arXiv:2601.17927v1 [cs.CV] 25 Jan 2026

RemEdit：基于黎曼几何的高效扩散编辑

Eashan Adhikarla
Lehigh University
美国宾夕法尼亚州伯利恒
eaa418@lehigh.edu

Brian D. Davison
Lehigh University
美国宾夕法尼亚州伯利恒
bdd3@lehigh.edu

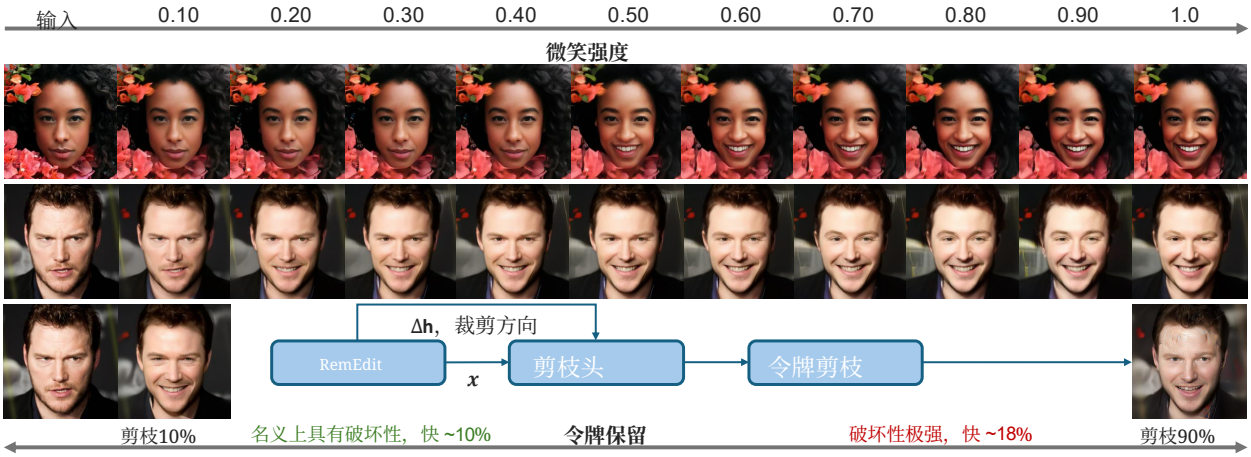

图1. RemEdit 在激进的令牌剪枝下仍能保持语义保真度；**90%** 剪枝可 **~18%** 更快，同时在视觉上仍可接受；而**10%** 剪枝几乎与未剪枝的编辑难以区分，且仍可 **~10%** 更快。

摘要

可控图像生成是现代生成式人工智能取得成功的基础，但它面临着语义保真度与推理速度之间的关键权衡。***RemEdit*** 这一基于扩散的框架通过两项协同创新应对这一权衡。首先，为了保证编辑保真度，我们将潜空间视为黎曼流形，并在其中进行探索。基于 *Mamba* 的模块能够高效学习流形结构，从而直接、准确地计算测地线路径，实现平滑的语义编辑。双 *SLERP* 混合技术以及由视觉语言模型执行的目标感知提示词增强流程，进一步优化了这种控制。其次，为了进一步加速，我们提出了一种新颖的任务特定注意力剪枝机制。轻量级剪枝头能够学习保留对编辑至关重要的令牌，从而实现有效优化，同时避免内容无关方法中常见的语义退化。*RemEdit* 超越了此前最先进水平的编辑框架，并在 *50%* 剪枝率下仍保持实时性能。因此，*RemEdit* 为实用且强大的图像编辑树立了新的标杆。源代码：*github.com/eashanadhikarla/RemEdit*。

1. 引言

扩散模型的图像合成能力已将生成式人工智能研究的重点转向一个更具挑战性的前沿：精准且高效的图像编辑。尽管扩散模型 [13, 41] 已经展现出SOTA级的合成质量，但要引导其生成过程完成精准、由用户定义的编辑，仍然是一项挑战。早期方法主要通过高成本的优化或微调 [19], 来实现文本引导的操控，而另一些方法则利用反演技术，使真实图像能够被编辑 [26]。一个关键突破是发现了 U-Net 瓶颈的特征空间，即所谓的 h空间；研究表明，该空间具有极高的稳定性和丰富的语义信息 [21]。与此同时，类似 [14, 17] 的方法开始探索这些模型的内部机制，通过编辑交叉注意力图实现更细粒度的控制。然而，要实现既保持高保真度、又尊重图像原始身份的编辑，往往需要复杂的调参。此外，迭代去噪过程所需的高昂计算成本，也推动了另一条加速研究路线的发展，从更快的采样调度 [13, 23] 到早期的模型剪枝尝试 [7]。这迫使实践者在功能强大但速度缓慢的方法与更快速的

但可靠性较低的替代方案之间做出选择。计算瓶颈在U-Net 的自注意力层中尤为严重，如图 4所示，在单次前向传播中，这些层消耗的 GFLOPs 可能超过总量的 80%。虽然这促使研究者将这些层作为加速目标，但我们认为，现有的剪枝方法从根本上与高保真图像编辑的目标并不匹配。

为解决这一权衡问题，我们提出了*RemEdit*，一种同时优化几何保真度与计算效率的新框架。据我们所知，*RemEdit* 是首个在黎曼扩散潜空间编辑中同时应对这两项挑战的方法。我们的工作建立在三项核心贡献之上：

1. **精准的测地线导航空间**。 我们提出了一种新颖且高效的方法，通过直接学习局部曲率来遍历语义流形。一个轻量级的基于Mamba的预测器能够高效估计克里斯托费尔符号，进而定义可学习的指数映射。该架构使我们能够求解测地线常微分方程，从而实现直接、平滑且几何上忠实的编辑。

2.**高保真语义控制**。 我们通过两项关键创新实现了更出色的控制与融合效果。双重 SLERP 机制在流形感知的潜空间中提供了有理论依据且无伪影的插值。该机制由目标感知的提示增强策略引导，该策略仅需通过视觉语言模型（Qwen2-VL）进行一次处理，即可生成细致且具上下文感知能力的语义方向。

3. **任务特定加速**。 我们提出了一种新颖的任务感知注意力剪枝方法，在不牺牲编辑质量的情况下大幅加速推理。一个由编辑的语义目标调节的轻量级“PruningHead”能够学习保留最相关的词元。这使得我们可以积极加速无关区域，同时维持编辑操作的完整性。

这些贡献的协同作用促成了一个比以往工作更强大且更实用的框架，弥合了理论稳健性与现实可用性之间的差距。关于推动本研究的关键挑战，更多讨论请参见**附录A**。

2. 相关工作

2.1. h-空间中的语义潜变量结构

Kwon等人首次将 U-Net 瓶颈（h-空间）形式化，并证明简单的偏移量无需重新训练即可实现可靠编辑 [21]。后续研究探索了组合性（InjectFusion） [16], PCA方向/语义轴 [10, 50],三维一致性 [10]，以及用于保持风格保真度的 h-空间反演 [15]。综述总结了这一趋势：转向可解释的 h空间控制 [30]。我们在此基础上展开研究，但认为通常归因于线性的稳定性其实更好……

这可以通过扩散动力学内在的黎曼结构来解释。

近期工作 [28, 46] 强调， h-空间并非仅仅是一种暂时性编码，而是一种结构化且语义对齐的表示。我们的工作以此为基础，提出其稳定性和线性源于扩散动力学所固有的黎曼几何。

2.2. 测地线与流形感知编辑

流形假设贯穿生成建模，从经典的 Isomap [43]/LLE [34] 到生成对抗网络的潜在空间几何 [8,11, 34, 38, 43]。近期研究通过自编码潜变量 [24, 32], 、评分雅可比度量 [1, 29, 35] 和最短路径 [30] 将这一观点应用于扩散模型（更多讨论见**附录 A**）。与之不同，我们学习 h 空间中的连接，并引入指数映射（基于 Mamba），以高效地沿测地线前进，同时避免雅可比估计并保持 O(N) 复杂度。

2.3. 语义引导与控制

无需训练的文生图编辑器通过注意力控制或隐式掩码来定位变化。Prompt-to-Prompt（P2P）通过复用交叉注意力图 [12] 来约束词级变化，MasaCtrl [3]则结合了快速反演与隐式注意力掩码。LEDITS++ [2] 将二者纳入比较，并为其设置相同的提示、反演深度和随机种子。结果表明，在运行时间相当或更短的情况下，RemEdit 的局部性和身份保持能力更强。这些方法在仅靠文本即可实现编辑时表现出色，但对措辞敏感，也缺少明确的几何约束。我们的方法与之互补：通过双重球面线性插值，沿学习到的 h 空间测地线约束编辑，从而平衡编辑强度与身份保持。基于现代文生图流程构建的零样本编辑器还能进一步降低开销：空文本反演（基于优化的反演） [26], 、负提示反演（单次反演） [25], 、采用修改后采样器的无反演编辑（InfEdit），无需显式反演 [45], ，以及通过引导式牛顿步实现的极速反演 [37]。这些方法兼具出色的身份保持能力与速度，但仍依赖提示。RemEdit 与之互补：它在无条件 h 空间中运行，借助学习到的连接和双重球面线性插值，即使在提示相同的情况下，也能从设计上确保身份保持与局部性。

2.4. 扩散模型加速

采样级加速方法（DDIM/DPM-Solver/蒸馏）减少步数 [13, 23, 36], ，而令牌级方法则通过免训练（SiTo [49]/AT-EDM [44]）、混合（DaTo [48]/CAT [5]）或合并（ToFu [20]）策略剪枝注意力。基于训练的方法（Dy-DiT [51]/LD-Pruner [4]）会重构模块。EffDiff [42] 等其他工作则着重于实时操控。编辑

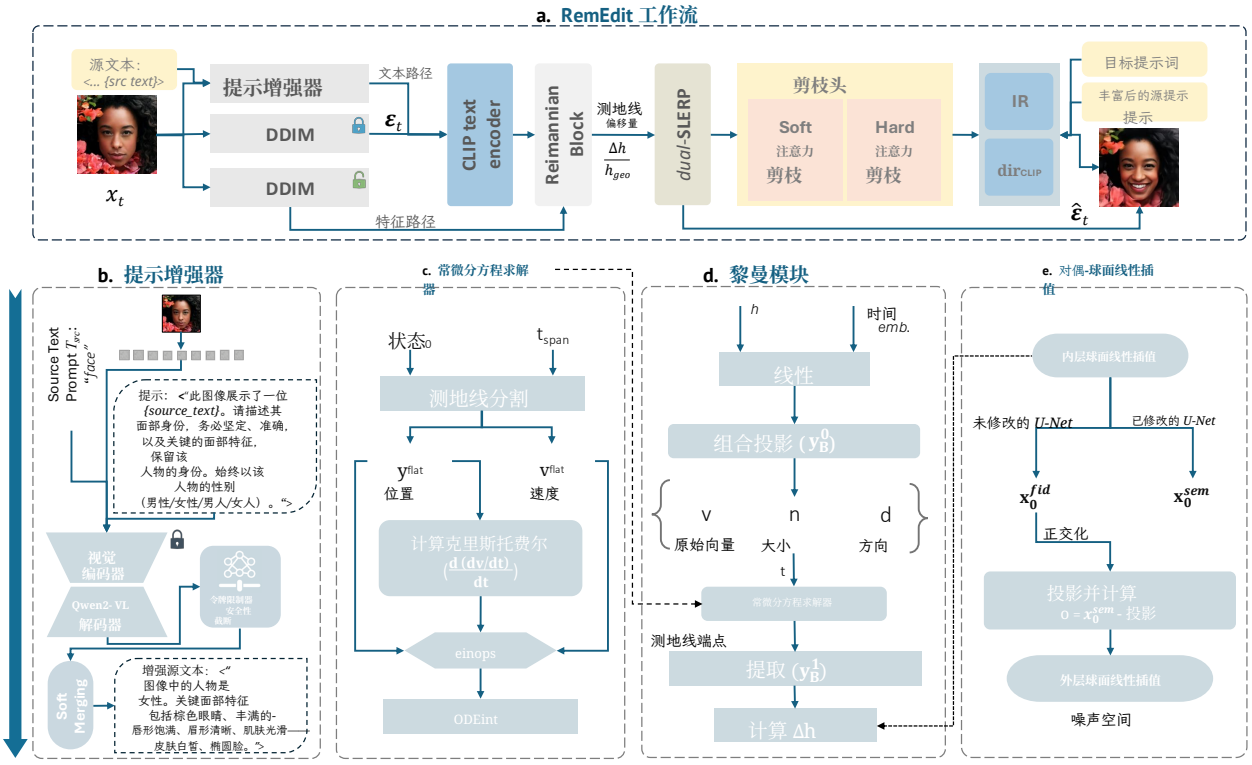

图 2。**RemEdit 架构图**。 我们的扩散编辑流水线概览，集成了用于测地线导航的指数映射、用于保真度控制的双重球面线性插值，以及用于提示词增强的 Qwen2-VL。各个模块自上而下流动。

与生成相比，编辑对局部性有更严格的要求；内容无关剪枝可能会损害未编辑区域中的身份。尽管上述方法颇具巧思，但其设计和评估主要面向文本到图像任务。语义编辑任务施加了严格得多的约束：必须在图像的所有未编辑区域中保持语义一致性。内容无关剪枝器本质上无法区分全局冗余的令牌（例如一片蓝天）与对维持未编辑物体身份至关重要的局部令牌。剪除后者可能导致源内容无法接受的劣化。我们提出了一种针对编辑局部性设计的任务感知剪枝器。

3. 方法

3.1. 黎曼编辑框架

我们的完整架构如图2所示。

语义编辑中的一个核心挑战是，潜空间中的线性操作 [21] 往往无法保持图像的真实感。H 空间方法通常应用一个学习得到的线性偏移量：$h' = h + \Delta h_{\text{linear}}$。我们的结果表明，测地线更新能够改善方向对齐和分割一致性，仅带来少量计算开销，而这一开销随后可通过剪枝收回。向潜代码添加向量可能会将其推离自然图像流形[1]，，从而产生伪影。这引出了一个关键问题：

> 如何在尊重数据流形的同时进行语义编辑，而不增加过高的计算成本？

虽然先前的方法承认这种几何结构，但通常依赖事后计算，例如根据得分函数雅可比矩阵估算度量，或求解计算开销巨大的优化问题 [29, 35]。

我们通过常微分方程求解器估算完整度量，以捕捉测地线的曲率和端点，同时直接学习流形的联络。[2] 这使我们能够借助指数映射这一数学工具，计算测地线——曲面上尽可能笔直的路径。

在黎曼几何中，指数映射记作$exp_p(v)$，它将流形上的一个点 p 和该点处的切向量v 映射到流形上的一个新点：沿着从 p出发、朝v 方向的测地线行进单位时间。该测地线路径 $\gamma(t)$由测地线方程决定，而测地线方程取决于流形的

[1]线性偏移量往往会脱离流形，导致编辑结果不再保留主体的身份。[2]向量沿曲线传输时所遵循的规则。

克里斯托费尔符号 Γ_{ij}^k:

$$\frac{d^2\gamma^k}{dt^2} + \mathbf{\Gamma}_{ij}^k \frac{d\gamma^i}{dt}\frac{d\gamma^j}{dt} = 0 \tag{1}$$

关键在于，通过学习克里斯托费尔符号，我们可以求解这个二阶常微分方程，从而高效地计算指数映射。如我们的架构所示（见附录，图 11），黎曼模块的实现步骤如下：**(a) 初始化：** 通过结合输入 y_0 空间特征 h特征 h 与时间嵌入，在流形上构造初始点。**(b) 切向量预测：** 基于 y_0，我们使用线性层和一个平滑且保持范数的基于 tanh 的回缩，预测初始切向速度向量 v_0。**(c) 常微分方程积分：** 我们使用轻量级 Mamba 网络学习克里斯托费尔符号 Γ。随后，将测地线方程（式 1）构造成一阶系统，并使用高精度自适应常微分方程求解器从 $t = 0$求解至 $t = 1$。**(d) 偏移量提取：** 常微分方程在 $t = 1$处的解给出测地线的终点，$\gamma(1) = exp_h(v_0)$。随后，我们将沿此路径的位移定义为几何上合理的编辑向量:

$$\Delta h = \exp_h(v_0) - h \tag{2}$$

这使我们能够以有原则且流形上的方式，构造 h-空间更新，其中最终编辑后的特征就是测地线终点本身:

$$h'_{\text{RemEdit}} = h + \Delta h_{\text{geo}} = \exp_h(v_0) \tag{3}$$

指数映射模块和常微分方程求解器配置的详细说明见**附录 C**。

3.2. 使用双重球面线性插值进行高保真控制

测地线偏移量 Δh （第 3.1节）为编辑提供了稳健且具备几何感知能力的方向。然而，直接将其应用为 $h' = h + \Delta h$ 无法控制编辑强度，并可能压制原始图像的身份特征。这引出了下一个关键问题:

> 我们如何缩放测地线偏移 Δh，以控制编辑强度而不引入伪影?

我们假设扩散模型中的潜特征近似服从超球面高斯分布。因此，插值应沿球面进行，以保持在数据流形上并保留统计特性。线性插值（LERP）无法保持潜向量的范数，可能会“脱离”超球面，导致质量下降。球面线性插值（SLERP） [39, 40], 沿两点之间的大圆弧移动，是适用于该空间的自然选择。我们将这一思路用于一种新颖的双重 SLERP 机制，如图3所示，可实现两种不同层级的控制。

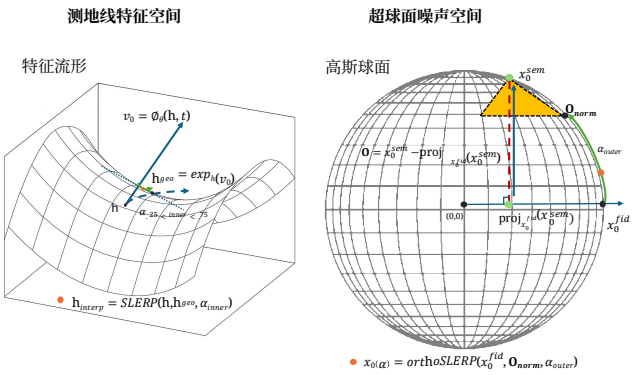

图 3. 双重球面线性插值的两阶段插值： **[左图]** 在黎曼特征流形上进行内层球面线性插值，通过插值参数 α_{inner}混合原始特征 h 与测地线偏移特征 $\exp_h(v_0)$。**[右图]** 在超球面噪声潜空间上进行外层正交球面线性插值，将语义预测 x_0^{sem}投影到保真潜变量 x_0^{fid}上，提取正交分量 o，并使用 α_{outer}在 x_0^{fid}与 o 之间进行插值，从而将属性编辑与身份保持解耦。

内层球面线性插值（调节编辑强度）。如图3左侧所示，融合策略的第一阶段在特征空间内进行。通过黎曼框架（第3.1节）计算出完整的测地线偏移量 Δh 后，我们在原始特征映射 h 与完全编辑后的特征映射 $h_{geo} = h + \Delta h$ 之间进行球面线性插值。该插值定义如下： $h' = SLERP(h, h_{geo}, \alpha_{inner})$。标量 $\alpha_{inner} \in [0, 1]$ 可作为精确的调节旋钮，[3] 用于控制特征空间中的语义编辑强度，然后再将其传递至 U-Net 的后续层。

外层 SLERP（保持全局保真度）。第二阶段发生在去噪步骤结束时，用于形成干净图像的最终预测， x_0。我们的框架自然地产生两种预测：来自原始 U-Net 路径的保真潜向量， x_0^{fid}，以及来自编辑路径的语义编辑结果， x_0^{sem}。为了在保持原始图像身份的同时稳健地融合二者，我们在噪声空间中使用正交化的 SLERP（如图3所示）。首先，我们将语义潜变量投影到保真潜变量上，以找出编辑中与原始身份正交的分量: $o = x_0^{sem} - proj_{x_0^{fid}}(x_0^{sem})$。随后，我们在保真预测与正交分量 o之间进行插值:

$$x_0(\alpha_{outer}) = \Psi(x_0^{fid}, o, \alpha_{outer}) \tag{4}$$

关于插值方法更广泛的概念性比较，请参见**附录 E**。

3.3. 目标感知提示词增强

先前方法面临的一个核心挑战是文本引导编辑存在歧义，因为简短的提示往往缺乏足够的特异性，无法避免无关属性发生变化。

[3]将编辑方向与其大小解耦（由球面线性插值控制）

例如，像化妆这样的属性，简单的源图像—目标图像对（如“面部”→“化妆后的面部”）提供的上下文并不足够。为了满足这一最低限度的约束，模型可以自由探索数据分布中的邻近流形，从而可能引入非预期的变化，例如改变对主体性别的感知。这就需要为编辑方向提供更丰富、针对具体实例的上下文，同时避免显著增加计算开销，也无需为图像对重新标注文本。我们将此称为“天下没有免费的午餐”问题，并通过引入一个轻量级、单次运行的提示词增强阶段来解决，该阶段使用预训练的视觉语言模型（Qwen2-VL）。我们不再对所有图像使用通用源文本，而是首先为特定源图像生成详细描述 x_0（详情见附录图12）。这种方法有效地将编辑锚定在输入图像的具体视觉上下文中（见图6），通过缩小模型的探索自由度来减少不必要的语义偏移。更多详情见**附录 D**。

3.4. 任务特定注意力剪枝

迄今为止，我们的框架实现了高保真、可控的编辑。然而，与所有现代扩散模型一样，它仍然计算开销巨大。自注意力模块是主要瓶颈，会消耗大量资源：在 256×256 分辨率下，其 GFLOPs 超过卷积和 ResNet 模块的 $4\times$ 倍；在 512×512 分辨率下则超过 $8\times$ 倍。因此，这些模块非常适合进行优化。

尽管令牌剪枝已成为一种可行的加速策略，但现有方法是为通用图像生成而设计的。这引出了一个关键问题：

> 我们如何才能剪除计算开销巨大的令牌，同时不违背图像编辑的核心约束：保留未编辑区域的语义内容？

对于纯生成任务，轻微的不一致或伪影或许可以接受；但对于编辑任务，它们就意味着失败。如果在面部上编辑出“微笑”却改变了主体的头发或背景，那么这次编辑就失败了。这意味着，用于编辑的剪枝器不能与内容无关；它必须明确感知编辑任务。成功的语义编辑剪枝策略必须以……为条件编辑向量本身。通过向剪枝机制提供有关所需语义变化的信息，它可以学会区分全局冗余的令牌与对于保留未编辑区域身份至关重要的令牌，即使这些区域的方差很低。

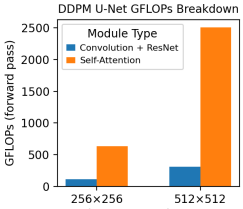

图 4. 256^2 和 512^2 下 DDPM U-Net 的 GFLOPs 分解。

我们提出了一种新颖的**任务感知注意力剪枝**机制。我们不依赖静态启发式方法，而是训练一个轻量级神经网络，根据输入特征和编辑的语义目标动态预测令牌重要性。

训练目标是两个相互竞争的损失的加权和，旨在平衡输出质量与计算开销：$\mathbf{L} = \mathbf{L}_{\text{fidelity}} + \lambda_{\text{sparsity}} \mathbf{L}_{\text{sparsity}}$。首先，保真度损失通过要求剪枝器的输出与原始未剪枝模型的输出一致，教会它应保留哪些内容。其次，稀疏性损失通过促使模型尽可能多地剪除令牌来提高效率。在推理阶段，训练好的剪枝器 P_θ 会执行不可微的硬剪枝。

给定输入特征 $X \in \mathbb{R}^{B\times C\times H\times W}$ 和定义任务的语义向量 $\mathbf{d}_{\text{edit}} \in \mathbb{R}^{D_{\text{clip}}}$，我们将 X 重塑为令牌序列 $T \in \mathbb{R}^{B\times N\times C}$，其中 $N = H \cdot W$。一个学习得到的剪枝函数 P_θ 会估计令牌重要性：

$$S = \mathcal{P}_\theta(T, \mathbf{d}_{\text{edit}}) \in [0,1]^{B\times N}$$

选取 Top-k 个索引，其中 $k = \lfloor N \cdot (1-\rho) \rfloor$，并且仅在保留令牌上计算注意力：

$$A = \text{Softmax}\left(\frac{Q_{\text{kept}} K_{\text{kept}}^{\top}}{\sqrt{C}}\right) V_{\text{kept}}$$

注意力输出会恢复至完整令牌空间并进行投影，从而得到最终输出 X_{out}：

$$X_{\text{out}} = X + W_o(\text{Scatter}(A, \mathcal{I}_{\text{keep}}))$$

更多关于剪枝的细节请参见**附录 B** 。

4. 实验

我们开展了一系列全面的实验，以验证*RemEdit*：其准确性、语义控制机制的有效性，以及任务感知剪枝器的效率。我们在多个具有挑战性的基准测试上评估了我们的方法，包括 CelebA-HQ [18]，、LSUN-Church [47]，和 AFHQ-Dog [6]，每项均涵盖 256×256。

4.1. 实现细节

我们的框架以高效的少样本方式进行训练。与 Asyrp 等基线方法不同，[21]后者需要对 1000+图像进行多轮迭代，RemEdit 仅需使用 500 张图像进行双样本设置训练，即可取得更优效果（n=2）。我们发现，单次训练迭代便足以让黎曼模块捕捉编辑的高层语义；不过，训练约三轮可显著改善输出，去除眉毛和头发纹理等细节区域中的轻微伪影。除非另有说明，所有实验均使用 DDIM 架构处理 256×256 分辨率的图像。我们报告了整个流程端到端推理的实际耗时。

表1。在 CelebA-HQ数据集上对基于扩散的图像编辑方法进行定量比较。结果展示了三种面部属性编辑任务（微笑、悲伤、晒黑）中的 (S_{dir})、分割一致性和推理时间。

方法	微笑		Sad		晒黑		推理时间
	S_{dir} ↑	分割一致性（%）↑	S_{dir} ↑	分割一致性（%）↑	S_{dir} ↑	分割一致性（%）↑	（秒）↓
StyleCLIP [31]	0.130	86.80	0.149	85.50	0.152	84.30	8.5
StyleGAN-NADA [9]	0.160	89.40	0.161	87.70	0.166	88.50	12.3
Diffusion-CLIP [19]	0.170	**93.70**	0.163	**89.93**	0.174	**92.85**	45.2
BoundaryDiffusion [53]	0.170	90.40	0.166	89.02	0.177	85.71	38.7
Asyrp [21]	0.190	87.90	0.159	88.90	0.177	89.31	28.9
Prompt-to-prompt（p2p） [12]	0.165	85.20	0.152	84.10	0.158	86.30	145.0
LEdits++ [2]	0.182	89.70	0.169	87.80	0.175	88.90	20.1
RemEdit	**0.1982**	92.41	**0.1792**	89.72	**0.1948**	92.18	**2.8**

表 2. 在**LSUN-Church** 数据集上的定量评估。

方法	百货商店		古代		红砖	
	S_{dir} ↑	分割一致性（%）↑	S_{dir} ↑	分割一致性（%）↑	S_{dir} ↑	分割一致性（%）↑
扩散 CLIP [19]	0.1300	54.50	0.1976	**64.82**	0.2085	65.83
BoundaryDiffusion [53]	0.1866	56.72	0.2034	60.13	0.2112	66.10
Asyrp [21]	0.1932	57.62	0.2087	62.65	0.2170	67.42
Prompt-to-prompt（p2p） [12]	0.1820	53.40	0.1890	58.75	0.1985	61.30
LEdits++ [2]	0.1895	58.85	0.2045	59.20	0.2130	64.95
RemEdit	**0.1959**	**59.03**	**0.2193**	63.91	**0.2200**	**68.20**



图5. RemEdit 在 CelebA-HQ数据集上的编辑结果，其中属性 {“悲伤”、“微笑”、“化妆”、“年轻”}属于人类分布内， {“僵尸”、“皮克斯”}属于人类分布外。

表 3. *RemEdit* 在不同反演深度 t_0、前向 DDIM 步数 S_{for}和生成细化步数 S_{gen}下的重建保真度。每个单元格报告平均绝对误差 / LPIPS / 结构相似性。平均绝对误差/LPIPS 越低、结构相似性越高，表示保真度越好。

	S_{gen} = 6	S_{gen} = 40	S_{gen} = 500
t_0 = 300			
S_{for} = 6	0.047 / 0.185 / 0.732	0.061 / 0.221 / 0.704	0.063 / 0.224 / 0.694
S_{for} = 40	0.027 / 0.110 / 0.863	0.023 / 0.091 / 0.891	0.023 / 0.086 / 0.895
S_{for} = 500	0.024 / 0.095 / 0.885	0.020 / 0.073 / 0.914	0.019 / 0.065 / 0.923
t_0 = 450			
S_{for} = 6	0.055 / 0.208 / 0.673	0.073 / 0.255 / 0.655	0.077 / 0.260 / 0.643
S_{for} = 40	0.031 / 0.128 / 0.827	0.025 / 0.100 / 0.880	0.024 / 0.093 / 0.885
S_{for} = 500	0.028 / 0.108 / 0.862	0.024 / 0.076 / 0.910	0.020 / 0.068 / 0.919
t_0 = 600			
S_{for} = 6	0.084 / 0.283 / 0.501	0.101 / 0.325 / 0.564	0.106 / 0.330 / 0.552
S_{for} = 40	0.047 / 0.175 / 0.706	0.029 / 0.120 / 0.852	0.028 / 0.108 / 0.862
S_{for} = 500	0.041 / 0.147 / 0.778	0.024 / 0.087 / 0.893	0.022 / 0.076 / 0.907

表4. RemEdit 核心组件的消融研究。结果表明，每个组件都能带来独特的益处。**时间（秒）为端到端测量，包括视觉语言模型图像描述、反演、测地线常微分方程求解和解码。**

方法配置	**方向 CLIP** ↑	**分割一致性（%）** ↑	**时间（秒）** ↓
欧几里得偏移（h空间基线）	0.190	87.9	**1.82**
+ 测地线导航（我们的方法）	0.196	89.5	2.76
线性插值（无球面线性插值）	0.188	86.7	2.74
仅内层球面线性插值	0.191	88.2	2.78
仅外层球面线性插值	0.192	88.4	2.77
双重球面线性插值混合（我们的方法，未剪枝）	**0.198**	**92.4**	2.89
完整 RemEdit + 剪枝（$\rho = 0.2$）	0.192	90.1	2.38
完整 RemEdit + 剪枝（$\rho = 0.5$）	0.184	89.5	2.31

4.2. 几何语义控制分析

4.2.1. 定量分析

我们首先评估 RemEdit 的核心编辑能力，着重展示测地线导航与双 SLERP 混合之间的协同作用。遵循近期最先进方法 [19, 21, 53],

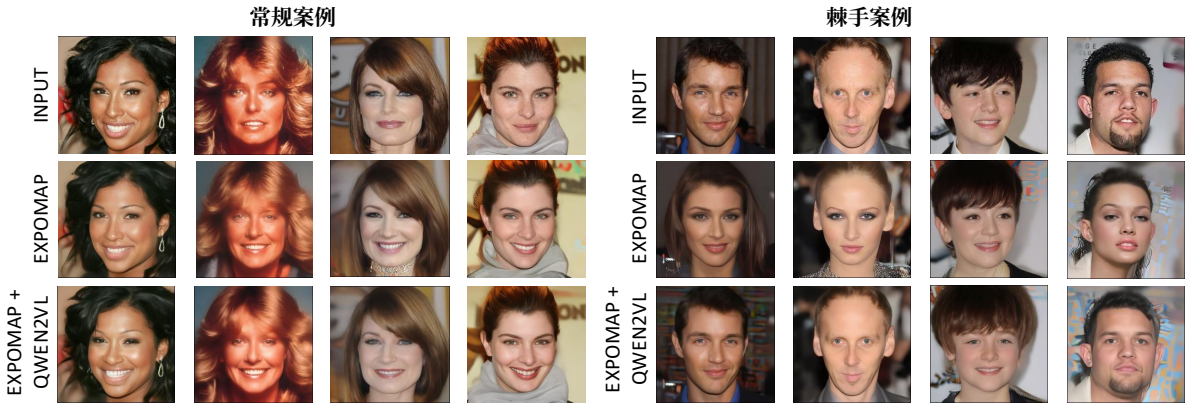

图6。通过“化妆 ”任务展示了使用 Qwen2 VL 进行细粒度文本注入如何纠正部分失败案例 s.

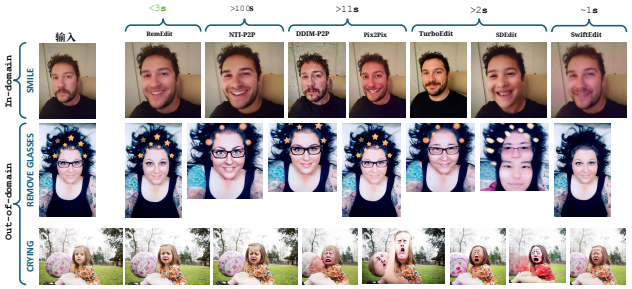

图7。RealEdit 基准数据集上的零样本定性比较。提示、反演深度和随机种子均相同。RemEdit 在保留身份的同时，以极小的附带变化实现了稳定的属性迁移。

表 1 和 2比较了 RemEdit 在 CelebA-HQ 和 LSUN-Church 上的表现，采用方向性 CLIP 相似度（S_{dir}） [33] 衡量语义对齐，采用分割一致性 [22] 衡量身份保持。我们为每种任务属性选取了 250 个样本。RemEdit 的表现始终优于以往方法（包括 Asyrp），在“微笑”“悲伤”和“晒黑”等属性上均是如此。这凸显了我们指数映射的有效性：通过求解测地线常微分方程，黎曼模块可确保编辑在语义上精确，而双重球面线性插值则能保持身份。

4.2.2. 定性分析

图 5 展示了各属性的定性结果。RemEdit 成功操控了从简单表达（“悲伤”、“微笑”）到更复杂变换（“僵尸”、“皮克斯”）的属性，同时保持了高度的照片级真实感和身份一致性。图 10 将其与 Asyrp 基线方法进行了直接比较。即使仅运行 40 个去噪步骤，RemEdit 生成的结果在定性上也优于运行 1000 步的 Asyrp，并成功编辑了基线方法难以处理的复杂案例。这凸显了我们几何方法的高效性和稳健性。如图 6所示，Qwen2-VL 能够提升描述不充分的提示的清晰度，而常规案例及其底层几何结构仍保持稳定。

4.2.3. 消融研究

重建质量。 表3 探究了模型在不同反演和生成设置下的重建保真度。我们分析了反演深度（t_0）、前向 DDIM 步数（S_{for}）以及生成步数（S_{gen}）的影响。采用平均绝对误差、LPIPS 和结构相似性进行衡量，结果表明，较深的反演（更高的（t_0）以及充足的生成步数对于实现高保真度重建至关重要，这证实了我们的方法符合基于扩散的反演流程的预期行为。

可控图像生成。我们还对各项贡献分别进行了消融研究，结果见表4。我们从一个实现 Asyrp 风格 h 空间编辑的强基线开始，然后逐步加入所提出的模块：先是测地线导航，再是双 SLERP 混合，构成完整且未经剪枝的 RemEdit 模型，最后是在两种不同剪枝率下进行任务专用剪枝（$\rho = 0.2$ 和 $\rho = 0.5$）。我们评估每种配置的方向性 CLIP 相似度 S_{dir}、分割一致性 SC以及每张图像的推理时间。

4.3. 任务特定加速分析

如图 1所示，RemEdit 即使在激进剪枝下仍能保持保真度。在 10% 的剪枝比例下，编辑后的输 ~18出在视觉上与未剪枝版本几乎无法区分，同时还能提升 10% 的速度。即使在激进的 90% 剪枝比例下，所带来的8% 速度提升，编辑结果仍然在语义上正确且在视觉上可接受。这种性能下降是我们任务感知方法的直接结果。令牌保留模式如图 8所示。重要性图清晰地表明，模型学会了聚焦于语义相关区域（例如，“微笑”对应嘴部，“睡眠”对应整个面部），同时正确地将背景识别为可剪枝区域。

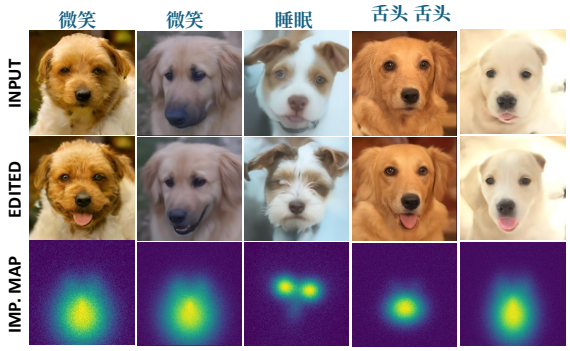

图8. 每张图像的令牌重要性图热力图，用于可视化剪枝头关注的内容。



图9. 我们还通过多属性编辑展示了 RemEdit 的语义控制能力。

5. 结果与讨论

我们的实验结果验证了 RemEdit框架 的有效性。表 1 和2 中的定量比较表明，在语义对齐和身份保持方面，我们的方法明显优于现有方法。

表 4 中的消融研究更深入地揭示了其中的权衡。我们提出的测地线方法和双SLERP混合模块显著提升了 S_{dir} 分数（从0.190 提升至 0.198）和分割一致性（从 87.9%提升至 92.4%），证实了这些方法在提升编辑质量和保真度方面的作用。然而，这一提升是以推理时间增加为代价的：由于常微分方程求解计算，推理时间从基线的 1.82 秒增加到未剪枝的完整 RemEdit 的 2.89 秒。

这正是任务专用剪枝发挥作用的地方。仅剪除 20% 的令牌，我们便将推理时间缩短至 $2.38s$，同时使各项性能指标保持接近完整模型。即使剪枝比例高达 50%，模型的质量仍与基线相当（$S_{dir} = 0.184$，而基线为 0.190），且速度显著快于未剪枝版本。这证实了我们的剪枝方法成功提升了几何编辑框架的实用性，在保留明显编辑质量优势的同时，缩小了与更简单方法之间的性能差距。

此外，图中令牌重要性图8 通过聚焦相关区域（嘴部、舌头、面部）并为背景分配较低权重，证实了我们的任务感知剪枝策略

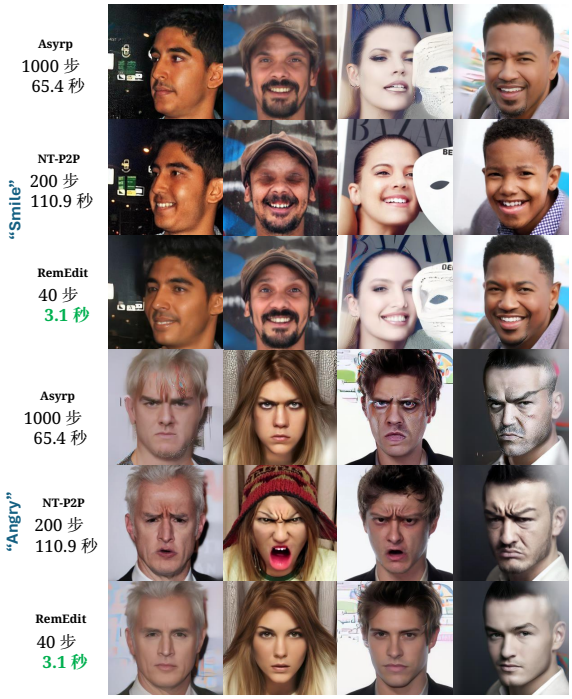

图10. 仅用40步即可解决棘手案例。

；剪枝比例为 20–50% 时，质量保持稳定。图9 展示了多属性示例（晒黑+微笑、化妆+微笑），未观察到干扰或身份变化。对于图10所示的困难案例，RemEdit 仅需 ˜40 步（˜3.1 秒），Asyrp 需 ˜1000 (˜65.4 秒），NT-P2P 需 ˜200 (˜110.9 秒）；各面板展示了身份和局部性的差异，但推理速度有所降低。

6. 结论

在本研究中，我们提出了 RemEdit，这是一种新颖的框架，可实现高保真、高效率的可控图像生成。我们从三个方面着手，解决准确性与速度之间的根本性权衡。首先，我们提出了一种导航语义 h-空间的新方法，将其建模为黎曼流形并求解测地线路径；我们证明，这种方法能够提升语义质量。其次，我们引入了一种基于双重球面线性插值和 Qwen2-VL 的文本注入混合机制，通过文本和图像特征对编辑强度和身份保持进行细粒度控制，进一步提升保真度。最后，我们开发了一种任务感知的注意力剪枝方法，使这一几何方法在计算上切实可行，在保留编辑质量的同时显著加快推理速度。

尽管 RemEdit 展现出较高的语义保真度和实用效率，但它依赖于 Qwen2-VL 和 CLIP 等预训练模型。未来的工作可以探索针对分布外任务的、经过学习的领域特定引导。

参考文献

[1] Simone Azeglio 和 Arianna Di Bernardo。扩散模型内部是什么？用于探索数据流形的基于得分的黎曼度量。arXiv 预印本 arXiv:2505 11128，2025 2。。[2] Manuel Brack、Felix Friedrich、Katharia Kornmeier、Linoy Tsaban、Patrick Schramowski、Kristian Kersting 和 Apolin´ario Passos。Ledits++：使用文本到图像模型进行无限制图像编辑。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8861–8870 页，2024 年。2、6[3] Mingdeng Cao、Xintao Wang、Zhongang Qi、Ying Shan、Xi- aohu Qie 和 Yinqiang Zheng。MasaCtrl：无需调参的互注意力控制，用于一致的图像合成与编辑。载于 IEEE/CVF计算机视觉国际会议论文集，第 22560–22570 页，2023 年。2[4] Thibault Castells、Hyoung-Kyu Song、Bo-Kyeong Kim 和 Shinkook Choi。LD-Pruner：利用与任务无关的洞见对潜扩散模型进行高效剪枝。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 821–830 页，2024 年。2[5] Xinle Cheng、Zhuoming Chen 和 Zhihao Jia。CAT 剪枝：面向文本到图像扩散模型的聚类感知令牌剪枝。arXiv 预印本 arXiv:2502.00433，2025 年。2[6] Yunjey Choi、Youngjung Uh、Jaejun Yoo 和 Jung-Woo Ha。StarGAN v2：面向多个领域的多样化图像合成。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8188–8197 页，2020 年。5[7] Gongfan Fang、Xinyin Ma 和 Xinchao Wang。扩散模型的结构化剪枝。载于神经信息处理系统进展，2023 年。1[8] Charles Fefferman、Sanjoy Mitter 和 Hariharan Narayanan。检验流形假设。《美国数学学会杂志》，29(4)：983–1049，2016 年。2[9] Rinon Gal、Or Patashnik、Haggai Maron、Amit H Bermano、Gal Chechik 和 Daniel Cohen-Or。StyleGAN-NADA：利用 CLIP 引导图像生成器的领域自适应。《ACM 图形学汇刊》（TOG），41(4)：1–13，2022 年。6[10] Ren´e Haas、Inbar Huberman-Spiegelglas、Rotem Mulayoff、Stella Graßhof、Sami S Brandt 和 Tomer Michaeli。发现扩散模型语义潜空间中的可解释方向。载于 2024 IEEE 第 18 届自动面部与手势识别国际会议（FG），第 1–9 页。IEEE，2024 年。2[11] Erik H¨ark¨onen、Aaron Hertzmann、Jaakko Lehtinen 和 Sylvain Paris。GANSpace：发现可解释的生成对抗网络控制。神经信息处理系统进展，33：9841–9850，2020 年。2[12] Amir Hertz、Ron Mokady、Jay Tenenbaum、Kfir Aberman、Yael Pritch 和 Daniel Cohen-Or。利用交叉注意力控制进行 Prompt-to-Prompt 图像编辑。arXiv 预印本 arXiv:2208.01626，2022 年。2、6[13] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。去噪扩散概率模型。神经信息处理系统进展，33：6840–6851，2020 年。1、2[14] Zhiyuan Hu、Jiancheng Lyu、Dashan Gao 和 Nuno Vascon-celos. Pop：用于持续学习的提示之提示。arXiv 预印本 arXiv:2306.08200，2023。1[15] Ayodeji Ijishakin 、Ming Liang Ang、Levente Baljer、Daniel Chee Hian Tan、Hugo Laurence Fry、Ahmed Abdulaal、Aen- g us Lynch 和 James H Cole。H 空间稀疏自编码器。载于 2024 年 NeurIPS 安全生成式人工智能研讨会，2024。2[16] Jaeseok Jeong、Mingi Kwon 和 Youngjung Uh。利用扩散模型中的 h 空间实现免训练内容注入。载于 IEEE/CVF 冬季计算机视觉应用会议论文集，第 5151–5161 页，2024。2[17] KJ Joseph、Prateksha Udhayanan、Tripti Shukla、Aishwarya Agarwal、Srikrishna Karanam、Koustava Goswami 和 Bal- aji Vasan Srinivasan。使用扩散模型进行迭代式多粒度图像编辑。载于 IEEE/CVF 冬季计算机视觉应用会议论文集，第 8107–8116 页，2024。1[18] Tero Karras、Timo Aila、Samuli Laine 和 Jaakko Lehtinen。通过 GAN 渐进式生长提升质量、稳定性和多样性。载于第六届国际学习表征会议（ICLR 2018），加拿大不列颠哥伦比亚省温哥华，2018 年 4 月 30 日至 5 月 3 日，会议论文集。OpenReview.net，2018。5[19] Gwanghyun Kim、Taesung Kwon 和 Jong Chul Ye。DiffusionCLIP：用于稳健图像操控的文本引导扩散模型。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 2426–2435 页，2022。1, 6 – 。[20] Minchul Kim、Shangqian Gao、Yen-Chang Hsu、Yilin Shen 和 Hongxia Jin。令牌融合：弥合令牌剪枝与令牌合并之间的差距。载于 IEEE/CVF 冬季计算机视觉应用会议论文集，第 1383 1392, 2024 2 – 页。[21] Mingi Kwon、Jaeseok Jeong 和 Youngjung Uh。扩散模型已经拥有语义潜空间。arXiv 预印本 arXiv:2210.10960，2022。1, 2, 3, 5, 6, 11[22] Cheng-Han Lee、Ziwei Liu、Lingyun Wu 和 Ping Luo。MaskGAN：迈向多样且交互式的面部图像操控。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 5549–5558 页，2020。7 – 。[23] Cheng Lu、Yuhao Zhou、Fan Bao、Jianfei Chen、Chongxuan Li 和 Jun Zhu。DPM-Solver：一种快速常微分方程求解器，可在约 10 步内完成扩散概率模型采样。神经信息处理系统进展，35：5775–5787，2022。1, 2[24] Zeyu Lu、Chengyue Wu、Xinyuan Chen、Yaohui Wang、Lei Bai、Yu Qiao 和 Xihui Liu。层次化扩散自编码器与解耦图像操控。载于 IEEE/CVF 冬季计算机视觉应用会议论文集，第 5374–5383 页，2024。2[25] Daiki Miyake、Akihiro Iohara、Yu Saito 和 Toshiyuki Tanaka。负提示反演：利用文本引导扩散模型实现快速图像反演与编辑。载于 2025 年 IEEE/CVF 冬季计算机视觉应用会议（WACV），第 2063–2072 页。IEEE，2025。2[26] Ron Mokady、Amir Hertz、Kfir Aberman、Yael Pritch 和 Daniel Cohen-Or。使用引导式扩散模型编辑真实图像的空文本反演。载于

IEEE/CVF 计算机视觉与模式识别会议（CVPR）论文集，第 6038–6047 页，2023。1, 2[27] Toan Nguyen、Kien Do、Duc Kieu 和 Thin Nguyen。h-edit：通过 Doob 的 h 变换实现高效灵活的基于扩散的编辑。载于计算机视觉与模式识别会议论文集，第 28490–28501 页，2025。11[28] Jinhyeong Park、Muhammad Shaheryar、Seangmin Lee 和 Soon Ki Jung。在扩散模型中探索 h 空间以进行多属性编辑。载于 2025 年人工智能与信息通信国际会议（ICAIIC），第 1129–1133 页。IEEE，2025。2[29] Sung Woo Park、Hyomin Kim、Kyungjae Lee 和 Junseok Kwon。黎曼神经随机微分方程：在流形上学习随机表示。神经信息处理系统进展，35：1434–1444，2022。2, 3[30] Yong-Hyun Park、Mingi Kwon、Jaewoong Choi、Junghyo Jo 和 Youngjung Uh。从黎曼几何视角理解扩散模型的潜空间。神经信息处理系统进展，36：24129–24142，2023。2[31] Or Patashnik、Zongze Wu、Eli Shechtman、Daniel Cohen-Or 和 Dani Lischinski。StyleCLIP：文本驱动的 StyleGAN 图像操控。载于 IEEE/CVF 国际计算机视觉会议论文集，第 2085–2094 页，2021。6 – 。[32] Konpat Preechakul、Nattanat Chatthee、Suttisak Wizad- wongsa 和 Supasorn Suwajanakorn。扩散自编码器：迈向有意义且可解码的表示。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 10619–10629 页，2022。2[33]Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等。通过自然语言监督学习可迁移的视觉模型。载于国际机器学习会议论文集，第 8748–8763 页。PMLR，2021。7[34] Sam T. Roweis 和 Lawrence K. Saul。通过局部线性嵌入进行非线性降维。《科学》，290（5500）：2323–2326，2000。2[35] Shinnosuke Saito 和 Takashi Matsubara。利用扩散模型的基于得分的黎曼度量进行图像插值。arXiv 预印本 arXiv:2504.20288，2025。2, 3[36] Tim Salimans 和 Jonathan Ho。用于扩散模型快速采样的渐进式蒸馏。arXiv 预印本 arXiv:2202.00512，2022。2[37] Dvir Samuel、Barak Meiri、Haggai Maron、Yoad Tewel、Nir Darshan、Shai Avidan、Gal Chechik 和 Rami Ben-Ari。用于文生图扩散模型的极速图像反演与编辑。arXiv 预印本 arXiv:2312.12540，2023。2[38] Yujun Shen、Ceyuan Yang、Xiaoou Tang 和 Bolei Zhou。InterfaceGAN：解读生成对抗网络学习到的解耦面部表示。IEEE模式分析与机器智能汇刊，44(4)：2004–2018，2020。2[39] Ken Shoemake。使用四元数曲线制作旋转动画。载于第 12 届计算机图形学与交互技术年度会议论文集，第 245–254 页，1985。4, 12

[40] Jiaming Song、Chenlin Meng 和 Stefano Ermon。《去噪扩散隐式模型》。arXiv 预印本 arXiv:2010.02502，2020。4、12[41] Yang Song、Jascha Sohl-Dickstein、Diederik P Kingma、Abhishek Kumar、Stefano Ermon 和 Ben Poole。《通过随机微分方程进行基于分数的生成建模》。arXiv 预印本 arXiv:2011.13456，2020。1[42] Nikita Starodubcev、Dmitry Baranchuk、Valentin Khrulkov 和 Artem Babenko。《借助无条件扩散模型实现实时文本驱动的图像操控》。arXiv 预印本 arXiv:2304.04344，2023。2[43] Joshua Tenenbaum。《感知观测流形的映射》。神经信息处理系统进展，10，1997。2[44] Hongjie Wang、Difan Liu、Yan Kang、Yijun Li、Zhe Lin、Niraj K Jha 和 Yuchen Liu。《注意力驱动的扩散模型免训练效率提升》。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 16080–16089 页，2024。2[45] Sihan Xu、Yidong Huang、Jiayi Pan、Ziqiao Ma 和 Joyce Chai。《利用自然语言进行无反演图像编辑》。arXiv 预印本 arXiv:2312.04965，2023。2[46] Zizheng Yang、Hu Yu、Bing Li、Jinghao Zhang、Jie Huang 和 Feng Zhao。《释放扩散模型语义潜空间在图像去雾中的潜力》。发表于欧洲计算机视觉会议，第 371–389 页。Springer，2024。2[47] Fisher Yu、Yinda Zhang、Shuran Song、Ari Seff 和 Jianxiong Xiao。《LSUN：利用深度学习和人在环路中的方式构建大规模图像数据集》。CoRR，abs/1506.03365，2015。5[48] Evelyn Zhang、Bang Xiao、Jiayi Tang、Qianli Ma、Chang Zou、Xuefei Ning、Xuming Hu 和 Linfeng Zhang。《令牌剪枝助力更佳缓存：免费实现 Stable Diffusion 9 倍加速》。arXiv 预印本 arXiv:2501.00375，2024。2[49] Evelyn Zhang、Jiayi Tang、Xuefei Ning 和 Linfeng Zhang。《通过基于相似度的令牌剪枝，实现扩散模型的免训练硬件友好型加速》。发表于 AAAI人工智能会议论文集，第 9878–9886 页，2025。2[50] Zijian Zhang、Luping Liu、Zhijie Lin、Yichen Zhu 和 Zhou Zhao。《在预训练扩散模型的 h空间中无监督发现可解释方向》。arXiv 预印本 arXiv:2310.09912，2023。2[51] Wangbo Zhao、Yizeng Han、Jiasheng Tang、Kai Wang、Yibing Song、Gao Huang、Fan Wang 和 Yang You。《动态扩散变换器》。arXiv 预印本 arXiv:2410 03456，2024 2。.[52] PengFei Zheng、Yonggang Zhang、Zhen Fang、Tongliang Liu、Defu Lian 和 Bo Han。《NoiseDiffusion：利用扩散模型校正图像插值中的噪声，超越球面线性插值》。发表于第十二届国际学习表征会议，2024。12[53] Ye Zhu、Yu Wu、Zhiwei Deng、Olga Russakovsky 和 Yan Yan。《利用扩散模型进行边界引导的免学习语义控制》。神经信息处理系统进展，36：78319–78346，2023。6

7. 附录

A. 问题陈述

基于扩散的图像编辑旨在开发一种同时忠实于用户意图、几何上准确且计算高效的框架。尽管先前的研究表明，将语义偏移量 Δh添加到瓶颈特征的 h空间中是一种有效的操控策略 [21, 27], 但该方法面临三重挑战，而以往研究尚未对其进行全面解决。

首先，标准操作 $h' = h + \Delta h$ 默认将这个空间视为欧几里得空间，这可能导致伪影和流形外结果。因此，一个根本性问题仍未解决：

尊重数据流形内在非欧几里得结构的编辑向量 Δh 的最优表述是什么?

其次，即使方向在几何上合理，应用编辑时仍需进行细粒度控制，以保留主体的身份并避免伪影。简单文本提示本身存在的模糊性使这一点更加困难，并可能导致意料之外的语义偏移。这引出了一个问题：

如何以可调节的强度应用编辑，并通过精准且考虑上下文的指令加以引导，从而确保高保真度?

最后，任何通过提升模型复杂度来提高保真度的方法，都可能加剧扩散模型本已高昂的计算成本，从而妨碍其实际应用。这引出了第三个挑战：

如何在不损害编辑质量和语义一致性的前提下，使高保真编辑过程具备计算效率?

本文针对这些相互关联的问题展开研究。我们假设，真正稳健的解决方案需要一个统一框架：将 h-空间建模为黎曼流形，以推导出有理论依据的编辑向量；引入先进的混合与引导机制，实现高保真控制；并整合任务感知加速策略，确保效率。我们的工作旨在将这一整体框架形式化并付诸实现。

B. 任务感知剪枝的细节

1. **重要性评分：** 输入 X 被重塑为令牌序列 $T \in \mathbb{R}^{B\times N\times C}$，其中 $N = H \times W$。PruningHead 函数 P_θ 计算重要性分数 S:

$$S = \mathcal{P}_\theta(T, \mathbf{d}_{\text{edit}}) \in [0,1]^{B\times N} \tag{5}$$

2. **索引选择：** 给定剪枝比例 ρ，我们保留 $k = \lfloor N \cdot (1-\rho) \rfloor$ 个令牌。这些令牌的索引 I_{keep} 被选出:

$$\mathcal{I}_{\text{keep}} = \text{topk}_{\text{indices}}(S, k) \tag{6}$$

3. **剪枝注意力：** 查询（Q）、键（K）和值（V）使用选定的索引收集投影，以形成经过剪枝的集合 Q_{kept}、K_{kept}、$V_{\text{kept}} \in \mathbb{R}^{B\times k\times C}$。注意力仅在这个缩减后的集合上计算:

$$A_{\text{pruned}} = \text{Softmax}\left(\frac{Q_{\text{kept}} K_{\text{kept}}^T}{\sqrt{C}}\right) V_{\text{kept}} \tag{7}$$

4. **散射与输出：** 注意力处理后的特征 $A_{\text{pruned}} \in \mathbb{R}^{B\times k\times C}$ 会在其原始位置 I_{keep} 处散射回原始大小的全零张量 $A_{\text{result}} \in \mathbb{R}^{B\times N\times C}$。最终输出 X_{out} 通过与输出投影 W_o 的残差连接计算得出:

$$X_{\text{out}} = X + W_o(A_{\text{result}}) \tag{8}$$

C. 指数映射架构

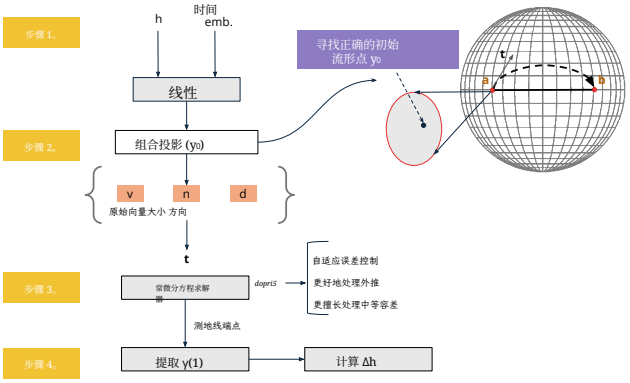

图11。我们的 ExpoMap 模块通过估计克里斯托费尔符号并使用数值积分器（Dopri5）求解测地线常微分方程，来学习 h-空间流形的局部几何。模块以语义方向和时间步为输入，通过可学习的投影将其映射为初始切向量 v_0，并沿流形积分以计算最终偏移量 Δh（$\gamma(1) - h$）。这使得无需事后估计雅可比矩阵即可进行几何感知编辑。

D. Qwen2-VL 的工作流程详情

尽管 CLIP嵌入已被证明能有效引导语义编辑，但它们往往会编码社会或审美偏见，且其概念分布相对狭窄。为缓解这一局限，我们利用冻结的Qwen2-VL模型提供的额外上下文丰富 CLIP嵌入，从而有效构建更广泛的 CLIP+Qwen2 联合

表 5. 扩散模型编辑中插值方法的比较。虽然 LERP 很简单，但由于偏离数据流形，通常会产生伪影。SLERP 通过保持超球面结构对此进行了改进，而 NoiseDiffusion 通过校正噪声分布，进一步针对自然图像优化了这一点。我们提出的双重球面线性插值通过同时作用于学习到的特征流形和噪声空间，提供了精细的控制，从而实现对编辑强度和身份保持的解耦控制。

特征	LERP（线性）	SLERP（球面） [39, 40]	NoiseDiffusion [52]	双重球面线性插值（我们的方法）
流形一致性	流形外	位于超球面上	校正（投影到有效的噪声分布）	**双流形感知**（黎曼与超球面）
语义融合	有限 / 纠缠	细粒度	细粒度（已校正）	**层次化与解耦**
伪影风险	High	中等（针对自然图像）	较低（针对自然图像）	**最低**（由于正交性）
控制机制	粗粒度（单个参数）	平滑且结构化	多参数（校正 + 插值）	**解耦控制**（编辑强度与保真度）

嵌入空间。这种扩展后的表示捕捉了细粒度、实例特定的视觉语义，尤其适用于“面部”或“年轻女性”等描述不充分的提示。我们的黎曼编辑和双重 SLERP 模块优化了噪声与特征空间的几何结构，而这种多模态增强则确保了文本层面的对齐。二者结合，使 RemEdit 即使面对模糊或带有偏见的属性条件，也能执行几何上连贯且语义上精确的编辑。

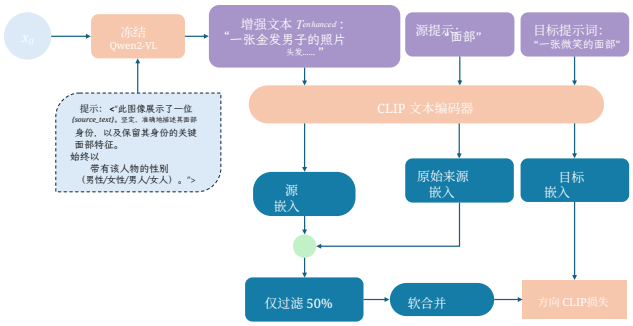

图 12. 源图像 x_0 和一条描述性提示被输入冻结的 Qwen2-VL模型，以生成详细且针对特定实例的说明文字，T_{enhanced}。这段增强文本提供了更丰富的源嵌入，从而约束编辑，使其更忠实于原始图像的核心属性。

E. 扩散模型中的图像插值

为更好地阐明我们的贡献，我们在表5中对不同的插值策略进行了概念性比较。NoiseDiffusion [52] 通过将噪声投影回有效先验来缓解这一问题。我们提出的双重 SLERP 机制更进一步，具备双流形感知能力：内层球面线性插值在学习到的黎曼 h空间中进行，而外层 SLERP 则在噪声空间中调节保真度。这种解耦方案能够分别控制语义强度和身份保持，从而实现精准、稳健的图像编辑。