方法介绍：视频高斯表示（VGR）

1. 方法介绍

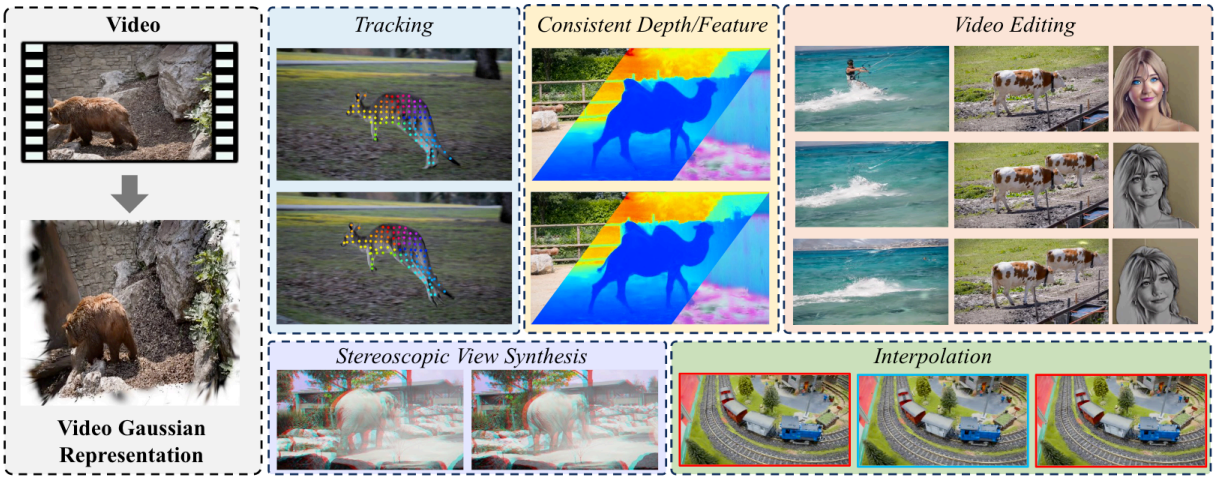

Figure 1: We propose an approach to convert a video into a Video Gaussian Representation (VGR), which can be used for versatile video processing tasks conveniently.

Drawing inspiration from the fact that a video is essentially a projection of the dynamic 3D world onto the 2D image plane at different moments, we pose the question: *is it possible to represent a video in its intrinsic 3D form*? By doing so, we could potentially bypass the limitations of 2D representations, such as occlusions, reduce the complexity of motion modeling, and support processing tasks that require 3D information. Recent work [45] has explored 3D representations, which employ an implicit radiance field to model a canonical 3D space and leverage a bi-directional mapping network for associating 2D pixels with 3D representations. While this approach demonstrates promising performance in dense tracking, it falls short in faithfully representing video appearance, making it incapable of performing video processing tasks that require generating new videos, such as video editing. Moreover, its implicit nature limits its applicability to a variety of video processing tasks that require explicit content or motion manipulations, such as the removal or addition of objects and adjustments to the motion patterns of objects.

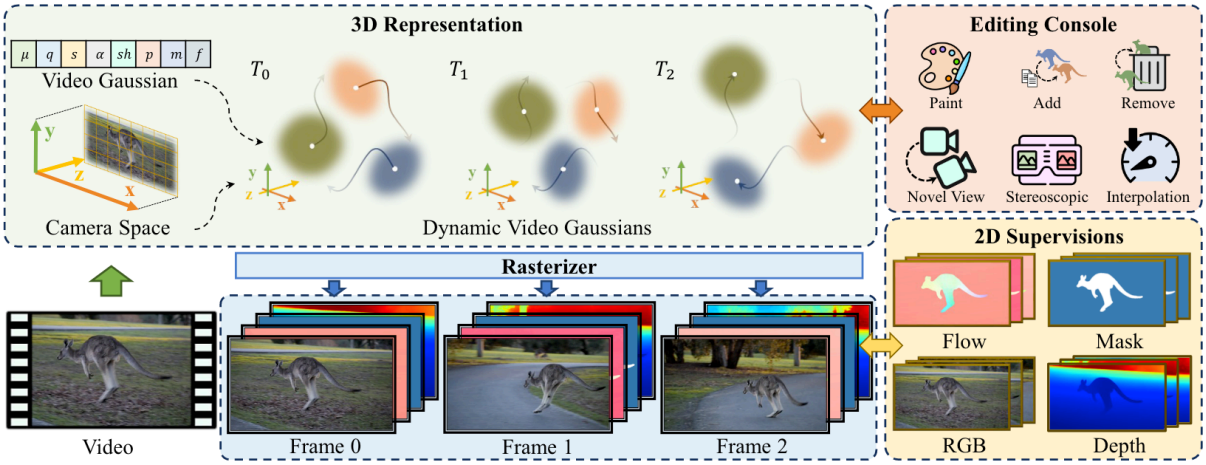

Figure 2: **Pipeline of our approach.** Given a video, we represent its intricate 3D content using video Gaussians in the camera coordinate space. By associating them with motion parameters, we enable video Gaussians to capture the video dynamics. These video Gaussians are supervised by RGB image frames and 2D priors such as optical flow, depth, and label masks. This representation makes it convenient for users to perform various editing tasks on the video.

1.1 背景与动机

- **现有方法不足**：
  1. **2D/2.5D 方法**：如光流或分层图像表示（Layered Atlas），在处理遮挡和复杂运动时表现不佳。主要依赖逐帧像素传播，难以建模三维信息。
  2. **隐式3D方法**：如基于辐射场的隐式表示，难以显式编辑视频内容，不能灵活操控运动模式。
- **本方法创新**：
  - 引入显式3D高斯表示，每帧视频以动态3D高斯的集合形式表示，能够同时建模外观与运动。

1.2 高斯表示的核心公式

1.2.1 3D 高斯定义

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{1}$$

- **参数含义**：
  - x：3D 空间中的任意点。

- μ：高斯的中心（3D 位置）。
- Σ：协方差矩阵，表示高斯分布的形状，分解为：

$$\Sigma = RSS^T R^T \tag{2}$$

其中：

- R：旋转矩阵（用四元数 q 表示）。
- S：缩放矩阵（由向量 s 表示）。

1.2.2 动态高斯的运动建模

$$\mu(t) = \mu_0 + \sum_{n=0}^{N} p_{np} t^n + \sum_{l=0}^{L} \left(p_{l\sin} \cos(lt) + p_{l\cos} \sin(lt)\right) \tag{3}$$

- **参数含义**：
  - $\mu(t)$：时间 t 时高斯的位置。
  - μ_0：高斯的初始位置。
  - p_{np}：多项式基的系数，捕捉非周期性运动趋势。
  - $p_{l\sin}, p_{l\cos}$：傅里叶基的系数，建模周期性运动。

---

2. 损失函数与优化

2.1 损失函数的公式与意义

2.1.1 光流蒸馏损失

$$L_{\text{flow}} = E_{t_1,t_2} \| \pi(\mu(t_2)) - \pi(\mu(t_1)) - \text{flow}_{t_1 \to t_2} \|_1 \tag{4}$$

- **公式解读**：
  - $\pi(\cdot)$：3D 点到 2D 图像平面的投影函数。
  - $\mu(t_1), \mu(t_2)$：高斯在时间 t_1, t_2 时的3D位置。
  - $\text{flow}_{t_1 \to t_2}$：由光流网络 RAFT 提供的 2D 光流估计。

- **作用**：约束高斯的运动与估计光流一致，确保2D-3D对齐。

2.1.2 深度蒸馏损失

$$L_{\text{depth}} = E_t \left\| \tau(D_t) - \tau(\hat{D}_t) \right\|_2 \tag{5}$$

- **公式解读**：
  - D_t：由高斯渲染的深度图。
  - $\hat{D}_t$：由单目深度估计网络（Marigold）生成的深度图。
  - $\tau(D)$：深度归一化函数，用于消除尺度差异：

$$\tau(D) = \frac{D - \text{median}(D)}{\text{mean}(|D - \text{median}(D)|)} \tag{6}$$

- **作用**：确保3D结构的合理性，提高深度估计的跨帧一致性。

2.1.3 局部刚性约束损失

$$L_{\text{arap}} = E(i, t_1, t_2) \sum_{k \in N_i} \left\| (\mu_i(t_1) - \mu_k(t_1)) - \hat{R}_i(\mu_i(t_2) - \mu_k(t_2)) \right\|_2 \quad (7$$

- **公式解读**：
  - μ_i, μ_k：高斯 i 和其邻居 k 的位置。
  - $\hat{R}_i$：最优刚性旋转矩阵，确保局部运动尽可能刚性。
- **作用**：防止高斯过拟合目标，生成非物理性运动。

2.1.4 颜色渲染损失

$$L_{\text{render}} = E_t \left\| R(\mu, q, s, \alpha, SH(sh, v)) - I_t \right\| \tag{8}$$

- **公式解读**：
  - $R(\cdot)$：高斯渲染函数。
  - $SH(sh, v)$：基于球谐系数 sh 和视角方向 v 的颜色表示。
  - I_t：时间 t 的目标视频帧。

2.1.5 总损失

$$L = \lambda_{\text{render}} L_{\text{render}} + \lambda_{\text{depth}} L_{\text{depth}} + \lambda_{\text{flow}} L_{\text{flow}} + \lambda_{\text{arap}} L_{\text{arap}} + \lambda_{\text{label}} L_{\text{label}}$$

---

3. 方法对比

方法	优点	不足
现有方法：2D/2.5D	简单，适用于层级编辑	难以建模复杂运动，受遮挡影响严重
现有方法：隐式3D（如NeRF）	能处理复杂3D结构	隐式表示难以编辑，无法显式操控内容
本文方法：VGR	显式3D建模，支持灵活编辑	对于快速非刚性运动仍有挑战

---

4. 实验分析

4.1 对比实验

- DAVIS 数据集上与 Omnimotion、CoDeF 等方法对比，在视频重建质量（PSNR）和一致性上取得最佳表现。

4.2 消融实验

- **去掉深度蒸馏**：高斯分布塌缩为2D层，失去3D效果。
- **去掉刚性约束**：高斯运动出现浮动效应，PSNR 降低 1.51 dB。