arXiv:2601.00796v1 [cs.CV] 2 Jan 2026

AdaGaR：动态场景重建自适应加博尔表示

Jiewen Chan[1] ZhenjunZhao[2] Yu-Lun Liu[1]1国立阳明交通大学 [2]萨拉戈萨大学

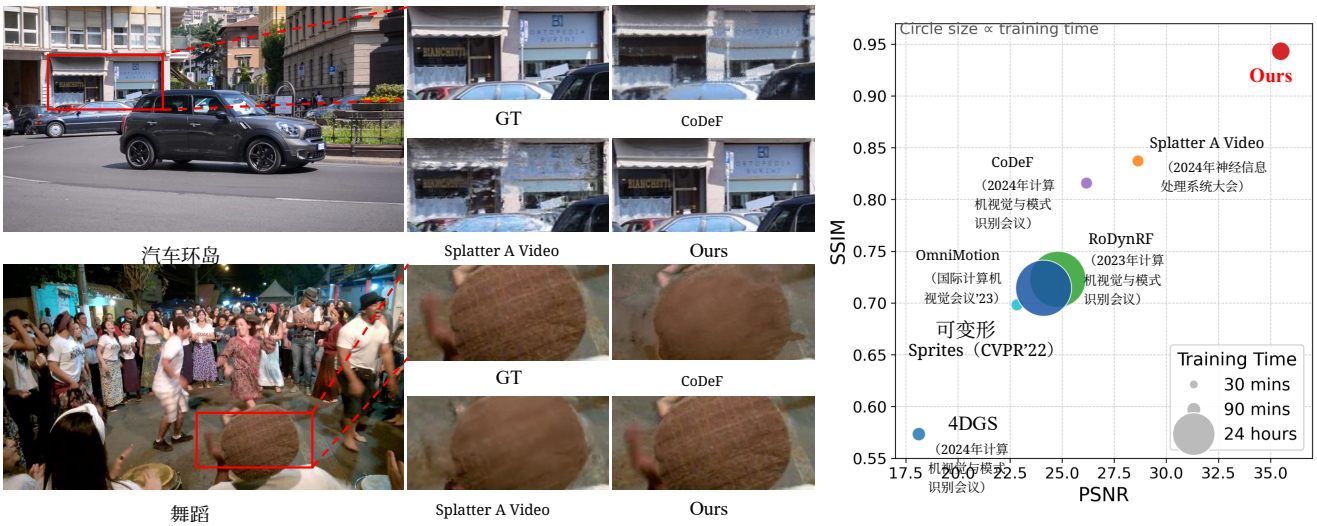


图 1. **DAVIS 上最先进的视频重建质量 [65] 数据集**。 我们的自适应 Gabor 表示在保留精细细节和时间一致性的同时，实现了卓越的渲染质量（峰值信噪比：35.49 dB，结构相似性指数：0.9433）。 （左） 定性比较表明，与 CoDeF 相比，我们的方法在具有挑战性的区域（汽车车窗、鼓面）呈现出更清晰的纹理 [62] 以及 Splatter A Video [76]。（右） 我们的方法（**红色** 点）在所有指标上均显著优于近期基线方法；与排名第二的方法相比，峰值信噪比提高了 6.86 dB，且训练时间合理（圆圈大小表示训练时长：30 分钟至 24 小时）。

摘要

从单目视频重建动态三维场景，需要同时捕捉高频外观细节和时间连续的运动。采用单一高斯基元的现有方法受限于其低通滤波特性，而标准 *Gabor* 函数会导致能量不稳定。此外，缺乏时间连续性约束往往会在插值过程中造成运动伪影。我们提出了***AdaGaR***，这是一个统一框架，旨在解决显式动态场景建模中的频率自适应性和时间连续性问题。我们提出自适应 *Gabor* 表示，通过可学习的频率权重和自适应能量补偿扩展高斯表示，从而兼顾细节捕捉与稳定性。为保证时间连续性，我们采用三次 *Hermite* 样条，并结合时间曲率正则化，确保运动平滑演变。自适应初始化机制融合深度估计、点跟踪和前景掩码，在训练初期建立稳定的点云分布。在 *Tap-Vid DAVIS* 上的实验表明，该方法达到了最先进的性能（峰值信噪比 *35.49*、结构相似性指数 *0.9433*、学习感知图像块相似度 *0.0723*），并且在帧插值、深度一致性、视频编辑和立体视图合成方面表现出强大的泛化能力。项目页面：*https://jiewenchan.github.io/AdaGaR/*

1. 引言

从单目视频重建动态三维场景是计算机视觉领域的一项基础挑战，在虚拟现实、增强现实和电影制作等领域有着广泛应用。关键难点在于同时实现时间连续性和丰富的频率表示：真实场景既要求运动随时间平滑变化，又要保留定义其外观的高频纹理。

现有方法 [6, 30, 41, 66, 75, 102] 分为两类。基于高斯基元的方法建模快速且显式，但存在严重的低通滤波问题，会抑制高频细节。引入频率调制 [87]（例如，类加博表示）可以提升纹理保真度，但往往会破坏能量平衡和渲染质量。此外，许多方法缺乏明确的

时间约束，导致运动不连续性和几何撕裂，尤其是在快速运动或发生遮挡时。

为弥补这些不足，我们提出了 ***AdaGaR***（面向动态场景重建的自适应 Gabor 表示），这是一个在显式动态表示中联合优化时间与频率的统一框架。我们的核心思路是将两个正交方面分开处理，同时又紧密耦合：(i) 通过可学习的自适应 Gabor 表示实现频率自适应，在平衡高频和低频分量的同时保持能量稳定性；(ii) 通过结合时间曲率正则化的三次 Hermite 样条实现时间连续性，约束运动轨迹平滑演化。此外，自适应初始化进一步为早期训练奠定稳定且时间连贯的几何基础。

我们在 Tap-Vid 上验证了 AdaGaR [65]，实现了最先进的视频重建效果，并在帧插值、深度一致性、视频编辑和立体视图合成方面展现出强大的泛化能力，如图所示。1。本研究为显式动态表示中的时间与频率建模提供了一种精简的端到端解决方案，有望为频率感知动态建模的未来发展提供指引。

我们的主要贡献总结如下：

•我们提出了一种新颖的 自适应 *Gabor* 表示，将传统高斯扩展到频域，如图 2所示。该方法具有以下特点：(i) 频率自适应；(ii) 能量稳定；(iii) 能够捕捉高频纹理细节，同时根据场景需求自动调整高频和低频分量；

• 我们引入时间曲率正则化 和三次 *Hermite* 样条插值，准确而有效地确保时间维度上的几何与运动连续性，实现平滑的时间演化并避免插值伪影；

• 我们提出一种自适应初始化 机制，结合深度估计、点跟踪和前景掩码，建立稳定且时间一致的点云分布，显著提升训练效率和最终重建质量。

2. 相关工作

动态三维高斯泼溅。3D 高斯泼溅（3DGS）[38] 推动了动态场景扩展方面的大量研究。早期工作[59]使用随时间变化的 MLP 进行形变。近期的规范空间方法 [2, 19, 27, 32, 53, 58, 85, 96]采用形变网络来处理压缩和镜面动态。时间建模策略包括流引导方法 [105], 、神经特征 [49], 、时间切片 [18],、时空正则化 [13, 45], 和哈希编码 [89]。专门的应用面向自动驾驶、稀疏重建 [71, 91, 104], 、无约束捕获 [61], 、加速 [40, 72], 和运动模糊 [77, 101], 。与我们的工作最相近的是 SplineGS [63]，它在多视图场景中应用三次厄米特样条。相比之下，我们将三次 *Hermite* 样条与基于加博尔的基元相结合，用于无需相机位姿估计的单目视频，并引入时间曲率正则化，以实现物理上合理的运动。

频率自适应渲染。传统 3D 高斯核充当低通滤波器，限制了高频细节的表示能力。3DGS 的抗锯齿方法包括多尺度滤波 [92, 98]，解析积分 [51],和不透明度场推导 [99]。神经辐射场的频率感知方法采用锥体追踪 [3], 频率正则化 [88, 93], 频率分解 [25], 以及结构与噪声分离 [67]。神经渲染中的 Gabor 表示 [1, 83, 87] 建立在程序化图形基础之上 [21, 44]。其他基元包括指数函数 [23], 表面元 [31], 和贝塔核 [54]。然而，现有的 Gabor 方法面向静态 场景，且使用固定频率。我们的自适应 Gabor 表示可扩展至动态 视频，采用可学习的频率权重，并能平滑退化为标准高斯函数。

时间建模与样条表示。经典样条 [20, 55] 可提供平滑的时间插值。近期的神经渲染通过埃尔米特形式 [15], B 样条 [63, 82],和时间调制权重 [22]引入了样条。粗细分解方法 [2, 90, 95] 可分离不同时间尺度。光流引导方法 [50, 60, 103, 105] 利用光流约束。其他时间模型包括卡尔曼滤波 [100],神经轨迹 [42], 帧插值 [33, 69],和鲁棒动态场 [57]。与架构或训练带来的隐式平滑性不同，我们显式通过基于二阶导数的时间曲率正则化来强制实现平滑性，从而确保运动在物理上合理且具有几何可解释性。

视频表示与规范空间。规范空间方法通过分层图集 [36], 变形场 [11, 62], 和规范体 [80]实现时间一致性处理。隐式神经视频表示 [8, 9, 39, 43, 47, 70, 73]通过逐图像函数实现压缩。显式表示采用四维高斯 [85], 2 D 特征流 [78], 层分解 [74], 学习式量化 [46], 哈希编码 [10],和场景修复 [84]。视频高斯泼溅方法针对单目 [6, 28, 30, 41, 52, 66, 75, 102] 和多视角 [12, 56] 场景。我们的方法在正交相机坐标系 [76], 中运行，无需姿态估计，同时借助 Gabor 基元保留高频信息，并维持显式三维结构，从而适用于多种应用。

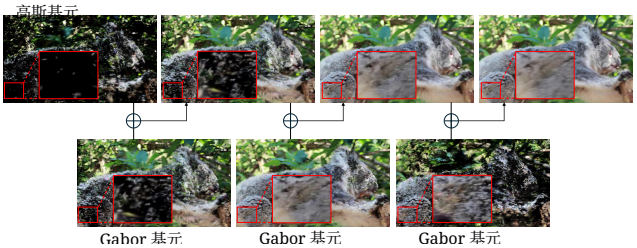


图 2. **分层频率自适应**。我们的基元可自适应地从高斯（左上）过渡到 Gabor（底部），实现由粗到细重建。每个基元通过可学习权重 ω_i 学习其最佳频率响应，在统一框架中兼顾几何稳定性与纹理细节。

单目深度与运动估计。 基础模型提供稳健的单目先验。深度估计方法 [5, 29, 37, 48, 64, 68, 79, 94]通过合成数据训练、扩散模型改造和多数据集学习实现零样本泛化。点跟踪方法 [14, 16, 17, 24, 34, 35, 81] 通过伪标签、自监督和局部相关性实现稠密对应关系。不同于以往独立使用这些信号的工作，我们的 自适应初始化联合推理深度、运动和分割，从而实现几何与时间上一致的初始化。

3. 预备知识：3D 高斯泼溅

3D 高斯泼溅（3DGS） [38]将三维场景表示为一组参数化的高斯基元 $\{\mathcal{G}_k \mid k = 1, \ldots, N\}$。每个 $\mathcal{G}_k$ 都有中心 $\boldsymbol{\mu}_k \in \mathbb{R}^3$、协方差$\boldsymbol{\Sigma}_k \in \mathbb{R}^{3\times3}$、不透明度 $\alpha_k \in [0, 1]$ 和颜色$\mathbf{c}_k$。其密度为

$$\mathcal{G}_k(\mathbf{x}) = \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right),$$

其中 $\boldsymbol{\Sigma}_k = \mathbf{R}_k\mathbf{S}_k\mathbf{S}_k^\top\mathbf{R}_k^\top$。

渲染会将高斯投影到图像平面上，并通过由前向后混合累积颜色：

$$C(\mathbf{x}) = \sum_{k=1}^{K} T_k\alpha_k\mathbf{c}_k, \quad T_k = \prod_{j<k}(1-\alpha_j).$$

一个关键局限在于，单个高斯相当于低通滤波器，会限制高频纹理细节。为了解决这一问题，我们引入 Gabor 核，作为高斯的周期性扩展，以增强空间频率表示。

4. 方法

4.1. 概述

我们提出 AdaGaR，这是一种显式的 3D 视频表示，能够保留高频外观，同时确保运动随时间平滑。如图 3 所示，视频被建模为正交相机坐标系中的一组动态自适应加博尔基元，其中基元编码空间纹理和结构，并通过三次厄米特样条插值时间演变，以确保几何一致性和时间一致性。自适应 Gabor 表示通过可学习的频率权重和能量补偿扩展了高斯基元，从而实现频率自适应细节捕获，同时保持能量稳定性。结合时间曲率正则化和多重监督损失，我们的方法实现了高视觉质量和稳健的时间一致性，并适用于帧插值、深度一致性、视频编辑等任务。

4.2. 自适应 Gabor 视频表示

相机坐标空间。 受 [80] 和 [76], 启发，我们采用正交相机坐标系，将宽度、高度和深度分别映射到 X、 Y 和 Z 轴，从而能够直接正交地表示三维视频结构。这样可以避免耗费大量计算的相机位姿估计和运动解耦，并将相机运动与物体运动视为同一类动态变化。视频表示为一组动态自适应 Gabor 基元，每个基元编码空间位置、时间变化和频率响应，并从固定的单位姿态进行渲染。

自适应 Gabor 表示。为了在图像平面上引入高频细节，可以将加博尔函数视为高斯函数的周期性扩展。其一般二维形式可定义为：

$$\mathcal{G}_{\text{Gabor}}(\mathbf{x}) = \exp\left(-\frac{1}{2}||\mathbf{x}-\boldsymbol{\mu}||^2_{\boldsymbol{\Sigma}^{-1}}\right)\cos(\mathbf{f}^\top\mathbf{x}+\phi), \quad (1)$$

其中，$\mathbf{x} = (x, y)^\top$ 表示图像平面坐标，$\mathbf{f} = (f_x, f_y)^\top$ 为中心频率向量， ϕ 表示相位偏移。该结构在高斯包络中引入正弦调制，使分布能够同时捕捉局部方向性纹理和高频细节变化。

为建模更丰富的频率分量，可以将多个加博尔波加权叠加：

$$S(\mathbf{x}) = \sum_{i=1}^{N}\omega_i\cos(f_i\langle\mathbf{d}_i,\mathbf{x}\rangle+\phi_i), \quad (2)$$

其中， $\omega_i \in \mathbb{R}$ 表示振幅权重， $f_i \in \mathbb{R}^+$ 表示频率幅值，而$\mathbf{d}_i \in \mathbb{R}^2$ 与 $||\mathbf{d}_i||_2 = 1$表示频率方向单位向量。这种结构会产生空间周期性纹理变化，与高斯结合后可呈现更丰富的纹理细节。

尽管 Gabor 结构增强了细节表示，但固定振幅的余弦调制会破坏高斯的能量稳定性。为解决这一问题，我们提出了自适应

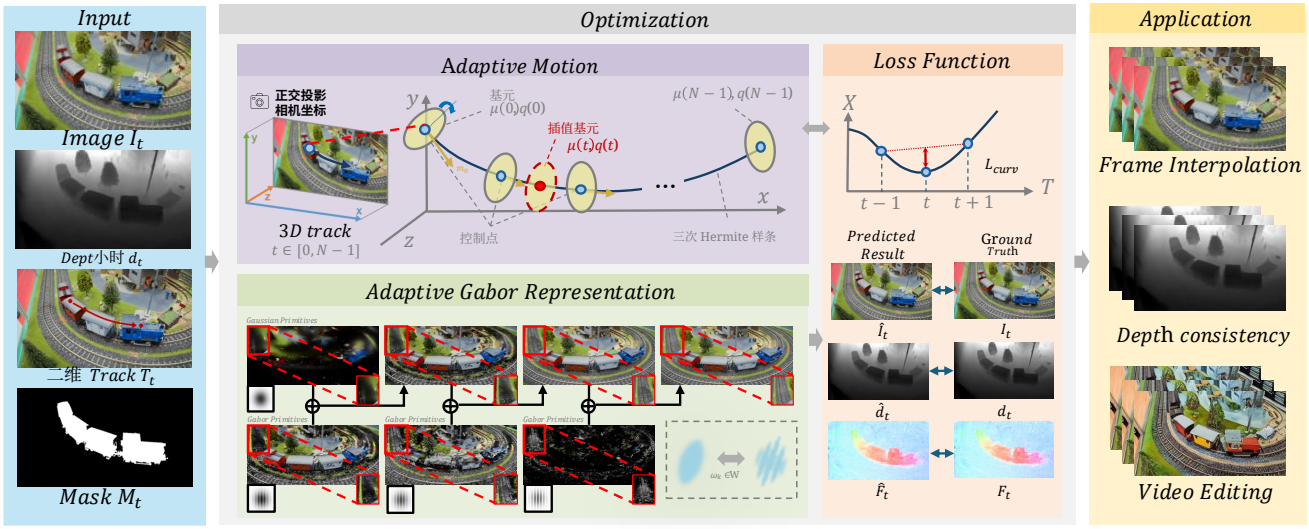


图 3. **方法概述**。我们的方法将动态视频表示为具有时间平滑运动的自适应 Gabor 基元。（输入）来自 RGB、深度、跟踪和掩码的多模态监督。（优化）两个核心组件：（1）自适应运动：使用三次 Hermite 样条对基元轨迹进行建模，控制点 $\mu(t)$、$q(t)$ 位于正交相机空间中，以确保 C^1 连续性。（2）自适应 *Gabor* 表示：可学习频率权重 ω_k 使基元能够自适应地覆盖从高斯（低频）到 Gabor（高频）的范围，从而实现分层细节重建。（损失）通过 RGB、深度、光流监督和曲率正则化 L_{curv}进行联合优化。（应用）支持帧插值、深度一致性和视频编辑。

Gabor，它会根据波能量自动调整强度，并在极端情况下自然退化为高斯。我们将原有的不透明度表达式扩展为$\alpha_{\text{Gabor}} = \mathcal{G}(\mathbf{x}) \cdot S(\mathbf{x})$。在实践中，我们设定相位项 $\phi_i = 0$，得到：

$$S(\mathbf{x}) = \sum_{i=1}^{N} \omega_i \cos(f_i \langle \mathbf{d}_i, \mathbf{x} \rangle), \tag{3}$$

其中我们固定频率参数 $f_i \in \{1, 2\}$，它们对应于两个正交基频波形。振幅权重 $\omega_i \in [0, 1]$ 是为 Gabor 结构引入的可学习参数，用于调整不同频率分量的能量权重。方向单位向量 $\mathbf{d}_i$ 与原始高斯的空间方向共享，从而确保频率调制与高斯形状方向的一致性。

为防止出现整体强度衰减，我们引入补偿项 b:

$$S_{adap}(\mathbf{x}) = b + \frac{1}{N} \sum_{i=1}^{N} \omega_i \cos(f_i \langle \mathbf{d}_i, \mathbf{x} \rangle), \tag{4}$$

$$b = \gamma + (1 - \gamma) \left(1 - \frac{1}{N} \sum_{i=1}^{N} \omega_i \right), \tag{5}$$

其中，$\gamma \in [0, 1]$ 是控制退化平滑度的固定超参数，而因子 $1/N$ 将多个波的加权平均归一化到稳定范围。当 $\omega_i \to 0$时，我们有 $b \to 1$，此时该公式自然退化为传统高斯。

4.3. 时变自适应 Gabor

三次 Hermite 样条插值。我们使用三次 Hermite 样条 [7, 26] 来插值动态基元的时间演化。给定在时刻 $\{t_0, t_1, \ldots, t_{M-1}\}$ 处的 M 个时间关键帧及其对应的控制点位置 $\{\mathbf{y}_0, \mathbf{y}_1, \ldots\ldots, \mathbf{y}_{M-1}\} \subset \mathbb{R}^3$，我们将相邻关键帧之间的时间间隔定义为 $\Delta_k = t_{k+1} - t_k$，并将斜率定义为 $\delta_k = (\mathbf{y}_{k+1} - \mathbf{y}_k)/\Delta_k$。为避免关键帧之间产生不必要的振荡，我们引入带有单调门控的自动斜率机制：

$$\mathbf{m}_k = \begin{cases} \beta \cdot \frac{\delta_{k-1} + \delta_k}{2}, & \text{if } \operatorname{sign}(\delta_{k-1}) = \operatorname{sign}(\delta_k), \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{6}$$

其中，$\beta \in (0, 1]$ 是控制插值曲线平坦度的平滑度系数。这种设计可避免关键帧处出现反向振荡，并确保视觉上平滑的插值。

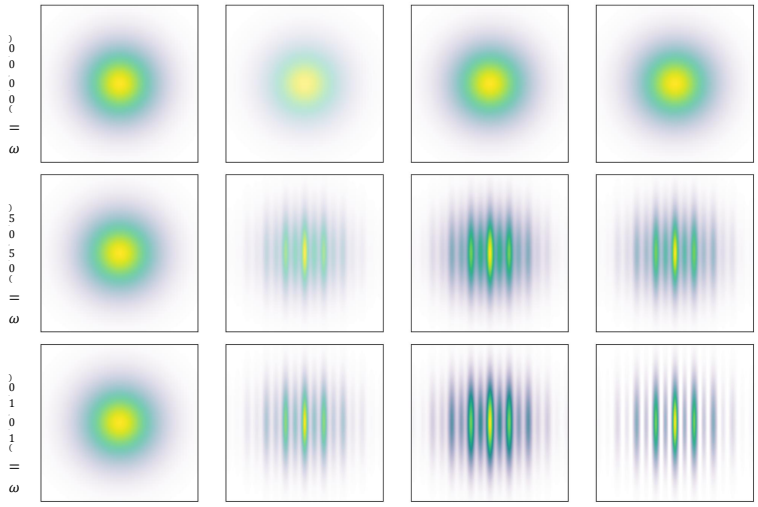
厄米特基函数定义如下：

$$\begin{aligned} H_{00}(s) &= 2s^3 - 3s^2 + 1, \quad H_{10}(s) = s^3 - 2s^2 + s, \\ H_{01}(s) &= -2s^3 + 3s^2, \quad H_{11}(s) = s^3 - s^2, \end{aligned} \tag{7}$$

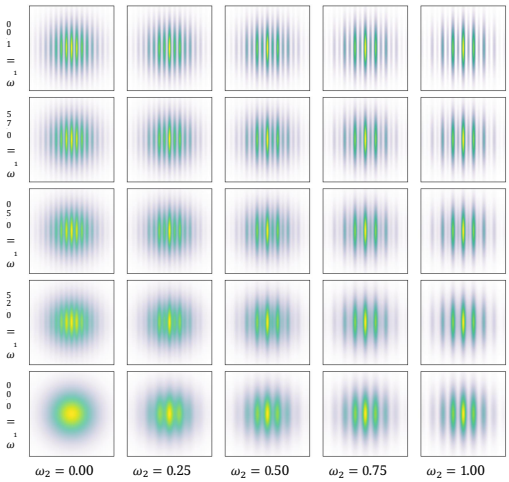
其中，$s = (t - t_k)/\Delta_k \in [0, 1]$ 是区间 $[t_k, t_{k+1}]$内的归一化时间。在时间 t 处的插值位移为：

$$\begin{aligned} \mathbf{\Delta}(t) = {} & H_{00}(s)\mathbf{y}_k + H_{10}(s)\Delta_k \mathbf{m}_k \\ & + H_{01}(s)\mathbf{y}_{k+1} + H_{11}(s)\Delta_k \mathbf{m}_{k+1}. \end{aligned} \tag{8}$$

为确保几何连续性一致，最终位置通过将插值位移加到



(a) $G(x)$高斯， $S(x)$，正弦部分



(b) 不同组合的影响 Gabor波系数（ω_0 和 ω_1）对空间频率纹理的影响。

图4。**自适应Gabor公式。(a) 高斯核与Gabor核之间的平滑过渡**。我们的方法（最右列，$S_{\text{ours}}(x)$）使用补偿项 b，在从纯高斯（$\omega = 0$，顶部）过渡到频率调制 Gabor（$\omega = 1$，底部）的过程中保持能量稳定性。朴素组合 $1 + S(x)$（第三列）会产生强度伪影。**(b) 频率权重组合**。不同的（ω_0，ω_1）配对可生成多样的空间模式，从平滑模式（低 ω）到高频纹理（高 ω），从而能够在不同场景区域自适应地捕捉细节。

基准位置上得到：

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}_{\text{base}} + \boldsymbol{\Delta}(t). \tag{9}$$

旋转插值。我们将相同原理扩展到旋转的时间插值。对于旋转参数，我们首先在 $\mathfrak{so}(3)$李代数空间中进行插值，然后通过指数映射转换为单位四元数：

$$\mathbf{q}(t) = \text{normalize}\big(\text{normalize}(\mathbf{q}_{\text{base}}) \otimes \exp(\boldsymbol{\Delta}_{\mathbf{q}}(t))\big), \tag{10}$$

其中，$\otimes$ 表示四元数乘法，角度环绕可确保旋转角保持在 $(-\pi,\ \pi]$。

时间曲率正则化。为确保时间演化的平滑性，我们在每个关键帧的轨迹上引入曲率惩罚。对于非均匀关键帧，二阶导数估计为

$$\mathbf{y}''_k = \frac{2(\mathbf{d}^+_k - \mathbf{d}^-_k)}{h_{k-1} + h_k}, \tag{11}$$

其中，$h_{k-1} = t_k - t_{k-1}$、$h_k = t_{k+1} - t_k$、$\mathbf{d}^+_k = (\mathbf{y}_{k+1} - \mathbf{y}_k)/h_k$、$\mathbf{d}^-_k = (\mathbf{y}_k - \mathbf{y}_{k-1})/h_{k-1}$以及 $D = 3$。曲率损失为

$$\mathcal{L}_{\text{curve}} = \frac{\sum_{k=1}^{M-2} w_k \|\mathbf{y}''_k\|_2^2}{\sum_{k=1}^{M-2} w_k D + \varepsilon}, \tag{12}$$

其中 $w_k = \frac{1}{2}(h_{k-1} + h_k)$，而 $\varepsilon > 0$ 是一个小常数。该项通过惩罚时间上的二阶能量来确保平滑性。

4.4. 优化

为在动态场景中同时保持逼真的外观和时间稳定性，我们采用多目标损失函数，对外观保真度、运动一致性、深度几何和时间平滑性进行约束。

渲染重建损失。我们结合 $\mathcal{L}_1$ 和结构相似性指数，以同时保留像素级准确性和结构特征：

$$\mathcal{L}_{\text{rgb}}(I_t, \hat{I}_t) = (1 - \lambda_{\text{ssim}})\mathcal{L}_1^{\text{rgb}}(I_t, \hat{I}_t) + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}}^{\text{rgb}}(I_t, \hat{I}_t), \tag{13}$$

其中，I_t 和 $\hat{I}_t$ 分别表示第 t帧的真实值图像和预测图像。

光流一致性损失。我们利用 Co-Tracker [34]提供跨帧监督。自适应 Gabor 基元的投影位置与 2D 轨迹对齐，并使用可见性加权 $\mathcal{L}_1$ 损失：

$$\mathcal{L}_{\text{flow}}(\hat{F}_{t_1,t_2}, F_{t_1,t_2}) = \frac{\sum_j w_j \left\| \hat{\mathbf{x}}^j_{t_2} - \mathbf{x}^j_{t_2} \right\|_1}{\sum_j w_j + \varepsilon}, \tag{14}$$

其中，$\mathbf{x}^j_{t_2}$ 和 $\hat{\mathbf{x}}^j_{t_2}$ 分别表示第 j 个跟踪点在帧 t_2 中的真实值像素位置和预测像素位置，w_j 表示其可见性权重。

深度损失。我们使用来自 DPT [68] 的单目深度估计作为具有尺度和平移不变性的几何先验

对齐：

$$\mathcal{L}_{\text{depth}}(D_t, \hat{D}_t) = \left\| \gamma(D_t) - \gamma(\hat{D}_t) \right\|_1, \tag{15}$$

其中 $\gamma(D_t) = (D_t - c_t(D_t))/\|D_t - c_t(D_t)\|_1$ 与$c_t(D_t)$ = 中位数(D_t)。

总损失。 总体优化目标结合了所有组成部分：

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rgb}}\mathcal{L}_{\text{rgb}} + \lambda_{\text{flow}}\mathcal{L}_{\text{flow}} + \lambda_{\text{depth}}\mathcal{L}_{\text{depth}} + \lambda_{\text{curv}}\mathcal{L}_{\text{curv}}. \tag{16}$$

这种多方面的监督使 AdaGaR 能够实现高保真渲染和时间稳定的动态场景表示。

4.5. 自适应初始化

我们提出了一种自适应初始化方法，以便在训练早期初始化时间连贯的三维点分布。该方法融合多模态线索，从输入视频中生成稠密、动态的初始点云，为后续显式表示奠定几何基础。与随机采样或单帧方法不同，我们的方法会根据场景运动和深度分布自适应调整采样密度，确保前景和背景均得到均衡覆盖。

时空自适应采样。 对于每个候选点 $\mathbf{p}_i$，其采样概率为

$$\Pi(\mathbf{p}_i) \propto \frac{1}{\tau_i + \epsilon} + \lambda_\tau \frac{1}{\rho_i + \epsilon}, \tag{17}$$

其中，τ_i 为时间支持度，ρ_i 为局部密度，$\lambda_\tau \in [0, 1]$ 用于平衡时间稳定性与空间均匀性，而$\epsilon > 0$。

基于网格的均匀覆盖。 为确保全局覆盖，我们将图像划分为固定网格 $\mathcal{G} = \{G_{u,v}\}$，并通过以下方式调节每个网格单元的采样

$$\Pi'(\mathbf{p}_i \mid G_{u,v}) = \frac{\Pi(\mathbf{p}_i)}{1 + \lambda_g C_{u,v}}, \tag{18}$$

其中，$C_{u,\,v}$ 为该网格单元的累积采样数，$\lambda_g > 0$。

边界感知补偿。 我们进一步通过以下方式对运动边界进行调整

$$\Pi''(\mathbf{p}_i) = \Pi'(\mathbf{p}_i \mid G_{u,v})\,(1 + \lambda_b \|\nabla M_t(\mathbf{p}_i)\|), \tag{19}$$

其中，M_t 为前景掩码，$\lambda_b > 0$。

该方案生成稠密且时间连贯的初始点云，并减少早期闪烁。

5. 实验

5.1. 评估

数据集与指标。 我们在 Tap-Vid DAVIS [65],上进行评估，该数据集包含丰富多样的动态场景和遮挡。定量指标包括峰值信噪比、结构相似性指数和学习感知图像块相似度，分别用于评估像素精度、结构保真度和感知质量。

表 1. **Tap-Vid DAVIS 上的定量结果 [65]**。我们的方法在所有指标上均取得了最先进的性能，PSNR 比此前最佳方法高出 6.86 dB [76], ，验证了频率自适应基元与平滑时序建模的有效性。

方法	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓
4DGS [86]	18.12	0.5735	0.5130
RoDynRF [57]	24.79	0.7230	0.3940
可变形精灵 [97]	22.83	0.6983	0.3014
OmniMotion [80]	24.11	0.7145	0.3713
CoDeF [62]	26.17	0.8160	0.2905
Splatter A Video [76]	28.63	0.8373	0.2283
Ours	35.49	0.9433	0.0723

实现细节。训练分为两个阶段：500 次迭代的预热阶段，以及 10K 次迭代的主要优化阶段；控制点每 100 次迭代更新一次。实验在 NVIDIA RTX 4090 上运行，每段视频序列耗时 90 分钟。

视频重建。如表1所示，我们的方法在 Tap- VidDAVIS 上的 PSNR/SSIM/LPIPS 指标均优于基线方法 [65]。与基于多层感知机的表示相比，我们的方法在图5中呈现出更清晰的纹理和连贯的运动。

5.2. 应用

深度一致性。如图6所示，我们在整个过程中实现了稳定的深度分布，大幅减少了深度闪烁和边界错位，并优于逐帧优化器。

帧插值。如图7所示，我们使用带曲率正则化的三次厄米特样条，在关键帧之间生成平滑的中间帧，同时保留纹理细节并避免边界伪影。

视频编辑。在规范空间中，通过对共享的自适应 Gabor 基元进行操作，风格迁移保持了时间上的连贯性，减少了风格漂移和闪烁，如图 8所示。

立体视图合成。 我们的显式表示支持从单目输入进行立体合成，并具有更好的视差一致性和合理的几何结构，如图 9 所示。

5.3. 消融研究

自适应 Gabor 表示。 我们使用相同的 1M 个基元，将自适应 Gabor 表示（AGR）与高斯表示和标准 Gabor 表示进行比较。如表 2 所示，AGR 提升了高频细节和能量稳定性，在三种配置中取得了最佳的峰值信噪比/结构相似性指数/学习感知图像块相似度。

样条插值。我们在表3中对 50 帧进行曲线插值消融实验。B 样条和三次样条提供了一定的时间连续性，但难以处理非线性运动。相比之下，所提出的三次 Hermite 样条实现了

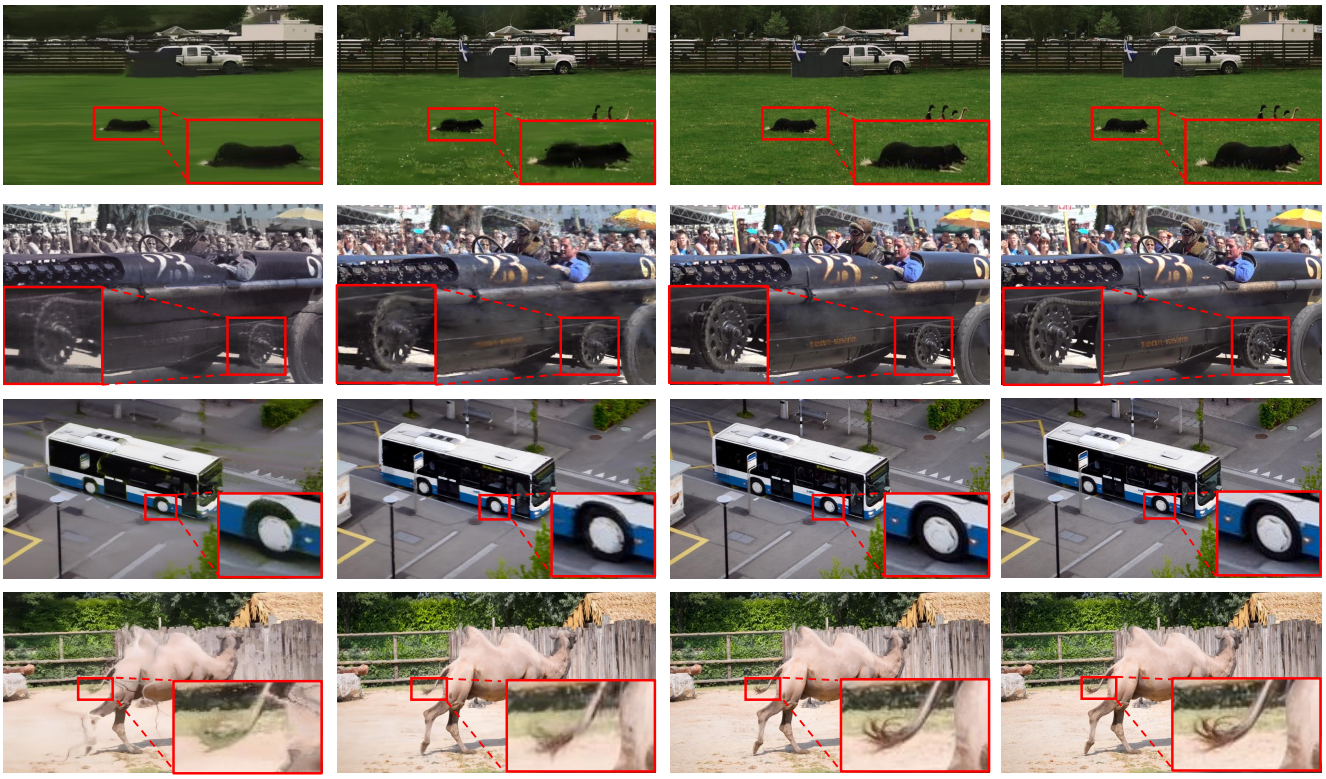


图 5. **DAVIS 数据集上的视觉比较**。 与 CoDeF 相比，我们的方法保留了更精细的细节（毛发、车辆边缘、车轮结构），并呈现出更清晰的运动边界 [62] 以及 Splatter A Video [76]。 红色 方框标出了关键区域，展示了我们在纹理重建和时间一致性方面的优势。放大后观看效果最佳。

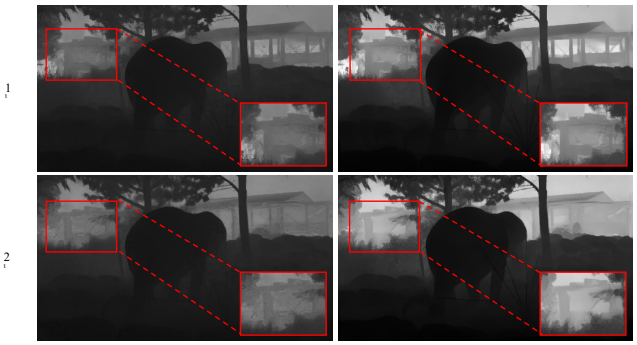


图 6.**跨时间的深度一致性**。 （左）我们的三维基元表示能够在不同帧中为静态元素保持一致的深度。 （右）逐帧估计（Marigold [37]）会出现时间闪烁（红色 方框）。通过显式三维几何和平滑运动建模，可确保时间连贯性，这对基于深度的视频应用至关重要。

最佳的综合指标表现，同时生成更平滑的轨迹并保留动态细节。

曲率正则化。 我们在图 $\mathcal{L}_{\text{curv}}$ 中比较了有无时间曲率项 10。没有 $\mathcal{L}_{\text{curv}}$，mo-

表 2. **Gabor 基元消融**。 带补偿项的自适应 Gabor b 优于标准高斯函数和朴素 Gabor 变体，验证了能量感知公式（式（5））对于稳定的频率建模至关重要。

方法	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓
高斯	36.66	0.9423	0.0421
标准 Gabor（$b = 0$）	36.65	0.9543	0.0345
$1 + S\ (x)$	36.50	0.9511	0.0322
自适应 Gabor（我们的方法）	**37.43**	**0.9620**	**0.0242**

表 3. **样条方法消融**。 我们的方法中的带单调门控三次 Hermite 样条优于 B 样条，并显著优于标准三次样条；后者会出现轨迹振荡。显式速度控制（式（6））对于实现平滑、无伪影的运动建模至关重要。

方法	峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓
B 样条	36.68	0.9573	0.0368
三次样条	32.42	0.9073	0.0818
三次 Hermite 样条（我们的方法）	**38.98**	**0.9697**	**0.0259**

出现伪影和撕裂现象，而使用 $\mathcal{L}_{\text{curv}}$ 时，插值更加平滑稳定，证实了其必要性



图 7. **帧插值结果**。 我们的方法通过在分数时间戳处查询三次 Hermite 样条，在输入关键帧 t 和 $t+1$ 之间生成时间上平滑的中间帧。插值序列（从第 1st 帧到第 4th 帧）保持了毛发纹理细节和自然运动，且没有重影伪影。红色方框标示了整个插值过程中高频细节的保留情况。这表明，我们的方法能够通过曲率正则化样条轨迹，以 C^1 平滑度实现连续运动。完整的时间连贯性请参阅补充视频。



图 8. **时间一致的视频编辑**。 (上) 逐帧编辑会导致帧间风格不一致，产生时间闪烁。(下) 我们在规范空间中的编辑通过对共享的自适应 Gabor 基元应用风格迁移，确保不同时间的场景元素受到一致处理，同时保留运动动态，从而维持时间一致性。 红色 方框突出显示了关键差异。请参阅补充视频。



图 9. **立体视图合成**。我们的三维表示能够根据单目视频进行立体可视化的新视角合成。这表明，正交相机坐标空间中的自适应 Gabor 基元能够捕捉精确的三维几何，从而支持沉浸式应用。

显式曲率控制。

自适应初始化。 我们将随机初始化与自适应初始化进行了比较。如图 11所示，自适应方法生成的初始几何更加稠密且在时间上连贯，从而减少闪烁并提升重建初期的质量。

6. 结论

我们提出 AdaGaR，这是一种面向动态场景建模的统一框架，可实现时间连续性和频率自适应性。通过将高斯基元扩展为自适应 Gabor



图 10. **曲率正则化消融实验**。（左）在缺少$\mathcal{L}_{\text{curv}}$时，插值帧会因轨迹振荡而出现运动伪影。（右）我们的方法通过约束二阶导数，实现平滑且无伪影的插值，验证了显式曲率控制对于时间一致性的必要性。



图 11。**自适应初始化消融实验**。（左）缺少运动感知初始化时，基元分布不佳，导致细节模糊。（右）我们基于深度、跟踪和掩码提出的自适应初始化（公式（17）至（19））能够提供更好的初始几何，使重建效果清晰，并提升 6.78 dB。

表示，并采用结合时间曲率正则化的三次厄米特样条，我们的方法在确保几何与运动连续性的同时，能够捕捉高频细节。实验表明，该方法在 Tap-Vid DAVIS 上达到了最先进的性能，并在帧插值、深度一致性、视频编辑和立体合成等任务中表现出强大的泛化能力。

局限性。 尽管 AdaGaR 性能优异，但仍存在局限性。基于样条的运动建模假设轨迹平滑，因此在突发或高度非线性运动下可能导致错位。此外，受能量约束影响，自适应 Gabor 表示在高频区域可能出现振荡。未来的研究可引入自适应时间控制点和运动感知频率调制。

致谢。 本研究由台湾国家科学及技术委员会资助，项目编号为 NSTC 112-2222-E-A49-004-MY2 和 113-2628-E-A49-023-。作者感谢 Google、NVIDIA 和联发科技股份有限公司慷慨捐赠。Yu-Lun Liu 感谢台湾教育部玉山青年学者计划的支持。

参考文献

[1] Ahmad AlMughrabi、Ricardo Marques 和 Petia Radeva。Momentsnerf：利用正交矩实现小样本神经渲染。arXiv 预印本 arXiv:2407.02668，2024。2[2] Jeongmin Bae、Seoha Kim、Youngsik Yun、Hahyun Lee、Gun Bang 和 Youngjung Uh。基于逐高斯嵌入的可变形三维高斯泼溅形变。载于《欧洲计算机视觉会议》，第 321–335 页。施普林格，2024。2[3] Jonathan T Barron 、Ben Mildenhall、Dor Verbin、Pratul P Srinivasan 和 Peter Hedman。Zip-NeRF：抗锯齿网格神经辐射场。载于《IEEE/CVF国际计算机视觉会议论文集》，第 19697–19705 页，2023。2[4] Yoshua Bengio、Nicholas L´ eonard 和 Aaron Courville。针对条件计算，估计或传播随机神经元的梯度。arXiv 预印本 arXiv:1308.3432，2013。14.[5] Aleksei Bochkovskii、AmaAG.l Delaunoy、Hugo Germain、Marcel Santos、Yichao Zhou、Stephan R Richter 和 Vladlen Koltun。Depth Pro：不到一秒即可实现清晰的单目度量深度。arXiv 预印本 arXiv:2410.02073，2024。3[6] Minh-Quan Viet Bui、Jongmin Park、Juan Luis Gonzalez Bello、Jaeho Moon、Jihyong Oh 和 Munchurl Kim。MoBGS：针对模糊单目视频的运动去模糊动态三维高斯泼溅。arXiv 预印本 arXiv:2504.15122，2025。1, 2[7] AKB Chand 和 P Viswanathan。三次埃尔米特与三次样条分形插值函数。载于《AIP会议论文集》，第 1467–1470 页。美国物理联合会，2012。4[8] Hao Chen、Bo He、Hanyu Wang、Yixuan Ren、Ser Nam Lim 和 Abhinav Shrivastava。NeRV：视频神经表示。《神经信息处理系统进展》，34：21557–21568，2021。2[9] Hao Chen、Matthew Gwilliam、Ser-Nam Lim 和 Abhinav Shrivastava。HNeRV：视频混合神经表示。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第 10270–10279 页，2023。2[10] Jie Chen、Zhangchi Hu、Peixi Wu、Huyue Zhu、Hebei Li 和 Xiaoyan Sun。DASH：结合自监督分解的四维哈希编码，用于实时动态场景渲染。载于《IEEE/CVF国际计算机视觉会议论文集》，第 26349–26359 页，2025。2[11] Ting-Hsuan Chen、Jie Wen Chan、Hau-Shiang Shiu、Shih-Han Yen、Changhan Yeh 和 Yu-Lun Liu。NARCan：自然

融入扩散先验的细化规范图像，用于视频编辑。《神经信息处理系统进展》，37:36097–36120，2024。2[12] Yuedong Chen、Haofei Xu、Chuanxia Zheng、Bohan Zhuang、Marc Pollefeys、Andreas Geiger、Tat-Jen Cham 和 Jianfei Cai。Mvsplat：从稀疏多视图图像高效实现 3D 高斯泼溅。载于《欧洲计算机视觉会议》，第 370–386 页。施普林格，2024。2[13] Hao-Jen Chien、Yi-Chuan Huang、Chung-Ho Wu、Wei-Lun Chao 和 Yu-Lun Liu。Splannequin：利用双检测泼溅冻结单目人体模型挑战视频素材。arXiv 预印本 arXiv:2512.05113，2025。2[14] Seokju Cho、Jiahui Huang、Jisu Nam、Honggyu An、Seungryong Kim 和 Joon-Young Lee。用于点跟踪的局部全对对应关系。载于 ECCV，第 306–325 页，2024。3[15] Ilya Chugunov、David Shustin、Ruyu Yan、Chenyang Lei 和 Felix Heide。用于连拍图像融合和层分离的神经样条场。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 25763–25773 页，2024。2[16] Carl Doersch、Yi Yang、Mel Vecerik、Dilara Gokay、Ankush Gupta、Yusuf Aytar、Joao Carreira 和 Andrew Zisserman。Tapir：通过逐帧初始化和时间细化实现任意点跟踪。载于 IEEE/CVF国际计算机视觉会议论文集，第 10061–10072 页，2023。3[17] Carl Doersch、Pauline Luc、Yi Yang、Dilara Gokay、Skanda Koppula、Ankush Gupta、Joseph Heyward、Ignacio Rocco、Ross Goroshin、Joao Carreira 等。Bootstap：用于任意点跟踪的自助训练。载于《亚洲计算机视觉会议》，第 3257–3274 页，2024。3[18] Yuanxing Duan、Fangyin Wei、Qiyu Dai、Yuhang He、Wenzheng Chen 和 Baoquan Chen。四维转子高斯泼溅：面向动态场景的高效新视角合成。载于 ACM SIGGRAPH 2024 会议论文集，第 – 页，2024。1 2[19] Cheng-De Fan、Chen-Wei Chang、Yi-Ruei Liu、Jie-Ying Lee、Jiun-Long Huang、Yu-Chee Tseng 和 Yu-Lun Liu。Spectromotion：镜面场景的动态三维重建。载于计算机视觉与模式识别会议论文集，第 21328 21338, 2025 2 – 页。Gerald E [20] Farin。《CAGD曲线与曲面：实用指南》。Morgan Kaufmann，2002。2[21] Bruno Galerne、Ares Lagae、Sylvain Lefebvre 和 George Drettakis。通过示例生成 Gabor噪声。ACM图形学汇刊（ToG），31(4):1–9，2012。2[22] Ivan Grega、William F Whitney 和 Vikram S Deshpande。神经渲染支持动态断层成像。arXiv 预印本 arXiv:2410.20558，2024。2[23] Abdullah Hamdi、Luke Melas-Kyriazi、Jinjie Mai、Guocheng Qian、Ruoshi Liu、Carl Vondrick、Bernard Ghanem 和 Andrea Vedaldi。Ges：用于高效辐射场渲染的广义指数泼溅。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 19812–19822 页，2024。2[24] Adam W Harley、Yang You、Xinglong Sun、Yang Zheng、Nikhil Raghuraman、Yunqi Gu、Sheldon Liang、Wen-Hsuan

Chu、Achal Dave、Suya You 等。Alltracker：高分辨率下的高效稠密点跟踪。载于 IEEE/CVF国际计算机视觉会议论文集，第 5253–5262 页，2025。3[25] Yisheng He、Weihao Yuan、Siyu Zhu、Zilong Dong、Liefeng Bo 和 Qixing Huang。Freditor：通过频率分解实现高保真且可迁移的神经辐射场编辑。载于《欧洲计算机视觉会议》，第 73–91 页。施普林格，2024。2[26] Niels T Hintzen、Gerjan J Piet 和 Thomas Brunel。使用位置配准数据的三次 Hermite 样条插值改进拖网轨迹估计。《渔业研究》，101(1-2):108–115，2010。4[27] Cheng-Yuan Ho、He-Bi Yang、Jui-Chiu Chiang、Yu-Lun Liu 和 Wen-Hsiao Peng。Ted-4dgs：用于 4DGS压缩的时间激活与基于嵌入的形变。arXiv 预印本 arXiv:2512.05446，2025。2[28] Hao-Yu Hou、Chia-Chi Hsu、Yu-Chen Huang、Mu-Yi Shen、Wei-Fang Sun、Cheng Sun、Chia-Che Chang、Yu-Lun Liu 和 Chun-Yi Lee。具有分组不确定性的 3D 高斯泼溅无约束图像。载于 ICASSP 2025-2025 IEEE国际声学、语音与信号处理会议（ICASSP），第 1–5 页。IEEE，2025。2[29] Mu Hu、Wei Yin、Chi Zhang、Zhipeng Cai、Xiaoxiao Long、Hao Chen、Kaixuan Wang、Gang Yu、Chunhua Shen 和 Shaojie Shen。Metric3D v2：用于零样本度量深度和表面法线估计的通用单目几何基础模型。IEEE模式分析与机器智能汇刊，2024。3[30] Shoukang Hu、Tao Hu 和 Ziwei Liu。Gauhuman：从单目人体视频进行关节式高斯泼溅。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 20418–20431 页，2024。1, 2[31] Binbin Huang、Zehao Yu、Anpei Chen、Andreas Geiger 和 Shenghua Gao。用于几何精确辐射场的 2D高斯泼溅。载于 ACM SIGGRAPH 2024 会议论文集，第 1–11 页，2024。2[32] Yi-Hua Huang、Yang-Tian Sun、Ziyi Yang、Xiaoyang Lyu、Yan-Pei Cao 和 Xiaojuan Qi。Sc-gs：用于可编辑动态场景的稀疏控制高斯泼溅。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 4220–4230 页，2024。2[33] Zhewei Huang、Tianyuan Zhang、Wen Heng、Boxin Shi 和 Shuchang Zhou。用于视频帧插值的实时中间光流估计。载于《欧洲计算机视觉会议》，第 624–642 页。施普林格，2022。2[34] Nikita Karaev、Ignacio Rocco、Benjamin Graham、Natalia Neverova、Andrea Vedaldi 和 Christian Rupprecht。Co-tracker：共同跟踪效果更佳。载于 ECCV，第 18 35 页，2024 3, 5 – 。[35] Nikita Karaev、Yuri Makarov、Jianyuan Wang、Natalia Neverova、Andrea Vedaldi 和 Christian Rupprecht。Co-tracker3：通过对真实视频进行伪标注，实现更简单、更优的点跟踪。载于 IEEE/CVF国际计算机视觉会议论文集，第 6013 6022 页，2025 3 – 。

[36] Yoni Kasten、Dolev Ofri、Oliver Wang 和 Tali Dekel。《用于一致视频编辑的分层神经图集》。ACM图形学汇刊（TOG），40(6):1–12，2021。2[37] Bingxin Ke、Anton Obukhov、Shengyu Huang、Nando Metzger、Rodrigo Caye Daudt 和 Konrad Schindler。《重新利用基于扩散的图像生成器进行单目深度估计》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 9492–9502 页，2024。3, 7 – 。[38] Bernhard Kerbl、Georgios Kopanas、Thomas Leimk¨uhler 和 George Drettakis。《用于实时辐射场渲染的 3D 高斯泼溅》。ACM图形学汇刊，42(1)，2023，139– 。2, 3（）：– 。[39] Jina Kim 、 Jihoo Lee 和 Je-Won Kang。《 SNERV：用于视频的保谱神经表示》。载于欧洲计算机视觉会议，第 332–348 页。施普林格，2024。2[40] Mijeong Kim、Jongwoo Lim 和 Bohyung Han。《通过不确定性感知正则化实现野外 4D 高斯泼溅》。神经信息处理系统进展，37：129209–129226，2024。2[41] Muhammed Kocabas、Jen-Hao Rick Chang、James Gabriel、Oncel Tuzel 和 Anurag Ranjan。《HUGS：人体高斯泼溅》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 505–515 页，2024。1, 2 – 。[42]Agelos Kratimenos、Jiahui Lei 和 Kostas Daniilidis。《DynMF：利用三维高斯泼溅实现实时动态视图合成的神经运动分解》。载于欧洲计算机视觉会议，第 252–269 页。施普林格，2024。2[43] Ho Man Kwan、Ge Gao、Fan Zhang、Andrew Gower 和 David Bull。《 NVRC：神经视频表示压缩》。神经信息处理系统进展，37：132440–132462，2024。2[44] Ares Lagae、Sylvain Lefebvre、George Drettakis 和 Philip Dutr´e。《使用稀疏 Gabor卷积的程序化噪声》。ACM图形学汇刊（TOG），28(3):1–10，2009。2[45] Deqi Li、Shi-Sheng Huang、Zhiyuan Lu、Xinran Duan 和 Hua Huang。《ST-4DGS：用于高效动态场景渲染的时空一致 4D 高斯泼溅》。载于 ACM SIGGRAPH

2024 1 11, 2024 2 会

[46]

议论文集，第 – 页。Jiahao Li、Bin Li 和 Yan Lu。《通过特征调制实现神经视频压缩》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 26099–26108 页，2024。2[47] Zizhang Li、Mengmeng Wang、Huaijin Pi、Kechun Xu、Jianbiao Mei 和 Yong Liu。《E-NERV：利用解耦时空上下文加速神经视频表示》。载于欧洲计算机视觉会议，第 267–284 2022 2 页。施普林格，。《Zhenyu Li、

[48]

Shariq Farooq Bhat 和 Peter Wonka。《Patch-fusion：用于高分辨率单目度量深度估计的端到端基于瓦片的框架》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 10016–10025 页，2024。3[49] Zhan Li、Zhang Chen、Zhong Li 和 Yi Xu。《用于实时动态视图合成的时空高斯特征泼溅》。

载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8508–8520 页，2024。2[50] Feng Liang、Bichen Wu、Jialiang Wang、Licheng Yu、Kunpeng Li、Yinan Zhao、Ishan Misra、Jia-Bin Huang、Peizhao Zhang、Peter Vajda 等。《Flowvid：驾驭不完美光流，实现一致的视频到视频合成》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8207–8216 页，2024。2[51] Zhihao Liang、Qi Zhang、Wenbo Hu、Lei Zhu、Ying Feng 和 Kui Jia。《Analytic-splatting：通过解析积分实现抗混叠 3D 高斯泼溅》。载于欧洲计算机视觉会议，第 281–297 页。施普林格，2024。2[52] Chin-Yang Lin、Cheng Sun、Fu-En Yang、Min-Hung Chen、Yen-Yu Lin 和 Yu-Lun Liu。《Longsplat：针对日常长视频的鲁棒无姿态三维高斯泼溅》。载于 IEEE/CVF国际计算机视觉会议论文集，第 27412–27422 页，2025。2[53]Youtian Lin、Zuozhuo Dai、Siyu Zhu 和 Yao Yao。《Gaussian-flow：利用动态三维高斯粒子进行四维重建》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 21136–21145 页，2024。2[54] Rong Liu、Dylan Sun、Meida Chen、Yue Wang 和 Andrew Feng。《可变形贝塔泼溅》。载于计算机图形学与交互技术特别兴趣小组会议论文集，第 1–11 页，2025。2 – 。[55] Songrun Liu、Alec Jacobson 和 Yotam Gingold。《蒙皮立方贝塞尔样条与 Catmull-Clark细分曲面》。ACM图形学汇刊（TOG），33(6):1–9，2014。2 Tianqi Liu、Guangcong Wang、Shoukang Hu、Liao Shen、Xinyi Ye、Yuhang Zang、Zhiguo Cao、Wei Li 和 Ziwei Liu。《MVSgaussian：基于多视图立体的快速通用高斯泼溅重建》。载于欧洲计算机视觉会议，第 37–53 页。施普林格，2024。2[56] Yu-Lun Liu、Chen Gao、Andreas Meuleman、Hung-Yu Tseng、Ayush Saraf、Changil Kim、Yung-Yu Chuang、Johannes Kopf 和 Jia-Bin Huang。《鲁棒动态辐射场》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 13–23 页，2023。2, 6[57] Zhicheng Lu、Xiang Guo、Le Hui、Tianrui Chen、Min Yang、Xiao Tang、Feng Zhu 和 Yuchao Dai。《用于动态视图合成的三维几何感知可变形高斯泼溅》。载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 8900–8910 页，2024。2[58] Jonathon Luiten、Georgios Kopanas、Bastian Leibe 和 Deva Ramanan。《Dynamic 3D Gaussians：通过持久动态视图合成进行跟踪》。载于 2024 年国际三维视觉会议（3DV），第 800–809 页。IEEE，2024。2[59] Wan-Duo Kurt Ma、John P Lewis 和 W Bastiaan Kleijn。《Trailblazer：用于基于扩散的视频生成的轨迹控制》。载于 SIGGRAPH Asia 2024 会议论文集，第 1–11 页，2024。2[60] Marko Mihajlovic 、 Sergey Prokudin、Siyu Tang、Robert Maier、Federica Bogo、Tony Tung 和 Edmond Boyer。《Splat-场：用于稀疏三维和四维重建的神经高斯泼溅。载于欧洲计算机视觉会议，页码 313–332。施普林格，2024。2[62] Hao Ouyang、Qiuyu Wang、Yuxi Xiao、Qingyan Bai、Jun-tao Zhang、Kecheng Zheng、Xiaowei Zhou、Qifeng Chen 和 Yujun Shen。CoDeF：用于时间一致的视频处理的内容变形场。载于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 8089–8099，2024。1、2、6、7[63] Jongmin Park、Minh-Quan Viet Bui、Juan Luis Gonzalez Bello、Jaeho Moon、Jihyong Oh 和 Munchurl Kim。SplineGS：一种用于从单目视频实时生成动态三维高斯的鲁棒运动自适应样条。载于计算机视觉与模式识别会议论文集，页码 26866–26875，2025。2[64] Luigi Piccinelli、Yung-Hsu Yang、Christos Sakaridis、Mattia Segu、Siyuan Li、Luc Van Gool 和 Fisher Yu。Unidepth：通用单目度量深度估计。载于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 10106–10116，2024。3[65] Jordi Pont-Tuset、Federico Perazzi、Sergi Caelles、Pablo Arbeláez、Alex Sorkine-Hornung 和 Luc Van Gool。2017年DAVIS挑战赛：视频目标分割。arXiv预印本 arXiv:1704.00675，2017。1、2、6[66] LIU Qingming、Yuan Liu、Jiepeng Wang、Xianqiang Lyu、Peng Wang、Wenping Wang 和 Junhui Hou。Modgs：利用深度先验从随手拍摄的单目视频中进行动态高斯泼溅。载于第十三届国际学习表征会议，2025。1、2[67] Zefan Qu、Ke Xu、Gerhard Petrus Hancke 和 Rynson WH Lau。Lush-nerf：为低光照场景提升神经辐射场的亮度与清晰度。arXiv预印本 arXiv:2411.06757，2024。2[68] Ren´e Ranftl、Alexey Bochkovskiy 和 Vladlen Koltun。用于密集预测的视觉变换器。载于 IEEE/CVF国际计算机视觉会议论文集，页码 12179–12188，2021。3、5[69] Fitsum Reda、Janne Kontkanen、Eric Tabellion、Deqing Sun、Caroline Pantofaru 和 Brian Curless。Film：用于大幅运动的帧插值。载于欧洲计算机视觉会议，页码 250–266。施普林格，2022。2[70] Jens Eirik Saethre、Roberto Azevedo 和 Christopher Schroers。结合帧和GOP嵌入的神经视频表示。载于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 9253–9263，2024。2[71] Mu-Yi Shen、Chia-Chi Hsu、Hao-Yu Hou、Yu-Chen Huang、Wei-Fang Sun、Chia-Che Chang、Yu-Lun Liu 和 Chun-Yi Lee。Driveenv-nerf：探索基于神经辐射场的自动驾驶环境，以验证真实世界性能。arXiv 预印本 arXiv:2403.15791，2024。2[72] Meng-Li Shih、Ying-Huan Chen、Yu-Lun Liu 和 Brian Curless。先验增强高斯泼溅：从随手拍摄的视频中重建动态场景。载于 SIGGRAPH Asia 2025 会议论文集，页码 –。2025。13、1、2[73] Seungjun Shin、Suji Kim 和 Dokwan Oh。通过时间一致调制实现高效神经视频表示。

载于欧洲计算机视觉会议，页码 179–195。施普林格，2024。2[74] Gaurav Shrivastava、Ser-Nam Lim 和 Abhinav Shrivastava。视频分解先验：一种将视频分解为多个图层的方法。arXiv预印本 arXiv:2412.04930，2024。2[75] Colton Stearns、Adam Harley、Mikaela Uy、Florian Dubost、Federico Tombari、Gordon Wetzstein 和 Leonidas Guibas。用于随手拍摄的单目视频新视角合成的动态高斯弹珠。载于 SIGGRAPH Asia 2024 会议论文集，页码 1–11，2024。1、2[76] Yang-Tian Sun、Yihua Huang、Lin Ma、Xiaoyang Lyu、Yan-Pei Cao 和 Xiaojuan Qi。Splatter A Video：面向多样化处理的视频高斯表示。神经信息处理系统进展，37：50401–50425，2024。1、2、3、6、7[77] Allen Tu、Haiyang Ying、Alex Hanson、Yonghan Lee、Tom Goldstein 和 Matthias Zwicker。快速可变形三维高斯泼溅：动态场景的快速渲染与压缩。arXiv预印本 arXiv:2506.07917，2025。2[78] Liao Wang、Kaixin Yao、Chengcheng Guo、Zhirui Zhang、Qiang Hu、Jingyi Yu、Lan Xu 和 Minye Wu。VideoRF：将动态辐射场渲染为二维特征视频流。载于 IEEE/CVF计算机视觉与模式识别会议论文集，页码 470–481，2024。2[79] Ning-Hsu Albert Wang 和 Yu-Lun Liu。Depth Anywhere：通过透视蒸馏和无标注数据增强提升360度单目深度估计。神经信息处理系统进展，37：127764，2024。127739：–。3[80] Qianqian Wang、Yen-Yu Chang、Ruojin Cai、Zhengqi Li、Bharath Hariharan、Aleksander Holynski 和 Noah Snavely。随时随地追踪一切。载于国际计算机视觉会议，页码 19795–19806，2023。2、3、6[81] Yihan Wang、Lahav Lipson 和 Jia Deng。SEA-RAFT：用于光流的简单、高效、准确的RAFT。载于欧洲计算机视觉会议，页码 36–54。施普林格，2024。3[82] Yikai Wang、Xinzhou Wang、Zilong Chen、Zhengyi Wang、Fuchun Sun 和 Jun Zhu。VIDU4D：从单个生成视频实现高保真四维重建与动态高斯表面元。神经信息处理系统进展，37：131316–131343，2024。2[83] Haato Watanabe、Kenji Tojo 和 Nobuyuki Umetani。三维Gabor泼溅：利用Gabor噪声重建高频表面纹理。arXiv预印本 arXiv:2504 11003，2025。2。[84] Chung-Ho Wu、Yang-Jung Chen、Ying-Huan Chen、Jie-Ying Lee、Bo-Hsu Ke、Chun-Wei Tuan Mu、Yi-Chuan Huang、Chin-Yang Lin、Min-Hung Chen、Yen-Yu Lin 等。Aurafusion360：面向基于参考的360度无界场景修复的增强未见区域对齐。载于计算机视觉与模式识别会议论文集，页码 16366–16376，2025。2[85] Guanjun Wu、Taoran Yi、Jiemin Fang、Lingxi Xie、Xiaopeng Zhang、Wei Wei、Wenyu Liu、Qi Tian 和 Xinggang Wang。用于实时动态场景渲染的4D高斯泼溅。

发表于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20310–20320页，2024年。 2[86] Renlong Wu、Zhilu Zhang、Mingyang Chen、Zifei Yan和Wangmeng Zuo。Deblur4dgs：基于模糊单目视频的4D 高斯泼溅。arXiv预印本 arXiv:2412 06424，2024年 2，6…… [87] Skylar Wurster、Ran Zhang和Changxi Zheng。用于高质量十亿像素图像表示的 Gabor 泼溅。载于《ACM SIGGRAPH 2024海报集》，第1–2页。2024年。1，2[88] Shuxiang Xie、Shuyi Zhou、Ken Sakurada、Ryoichi Ishikawa、Masaki Onishi和Takeshi Oishi。G 2 f r：基于网格的特征编码神经辐射场中的频率正则化。载于《欧洲计算机视觉会议》，第186–203页。施普林格，2024年。2[89] Jiawei Xu、Zexin Fan、Jian Yang和Jin Xie。Grid4d：用于高保真动态高斯泼溅的4D 分解哈希编码。神经信息处理系统进展，37：123787–123811，2024年。2[90] Hao Yan、Zhihui Ke、Xiaobo Zhou、Tie Qiu、Xidong Shi和 Dadong Jiang。Ds-nerv：采用分解静态与动态编码的隐式神经视频表示。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第23019–23029页，2024年。2[91] Yunzhi Yan、Haotong Lin、Chenxu Zhou、Weijie Wang、Haiyang Sun、Kun Zhan、Xianpeng Lang、Xiaowei Zhou和Sida Peng。街景高斯：利用高斯泼溅对动态城市场景建模。载于《欧洲计算机视觉会议》，第156–173页。施普林格，2024年。2[92] Zhiwen Yan、Weng Fei Low、Yu Chen和Gim Hee Lee。用于抗锯齿渲染的多尺度 3D 高斯泼溅。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20923–20931页，2024年。2[93] Jiawei Yang、Marco Pavone和Yue Wang。Freenerf：通过自由频率正则化改进小样本神经渲染。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第8254 8263页，2023年 2 – 。[94] Lihe Yang、Bingyi Kang、Zilong Huang、Xiaogang Xu、Jiashi Feng和Hengshuang Zhao。Depth Anything：释放大规模无标签数据的潜力。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第10371–10381页，2024年。3[95]Zeyu Yang、Hongye Yang、Zijie Pan和Li Zhang。利用4D 高斯泼溅实现实时逼真的动态场景表示与渲染。arXiv预印本 arXiv:2310.10642，2023年。2[96] Ziyi Yang、Xinyu Gao、Wen Zhou、Shaohui Jiao、Yuqing Zhang和 Xiaogang Jin。用于高保真单目动态场景重建的可变形 3 D 高斯。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第20331–20341页，2024年。2[97] Vickie Ye、Zhengqi Li、Richard Tucker、Angjoo Kanazawa和 Noah Snavely。用于无监督视频分解的可变形精灵。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第2657–2666页，2022年。6

[98] Zehao Yu、Anpei Chen、Binbin Huang、Torsten Sattler和Andreas Geiger。Mip-splatting：无混叠 3D 高斯泼溅。载于《 IEEE/CVF 计算机视觉与模式识别会议论文集》，第19447 19456页，2024年 2 – 。[99] Zehao Yu、Torsten Sattler和Andreas Geiger。高斯不透明度场：无界场景中的高效自适应表面重建。《ACM图形学汇刊》（ToG），43(6)：1–13，2024年。2[100] Yifan Zhan、Zhuoxiao Li、Muyao Niu、Zhihang Zhong、Shohei Nobuhara、Ko Nishino和Yinqiang Zheng。Kfd-nerf：使用卡尔曼滤波器重新思考动态神经辐射场。载于《欧洲计算机视觉会议》，第1–18页。施普林格，2024年。2[101] Yu-Ting Zhan、Cheng-Yuan Ho、Hebi Yang、Yi-Hsin Chen、Jui Chiu Chiang、Yu-Lun Liu和Wen-Hsiao Peng。Cat-3dgs：一种面向率失真优化 3DGS 压缩的上下文自适应三平面方法。arXiv预印本 arXiv:2503.00357，2025年。2[102] Tingyang Zhang、Qingzhe Gao、Weiyu Li、Libin Liu和Baoquan Chen。Bags：利用扩散先验从单目视频构建可动画高斯泼溅。arXiv预印本 arXiv:2403.11427，2024年。1，2[103] Shangchen Zhou、Peiqing Yang、Jianyi Wang、Yihang Luo和Chen Change Loy。Upscale-A-Video：面向真实世界视频超分辨率的时间一致扩散模型。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第2535–2545页，2024年。2[104] Xiaoyu Zhou、Zhiwei Lin、Xiaojun Shan、Yongtao Wang、Deqing Sun和Ming-Hsuan Yang。DrivingGaussian：面向周围动态自动驾驶场景的复合高斯泼溅。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第21634–21643页，2024年。2[105] Ruijie Zhu、Yanzhe Liang、Hanzhi Chang、Jiacheng Deng、Jiahao Lu、Wenfei Yang、Tianzhu Zhang和Yongdong Zhang。MotionGS：探索面向可变形 3D 高斯泼溅的显式运动引导。神经信息处理系统进展，37：101790–101817，2024年。2

目录

1. 引言 1

2. 相关工作 2

3. 预备知识：3D 高斯泼溅 3

4. 方法 3
4.1. 概述 .4.2. 自适应 Gabor 视频表示 3
4.3. 时序动态自适应 Gabor . . . .4.4. 优化 . 3
4.5. 自适应初始化 . 4
5
6

5. 实验 6
5.1. 评估 .5.2. 应用 .5.3. 消融研究 . 6
6
6
6

6. 结论 8

A. Gabor 系数的激活函数 1A.1. 频率权重的直通硬 Sigmoid .1 4
4

B. 向高斯的自适应退化证明 14
B.1. 数学表述 . . . . . .B.2. 退化为高斯分布 .B.3. 对不透明度的影响 .B.4. 结论 . 14
14
15
15

C. 更多视觉对比与结果 15

A. Gabor系数激活

A.1. 频率权重的直通硬 Sigmoid

在 Gabor 基元中，频率系数 ω_i 必须满足两个要求：(1) 将数值限制在可学习的范围内；(2) 梯度必须能够通过激活函数反向传播，以实现端到端优化。

为此，我们采用直通估计器（STE）[4]和硬Sigmoid激活。在前向传播过程中，我们应用硬 Sigmoid 将 ω_i 裁剪到范围 $[0,1]$:

$$\hat{\omega} = \text{clip}\left(\frac{\omega+1}{2}, 0, 1\right). \tag{20}$$

这确保 Gabor核的频率调制保持有界，避免无限增长导致能量平衡不稳定。

然而，由于硬裁剪操作不可微，我们无法直接通过它进行反向传播。

因此，在反向传播过程中，我们使用 Sigmoid函数的梯度作为替代:

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial \hat{\omega}} \cdot \sigma(\omega)(1-\sigma(\omega)), \tag{21}$$

其中 $\sigma(\cdot)$ 是标准的 Sigmoid函数。这提供了平滑且有界的梯度信号。

有界的前向传播与平滑的反向传播相结合，实现了稳定训练：前向传播通过约束频率权重来防止伪影，而反向传播则支持有效的基于梯度的优化。这种方法避免了无界激活可能导致的梯度爆炸。

B. 向高斯的自适应退化的证明

我们证明，当所有频率权重都消失时，我们的自适应 Gabor 表示会自然退化为传统高斯，从而展现其在高斯模式和Gabor模式之间自适应的能力。

B.1. 数学表述

回顾正文中的公式 (4) 和公式 (5)，自适应调制函数定义如下:

$$S_{\text{adap}}(\mathbf{x}) = b + \frac{1}{N}\sum_{i=1}^{N} \omega_i \cos(f_i \langle \mathbf{d}_i, \mathbf{x} \rangle), \tag{22}$$

其中补偿项 b 为:

$$b = \gamma + (1-\gamma)\left(1 - \frac{1}{N}\sum_{i=1}^{N}\omega_i\right), \tag{23}$$

其中，$\gamma \in [0,1]$ 是控制退化平滑度的固定超参数，$1/N$ 用于对多个波的加权平均进行归一化。

B.2. 退化为高斯分布

考虑所有频率权重趋近于零的极限情况：对于所有 $i \in \{1,\ldots,N\}$，$\omega_i \to 0$。
在这种情况下:

$$\sum_{i=1}^{N}\omega_i \to 0. \tag{24}$$

代入补偿项:

$$b \to \gamma + (1-\gamma)\left(1 - \frac{1}{N}\cdot 0\right) = \gamma + (1-\gamma)\cdot 1 = 1. \tag{25}$$

因此，调制项变为:

$$\frac{1}{N}\sum_{i=1}^{N}\omega_i \cos(f_i\langle \mathbf{d}_i, \mathbf{x}\rangle) \to 0. \tag{26}$$

因此:

$$S_{\text{adap}}(\mathbf{x}) \to 1 + 0 = 1. \tag{27}$$

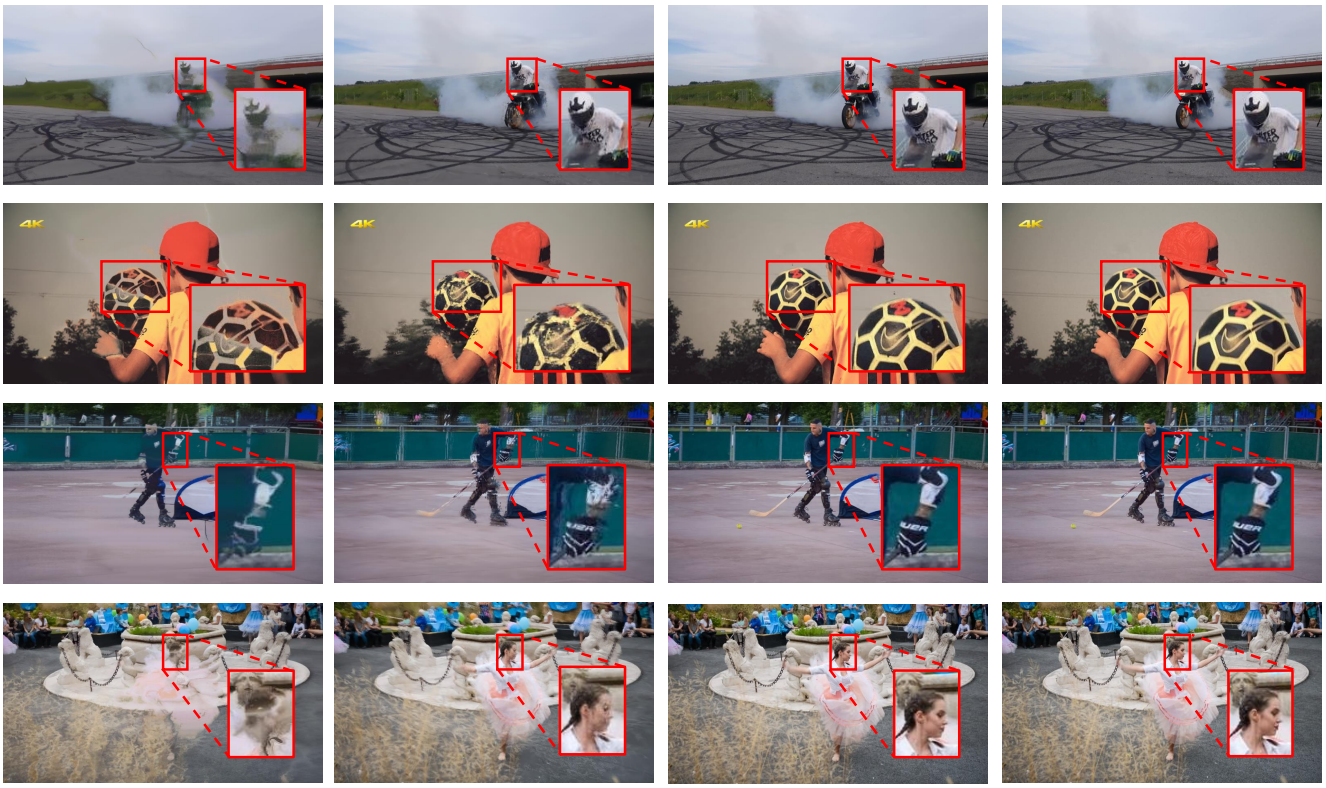


图 12。**在 DAVIS 数据集上的视觉对比。**

B.3. 对不透明度的启示

由于 Gabor 调制不透明度定义为：

$$\alpha_{\text{Gabor}}(\mathbf{x}) = \mathcal{G}(\mathbf{x}) \cdot S_{\text{adap}}(\mathbf{x}), \tag{28}$$

当 $S_{\text{adap}}(\mathbf{x}) = 1$时，我们得到：

$$\alpha_{\text{Gabor}}(\mathbf{x}) = \mathcal{G}(\mathbf{x}) \cdot 1 = \mathcal{G}(\mathbf{x}), \tag{29}$$

这正是没有频率调制的传统高斯基元。

B.4. 结论

该证明表明，当不需要频率内容时，我们的自适应 Gabor 表示会平滑退化为标准高斯函数（$\omega_i \to 0$）；而当需要高频细节时，则会平滑过渡到增强频率的 Gabor 模式（$\omega_i > 0$）。这种自适应行为对于在频率特征各异的不同场景区域中维持能量稳定性至关重要。

C. 更多视觉对比与结果

如需全面比较不同动态场景下我们的方法与基线方法的视觉效果，请参阅图 12 至 14。这些图展示了我们的方法在具有挑战性的场景中表现出色，能够保留高频纹理细节并维持时间一致性，这些场景包括快速运动、遮挡和复杂形变。

如需交互式查看下游应用结果，包括帧插值、视频编辑和立体视图合成，请参阅补充 HTML 页面（index.html）。交互式查看器支持逐帧检查和视频播放，便于更好地欣赏我们方法的时间连贯性和视觉质量。

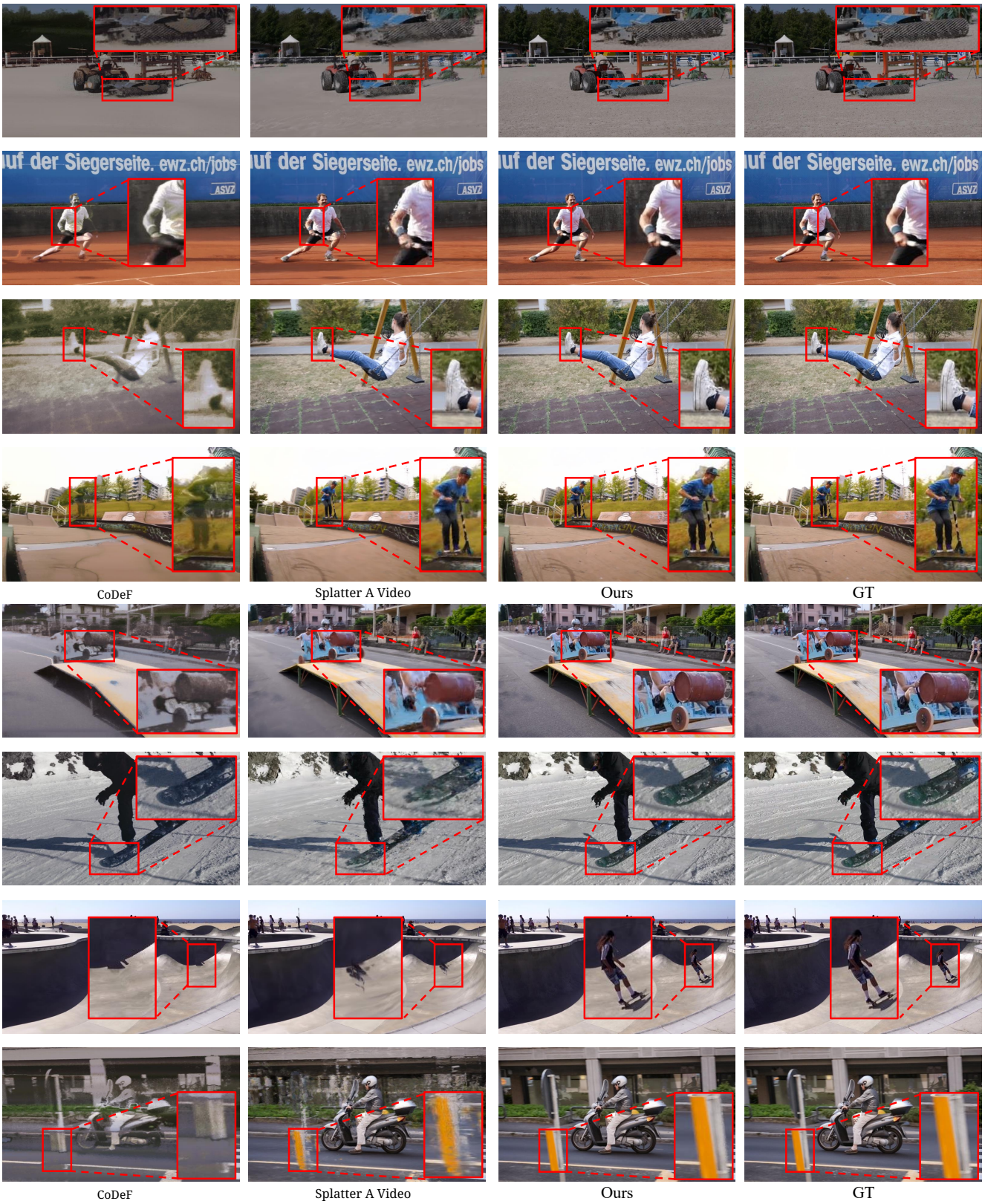


图13。**DAVIS 数据集上的视觉比较。**

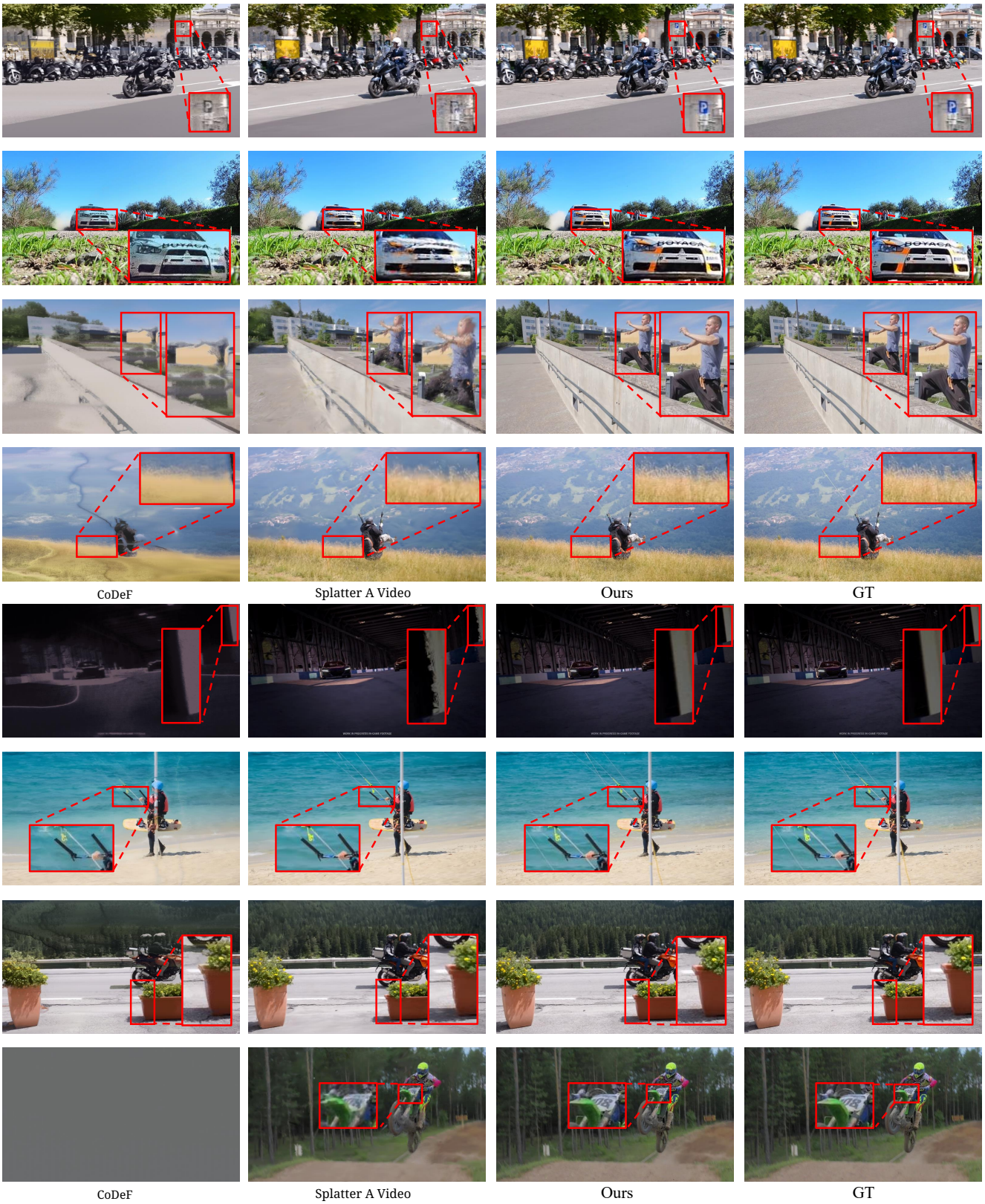


图14。**DAVIS 数据集上的视觉比较。**