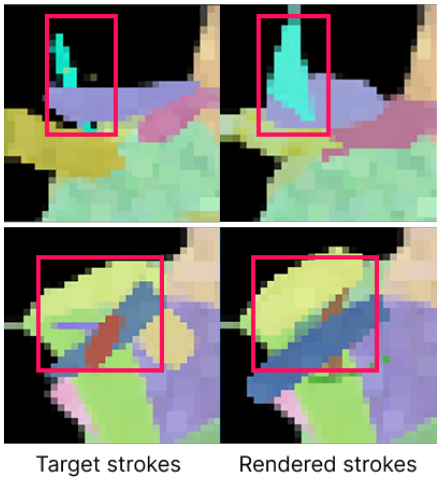


图 5. 仅进行最优匹配的笔触会产生无序序列，导致重叠错误。

3.4 绘画流程集成

为了渲染完整的绘画作品，我们将 StrokeDiff 集成到基于笔触的渲染流程中。该过程包括：(i) 从目标图像中预测笔画参数（位置、形状、大小），(ii) 使用我们的扩散模型生成相应的笔触，以及 (iii) 借助渲染器在画布上依次合成这些笔触，如图 4 所示。

笔触预测器。我们采用类似于 PaintTransformer [33], 的 DETR 风格架构，为每个图像区域输出一组候选笔触。每个笔触预测表示为 $(c_p,\ x_{\text{shift}},\ y_{\text{shift}},\ \text{scr}_r)$，其中$x_{\text{shift}},\ y_{\text{shift}}$ 指定位置偏移量，$\text{scr}_r \in (0,\ 1)$是表示相对绘制顺序的排序分数。在训练时，我们首先重新定义匹配损失，在最优分配下将每个预测笔触与真实笔触对齐：

$$\mathcal{L}_{match} = \min_{\sigma} \sum_{i} \Big(\lambda_{\text{L1}} \|P_i^- - \hat{P}_{\sigma(i)}^-\|_1 + \lambda_{\cos} \operatorname{Cosine}(P_i^-, \hat{P}_{\sigma(i)}^-) + \lambda_{\text{D}} \operatorname{BCE}(d_i, \hat{d}_{\sigma(i)}) \Big). \tag{16}$$

其中 $P^- = (c_p,\ x_{\text{shift}},\ y_{\text{shift}})$不包括排序分数，$\sigma$是最优二分图匹配，而 $d \in \{0,\ 1\}$表示笔触是否存在。

然而，基于集合的预测不施加任何排序，这往往会导致直接绘制笔触时出现不连贯的重叠，如图 5 所示。为解决这一问题，我们引入排序损失。假设真实值提供了一个按正确绘制顺序编入索引的笔触序列 $\{1,\ 2,\ \ldots\ldots,\ vs\}$，其中索引较小表示该笔触应更早绘制。对于每个有效序列，预测器输出 $\{\text{scr}_{r,\ u}\}_{u=1}^{vs}$，我们对预测分数进行正则化：

$$\mathcal{L}_{\text{rank}} = \sum_{i<j} \max\Big(0,\ \text{scr}_{r,\sigma(i)} - \text{scr}_{r,\sigma(j)} + (j-i)\cdot \text{margin}\Big), \tag{17}$$

其中 i 和 j 是真实笔触的索引。该损失会惩罚这样一种情况：本应较早出现（ i较小）的笔触，其得分却高于或几乎等于较晚出现（ j较大）的笔触。间隔项



图6. 收集的笔触示例。

要求间隔与索引差成正比，从而防止序列中相距较远的笔触得分趋于相同。

整体训练目标为

$$\mathcal{L}_{\text{total}} = \lambda_m \mathcal{L}_{\text{match}} + \lambda_r \mathcal{L}_{\text{rank}}, \tag{18}$$

构图。 训练完成后，预测器会同时输出笔画参数及其推断出的顺序。随后，StrokeDiff 生成笔触，并逐层依次渲染，从而创作出连贯的绘画，减少重叠伪影，并呈现更自然的时间结构。

4 实验

4.1 数据集

为了更贴近人类的绘画行为，我们从一位专业艺术家处收集了 520 个独特的油画笔触，并整理成一个数据集。每个笔触都在一个 20×26 网格画布上自由绘制，如图6所示，图中展示了原始扫描稿。随后，我们从画布上裁剪出单个笔触，并将其统一调整为295×295 像素。我们将数据集划分为 470 个训练笔触和 50 个测试笔触。为缓解样本量有限的问题，我们采用了一系列预处理和增强步骤，包括背景去除、颜色归一化、翻转和旋转。经过这些处理，数据集扩充至 9,400 张训练图像和 1,000 张测试图像。该数据集已公开发布，数字对象标识符为DOI: 10.21227/t6bd-cs26 ，以支持后续研究。

4.2 实施细节

StrokeDiff 使用包含 9,400 张增强笔触图像的数据集进行训练。对于每个训练实例，我们从数据集中随机采样另外 32 个笔触作为视觉先验，共得到 300,800 对训练样本。该模型以 Stable Diffusion v1_5 为基础，生成分辨率为 512×512的图像。我们采用 Stable Diffusion 原有的噪声调度，其中 $\bar{\alpha}_t = \prod_{i=0}^{t}(1-\beta_i)$，$\beta_t$ 在 $\sqrt{0.00085}$ 与$\sqrt{0.012}$之间进行线性插值。η 的先验上界 Υ 设为 0.5。训练使用 PyTorch 1.9.0 完成。模型首先在不进行参数条件化的情况下训练 4 个轮次，随后使用条件输入进行 5 个轮次的微调。

笔触预测器处理尺寸为 32×32 的输入画布，并预测每张图像最多 8 个笔触。匹配损失权重向量 $\lambda_m = [5, 10, 10]$ 分别对应 L_1、余弦损失和决策损失项。$\lambda_r = 5$，间隔值为 0.125。模型训练了 220 个训练轮次。

表1. 基于扩散的方法在笔触生成上的比较，所有方法均基于 SD1.5 训练。 $n.$ 和 $r.$ 分别表示每个笔触的闭合区域平均数量**n**以及区域面积与完整图像面积的比值**r**。 Δ 表示相对于仅使用数据集的基线的变化（FID = 47.3，在训练集和测试集之间计算；CRD (n./r.) = 1.09 (0.188)）。

方法	FID ↓	Δ	CRD（名词/可数名词）↓	Δ
INS b=0.3[5]	273	+225	48.51 (0.675)	+47. (+0.49)
INS b=0.6[5]	251	+203	44.14 (0.732)	+43. (+0.54)
SD1.5_LoRA[15]	285	+237	88.13 (0.760)	+87. (+0.57)
SmR（我们的方法）	**54**	**+6**	**1.24 (0.254)**	**+0.15（+0.07）**

表2. 与先前用于可控生成的笔触生成方法进行比较，所有方法均在我们的数据集上训练。
† StrokeGAN 和 Stylized 在训练过程中以笔画参数为条件。RobPaint 不支持条件生成，因此我们使用其重建输出进行评估。

模型	Size	LPIPS↓	MSE↓	FID↓
StrokeGAN [53]	64 × 64 × 3	0.285	0.064	245
Stylized [61]	128 × 128 × 3	0.242	0.036	273
RobPaint [1]†	64 × 32 × 1	0.234	0.085	322
StrokeDiff	512 × 512 × 3	**0.232**	**0.031**	**52**

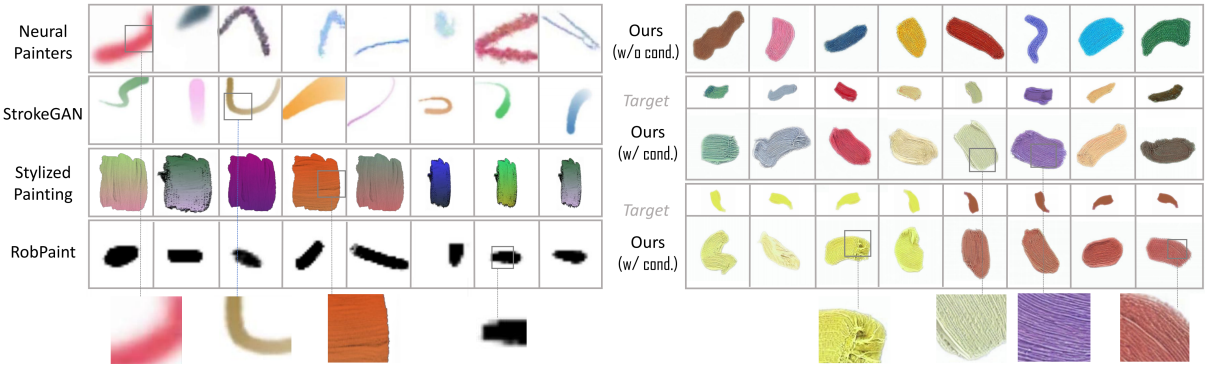


图7. 与其他方法的定性比较：1）Neural Painters [36], 使用基于合成 MyPaint 笔触训练的变分自编码器和生成对抗网络；2）StrokeGAN [53], 一种基于四阶贝塞尔曲线训练的生成对抗网络；3）模板笔触 [33, 47, 50, 61], 取自基于笔触的渲染中普遍采用的一组形状；以及 4）RobPaint [1], 一种使用人工绘制的灰度笔触训练的变分自编码器。我们的方法在保留笔刷纹理的同时，生成了更清晰的笔触。

4.3 平滑正则化对笔触生成的影响

4.3.1 评估少样本扩散中的泛化。 表1 报告了将 SmR 与无条件笔触生成的代表性低资源训练策略进行比较的定量结果，包括噪声调度调整和基于 LoRA 的微调。评估使用 FID [11], 以及我们提出的一种结构敏感指标——闭合区域检测（CRD）。提出 CRD 的动机在于，笔触生成失败通常表现为两类情况：要么是类似多个小笔刷斑块的碎片化伪影，要么是笔触融合成难以分辨的大块团块。为捕捉这些结构失败，CRD 度量（i）每幅图像中的闭合区域数量，以及（ii）笔触面积相对于画布的比例。二者均通过 OpenCV 中的连通分量分析计算。对于噪声调度，我们采用 策略 *2* 来自 INS [5], 修改缩放因子 b 在前向噪声项中，以减轻其抵消效应。设置 $b = 1$ 对应于默认调度。我们评估 b 取值为 0.3 和 0.6。尽管 LoRA [15] 最初是为高效微调而提出的，但由于其参数效率较高，我们将其作为一种强有力的低数据基线纳入比较。尽管仅使用 470 个手绘笔触，SmR 仍取得了显著更低的 FID。基于 INS 的噪声调度和 LoRA 均未能防止结构坍缩或模式增殖：检测到的区域数量被显著高估，且 FID 仍然较高（定性示例见图 11 第 2–4 列）。



图 8. 不同基于笔触的渲染管线生成绘画的定性比较：Diffvg [29],Learn2Paint [22],Stylized [61],PT [33],和 StrokeGAN [53]。该图展示了不同笔画生成策略如何影响最终构图中的纹理外观、笔触不规则性和分层行为。放大查看差异。

4.3.2 与笔画生成方法的定量比较。我们进一步在可控笔触生成的背景下评估 SmR，并与先前工作进行比较，包括 StrokeGAN [53], Stylized [61], 和 RobPaint [1]。表 2 报告了 LPIPS [60], MSE 和 FID 分数，其中应用 SmR 的模型在所有指标上均优于所有方法。我们注意到，与自然图像不同，笔触数据通常表现出较低的感知方差和像素级方差，这限制了 LPIPS 和 MSE 在区分结构质量方面的敏感性。我们发现，在实验中 FID 在捕捉分布失配方面仍更具判别力，而 SmR 取得了最低的 FID，表明其输出与训练数据分布更为一致。定性结果见附录。

4.3.3 其他方法生成笔触的定性比较。图7展示了几种具有代表性的笔触生成模型所产生的定性结果。Neural Painters [36] 和 StrokeGAN [53] 生成的结果较为平滑，但纹理保真度有限。基于模板的方法像素准确度较高，但多样性受限；RobPaint [1] 生成的笔触更具自然感，但由于采用灰度训练，缺乏色彩表现力。相比之下，StrokeDiff 兼具无条件生成和条件生成能力：在无条件单次生成设置下，它能生成特征多样、带纹理且结构完整性清晰的笔触。以特定参数为条件时，该模型在形状和外观上都能准确匹配目标。最后一行展示了训练中使用的增强目标，凸显了模型捕捉丰富多样笔触风格的能力。

4.4 StrokeDiff 在下游绘画任务中的应用

接下来，我们评估 第3.4节所述的渲染阶段，重点关注完整画作的质量。

我们比较我们的方法 mer

与 Diffvg [29], Learn2Paint [22], PaintTransfor 等具有代表性的方法进行对比

表 3. 不同方法的 CLIP 分数比较。Ori. 表示参考图像。

方法		Ori.	Diffvg	L2P	风格	PT	SGAN	Ours
CLIP 分数	第1章	0.125	0.1704	0.1879	0.1981	0.1751	0.1886	0.1951
	第2章	0.117	0.1757	0.2004	0.2059	0.1896	0.2019	0.2089
	第3章	0.106	0.1710	0.1970	0.2079	0.1901	0.1960	0.2226
	第4章	0.131	0.1950	0.2254	0.2294	0.2079	0.2097	0.2363

表 4. 绘画质量人工评估（1–5 分制）。

	Diffvg	L2P 风格化	PT	S.GAN	Ours
风格	3.21	3.09	3.35	3.18	3.26	**3.38**
美感	2.87	2.50	3.00	2.76	3.11	**3.38**
纹理	2.84	2.39	3.18	2.79	2.94	**3.50**
内容	**4.35**	4.12	3.37	3.48	3.39	2.59
总计	**3.32**	3.03	3.24	3.05	3.17	3.21

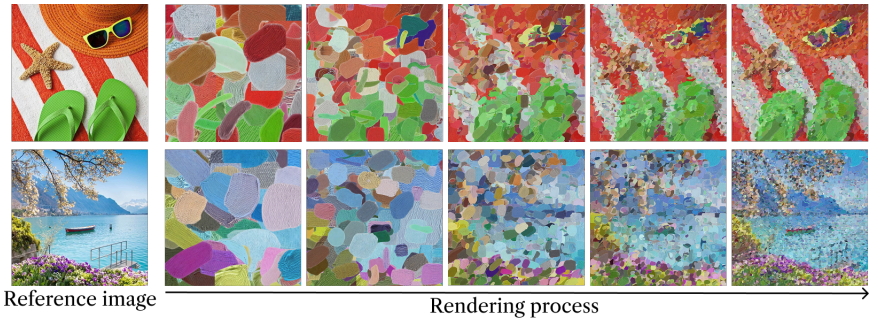


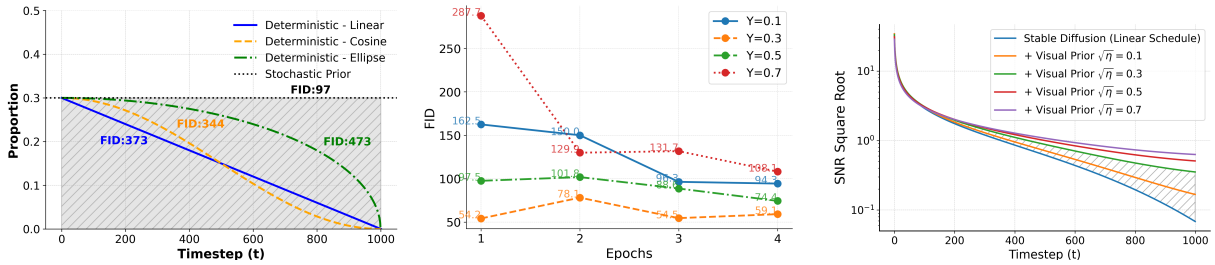
图 9. 渲染过程中的中间画布。

(PT) [33], Stylized [61], 和 StrokeGAN [53]。这些方法涵盖可微矢量图形、强化学习、前馈预测和基于 GAN 的笔触生成。

4.4.1 绘画的定性评估。图 8 比较了不同基于笔触的渲染方法生成的绘画结果。Diffvg 和 Learn2Paint 专注于精确的像素级重建，生成的图像与参考图像十分接近，但依赖相对均匀且由几何形状驱动的笔触模式。Stylized 和 StrokeGAN 增强了形状的多样性，但其笔触往往呈现出简化的纹理或不够一致的分层行为。我们的方法则采用了不同的侧重点，将笔触建模为纹理丰富且不规则的视觉基元。因此，渲染出的绘画呈现出更为明显的笔触纹理、更清晰的图层结构，其视觉效果也更符合油画的特点。图 9 进一步展示了这些笔触层面的属性如何在序列渲染过程中逐渐累积。

4.4.2 绘画的定量评估。表3 报告了通过 CLIP 分数衡量的语义对齐情况。我们使用以下四条描述性文本对每种方法在五张图像上进行评估：“由人类创作的油画”、“笔触油画”、“纹理油画”，以及 “由人类创作、具有强烈笔触纹理的风格化油画”。结果表明，在第一条描述下，我们的方法取得了与 Stylized 相近的分数。在其余三条强调笔触纹理和风格化的描述下，我们的方法均稳定优于所有其他方法。这些提升表明，StrokeDiff 与面向人类的油画描述具有更好的对齐效果，尤其是在笔触质量和艺术表现方面。

4.4.3 用户研究。在 [52], 我们邀请 51 名参与者进行了人工评估，其中 26 人具有正规艺术背景，25 人没有。参与者根据四项标准对每种方法进行评分：油画风格、审美价值、纹理和内容保留度。如表4所示，我们的方法在风格、美感和纹理方面获得了最高评分。然而，该方法对艺术表现的侧重导致内容保留度评分较低，这表明保真度与风格化之间存在权衡。问卷见附录。



(a) 随机与确定性对比。 (b) 对 Υ 值的评估。 (c) 信噪比变化。

图 10. 消融结果。(a) 随机先验与确定性先验的注入趋势；(b) 不同Υ 值下各训练轮次的 FID；(c) 信噪比变化。我们对标准信噪比取平方根。

表 5. 对每个数据样本使用不同数量目标先验对的评估。左侧和右侧组分别对应 Υ 设为 0.3 和 0.5 的情况。

# 多组配对 →	**10**	**20**	**32**	**10**	**20**	**32**
FID	148.6	91.6	54.2	108	97.5	74.4

4.5 消融研究

4.5.1 分析视觉先验注入的作用。

随机视觉先验评估。我们将随机先验注入策略与确定性替代方案进行了比较。在我们的方法中，注入因子随机采样，且不随步数变化；而确定性先验则遵循随时间变化的固定调度——类似于标准扩散模型中常用的噪声调度（例如线性、余弦或缩放余弦调度），通过预定义的 α_t 或 β_t 曲线控制扩散强度。我们测试了若干此类确定性变体，包括线性、余弦和椭圆形先验权重。如图10a所示，我们的方法显著降低了 FID，因为它避免模型传播固定信息，使模型能够更好地适应笔画变化并学习更广泛的样本分布。图中的灰色阴影区域表示随机方法引入的先验值可能范围，突显其不再遵循固定轨迹。

先验上界评估。我们评估了不同的先验上界取值 Υ。如图10b所示，在测试的四种取值下，模型都能收敛；其中，Υ =00.3 和 0.5 时，尤其是 0.3 时，模型在训练初期便呈现出明显的收敛趋势，并在第一个训练周期内趋于稳定。在其他实验中，我们发现Υ = 0.5 在条件训练后的表现与 = 0.3 相当，但在低数据场景下更加稳健。因此，设为 Υ =00.3 可在数据充足时加快训练，而较高的上界则能在低数据场景下提供更强的指导。表5 展示了每个样本采用不同先验匹配数量时的结果；可用配对越少，较大的 Υ 越有利。我们未测试超过 32 组配对的情况，因为这已足以生成高质量笔画。

对信噪比动态的影响。尽管我们的方法并未显式执行噪声调度，但在前向过程中注入视觉先验会固有地改变信噪比（SNR）。这种联系为理解 SmR 如何改进训练提供了一个互补视角。图10c 展示了在各个时间步统一应用$\sqrt{\eta} = 0.3$ 时的预期信噪比轨迹，并将先验视为信号的一部分（绿色曲线）。实际情况下，

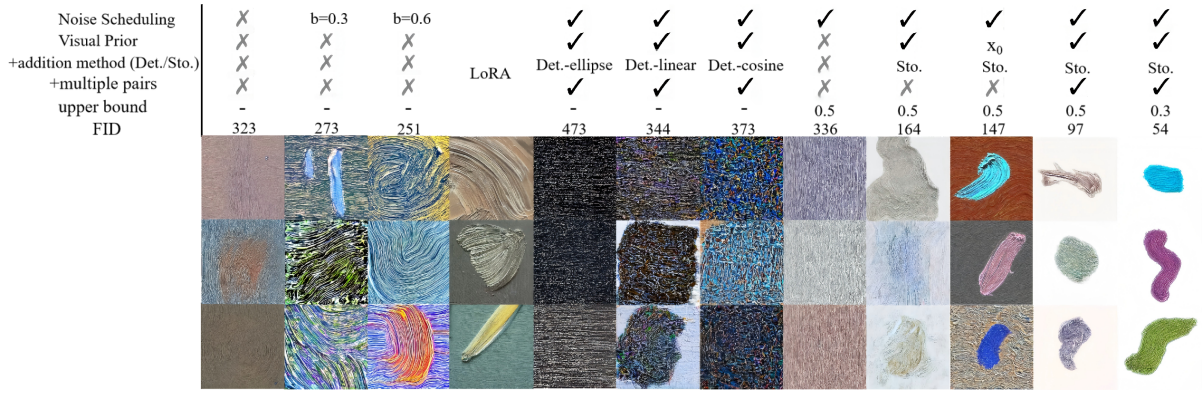


图 11. 消融实验的示例结果。由于默认的先验-样本对数量为 32，所有“无多组配对”设置均训练了 32 个训练轮次，以匹配更新总次数，而其他设置训练了 1 个训练轮次。 *Det.* 和 *Sto.*分别表示确定性和随机先验注入方法。虽然 LoRA 并非消融实验的一部分，我们仍在第四列中列出其结果以供定性比较。

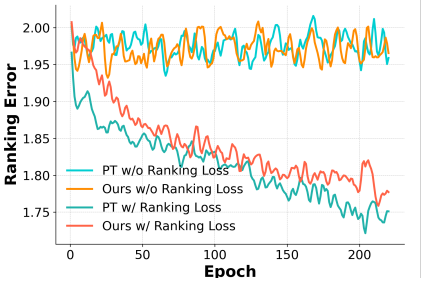


图 12. 使用和不使用排序损失时的排序误差。对于不使用排序损失的方法， src_r 替换为预测索引 ×间隔。

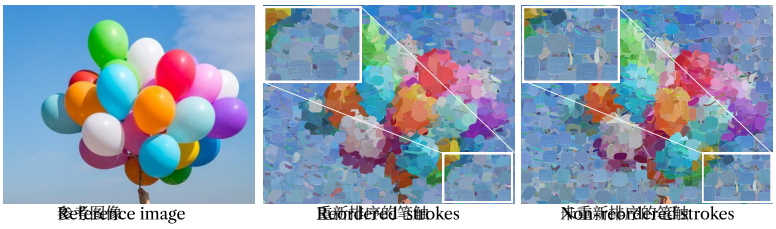


图13. 使用预测排序分数、带有与不带重新排序笔触的绘画。

由于随机注入，实际信噪比会在阴影区域内波动，并始终高于标准 Stable Diffusion。

增强效果。 为了分离SmR的增强效果，我们进一步测试了一种退化情形，其中先验 x_s 被替换为 x_0（真实值）。如图11 （倒数第三列）所示，即使没有引入新信息，该变体也显著降低了FID，表明SmR除数据增强之外还充当了有效的训练时正则化器。在其他领域探索这一性质是未来工作的一个有趣方向。

4.5.2 排序损失消融。我们通过衡量有无排序损失时预测笔触顺序与真实笔触顺序之间的序列误差，评估了笔触预测器维持连贯渲染顺序的能力。如图所示。12，没有排序损失时，模型采用的基于集合的预测方法会产生完全无序的序列，进而影响绘画整体的连贯性。加入排序损失后，笔画顺序会逐渐与正确序列对齐。我们还在 PT [33] 上测试了这一方法，并观察到类似的

表 6. 在 Open-Images 子集上使用 Paint-by-Example 的结果。

	8% Open-Images			1% Open-Images		
方法	FID ↓	QS ↑	CLIP ↑	FID ↓	QS ↑	CLIP ↑
Paint-by-Example	11.667	73.254	**74.120**	27.421	27.473	**58.827**
+ SmR (x_0)	**11.467**	**73.981**	73.544	**26.858**	**31.148**	57.179

改进。图13 进一步表明，如果不重新排序，笔触会出现明显的分层伪影；而重新排序后，融合效果则更加连贯。

4.6 笔画生成之外的通用性

为考察平滑正则化是否也能使笔触合成之外的领域受益，我们在图像修复任务 Paint-by-Example 上进行了初步测试 [57], ，该任务提供了完整的训练与评估框架。按照相同的训练流程，我们采用了 SmR 的一种退化形式，将视觉先验设为输入 x_0。基线模型和加入 SmR 的模型均在 Open- Images [27] 数据集的子集上训练，使用了 8% 或 1% 的训练数据。表 6 总结了结果。当数据量为 8% 时，两者的性能总体相当；而在数据极度稀缺（1%）时，SmR 带来了小幅提升，FID 更低，质量评分更高。这些初步结果表明，SmR 或可作为笔触生成之外的一种有效训练时正则化器，不过，未来仍需在多种不同任务上进行更广泛的验证。

5 讨论

5.1 潜在应用场景

除了当前实验的范围之外，也值得探讨如何将笔画级生成应用于相关领域。

机器人绘画。一个自然的方向是机器人绘画，即将基于扩散的笔触映射到机械臂轨迹，以便以更高的保真度再现类似画笔的外观。FRIDA [41] 和 COFRIDA [43] 等近期系统已经证明了将生成模型与机器人绘画相结合的可行性，而笔触级合成可通过提供更多样、更具纹理的视觉单元，进一步丰富这一过程。

创意辅助工具。 创意辅助工具是一类交互式系统，通过提供实时、可控的增强来辅助创作过程，同时让用户保持主要的创作控制权。在此背景下，我们笔触的纹理表现力可以扩展为嵌入数字绘画工具链中的渲染单元，例如触控笔或触摸屏画笔。这一思路与 Diffusion Texture Painting [14], 相关，后者展示了扩散模型如何以带纹理的反馈来丰富用户输入。StrokeDiff 的不同之处在于它专注于笔触级合成，因此未来可被探索为一种互补组件，用于在教育或创意软件中生成具有绘画感的分层笔刷效果。

2.5D打印。2.5D打印和浮雕印刷技术（例如佳能的表面浮雕打印）取得的进展，使再现艺术作品的触觉质感成为可能。然而，现有流程通常依赖将 RGB值启发式映射为高度图，这可能无法捕捉细腻的笔刷纹理。由于我们的笔触明确编码了分层且富有纹理的外观，因此可以将其转换为可打印的高度场，保留油画的实体质感。此类应用将使笔触合成从视觉渲染扩展到实体再现。

5.2 人工智能绘画中的评估挑战

在我们的用户研究中，我们请参与者评价："从照片到绘画，作品对内容的呈现效果如何？"与风格、美学和纹理相比，这个与内容相关的问题得分明显较低。这引出了一个重要问题：内容得分较低是否必然意味着表现较差，还是说它反映了非照片级渲染中至关重要的抽象与保真度之间的张力？在我们的研究中，生成的笔触没有改变物体的存在与布局，而是以不同程度的抽象重新诠释相同的内容，因此"内容保留"这一概念本身有些模糊。从一个角度看，保真度降低是一个缺点；从另一个角度看，这却是艺术风格化的预期特性。这种模糊性揭示了人工智能绘画系统评估中一个更广泛的问题。现有关于美学评估 [24] 或艺术评论 [2] 的研究大多关注完成后的艺术作品，往往聚焦于已完成的杰作或历史绘画，与面向过程的生成式绘画有显著差异。因此，目前仍缺乏全面的基准测试，能够根据任务需求，从多个维度评估绘画系统，例如语义连贯性、抽象程度、笔触表现力和风格意图。常用的图像指标，包括 FID 或基于 CLIP 的相似度，主要衡量分布对齐或粗略的语义对应，却无法敏锐捕捉艺术渲染所重视的非刚性、感知层面的特质。近期针对生成任务的评估方案已开始将大型多模态语言模型纳入其中，作为灵活的评判工具，例如在 T2I-CompBench 等基准测试中采用基于 GPT 或视觉问答式的评估++ [20]，以及使用 LLaVA 系列 [31] 和 Qwen2-VL [51]等通用视觉语言模型。这些模型能够对难以通过数值形式加以定义的定性维度进行评估。尽管仍不够完善，基于多模态大语言模型的评估为超越传统像素级或嵌入级指标、评估人工智能绘画系统提供了一条前景可期的补充路径。

6 局限性与未来工作

6.1 渲染管线

在本研究中，我们采用了类似 PaintTransformer 的预测器作为下游渲染管线，主要是因为其前馈设计具有较高的计算效率。然而，这一设计选择也继承了 Learn2Paint 和 StrokeGAN 等类似方法的常见局限性。具体而言，基于网格的逐层渲染策略会在第 k 层将画布划分为 $2^k \times 2^k$ 个图块，导致笔触受到局部约束，且尺寸往往较为一致。这种行为不同于人类绘画实践：艺术家会根据不同区域和预期效果，自然而然地改变笔触长度、方向和连续性。另一种思路是将 StrokeDiff 集成到 Diffvg 等框架中，在画布范围内联合优化笔触位置。原则上，这种集成可以生成更长、贴合物体轮廓的笔触，并减少网格伪影，但需要调整笔触参数化方式并将其集成到渲染管线中。

6.2 风格多样性与泛化

StrokeDiff 特别适用于纹理丰富的彩色绘画风格，例如油画；在这类风格中，不规则边界和层叠颜料是视觉效果的核心。素描笔触等其他领域的视觉特征则截然不同：它们通常较为稀疏、呈单色，且缺乏连续纹理，因此，与基于扩散的纹理合成相比，采用光栅化的矢量域建模或许是更自然的选择。对于水彩和水墨等其他彩色媒介，我们尚未尝试进行泛化。这里的主要局限在于缺乏特定领域的数据集，因此难以捕捉透明度、颜料混合和负空间效果等现象。为了探索这一方向，我们目前正在构建一个包含 1,008 个水墨风格笔触的数据集。这项资源将使我们能够研究

在具有不同物理属性的媒介中进行笔触生成，并为今后调整笔触参数化和渲染规则提供参考。

6.3 SmR 的泛化能力

尽管我们在第 4.6 节中使用图像修复框架进行了初步实验，但仅凭这一实验还不足以证明 SmR 能够跨领域泛化。它能否作为一种通用正则化策略，仍需通过更广泛的验证来确定。尽管如此，对于数据收集困难，或旨在生成可用作构建模块的基础基元的任务，SmR 可能具有潜在价值，值得进一步研究。

7 结论

本研究探讨了数据有限条件下的笔触生成问题，并提出了带平滑正则化的 StrokeDiff，这是一种高效利用数据的扩散框架。结果表明，笔触级建模既能呈现丰富的结构多样性，也能充分表达风格；此类基元还可组合成令人信服的绘画作品。更广泛地说，本研究展示了生成模型经领域特定基元适配后，如何为非真实感渲染作出贡献。未来的工作可将这一方向拓展至其他艺术媒介，并探索更全面的评估方案，以兼顾抽象程度与保真度。

致谢

本研究由香港研究资助局通过主题研究计划（T45-205/21-N）资助。

参考资料

[1] Ardavan Bidgoli、Manuel Ladron De Guevara、Cinnie Hsiung、Jean Oh 和 Eunsu Kang。2020。机器人绘画中的艺术风格：一种从人类艺术家处学习笔触的机器学习方法。载于《2020年第29届IEEE机器人与人类交互通信国际会议（RO-MAN）》。IEEE，第412–418页。[2] Yi Bin、Wenhao Shi、Yujuan Ding、Zhiqiang Hu、Zheng Wang、Yang Yang、See-Kiong Ng 和 Heng Tao Shen。2024。Gallerygpt：利用大型多模态模型分析绘画作品。载于《第32届ACM国际多媒体会议论文集》。第7734–7743页。[3] Eva Cetinic 和 James She。2022。借助人工智能理解与创作艺术：综述与展望。《ACM 多媒体计算、通信与应用汇刊》（TOMM）18卷2期（2022），第1–22页。[4] Lawrence Chen、Peter Schaldenbrand、Tanmay Shankar、Lia Coleman 和 Jean Oh。2025。Spline-FRIDA：借助高效采样、可微笔触模型，实现多样且类人的机器人绘画风格。《IEEE机器人与自动化快报》（2025）。[5] Ting Chen。2023。噪声调度对于扩散模型的重要性。arXiv预印本 arXiv:2301.10972（2023）。[6] Kevin Clark 和 Priyank Jaini。2023。文本到图像扩散模型是零样本分类器。载于《神经信息处理系统进展》，A. Oh、T. Naumann、A. Globerson、K. Saenko、M. Hardt 和 S. Levine（编），第36卷。Curran Associates, Inc.，第58921–58937页。https://proceedings.neurips.cc/paper_files/paper/2023/file/b87bdcf963cad3d0b265fcb78ae7d11e-Paper-Conference.pdf [7] Prafulla Dhariwal 和 Alexander Nichol。2021。生成对抗网络在图像合成方面不及扩散模型。《神经信息处理系统进展》34卷（2021），第8780–8794页。[8] Wenkai Dong、Song Xue、Xiaoyue Duan 和 Shumin Han。2023。利用扩散模型进行文本驱动图像编辑的提示调优反演。载于《IEEE/CVF国际计算机视觉大会论文集》。第7430–7440页。[9] Jiatao Gu、Shuangfei Zhai、Yizhe Zhang、Miguel Angel Bautista 和 Josh Susskind。2022。f-dm：通过渐进式信号变换实现的多阶段扩散模型。arXiv预印本 arXiv:2210.04955（2022）。[10] Amir Hertz、Ron Mokady、Jay Tenenbaum、Kfir Aberman、Yael Pritch 和 Daniel Cohen-Or。2022。通过交叉注意力控制实现提示词到提示词图像编辑。（2022）。网址：https://arxiv. org/abs/2208.01626（2022）。[11] Martin Heusel、Hubert Ramsauer、Thomas Unterthiner、Bernhard Nessler 和 Sepp Hochreiter。2017。采用双时间尺度更新规则训练的生成对抗网络收敛于局部纳什均衡。《神经信息处理系统进展》30卷（2017）。[12] Jonathan Ho 和 Tim Salimans。2022。无分类器扩散引导。arXiv预印本 arXiv:2207.12598（2022）。[13] Emiel Hoogeboom、Jonathan Heek 和 Tim Salimans。2023。simple diffusion：面向高分辨率图像的端到端扩散。载于《国际机器学习会议》。PMLR，第13213–13232页。

[14] Anita Hu、Nishkrit Desai、Hassan Abu Alhaija、Seung Wook Kim 和 Maria Shugrina。2024。扩散纹理绘画。载于《ACM SIGGRAPH 2024会议论文集》。第1–12页。[15] Edward J Hu、Yelong Shen、Phillip Wallis、Zeyuan Allen-Zhu、Yuanzhi Li、Shean Wang、Lu Wang、Weizhu Chen 等。2022。LoRA：大语言模型的低秩适配。ICLR 1卷2期（2022），第3页。[16] Hexiang Hu、Kelvin CK Chan、Yu-Chuan Su、Wenhu Chen、Yandong Li、Kihyuk Sohn、Yang Zhao、Xue Ben、Boqing Gong、William Cohen 等。2024。Instruct-imagen：基于多模态指令的图像生成。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。第4754–4763页。[17] Juncheng Hu、Ximing Xing、Jing Zhang 和 Qian Yu。2024。VectorPainter：利用笔触风格先验实现先进的风格化矢量图形合成。arXiv预印本 arXiv:2405.02962（2024）。[18] Teng Hu、Ran Yi、Haokun Zhu、Liang Liu、Jinlong Peng、Yabiao Wang、Chengjie Wang 和 Lizhuang Ma。2023。基于笔触的神经绘画与风格化，以及动态预测绘画区域。载于《第31届ACM国际多媒体会议论文集》。第7470–7480页。[19] Zhangli Hu、Ye Chen、Zhongyin Zhao、Jinfan Liu、Bilian Ke 和 Bingbing Ni。2024。通过多粒度语义对齐迈向类艺术家绘画智能体。载于《第32届ACM国际多媒体会议论文集》。第10191–10199页。[20] Kaiyi Huang、Chengqi Duan、Kaiyue Sun、Enze Xie、Zhenguo Li 和 Xihui Liu。2025。T2i-compbench++：用于组合式文生图生成的增强型综合基准测试。《IEEE模式分析与机器智能汇刊》（2025）。[21] Nisha Huang、Weiming Dong、Yuxin Zhang、Fan Tang、Ronghui Li、Chongyang Ma、Xiu Li、Tong-Yee Lee 和 Changsheng Xu。2025。CreativeSynth：利用多模态扩散中的跨艺术注意力进行艺术图像合成。《IEEE可视化与计算机图形学汇刊》（2025）。[22] Zhewei Huang、Wen Heng 和 Shuchang Zhou。2019。通过基于模型的深度强化学习学习绘画。载于《IEEE/CVF国际计算机视觉大会论文集》。第8709–8718页。[23] Francisco Ibarrola、Tomas Lawton 和 Kazjon Grace。2023。用于协同创作设计的协作式、交互式、情境感知绘图智能体。《IEEE可视化与计算机图形学汇刊》30卷8期（2023），第5525–5537页。[24] Xin Jin、Qianqian Qiao、Yi Lu、Huaye Wang、Heng Huang、Shan Gao、Jianfei Liu 和 Rui Li。2024。APDDv2：绘画与素描美感数据集，包含艺术家标注评分与评论。arXiv预印本 arXiv:2411.08545 8卷9期（2024）。[25] Dmytro Kotovenko、Matthias Wright、Arthur Heimbrecht 和 Bjorn Ommer。2021。重新思考风格迁移：从像素到参数化笔触。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。第12196–12205页。[26] Nupur Kumari、Bingliang Zhang、Richard Zhang、Eli Shechtman 和 Jun-Yan Zhu。2023。文本到图像扩散的多概念定制。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。第1931–1941页。[27] Alina Kuznetsova、Hassan Rom、Neil Alldrin、Jasper Uijlings、Ivan Krasin、Jordi Pont-Tuset、Shahab Kamali、Stefan Popov、Matteo Malloci、Alexander Kolesnikov 等。2020。Open Images数据集v4：大规模统一图像分类、目标检测和视觉关系检测。《国际计算机视觉杂志》128卷7期（2020），第1956–1981页。[28] Jing Li、Jun Guo、M Shamim Hossain 和 Ning Yu。2025。用于艺术治疗的多媒体沉浸式水墨山水设计。《ACM 多媒体计算、通信与应用汇刊》21卷11期（2025），第1–26页。[29] Tzu-Mao Li、Michal Lukáč、Michaël Gharbi 和 Jonathan Ragan-Kelley。2020。用于编辑与学习的可微矢量图形光栅化。《ACM图形学汇刊》（TOG）39卷6期（2020），第1–15页。[30] Shanchuan Lin、Bingchen Liu、Jiashi Li 和 Xiao Yang。2024。常见的扩散噪声调度与采样步数存在缺陷。载于《IEEE/CVF冬季计算机视觉应用会议论文集》。第5404–5411页。[31] Haotian Liu、Chunyuan Li、Qingyang Wu 和 Yong Jae Lee。2023。视觉指令微调。《神经信息处理系统进展》36卷（2023），第34892–34916页。[32] Ruoxue Liu、Linjiajie Fang、Wenjia Wang 和 Bing-Yi Jing。2024。D2R2：利用随机距离匹配的基于扩散的表征，用于表格小样本学习。载于《神经信息处理系统进展》，A. Globerson、L. Mackey、D. Belgrave、A. Fan、U. Paquet、J. Tomczak 和 C. Zhang（编），第37卷。Curran Associates, Inc.，第36890–36913页。https://proceedings.neurips.cc/paper_files/ paper/2024/file/40eff1670d6b08bb1bda48b0c5f30110-Paper-Conference.pdf [33] Songhua Liu、Tianwei Lin、Dongliang He、Fu Li、Ruifeng Deng、Xin Li、Errui Ding 和 Hao Wang。2021。Paint Transformer：利用笔触预测进行前馈神经绘画。载于《IEEE/CVF国际计算机视觉大会论文集》。第6598–6607页。[34] Xiao-Chang Liu、Yu-Chen Wu 和 Peter Hall。2023。利用学习得到的笔触进行绘画风格迁移。《IEEE可视化与计算机图形学汇刊》30卷9期（2023），第6309–6320页。[35] Ron Mokady、Amir Hertz、Kfir Aberman、Yael Pritch 和 Daniel Cohen-Or。2023。利用引导扩散模型进行真实图像编辑的空文本反演。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。第6038–6047页。[36] Reiichiro Nakano。2019。Neural Painters：一种用于生成笔触绘画的可学习可微约束。arXiv预印本 arXiv:1904.08410（2019）。[37] Hyelin Nam、Jaemin Kim、Dohun Lee 和 Jong Chul Ye。2024。用于连贯视频生成的光流引导提示优化。arXiv预印本 arXiv:2411.15540（2024）。

[38] Yunfang Niu、Lingxiang Wu、Yufeng Zhang、Yousong Zhu、Guibo Zhu 和 Jinqiao Wang。2024。《用于语义图像合成的多模型风格感知扩散学习》。ACM 多媒体计算、通信与应用汇刊 20，11（2024），1–21。[39] Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。2022。使用潜在扩散模型进行高分辨率图像合成。载于 IEEE/CVF计算机视觉与模式识别会议论文集。10684–10695。[40] Nataniel Ruiz、Yuanzhen Li、Varun Jampani、Yael Pritch、Michael Rubinstein 和 Kfir Aberman。2023。Dreambooth：微调文本到图像扩散模型以实现主体驱动生成。载于 IEEE/CVF计算机视觉与模式识别会议论文集。22500–22510。[41] Peter Schaldenbrand、James McCann 和 Jean Oh。2022。Frida：一款采用可微真实到仿真再到真实规划环境的协作式机器人画家。arXiv预印本 arXiv:2210.00664（2022）。[42] Peter Schaldenbrand 和 Jean Oh。2021。内容掩码损失：强化学习绘画智能体中的类人笔触规划。载于 AAAI人工智能会议论文集，第35卷。505–512。[43] Peter Schaldenbrand、Gaurav Parmar、Jun-Yan Zhu、James McCann 和 Jean Oh。2024。Cofrida：面向人机协同绘画的自监督微调。载于2024年IEEE国际机器人与自动化会议（ICRA）。IEEE，2296–2302。[44] Jaskirat Singh、Cameron Smith、Jose Echevarria 和 Liang Zheng。2022。Intelli-Paint：致力于开发更易于人类理解的绘画智能体。载于欧洲计算机视觉会议。Springer，685–701。[45] Jaskirat Singh 和 Liang Zheng。2021。结合语义引导与深度强化学习，生成达到人类水平的绘画作品。载于 IEEE/CVF计算机视觉与模式识别会议论文集。16387–16396。[46] Kihyuk Sohn、Nataniel Ruiz、Kimin Lee、Daniel Castro Chin、Irina Blok、Huiwen Chang、Jarred Barber、Lu Jiang、Glenn Entis、Yuanzhen Li 等。2023。Styledrop：以任意风格进行文生图。arXiv预印本 arXiv:2306.00983（2023）。[47] Yiren Song、Shijie Huang、Chen Yao、Xiaojun Ye、Hai Ci、Jiaming Liu、Yuxuan Zhang 和 Mike Zheng Shou。2024。ProcessPainter：从序列数据中学习绘画过程。arXiv预印本 arXiv:2406.06062（2024）。[48] Yizhe Tang、Yue Wang、Teng Hu、Ran Yi、Xin Tan、Lizhuang Ma、Yu-Kun Lai 和 Paul L Rosin。2024。AttentionPainter：用于场景绘画的高效自适应笔触预测器。arXiv预印本 arXiv:2410.16418（2024）。[49] Zhengyan Tong、Xuanhong Chen、Bingbing Ni 和 Xiaohang Wang。2021。由矢量流和灰度引导绘制过程的草图生成。载于AAAI人工智能会议论文集，第35卷。609–616。[50] Zhengyan Tong、Xiaohang Wang、Shengchao Yuan、Xuanhong Chen、Junjie Wang 和 Xiangzhong Fang。2022。Im2Oil：通过自适应采样实现线性可控精细度的基于笔触的油画渲染。载于第30届ACM国际多媒体会议论文集。1035–1046。[51] Peng Wang、Shuai Bai、Sinan Tan、Shijie Wang、Zhihao Fan、Jinze Bai、Keqin Chen、Xuejing Liu、Jialin Wang、Wenbin Ge 等。2024。Qwen2-vl：增强视觉语言模型对任意分辨率世界的感知能力。arXiv预印本 arXiv:2409.12191（2024）。[52] Qian Wang、Hong-NingH Dai、Jinghua Yang、Cai Guo、Peter Childs、Maaike Kleinsmann、Yike Guo 和 Pan Wang。2024。基于学习的人工智能艺术作品：方法学分类与质量评估。计算机调查（2024）。[53] Qian Wang、Cai Guo、Hong-Ning Dai 和 Ping Li。2023。Stroke-GAN Painter：利用笔触风格生成对抗网络学习绘制艺术作品。计算视觉媒体 9，4（2023），787–806。[54] Quan Wang、Sheng Li、Xinpeng Zhang 和 Guorui Feng。2022。用于通用风格迁移的多粒度笔触网络。ACM 多媒体计算、通信与应用汇刊（TOMM）18，4（2022），1–17。[55] Yu-Ao Wang、James She、Troy TianYu Lin 和 Kang Zhang。2025。人工智能视觉艺术史：一场拓展艺术视野的艺术运动。ACM 多媒体计算、通信与应用汇刊 21，7（2025），1–16。[56] Peizhi Yan、Rabab K Ward、Qiang Tang 和 Shan Du。2025。基于弱监督学习和单一风格模板的神经网络三维人脸形状风格化。IEEE可视化与计算机图形学汇刊（2025）。[57] Binxin Yang、Shuyang Gu、Bo Zhang、Ting Zhang、Xuejin Chen、Xiaoyan Sun、Dong Chen 和 Fang Wen。2023。基于示例绘画：使用扩散模型进行基于范例的图像编辑。载于 IEEE/CVF计算机视觉与模式识别会议论文集。18381–18391。[58] Ruofeng Yang、Bo Jiang、Cheng Chen、Ruinan Jin、Baoxiang Wang 和 Shuai Li。2024。少样本扩散模型摆脱维数灾难。载于《神经信息处理系统进展》，A. Globerson、L. Mackey、D. Belgrave、A. Fan、U. Paquet、J. Tomczak 和 C. Zhang（编），第37卷。Curran Associates, Inc.，68528–68558。https://proceedings.neurips.cc/paper_files/paper/2024/file/7eb6233e02f7d9efbb84acd839a996fb-Paper-Conference.pdf[59] Xiaoyu Yue、Zidong Wang、Zeyu Lu、Shuyang Sun、Meng Wei、Wanli Ouyang、Lei Bai 和 Luping Zhou。2024。扩散模型需要视觉先验来生成图像。arXiv预印本 arXiv:2410.08531（2024）。[60] Richard Zhang、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。2018年。深层特征作为感知度量的惊人效用。载于IEEE计算机视觉与模式识别会议论文集。586–595。[61] Zhengxia Zou、Tianyang Shi、Shuang Qiu、Yi Yuan 和 Zhenwei Shi。2021。风格化神经绘画。载于 IEEE/CVF计算机视觉与模式识别会议论文集。15689–15698。

在线附录

SmR 的公式推导

第二 Zonklar 方程的证明

扩散模型中的前向过程（例如去噪扩散概率模型）描述了数据如何随时间推移，通过逐步添加噪声而发生变化。时间步之间的转换可表示为：

$$x_t = \sqrt{\alpha} x_{t-1} + \sqrt{1-\alpha}\,\epsilon_{t-1}, \tag{19}$$

$$x_{t-1} = \sqrt{\alpha} x_{t-2} + \sqrt{1-\alpha}\,\epsilon_{t-2}, \tag{20}$$

$$\vdots \tag{21}$$

$$x_2 = \sqrt{\alpha} x_1 + \sqrt{1-\alpha}\,\epsilon_1, \tag{22}$$

$$x_1 = \sqrt{\alpha} x_0 + \sqrt{1-\alpha}\,\epsilon_0. \tag{23}$$

我们在扩散过程的不同阶段引入额外的视觉先验 x_s，使语义信息能够引导这一过程，从而让修改后的前向过程部分依赖于该视觉先验：

$$x'_t = x_t + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta} x_s - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\cdot\epsilon^*_t, \tag{24}$$

$$x'_{t-1} = x_{t-1} + \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta} x_s - \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}\cdot\epsilon^*_{t-1}, \tag{25}$$

$$\vdots \tag{26}$$

其中，$\epsilon^* \sim \mathrm{N}(0, I)$ 是独立采样的噪声，$\sqrt{\eta} \sim Uni[0, 0.5)$ 是缩放因子。x_s在每个时间步直接添加，但不会沿序列迭代演变或传播。因此，x'_t 由 x_t 推导得出，并与 x'_{t-1}存在隐式关联。

证明：序列 $X = \{x_0, x'_1, x'_2, \ldots, x_{\prime t}\}$具有马尔可夫性

要利用贝叶斯定理计算新的采样分布，$q(x'_t|x'_{t-1}, x_0)$ 和 $q(x'_t|x'_{t-1})$ 必须等价。因此，有必要证明序列 $X = \{x_0, x'_1, x'_2, \ldots\ldots, x'_t\}$具有马尔可夫性。

序列依赖关系定义如下：

$$x_{t-1} \to x'_{t-1} = f(x_{t-1}, \epsilon^*_{t-1}, x_s), \quad x_{t-1} \to x_t = g(x_{t-1}, \epsilon_{t-1}), \quad x_t \to x'_t = h(x_t, \epsilon^*_t, x_s), \tag{27}$$

x_s作为固定的视觉先验提供，并在所有步骤中共享。因此，x'_t仅依赖于 x_t，而 x_t仅依赖于 x_{t-1}；x'_{t-1}则直接由 x_{t-1}推导而来，包含了 x_{t-1}中的所有相关信息。由于 x'_t仅通过中间状态 x_t依赖于 x'_{t-1}，对较早状态（x'_{t-2}，x_{t-2}，……）的任何依赖都由 x'_{t-1}传递。每个高斯噪声项（ϵ^*_{t-1}、ϵ_{t-1}、ϵ^*_t）均独立采样。这确保了除 x'_{t-1}之外，x'_t与较早状态之间不存在隐藏依赖关系。

鉴于信息的封装性和噪声的独立性，条件依赖可简化为：

$$P(x'_t|x_0, x'_1, \ldots, x'_{t-1}) = P(x'_t|x'_{t-1}).$$

因此，序列 $X = \{x_0, x'_1, x'_2, \ldots\ldots, x'_t\}$ 满足马尔可夫性质。

新的采样分布x_t'和 x_{t-1}'可表示
如下:

$$x_t' = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}x_s + \sqrt{1-\bar{\alpha}_t}\,\bar{\epsilon}_t - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\,\epsilon_t^* \tag{28}$$

$$x_{t-1}' = \sqrt{\bar{\alpha}_{t-1}}x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}x_s + \sqrt{1-\bar{\alpha}_{t-1}}\,\bar{\epsilon}_{t-1} - \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}\,\epsilon_{t-1}^* \tag{29}$$

$$x_t' = x_t + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}x_s - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\cdot\epsilon_t^* \tag{30}$$

$$= \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}x_s + \sqrt{1-\alpha_t}\,\epsilon_{t-1} - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\epsilon_t^* \tag{31}$$

$$= \sqrt{\alpha_t}\left(x_{t-1}' - \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}x_s + \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}\epsilon_{t-1}^*\right) \tag{32}$$

$$+ \sqrt{1-\alpha_t}\,\epsilon_{t-1} - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\epsilon_t^* + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}x_s$$

$$= \sqrt{\alpha_t}x_{t-1}' + \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s \tag{33}$$

$$+ \sqrt{\alpha_t}\,\sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}\epsilon_{t-1}^* + \sqrt{1-\alpha_t}\,\epsilon_{t-1} - \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}\epsilon_t^*.$$

因此,

$$q(x_t'|x_0) = \mathcal{N}\left(x_t'; \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\,\sqrt{\eta}x_s,\ (1+\eta)\,(1-\bar{\alpha}_t)\mathbf{I}\right) \tag{34}$$

$$q(x_{t-1}'|x_0) = \mathcal{N}\left(x_{t-1}'; \sqrt{\bar{\alpha}_{t-1}}x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\sqrt{\eta}x_s,\ (1+\eta)(1-\bar{\alpha}_{t-1})\mathbf{I}\right) \tag{35}$$

$$q(x_t'|x_{t-1}') = \mathcal{N}\left(x_t';\ \sqrt{\alpha_t}x_{t-1}' + \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s,\ \left((1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)\mathbf{I}\right) \tag{36}$$

贝叶斯定理可以计算后验分布:

$$\begin{aligned}
q(x_{t-1}'|x_t', x_0) &= \frac{q(x_t'|x_{t-1}', x_0)\,q(x_{t-1}'|x_0)}{q(x_t'|x_0)} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(x_t' - \sqrt{\alpha_t}x_{t-1}' - \left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s\right)^2}{(1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t} + \frac{\left(x_{t-1}' - \sqrt{\bar{\alpha}_{t-1}}x_0 - \sqrt{1-\bar{\alpha}_{t-1}}\sqrt{\eta}x_s\right)^2}{(1+\eta)(1-\bar{\alpha}_{t-1})} - \frac{\left(x_t' - \sqrt{\bar{\alpha}_t}x_0 - \sqrt{1-\bar{\alpha}_t}\sqrt{\eta}x_s\right)^2}{(1+\eta)(1-\alpha_t)}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t x_{t-1}^2 - 2\sqrt{\alpha_t}x_t'x_{t-1}' + 2\sqrt{\alpha_t}\left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s x_{t-1}'}{(1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t} + \frac{{x_{t-1}'}^2 - 2\sqrt{\bar{\alpha}_{t-1}}x_0x_{t-1}' - 2\sqrt{1-\bar{\alpha}_{t-1}}\sqrt{\eta}x_s x_{t-1}'}{(1+\eta)(1-\bar{\alpha}_{t-1})} + \mathrm{C}(x_t', x_0)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t}{(1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t}{x_{t-1}'}^2 - 2\frac{\sqrt{\alpha_t}x_t' - \sqrt{\alpha_t}\left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s}{(1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t}x_{t-1}' + \frac{1}{(1+\eta)(1-\bar{\alpha}_{t-1})}{x_{t-1}'}^2 - 2\frac{\sqrt{\bar{\alpha}_{t-1}}x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\sqrt{\eta}x_s}{(1+\eta)(1-\bar{\alpha}_{t-1})}x_{t-1}'\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}{\left((1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)(1+\eta)(1-\bar{\alpha}_{t-1})}{x_{t-1}'}^2 - 2\left[\frac{\sqrt{\alpha_t}x_t' - \sqrt{\alpha_t}\left(\sqrt{1-\bar{\alpha}_t} - \sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_{t-1}}\right)\sqrt{\eta}x_s}{(1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\sqrt{\eta}x_s}{(1+\eta)(1-\bar{\alpha}_{t-1})}\right]x_{t-1}'\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\frac{1}{\frac{((1+\alpha_t-2\bar{\alpha}_t)\eta+1-\alpha_t)(1+\eta)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t+(1+2\alpha_t-3\bar{\alpha}_t)\eta}}\left[{x_{t-1}'}^2 - 2\frac{[\ \cdots\ ]\left((1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)(1+\eta)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}x_{t-1}'\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\frac{1}{\frac{((1+\alpha_t-2\bar{\alpha}_t)\eta+1-\alpha_t)(1+\eta)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t+(1+2\alpha_t-3\bar{\alpha}_t)\eta}}\left[{x_{t-1}'}^2 - 2\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})(1+\eta)x_t' + \sqrt{\bar{\alpha}_{t-1}}\left((1-\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)x_0 + f(\eta)}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}x_{t-1}'\right]\right\} \\
&\propto \mathcal{N}\left(x_{t-1}';\ \underbrace{\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})(1+\eta)x_t' + \sqrt{\bar{\alpha}_{t-1}}\left((1-\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)x_0 + f(\eta)}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}}_{\mu_q(x_t', x_0)},\ \underbrace{\frac{\left((1+\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)(1+\eta)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta}\mathbf{I}}_{\Sigma_q(t)}\right)
\end{aligned}$$

T函数 $f(\eta)$ 表示依赖于 η的独立项，可表示为 s:

$$f(\eta) = \sum_{k=1}^{3} A_k\sqrt{\eta}^k, \tag{37}$$

其中 A_k为系数，由 α_t、 α_{t-1}、 $\bar{\alpha}_t$、 $\bar{\alpha}_{t-1}$和 x_s等变量决定。

模型参数化

$q(x'_{t-1}|x'_t, x_0)$ 的均值为:

$$\mu_q(x'_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})(1+\eta)x'_t + \sqrt{\bar{\alpha}_{t-1}}\left((1-\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)x_0 + f(\eta)}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta} \tag{38}$$

推导　　根据前向过程 $q(x'_t|x_0)$ 的推导，我们可以对 x_0 作出如下等价解释:

$$x_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\cdot\sqrt{1-\eta}\cdot\epsilon - \sqrt{1-\bar{\alpha}_t}\cdot\sqrt{\eta}\cdot x_s}{\sqrt{\bar{\alpha}_t}} \tag{39}$$

Let

$$\tau = \sqrt{1+\eta}\epsilon + \sqrt{\eta}x_s, \tag{40}$$

真实去噪转移均值 $\mu_q(x'_t, x_0)$ 的替代参数化为:

$$\mu_q(x'_t, t) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})(1+\eta)x'_t + \sqrt{\bar{\alpha}_{t-1}}\left((1-\alpha_t-2\bar{\alpha}_t)\eta + 1 - \alpha_t\right)\frac{x'_t-\sqrt{1-\bar{\alpha}_t}\tau}{\sqrt{\bar{\alpha}_t}} + f(\eta)}{1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta} \tag{41}$$

$$= \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_t)x'_t + (C_1)x'_t - \sqrt{\bar{\alpha}_{t-1}}\sqrt{1-\bar{\alpha}_t}(1-\alpha_t)\tau + (C_2)\tau + f(\eta)}{\sqrt{\bar{\alpha}_t}\left(1-\bar{\alpha}_t + (1+2\alpha_t-3\bar{\alpha}_t)\eta\right)}, \tag{42}$$

其中 C_1 和 C_2 也是系数，每一项都乘以 η。

我们使用神经网络对$\hat{}\tau_\theta(x'_t,\ t)$进行参数化，使其能够同时预测底层噪声和先验的影响，输入为噪声图像 x'_t。于是，优化问题可简化为:

$$\mathbf{L}_\tau(\theta) = \mathbb{E}\left[||\tau - \hat{\tau}_\theta(x'_t, t)||^2\right]. \tag{43}$$

在推理过程中，令 $\eta = 0$，则:

$$\mu_q(x'_t, t) = \mu_q(x_t, t) \quad \text{in DDPM}, \tag{44}$$

$$\Sigma_q(t) = \Sigma_q(t) \quad \text{in DDPM} \tag{45}$$

我们仍可利用DDPM采样公式，从高斯噪声生成图像$\hat{}x$，而无需额外的视觉先验 x_s。

DDPM中μ_q（x_t，　t）的证明。

$$\mu_\theta(x_t, t) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\frac{x_t-\sqrt{1-\bar{\alpha}_t}\epsilon}{\sqrt{\bar{\alpha}_t}}}{1-\alpha_t} \tag{46}$$

$$= \frac{\sqrt{\alpha_t}\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon)}{\sqrt{\bar{\alpha}_t}(1-\alpha_t)} \tag{47}$$

$$= \frac{\alpha_t\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)x_t - \sqrt{\bar{\alpha}_{t-1}}\sqrt{1-\bar{\alpha}_t}(1-\alpha_t)\epsilon}{\sqrt{\bar{\alpha}_t}(1-\alpha_t)} \tag{48}$$

$$= \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_t)x_t - \sqrt{\bar{\alpha}_{t-1}}\sqrt{1-\bar{\alpha}_t}(1-\alpha_t)\epsilon}{\sqrt{\bar{\alpha}_t}(1-\alpha_t)} \tag{49}$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right) \tag{50}$$

$\mu_q(x'_t,\ t)$ 在 $\eta = 0$ 时等于方程（31）。

维度	$p_{0,x}$	$p_{0,y}$	$p_{1,x}$	$p_{1,y}$	$p_{2,x}$	$p_{2,y}$	$p_{3,x}$	$p_{3,y}$	R	G	B	不透明度	宽度
Min.	12	22	-100	-195	-84	-140	-9	22	0	0	0	0	6
Max.	268	273	450	399	465	448	267	305	255	255	255	1	106

表 7. 图像分辨率为 295×295 时，各维度笔画参数的取值范围。



图 14. 拟合得到的贝塞尔曲线，其形状与我们的笔触数据相同。

```
from scipy.special import comb def cal_ranking_loss(pred_rank, gt_order, margin=0.125): dif_gt
 = gt_order - gt_order.t() dif_pred  = pred_rank - pred_rank.t() # 所有成对组合：值为 1 的位置表示
该行对应的真实渲染顺序早于该列。 mask  = (dif_gt  < 0).float() loss_matric  =
torch.max(torch.tensor(0.0), (dif_pred -dif_gt * margin) * mask) loss  = loss_matric.sum() /
comb(pred_rank.shape[0], 2) return loss loss_r  = 0.0 param, _  = stroke_predictor(canvas1,
canvas2) # matching_idx 来自匈牙利匹配 # -1 表示取排序维度 pred_rank  = param[
matching_idx, -1]gt_order  = gt_param[valid, -1]loss_r  = cal_ranking_loss(pred_rank, gt_order)
```

代码清单 1 单个笔画序列排序损失计算的简化代码。

笔画参数拟合（正文第 3.3 节相关）

在可微光栅化器中 [29], ，我们将拟合条件设为 [路径=1、线段=1]，以确保一条三次贝塞尔曲线对应单个笔画图像。由于曲线采用随机初始化，当曲线与笔画没有重叠时，更新过程中常会出现梯度消失。为解决这一问题，我们最初按照 [61], 的提议，在像素损失优化函数中加入最优传输损失 $Loss_{OT}$，以最小化分布间的距离。然而，这种方法并未奏效。事实上，我们还在 [61] 中测试了 $Loss_{OT}$，并观察到，当笔画的初始采样点与图像中的对象（使用白底图像）没有重叠时，笔触不会朝对象移动，而是在反向传播过程中逐渐消失。

最终，我们采用了更实用的方案：检测图像中存在笔触的区域，并在这些区域初始化曲线。每个笔触的参数都经过 300 次迭代优化，以完成参数化。图 14展示了一些拟合示例。表 7列出了基于分辨率为 295×295 像素的图像所得到的各参数维度取值范围。

排序损失代码（与正文第3.4 节相关）

清单 1呈现 了用于计算画布块上一组笔触的排序损失的核心代码。批次 h
操作为 为简洁起见省略。

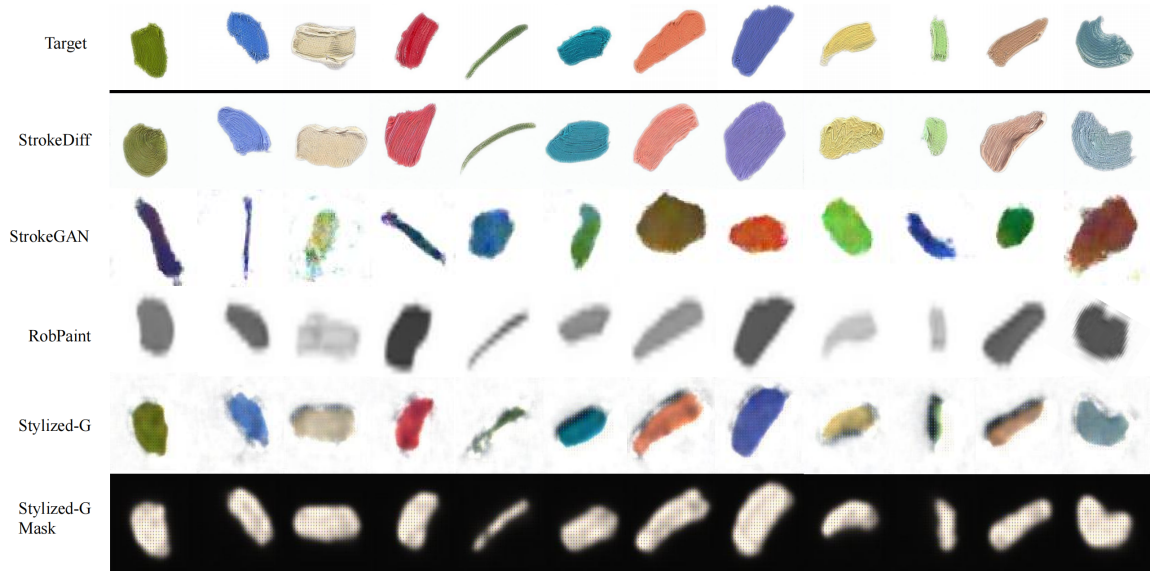


图15. 我们的笔触生成器与 StrokeGAN [53],Stylized [61]和 RobPaint 之间的比较 [1]。

笔触生成比较（与正文第4.3.2 节相关）
在图15中，我们对比了本方法的结果与其他笔触生成器生成的结果，进行了定性比较。对于 StrokeGAN，我们训练模型 1000 个训练轮次，并选取第 700 个训练轮次的最佳权重进行比较。对于 Stylized-G，我们遵循该方法的原始设置，将除 RGB 以外的笔触参数输入“着色网络”以预测蒙版，并将所有参数输入“光栅化网络”以预测前景。模型训练了 370 个训练轮次。对于 RobPaint，我们使用默认设置，输出尺寸为 $1 \times 16 \times 32$，并训练了 255 个训练轮次。

用户研究（参见正文第4.4.3 节）
绘画质量通过以下四个问题进行评估：风格（“这件艺术作品在多大程度上突出了其油画风格？”）、美感（“作为一件艺术作品，从美学角度来看，它们有多大价值？”）、纹理（“作品中的油画笔触纹理表现得如何？”）以及内容（“从照片到绘画，作品对内容的呈现效果如何？”）。参与者评估了六种方法生成的绘画，每种方法有五个示例。

收稿日期：2007年2月20日；修订日期：2009年3月12日；接受日期：2009年6月5日