Region-Reg-DPO

面向视频补全的分区正则化直接偏好优化

从 DiffuEraser 与 DPO 的结构性缺陷出发的方法论创新

完整理论推导与 DiffuEraser 融合方案设计

Technical Report

2026 年 2 月 19 日

摘要

视频补全（Video Inpainting）是计算机视觉中的核心任务，其目标是在保持时空一致性的前提下修复视频中的缺失区域。当前主流方案分为两条路线：以 ProPainter 为代表的传统光流传播方法，和以 DiffuEraser 为代表的生成式扩散模型方法。前者在高时间可见性场景下表现优异但在低可见性场景下彻底失效；后者虽具备" 无中生有" 的生成能力，但面临**六大系统性痛点**：Prompt 语义困境、Mask 双刃剑效应、时空异质性感知盲区、L2 损失的固有局限、条件信号耦合导致的人类偏好失配、以及边界融合的不可控性。

另一方面，直接偏好优化（DPO）作为对齐人类偏好的主流方法，在应用于扩散模型时暴露出**两大结构性缺陷**：DPO 梯度比率（DGR）快速衰减导致梯度消失、以及缺乏绝对分布约束导致的" 作弊路径"。Reg-DPO 通过引入 SFT 正则化修复了这两个缺陷。

本报告提出 **Region-Reg-DPO**——一种面向视频补全任务的分区正则化直接偏好优化方法。该方法将帧空间分解为洞内、边界、上下文三个不相交区域，使用两组**独立**的区域权重（α_r 控制 DPO 偏好学习，ρ_r 控制 SFT 锚定）实现全链路区域感知。我们发现了分区 DPO 的**第三缺陷**——区域间优化需求不对称被共享 DGR 掩盖，并通过常数 ρ_r 设计加以修复。理论上，Region-Reg-DPO 是 Reg-DPO 的**严格泛化**，当所有权重统一时退化为标准 Reg-DPO。

我们详细分析了 Region-Reg-DPO 与 DiffuEraser 双分支架构的兼容性，提出了三阶段训练方案和多层次负样本构建策略，旨在将 DiffuEraser 从” 最小化重建误差” 升级为” 对齐人类偏好” 的视频补全系统。

关键词：视频补全、直接偏好优化、扩散模型、区域正则化、DiffuEraser

目录

第 1 章　引言

1.1　研究动机

视频补全（Video Inpainting）在影视后期制作、短视频编辑和计算机视觉研究中扮演着至关重要的角色。其核心任务可分为两类：**物体移除**（Object Removal, OR）——从视频中移除不需要的物体，以及**背景恢复**（Background Restoration, BR）——修补视频中的缺失或破损区域。无论哪种任务，核心挑战都可以概括为八个字：**空间合理，时间一致**。

当前，学术界和工业界演化出两条截然不同的技术路线来应对这一挑战：

- **传统光流补全**（以 ProPainter 为代表）：核心机制是” 搬运与传播”。利用光流（Optical Flow）在时间轴上精准寻找” 曾经露面” 的真实像素，并将它们对齐、传播到当前帧的缺失区域。
- **生成式扩散模型**（以 DiffuEraser 为代表）：核心机制是” 无中生有”。不再苦苦寻找真实像素，而是依靠模型庞大的先验知识，根据上下文重新” 画” 出缺失区域。

然而，这两条路线都存在根本性局限。传统方法极度依赖” 时间可见性”——当目标背景从头到尾都被遮挡时彻底失效；生成式方法虽然理论上可以处理任意遮挡场景，但在工程实践中面临严重的系统性错配。

与此同时，**直接偏好优化**（Direct Preference Optimization, DPO）作为对齐人类偏好的前沿方法，也在扩散模型领域暴露出结构性缺陷，limiting 了其在实际训练中的有效性。

本报告的核心贡献在于：**系统性地分析 DiffuEraser 和 DPO 各自的结构性缺陷，并提出 Region-Reg-DPO 方法，以一种数学严谨且工程可行的方式将偏好优化融入视频补全流水线，同时修复双方的缺陷**。

1.2　本文结构

本报告的组织结构如下：第 2 章深入分析 DiffuEraser 的六大系统性痛点；第 3 章完整推导 DPO 在扩散模型中的损失函数与梯度公式，揭示其两大结构性缺陷及 Reg-DPO 的修复方案；第 4 章提出 Region-Reg-DPO 方法，包含三区分解、完整数学推导、第三缺陷的发现与修复；第 5 章分析 Region-Reg-DPO 如何逐一解决 DiffuEraser 的痛点；第 6 章给出完整的融合方案设计，包括训练策略、数据构建和参数推荐；第 7 章进行风险评估与实施规划。

第 2 章 DiffuEraser 的系统性痛点分析

DiffuEraser 采用双分支架构完成视频补全任务：

- **主 UNet**：基于 Stable Diffusion 1.5 的去噪网络，包含空间注意力层和时序运动模块（Motion Module）
- **BrushNet**：辅助分支，提取 Masked Image 的特征并通过跨注意力（Cross-Attention）注入主 UNet
- **Prior 注入**：将 ProPainter 输出通过 DDIM Inversion 转换为噪声先验，作为去噪的初始状态
- **Blending**：最终通过高斯模糊的 Mask 将补全结果与原始帧进行像素级混合

尽管这一架构在特定场景下表现出色，但它面临着以下六大系统性痛点。

2.1 痛点一：Prompt 的语义困境

Prompt（文本提示）作为条件信号天然是**全局性**、**粗粒度的**（如” 一条阳光明媚的公园小路”），它难以精确控制局部被遮挡区域所需的纹理延伸或光照渐变。

工程实战发现

在实际部署中（参见 `run_OR.py` 的 `text_guidance_scale` 参数），对于物体移除（OR）任务，**不加 Prompt 往往比加 Prompt 效果更好**。加入文本引导后，模型获得的信号过于宽泛，不仅无法为局部修复提供有效指导，反而容易引发不连贯的幻觉；而将 Guidance Scale 设为零（空 Prompt），让扩散模型单纯依赖 Mask 周围的上下文特征进行无条件去噪，融合度反而更高。

这一现象的本质原因在于：视频补全是一个**强条件局部生成**任务，与全局文生图任务的条件信号需求存在根本性错配。Prompt 提供的全局语义信号与局部修复所需的精细纹理延伸之间存在语义鸿沟。

2.2 痛点二：Mask 的双刃剑效应

Mask 作为独立通道送入 BrushNet 分支（参见 `pipeline_diffueraser.py` 中的条件注入逻辑），携带了强烈的物体轮廓先验。特别是在物体移除任务中，Mask 恰好就是目标物体的精确轮廓。

双刃剑效应表现为：扩散模型会隐式” 感知” 到 Mask 通道中的轮廓信息，倾向于在被遮挡区域内生成一个与原物体形状一致的伪影——用户想要删除一辆车，模型却顺

着 Mask 的轮廓画了个” 幽灵车” 留在原地。这种”Mask 泄露” 问题在当前 L2 训练范式下**无法通过增加训练数据来解决**，因为 L2 损失从未教过模型” 什么不该生成”。

2.3 痛点三：时空异质性的感知盲区

DiffuEraser 缺乏对**时间可见性**（Temporal Visibility）维度的感知能力。在视频补全的实际场景中，同一帧内的不同区域可能具有截然不同的时间可见性特征：

- 高可见性区域：目标背景在相邻帧中大量暴露，ProPainter 的光流传播已经恢复了精确像素
- 低可见性区域：背景从头到尾被遮挡，ProPainter 只能产出模糊失真的填充

这导致了一个根本性矛盾：在高可见性区域，扩散模型的” 创造性介入” 反而**破坏**了光流已经恢复的精确像素；在低可见性区域，又因缺乏足够的语义引导而**无法**生成合理内容。这种**” 该保守时过度创造，该创造时能力不足”** 的矛盾是现有框架对时空特性无感的直接后果。

2.4 痛点四：L2 损失的固有局限

DiffuEraser 当前使用标准的 L2 损失（均方误差）训练噪声预测网络：

$$\mathcal{L}_{\text{L2}} = \mathbb{E}_{t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right] \tag{1}$$

L2 损失在统计意义上等价于对所有可能输出的**均值**进行回归。当目标分布是多模态的（即对于同一个被遮挡区域存在多种合理的补全方案），L2 损失会驱动模型生成这些方案的加权平均——一个模糊的、缺乏细节和多样性的” 安全解”。

核心洞察

L2 损失教会了模型**” 什么是正确的平均答案”**，但从未教过模型**” 在多个正确答案中，哪个更符合人类视觉偏好”**，也从未教过模型**” 什么不该生成”**。这两个能力缺口正是 DPO 偏好优化可以填补的空间。

2.5 痛点五：条件信号耦合导致的人类偏好失配

前述痛点一至痛点四并非孤立存在，它们在实际推理中相互**耦合放大**，共同造成了一个更深层的系统性问题：**生成结果系统性地偏离人类视觉偏好**。

具体而言，Prompt 的全局语义信号、Mask 的轮廓先验信息、以及扩散模型自身不可控的生成能力三者形成了一个” 失控三角”：

1. **语义漂移（Prompt × 生成能力）**：当 Prompt 提供了模糊的全局描述（如” 城市街道”），扩散模型的强大生成能力会过度” 脑补”——在本应只需延伸路面纹理的 Mask 区域内，凭空生成新的建筑物、车辆甚至行人。这些生成物在像素级别可能是” 合理” 的（纹理清晰、光照一致），但在语义层面完全违背了用户” 让物体消失” 的意图。L2 损失无法惩罚这类语义级错误，因为它只衡量像素重建精度，不衡量语义合理性。

2. **轮廓幽灵（Mask × 生成能力）**：Mask 通道的轮廓信息与扩散模型的条件生成机制发生” 共振”。BrushNet 分支将 Mask 编码为条件特征注入 UNet，模型隐式地将 Mask 轮廓**解读为一个” 需要在此处生成某种物体” 的空间先验**，而非” 此处需要移除物体” 的语义指令。越强大的生成能力反而越会” 填充” 这个轮廓，产出与原物体形状高度吻合的幻觉伪影。

3. **风格断裂（Prompt × Mask × 生成能力）**：当三者同时作用时，Mask 内部区域可能呈现与周围上下文**风格不一致**的生成结果。例如，在一段手持视频中移除前景人物后，Mask 区域内的背景可能呈现出更平滑、更”AI 风格” 的纹理——与 Mask 外的真实像素在颗粒感、噪声分布、色彩饱和度上存在可察觉的差异。人眼对这种微妙的风格断裂极为敏感，即使 PSNR/SSIM 等指标可能正常。

核心矛盾

上述三种耦合效应的共同本质是：**DiffuEraser 的训练目标（最小化 L2 重建误差）与用户的真实需求（视觉上” 看起来像是物体从未存在过”）之间存在根本性错位**。L2 损失优化的是像素级统计量，而人类偏好是一个涉及语义合理性、风格一致性、视觉自然度等多维度的高层判断。这种错位无法通过增加训练数据、调整超参数或改进网络架构来弥合——它需要一种从**目标函数层面**引入人类偏好信号的全新训练范式。

2.6 痛点六：边界融合的不可控性

DiffuEraser 的最终输出通过 Blending 步骤完成（参见 `diffueraser_OR.py` 中的 `blended` 参数），即使用高斯模糊处理后的 Mask 将模型补全结果与原始帧进行像素级线性混合：

$$y_{\text{out}} = M_{\text{blur}} \odot y_{\text{generated}} + (1 - M_{\text{blur}}) \odot y_{\text{original}} \tag{2}$$

这一策略存在两个固有缺陷：

- **高斯核参数硬编码**：模糊核的标准差和膨胀迭代次数（`mask_dilation_iter`）在推理时固定，无法根据局部内容复杂度自适应调整。在纹理丰富的边界区域（如树叶、

头发丝），固定的高斯模糊导致可见的”光晕”伪影；在平坦区域（如天空、墙壁），同样的模糊又显得过于保守。

- **混合信号与训练信号脱节**：Blending 作为推理时的后处理步骤，在训练阶段是**不可见的**——L2 损失直接作用于 UNet 的去噪输出，而非最终混合结果。这意味着模型在训练时从未”见过”自己的输出经过 Blending 后的样子，更无法针对边界融合质量进行优化。

Region-Reg-DPO 的分区 SFT 正则化（特别是边界区域的 ρ_b 最强锚定）正是针对这一痛点：通过在训练阶段对边界区域施加最强的重建精度约束，确保 UNet 的去噪输出在 Mask 边缘处就已经与原始像素高度一致，从而**减轻对 Blending 后处理的依赖**。

2.7 DiffuEraser 的最佳工作区间

综合以上分析，DiffuEraser 当前的最佳工作区间是一个**极窄的窗口**：**物体缓慢运动 + 相机静止**的 OR 任务。此场景恰好避开了所有痛点：Mask 形状在帧间平滑过渡（痛点二缓解）；ProPainter 提供了方向正确的低频先验（痛点三缓解）；固定机位背景本身几乎不变（痛点三/四缓解）。

Region-Reg-DPO 的引入目标正是**拓宽这个窗口**。

第 3 章 DPO 的结构性缺陷与 Reg-DPO 修复

3.1 从 RLHF 到 DPO 的推导简述

DPO 源自 RLHF 框架。PPO 的目标是最大化期望奖励同时约束策略不偏离参考策略：

$$\max_{\pi_\theta} \mathbb{E}_{x\sim\mathcal{D},\, y\sim\pi_\theta(y|x)} \left[r_\phi(x,y)\right] - \beta\, D_{\mathrm{KL}}\left[\pi_\theta(y|x) \,\|\, \pi_{\mathrm{ref}}(y|x)\right] \tag{3}$$

通过求解最优策略 $\pi^* = \frac{1}{Z(x)}\pi_{\mathrm{ref}}(y|x)\exp\Big(\frac{1}{\beta}r_\phi(x,y)\Big)$，反解出隐式奖励函数 $r_\phi(x,y) = \beta\log\frac{\pi_\theta(y|x)}{\pi_{\mathrm{ref}}(y|x)} + \beta\log Z(x)$，代入 Bradley-Terry 偏好模型，$Z(x)$ 项相消，得到 DPO 的目标函数：

$$\mathcal{L}_{\mathrm{DPO}} = -\mathbb{E}_{(x,y^w,y^l)}\left[\log\sigma\left(\beta\cdot s(\theta)\right)\right] \tag{4}$$

其中 $s(\theta) = \big[\log\pi_\theta(y^w|x) - \log\pi_\theta(y^l|x)\big] - \big[\log\pi_{\mathrm{ref}}(y^w|x) - \log\pi_{\mathrm{ref}}(y^l|x)\big]$。

3.2 扩散模型中的 DPO

在扩散模型中，概率通过预测误差间接表达：$\log\pi_\theta(y|x) \Leftrightarrow -\|y - y_\theta(x_t,t)\|^2$（方向相反）。代入并取反，得到扩散模型版本的 DPO 损失：

$$\mathcal{L}_{\text{DPO}}(\theta) = -\mathbb{E}\left[\log \sigma(-\beta \cdot s(\theta))\right] \tag{5}$$

其中：

$$s(\theta) = \underbrace{\left(\|\epsilon^w - \epsilon_\theta^w\|^2 - \|\epsilon^l - \epsilon_\theta^l\|^2\right)}_{\Delta_\theta(x^w, x^l)} - \underbrace{\left(\|\epsilon^w - \epsilon_{\text{ref}}^w\|^2 - \|\epsilon^l - \epsilon_{\text{ref}}^l\|^2\right)}_{\Delta_{\text{ref}}(x^w, x^l)(\text{常量})} \tag{6}$$

优化方向：最小化 $\mathcal{L}_{\text{DPO}}$ 等价于让 $s(\theta) \to -\infty$，即让模型更精确地预测正样本噪声（$\|\epsilon^w - \epsilon_\theta^w\|^2 \downarrow$），同时降低对负样本的预测精度（$\|\epsilon^l - \epsilon_\theta^l\|^2 \uparrow$）。

3.3 DPO 梯度推导（完整 5 步）

令 $S = -\beta \cdot s(\theta)$，则 $\mathcal{L} = -\mathbb{E}[\log \sigma(S)]$。

Step 1：对 $\log \sigma$ 求导

$$\nabla_\theta \mathcal{L} = -\mathbb{E}\left[(1 - \sigma(S)) \cdot \nabla_\theta S\right] \tag{7}$$

Step 2：$\nabla_\theta S = -\beta \cdot \nabla_\theta s(\theta)$，参考模型冻结故 $\nabla_\theta \Delta_{\text{ref}} = 0$

Step 3：L2 范数梯度

$$\nabla_\theta \|y - y_\theta(x_t, t)\|^2 = -2\left(y - y_\theta(x_t, t)\right)^T \nabla_\theta y_\theta(x_t, t) \tag{8}$$

Step 4–5：代入正负样本并定义 **DPO Gradient Ratio (DGR)**：

$$\text{DGR} = \beta(1 - \sigma(S)) = \frac{\beta}{1 + e^{-\beta s(\theta)}} \tag{9}$$

$$\boxed{\nabla_\theta \mathcal{L}_{\text{DPO}} = -\mathbb{E}\left[2\,\text{DGR} \cdot \left\{(\epsilon^w - \epsilon_\theta^w)^T \nabla_\theta \epsilon_\theta^w - (\epsilon^l - \epsilon_\theta^l)^T \nabla_\theta \epsilon_\theta^l\right\}\right]} \tag{10}$$

3.4 缺陷一：DGR 快速衰减——自毁循环

定理 3.1 (DGR 自毁循环). *在 DPO 训练过程中，DGR 必然经历指数级衰减：*

1. *训练开始时 $\theta \approx \theta_{ref}$，故 $s(\theta) \approx 0$，$DGR \approx \beta/2$（很大），梯度很强；*
2. *强梯度迅速拉开正负样本预测误差差距，$s(\theta)$ 快速变负；*
3. *$e^{-\beta s(\theta)}$ 指数爆炸，DGR 急剧衰减至接近零；*
4. *梯度消失，训练停滞（过早收敛）。*

整个过程在 $\beta = 1000$ 时可能仅需数百步即完成。

以 $\beta = 1000$ 的数值验证：

$s(\theta)$ 仅从 0 变到 -0.01——一个极其微小的变化——DGR 就暴跌了超过四个数量级。

表 1: DGR 衰减的数值分析（$\beta = 1000$）

$s(\theta)$	$e^{-\beta s}$	DGR	相对初始值
0	1	500	100%
-0.001	2.72	≈ 269	53.8%
-0.005	148.41	6.69	1.34%
-0.01	22026	0.045	0.009%
-0.02	4.9×10^8	≈ 0	≈ 0

3.5 缺陷二：无绝对分布约束——" 作弊路径"

DPO 只约束正负样本预测误差的**差值** $A-B$（其中 $A = \|\epsilon^w - \epsilon_\theta^w\|^2$, $B = \|\epsilon^l - \epsilon_\theta^l\|^2$），不约束**绝对值**。

表 2: 理想路径 vs " 作弊" 路径

	A（正样本误差）	B（负样本误差）	$A - B$	模型状态
初始	10	10	0	正常
理想路径	5	15	-10	**健康**
作弊路径	500	510	-10	**崩溃**

在高维参数空间中，" 作弊路径" 的可行方向数量远多于" 理想路径"，因此模型在无额外约束时更倾向于走捷径。

3.6 Reg-DPO 的修复方案

Reg-DPO 的设计思路是**从梯度缺陷出发，做最小修补**：为正样本的梯度系数增加一个不随 DGR 衰减的常数项 $r > 0$，使正样本始终保持非零梯度信号。

从修改后的梯度**反推**损失函数：

$$\mathcal{L}_{\text{Reg-DPO}}(\theta) = \underbrace{-\mathbb{E}\left[\log \sigma(-\beta \cdot s(\theta))\right]}_{\text{DPO 损失}} + \underbrace{r \cdot \mathbb{E}\left[\|\epsilon^w - \epsilon_\theta(x_t^w, t)\|^2\right]}_{\text{SFT 正则化}} \tag{11}$$

三重稳定机制：

1. **持续梯度信号**：即使 DGR $\to 0$，正样本仍保有 $r > 0$ 的梯度系数
2. **间接约束负样本**：SFT 项锚定 $A = \|\epsilon^w - \epsilon_\theta^w\|^2$ 不能膨胀，迫使模型走" 理想路径"
3. **控制分布偏移**：SFT 项将 θ 锚定在 θ_{ref} 附近，防止灾难性遗忘

第 4 章 Region-Reg-DPO：面向视频补全的分区正则化 DPO

4.1 核心动机：视频补全的空间异质性

将 Reg-DPO 直接应用于 DiffuEraser 的最大问题在于：视频补全任务中，帧空间的不同区域具有**完全不同的优化目标和约束需求**：

表 3: 视频补全中不同区域的优化需求

区域	优化目标	关键约束	对持续梯度的需求
洞内 M_h	生成合理新内容	语义一致性	中等
边界 M_b	无缝融合过渡	与原始像素严格对齐	**最高**
上下文 M_c	保持原始不变	完全保护	最低（由 Blending 保证）

标准 Reg-DPO 使用**全局统**一的损失权重，无法区分这些不同需求。

4.2 三区 Mask 分解

定义 4.1 (三区分解)**.** 将视频帧空间分为三个不相交区域：

- **洞内区域** R_{hole}，$mask\ M_h$：需要生成的核心区域
- **边界带** R_{bd}，$mask\ M_b = Dilate(M_h, k) - M_h$：融合过渡区
- **上下文区域** R_{ctx}，$mask\ M_c = 1 - M_h - M_b$：未遮挡的原始像素

满足空间完整划分：$M_h + M_b + M_c = 1$。

4.3 偏好对定义（GT-Pair 策略）

训练时通过 `create_random_shape_with_random_motion` 生成随机 mask 序列：

- **正样本（win）**：原始未遮挡视频帧（Ground Truth），对应噪声目标 ϵ^w
- **负样本（lose）**：模型补全后的视频帧对应的噪声预测目标 ϵ^l

这与 Reg-DPO 的 GT-Pair 策略天然对应：原始帧即为” 完美补全”。

4.4 Region-Reg-DPO 损失函数

Region-Reg-DPO 完整损失函数

$$\mathcal{L}(\theta) = \underbrace{-\mathbb{E}\left[\log\sigma(-\beta' \cdot s_{\text{region}}(\theta))\right]}_{\text{Region-DPO 损失}} + \underbrace{\mathbb{E}\left[\sum_{r\in\{h,b,c\}} \rho_r \cdot \|M_r \odot (\epsilon^w - \epsilon_\theta(x_t^w, t))\|^2\right]}_{\text{分区 SFT 正则化}} \tag{12}$$

其中区域加权的偏好得分为：

$$s_{\text{region}}(\theta) = \sum_r \alpha_r \left[\left(\|M_r \odot (\epsilon^w - \epsilon_\theta^w)\|^2 - \|M_r \odot (\epsilon^w - \epsilon_{\text{ref}}^w)\|^2\right) - \left(\|M_r \odot (\epsilon^l - \epsilon_\theta^l)\|^2 - \|M_r \odot (\epsilon^l - \epsilon_{\text{ref}}^l)\|^2\right)\right] \tag{13}$$

两组独立的区域权重是 Region-Reg-DPO 的核心设计：

表 4: α_r 与 ρ_r 的独立性分析

权重	所在位置	物理含义	是否随 DGR 衰减
α_r	DPO 损失（σ 函数内部）	区域对偏好学习的贡献	是（随 $\text{DGR}_{\text{region}}$ 衰减）
ρ_r	SFT 损失（σ 函数外部）	区域对正样本锚定的强度	**否（常数）**

4.5 完整梯度推导

令 $\beta' = \beta T \omega(\lambda_t)$，$S = -\beta' \cdot s_{\text{region}}(\theta)$。

Step 1–2：与标准 DPO 推导类似，利用链式法则和参考模型冻结条件得：

$$\nabla_\theta S = -\beta' \sum_r \alpha_r \left[\nabla_\theta \|M_r \odot (\epsilon^w - \epsilon_\theta^w)\|^2 - \nabla_\theta \|M_r \odot (\epsilon^l - \epsilon_\theta^l)\|^2\right] \tag{14}$$

Step 3：Mask 加权 L2 范数的梯度（关键步骤）

当 M_r 为二值 mask 时 $M_{r,i}^2 = M_{r,i}$，故：

$$\boxed{\nabla_\theta \|M_r \odot (\epsilon - \epsilon_\theta)\|^2 = -2\left(M_r \odot (\epsilon - \epsilon_\theta)\right)^T \nabla_\theta \epsilon_\theta} \tag{15}$$

Step 4–5：代入正负样本，定义区域 DGR 并合并：

$$\text{DGR}_{\text{region}} = \frac{\beta'}{1 + e^{-\beta' \cdot s_{\text{region}}(\theta)}} \tag{16}$$

$$\nabla_\theta \mathcal{L}_{\text{Region-DPO}} = -\mathbb{E}\left[2\text{DGR}_{\text{region}} \sum_r \alpha_r \left\{(M_r \odot (\epsilon^w - \epsilon_\theta^w))^T \nabla_\theta \epsilon_\theta^w - (M_r \odot (\epsilon^l - \epsilon_\theta^l))^T \nabla_\theta \epsilon_\theta^l\right\}\right] \tag{17}$$

4.6 第三缺陷的发现：区域优化需求不对称

缺陷 3（新发现）：区域优化需求不对称被共享 DGR 掩盖

这是引入分区权重后出现的**新问题**，标准 Reg-DPO 中不存在：
洞内区域面积大、信号强，往往最先学到偏好差异 → 驱动 s_{region} 快速变负 → DGR 全局衰减 → 边界区域尚未充分优化就失去梯度。
边界是最需要持续优化的区域（对人眼最敏感），却最容易被洞内的快速收敛"拖垮"。

4.7 梯度修复：分区独立 SFT 正则化

为正样本在每个区域添加**不随 DGR 衰减的常数**项 $\rho_r > 0$：

$$\nabla_\theta \mathcal{L}_{\text{修改}} = -\mathbb{E}\left[\sum_r 2\left\{(\alpha_r \cdot \text{DGR}_{\text{region}} + \rho_r)(M_r \odot (\epsilon^w - \epsilon_\theta^w))^T \nabla_\theta \epsilon_\theta^w - \alpha_r \cdot \text{DGR}_{\text{region}} \cdot (M_r \odot (\epsilon^l - \epsilon_\theta^l))^T \nabla_\theta \epsilon_\theta^l\right.\right. \tag{18}$$

拆解为 Region-DPO 梯度 + 分区 SFT 梯度，后者恰好为 $\nabla_\theta \sum_r \rho_r \|M_r \odot (\epsilon^w - \epsilon_\theta^w)\|^2$，反推得到公式 (12) 的完整损失函数。

4.8 ρ_r 必须为常数的必要性

命题 4.1 (ρ_r 不可依赖 DGR). *若令 $\rho_r = \gamma_r \cdot DGR_{region}$，则正样本梯度系数变为 $(\alpha_r + \gamma_r) \cdot DGR_{region}$，当 $DGR_{region} \to 0$ 时**全部归零**，SFT 保护完全失效。ρ_r 必须为不依赖 DGR_{region} 的常数，才能在 DGR 衰减后接管梯度信号。*

动态验证（$\alpha_r = 1$，$\rho_r = 100$，$\beta' = 1000$）：

4.9 与 Reg-DPO 的统一性

定理 4.1 (严格泛化). *当 $\alpha_h = \alpha_b = \alpha_c = 1$ 且 $\rho_h = \rho_b = \rho_c = \rho$ 时，由于三区不重叠且 $M_h + M_b + M_c = 1$：*

$$\sum_r \|M_r \odot (\epsilon - \epsilon_\theta)\|^2 = \|\epsilon - \epsilon_\theta\|^2, \quad \sum_r \rho \|M_r \odot (\epsilon^w - \epsilon_\theta)\|^2 = \rho \|\epsilon^w - \epsilon_\theta\|^2$$

Region-Reg-DPO 严格退化为标准 Reg-DPO。

表 5: 训练过程中 DPO 与 SFT 的动态主导权交接

训练时期	DGR	$\alpha_r \cdot \mathrm{DGR}$	ρ_r	主导力量
初始（$s \approx 0$）	500	500	100	DPO 偏好学习（83%）
早期（$s = -0.005$）	6.69	6.69	100	过渡 → SFT 接管
中后期（$s = -0.01$）	0.045	0.045	100	SFT 锚定（≈ 100%）

第 5 章　Region-Reg-DPO 如何解决 DiffuEraser 的痛点

5.1　痛点一（幻觉抑制）→ 偏好对引导

DiffuEraser 的幻觉问题本质上是模型在 L2 损失下从未学过” 什么不该生成”。Region-Reg-DPO 通过**负样本**显式提供反面教材：

- 移除 Prior 注入的 DiffuEraser 输出作为**幻觉负样本**（包含幽灵物体、Mask 轮廓伪影）
- 模型通过偏好学习显式学到” 避免在 Mask 区域内生成与轮廓形状一致的伪影”

5.2　痛点二（模糊抑制）→ SFT 分区锚定

L2 导致的模糊源于模型倾向于生成均值解。DPO 偏好信号让模型学会选择**纹理更丰富、细节更清晰**的方案。同时，ρ_b（边界最强 SFT 锚定）确保边界重建误差被严格控制，防止 Blending 步骤出现接缝伪影。

5.3　痛点三（时序一致性）→ 多帧联合偏好

构建偏好对时以**连续多帧序列**为单位，让 DPO 梯度信号天然包含时序一致性信息。收集 DiffuEraser 在 clip 边界处的时序跳变帧作为闪烁负样本。

5.4　痛点四（工作区间过窄）→ 多源负样本覆盖

通过构建覆盖不同失败模式的**三层次负样本**（幻觉型、模糊型、闪烁型），让模型在更广泛的场景下学到人类偏好，从而拓宽有效工作区间。

第 6 章　融合方案设计

表 6: 完整三阶段训练流程

阶段	目标	微调对象	损失函数
Stage 1	内容生成能力	BrushNet + UNet（无运动模块）	L2（已完成）
Stage 2	时序一致性	UNet 运动模块	L2（已完成）
Stage 3	**偏好对齐**	主 UNet 的 LoRA	**Region-Reg-DPO**（新增）

6.1 三阶段训练流程

6.2 Stage 3 架构设计

- **微调范围**：仅主 UNet 的空间/时序注意力层，使用 LoRA（rank=16，α=32）
- **冻结组件**：BrushNet（保护 Prior 编码）、运动模块（保护时序一致性）
- **参考模型**：Stage 2 输出的完整模型（冻结），接收与当前模型相同的 Prior 输入
- **Prior 保持**：训练期间必须保留 ProPainter Prior 的 DDIM Inversion 注入流程

6.3 偏好对数据构建

表 7: 三层次负样本设计

类型	生成方式	覆盖的失败模式
幻觉负样本	移除 Prior 的 DiffuEraser 输出	幽灵物体、噪声伪影、Mask 轮廓伪影
模糊负样本	低可见性场景的 ProPainter 输出	边缘拉伸模糊、纹理缺失
闪烁负样本	DiffuEraser 长序列 clip 边界帧	时序跳变、帧间不一致

6.4 参数推荐

第 7 章 风险评估与实施规划

表 8: Region-Reg-DPO 完整参数推荐

参数	推荐值	位置	说明
α_h	1.0	DPO（σ 内）	洞内偏好基准
α_b	1.5–2.0	DPO（σ 内）	边界偏好优先
α_c	0.05–0.1	DPO（σ 内）	上下文几乎不参与
ρ_h	50–150	SFT（σ 外）	洞内中等锚定
ρ_b	100–250	SFT（σ 外）	边界最强锚定
ρ_c	0–10	SFT（σ 外）	上下文由 Blending 保护
β	500–1000	DPO 偏好锐度	与 Reg-DPO 一致
lr	1×10^{-6}–5×10^{-6}	全局学习率	需结合 β' 调整
LoRA rank	16	微调维度	Reg-DPO 验证最优
膨胀核 k	5–15 px	边界带宽度	约为 mask 半径 5%–10%
训练帧数	22 帧	clip 长度	与 DiffuEraser 推理一致

7.1 风险矩阵

表 9: 风险矩阵

	风险	概率	影响	缓解措施
R1	DPO 破坏 Prior 注入	中	高	冻结 BrushNet；LoRA 微调
R2	边界接缝伪影	中	高	ρ_b 最强锚定
R3	显存不足	高	中	LoRA + 参考模型 CPU offload
R4	负样本质量不足	中	中	三层次 + 确保正负差距
R5	时序一致性退化	低	高	多帧偏好 + 运动模块冻结

7.2 实施时间线

总周期：**11–15 周**。

7.3 评估指标体系

第 8 章 结论

表 10: 实施路线图

优先级	周期	任务	依赖
P0	3–4 周	构建 mask-aware 偏好对数据集	基线模型 + 训练数据
P1	1–2 周	实现 Region-Reg-DPO 损失	P0 数据集
P2	2–3 周	LoRA + 显存优化 + 小规模验证	P1 + GPU 集群
P3	3–4 周	全量训练 + 消融实验	P2 验证通过
P4	2 周	推理优化协调	P3 模型

表 11: 评估指标

维度	指标	说明
补全质量	PSNR, SSIM, LPIPS	仅在 mask 区域内计算
时序一致性	光流一致性误差	相邻帧 mask 区域像素变化
幻觉抑制	幻觉检测率	VLM 检测补全区域异常物体
融合质量	边界梯度连续性	mask 边界处纹理/颜色一致性
人类评估	GSB 方法	A/B 对比（质量 + 时序平滑度）

8.1 总体判断

将 Region-Reg-DPO 融入 DiffuEraser 在理论上完全可行。两者共享扩散模型基础框架；DiffuEraser 现有的 L2 训练损失在形式上等价于 Reg-DPO 的 SFT 正则化项，为偏好学习提供了天然的 SFT 基础。然而，视频补全的局部性、Mask 依赖性和 Prior 注入机制要求对原始 Reg-DPO 进行深度适配。

8.2 核心创新总结

1. **Region-Reg-DPO 损失函数**：将帧空间分为洞内、边界、上下文三区，使用 α_r（随 DGR 衰减）和 ρ_r（常数不衰减）两组独立权重，实现全链路区域感知
2. **第三缺陷的发现**：区域间优化需求不对称被共享 DGR 掩盖——洞内快速收敛拖垮边界优化，这是分区 DPO 的特有问题
3. ρ_r **常数设计**：ρ_r 必须不依赖 DGR，否则 SFT 保护随 DGR 同步失效
4. **严格泛化性**：当所有区域权重统一时退化为标准 Reg-DPO

8.3 预期收益

- **幻觉抑制**：通过偏好学习显式教会模型” 不该生成什么”
- **纹理提升**：从”MSE 均值解” 转向” 人类偏好解”
- **时序稳定**：以闪烁序列为负样本学习平滑过渡
- **边界保护**：ρ_b 最强锚定确保 Blending 融合质量
- **工作区间拓宽**：多源负样本覆盖更广泛的失败模式

建议

启动融合项目，但**优先投入数据构建（P0）**。只有在偏好对质量达标后，才推进算法实现和训练。该路径有潜力将 DiffuEraser 从” 最小化重建误差” 升级为” 对齐人类偏好” 的视频补全系统，拓宽其目前极窄的最佳工作区间。