arXiv:2611.31140v1 [cs.CV] 23 Nov 2026

HoloSplat：用于前馈式人体重建的实时令牌流

李超凡* 丁启涵* ShihanWang ZhehaoShen BoyuShen 陈全宇
HaoyuWu Lan Xu† 娄欣†
上海科技大学
{lichf2025, dingqh2025, wangshh2023, shenzhh2025, shenby2024,
chenqy2026, wuhy2025, xulan1, louxin}@shanghaitech.edu.cn

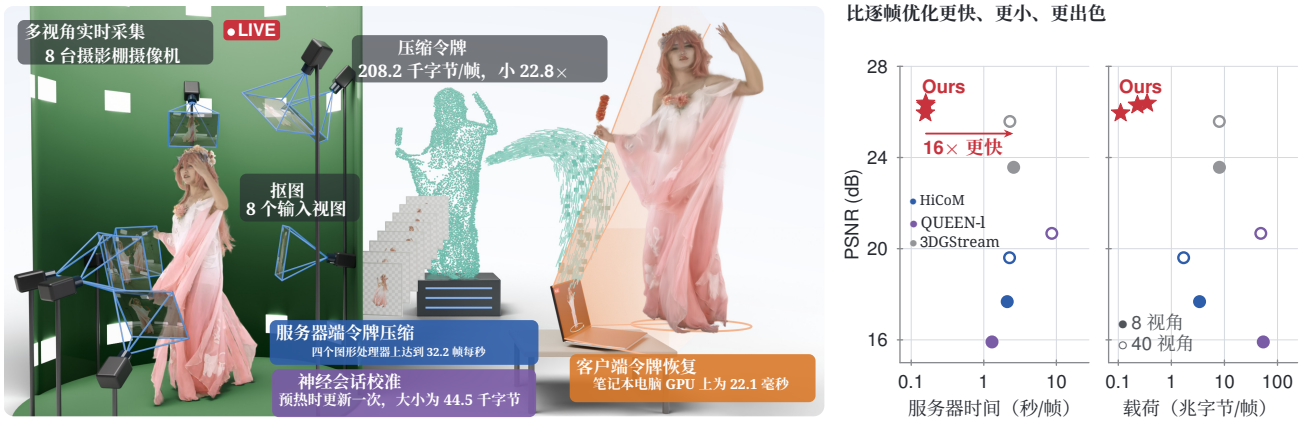

图 1. **HoloSplat 通过紧凑令牌而非显式三维高斯，实时传输以人为中心的体积视频。** 在高斯头之前拆分前馈式重建器后，服务器端令牌压缩发送22。8× 比 float32 高斯更少的数据；客户端令牌恢复通过与高斯头融合的特征重建超网络对其进行解码，并在笔记本电脑 GPU 上运行；神经会话校准则在预热时对该头部进行一次微调（左侧，DNA-Rendering 0678 02）。右侧（表 3）显示，所有层级在速度（16×）、比特率和平均峰值信噪比方面均优于逐帧优化。

摘要

实时传输人体自由视点视频，可实现沉浸式远程临场和交互式扩展现实应用。然而，这需要实时重建每一帧、在严格的带宽预算内传输数据，并在轻量级客户端图形处理器上进行解码。现有方法无论依赖逐帧优化还是前馈推理，都会生成海量三维高斯载荷，超出实际流传输限制；与此同时，激进压缩会严重降低渲染质量。在本研究中，我们提出 *HoloSplat*，这是一种面向以人为中心的体积视频的实时流媒体框架。*HoloSplat* 在前馈重建器的高斯头之前将其拆分，并引入一种由服务器端令牌压缩和客户端令牌恢复驱动的多速率流媒体编解码器。在服务器端，我们采用不透明度引导的令牌剪枝、基于主成分分析的降维，以及无损几何编码，以

令牌锚点。在客户端，将特征重建超网络融合到高斯头中，以恢复压缩造成的质量损失。此外，我们提出神经会话校准（*NSC*），在预热阶段利用输入相机视图在线快速微调高斯头。此方法无需增加每帧开销，即可解决新相机参数和不同采集条件导致的质量下降。在评估中，*HoloSplat* 在四块 *RTX 5090 GPU* 上以 *50* 兆比特/秒的速率平均达到 *32.2* 帧每秒。它在 *RTX 3050 Ti* 笔记本 *GPU* 上以 *22.1* 毫秒将每帧解码为三维高斯，传输的数据量比未压缩 *float32* 高斯少 *22.8*×，而经过校准的解码可将峰值信噪比提升 *1.050 dB*。

*共同贡献。†通讯作者。



图2. **现有方案均无法满足实时自由视点视频的三项逐帧预算；HoloSplat 则全部达标。** 红色标出未达标的阶段；缩略图展示了各方案发送的内容。码率为 30 帧每秒下 DNA-Rendering 的有效载荷，HoloSplat 使用 Pro（表 2 和 3）；除虚拟形象训练（表 S30）和运行在 RTX 3050 Ti 笔记本 GPU 上的客户端外，计时均在 RTX 5090 GPU 上进行。摄像机视频为八路 1024×1024 HEVC [64] 码流，并配有无损遮罩。

1. 引言

人类实时自由视点视频流传输可实现沉浸式远程临场和交互式扩展现实应用，让观看者能够实时渲染任意视点。在 30 帧每秒下，串联的三个阶段必须满足严格的每帧预算（图 2）。具体而言，服务器必须重建每一帧（在每 33 毫秒内完成，并将延迟控制在界限内），网络必须传输每一帧，并满足实际带宽预算，例如 100 兆比特/秒的 5G 下行链路 [25],，而轻量级客户端必须将每一帧解码为可渲染的三维高斯 [33]，并在 33 毫秒内完成。

然而，无论是通过逐帧优化还是前馈推理进行重建，每一帧都会生成大量三维高斯载荷，超出实际可用的流式传输带宽，而激进压缩又会降低渲染质量（图 2）。逐帧优化方法 [16, 17, 34,65, 83]可在客户端实时渲染 [17, 31],，但逐帧拟合 t 会使 $t-1$ 服务器端重建变成顺序优化过程，每帧耗时 1.2–23.4 秒，无法在多个图形处理器之间并行执行，且需要 ≥154兆比特/秒 [16, 17, 65]。前馈推理 [21, 70, 93]只需一次处理即可重建每一帧，但仍会输出稠密的三维高斯（≥173使用 3D高斯泼溅（3DGS）压缩方法时为 兆比特/秒 [7, 51]），或者在激进的视频编码下损失质量 [70]。在其他范式中，瓶颈只是转移了位置（第 2 节）。静态场景特征编解码器 [87] 会导致客户端过载（199 毫秒）；多视角视频流会在每个客户端重复执行重建（85–126 毫秒）；服务器渲染的视频则使视点变化受制于一次网络往返；而特定人物化身 [40, 44] 需要训练 2 天。

我们提出 HoloSplat，这是一种用于以人为中心的体积视频的实时流媒体框架。该框架在高斯头之前拆分逐帧运行的前馈式重建器：服务器流式传输压缩令牌，客户端则恢复令牌并运行头部，以生成三维高斯。现有 3DGS 流媒体方法会在头部之后压缩显式高斯；此时，非线性激活——尤其是指数缩放——会将紧凑且相关的令牌映射到扩展后的属性空间中，使微小的量化误差导致渲染失真放大四个数量级（第 4.3节）。与压缩头部后的属性相比，压缩头部前的令牌可实现更高的压缩比，同时将质量损失降至最低；尽管 PointSplat 中每个 1024 维令牌的维数高于其 16 个高斯的 224 个属性，但使用相同编解码器的效果仍优于头部后属性压缩 2.8–10.3 dB。

基于这一划分，我们设计了一种多速率流式编解码器，由服务器端令牌压缩和客户端令牌恢复组成。在服务器端，我们进行不透明度引导的令牌剪枝、基于主成分分析的令牌降维，以及令牌锚点坐标的无损几何编码。在客户端，我们设计了一个由 FiLM 调节的特征重建超网络，以恢复压缩造成的质量损失，并将逆主成分分析和超网络融合到客户端高斯头中。借助保留的 PCA 系数、局部锚点几何结构和空间邻居，该超网络及其一次性更新可在每帧码率开销为零的情况下恢复 80.4% 的压缩损失，无需显式扩展 1024 维令牌。

此外，我们引入神经会话校准（NSC），在预热阶段利用输入相机视图对高斯头进行在线快速微调，从而在每帧开销为零的情况下，解决新相机参数和变化的采集条件所导致的高斯头质量损失。在 HiFi4G [27]和 ActorsHQ [24],上，在校准前移除令牌压缩后，PSNR 的变化最多为 0.13 dB （图 4 和表 S37），这证实即使没有压缩也会出现这种质量损失，而非源于编解码器失真；NSC 使用八个输入视角的首秒数据拟合出一个紧凑的 44.5 千字节低秩更新，并在流式传输前将其融合到客户端高斯头中。

HoloSplat 满足全部三项预算要求（图 2，下）。在 Pro 级别以 50 兆比特/秒传输时，四块 RTX 5090 GPU 的平均速度为 32.2 帧每秒；在 RTX 3050 Ti 笔记本 GPU 上，每帧解码为三维高斯仅需 22.1 毫秒；五块 GPU 通过 Wi-Fi 以 30 帧每秒进行实时流式传输，端到端延迟为 204–304 毫秒。Pro 传输的数据量比 float32 高斯少 228× %，而校准解码可将峰值信噪比提高 1.050 dB，并且相比 3DGS 压缩方法节省 67–85% 的比特率

，在质量相同的情况下。在相同的八个视角下，三个层级在速度 16× 方面的每个序列中，质量和比特率均优于 HiCoM、QUEEN 和 3DGStream——Pro 级别相比离线 4C4D [97] 高出 1.62 dB，且58× 有效载荷大小更小——而 NSC 在 DNA-Rendering 上额外提升 +10.154 dB [8]，并在 +1未见过的设备上分别提升 0.419 和 +10.436 dB。

贡献。

1. **多速率令牌流式编解码器。** 我们设计了一种面向以人为中心的体积视频的多速率流式编解码器，由服务器端令牌压缩和客户端令牌恢复组成。在服务器端，我们进行不透明度引导的令牌剪枝、基于 PCA 的令牌降维，以及令牌锚点坐标的无损几何编码；与 float32 高斯相比，传输的数据量减少 22.8×，并且借助我们的客户端解码器，在质量相同的情况下，比 3DGS 压缩方法节省 67–85% 的码率。

2. **特征重建超网络与融合解码。** 在客户端，我们设计了一种特征重建超网络，利用保留的 PCA 系数、局部锚点几何结构和空间邻域，通过一次性更新恢复压缩造成的质量损失的 80.4%；并将逆主成分分析和超网络融合到高斯头中，使其能够在笔记本电脑 GPU 上以 22.1 毫秒完成解码，从而使整个系统在质量、码率和速度方面均优于逐帧优化方法。

3. **神经会话校准。** 在预热阶段，我们利用输入的摄像机视图在线快速微调高斯头（将 44.5 千字节的低秩更新融合至客户端头部，且不产生每帧开销），从而解决高斯头因新的相机参数和不断变化的采集条件而导致的质量损失。

2. 相关工作

逐帧优化。 3DGStream [65]训练逐帧变换缓存，而 HiCoM [16] 和 QUEEN [17] 通过层级运动和量化残差加速逐帧优化；ReCon-GS [15]采用运动层级结构，ComGS [3]仅更新对运动敏感的关键点，iFVC [67] 则对二进制变换缓存进行熵编码。Instant Gaussian Stream [83] 利用可泛化网络预测运动，但仍会细化关键帧；MoRGS [34]利用关键视图上的光流对逐高斯运动进行正则化；4DGC [23]针对码率与失真优化运动网格；StreamSTGS [31] 将规范属性编码为图像，并将时间特征编码为视频。V^3 [75],VideoRF [74], 和 TeTriRF [78]将属性或特征打包到二维映射中，以便使用视频编解码器；ReRF [73]则对残差辐射场进行编码。HiFi4G [27] 和 DualGS [26] 等以人为中心的系统会跟踪表演者，并压缩由此生成的四维高斯；SharpTimeGS [41]则调节高斯生命周期；NeuS2 [77] 对一个逐帧优化的哈希编码表面。由于每一帧都要以前一帧为基础依次优化，耗时从数秒到数分钟不等（表S29 和 S30），逐帧优化无法实现实时流媒体。Sparse-camera 4C4D [97] 和 SparseCam4D [55] 则利用未来帧离线优化整个序列。

前馈重建与 3DGS 压缩。 PointSplat [21] 在单次处理中，使用 Point-Image Transformer 围绕粗略点代理回归三维高斯；我们将其冻结并用作重建器。GPS-Gaussian [2, 93, 96]根据相邻视图对实时回归逐像素高斯图，HiReFF [28]则根据四个未经标定的视图回归高斯。Concurrent Tele360 [70] 在单个图形处理器上以超过 25 帧每秒的速度，从四到六个非摆拍摄像机重建高斯图，并将输出高斯作为二维映射通过 H.265 视频流传输 [75, 85], ，但画质比视频压缩低约 2 dB，且快速运动时会出现闪烁。LiveGS [6] 将实时自由视点视频带到移动设备上。SPZ [51], 等通用 3DGS 压缩方法、剪枝和熵编码流程 [4, 14, 49, 79], 2、基于视频编解码器的网格 [47, 85],，以及无需训练或采用前馈方式的压缩器 [7, 63, 91]，都在高斯头将令牌扩展为非线性属性后才进行处理。

潜在特征编码与拆分计算。拆分计算在一台设备上执行网络的前几层，并将中间激活函数传输到另一台设备 [9, 11, 30, 46]。GIF- Stream [36] 使用学习式网络压缩四维表示的时序特征流。CodecSplat [87]和 FeCoSplat [39] 为静态场景前馈模型的中间特征训练熵编解码器；并行的 MFGS [10] 则在高斯预测前编码多视图静态场景特征。GS-SCNet [84]将立体图像编解码器与双视图高斯预测器联合训练，在 A100 上达到 19.2帧每秒；Feat2GS [5] 则在读取前使用主成分分析投影基础模型特征。HoloSplat 不同于为静态场景训练大型熵编解码器，而是在高斯头之前拆分冻结的多视图重建器，将服务器端的主成分分析令牌压缩与融合客户端特征重建超网络配对，实现多图形处理器流式传输，并支持帧并行处理。FeCoSplat 会针对每个场景，将发送端的重建反馈编码为第二条比特流；而 HoloSplat 则将会话适应摊销为一次性的客户端更新，因此没有每帧开销。

实时三维远程呈现与人类虚拟化身。Holoporta- tion [54], Fusion4D [13], DoubleFusion [88], 和 Func- tion4D [89] 通过融合深度传感器实时采集表演者。FromRGB 摄像机、Live4D [99] 传输由服务器融合的网格，Tele-Aloha [69] 利用四台摄像机渲染远程呈现视图，LiveStre4m [59] 渲染新视角视频

通过帧插值处理两路非摆拍视频流，而 StreamSplat [62] 则以因果方式更新静态前馈重建结果。特定人物化身每帧都会传输紧凑的驱动码，但需要针对每个主体单独采集，并训练数天 [40, 44, 61]。相反，HoloSplat 会传输任意主体的头部前标记，使客户端能够解码三维高斯，并在本地渲染自由视点视频。

测试时与会话适应。 实例自适应神经压缩会针对每个序列或静态场景微调编解码器，并传输权重更新 [22, 29, 53, 71], ; 测试时训练则通过调整重建骨干网络来适应分布偏移 [45, 66, 72, 76, 90, 98]。神经会话校准则保持重建器和令牌流不变，在预热阶段利用输入视图快速在线微调高斯头，并融合更新，且不增加每帧开销。

3. 方法

图 3 展示了系统。服务器每帧接收八个同步且经过标定的 1024×1024 RGB视图，并对其进行区域平均，得到 512×512，使用在带掩码标注的 DNA 渲染数据上微调的 BiRefNet Lite 预测软阿尔法 [92] ，再将 RGB 和阿尔法传递给冻结的 PointSplat [21], ; PointSplat 联合融合全部八个标定视图，而非针对每个目标视图使用一个立体对 [93]。

我们在高斯头部之前立即将网络拆分。每个令牌 $f_i \in \mathbb{R}^{1024}$ 围绕锚点生成一组包含 16 个高斯槽位的集合 $X_i \in \mathbb{R}^{16\times 3}$ ，头部根据这些槽位预测有界偏移量。因此，一帧由令牌组成$F = （f_1, ..., f_N）$以及锚点 $X = （X_1, ..., X_N）$。服务器（第3.1节）执行不透明度引导的令牌剪枝、基于 PCA 的令牌降维和无损锚点编码；客户端（第3.2节）将逆主成分分析和特征重建超网络融合到客户端头部 h_{ψ}ˆ中，以从（F， X）解码三维高斯，而神经会话校准（第3.3节）则在预热期间快速进行在线微调 h_{ψ} 。一次会话在预热时发送其 PCA 基、超网络更新和 NSC 更新；随后每个数据包仅携带系数、范围和锚点，从而实现无状态帧并行流传输（第3.4节）和独立解码。

3.1. 服务器端令牌压缩

基于 PCA 的令牌降维。 令 $\mu \in \mathbb{R}^{1024}$ 和 $U_r \in \mathbb{R}^{1024\times r}$ 分别为会话首帧令牌的均值和前 r 个主成分。每个令牌的编码与解码方式如下

$$c_i = U_r^{\top}(f_i - \mu), \qquad \hat{f}_i = \mu + U_r \hat{c}_i, \tag{1}$$

其中$\hat{c}_i$ 为量化系数。在整个会话期间保持 (μ, U_r) 不变，可确保编码严格遵循因果顺序。

每帧为每个分量发送一个取值范围，将范围内的系数均匀量化为 b 位，并按分量优先的顺序使用 Zstd 压缩。轻量版层级 (r, b) =(32, 4)、默认的 Pro (48,6) 和 Ultra (80,6) 共用这一表示形式；其中，一个量化后的 Pro 令牌占用 $48 \times 6 = 288$ 位，而 BF16 令牌占用 1024×16 位。由于 Karhunen–Lo`eve 变换 [19], PCA 可在拟合帧上以 r 秩最小化平方线性投影误差，且无需针对冻结的重建器训练编码器，因此 Pro 基仅占 201 千字节，且只需发送一次。

无损锚点编码。 由于锚点误差会使整个组发生偏移，因此锚点坐标采用无损编码。在体素网格上，每个数据包存储坐标表、每组占用的 $3 \times 3 \times 3$ 邻域（以 32位掩码表示），以及每个槽位的一个索引，并通过 Zstd 编码，大小为 35.0 千字节/帧。

不透明度引导的令牌剪枝。 许多令牌组仅生成近乎透明的高斯，服务器无需运行头部即可识别这些组。轻量级预测器g_{ϕ} 对原始令牌的每个维度进行标准化，并通过无偏置的 $1024 \to 16$投影和 $16 \to 32 \to 16$SiLU层，回归未压缩头部的 16 个槽位不透明度逻辑值。当且仅当组 i满足以下条件时，才予以保留

$$q_i = \max_k \sigma\big(g_{\phi}(f_i)_k\big) \;\geq\; \delta, \qquad \delta = 0.075. \tag{2}$$

最大值会保留至少有一个可见槽位的每个组。通过未压缩头部的 logits 进行训练后，同一个预测器和阈值适用于所有层级、序列和拍摄装置（第 S2节）。

3.2. 客户端令牌恢复

将逆主成分分析融合到高斯头部中。 头部以一个线性映射 $y = Wf + b_0$ 开始，随后才是其第一个激活函数。将式（1）代入可得

$$y = (WU_r)\,\hat{c} + (W\mu + b_0). \tag{3}$$

预先计算这两项并将逆主成分分析融合到高斯头部中，从而避免显式扩展为 1024 维令牌；融合解码与非融合解码仅相差 0.002 dB。

特征重建超网络。 尽管主成分分析会独立地对每个令牌进行线性投影，但保留的系数、局部锚点几何和相邻组仍包含关于被丢弃分量的非线性空间线索。客户端通过残差超网络H_{θ}恢复这些信息，

$$\tilde{f}_i = \hat{f}_i + H_{\theta}\Big(\hat{f}_i,\; X_i - \bar{X}_i,\; \tfrac{1}{8}\textstyle\sum_{j\in\mathcal{N}_i}\hat{f}_j;\; r, b\Big), \tag{4}$$

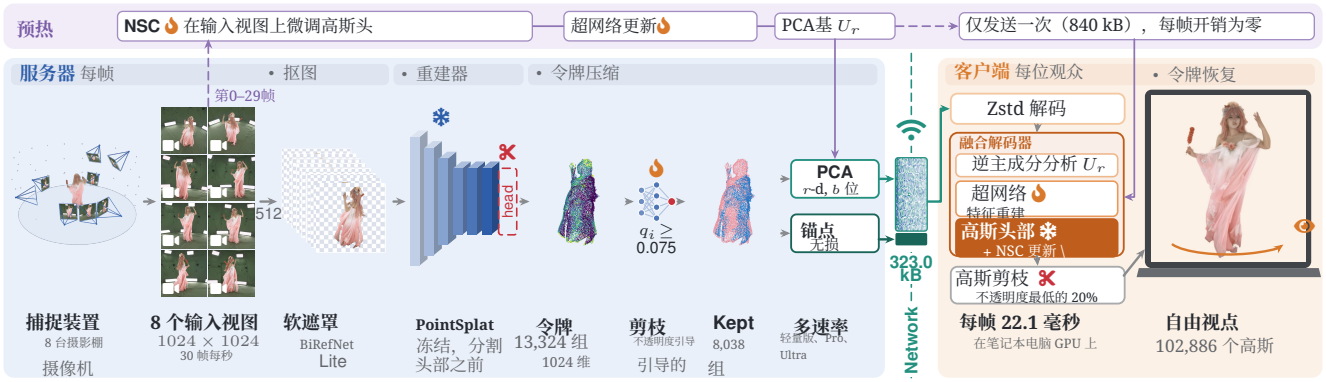

图3. **HoloSplat 在高斯头之前拆分冻结的前馈式重建器，并通过结合服务器端令牌压缩和客户端令牌恢复的多速率编解码器传输令牌。** 在服务器端（蓝色），每帧依次进行八视图抠图、PointSplat（直至其头部）、不透明度引导的令牌剪枝、主成分分析降至秩 r 并进行 b位量化，以及无损锚点编码。在客户端（橙色），融合解码器将逆主成分分析、轻量级特征重建超网络和 NSC 更新整合到高斯头中，然后剪除不透明度最低的 20% 高斯。在预热阶段（紫色），NSC 使用第0–29帧的输入视图对高斯头进行微调；其更新、超网络更新以及零帧主成分分析基仅发送一次。按比例绘制的 Pro 数据包（0678 02的第90帧）总计 323.0 千字节（其中 270.2 千字节为系数、51.4 千字节为锚点，1.4 千字节为头部；❄冻结，\ 学习）。

其中，$X_i - \bar{X}_i$ 是以质心 X_i 为中心的锚点布局，$\mathcal{N}_i$ 则包含同一帧中最近的八个组。层级（r、b）的嵌入会为共享的四层、宽度为1024的多层感知机生成 FiLM [56] 参数，因此一个网络即可适用于所有层级。由于头部的第一层通过 W投影令牌，H_θ通过直接针对未压缩令牌最小化 $\|W(f_i - \tilde{f}_i)\|_2^2$进行训练，无需渲染。由于 f_i和邻居均值都是的线性函数，H_θ 的第一层与 U_r 融合（在 r 维空间中对邻居取平均），其输出层则融合进 W，因此恢复后的令牌也不会被显式展开。H_θ 仅需在160个 DNA-Rendering 训练场景上训练一次，并在三个留出主体上进行选择，随后与客户端一同交付。

超网络更新与客户端剪枝。在预热阶段，服务器针对每个流式层级，使用第0–29帧上的未压缩令牌（FP16 格式下为 593.8 kB），为 H_θ 的四层拟合一个秩为32的子空间更新 [22]；客户端将其融合到 H_θ 中，每帧无需额外开销；加上主成分分析基和 NSC 更新后，Pro 会话状态总计为 840 kB。随后，客户端剪枝掉不透明度最低的 20% 高斯。

3.3. 神经会话校准

预训练的高斯头即使未经压缩，在新的相机参数和变化的采集条件下也会出现质量损失（图4）。预热期间，我们在四项流式传输约束下，利用输入相机视图在线快速微调高斯头：(i) 所有观看者的令牌流保持不变；(ii)自监督具有因果性，使用会话中评估帧之前的输入视图；(iii)不增加每帧开销；(iv) 一次性更新保持精简。

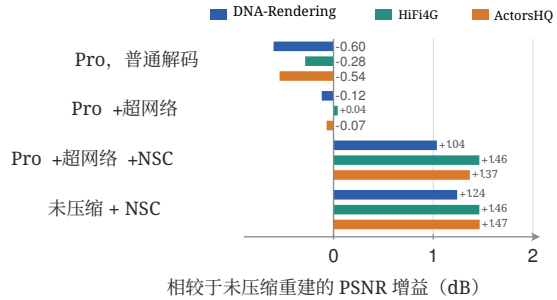

图 4. **神经会话校准可提升各数据集上的质量，而令牌编码的成本很低。** 仅通过逆主成分分析解码时，Pro 码流在采用相同剪枝的情况下，峰值信噪比比我们的未压缩重建最多低 0.60 dB；超网络将差距控制在 0.12 dB 以内，而 NSC 在此基础上分别提升 +1.154、+1.419 和 +1.436 dB。在 16 个 DNA-Rendering 测试场景以及所有帧的留出相机上进行评估；HiFi4G 和 ActorsHQ 均在一秒预热后进行评估。NSC 在输入视图上拟合，未压缩 + NSC 不使用编解码器。

自监督目标。令 $\mathcal{W}$为会话的前 30 帧（1 秒的采集），F_t 为流式层级下帧 t 的恢复令牌，$\mathcal{R}_v$ 为剪枝后渲染至输入视图 v 所得的图像$I_{v,\,t}$。服务器求解

$$\min_{\Delta} \sum_{t\in\mathcal{W}} \sum_{v=1}^{8} \mathcal{L}\Big(\mathcal{R}_v\big(h_{\psi+\Delta}(\tilde{F}_t, X_t)\big),\, I_{v,t}\Big) \qquad (5)$$

对于每个流式层级，其 $\mathcal{L} = 0.8\,\ell_1 + 0.2\,(1 - $结构相似性指数$)$ +0。1 学习感知图像块相似度，同时冻结编码器、剪枝预测器、编解码器和超网络，以满足 (i) 和 (ii)。

低秩高斯头部更新。 为了联合调整高斯位置、尺度、不透明度和颜色，我们将 Δ 参数化为秩5更新 $\Delta W_\ell = A_\ell B_\ell$，应用于头部的三个逐槽位多层感知机层 ℓ [22], 仅需 20,480 个 FP16 参数（41.0 千字节；包含头部和校验和时为 44.5 千字节，占平均 Pro 数据包的 21%），从而满足 (iv)；完整头部更新则会 814× 大得多。在逆 PCA 融合（公式（3））之前，将每个 $A_\ell B_\ell$ 融合到对应层中一次，即可保持头部形状和逐帧开销不变，从而满足 (iii)。

3.4. 实时帧并行流传输

由于帧是独立编码的，服务器会跨帧并行处理。每个图形处理器在 CUDA MPS [52] 下运行三个常驻工作线程进程；传入帧会分发到待处理帧数最少的图形处理器上，使抠图、推理、令牌剪枝和 PCA 投影能够重叠执行，而无需设置全局屏障；与此同时，由一个中央编码器负责量化数据包并使用 Zstd 编码。最多允许 30 帧同时在途，以限制内存占用；重排序缓冲区则严格按照采集顺序释放数据包。

4. 实验

4.1. 实验设置

数据集与指标。 我们遵循 DNA-Rendering 上的官方 PointSplat 评估协议 [8]（16 个测试场景，8 个输入相机和 48 个评估相机，取 512×512、第 30 帧、预热后），此外还在三个未见过的设备上进行评分：在一秒预热后，基于所有帧的留出视图进行评估：HiFi4G [27], ActorsHQ [24], 和合成 X-Humans [60]（第 S5节）。在 DNA-Rendering 上，使用 BiRefNet Lite 生成的软掩码 [92] 针对 DNA-Rendering 掩码进行微调后，用于合成采集的输入和真实值。我们在黑色背景上、真实值掩码的边界框内报告 PSNR、SSIM 和 LPIPS（VGG），并对各场景赋予相同权重。

基线方法与统计口径。 标准 PointSplat512 是已发布的模型，每个高斯均采用原生不透明度剪枝，并包含 14 个 32 位浮点数属性。所有层级均采用不透明度引导的令牌剪枝、融合式特征重建超网络及其一次性更新、NSC 和 20% 的客户端剪枝；数据包大小包含每个编码帧的所有辅助信息。服务器使用一至六块 RTX 5090 GPU 运行完整的流传输流水线；我们报告稳态吞吐量，单位为帧/秒（第 S7节）。

4.2. 令牌流传输质量与码率

在 DNA-Rendering（表 1）上，我们的未压缩路径与 PointSplat512 相当（+0.000 dB；表 S4）。头部前令牌压缩将 32 位浮点数高斯缩减 22.8× ，在 Pro 层级下达到 208.2 千字节/帧；而特征重建超网络和 NSC 使 Pro 相较于该方案提升 1.050 dB；在全部 16 个序列的第 30–89 帧中，Lite/Pro/Ultra

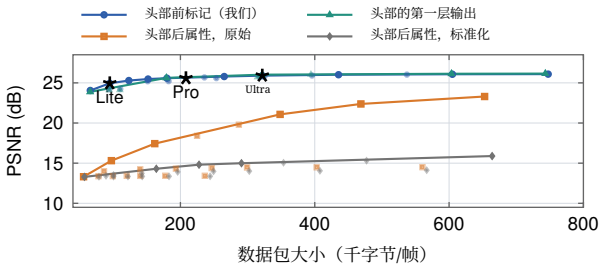

图 5. **在高斯头之前对令牌进行编码，在各种码率下比头部后属性编码高出 2.8–10.3 dB。** 在 16 个 DNA-Rendering 测试场景上，采用相同的会话PCA、量化和 Zstd编码器，在冻结的重建器的三个拆分点进行测试，并使用不透明度引导的令牌剪枝且不使用超网络。曲线连接每个拆分点的最佳设置；星号标记轻量版、Pro 和 Ultra 设置。

分别保持 25.95/26.30/26.38 dB（表 3）。在未见过的 HiFi4G、ActorsHQ 和 X-Humans 数据集上，每个层级的 PSNR 和 SSIM 均优于 PointSplat512；Pro 和 Ultra 在这三个数据集上的 LPIPS 也都有所改善，分别提升 1.34–1.50、1.28–1.42 和 0.54–1.08 dB，同时负载缩小 11.7–39.5×、16.9–59.8×和 11.7–39.3× 倍。

不透明度引导的令牌剪枝。 令牌剪枝移除了 23.6% 的令牌组而不损失峰值信噪比，并使 Ultra 数据包缩小 23.5%。在六种编解码器设置下（表 S8），剪枝后的码流 BD-PSNR 提高 0.12 dB，BD 码率降低 25.1% [1]。

特征重建超网络与融合解码。 对于相同数据包，超网络（经过一次性更新）使轻量版、Pro 和 Ultra 的性能分别提升 +0.822,+0.483 和 +0.191 dB，相比普通 PCA解码分别恢复了 66.7%、80.4% 和 68.5% 的压缩损失（表 S42）；使用 NSC 时，提升分别达到 +1.890, +1.636 和 +1.369 dB（在全部 16 个场景上均为正值），且在所有层级上都优于不使用 NSC 的未压缩令牌（图4 和表 S41）。

4.3. 头部前标记与头部后属性压缩

我们在重建器的三个拆分点处应用第 3.1 节中的编码器，但不使用超网络（图 5）。无论是原始形式还是标准化后，每组 224 个预激活属性在秩为 80 时都能保留 97.9–99.9% 的方差，而令牌只能保留 81.9%；然而，对这些属性进行编码并不可行，因为 1% 的相对属性扰动会造成 14，446× 这相当于等能量令牌扰动所造成的渲染均方误差，其影响由指数尺度驱动（第S3节）。在 653 千字节时，它们达到 23.30 dB，比 Lite 在其有效载荷大小的 6。9×低 1.65 dB。对头部 16,384 维第一层输出进行编码，在秩为 80 和 160 时仅提升 0.04–0.12 dB；对数尺度量化的 3DGS 编解码器需要 3。47× Pro 的付

方法	DNA-Rendering（16 个场景）				HiFi4G				ActorsHQ				X-Humans（合成）				
	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	千字节/帧 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	千字节/帧 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	千字节/帧 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	千字节/帧 ↓	
PointSplat512	26.169	0.8650	0.1034	4753.1	22.810	0.7017	0.2137	2564.6	24.438	0.8131	0.1749	10631.7	26.189	0.8692	0.1057	2552.0	
未压缩 + NSC	27.423	0.8857	0.0884	–	24.271	0.7553	0.1895	–	25.904	0.8355	0.1642	–	27.445	0.9014	0.0878	–	
我们的轻量版 (32/4)	26.840	0.8700	0.1043	**94.8**	24.148	0.7414	0.2069	**65.0**	25.721	0.8274	0.1809	**177.9**	26.733	0.8877	0.1073	**65.0**	
我们的专业版 (48/6)	27.219	0.8802	0.0941	208.2	24.270	0.7511	0.1953	141.6	25.806	0.8320	0.1703	404.4	27.188	0.8965	0.0947	141.1	
我们的 Ultra 版 (80/6)	**27.273**	**0.8818**	**0.0923**	321.9	**24.306**	**0.7519**	**0.1931**	219.2	**25.854**	**0.8333**	**0.1680**	628.3	**27.267**	**0.8982**	**0.0922**	217.5	

表 1. **HoloSplat 发送的高斯字节数仅为 PointSplat 的一小部分，却在四种拍摄装置的所有层级上取得了高于 PointSplat 的峰值信噪比和结构相似性指数。** 在黑色背景下，于真实值掩码边界框内评估图像质量，并统计每帧平均千字节数；留出相机来自 HiFi4G、ActorsHQ 和合成 X-Humans。所有层级均使用融合超网络解码器和 NSC；第二行（未压缩令牌与 NSC）为参考结果，粗体表示最佳的非参考结果。

格式	千字节/帧↓	峰值信噪比 ↑	结构相似性指数 ↑	解码（毫秒）↓
PointSplat512，32 位浮点数	4753.1	26.169	0.8650	–
未压缩 +NSC，32 位浮点数	3630.7	27.423	0.8857	–
SPZ v4	852.7	27.357	0.8844	27.9
PlayCanvas 压缩 PLY	1056.3	27.364	0.8846	209.4
FCGS [7]	866.5	27.207	0.8777	4562.7
ksplat	1564.2	27.197	0.8824	154.6
Draco	723.0	27.415	0.8854	15.3
PlayCanvas SOG	841.2	27.226	0.8827	221.2
CodecSplat [87]	45.4	23.143	0.7434	199.1
我们的轻量版	94.8	26.840	0.8700	21.7
我们的专业版	208.2	27.219	0.8802	22.1
我们的 Ultra 版	321.9	27.273	0.8818	22.8

表2。 **采用相同标定时，最小的高斯格式至少需要 Ultra 字节数的 2。24×倍，以及 Pro 字节数的 3。47× 倍。** 各格式均以默认或推荐设置下的 NSC 对我们未压缩的剪枝高斯进行编码；CodecSplat 则以最高码率对来自我们八个视图的特征进行编码，不包括输入掩码。解码：在 RTX 3050 Ti 笔记本电脑（我们的方法）、RTX 5090（CodecSplat）或 CPU 参考解码器（各格式；FCGS 使用图形处理器）上逐帧解码为高斯；表 S12未列出 NSC。

方法	视图	千字节/帧↓	秒/帧 ↓	峰值信噪比↑	学习感知图像块相似度 ↓
我们的轻量版	8	82–139	0.13–0.17	25.95	0.118
我们的专业版	8	166–296	0.13–0.18	26.30	0.107
我们的 Ultra 版	8	252–455	0.13–0.18	26.38	0.104
HiCoM [16]	8	2969–3994	2.0–2.8	17.68	0.389
QUEEN-l [17]	8	50038–57680	1.2–1.4	15.91	0.399
3DGStream [65]	8	8019–8133	2.2–5.2	23.57	0.268
4C4D [97]（离线）	8	3873–39878	30–124	24.68	0.184
HiCoM [16]	40	644–3309	2.1–2.4	19.60	0.346
QUEEN-l [17]	40	1707–199177	4.3–23.4	20.67	0.313
3DGStream [65]	40	7916–8105	2.1–2.6	25.58	0.210

表3。 **在相同的八个视角下，HoloSplat 的每个层级在峰值信噪比、学习感知图像块相似度和字节数方面，均优于所有逐帧优化方法；Pro 在这三项指标上均优于离线 4C4D，且覆盖全部 16 个测试序列。** DNA-Rendering 第 30–89 帧的官方设置（第 S5节）。

加载大小（第4.4节），且在 DepthSplat 和 GPS-Gaussian 上，头部前令牌编码同样优于头部后属性编码（第S5节）。

4.4. 与最先进方法的比较

逐帧优化。 HiCoM [16], QUEEN [17], 和 3DGStream [65] 在共享 GPU 上处理第零帧耗时 50–611 秒，处理每个后续帧耗时 1.2–23.4 秒；相比之下，我们在单个 GPU 上仅需 0.13–0.18 秒， 16× 在 8 和 40 个视角下平均速度更快。在空闲 GPU 上，最快的方法处理每帧也需 0.87 秒，因此即使理想扩展，达到 30 帧每秒至少也需要 27 个 GPU （表 S26 ）；较新的方法 [34, 83] 耗时 2.67 秒或更久（表 S29）。

在相同的八个视角下，HoloSplat 的每个层级在全部 16 个序列上的 PSNR、学习感知图像块相似度和比特率均优于这三种方法（表 3和图 6；图1，右），时间差分峰值信噪比也提升了 2.55–4.59 dB（第 S4节）；即使使用 40 个视角，这三种方法的平均 PSNR 仍低于使用八个视角的轻量版。离线 4C4D [97], 一次性拟合全部 60 帧，在每个序列上的表现都不如 Pro——在其中位58×Pro 有效载荷大小下，平均低 1.62 dB（表S16）。

3DGS 压缩方法。我们对经过相同 NSC 处理的未压缩高斯（3630.7 千字节/帧）采用默认或推荐设置，使用 SPZ v4 [51]、PlayCanvas 压缩 PLY、SOG、ksplat、Draco 和 FCGS [7] 进行压缩。体积最小的 Draco 需要 3。47× Pro 有效载荷大小的 173 兆比特/秒（30 帧每秒）（表 2）；与 Draco、FCGS 和均匀量化相比，我们的层级可将 BD-rate 降低 67–85%（图 S8），而 V^3-style x265 [75] 峰值为 16.36 dB（表 S13）。

潜在特征编码。 在我们的八个视角上，CodecSplat 发布的静态场景模型 [87] 在最高码率下达到 23.1 dB 后便停滞，比 Lite 低 3.7 dB ；而 Pro 在每个序列上的表现都更优，平均高出 4.1 dB（表 2 和 S32）。其解码器在 RTX 5090 上每帧耗时 199 毫秒，超出每帧客户端预算 6。0× 我们的融合客户端则能在同一图形处理器上以 3.9 毫秒解码 Pro， 51× 速度更快（表 S31）。

数字人系统。 逐序列人体重建方法每帧耗时 6 秒至 12 分钟（表 S30）；HiReFF 在 RTX 5090 上达到 4.0 帧/秒，低于我们的单 GPU 吞吐量（表 S31），而 H.265 高斯映射流相比视频编码约损失 2dB [70]。

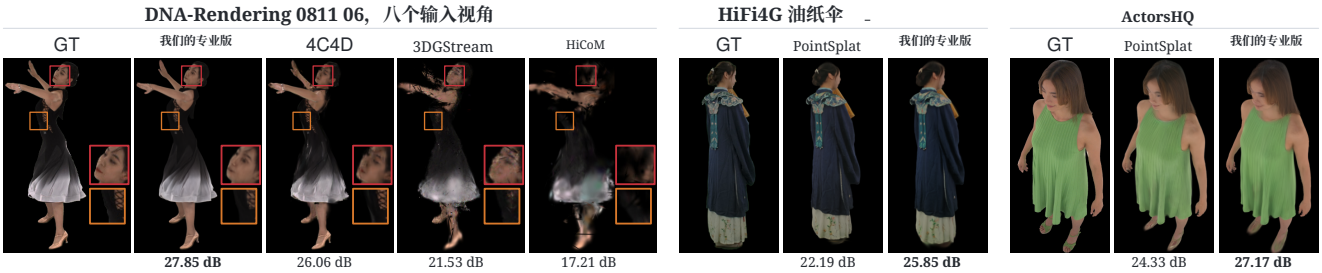

图 6. **在八个视图下，HoloSplat 保留了面部细节和服装系带；离线 4C4D 会模糊面部，3DGStream 会将系带碎裂成针状，HiCoM 则会将两者都模糊，而 NSC 能纠正 PointSplat 在未见过的设备上的外观误差。**留出视图及各视图的峰值信噪比；插图放大展示 2× 面部和系带。基线方法适配我们的八个输入视图；我们的 Pro 使用融合超网络解码器和 NSC。

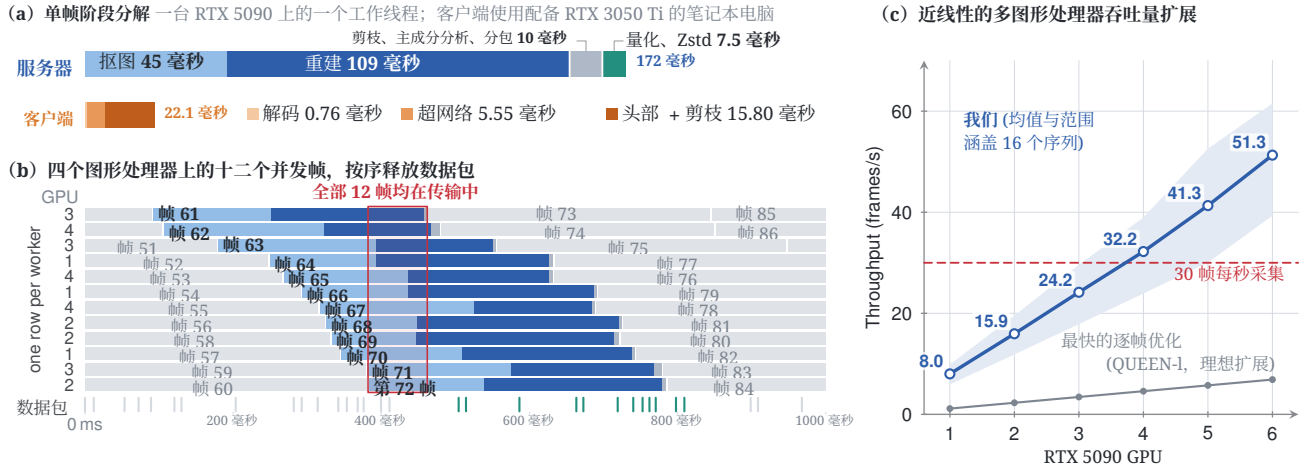

图 7. **单个图形处理器处理一帧需要 172 毫秒，但帧并行工作进程可通过重叠执行，使吞吐量随图形处理器数量近乎线性增长。**(a) 服务器各阶段耗时的均值和客户端耗时的中位数；(b) 四个图形处理器上持续一秒的时间窗口（每行对应一个工作进程），其中第 61–72 帧（红色）并发执行，数据包按采集顺序发出；(c) 所有 16 个序列的吞吐量。

4.5. 实时帧并行流传输

在一块 RTX 5090 上，一帧在抠图中耗时 45 毫秒，在重建中耗时 109 毫秒，随后用 10 毫秒剪枝、投影并分包令牌，再用 7.5 毫秒对其进行量化和压缩（图 7a）。在四块图形处理器上，全部十二个工作线程在一秒窗口内有 99.8% 的时间处于活动状态，并按采集顺序释放数据包（图 7b）。在全部 16 个序列上，吞吐量从单 GPU 时的 8.0 帧/秒扩展到四 GPU 时的 32.2 帧/秒（超过 30 FPS），并在六 GPU 时达到 51.3 帧 / 秒，相当于理想线性缩放的 97–115%（图 7c）；从五块图形处理器开始，全部 16 个序列均超过 30 FPS，此时通过 Wi-Fi 将实时 30 FPS 流传输至 RTX 3050 Ti 笔记本电脑的平均端到端延迟为 204–304 毫秒，且无漂移（第 S7）。无状态数据包还支持每一帧的随机访问、位精确的层级切换以及单帧丢失恢复。

4.6. 神经会话校准

在 HiFi4G 和 ActorsHQ 上，以八个输入视图的首秒进行校准后，在留出相机上评分（第 S6 节），普通 PCA 解码的峰值信噪比比我们的方案低至多 0.76 dB

未压缩重建，而在 DNA-Rendering 上训练的超网络将这一差距缩小至至多 0.13 dB（图4 和表 S37）。相比之下，12 参数的颜色仿射变换仅使峰值信噪比变化 +0.076 和 +0.119 dB，而 NSC 在 Pro 下提升了 +1.419 和 +1.436 dB，因此即使是轻量版，其表现也优于未经校准、未剪枝的重建 +1.327 和+1.282 dB；即使不再进行更新，NSC 在每个序列最后评估的帧上仍可提升 +1.19 和+1.69 dB。

5. 结论

HoloSplat 通过将前馈式重建器在其高斯头部之前拆分为多速率令牌编解码器，并结合不透明度引导剪枝、PCA降维和无损锚点编码，满足实时自由视点视频的三项逐帧预算。Pro 传输的数据量比 float32 高斯少 22.8×，而融合式特征重建超网络及其一次性更新可在笔记本电脑 GPU 上用 22.1 毫秒恢复 80.4% 的压缩损失；NSC 则在预热期间在线微调头部，更新大小为 44.5 千字节，并且

帧并行编码在四个图形处理器上的平均速度为 32.2 帧每秒。

局限性。 编码帧时不跟踪高斯标识，以牺牲时间对应关系为代价，实现无屏障并行、每帧随机访问以及零误差传播。基于采集的第一秒内容，会话设置在四个图形处理器上耗时 24 秒，之后即可开始实时运行；每帧的流传输或客户端开销均为零。

大语言模型使用声明

人工智能助手（Anthropic Claude 和 OpenAI GPT 模型）协助作者讨论研究思路和论点、检索文献并撰写相关工作、设计实验、编写代码并运行实验和评估、生成合成评估数据并处理数据、分析结果、制作图表，以及起草和编辑论文及补充材料的文本。

参考结果

[1] Gisle Bjøntegaard. RD-curves 的平均 PSNR 差异计算。技术报告 VCEG-M33，ITU-T SG16 Q.6 视频编码专家组，2001 年。6, 19[2] Athanasios Charisoudis, Simone Croci, Kit Yung Lam, Pascal Frossard 和 Aljosa S molic. 面向动态点云和高斯泼溅的快速体积捕获与重建流程。载于 ACM SIGGRAPH 欧洲视觉媒体制作会议（CVMP）论文集，第 1–11 页，2025 年。3 – . [3] Jiacong Chen, Qingyu Mao, Youneng Bao, Xiandong Meng, Fanyang Meng, Ronggang Wang 和 Yongsheng Liang. 运动至关重要：用于自由视点视频重建的紧凑型高斯流传输。载于神经信息处理系统进展（NeurIPS），第 133458–133482 页，2025 年。3, 33[4] Yihang Chen, Qianyi Wu, Weiyao Lin, Mehrtash Harandi 和 Jianfei Cai. HAC：用于三维高斯点元压缩的哈希网格辅助上下文。载于欧洲计算机视觉会议（ECCV）论文集，第 422–438 页，2024 年。3[5] Yue Chen, Xingyu Chen, Anpei Chen, Gerard Pons-Moll 和 Yuliang Xiu. Feat2GS：利用高斯泼溅探测视觉基础模型。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 6348–6361 页，2025 年。3[6] Yuzhong Chen, Yuqin Liang, Zihao Wang, Danying Wang, Cong Xie, Shaohui Jiao 和 Li Zhang. LiveGS：通过面向移动设备的高性能高斯泼溅实现实时自由视点视频。载于 ACM SIGGRAPH 2025 新兴技术，第 1–2 页，2025 年。3[7] Yihang Chen, Qianyi Wu, Mengyao Li, Weiyao Lin, Mehrtash Harandi 和 Jianfei Cai. 快速前馈式三维高斯泼溅压缩。载于国际学习表征会议（ICLR），2025 年。2, 3, 7, 22, 23, 24, 36, 37

[8] Wei Cheng, Ruixiang Chen, Siming Fan, Wanqi Yin, Keyu Chen, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, Daxuan Ren, Lei Yang, Ziwei Liu, Chen Change Loy, Chen Qian, Wayne Wu, Dahua Lin, Bo Dai 和 Kwan-Yee Lin. DNA-Rendering：用于高保真人本位渲染的多样化神经演员库。载于*IEEE/CVF*国际计算机视觉会议（*ICCV*）论文集，第 19982–19993 页，2023 年。3，6

[9] Hyomin Choi、Robert A. Cohen 和 Ivan V. Baji´c。用于深度张量压缩的往返预测。载于《2020 IEEE国际声学、语音与信号处理会议》（ICASSP），第4467–4471页，2020年。3[10] Seungmi Choi、Ji-Hyoun Choi、Suhyun Jo、Tae-Gwon Park 和 Hui Yong Kim。面向3D高斯泼溅的前馈式多视图特征编码，3 20590，2026 2610。arXiv预印本 arXiv：. . [11]Robert A. Cohen、Hyomin Choi 和 Ivan V. Baji´c。用于协同智能的神经网络特征张量轻量化压缩。载于《IEEE国际多媒体与博览会会议》（ICME）论文集，第1–6页，2020年。3[12] Yann Collet 和 Murray S. Kucherawy。Zstandard压缩与application/zstd媒体类型。RFC 8878，互联网工程任务组，2021年。14[13] Mingsong Dou、Sameh Khamis、Yury Degtyarev、Philip Davidson、Sean Ryan Fanello、Adarsh Kowdle、Sergio Orts Escolano、Christoph Rhemann、David Kim、Jonathan Taylor、Pushmeet Kohli、Vladimir Tankovich 和 Shahram Izadi。Fusion4D：具有挑战性场景的实时性能捕捉。《ACM图形学汇刊》，35(4)：114:1–114:13，2016年。3[14] Zhiwen Fan、Kevin Wang、Kairun Wen、Zehao Zhu、Dejia Xu 和 Zhangyang Wang。LightGaussian：无界3D高斯压缩，实现15倍缩减和 200+ 帧每秒。载于《神经信息处理系统进展》（NeurIPS），第140138–140158页，2024年。3[15] Jiaye Fu、Qiankun Gao、Chengxiang Wen、Yanmin Wu、Siwei Ma、Jiaqi Zhang 和 Jian Zhang。ReCon-GS：用于动态场景快速紧凑重建的连续性保持高斯流传输。载于《神经信息处理系统进展》（NeurIPS），第149397–149424页，2025年。3, 33[16] Qiankun Gao、Jiarui Meng、Chengxiang Wen、Jie Chen 和 Jian Zhang。HiCoM：基于3D高斯泼溅的动态可流式传输场景的分层相干运动。载于《神经信息处理系统进展》（NeurIPS），第80609–80633页，2024年。2, 3, 7, 32, 33[17] Sharath Girish、Tianye Li、Amrita Mazumdar、Abhinav Shri-vastava、David Luebke 和 Shalini De Mello。QUEEN：用于自由视点视频流式传输的动态高斯量化高效编码。载于《神经信息处理系统进展》（NeurIPS），第43435 43467页，2024年 2, 3, 7, 32, 33 – 。[18] Google。Draco。https://github.com/google/draco，2017年。17[19] Vivek K. Goyal。变换编码的理论基础。《IEEE信号处理杂志》，18(5)：9–21，2001年。4

[20] Antoine Gu´edon、Tomoki Ichikawa、Kohei Yamashita 和 Ko Nishino。MAtCha Gaussians：从稀疏视图获取高质量几何与照片级真实感的图表图集。载于《IEEE/CVF计算机视觉与模式识别会议》（CVPR）论文集，第6001–6011页，2025年。21[21] Yujie Guo、Yudong Jin、Lingteng Qiu、Zehong Shen、Zhen Xu、Jing Zhang、Xianchao Shen、Hujun Bao、Sida Peng 和 Xiaowei Zhou。PointSplat：通过以人为中心的预测实现紧凑高斯泼溅。载于《欧洲计算机视觉会议》（ECCV）论文集，第615 632页，2026年 2, 3, 4, 40, 42, 43 – 。[22] Edward J. Hu、Yelong Shen、Phillip Wallis、Zeyuan Allen-Zhu、Yuanzhi Li、Shean Wang、Lu Wang 和 Weizhu Chen。LoRA：大语言模型的低秩适应。载于《国际学习表征会议》（ICLR），2022年。4, 5, 6[23] Qiang Hu、Zihan Zheng、Houqiang Zhong、Sihua Fu、Li Song、Xiaoyun Zhang、Guangtao Zhai 和 Yanfeng Wang。4DGC：面向高效可流式传输自由视点视频的速率感知4D高斯压缩。载于《IEEE/CVF计算机视觉与模式识别会议》（CVPR）论文集，第875–885页，2025年。3, 33[24] Mustafa Isık、Martin R¨unz、Markos Georgopoulos、Taras Khakhulin、Jonathan Starck、Lourdes Agapito 和 Matthias Nießner。HumanRF：面向运动中人体的高保真神经辐射场。《ACM图形学汇刊》，42(4)：98:1–98:12，2023年。2, 6, 34[25] ITU-R。IMT-2020无线接口的技术性能最低要求。报告ITU-R M.2410-0，国际电信联盟，2017年。2[26] Yuheng Jiang、Zhehao Shen、Yu Hong、Chengcheng Guo、Yize Wu、Yingliang Zhang、Jingyi Yu 和 Lan Xu。用于沉浸式以人为中心的体积视频的鲁棒双高斯泼溅。《ACM图形学汇刊》，43(6)：265:1–265:15，2024年。3, 33, 34[27] Yuheng Jiang、Zhehao Shen、Penghao Wang、Zhuo Su、Yu Hong、Yingliang Zhang、Jingyi Yu 和 Lan Xu。HiFi4G：通过紧凑高斯泼溅实现高保真人体表演渲染。载于《IEEE/CVF计算机视觉与模式识别会议》（CVPR）论文集，第19734–19745页，2024年。2, 3, 6, 33, 34[28] Yiming Jiang、Hanzhang Tu、Wenfeng Song、Siyou Lin、Liang An、Shuai Li、Aimin Hao 和 Yebin Liu。HiReFF：从未标定稀疏视图视频进行高分辨率前馈人体重建。载于《欧洲计算机视觉会议》（ECCV）论文集，第279 297页，2026年 3, 34 – 。[29] Gyeongjin Kang、Younggeun Lee、Seungjun Oh 和 Eun-byung Park。CodecNeRF：迈向快速编码与解码、紧凑且高质量的新视角合成。载于《AAAI人工智能会议》（AAAI）论文集，第4203–4211页，2025年。4[30] Yiping Kang、Johann Hauswald、Cao Gao、Austin Rovinski、Trevor Mudge、Jason Mars 和 Lingjia Tang。Neurosurgeon：云端与移动边缘之间的协同智能。载于《编程语言与操作系统架构支持国际会议》（ASPLOS）论文集，第615–629页，2017年。3[31] Zhihui Ke、刘宇阳、周晓波和邱铁。Stream-STGS：用于实时自由视点视频的时空高斯网格。在AAAI人工智能会议（AAAI）论文集，第5629–5638页，2026。2、3、33[32] Mark Kellogg。GaussianSplats3D：ksplat格式。https://github.com/mkkellogg/GaussianSplats3D，2023。17[33] 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(4)：139:1–139:14，2023。2[34] 李Wonjoon、Woo Sungmin、金东亨、李俊浩、朴尚宪和李尚允。MoRGS：面向可流式传输的动态3D场景的高效逐高斯运动推理。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第41044–41053页，2026。2、3、7、33。[35] Vincent Leroy、Yohann Cabon和Jerome Revaud。利用MASt3R在3D中进行图像匹配。在欧洲计算机视觉会议（ECCV）论文集，第71–91页，2024。21[36] Hao Li、李思成、高翔、阿布都艾哈提·巴图尔、于璐和廖一一。GIFStream：基于特征流的4D高斯沉浸式视频。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第21761–21770页，2025。3[37] 李灵芝、沈振、王中书、沈力和谭平。用于3D视频合成的流式辐射场。在神经信息处理系统进展（NeurIPS）论文集，第13485–13498页，2022。33[38] Tianye Li、Mira Slavcheva、Michael Zollh¨ofer、Simon Green、Christoph Lassner、金昌日、Tanner Schmidt、Steven Lovegrove、Michael Goesele、Richard Newcombe和吕兆阳。基于多视角视频的神经3D视频合成。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第5521–5531页，2022。21、33[39] Yuxuan Li、Yihang Chen、张宇峰、Jianfei Cai和Weiyao Lin。FeCoSplat：用于前馈式3D高斯泼溅的反馈引导压缩。arXiv预印本arXiv:2609.33330，2026。3[40] 李哲、Zerong Zheng、王立真和Yebin Liu。可动画高斯：学习姿态相关高斯映射，实现高保真数字人建模。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第19711–19722页，2024。2、4、34[41] Zhanfeng Liao、Jiajun Zhang、Hanzhang Tu、王志熙、高云琪、张宏文和Yebin Liu。SharpTimeGS：通过生命周期调制实现清晰稳定的动态高斯泼溅。在IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第11798–11807页，2026。3[42] Haotong Lin、陈思力、Jun Hao Liew、Donny Y. Chen、李振宇、赵阳、Sida Peng、郭恒凯、Xiaowei Zhou、石光、冯家石和康秉义。Depth Anything 3：从任意视图恢复视觉空间。在学习表征国际会议（ICLR）论文集，2026。41、42

Shanchuan Lin、Linjie Yang、Imran Saleemi和Soumyadip Sengupta。具有时间引导的稳健高分辨率视频抠像。在IEEE/CVF冬季计算机视觉应用会议（WACV）论文集，第238–247页，2022。40[43] Stephen Lombardi、Jason Saragih、Tomas Simon和Yaser Sheikh。用于人脸渲染的深度外观模型。《ACM图形学汇刊》，37(4)：68:1–68:13，2018。2、4[44] Ziqi Lu、Heng Yang、Danfei Xu、Boyi Li、Boris Ivanovic、Marco Pavone和Yue Wang。LoRA3D：三维几何基础模型的低秩自标定。在学习表征国际会议（ICLR）论文集，2025。4[45] Yoshitomo Matsubara、Ruihan Yang、Marco Levorato和Stephan Mandt。面向资源受限边缘计算系统的监督式压缩。在IEEE/CVF冬季计算机视觉应用会议（WACV）论文集，第2685–2695页，2022。3[46] Wieland Morgenstern、Florian Barthel、Anna Hilsmann和Peter Eisert。通过自组织高斯网格实现紧凑的3D场景表示。在欧洲计算机视觉会议（ECCV）论文集，第18–34页，2024。3[47] MulticoreWare。x265：免费的H.265/HEVC编码器。https://www.videolan.org/developers/x265.html，2013。18[48]K. L. Navaneet、Kossar Pourahmadi Meibodi、Soroush Abbasi Koohpayegani和Hamed Pirsiavash。CompGS：通过矢量量化实现更小、更快的高斯泼溅。在欧洲计算机视觉会议（ECCV）论文集，第330–349页，2024。3[49] Vu Trung Nghia Nguyen、Yuchen Wang、Kyung Chul Lee和Kevin C. Zhou。QuARC-GS：用于紧凑动态场景流传输的高斯泼溅量化锚定残差编码。arXiv预印本arXiv:[50]。 *2608 18285*, 2026 33

[51] Niantic Labs。SPZ：用于压缩3D高斯点元的文件格式。[3] https://github.com/nianticlabs/spz，2026。我们的比较中使用了版本4和SH阶数0。2、3、7、17[52] 英伟达。CUDA多进程服务。https://docs.nvidia.com/deploy/mps/index.html，2026。访问日期：2026-10-16。6[53] Seungjun Oh、Hyunmo Yang和Eunbyung Park。参数高效的实例自适应神经视频压缩。在亚洲计算机视觉会议（ACCV）论文集，第250–267页，2024。4[54] Sergio Orts-Escolano、Christoph Rhemann、Sean Fanello、Wayne Chang、Adarsh Kowdle、Yury Degtyarev、David Kim、Philip L. Davidson、Sameh Khamis、Mingsong Dou、Vladimir Tankovich、Charles Loop、Qin Cai、Philip A. Chou、Sarah Mennicken、Julien Valentin、Vivek Pradeep、Shenlong Wang、Sing Bing Kang、Pushmeet Kohli、Yuliya Lutchyn、Cem Keskin和Shahram Izadi。Holoportation：实时虚拟3D传送。在ACM用户界面软件与技术研讨会（UIST）论文集，第741–754页，2016。3[55] Weihong Pan、Xiaoyu Zhang、Zhuang Zhang、Zhichao Ye、Nan Wang、Haomin Liu 和 Guofeng Zhang。SparseCam4D：基于稀疏摄像机的时空一致4D重建。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 33237–33247 页，2026 年。3, 21[56] Ethan Perez、Florian Strub、Harm de Vries、Vincent Dumoulin 和 Aaron Courville。FiLM：采用通用条件层进行视觉推理。载于 AAAI人工智能会议（AAAI）论文集，第 3942–3951 页，2018 年。5[57] PlayCanvas。压缩 PLY 格式。https : // developer.playcanvas.com/user- manual/supersplat/editor/import-export，2023 年。17[58] PlayCanvas。SOG 格式。https://developer. playcanvas . com / user - manual / gaussian - splatting/formats/sog，2025 年。17[59] Pedro Quesado、Erkut Akdag、Yasaman Kashefbahrami、Willem Menu 和 Egor Bondarev。LiveStre4m：从非摆拍的多视角视频前馈式实时直播新视角。载于 IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）论文集，第 279–288 页，2026 年。3[60] Kaiyue Shen、Chen Guo、Manuel Kaufmann、Juan Jose Zarate、Julien Valentin、Jie Song 和 Otmar Hilliges。X-Avatar：富有表现力的人类虚拟化身。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 16911–16921 页，2023 年。6, 24[61] Ashwath Shetty、Marc Habermann、Guoxing Sun、Diogo Luvizon、Vladislav Golyanik 和 Christian Theobalt。Holoported Characters：利用稀疏RGB相机实时渲染人类自由视点画面。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 1206–1215 页，2024 年。4[62] Changhao Song、Yuxuan Wang、Qibiao Li、Youcheng Cai 和 Ligang Liu。StreamSplat：流式前馈式3D高斯泼溅。arXiv 预印本 arXiv:2608.01659，2026 年。4[63] Zetian Song、Jiaye Fu、Jiaqi Zhang、Xiaohan Lu、Chuanmin Jia、Siwei Ma 和 Wen Gao。TinySplat：生成紧凑型3D场景表示的前馈方法。IEEE视频技术电路与系统汇刊，36(4):5567–5580，2026 年。3[64] Gary J. Sullivan、Jens-Rainer Ohm、Woo-Jin Han 和 Thomas Wiegand。高效视频编码（HEVC）标准概述。IEEE视频技术电路与系统汇刊，22(12):1649–1668，2012 年。2[65] Jiakai Sun、Han Jiao、Guangyuan Li、Zhanjie Zhang、Lei Zhao 和 Wei Xing。3DGStream：即时训练三维高斯，实现逼真自由视点视频的高效流式传输。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 20675–20685 页，2024 年。2, 3, 7, 32, 33[66] Yu Sun、Xiaolong Wang、Zhuang Liu、John Miller、Alexei Efros 和 Moritz Hardt。利用自监督进行测试时训练，以适应分布偏移下的泛化。载于国际机器学习会议（ICML）论文集，第 9229–9248 页，2020 年。4

[67] Luyang Tang、Jiayu Yang、Rui Peng、Yongqi Zhai、Shihe Shen 和 Ronggang Wang。将可流式传输的自由视点视频压缩至每帧 0.1 兆字节。载于 AAAI人工智能会议（AAAI）论文集，第 7257–7265 页，2025 年。3, 33[68] Team HY-World、Chenjie Cao、Xuhui Zuo、Zhenwei Wang、Yisu Zhang、Junta Wu、Zhenyang Liu、Yuning Gong、Yang Liu、Bo Yuan、Chao Zhang、Coopers Li、Dongyuan Guo、Fan Yang、Haiyu Zhang、Hang Cao、Jianchen Zhu、Jiaxin Lin、Jie Xiao、Jihong Zhang、J unlin Yu、Lei Wang、Lifu Wang、Lilin Wang、Linus、Minghui Chen、Peng He、Penghao Zhao、Qi Chen、Rui Chen、Rui Shao、Sicong Liu、Wangchen Qin、Xiaochuan Niu、Xiang Yuan、Yi Sun、Yifei Tang、Yifu Sun、Yihang Lian、Yonghao Tan、Yuhong Liu、Yuyang Yin、Zhiyuan Min、Tengfei Wang 和 Chunchao Guo。HY-World 2.0：用于重建、生成和模拟3D世界的多模态世界模型。arXiv 预印本 arXiv:2604.14268，2026 年。41, 42[69] Hanzhang Tu、Ruizhi Shao、Xue Dong、Shunyuan Zheng、Hao Zhang、Lili Chen、Meili Wang、Wenyu Li、Siyan Ma、Shengping Zhang、Boyao Zhou 和 Yebin Liu。Tele-Aloha：采用稀疏RGB相机、低预算且高度逼真的远程临场系统。载于 ACM SIGGRAPH 2024 会议论文集，第 1–12 页，2024 年。3, 34[70] Hanzhang Tu、Zhanfeng Liao、Wei Min、Jiajun Zhang 和 Yebin Liu。Tele360：基于稀疏非摆拍摄像机的实时前馈式人体重建。ACM图形学汇刊，45(6):1–18，2026 年。2, 3, 7, 18, 34, 42[71] Ties van Rozendaal、Iris A. M. Huijben 和 Taco S. Cohen。为乐趣与收益而过拟合：实例自适应数据压缩。载于国际学习表征会议（ICLR），2021 年。4[72] Chen Wang、Hao Tan、Wang Yifan、Zhiqin Chen、Yuheng Liu、Kalyan Sunkavalli、Sai Bi、Lingjie Liu 和 Yiwei Hu。tt-tLRM：面向长上下文和自回归三维重建的测试时训练。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 36582–36592 页，2026 年。4[73] Liao Wang、Qiang Hu、Qihan He、Ziyu Wang、Jingyi Yu、Tinne Tuytelaars、Lan Xu 和 Minye Wu。适用于流式自由视点视频的神经残差辐射场。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第 76–87 页，2023 年。3[74] 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（CVPR）论文集，第 470–481 页，2024 年。3[75] Penghao Wang、Zhirui Zhang、Liao Wang、Kaixin Yao、Siyuan Xie、Jingyi Yu、Minye Wu 和 Lan Xu。V^3：通过可流式传输的二维动态高斯，在移动设备上观看体积视频。AC M图形学汇刊，43(6):187:1–187:13，2024 年。3, 7, 18, 33[76] Renhao Wang、Yu Sun、Arnuv Tandon、Yossi Gandelsman、Xinlei Chen、Alexei A. Efros 和 Xiaolong Wang。视频流上的测试时训练。机器学习研究期刊，26(9):1–29，2025 年。4[77] Yiming Wang、Qin Han、Marc Habermann、Kostas Daniilidis、Christian Theobalt 和 Lingjie Liu。NeuS2：用于多视图重建的神经隐式表面快速学习。载于 IEEE/CVF国际计算机视觉会议（ICCV）论文集，第3295–3306页，2023年。3, 34[78] Minye Wu、Zehao Wang、Georgios Kouros 和 Tinne Tuyte-laars。TeTriRF：用于高效自由视点视频的时序三平面辐射场。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第6487–6496页，2024年。3，33[79] Shuzhao Xie、Weixiang Zhang、Chen Tang、Yunpeng Bai、Rongwei Lu、Shijia Ge 和 Zhi Wang。MesonGS：通过高效属性变换实现三维高斯的后训练压缩。载于欧洲计算机视觉会议（ECCV）论文集，第434–452页，2024年。3[80] Haofei Xu、Songyou Peng、Fangjinhua Wang、Hermann Blum、Daniel Barath、Andreas Geiger 和 Marc Pollefeys。Depth-Splat：连接高斯泼溅与深度。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第16453–16463页，2025年。22[81] Zhen Xu、Sida Peng、Haotong Lin、Guangzhao He、Jiaming Sun、Yujun Shen、Hujun Bao 和 Xiaowei Zhou。4K4D：4K分辨率下的实时四维视图合成。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第20029–20040页，2024年。34[82] Zhen Xu、Yinghao Xu、Zhiyuan Yu、Sida Peng、Jiaming Sun、Hujun Bao 和 Xiaowei Zhou。利用时序高斯层次结构表示长时体积视频。ACM图形学汇刊，43(6)：171:1–171:18，2024年。34[83] Jinbo Yan、Rui Peng、Zhiyan Wang、Luyang Tang、Jiayu Yang、Jie Liang、Jiahao Wu 和 Ronggang Wang。Instant Gaussian stream：通过高斯泼溅实现动态场景重建的快速且可泛化流式传输。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第16520–16531页，2025年。2，3，7，33[84] Dingxi Yang、Wenqi Guo、Yue Liu、Jungong Han 和 Zhijin Qin。支持实时沉浸式视频通信的可泛化三维高斯泼溅语义编码。arXiv预印本 arXiv:2604.25330，2026年。3[85] Qi Yang、Geert Van Der Auwera 和 Zhu Li。通过视频编解码器实现轻量级三维高斯泼溅压缩。载于数据压缩会议（DCC）论文集，第362–371页，2026年。3[86] Vickie Ye、Ruilong Li、Justin Kerr、Matias Turkulainen、Brent Yi、Zhuoyang Pan、Otto Seiskari、Jianbo Ye、Jeffrey Hu、Matthew Tancik 和 Angjoo Kanazawa。gsplat：一个用于高斯泼溅的开源库。机器学习研究期刊，26(34)：1–17，2025年。40[87] Pengpeng Yu、Runqing Jiang、Qi Zhang、Dingquan Li、Jing Wang 和 Yulan Guo。CodecSplat：用于前馈式3D高斯泼溅的超紧凑潜编码。载于神经信息处理系统进展（NeurIPS），2026年。2，3，7，22，34[88] Tao Yu、Zerong Zheng、Kaiwen Guo、Jianhui Zhao、Qionghai Dai、Hao Li、Gerard Pons-Moll 和 Yebin Liu。DoubleFusion：利用单个深度传感器实时捕捉带有内部人体形状的人体表演。载于

IEEE计算机视觉与模式识别会议（CVPR）论文集，第7287–7296页，2018年。3[89] Tao Yu、Zerong Zheng、Kaiwen Guo、Pengpeng Liu、Qionghai Dai 和 Yebin Liu。Function4D：利用极稀疏的消费级RGBD传感器实时进行人体体积捕获。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第5746–5756页，2021年。3[90] Yuheng Yuan、Qiuhong Shen、Shizun Wang、Xingyi Yang 和 Xinchao Wang。Test3R：学习在测试时进行三维重建。载于神经信息处理系统进展（NeurIPS），第111375–111394页，2025年。4[91] Wenkang Zhang、Yan Zhao、Qiang Wang、Zhixin Xu、Li Song 和 Zhengxue Cheng。D-FCGS：面向自由视点视频的动态高斯泼溅前馈压缩。载于 AAAI人工智能会议（AAAI）论文集，第16361–16369页，2026年。3[92] Peng Zheng、Dehong Gao、Deng-Ping Fan、Li Liu、Jorma Laaksonen、Wanli Ouyang 和 Nicu Sebe。高分辨率二分图像分割的双边参考。CAAI人工智能研究，3：9150038，2024年。4，6[93] Shunyuan Zheng、Boyao Zhou、Hanzhang Tu、Ruizhi Shao、Boning Liu、Shengping Zhang、Liqiang Nie 和 Yebin Liu。GPS-Gaussian：用于实时人体新视角合成的可泛化逐像素三维高斯泼溅。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第19680–19690页，2024年。2，3，4，22，34，41，42，43[94] Zihan Zheng、Houqiang Zhong、Qiang Hu、Xiaoyun Zhang、Li Song、Ya Zhang 和 Yanfeng Wang。JointRF：动态神经辐射场表示与压缩的端到端联合优化。载于 IEEE国际图像处理会议（ICIP）论文集，第3292–3298页，2024年。33[95] Zihan Zheng、Zhenlong Wu、Houqiang Zhong、Yuan Tian、Ning Cao、Lan Xu、Jiangchao Yao、Xiaoyun Zhang、Qiang Hu 和 Wenjun Zhang。4DGCPro：用于渐进式体积视频流传输的高效分层四维高斯压缩。载于神经信息处理系统进展（NeurIPS），第111881–111906页，2025年。33[96] Boyao Zhou、Shunyuan Zheng、Hanzhang Tu、Ruizhi Shao、Boning Liu、Shengping Zhang、Liqiang Nie 和 Yebin Liu。GPS-Gaussian+：用于从稀疏视图进行实时人景渲染的可泛化逐像素三维高斯泼溅。IEEE模式分析与机器智能汇刊，2025年。3[97] Junsheng Zhou、Zhifan Yang、Liang Han、Wenyuan Zhang、Kanle Shi、Shenkun Xu 和 Yu-Shen Liu。4C4D：四相机四维高斯泼溅。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第11829–11839页，2026年。3，7，21[98] Lei Zhou、Haoyu Wu、Akshat Dave 和 Dimitris Samaras。学习在测试时利用先验进行三维重建。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，第36550–36560页，2026年。4[99] Yifeng Zhou、Shuheng Wang、Wenfa Li、Chao Zhang、Li Rao、Pu Cheng、Yi Xu、Jinle Ke、Wenduo Feng、Wen Zhou、Hao Xu、Yukang Gao、Yang Ding、Weixuan Tang 和 Shaohui Jiao。Live4D：用于可流式传输的体积视频的实时捕获系统。载于 *SIGGRAPH Asia 2023* 技术通讯，第 1–4 页，2023 年。 3，34

HoloSplat：实时 令牌流式传输，实现前馈式人体重建

补充材料

S1. 更多质量结果

图 S1 展示了使用主论文中的冻结重建器对两位表演者进行实时流式传输的效果。我们的未压缩重建与标准 PointSplat512 相匹配（+0.000dB，表 S4）。本节所有表格均采用 512×512 全图重建、经带掩码标注的 DNA-Rendering数据微调的 BiRefNet Lite 生成的软掩码，以及黑色背景。表 S1 评估整幅图像，表S2测量真实值（GT）掩码之外的透明度和颜色，表 S3列出每个场景的峰值信噪比和字节数，表S4还列出了客户端高斯数量。

S2. 编解码器与剪枝对照

令牌剪枝会移除令牌组及其对应字节，而客户端剪枝则在数据包大小不变的情况下移除高斯（图 S2 和表 S5）。表 S7 在令牌数量相同的条件下，比较学习得到的组评分与精确不透明度选择和随机选择，图S4则对令牌保留比例进行扫描。表 S6比较主成分分析编码、SPZ 与乘积量化。将 PCA 基融合到头部中（公式（3））会使峰值信噪比变化 0.002 dB，在 (r, b)=(80,6)；因此，表 S5和 S8 以及图 5 使用融合头解码普通数据包，其他普通行则不使用融合头解码。所有特征编解码器共用无损锚点流，如图S3 所示，该图将其与各层级的系数分开。

熵编码与帧间编码对照。帧内编码生成的数据包比帧间编码更小，且每个数据包都保持无状态。在 16 个 DNA-Rendering 测试场景中（帧 1–30，对真实 Lite/Pro/Ultra 数据包进行位精确往返），在 1 cm 世界体素网格上匹配相邻帧中的令牌，并对匹配的整数主成分分析编码进行差分编码，可使 Lite/Pro/Ultra 的特征字节分别减少 1.82%/0.59%/−0.51%。然而，无序表面组需要显式的帧间令牌置换，这会使几何字节增加 31.6%，并使整个数据包增大 6.12%/4.98%/4.00%。在第 0 帧上训练、并由客户端以确定性方式重建且不产生网络传输字节的因果 8kB Zstd [12] 字典，可使帧 1 的整个数据包字节数减少 0.19%/0.14%/0.09%

–30. 对于熵编码（第 10、20 和 30 帧），采用第零帧概率表的位精确、32 位、面向字节的范围非对称数制（rANS）编码，配合静态零阶模型，在 Pro 下可节省整个数据包 9.35% 的字节（Zstd 为每个系数 5.00 位，而对照为 5.65 位；Lite 下为 −0.59 %，Ultra 下为 10.42%）；令牌内前缀能量条件化最多可额外节省每个系数 0.03 位。在单核 CPU 上，rANS 解码每个 Pro 帧耗时 2.38 毫秒，而 Zstd 解包和上传耗时 0.76 毫秒。在笔记本电脑上，Zstd 解包和上传耗时低于一毫秒，因此我们保留 Zstd，且这一阶段的耗时也低于一毫秒。

令牌剪枝训练与选择。预测器会对每个输入维度进行标准化，并将标准差下限设为 0.05。标签来自未压缩头部，未经主成分分析或量化。我们在 0008 01、0019 06、0123 02、0206 04 和 0219 07 各自 50 帧的全部 384,000 个令牌上，对 16 个槽位不透明度逻辑值进行回归，使用 smooth-ℓ_1 损失，并加上权重为 0.5 的最大逻辑值损失（AdamW，学习率 3×10^{-3}，权重衰减 10^{-4}，批大小 8192，训练 48 个周期），并在 0165 08 和 0235 11 上进行检查点选择。部署模型和阈值从 53 个候选方案中选出。每个候选方案在 0219 07 和 0235 11 上进行 20% 客户端剪枝后，相较于未压缩重建的峰值信噪比下降不得超过 0.02 dB；在令牌数量不超过最低值 1% 的候选方案中，我们选择这两个场景上峰值信噪比最高的方案。

S3. 码率—质量与剪枝研究

头部前标记能够容忍高斯属性无法承受的编码误差。图 S5扫描了 48 种秩和位宽设置，据此确定 Lite 为 32/4、Pro 为 48/6、Ultra 为 80/6；表S8则对令牌剪枝、编解码器和客户端剪枝进行交叉比较。为单独分析高属性方差为何无法保留渲染质量（第 4.3 节），我们在四个场景（0012 09、0345 08、0536 06、0811 06；48 个摄像机）中，于进行 20% 客户端剪枝前，在每个拆分点扰动第 30 帧的保留组。对于每个保留组的分割表示 x、令牌 T、隐藏特征 H 或 224 个预激活属性 A，我们添加$\sqrt{\rho}\, z$，其中 z 是一个固定的高斯方向，其缩放方式使得$\|z\|^2 = \sum_{i,c}(x_{ic} - \bar{x}_c)^2$，相当于 x 在保留组内相对于其逐通道均值的能量。因而，该扰动携带每种表示自身中心化能量的 $\rho \in \{0.1\%, 1\%, 3\%, 10\%, 30\%\}$。在 ρ=1% 时，扰动T 或 H 使峰值信噪比变化不到 0.01 dB。在相同的相对扰动幅度 ρ下，扰动 A 会使峰值信噪比下降 16.38 dB（$26.46 \rightarrow 10.08$ dB）；对摄像机和场景取平均后，扰动后的渲染结果与未压缩渲染结果之间的均方误差，是扰动 T时的 14,446× 倍。对数尺度是导致这一失败的主要因素。以 (r, b)=(80,6) 对单独的属性类型进行编码时，仅对尺度进行有损编码会使峰值信噪比降至 11.25 dB；而仅对其他某种属性类型进行有损编码时，峰值信噪比为 21.81–24.68 dB。仅保留精确的尺度、并对其余四种类型进行有损编码，可将峰值信噪比恢复至 19.75 dB；相比之下，所有类型均进行有损编码时为 14.03 dB。指数函数 exp(log-scale) 会将微小的预激活误差放大为显著的协方差和屏幕覆盖范围偏移。即使在 (160,8) 设置下，仅对尺度进行有损编码也会使峰值信噪比从



图 S1。**HoloSplat 使用同一个冻结的重建器流式处理不同的表演者和动作**。DNA- Rendering 0678 02 和 0013 08 在客户端剪枝 20% 后的客户端高斯，展示了从五个视角观察的第 90 帧（左），以及实时会话流中第 70–110 帧期间从一个视角观察的画面（右）。

方法	设置	PSNR	SSIM	学习感知图像块相似度
PointSplat512	黑色，全图	33.786	0.9774	0.0208
未压缩 + NSC	黑色，全图	35.042	0.9809	0.0174
我们的轻量版 (32/4)	黑色，全图	34.459	0.9784	0.0207
我们的专业版 (48/6)	黑色，全图	34.837	0.9800	0.0185
我们 Ultra 版 (80/6)	黑色，全图	34.892	0.9803	0.0182

表 S1. **每个层级的全图峰值信噪比、结构相似性指数和学习感知图像块相似度均优于标准 PointSplat512**。 表 1的预测结果是在完整黑色背景图像上评分的；流式层级使用超网络和 NSC。

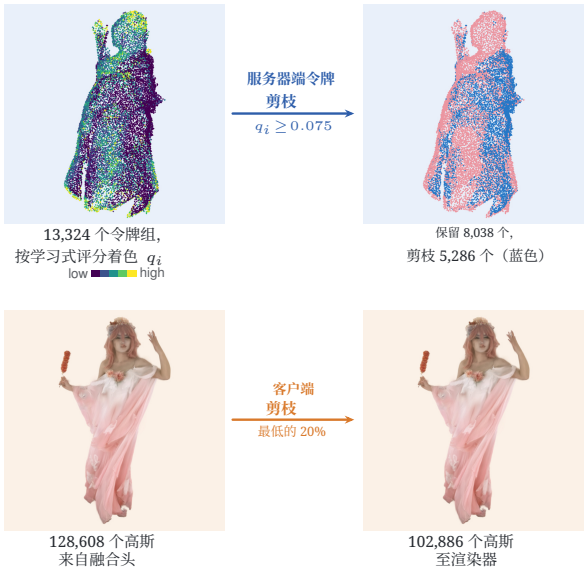

图 S2. **服务器端令牌剪枝和客户端剪枝作用于互补阶段**。 服务器端令牌剪枝会在投影前丢弃低不透明度的令牌组，以减小数据包大小（上方；平均峰值信噪比变化为 +0.012 dB）；客户端剪枝则在头部之后、传递给渲染器之前，丢弃生成的高斯中不透明度最低的 20%（下方； +0.002dB）。DNA-Rendering 0678_02，第 90 帧。

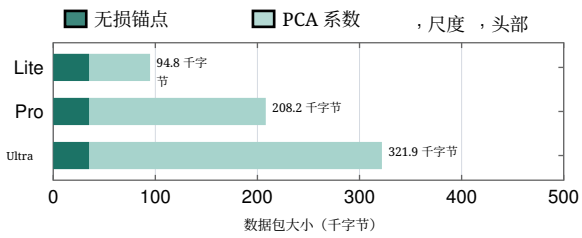

图 S3。**所有层级在无损锚点上均使用相同的 35.0 千字节，其差异在于 PCA 系数**。16 个 DNA-Rendering 测试场景的平均数据包大小。float32 PCA 状态在每个会话中仅发送一次（轻量版为 136 千字节，Pro 为 201 千字节，Ultra 为 333 千字节）；标准 PointSplat512 的 float32 高斯每帧占用 4753 千字节。

26.46 降至 22.55 dB（对任意其他单一属性类型进行有损编码时为 25.69–26.41 dB），因此，属性敏感性是结构性问题，而非低位宽造成的假象。

S4. 分辨率和时间行为

表S9 仅将重建输入设为 224、256、384 和 512 之一，从而改变点和令牌的数量；渲染和评分则保持在 512×512。时间研究对三个序列的第 30–149 帧进行重放，使用八个输入摄像机和四个留出相机。每个层级都基于第零帧的完整令牌集拟合其主成分分析状态，并将其固定；此外，还设置了一个不进行令牌剪枝的 Pro 变体，以单独考察其影响。

方法	平均 α	α 份额	MAEonblack
PointSplat512	0.000279	0.002894	0.000102
我们的未压缩版本	0.000279	0.002898	0.000102
我们的轻量版 (32/4)	0.000431	0.004564	0.000120
我们的专业版 (48/6)	0.000372	0.003916	0.000104
我们的 Ultra 版 (80/6)	0.000368	0.003871	0.000103

表 S2. **所有层级的掩码外颜色误差均不超过** 1.2×10^{-4}。真实值掩码外的平均阿尔法、掩码外预测 alpha 质量占比，以及黑色背景上的掩码外平均绝对误差；流式层级使用超网络和 NSC。

场景	PointSplat512	未压缩	Pro	Δ PSNR	Pro kB
0012_09	21.566	21.569	21.810	+0.244	212.3
0022_10	27.707	27.707	29.392	+1.685	239.0
0166_04	22.579	22.580	23.078	+0.498	255.1
0173_02	22.426	22.426	22.505	+0.079	206.2
0219_07	26.030	26.030	26.688	+0.659	198.1
0235_11	24.632	24.632	25.592	+0.960	252.3
0326_07	25.040	25.040	25.225	+0.185	189.2
0345_08	26.686	26.686	27.780	+1.094	209.7
0379_09	29.088	29.089	30.393	+1.305	151.1
0410_05	28.695	28.695	29.867	+1.173	193.4
0438_02	28.309	28.309	29.834	+1.525	187.1
0536_06	29.351	29.351	31.660	+2.308	162.4
0647_02	28.149	28.149	30.010	+1.861	187.8
0657_05	28.568	28.568	30.862	+2.293	168.0
0718_06	21.696	21.696	21.925	+0.230	249.9
0811_06	28.177	28.177	28.879	+0.702	269.2

表 S3. **Pro 在全部 16 个测试场景中的峰值信噪比均优于标准 PointSplat512**。各场景的峰值信噪比；Pro 使用超网络和 NSC，且 Δ 为 Pro 与标准 PointSplat512 之差。

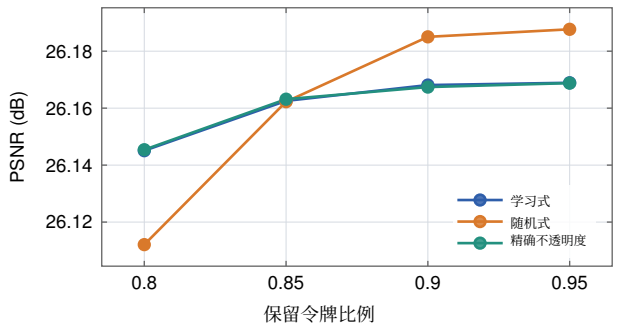

图 S4. **在已部署系统采用的 76.4% 保留率下，学习式令牌选择比随机选择高出 0.201 dB（表 S7），且在所有绘制的比例下，与精确不透明度排序的差距均不超过 0.001 dB**。未进行客户端剪枝的未压缩输出：每帧保留固定比例的令牌组，而非使用阈值；表 S7 比较了三种选择方式在已部署阈值对应令牌数量下的表现，随后进行 20% 客户端剪枝。

在最后 30 帧中，采用超网络的 Pro 的全图峰值信噪比与未压缩版本相差不超过 0.26 dB

重建（图S6 和表S10）。

长序列与会话状态刷新。已部署系统在整个会话期间保留其零帧状态。我们将五个 DNA-Rendering 测试序列延长至第 224 帧，并将 HiFi4G 和 ActorsHQ 序列延长至第 149 帧，同时采用主协议中的转换步数。冻结零帧主成分分析基和超网络更新（其在第 0–29 帧上的一次性微调，见第 S7节）后，Pro 在 10 个序列上的平均结果显示，相较于未压缩重建，从最初和最后 30 个评估帧之间变化了 +0.04 dB；各序列的变化趋势介于每 100 帧 −0.11 至 +0.27 dB 之间（图 S7）。NSC 在第 0–29 帧上拟合一次后不再更新，直到未见过的设备序列结束时，相较于仅使用超网络仍保持增益；在 HiFi4G 和 ActorsHQ 上，最后 30 个评估帧分别增加了 +1.19 和 +1.69 dB。每 120 帧刷新一次基，每帧会增加 1.4 千字节，并使峰值信噪比变化 −0.07 dB；若在每次刷新时也重新拟合超网络更新，则净变化为 +0.00 dB（表 S11）。

帧间稳定性。与所有逐帧优化基线相比，部署的 Pro 能更紧密地跟随帧间运动。我们在八个留出相机上回放 0012 09、0022 10 和 0166 04 的第 40–70 帧（每个相机 30 对相邻帧，每种变体共 720 对）。在相邻真实值掩码并集范围内的带符号帧间 RGB 差异上，部署的 Pro 达到 24.27 dB 的时间差分峰值信噪比（普通 Pro 为 23.98 dB，未压缩令牌为 24.48 dB），分别比 3DGStream（21.71 dB）、QUEEN-l（20.59 dB）和 HiCoM（19.67 dB）高 2.55、3.68 和 4.59 dB。在静态像素上（两帧的 α>0.5及真实值 |ΔRGB|<2/255；真实值下限 0.855/255），部署的 Pro 平均 3.44/255 帧间绝对变化，将静态闪烁相较 HiCoM（6.44/255）降低 46.5%，相较 3DGStream（8.07/255）降低 57.4%。

S5. 对比细节

使用相同的八个输入相机，在全部 16 个序列上，每个层级的峰值信噪比都高于各逐帧优化方法，学习感知图像块相似度更低，字节数更少；Pro 和 Ultra 在每个序列上的峰值信噪比、结构相似性指数均优于离线 4C4D，

方法	16 个场景 峰值信噪比↑	结构相似性指数↑	学习感知图像块相似度↓	kB ↓	高斯数量（千）↓
PointSplat512	26.169	0.8650	0.1034	4753.1*	84.9
我们的未压缩版本	26.169	0.8650	0.1034	5941.4*	106.1
我们的轻量版（32/4）	26.840	0.8700	0.1043	94.8	64.8
我们的专业版（48/6）	27.219	0.8802	0.0941	208.2	64.8
我们的 Ultra 版（80/6）	27.273	0.8818	0.0923	321.9	64.8

表 S4。**我们的各层级在峰值信噪比和结构相似性指数方面均超越 PointSplat512，而高斯字节数仅为其一小部分。** DNA-Rendering，16 个测试场景；所有层级均使用超网络和 NSC（表。 S41 给出了普通解码器）。 *每个高斯的 14 个属性采用 32 位浮点数存储。

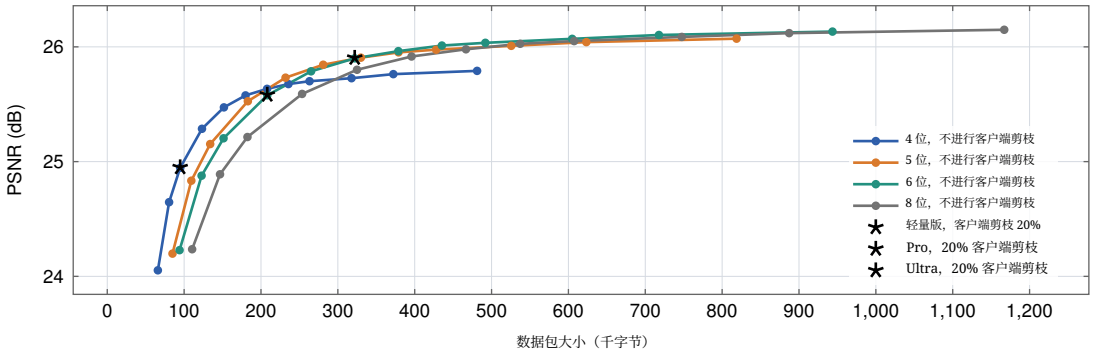

图 S5。**六位量化几乎可以媲美 8 位，而在字节数不变的情况下，客户端剪枝会使前沿向上移动。**48 个不使用客户端剪枝的主成分分析设置；星号表示采用 20% 客户端剪枝的已部署层级。所有点均使用普通解码器。

编解码器	剪枝	PSNR	千字节令牌	高斯 (k)
未压缩	无	26.169	5941.4	6631	106.1
未压缩	客户端	26.169	4753.1	6631	84.9
未压缩	令牌	26.181	4538.5	5065	81.0
未压缩	两者	26.183	3630.7	5065	64.8
主成分分析 80/6	客户端	25.894	420.7	6631	84.9
主成分分析 80/6	令牌	25.900	321.9	5065	81.0
主成分分析 80/6	两者	25.902	321.9	5065	64.8

表 S5。**令牌剪枝和客户端剪枝会降低不同的成本。**令牌剪枝会移除令牌组并缩小数据包；客户端剪枝则会移除生成的高斯中的 20%，而数据包大小保持不变。未压缩千字节数统计的是原始 32 位浮点数高斯；普通解码器。

学习感知图像块相似度和字节数也更优。以下段落介绍了正文中各项比较所采用的设置，并将比较扩展至全部序列。

现有格式。每种格式均采用其编码器的官方默认设置，或论文中推荐的设置：SPZ v4 [51] 使用其默认打包方式，PlayCanvas 压缩 PLY 写入器 [57]，SOG [58] 使用 libwebp 的默认无损设置，ksplat [32] 压缩级别 1，Google Draco [18] 编码器默认设置（位置 11 位，通用

方法	PSNR	SSIM	学习感知图像块相似度	kB
未压缩	26.169	0.8650	0.1034	4753.1
SPZ v4（SH0）	26.165	0.8639	0.1087	1089.3
乘积量化	25.712	0.8388	0.1463	356.9
轻量版，普通	24.950	0.8198	0.1809	94.8
轻量版，超网络	25.772	0.8491	0.1239	94.8
轻量版，超网络 +NSC	26.840	0.8700	0.1043	94.8
Pro，普通版	25.582	0.8419	0.1459	208.2
Pro，超网络	26.065	0.8598	0.1121	208.2
Pro，超网络 +NSC	27.219	0.8802	0.0941	208.2
Ultra，普通	25.904	0.8528	0.1256	321.9
Ultra，超网络	26.095	0.8614	0.1097	321.9
Ultra，超网络 +NSC	27.273	0.8818	0.0923	321.9

表 S6。**Pro 超网络在使用更少字节的情况下，三项指标均优于乘积量化（PQ），而 Ultra 使用 3。38× 比 SPZ 少的字节；加入 NSC 后，三个层级的表现均超过二者。** SPZ 在客户端剪枝 20% 后，无需令牌剪枝即可压缩我们的未压缩高斯元，而表 S12则在令牌剪枝后应用 SPZ；PQ 使用 64 个维度为 16 的子向量和 256 个质心，并采用 Pro 的令牌剪枝。普通行和超网络行解码相同的令牌。

属性 8 位，压缩级别 7）以及采用默认设置的 FCGS λ=10^{-4}。与表格中一样。2，每个高斯集都会应用我们各层级使用的 NSC，并在编码前将尺度限制为最小正规 32 位浮点数。表格 S14 列出了每种

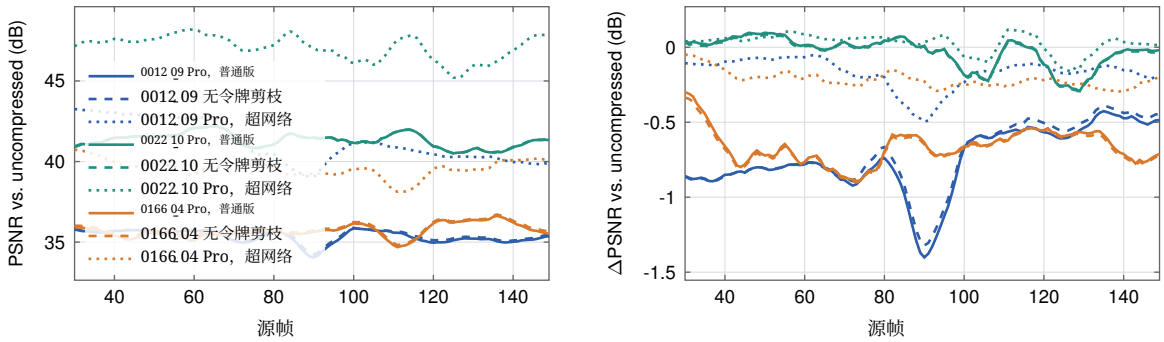

图 S6。**冻结的零帧主成分分析和超网络解码在 120 帧内保持质量**。在四个留出视图上，展示帧 30–149 之间的三个序列、全图峰值信噪比，以及流式渲染结果与未压缩渲染结果的对比（左），还有相对于真实值的未压缩重建的峰值信噪比差异（右）；对象为采用普通解码（实线）、无令牌剪枝（虚线）和超网络（点线）的 Pro。

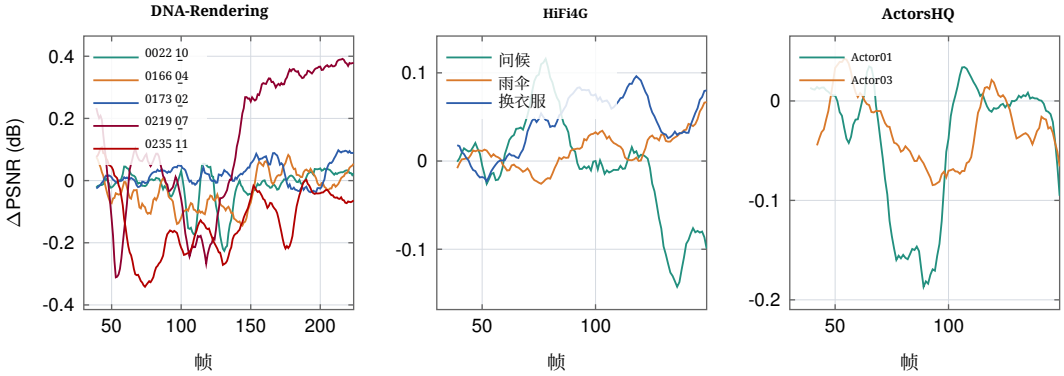

图 S7。**在第 120–195 帧期间，零帧会话状态未显示出系统性漂移**。Pro 经超网络解码后的时间 PSNR 相对于未压缩重建的变化（其零帧主成分分析基和超网络更新均被冻结），以每个序列最初评估的 30 帧为参照，并以 10 帧移动平均值显示。DNA-Rendering 序列扩展至第 224 帧，在四个留出相机上进行评估；HiFi4G 和 ActorsHQ 序列扩展至第 149 帧，在八个留出相机上进行评估。

组选择	峰值信噪比 PSNR 高斯元（千）	令牌	
仅客户端剪枝	26.169	84.9	6631
随机令牌组 25.982		64.8	5065
精确不透明度组	26.177	64.8	5065
学习式令牌组 26.183		64.8	5065

表 S7。**学习式组选择与精确不透明度排序效果相当，且明显优于随机选择**。令牌数量匹配，不进行特征压缩，随后进行 20% 客户端剪枝。精确不透明度排序需要运行完整头部。

格式的设置、学习感知图像块相似度和编码时间；图S9 绘制了其与字节数的关系，表格S15 标明了每种格式的默认设置可达到的各个层级。表格S12 重复了不含 NSCon 两侧数据的对比，图S8 则对每种格式在启用和未启用 NSCon 时的设置进行了遍历。

视频中的高斯元。 继 V^3 [75] 的打包以及二维高斯视频远程呈现流 [70], 之后，其中量化——

加上 H.265 总共使 Tele360 的代价约为 2 dB（其表 11 中为 32.65 降至 30.57 dB ），我们将未压缩高斯元量化为 17 个灰度平面，并使用 x265 [48]在五个恒定码率因子（CRF）下，按照 Morton 顺序或令牌网格进行编码（表 S13）。最佳结果在 185.69 兆比特/秒下达到 16.36 dB。8 位尺度无法容纳 8 个场景的动态范围；对于其他场景，最佳结果达到 23.65 dB，但仍低于每个层级，且字节数高出数倍。在 CPU 上解码这些平面，每帧需要 112.79–143.86 毫秒。

逐序列协议。 表 3涵盖全部 16 个测试序列的第 30–89 帧。表中的范围覆盖各序列；千字节/帧包含首帧或我们的会话状态，秒/帧表示在一台 RTX5090 上处理每个后续帧所需的时间。HiCoM、QUEEN-l 和 3DGStream 均分别基于我们的 8 台输入摄像机和 40 台摄像机进行拟合，共计 96 次运行，并采用各自的官方设置：HiCoM 的训练计划为首帧 15,000 步、每个后续帧 200 步；QUEEN-l 使用 MiDaS 深度初始化的配置；3DGStream 使用

令牌阈值	秩/位数	客户端剪枝	PSNR	SSIM	学习感知图像块相似度	kB	高斯数量（千）
None	64/6	20%	25.783	0.8492	0.1314	346.4	84.9
None	64/6	30%	25.786	0.8492	0.1315	346.4	74.3
None	64/6	none	25.784	0.8491	0.1314	346.4	106.1
None	80/4	20%	25.568	0.8429	0.1341	235.5	84.9
None	80/4	30%	25.573	0.8429	0.1343	235.5	74.3
None	80/4	none	25.569	0.8429	0.1341	235.5	106.1
None	80/6	20%	25.894	0.8532	0.1243	420.7	84.9
None	80/6	30%	25.897	0.8532	0.1245	420.7	74.3
None	80/6	none	25.894	0.8532	0.1243	420.7	106.1
None	96/6	20%	25.955	0.8557	0.1200	494.9	84.9
None	96/6	30%	25.959	0.8557	0.1202	494.9	74.3
None	96/6	none	25.956	0.8557	0.1200	494.9	106.1
None	128/6	20%	26.027	0.8584	0.1150	643.1	84.9
None	128/6	30%	26.031	0.8584	0.1152	643.1	74.3
None	128/6	none	26.028	0.8584	0.1150	643.1	106.1
None	160/6	20%	26.061	0.8600	0.1122	791.0	84.9
None	160/6	30%	26.064	0.8600	0.1124	791.0	74.3
None	160/6	none	26.061	0.8600	0.1122	791.0	106.1
0.075	64/6	20%	25.789	0.8487	0.1329	265.2	64.8
0.075	64/6	30%	25.790	0.8486	0.1331	265.2	56.7
0.075	64/6	none	25.787	0.8486	0.1327	265.2	81.0
0.075	80/4	20%	25.580	0.8426	0.1356	179.9	64.8
0.075	80/4	30%	25.586	0.8426	0.1359	179.9	56.7
0.075	80/4	none	25.577	0.8425	0.1355	179.9	81.0
0.075	80/6	20%	25.902	0.8528	0.1258	321.9	64.8
0.075	80/6	30%	25.905	0.8527	0.1260	321.9	56.7
0.075	80/6	none	25.900	0.8527	0.1257	321.9	81.0
0.075	96/6	20%	25.964	0.8553	0.1214	378.7	64.8
0.075	96/6	30%	25.968	0.8552	0.1218	378.7	56.7
0.075	96/6	none	25.963	0.8553	0.1213	378.7	81.0
0.075	128/6	20%	26.037	0.8581	0.1164	491.9	64.8
0.075	128/6	30%	26.042	0.8580	0.1167	491.9	56.7
0.075	128/6	none	26.035	0.8580	0.1162	491.9	81.0
0.075	160/6	20%	26.072	0.8597	0.1136	604.9	64.8
0.075	160/6	30%	26.077	0.8597	0.1139	604.9	56.7
0.075	160/6	none	26.070	0.8596	0.1134	604.9	81.0
0.075	32/4	20%，超网络	25.772	0.8491	0.1239	94.8	64.8
0.075	32/4	20%，超网络 +NSC	26.840	0.8700	0.1043	94.8	64.8
0.075	48/6	20%，超网络	26.065	0.8598	0.1121	208.2	64.8
0.075	48/6	20%，超网络 +NSC	27.219	0.8802	0.0941	208.2	64.8
0.075	80/6	20%，超网络	26.095	0.8614	0.1097	321.9	64.8
0.075	80/6	20%，超网络 +NSC	27.273	0.8818	0.0923	321.9	64.8

表 S8. **在峰值信噪比相同的情况下，令牌剪枝可节省 25.1% 的字节**。 在 16 个测试场景上使用的令牌剪枝阈值、编解码器和客户端剪枝规则；此外，还展示了 Lite、Pro 和 Ultra 在客户端剪枝率为 20% 时的超网络和 NSC。节省量以 Bjøntegaard 增量码率（BD 码率）表示 [1]，统计范围为六种编解码器、客户端剪枝率为 20% 且采用普通解码器的情况。

SH 一阶，并在 30,000 步的首帧拟合后，每帧执行 150 + 100 步更新。计时结果来自在第 0–59 帧上使用相同设置、序列和输入进行的运行。RGB 图像缩至 512，并使用与我们相同的软掩码合成到黑色背景上；姿态则转换为 OpenCV 格式，内参减半。三种方法均从利用已知标定对第 30 帧进行三角测量得到的 COLMAP点开始。对于全部 16 个八摄像机

摄像机，以及 16 种四摄像机设置中的 4 种；我们在输入掩码的视觉外壳内绘制 100,000 个点（表 S17）。字节数对应可解码码流，其中首帧摊销到 60 帧：QUEEN 的压缩数据包、3DGStream 的初始点及逐帧缓存和新增高斯，以及 HiCoM 层级结构和新增高斯的 32 位浮点数序列化，因为 HiCoM 的代码没有比特流写入器。我们的各行采用了从未见过

输入（像素）配置		PSNR	SSIM	学习感知图像块相似度	kB	高斯数量（千）
224	未压缩	24.025	0.7831	0.1882	5622.4*	100.4
224	令牌 +客户端剪枝	24.013	0.7824	0.1896	3210.0*	57.3
224	Pro，普通版	23.456	0.7668	0.2315	187.9	57.3
224	Pro，超网络	23.781	0.7747	0.2086	187.9	57.3
224	Pro，超网络 +NSC	24.083	0.7878	0.1921	187.9	57.3
256	未压缩	24.627	0.8040	0.1642	5720.7*	102.2
256	令牌 +客户端剪枝	24.617	0.8033	0.1655	3298.2*	58.9
256	Pro，普通版	24.021	0.7848	0.2112	193.7	58.9
256	Pro，超网络	24.416	0.7955	0.1839	193.7	58.9
256	Pro，超网络 +NSC	24.880	0.8115	0.1663	193.7	58.9
384	未压缩	25.961	0.8527	0.1170	5920.0*	105.7
384	令牌 +客户端剪枝	25.969	0.8523	0.1186	3514.6*	62.8
384	Pro，普通版	25.236	0.8276	0.1654	204.0	62.8
384	Pro，超网络	25.798	0.8453	0.1302	204.0	62.8
384	Pro，超网络 +NSC	26.696	0.8652	0.1107	204.0	62.8
512	未压缩	26.169	0.8650	0.1034	5941.4*	106.1
512	令牌 +客户端剪枝	26.183	0.8649	0.1048	3630.7*	64.8
512	Pro，普通版	25.582	0.8419	0.1459	208.2	64.8
512	Pro，超网络	26.065	0.8598	0.1121	208.2	64.8
512	Pro，超网络 +NSC	27.219	0.8802	0.0941	208.2	64.8

表 S9。**输入分辨率提升至 512 时，质量单调上升，而增加的字节数极少**。仅重建输入尺寸发生变化；渲染和评分仍保持在 512。Pro 在每种输入尺寸下都使用第 0 帧拟合其主成分分析，并复用 512 输入会话的超网络更新和 NSC。 *原始 float32 高斯。

场景	配置	PSNR	SSIM	学习感知图像块相似度	kB	Δ，最后 30 帧 斜率（dB/100 帧）	
0012_09	未压缩	21.779	0.8099	0.1053	5236.8*	–	–
0012_09	轻量版，普通	20.176	0.7426	0.1806	93.6	−1.246	+0.584
0012_09	Pro，普通版	21.030	0.7757	0.1404	207.7	−0.516	+0.380
0012_09	Ultra，普通	21.340	0.7901	0.1248	321.6	−0.334	+0.114
0012_09	Pro，普通版，无令牌剪枝	21.058	0.7763	0.1399	239.6	−0.472	+0.424
0012_09	Pro，超网络	21.608	0.8030	0.1145	207.7	−0.164	−0.073
0022_10	未压缩	27.471	0.8974	0.1145	7148.7*	–	–
0022_10	轻量版，普通	27.216	0.8789	0.1846	112.1	−0.395	−0.176
0022_10	Pro，普通版	27.441	0.8898	0.1593	246.2	−0.128	−0.176
0022_10	Ultra，普通版	27.602	0.8955	0.1401	380.1	+0.054	−0.194
0022_10	Pro，普通版，不进行令牌剪枝	27.435	0.8898	0.1586	327.6	−0.134	−0.172
0022_10	Pro，超网络	27.500	0.8967	0.1249	246.2	−0.033	−0.079
0166_04	未压缩	24.449	0.8376	0.0911	6741.8*	–	–
0166_04	轻量版，普通	23.096	0.7782	0.1568	117.2	−1.364	+0.039
0166_04	Pro，普通版	23.787	0.8074	0.1256	258.3	−0.641	+0.010
0166_04	Ultra，普通版	24.134	0.8208	0.1103	400.0	−0.295	−0.017
0166_04	Pro，普通版，不进行令牌剪枝	23.776	0.8074	0.1248	307.6	−0.653	+0.018
0166_04	Pro，超网络	24.234	0.8258	0.1021	258.3	−0.251	−0.102

表 S10。**零帧基底可持续 120 帧**。普通解码器和超网络在第 30–149 帧的真实值掩码边界框内的质量；最后两列分别给出第 120–149 帧全图峰值信噪比相对于未压缩重建的差值，以及该差值在第 30–149 帧内每 100 帧的斜率。 *原始高斯存储。

测试场景的超网络，并采用超网络更新和 NSC，两者均在第 0–29帧上拟合。

我们修复了基线方法公开代码中的错误，但没有更改其优化过程。QUEEN 附带的光栅化器会剔除所有距离摄像机不足 4 米的点，这会移除

在 DNA-Rendering 距离下，身体的大部分区域都受到影响；我们恢复了原始的 3DGS 规则（0.2 m），并重新训练了使用相同光栅化器的 HiCoM。我们还修复了两个绑定错误，以及一个空后缀切片（[-0:]）：在不添加任何点的帧之后，它会选中所有内容，而不是不选中任何内容。

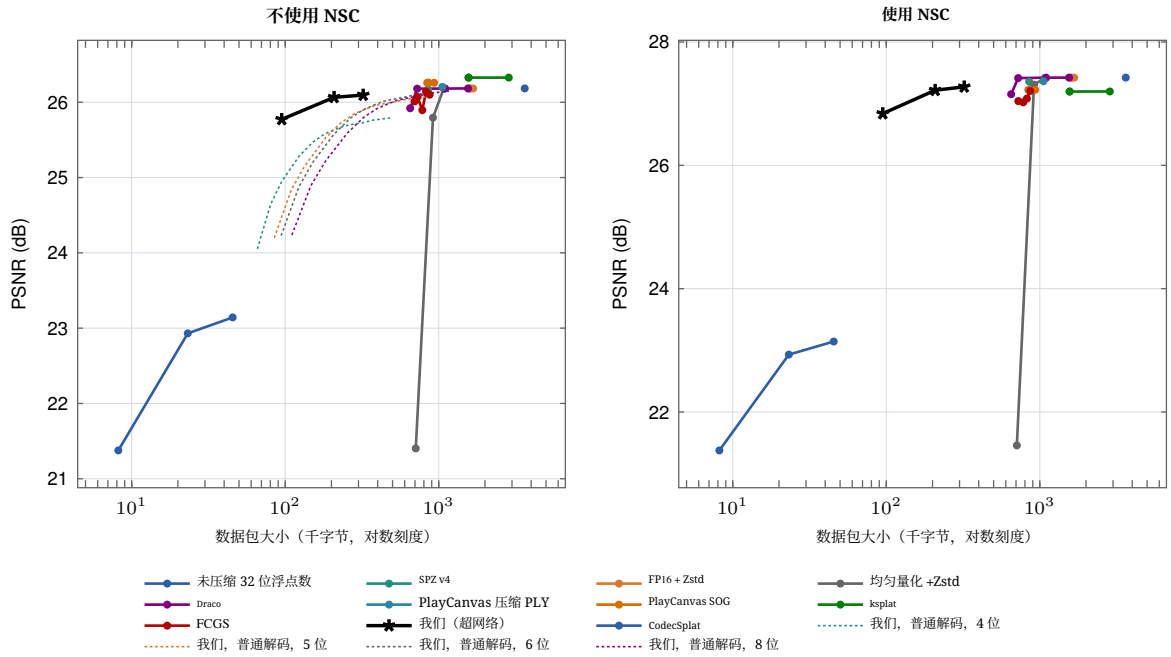

图 S8。**在峰值信噪比相同的情况下，与均匀量化、Draco 和采用 NSC 的 FCGS 相比，我们的层级平均所需字节数减少 67–85%（不采用 NSC 时减少 72–86%）**。DNA-Rendering，16 个测试场景，48 个评估相机。格式曲线展示了各编码器在我们的未压缩剪枝高斯上的不同设置；虚线曲线展示了我们的编解码器在 4、5、6 和 8 位下采用普通解码且不进行客户端剪枝时的效果，如图S5所示。右侧的高斯格式与我们的层级采用相同的 NSC，而 CodecSplat 则像左侧一样对其自身特征进行编码。与均匀量化、Draco 和采用 NSC 的 FCGS 相比，我们的层级分别节省 85%、67% 和 82% 的 BD 码率；不采用 NSC 时，分别节省 86%、75% 和 72%，比较范围为各格式与我们的层级共有的峰值信噪比区间。其他格式的帕累托最优点少于三个，不足以计算差值；CodecSplat 的峰值信噪比始终低于我们的范围。

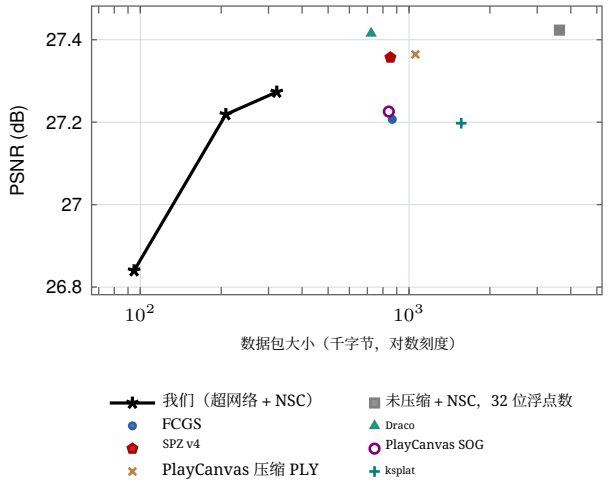

图 S9. **在六种高斯压缩格式中，Ultra 的表现仅比最佳格式低 0.14 dB，且至少少用 2 个。24× 字节。** 表 S14 的码率–峰值信噪比S14 以对数刻度表示字节数。

结果。 在相同的八个输入相机下，每个层级在全部 16 个序列上的峰值信噪比均高于各方法，学习感知图像块相似度均更低，且字节数均更少。轻量版优于 HiCoM 低 5.31–10.14 dB，3DGStream 低 0.62–4.79 dB，QUEEN-l 低 7.35–12.62 dB。使用全部 40 个摄像机时，QUEEN-l 和 HiCoM 在每个序列上的表现仍低于 Lite，分别低 1.96–8.42 dB 和 3.81–8.45 dB；而使用我们五倍数量摄像机的 3DGStream，平均仍比 Lite 低 0.37 dB（表 3），但消耗的字节数为其 17.6–31 倍。8× Ultra 的字节数。在全部 96 次运行中，每个层级在相同序列上的 LPIPS 均更低，平均低 0.211。每个基线流都大于各层级的流，中位数分别为 62× Lite 大小的 30× Pro 大小的 19× Ultra 大小。表 S25列出了 0012 09、0022 10 和 0166 04；表 S18至 S21 列出了其他 13 个序列。在这 39 次使用 8 个摄像机的运行中，每个层级在峰值信噪比、LPIPS 和字节数方面都更优，Lite 的峰值信噪比领先 0.62 至 12.56 dB。

离线稀疏视图基线。4C4D [97], 类似的 Sparse-Cam4D [55], 会离线拟合一个 4D 高斯模型来表示整个捕获序列。我们在同样的八个输入相机拍摄的全部 60 帧上运行该方法，采用其官方 DyNeRF [38] 配置，训练 30,000 步；初始化点云由其代码库通过 MAtCha 代码库 [20],中的基于 MASt3R 的运动恢复结构 [35]，根据输入视图重建得到。然后，我们像处理流式运行一样渲染并评估最终检查点。其每帧千字节数按其 32 位浮点数参数计算

刷新策略	Δ峰值信噪比（dB）	Δ字节/帧
(a) PCA基		
不刷新	+0.0000	0.0
每60帧	−0.1750	3,198.0
每120帧	−0.0740	1,359.4
方差 < 0.90	−0.1282	4,160.8
方差 < 0.95	−0.1853	23,603.0
方差 < 0.98	−0.1830	84,009.9
(b) PCA基 + 超网络更新		
不刷新（PCA）	+0.0000	0.0
每120帧（PCA）	−0.0571	1,413.0
每60帧 + 更新	+0.0002	12,869.8
每120帧 + 更新	+0.0038	5,551.6
方差 < 0.90 + 更新	+0.0043	8,785.9
方差 < 0.95 + 更新	−0.0011	63,605.9
方差 < 0.98 + 更新	+0.0136	278,359.6

表 S11。**零帧会话状态无需刷新**。部署的 Pro 相对于保留零帧状态的峰值信噪比变化（各序列权重相同），以及刷新所增加的每帧净字节数，其中包括数据包大小的变化。(a) 每 60 帧或 120 帧重新拟合 PCA 基，或者在当前帧的解释方差低于零帧解释方差的指定比例时重新拟合；在帧 t 检测到的刷新从 $t{+}1$起生效。(b) 此外，在每次刷新时，使用最近 30 帧重新拟合超网络更新。每个区块对包含其所有行的序列取平均值。

格式	千字节 ↓	峰值信噪比↑	结构相似性指数 ↑	解码（毫秒）↓
PointSplat512，32 位浮点数	4753.1	26.169	0.8650	–
未压缩，32 位浮点数	3630.7	26.183	0.8649	–
SPZ v4	852.2	26.172	0.8638	27.9
PlayCanvas 压缩 PLY 1056.3		26.204	0.8642	209.4
FCGS [7]	872.2	26.100	0.8626	4562.7
ksplat	1564.2	26.327	0.8648	154.6
Draco	723.9	26.180	0.8647	15.3
PlayCanvas SOG	840.9	26.258	0.8638	221.2
CodecSplat [87]	45.4	23.143	0.7434	199.1
我们的轻量版	94.8	25.772	0.8491	21.7
我们的专业版	208.2	26.065	0.8598	22.1
我们的 Ultra 版	321.9	26.095	0.8614	22.8

表 S12。**双方均不使用 NSC 时，高斯压缩格式至少需要 3。47× 在峰值信噪比相近时，Pro 的字节数**。表 2 对未校准高斯和仅由超网络解码的我们的层级进行了比较；NSC 不会改变我们的字节数和客户端耗时。

在 60 帧上使用 0.4–3.6 M 个高斯，其秒/帧将整个拟合过程（无需初始化）的耗时除以 60。Pro 和 Ultra 在每个序列的峰值信噪比、结构相似性指数、学习感知图像块相似度和字节数方面均优于它（表 S16），Ultra 平均高出 1.70 dB。Lite 在全部 16 个序列上的学习感知图像块相似度均优于它，平均峰值信噪比高出 1.27 dB。图 S10 展示了图 6 所示视角下的每种八视图方法。

解码。我们的各行在每个 DNA 序列和层级上都使用超网络和 NSC。表S22至 S24 将其与表 1 中 48 个评估相机上对相同数据包进行的普通解码进行比较；在全部 16 个序列中，超网络和 NSC 使 Lite 的性能提升了 0.696 至 3.273 dB。

速度。表 S26 在空闲的 RTX 5090 上，以 0012 09 的官方设置和两种摄像机数量对三种方法进行计时。计时不包括首帧拟合、序列化和评估，并假设它们可实现理想的线性缩放 $G/t_{\min}$，尽管它们的每一帧都依赖于前一帧。对于其专为 40 个视角设计的设置，3DGStream 的表现最接近我们，而我们的平均吞吐量超过它 17。0× 在图形处理器数量相同的情况下。表 S27 将相同的缩放应用于全部 96 次网格运行，这些运行记录于共享 GPU 上；其中最快的方法在使用一块和四块 GPU 时分别达到 0.81 和 3.25 帧每秒，而我们的平均值为 8.0 和 32.2 帧每秒。表 S28 对前馈重建后接入各现有格式的流程进行建模；此类流程仍会发送表 S14中的每帧字节数。

报告的运行点与特征编码。表 S29 列出了流式体积视频方法在其论文重点介绍的配置下的表现，表 S30列出了数字人方法。表S31 展示了在空闲的 RTX 5090 上，用我们的输入对 CodecSplat 和 HiReFF 进行计时的结果。CodecSplat 使用其发布的 DL3DV 八视角模型，并采用官方测试配置处理与我们的重建器接收的相同八个输入视角和软掩码，评分方式也与我们的结果完全一致。在 512×512 和256×256两种输入中，未遮罩的 256 在其最高码率下得分最高，因此报告该结果。我们丢弃输入阿尔法范围之外像素对应的高斯，否则其背景高斯会遮挡表演者。表 S32将其最高码率与我们的层级进行比较。

其他重建器。头部前编码在另外两种像素对齐前馈重建器上也表现更佳，我们在这些重建器上重复了第 4.3 节的分割点分析。在 CodecSplat 基于 DepthSplat [80] 发布的模型上，针对 16 个测试场景（表 S33），在共享接收器之后进行分割，能够逐位复现原始输出。采用主成分分析编码的 64 维头部前特征在 310.1 千字节下达到 15.64 dB，优于对 38 个头部后高斯属性的所有编码方式（最佳为 15.39 dB，耗用 9.6× 字节；绕过其联合训练的熵编解码器后为 15.28 dB，而使用该编解码器时为 23.14 dB）。对于发布的第二阶段 GPS-Gaussian [93] 模型，我们比较了其三个参数头之前的 32 维共享特征 F与 8 维预激活旋转、尺度和不透明度属性 A；两者使用相同的无损 XYZ 和 RGB 码流，并在 25.07 dB 原生全图峰值信噪比下实现逐位精确的直通。原始属性和标准化属性在秩为 1–8 时保留了 99.27–100% 的方差，而 F 为 98.58–99.90%；然而，在共享对数码率范围内，头部前编码的表现分别比原始属性编码和标准化属性编码高出 1.34 dB 和 0.65 dB

布局	标定	CRF	千字节/帧	兆比特/秒	PSNR	SSIM	学习感知图像块相似度	CPU毫秒/帧
L1	无 NSC	18	531.9	127.66	14.17	0.4681	0.3959	131.02
L1	无 NSC	23	439.0	105.36	13.77	0.4521	0.4118	126.50
L1	无 NSC	28	405.8	97.40	12.21	0.4238	0.4199	121.12
L1	无 NSC	33	371.0	89.04	10.56	0.3906	0.4334	120.18
L1	无 NSC	38	337.7	81.05	9.32	0.3615	0.4524	112.79
L1	NSC	18	530.9	127.42	15.51	0.4895	0.3808	–
L1	NSC	23	437.9	105.10	15.00	0.4724	0.3962	–
L1	NSC	28	404.8	97.14	13.25	0.4406	0.4065	–
L1	NSC	33	370.0	88.80	11.28	0.4042	0.4240	–
L1	NSC	38	337.0	80.88	9.86	0.3748	0.4441	–
L2	无 NSC	18	775.8	186.18	15.15	0.4790	0.3892	143.86
L2	无 NSC	23	660.1	158.43	14.82	0.4642	0.4042	137.86
L2	无 NSC	28	619.1	148.59	13.10	0.4343	0.4128	134.24
L2	无 NSC	33	576.8	138.44	11.38	0.4011	0.4264	129.41
L2	无 NSC	38	537.8	129.08	10.71	0.3849	0.4356	121.59
L2	NSC	18	773.7	185.69	16.36	0.5004	0.3739	–
L2	NSC	23	658.1	157.94	15.96	0.4832	0.3913	–
L2	NSC	28	617.1	148.11	14.25	0.4504	0.4025	–
L2	NSC	33	575.1	138.02	12.39	0.4153	0.4182	–
L2	NSC	38	536.5	128.75	11.59	0.3981	0.4278	–

表 S13。 **将高斯元打包成视频会以数倍于我们字节数的代价损失质量**。 V^3-样式打包我们未压缩的高斯元，覆盖 16 个测试场景中 DNA-Rendering 的第 30–89 帧，在第 40、50、.....、80 帧进行评分，使用八个留出相机；位置采用 16 位编码，高字节无损，其他属性采用 8 位编码，并放入 17 个灰度平面，以 x265 编码，图像组（GOP）长度为 30，按 Morton 顺序（L1）或在令牌网格上（L2）排列，分别不使用和使用 NSC。字节数涵盖所有码流和全部辅助信息。均值涵盖全部 16 个场景，包括尺度超出 8 位范围的 8 个场景（第 S5 节）；CPU 时间测量使用 ffmpeg 解码所有平面所需的时间，测试涵盖三个场景，并计入进程启动时间。

格式	设置	kB	PSNR	SSIM	学习感知图像块相似度	编码（秒）	解码（秒）
我们的轻量版	–	94.8	26.840	0.8700	0.1043	–	–
我们的专业版	–	208.2	27.219	0.8802	0.0941	–	–
我们的 Ultra 版	–	321.9	27.273	0.8818	0.0923	–	–
未压缩 +NSC，32 位浮点数	–	3630.7	27.423	0.8857	0.0884	–	–
SPZ v4	–	852.7	27.357	0.8844	0.0914	0.06	0.03
PlayCanvas 压缩 PLY	–	1056.3	27.364	0.8846	0.0907	0.29	0.21
FCGS [7]	λ=10^{-4}	866.5	27.207	0.8777	0.0991	4.20	4.56
ksplat	第 1 级	1564.2	27.197	0.8824	0.0925	0.28	0.15
Draco	位置 11 / 属性 8 位	723.0	27.415	0.8854	0.0889	0.07	0.02
PlayCanvas SOG	默认	841.2	27.226	0.8827	0.0925	1.88	0.22

表 S14。**在相同 NSC 下，最佳格式比 Ultra 高出 0.14 dB，且数据量达到 2。24× 字节或更多。**每种格式都对 16 个 DNA-Rendering 测试场景的最终未压缩高斯元进行编码，并采用 NSC 的官方默认设置或推荐设置（第 S5 节）；我们的层级使用超网络和 NSC。编码和解码栏给出了在不使用 NSC 的高斯元上，通过参考实现测得的每帧耗时（秒）；除 FCGS 外，均在 CPU 上运行。命令行版 PlayCanvas、ksplat 和 FCGS 工具的耗时包含进程启动和文件 I/O。

的平均值。它在 (r，b)=(1，2) 处以 0.91× 字节达到 25.21 dB，与 (8，8) 处最佳属性设置的字节数相同（25.06dB）。

未见过的设备。表S34 列出了普通解码、超网络、超网络加 NSC，以及普通解码加 NSC 下的所有层级；与标准 PointSplat512 相比，每个已部署层级都能在数据包更小的同时提升峰值信噪比和结构相似性指数。表——

bles S35 和S36 将全部 24 种格式设置应用于未见过的设备。每个已部署层级的峰值信噪比和结构相似性指数都优于应用于未校准高斯的每种格式设置，也优于未压缩输入。由于整个会话状态分布在评估帧中，即使是 Ultra 也小 2。02× 和 2。01× 比 HiFi4G 和 ActorsHQ 上最小的格式设置还小。

格式	Reach Pro 画质（208.2 千字节） PSNR	SSIM	学习感知图像块相似度	Reach Ultra 画质（321.9 千字节） PSNR	SSIM	学习感知图像块相似度
SPZ v4	852.7	852.7	852.7	852.7	852.7	852.7
PlayCanvas 压缩 PLY	1056.3	1056.3	1056.3	1056.3	1056.3	1056.3
FCGS [7]	–	–	–	–	–	–
ksplat	–	1564.2	1564.2	–	1564.2	–
Draco	723.0	723.0	723.0	723.0	723.0	723.0
PlayCanvas SOG	841.2	841.2	841.2	–	841.2	–

表 S15。在相同的 NSC 下，每种达到某一层级的格式至少会传输 2。**24**× 其字节。每个单元格给出某种格式的默认值在仅按该指标达到相应层级时的千字节/帧，否则为“–”。

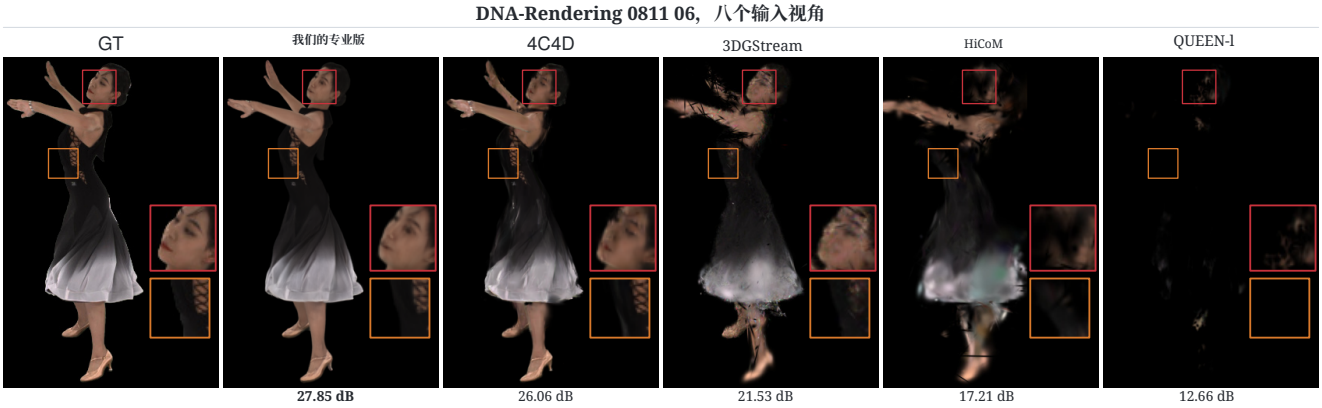

图 S10. **基于八个视角，QUEEN-l 几乎无法还原身体，而 HoloSplat 则保留了完整的几何结构和精细细节**。 图中所有八视图方法在留出视图上的结果。 6 并标注逐视图峰值信噪比；插图放大显示 2× 面部和系带。

序列	4C4D（离线） 峰值信噪比	结构相似性指数	学习感知图像块相似度	千字节/帧	我们的专业版 峰值信噪比	结构相似性指数	学习感知图像块相似度	千字节/帧
0012_09	19.21	0.712	0.222	15111	22.51	0.824	0.105	224
0022_10	27.06	0.879	0.167	12927	29.41	0.922	0.095	257
0166_04	22.18	0.760	0.177	25542	22.41	0.787	0.115	270
0173_02	20.47	0.728	0.211	13585	21.77	0.781	0.128	224
0219_07	23.81	0.823	0.174	7128	26.49	0.886	0.082	211
0235_11	22.03	0.793	0.187	11833	24.16	0.846	0.098	281
0326_07	23.04	0.765	0.252	5304	24.42	0.856	0.123	205
0345_08	25.42	0.874	0.144	10464	26.83	0.908	0.088	235
0379_09	27.24	0.822	0.200	3873	29.14	0.894	0.100	166
0410_05	25.60	0.804	0.228	16125	27.63	0.882	0.126	208
0438_02	26.50	0.829	0.175	15217	28.36	0.883	0.107	211
0536_06	28.47	0.872	0.164	10208	30.15	0.915	0.100	183
0647_02	28.49	0.906	0.146	11292	28.78	0.924	0.111	227
0657_05	27.61	0.857	0.191	12928	28.79	0.903	0.126	202
0718_06	20.80	0.749	0.175	39878	21.63	0.778	0.128	267
0811_06	26.96	0.900	0.124	20266	28.28	0.924	0.080	296
Mean	24.68	0.817	0.184	14480	26.30	0.870	0.107	229

表 S16。 **Pro 在每个序列的峰值信噪比、结构相似性指数、学习感知图像块相似度和字节数方面均优于离线 4C4D**。 八个输入视图以及表3中的协议。

合成 X-Humans。 我们评估了 X-Humans 的三个 150 帧测试序列 [60] (00035 test Take10, 00041 test Take8, 00088 test Take10)，渲染分辨率为 1024 ×

1024 （4×多重采样抗锯齿）来自三个仰角环上的 48 个固定摄像机（−12°、 6°、 24°；每个环上有 16 个方位角）。输入为位于 45° 方位角间距的 6° 环上的八个摄像机；在使用主协议（第S6节）对第 0–29 帧进行校准后，对其余 40 个摄像机在第 30–149 帧上的表现进行评估。姿态归一化遵循主协议，并以第一个摄像机的右轴作为其水平轴。

序列	8 个输入相机		40 个输入相机	
	COLMAP点	初始化	COLMAP点	初始化
0012_09	399	视觉外壳	6044	COLMAP
0022_10	63	视觉外壳	1218	COLMAP
0166_04	613	视觉外壳	4159	COLMAP
0173_02	183	视觉外壳	5313	COLMAP
0219_07	91	视觉外壳	3488	COLMAP
0235_11	135	视觉外壳	2197	COLMAP
0326_07	93	视觉外壳	1360	COLMAP
0345_08	35	视觉外壳	1640	COLMAP
0379_09	0	视觉外壳	384	视觉外壳
0410_05	33	视觉外壳	699	视觉外壳
0438_02	43	视觉外壳	1072	COLMAP
0536_06	25	视觉外壳	470	视觉外壳
0647_02	30	视觉外壳	948	视觉外壳
0657_05	253	视觉外壳	1060	COLMAP
0718_06	175	视觉外壳	5148	COLMAP
0811_06	100	视觉外壳	1588	COLMAP

表 S17. **每个序列的基线运行均使用经过校准的 COLMAP点，或使用 100,000 点的视觉外壳作为备用方案**。 每个序列分别在 8 台和 40 台输入摄像机的情况下进行初始化；当 COLMAP点少于 1000 个时，所有三种方法均从输入掩码视觉外壳中的 100,000 个随机点开始。

方法	相机序列	后续帧 kB	摊销 kB	PSNR	SSIM	学习感知图像块相似度	第一帧（秒）	秒/帧	序列化 (ms)	解码 (ms)	渲染 FPS
我们的轻量版	8 0173 02	96.051	108.953	21.449	0.7622	0.1404	–	–	–	–	–
我们的专业版	8 0173 02	209.782	223.779	21.771	0.7810	0.1282	–	–	–	–	–
我们的 Ultra 版	8 0173 02	324.539	340.722	21.835	0.7845	0.1250	–	–	–	–	–
我们的轻量版	8 0219 07	89.008	101.911	25.954	0.8701	0.0956	–	–	–	–	–
我们的专业版	8 0219 07	196.755	210.752	26.494	0.8864	0.0818	–	–	–	–	–
我们的 Ultra 版	8 0219 07	304.415	320.598	26.568	0.8888	0.0796	–	–	–	–	–
我们的轻量版	8 0235 11	119.386	132.288	23.924	0.8367	0.1083	–	–	–	–	–
我们的专业版	8 0235 11	266.591	280.587	24.159	0.8459	0.0980	–	–	–	–	–
我们的 Ultra 版	8 0235 11	412.975	429.158	24.362	0.8500	0.0939	–	–	–	–	–
我们的轻量版	8 0326 07	87.120	100.022	23.959	0.8429	0.1394	–	–	–	–	–
我们的专业版	8 0326 07	190.984	204.980	24.419	0.8564	0.1227	–	–	–	–	–
我们的 Ultra 版	8 0326 07	295.063	311.247	24.527	0.8595	0.1184	–	–	–	–	–
我们的轻量版	8 0345 08	100.487	113.389	26.397	0.8961	0.1000	–	–	–	–	–
我们的专业版	8 0345 08	220.564	234.561	26.832	0.9080	0.0881	–	–	–	–	–
我们的 Ultra 版	8 0345 08	340.767	356.950	27.021	0.9116	0.0839	–	–	–	–	–
我们的轻量版	8 0379 09	69.210	82.113	28.803	0.8844	0.1110	–	–	–	–	–
我们的专业版	8 0379 09	152.256	166.253	29.137	0.8936	0.0997	–	–	–	–	–
我们的 Ultra 版	8 0379 09	235.781	251.965	29.181	0.8959	0.0969	–	–	–	–	–
我们的轻量版	8 0410 05	87.917	100.819	27.278	0.8720	0.1407	–	–	–	–	–
我们的专业版	8 0410 05	193.878	207.875	27.626	0.8819	0.1265	–	–	–	–	–
我们的 Ultra 版	8 0410 05	299.393	315.577	27.679	0.8839	0.1238	–	–	–	–	–
我们的轻量版	8 0438 02	89.485	102.387	27.992	0.8709	0.1183	–	–	–	–	–
我们的专业版	8 0438 02	196.924	210.920	28.358	0.8834	0.1067	–	–	–	–	–
我们的 Ultra 版	8 0438 02	303.919	320.103	28.428	0.8866	0.1033	–	–	–	–	–
我们的轻量版	8 0536 06	77.619	90.521	29.947	0.9101	0.1109	–	–	–	–	–
我们的专业版	8 0536 06	169.456	183.452	30.151	0.9151	0.0998	–	–	–	–	–
我们的 Ultra 版	8 0536 06	261.757	277.941	30.227	0.9170	0.0962	–	–	–	–	–
我们的轻量版	8 0647 02	97.229	110.132	28.076	0.9162	0.1226	–	–	–	–	–
我们的专业版	8 0647 02	213.266	227.262	28.782	0.9238	0.1110	–	–	–	–	–
我们的 Ultra 版	8 0647 02	329.755	345.938	28.820	0.9249	0.1079	–	–	–	–	–

表 S18。 **在其余 13 个测试序列中，每个层级在峰值信噪比、学习感知图像块相似度和字节数方面均优于八视图逐帧优化**。 表 S25 列出了 0012 09、0022 10 和 0166 04。后续帧千字节数不包含每种方法的首帧，或我们的会话状态；摊销千字节数则取第 0 帧与 59 个后续帧的平均值。首帧（秒）和秒/帧分别表示首帧的拟合时间和后续帧的平均处理时间，数据来自共享 GPU 上对第 0–59 帧的运行，使用一块 RTX 5090；序列化表示每帧的序列化时间。

方法	相机序列	后续帧 kB	摊销 kB	PSNR	SSIM	学习感知图像块相似度	第一帧（秒）	秒/帧	序列化 (ms)	解码 (ms)	渲染 FPS
我们的轻量版	8 0657 05	85.851	98.753	28.385	0.8952	0.1382	–	–	–	–	–
我们的专业版	8 0657 05	187.788	201.785	28.792	0.9031	0.1255	–	–	–	–	–
我们的 Ultra 版	8 0657 05	289.721	305.905	28.932	0.9056	0.1213	–	–	–	–	–
我们的轻量版	8 0718 06	114.496	127.398	21.431	0.7648	0.1392	–	–	–	–	–
我们的专业版	8 0718 06	253.118	267.114	21.629	0.7778	0.1276	–	–	–	–	–
我们的 Ultra 版	8 0718 06	392.232	408.416	21.678	0.7811	0.1239	–	–	–	–	–
我们的轻量版	8 0811 06	125.744	138.646	27.963	0.9181	0.0897	–	–	–	–	–
我们的专业版	8 0811 06	282.199	296.195	28.280	0.9240	0.0802	–	–	–	–	–
我们的 Ultra 版	8 0811 06	438.361	454.545	28.328	0.9255	0.0777	–	–	–	–	–
HiCoM	8 0173 02	2687.218	2968.770	13.313	0.3011	0.5255	100.122	2.050	6.713	5.642	779.8
HiCoM	8 0219 07	2867.800	3173.828	17.354	0.5853	0.3534	100.082	1.964	6.481	5.441	878.0
HiCoM	8 0235 11	3352.303	3707.152	13.955	0.5019	0.4481	103.358	2.056	7.446	5.650	773.3
HiCoM	8 0326 07	2873.498	3181.108	16.863	0.5197	0.4096	104.805	1.984	6.989	5.482	802.4
HiCoM	8 0345 08	3603.752	3994.407	16.506	0.6446	0.3170	97.292	2.036	7.175	5.467	769.1
HiCoM	8 0379 09	2922.214	3243.183	22.247	0.6658	0.3152	103.069	2.110	6.954	5.743	703.6
HiCoM	8 0410 05	3029.779	3362.997	18.003	0.5044	0.4621	107.607	2.055	6.723	5.200	677.7
HiCoM	8 0438 02	2854.410	3160.745	18.434	0.5632	0.4133	156.574	2.771	16.202	5.758	776.6
HiCoM	8 0536 06	2883.696	3198.811	23.788	0.7069	0.3303	108.969	2.073	6.524	5.451	765.2
HiCoM	8 0647 02	2977.052	3301.044	18.858	0.7455	0.2960	103.125	2.016	6.861	5.531	778.3
HiCoM	8 0657 05	2963.212	3289.517	21.756	0.6828	0.3642	95.675	2.089	7.617	5.619	782.7
HiCoM	8 0718 06	2953.084	3260.782	16.121	0.4902	0.4333	86.496	1.959	6.966	5.373	804.8
HiCoM	8 0811 06	3545.691	3920.234	17.820	0.6657	0.3477	73.698	2.158	7.905	6.628	618.9
QUEEN-l	8 0173 02	54031.257	55316.626	11.601	0.3320	0.5296	49.681	1.284	6859.330	6106.814	420.7
QUEEN-l	8 0219 07	51697.802	52924.233	16.116	0.6412	0.3305	51.187	1.253	7193.362	6433.909	412.0
QUEEN-l	8 0235 11	48879.730	50037.696	12.431	0.5185	0.4384	54.972	1.277	6850.869	6143.515	390.1
QUEEN-l	8 0326 07	54497.476	55794.250	15.658	0.5287	0.4257	52.516	1.314	6890.727	6170.218	394.0
QUEEN-l	8 0345 08	50866.091	52073.056	14.108	0.6399	0.3529	52.392	1.301	7384.708	6647.156	384.5
QUEEN-l	8 0379 09	51038.252	52247.990	21.455	0.6586	0.3317	53.540	1.376	7526.054	6366.469	359.4
QUEEN-l	8 0410 05	51038.960	52246.742	17.335	0.5476	0.4246	54.389	1.290	7614.331	6431.493	394.6

表 S19。**表S18，续。**

方法	相机序列	后续帧 kB	摊销 kB	PSNR	SSIM	学习感知图像块相似度	首个帧（秒）	秒/帧	序列化 (ms)	解码 (ms)	渲染 FPS
QUEEN-l	8 0438 02	53817.035	55093.921	16.551	0.5103	0.4171	52.980	1.317	7736.324	6896.579	405.4
QUEEN-l	8 0536 06	54902.643	56209.136	22.478	0.6794	0.3413	50.013	1.351	8127.081	6875.982	398.3
QUEEN-l	8 0647 02	52490.131	53734.412	17.656	0.6727	0.3344	52.147	1.301	7610.390	6791.931	398.3
QUEEN-l	8 0657 05	56340.206	57680.325	18.966	0.5077	0.3890	53.331	1.318	8008.575	7065.417	415.6
QUEEN-l	8 0718 06	52408.231	53650.186	13.436	0.4668	0.4470	50.259	1.229	7381.049	6520.242	431.2
QUEEN-l	8 0811 06	52970.933	54224.639	15.399	0.6440	0.3415	51.753	1.283	5959.987	5369.567	379.5
3DGStream	8 0173 02	7995.881	8066.231	20.202	0.6901	0.2738	206.558	2.217	9.493	15.875	729.8
3DGStream	8 0219 07	7983.656	8049.417	23.488	0.7952	0.2260	217.192	2.259	8.590	15.879	649.3
3DGStream	8 0235 11	7990.582	8052.323	22.328	0.7701	0.2429	215.996	2.598	11.176	18.332	576.5
3DGStream	8 0326 07	7980.226	8039.286	23.000	0.7546	0.2914	202.970	2.415	8.624	40.734	723.3
3DGStream	8 0345 08	7982.142	8051.220	24.167	0.8391	0.2140	216.018	5.206	10.889	58.089	296.3
3DGStream	8 0379 09	7973.656	8024.221	27.223	0.8122	0.2372	171.835	2.322	8.405	37.788	685.9
3DGStream	8 0410 05	7989.588	8039.952	23.973	0.7652	0.3347	228.208	2.377	8.500	39.808	662.9
3DGStream	8 0438 02	7985.086	8029.767	25.574	0.7757	0.2989	187.813	2.348	8.404	41.496	695.1
3DGStream	8 0536 06	7980.374	8024.118	27.297	0.8527	0.2421	209.227	2.340	9.664	38.671	713.5
3DGStream	8 0647 02	7985.491	8029.215	27.458	0.8904	0.2108	198.333	2.270	9.412	40.506	747.6
3DGStream	8 0657 05	7982.456	8019.030	25.538	0.8265	0.2881	204.992	2.283	8.102	38.465	703.1
3DGStream	8 0718 06	8042.251	8132.599	18.915	0.6230	0.3379	199.549	2.419	12.143	41.078	767.4
3DGStream	8 0811 06	8005.638	8069.395	23.169	0.8209	0.2548	202.777	2.671	11.701	38.760	593.6
HiCoM	40 0173 02	1248.996	1365.100	15.701	0.4357	0.4574	95.109	2.179	4.473	4.890	957.9
HiCoM	40 0219 07	1168.829	1281.629	18.758	0.6164	0.3423	95.509	2.167	3.726	4.779	1234.9
HiCoM	40 0235 11	1051.011	1150.329	16.954	0.6239	0.3806	98.834	2.178	4.383	5.003	1067.5
HiCoM	40 0326 07	993.080	1089.033	20.145	0.6696	0.3322	103.674	2.334	4.684	5.399	1080.9
HiCoM	40 0345 08	1255.431	1381.497	18.372	0.6970	0.2996	99.830	2.253	4.118	4.683	1014.8
HiCoM	40 0379 09	2901.087	3219.697	23.584	0.7252	0.2856	83.136	2.332	6.397	5.179	767.6
HiCoM	40 0410 05	2982.209	3308.884	21.242	0.6969	0.3755	97.311	2.336	6.921	5.233	745.5
HiCoM	40 0438 02	644.502	705.035	21.658	0.6743	0.3581	98.307	2.283	3.300	4.795	1142.5
HiCoM	40 0536 06	2943.791	3267.711	24.720	0.7811	0.2891	88.022	2.392	6.584	5.272	792.5
HiCoM	40 0647 02	2954.072	3274.551	21.014	0.7904	0.2799	105.265	2.397	6.356	5.097	812.7
HiCoM	40 0657 05	586.328	643.746	21.660	0.7397	0.3486	100.336	2.325	2.956	4.624	1111.1

表S20。**表S18，续。**

方法	相机序列	后续帧 kB	摊销 kB	PSNR	SSIM	学习感知图像块相似度	第一帧（秒）	秒/帧	序列化 (ms)	解码 (ms)	渲染 FPS
HiCoM	40 0718 06	1015.610	1112.433	16.911	0.5399	0.4100	131.049	2.409	4.355	5.546	1006.4
HiCoM	40 0811 06	1264.599	1391.049	19.515	0.7510	0.3017	97.912	2.072	4.353	4.560	890.5
QUEEN-l	40 0173 02	4011.063	4070.350	17.982	0.6097	0.3387	135.636	4.547	494.910	360.807	794.5
QUEEN-l	40 0219 07	2258.409	2294.353	22.778	0.8218	0.2016	134.288	4.571	287.350	165.868	841.5
QUEEN-l	40 0235 11	1889.023	1908.310	20.688	0.7797	0.2265	124.676	4.488	364.870	245.978	834.2
QUEEN-l	40 0326 07	1685.625	1707.861	22.003	0.7933	0.2419	142.899	4.608	282.581	166.742	845.9
QUEEN-l	40 0345 08	2267.771	2292.600	20.844	0.7645	0.2676	149.622	4.732	287.185	184.402	835.1
QUEEN-l	40 0379 09	183690.716	188133.781	20.983	0.6502	0.3581	545.174	20.586	7302.344	6380.865	190.1
QUEEN-l	40 0410 05	158954.986	162794.101	20.363	0.6241	0.4001	610.550	18.400	6715.998	6058.561	208.4
QUEEN-l	40 0438 02	1941.090	1966.346	21.793	0.7127	0.3256	155.171	4.746	284.267	157.668	849.5
QUEEN-l	40 0536 06	194463.182	199176.677	22.693	0.6628	0.3548	602.551	23.431	8194.096	7355.756	167.2
QUEEN-l	40 0647 02	183560.819	188001.519	19.660	0.6911	0.3482	598.649	21.293	7459.476	6599.103	186.7
QUEEN-l	40 0657 05	2587.927	2629.904	23.192	0.7475	0.3237	152.203	4.280	182.587	113.020	891.5
QUEEN-l	40 0718 06	4079.114	4131.700	18.013	0.5693	0.3725	141.980	4.675	634.924	426.831	772.7
QUEEN-l	40 0811 06	2686.526	2712.549	21.532	0.7741	0.2911	149.128	4.571	511.741	343.377	815.3
3DGStream	40 0173 02	8022.227	8034.606	22.014	0.7884	0.1908	200.729	2.155	9.704	38.941	754.8
3DGStream	40 0219 07	8000.885	7985.765	25.360	0.8602	0.1781	204.988	2.142	8.979	42.788	862.3
3DGStream	40 0235 11	8003.732	7974.120	24.523	0.8512	0.1607	197.319	2.175	9.184	40.798	804.1
3DGStream	40 0326 07	7988.946	7945.213	27.474	0.8898	0.1579	188.441	2.145	25.565	40.802	851.9
3DGStream	40 0345 08	7988.476	7982.580	25.530	0.8827	0.1666	193.588	2.238	17.306	37.018	748.3
3DGStream	40 0379 09	7974.280	8024.764	30.156	0.9136	0.1300	210.423	2.257	7.892	47.291	720.8
3DGStream	40 0410 05	7993.785	8042.653	25.941	0.8244	0.2914	204.280	2.298	8.162	40.940	710.3
3DGStream	40 0438 02	7993.725	7932.900	26.573	0.8153	0.2663	188.871	2.128	8.274	40.139	912.8
3DGStream	40 0536 06	7983.984	8028.786	29.645	0.8986	0.1907	208.672	2.404	10.053	40.567	615.6
3DGStream	40 0647 02	7988.349	8056.197	29.018	0.9134	0.1930	216.001	2.593	12.343	41.981	678.3
3DGStream	40 0657 05	7987.085	7915.963	26.528	0.8579	0.2669	180.661	2.171	8.774	62.026	853.8
3DGStream	40 0718 06	8063.203	8095.365	19.741	0.6650	0.3122	203.321	2.408	13.518	44.273	734.6
3DGStream	40 0811 06	8013.570	8002.455	24.208	0.8514	0.2300	202.440	2.378	8.502	38.454	637.0

表 S21. **表S18，续。**

序列	层级	数据包千字节	摊销千字节	普通 PSNR	普通 SSIM	普通 学习感知图像块相似度	超网络 +NSC PSNR	超网络 +NSC SSIM	超网络 +NSC 学习感知图像块相似度
0173_02	Lite	94.766	107.668	21.436	0.7215	0.2179	22.138	0.7770	0.1362
0173_02	Pro	206.238	220.235	22.018	0.7557	0.1652	22.505	0.7951	0.1229
0173_02	Ultra	319.001	335.185	22.246	0.7717	0.1397	22.581	0.7980	0.1209
0219_07	Lite	90.404	103.306	24.793	0.8250	0.1452	26.171	0.8677	0.0904
0219_07	Pro	198.138	212.135	25.507	0.8535	0.1122	26.688	0.8842	0.0780
0219_07	Ultra	306.373	322.557	25.776	0.8632	0.0981	26.753	0.8859	0.0764
0235_11	Lite	114.291	127.193	23.692	0.7918	0.1624	25.326	0.8523	0.0953
0235_11	Pro	252.315	266.312	24.312	0.8213	0.1268	25.592	0.8635	0.0855
0235_11	Ultra	390.828	407.012	24.482	0.8324	0.1080	25.627	0.8659	0.0837
0326_07	Lite	85.939	98.841	23.107	0.8183	0.2107	24.843	0.8721	0.1195
0326_07	Pro	189.172	203.169	23.715	0.8392	0.1712	25.225	0.8833	0.1046
0326_07	Ultra	292.412	308.596	24.161	0.8532	0.1442	25.319	0.8861	0.1011
0345_08	Lite	95.443	108.345	24.605	0.8533	0.1745	27.121	0.9071	0.0843
0345_08	Pro	209.692	223.689	25.766	0.8833	0.1252	27.780	0.9195	0.0730
0345_08	Ultra	324.073	340.257	26.493	0.9010	0.0987	27.871	0.9218	0.0710
0379_09	Lite	68.451	81.353	28.157	0.8593	0.1919	30.053	0.8981	0.1032
0379_09	Pro	151.100	165.097	28.679	0.8802	0.1474	30.393	0.9067	0.0921
0379_09	Ultra	234.121	250.305	28.849	0.8887	0.1254	30.405	0.9081	0.0899
0410_05	Lite	88.073	100.975	27.036	0.8472	0.2218	29.444	0.8949	0.1211
0410_05	Pro	193.390	207.387	27.655	0.8659	0.1831	29.867	0.9044	0.1104
0410_05	Ultra	298.339	314.523	28.102	0.8776	0.1582	29.902	0.9055	0.1082

表 S22。 **超网络和 NSC 在每个序列和层级上都能改进相同数据包（其余 13 个序列，第一部分）**。 48 个评估相机，如表 1所示；数据包千字节是第 30 帧有效载荷，摊销千字节则加上会话状态的六十分之一。

序列层级		数据包千字节	摊销千字节	普通			超网络 +NSC		
				PSNR	SSIM	学习感知图像块相似度	PSNR	SSIM	学习感知图像块相似度
0438 02	Lite	85.680	98.582	27.183	0.8342	0.1815	29.302	0.8905	0.1004
0438 02	Pro	187.099	201.096	27.788	0.8572	0.1491	29.834	0.9022	0.0906
0438 02	Ultra	288.853	305.037	28.159	0.8691	0.1269	29.892	0.9037	0.0886
0536 06	Lite	74.995	87.897	28.715	0.8803	0.1877	31.365	0.9212	0.0999
0536 06	Pro	162.365	176.362	29.044	0.8895	0.1618	31.660	0.9268	0.0898
0536 06	Ultra	250.635	266.819	29.141	0.8961	0.1439	31.704	0.9276	0.0881
0647 02	Lite	85.814	98.716	26.024	0.8768	0.1710	29.297	0.9324	0.0882
0647 02	Pro	187.779	201.776	27.105	0.8931	0.1408	30.010	0.9396	0.0795
0647 02	Ultra	289.971	306.155	27.736	0.9020	0.1198	30.099	0.9407	0.0777
0657 05	Lite	77.179	90.081	27.194	0.8498	0.2044	30.443	0.9153	0.1040
0657 05	Pro	167.997	181.994	27.827	0.8644	0.1786	30.862	0.9221	0.0957
0657 05	Ultra	259.313	275.497	28.351	0.8746	0.1552	30.921	0.9229	0.0946
0718 06	Lite	113.828	126.730	21.028	0.7175	0.1784	21.725	0.7712	0.1267
0718 06	Pro	249.904	263.901	21.386	0.7419	0.1472	21.925	0.7851	0.1166
0718 06	Ultra	386.825	403.009	21.546	0.7549	0.1303	21.969	0.7873	0.1148
0811 06	Lite	121.913	134.815	27.084	0.8906	0.1351	28.629	0.9200	0.0800
0811 06	Pro	269.215	283.212	27.787	0.9054	0.1114	28.879	0.9247	0.0723
0811 06	Ultra	416.944	433.128	28.013	0.9110	0.0976	28.915	0.9258	0.0709

表 S23。 **超网络和 NSC 在每个序列和层级上都能改进相同数据包（其余 13 个序列，第二部分）**。 48 个评估相机，与表 1相同；数据包千字节为第 30 帧有效载荷，摊销千字节则加上六十分之一的会话状态。

序列层级		数据包千字节	摊销千字节	普通			超网络 +NSC		
				PSNR	SSIM	学习感知图像块相似度	PSNR	SSIM	学习感知图像块相似度
0012 09	Lite	95.742	108.644	20.338	0.7602	0.1607	21.678	0.8081	0.1113
0012 09	Pro	212.258	226.255	21.023	0.7863	0.1270	21.810	0.8159	0.1029
0012 09	Ultra	329.051	345.235	21.239	0.7966	0.1130	21.830	0.8167	0.1022
0022 10	Lite	108.916	121.818	27.066	0.8694	0.1957	29.053	0.9107	0.0997
0022 10	Pro	238.999	252.996	27.425	0.8842	0.1620	29.392	0.9176	0.0910
0022 10	Ultra	368.989	385.173	27.698	0.8916	0.1409	29.439	0.9188	0.0896
0166 04	Lite	115.911	128.813	21.738	0.7220	0.1548	22.850	0.7813	0.1090
0166 04	Pro	255.124	269.121	22.280	0.7495	0.1246	23.078	0.7923	0.1005
0166 04	Ultra	395.382	411.566	22.467	0.7618	0.1094	23.146	0.7948	0.0989

表 S24。 **超网络和 NSC 在每个序列和层级（0012 09、0022 10 和 0166 04）中都能改善相同数据包**。。 48 个评估相机，与表 1相同；数据包千字节指第 30 帧有效载荷，摊销千字节则在此基础上加上六十分之一的会话状态。

方法	相机序列	后续帧 kB	摊销 kB	PSNR	结构相似性指数	学习感知图像块相似度	第一帧（秒）	秒/帧	解码 (ms)	渲染 FPS
我们的轻量版	8 0012 09	94.7	107.6	22.367	0.8144	0.1125	–	0.13	–	–
我们的轻量版	8 0022 10	110.5	123.4	29.168	0.9154	0.1024	–	0.17	–	–
我们的轻量版	8 0166 04	116.6	129.5	22.142	0.7752	0.1245	–	0.16	–	–
我们的专业版	8 0012 09	209.9	223.9	22.512	0.8236	0.1045	–	0.13	–	–
我们的专业版	8 0022 10	242.8	256.8	29.415	0.9218	0.0946	–	0.18	–	–
我们的专业版	8 0166 04	256.2	270.2	22.408	0.7870	0.1148	–	0.17	–	–
我们的 Ultra 版	8 0012 09	325.1	341.2	22.541	0.8249	0.1036	–	0.13	–	–
我们的 Ultra 版	8 0022 10	374.9	391.1	29.471	0.9228	0.0921	–	0.18	–	–
我们的 Ultra 版	8 0166 04	396.8	413.0	22.457	0.7899	0.1119	–	0.17	–	–
HiCoM [16]	8 0012 09	3422.3	3792.2	14.013	0.5292	0.3217	112	2.16	6.5	732
HiCoM [16]	8 0022 10	3308.5	3667.7	19.206	0.6361	0.3759	88	2.11	5.5	695
HiCoM [16]	8 0166 04	2877.9	3177.2	14.589	0.3614	0.5144	72	2.06	5.6	823
QUEEN-l [17]	8 0012 09	52292.3	53534.2	11.054	0.5339	0.3891	52	1.31	4800.9	369
QUEEN-l [17]	8 0022 10	53149.8	54410.8	16.543	0.5408	0.4125	52	1.30	5905.5	367
QUEEN-l [17]	8 0166 04	52971.9	54234.1	13.762	0.4478	0.4792	51	1.25	5915.4	447
3DGStream [65]	8 0012 09	8001.3	8095.9	18.848	0.6870	0.2815	162	2.50	42.6	642
3DGStream [65]	8 0022 10	7985.1	8042.6	26.640	0.8554	0.2365	191	2.57	38.5	654
3DGStream [65]	8 0166 04	8015.1	8106.2	19.300	0.6318	0.3228	211	2.47	37.7	589
HiCoM [16]	40 0012 09	1644.2	1804.5	16.100	0.5770	0.3052	101	2.20	5.0	871
HiCoM [16]	40 0022 10	894.2	982.2	20.754	0.7143	0.3456	103	2.21	4.9	975
HiCoM [16]	40 0166 04	1158.5	1271.0	16.579	0.5070	0.4274	92	2.16	5.0	925
QUEEN-l [17]	40 0012 09	6373.7	6493.2	16.836	0.6844	0.2957	148	4.56	482.1	861
QUEEN-l [17]	40 0022 10	1685.9	1707.3	23.736	0.7951	0.2816	148	4.28	172.9	901
QUEEN-l [17]	40 0166 04	5287.4	5361.5	17.651	0.5555	0.3801	142	4.68	401.6	734
3DGStream [65]	40 0012 09	8042.3	8084.8	23.214	0.8238	0.1883	199	2.36	38.5	691
3DGStream [65]	40 0022 10	7992.8	7944.3	27.838	0.8844	0.2068	187	2.25	40.3	757
3DGStream [65]	40 0166 04	8042.5	8104.7	21.556	0.7460	0.2373	181	2.51	39.2	702

表 S25. **在 0012 09、0022 10 和 0166 04 上，各层级在质量、字节数和速度方面均优于八视图逐帧优化。** 后续帧千字节数不包括每种方法的首帧，或我们的会话状态；摊销千字节数则是第 0 帧与 59 个后续帧的平均值。时间是在一台 RTX 5090 上测得的；基线方法的数据来自它们在共享 GPU 上对第 0 至 59 帧的运行结果；我们的秒/帧是单个工作线程下每帧的稳态服务器时间。其他序列见表 S18。

方法	摄像机	空闲 GPU 上的优化（秒/帧） 试验 1	试验 2	试验 3	理想帧每秒 $G/t_{\min}$ $G{=}1$	$G{=}4$	我们 / $(G/t_{\min})$ $G{=}1$	$G{=}4$
HiCoM	8	2.004	2.003	2.065	0.499	1.997	16.05	16.13
HiCoM	40	2.200	2.143	2.267	0.467	1.867	17.17	17.26
QUEEN-l	8	0.918	0.872	0.909	1.147	4.589	6.98	7.02
QUEEN-l	40	4.499	4.649	4.696	0.222	0.889	36.05	36.24
3DGStream	8	1.967	1.932	1.981	0.517	2.070	15.48	15.56
3DGStream	40	2.117	2.128	2.156	0.472	1.889	16.96	17.05
Ours	8	–	–	–	8.01	32.22	–	–

表 S26。**在采用专门设计的 40 摄像头输入并实现理想扩展的情况下，拟合基线方法仍比我们的流式吞吐量低 17。0× 在相同图形处理器数量下。** 在空闲的 RTX 5090 上测试序列 0012 09；$t_{\min}$ 是每个后续帧平均拟合时间最快的方法，而 $G/t_{\min}$ 可在 G 个图形处理器上实现理想扩展。我们是在 16 个测试序列上测得的整台服务器吞吐量。

方法	摄像机	t 范围（秒/帧）	最大值 G/t（帧每秒）		我们 / (G/t)	
			G=1	G=4	G=1	G=4
Ours	8	–	8.01	32.22	–	–
HiCoM	8	1.96–2.77	0.51	2.04	15.7	15.8
QUEEN-l	8	1.23–1.38	0.81	3.25	9.9	9.9
3DGStream	8	2.22–5.21	0.45	1.80	17.8	17.9
HiCoM	40	2.07–2.41	0.48	1.93	16.6	16.7
QUEEN-l	40	4.28–23.43	0.23	0.93	34.3	34.5
3DGStream	40	2.13–2.59	0.47	1.88	17.1	17.1

表 S27. **在全部 96 次网格运行中，我们的服务器吞吐量均超过所有基线的理想扩展速率**。16 个序列中，共享 GPU 上每个后续帧的平均拟合时间为 t，每种方法和输入设置的理想吞吐量为 G/t；我们给出的是 16 个测试序列的整台服务器吞吐量，最后两列则将其分别除以每种方法的最大值。

格式	默认设置	编码时间（秒）	建模帧率（1 个图形处理器）	建模帧率（4 个图形处理器）
SPZ	v4	0.058	11.71	46.83
PlayCanvas 压缩 PLY	压缩 PLY	0.286	11.71	46.83
ksplat	第 1 级	0.279	11.71	46.83
FCGS	λ=10^{-4}	4.202	0.24	0.95
Draco	默认	0.072	11.71	46.83
PlayCanvas SOG	默认	1.882	11.71	46.83
FCGS，不含编码	–	0	11.71	46.83

表 S28。 **即使假设有 192 个并行 CPU 编码器且阶段间重叠达到理想状态，格式流水线仍至少会发送 173 兆比特/秒，高于 100 兆比特/秒的下行链路速率**。 表中各格式的建模吞吐量 S14 基于记录的编码时间、理想的阶段重叠和图形处理器扩展性，以及建模得到的头部和剪枝开销，在其默认设置下计算。

方法	每帧时间	每帧大小	数据集
Ours	前馈（32.2 帧/秒，4 个图形处理器）	94.8–321.9 千字节	DNA-Rendering
3DGStream [65]	12 s	7.6 兆字节	N3DV
HiCoM [16]	6.7 秒	0.7 兆字节	N3DV
QUEEN-l [17]	7.9 秒	0.75 兆字节	N3DV
Instant Gaussian Stream [83]	2.67 秒	7.90 兆字节	N3DV
4DGC [23]	–	0.5 兆字节	N3DV
MoRGS-l [34]	4.0 秒	3.80 兆字节	N3DV
QuARC-GS [50]	5.0 秒	36 千字节	N3DV
ReCon-GS [15]	6.4 秒	0.40 兆字节	N3DV
ComGS-s [3]	37 s	0.049 兆字节	N3DV
iFVC [67]	0.22 分钟	0.09 兆字节	N3DV
StreamSTGS [31]	–	173.6 千字节	N3DV
StreamRF [37]	0.25 分钟	17.7 兆字节	N3DV 火焰鲑鱼
TeTriRF [78]	0.65 分钟	62.5 千字节	DyNeRF
V^3 [75]	0.82 分钟	0.532 兆字节	ReRF（Kpop）
JointRF [94]	–	0.47 兆字节	ReRF
4DGCPro [95]	–	0.64 兆字节	N3DV
HiFi4G [27]	4 分钟	1.648 兆字节	自己的采集
DualGS [26]	12 分钟	0.323 兆字节	DualGS

表 S29。 **捕获时间尺度下的前馈流式处理，与数秒至数分钟的拟合**。 其他流式体积视频方法均采用其论文中列出的配置，数据和硬件也按论文所述；其中每帧时间为拟合时间；我们的方法则为实测值（第 4.5 节）。每帧时间栏中的短横线表示该配置未报告逐帧拟合时间。N3DV 和 DyNeRF 指的是同一个多视角视频数据集 [38]。

方法	输入视图	每个主体或序列的设置	每帧时间	GPU
每序列体积视频				
4K4D [81]	54 次训练（DNA-渲染；90% 共 60 个摄像机）	点云 + 序列拟合（~24 小时 / 200 帧）	~7.2 分钟（训练）†	RTX 4090
HumanRF [24]	124 训练（160 rig）	占用网格 +序列拟合（~1 天 / 1000 帧）	~86.4 秒（训练；4× 缩小后)†	RTX 3090（24 GB）
LongVolCap [82]	19–21（N3DV rig）	序列拟合（~2小时 / 1200 帧）	~6秒（训练）†	RTX 4090
HiFi4G [27]	81	网格化、跟踪 +逐帧拟合	< 7分钟（预处理 + 训练）	RTX 3090
DualGS [26]	81	每帧 + 关节/皮肤高斯 fit	12 分钟	RTX 3090
NeuS2 [77]	45–91 训练（DynaCap）	首帧拟合（~40秒） + 逐帧 fit	20 秒（训练）	RTX 3090
个体化虚拟形象				
可动画高斯 [40]	16–47	SMPL-X +模板；化身训练 ~2 天	–	RTX 4090
前馈				
GPS-Gaussian [93]	2（相邻）	none	25 帧/秒（TensorRT；一个视角来自两个 1024^2 输入）‡	RTX 3090
HiReFF [28]	4（未校准）	none	3.01 帧/秒（2K）‡	RTX 4090
Tele-Aloha [69]	4（上半身）	none	30 帧/秒（2K 系统）‡	RTX 4090
远摄360 [70]	4–6（非摆拍）	none	>25 帧/秒（2K 系统）‡	RTX 5090（推理）
实时网格系统				
Live4D [99]	20（次要）	none	24 帧/秒（实时网格）‡	4×RTX 3090 + 1× RTX 4090
HoloSplat（我们）	8	会话设置（~24秒）	8.0 帧/秒（1 个图形处理器）；32.2（4 个图形处理器）	RTX 5090

表 S30。 **HoloSplat 达到 32.2 帧每秒，而每个序列的人体重建每帧需要 6 秒至 12 分钟**。 作者报告的数字人方法：“每个主体或序列的设置”给出了显示前的训练或拟合时间，每帧时间不包括渲染，但标记为系统速率的项目除外 ‡。 †根据报告的总时间推导得出。Live4D 的摄像机和图形处理器数量来自 Tele-Aloha [69]；我们的行数据来自图 7。

方法	输入视图，分辨率	发送端	接收端
CodecSplat [87]	8, 256×256	276 毫秒	199 毫秒
HiReFF [28]	4，2K	4.0 帧每秒	–
HoloSplat（我们）	8， 512×512	8.0 帧每秒	3.9 毫秒

表 S31。 **在相同的 RTX 5090 和 DNA-Rendering 输入下，我们的客户端解码速度 51× 快于 CodecSplat 的接收端，且我们的服务器速度超过 HiReFF**。 已发布模型采用默认设置，CodecSplat 采用表 S32 的输入设置。CodecSplat 的发送端生成并压缩特征，其接收端对这些特征进行解码并预测高斯；我们的发送端包含在线抠图，而我们的接收端对 Pro 数据包进行解码，在其令牌上运行超网络，并生成和剪枝高斯。CodecSplat 的耗时和我们的服务器速率取 16 个测试场景的平均值（CodecSplat 取三次运行中的最佳结果）；HiReFF 的速率取 0012 09、0022 10 和 0166 04 上三次运行中的最佳结果，而我们的接收端耗时则取相同三个序列的中位数。

方法	设置	千字节 ↓	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓	解码（毫秒） ↓
CodecSplat [87]	λ=1024	45.4	23.14	0.743	0.268	199
我们的轻量版	–	94.8	26.84	0.870	0.104	–
我们的专业版	–	208.2	27.22	0.880	0.094	3.9
我们的 Ultra 版	–	321.9	27.27	0.882	0.092	–

表 S32。 **CodecSplat 的性能停滞在比我们最轻量层级低 3.7 dB 的水平，且解码速度更 51× 慢**。 DNA-Rendering，16 个测试场景，采用我们的评分方式和输入；CodecSplat 在我们的八个视角上运行其发布的模型，并使用其最佳输入设置，未应用掩码，数值为 256×256，其千字节数未计入用于移除其背景高斯的输入掩码。解码结果给出接收端在单张 RTX 5090 上的每帧时间，计时方式与表 S31相同。

拆分点 / 编解码器	设置	kB	PSNR	SSIM	学习感知图像块相似度
T（主成分分析）	主成分分析 $r = 1$，4 位	127.5	14.73	0.265	0.651
T（主成分分析）	主成分分析 $r = 2$，4 位	153.0	15.48	0.270	0.658
T（主成分分析）	主成分分析 $r = 4$，4 位	310.1	15.64	0.268	0.654
A（原始）	主成分分析 $r = 4$，4 位	379.7	13.51	0.226	0.670
A（原始）	主成分分析 $r = 4$，6 位	1115.8	13.58	0.234	0.668
A（原始）	主成分分析 $r = 16$，4 位	1259.0	15.38	0.233	0.649
A（原始）	主成分分析 $r = 38$，4 位	2982.4	15.39	0.233	0.641
A（标准化）	主成分分析 $r = 4$，4 位	323.2	13.19	0.205	0.682
A（标准化）	主成分分析 $r = 4$，6 位	991.8	13.43	0.219	0.679
A（标准化）	主成分分析 $r = 16$，4 位	1293.9	15.26	0.231	0.656
A（标准化）	主成分分析 $r = 16$，6 位	4341.9	15.30	0.242	0.650
A（标准化）	主成分分析 $r = 38$，6 位	10631.7	15.31	0.242	0.637
学习型编解码器	$\lambda = 16$	74.1	21.38	0.664	0.350
学习型编解码器	$\lambda = 128$	89.1	22.93	0.737	0.276
学习型编解码器	$\lambda = 1024$	111.4	23.14	0.743	0.268
未压缩	–	–	15.28	0.242	0.637

表 S33。**在第二个像素对齐的重建器上，先编码再输入头部，效果再次优于编码高斯属性**。在 16 个 DNA-Rendering 测试场景上，采用 CodecSplat 发布的基于 DepthSplat 的模型，将 64 维特征编码后输入高斯头（T），或将其后的 38 个属性（原始或标准化，A）进行编码，并用我们的主成分分析编解码器与其自有的学习型特征编解码器进行比较。率失真前沿；旁路行对未压缩特征进行解码。码率按完整数据包计算，其中包括八个输入视角的 1 位前景掩码，用于移除背景高斯；表 S32 未计入 CodecSplat 的字节数。发布的头部是在学习型编解码器之后训练的，因此该编解码器也会还原特征。

解码器 / 方法	Tier	PSNR	SSIM	学习感知图像块相似度	千字节/帧	压缩比
HiFi4G						
超网络 + NSC（已部署）	Lite	24.148	0.7414	0.2069	65.0	39.46
超网络 + NSC（已部署）	Pro	24.270	0.7511	0.1953	141.6	18.11
超网络 + NSC（已部署）	Ultra	24.306	0.7519	0.1931	219.2	11.70
NSC，普通解码（对照）	Lite	23.861	0.7278	0.2329	65.0	39.46
NSC，普通解码（对照）	Pro	24.059	0.7423	0.2104	141.6	18.11
NSC，普通解码（对照）	Ultra	24.184	0.7478	0.2011	219.2	11.70
仅超网络（对照）	Lite	22.771	0.6951	0.2334	65.0	39.46
仅超网络（对照）	Pro	22.851	0.7008	0.2208	141.6	18.11
仅超网络（对照）	Ultra	22.832	0.7008	0.2183	219.2	11.70
普通（对照）	Lite	22.396	0.6792	0.2903	65.0	39.46
普通（对照）	Pro	22.524	0.6899	0.2529	141.6	18.11
普通（对照）	Ultra	22.670	0.6947	0.2337	219.2	11.70
标准 PointSplat512	–	22.810	0.7017	0.2137	2564.6*	1.00
ActorsHQ						
超网络 + NSC（已部署）	Lite	25.721	0.8274	0.1809	177.9	59.76
超网络 + NSC（已部署）	Pro	25.806	0.8320	0.1703	404.4	26.29
超网络 + NSC（已部署）	Ultra	25.854	0.8333	0.1680	628.3	16.92
NSC，普通解码（对照）	Lite	25.343	0.8172	0.2016	177.9	59.76
NSC，普通解码（对照）	Pro	25.632	0.8272	0.1824	404.4	26.29
NSC，普通解码（对照）	Ultra	25.798	0.8315	0.1738	628.3	16.92
仅超网络（对照）	Lite	24.316	0.8084	0.1943	177.9	59.76
仅超网络（对照）	Pro	24.370	0.8118	0.1821	404.4	26.29
仅超网络（对照）	Ultra	24.411	0.8125	0.1792	628.3	16.92
普通（对照）	Lite	23.683	0.7962	0.2427	177.9	59.76
普通（对照）	Pro	23.898	0.8057	0.2140	404.4	26.29
普通（对照）	Ultra	24.195	0.8107	0.1938	628.3	16.92
标准 PointSplat512	–	24.438	0.8131	0.1749	10631.7*	1.00

表 S34。**在两种未见过的设备上，每个已部署层级都以少得多的字节数，在峰值信噪比和结构相似性指数方面优于标准 PointSplat512**。留出相机以及 NSC 研究中所有评估帧；其中“已部署”指超网络加 NSC。压缩比的计算方式是将标准 PointSplat512 的千字节/帧数值除以各行对应的数值。 *原始 float32 高斯；我们的数据包不包含会话状态。

解码器 / 格式	层级 / 设置	数据包千字节	会话千字节/帧	PSNR	SSIM	学习感知图像块相似度
超网络 + NSC	Lite	65.777	6.451	23.929	0.7414	0.2096
超网络 + NSC	Pro	143.330	6.998	24.048	0.7510	0.1976
超网络 + NSC	Ultra	221.623	8.091	24.123	0.7522	0.1952
普通	Lite	65.777	1.132	22.141	0.6749	0.2921
普通	Pro	143.330	1.679	22.266	0.6846	0.2547
普通	Ultra	221.623	2.772	22.406	0.6892	0.2366
仅超网络	Lite	65.777	6.080	22.494	0.6905	0.2368
仅超网络	Pro	143.330	6.627	22.579	0.6960	0.2242
仅超网络	Ultra	221.623	7.720	22.564	0.6960	0.2215
NSC，普通解码	Lite	65.777	1.502	23.648	0.7263	0.2353
NSC，普通解码	Pro	143.330	2.048	23.884	0.7422	0.2128
NSC，普通解码	Ultra	221.623	3.142	24.008	0.7476	0.2033
未压缩 32 位浮点数	–	2487.987	0.000	22.567	0.6966	0.2176
SPZ v4	–	578.908	0.000	22.575	0.6969	0.2181
PlayCanvas 压缩 PLY	–	724.047	0.000	22.567	0.6966	0.2177
PlayCanvas SOG	WebP 压缩级别 0	650.106	0.000	22.566	0.6966	0.2178
PlayCanvas SOG	WebP 压缩级别 4	606.978	0.000	22.566	0.6966	0.2178
PlayCanvas SOG	WebP 压缩等级 9	605.583	0.000	22.566	0.6966	0.2178
ksplat	层级 0	1959.967	0.000	22.584	0.6970	0.2175
ksplat	第 1 级	1073.493	0.000	22.584	0.6970	0.2176
ksplat	第 2 级	1073.493	0.000	22.584	0.6970	0.2176
Draco	8 位	466.166	0.000	22.591	0.6959	0.2248
Draco	12 位	777.045	0.000	22.567	0.6966	0.2177
Draco	16 位	1087.836	0.000	22.567	0.6966	0.2176
FCGS [7]	λ=1.6×10^{-3}	494.278	0.000	22.415	0.6874	0.2407
FCGS [7]	λ=8×10^{-4}	517.825	0.000	22.495	0.6915	0.2331
FCGS [7]	λ=4×10^{-4}	550.084	0.000	22.475	0.6914	0.2311
FCGS [7]	λ=2×10^{-4}	588.233	0.000	22.514	0.6934	0.2271
FCGS [7]	λ=10^{-4}	621.509	0.000	22.557	0.6943	0.2266
均匀 + Zstd	4 位	495.313	0.000	22.634	0.6885	0.2440
均匀 + Zstd	6 位	633.680	0.000	22.574	0.6963	0.2190
均匀 + Zstd	8 位	752.430	0.000	22.571	0.6966	0.2177
FP16 + Zstd	第 1 级	1167.392	0.000	22.567	0.6966	0.2176
FP16 + Zstd	等级 5	1139.390	0.000	22.567	0.6966	0.2176
FP16 + Zstd	等级 9	1133.906	0.000	22.567	0.6966	0.2176
Draco	位置 11 / 属性 8 位	516.179	0.000	22.571	0.6966	0.2178
PlayCanvas SOG	默认	607.919	0.000	22.566	0.6966	0.2178

表 S35。**在 HiFi4G 上，每个已部署层级在未校准高斯上的峰值信噪比和结构相似性指数均高于所有格式设置，同时数据包大小更小**。每个会话的第一帧、中间帧和最后一帧评估帧，所有留出相机。数据包千字节不包括会话状态；会话千字节/帧将每个解码器的 PCA基、超网络更新和 NSC 更新分摊到其会话中评估的帧上。各格式对未校准高斯进行编码。

解码器 / 格式	层级 / 设置	数据包千字节	会话千字节/帧	PSNR	SSIM	学习感知图像块相似度
超网络 + NSC	Lite	180.606	6.709	25.722	0.8278	0.1809
超网络 + NSC	Pro	408.830	7.277	25.793	0.8323	0.1704
超网络 + NSC	Ultra	634.597	8.414	25.882	0.8345	0.1681
普通	Lite	180.606	1.177	23.777	0.7994	0.2401
普通	Pro	408.830	1.746	24.035	0.8093	0.2124
普通	Ultra	634.597	2.883	24.386	0.8157	0.1917
仅超网络	Lite	180.606	6.323	24.457	0.8126	0.1925
仅超网络	Pro	408.830	6.891	24.518	0.8164	0.1796
仅超网络	Ultra	634.597	8.028	24.606	0.8183	0.1764
NSC，普通解码	Lite	180.606	1.562	25.328	0.8176	0.2021
NSC，普通解码	Pro	408.830	2.130	25.606	0.8276	0.1830
NSC，普通解码	Ultra	634.597	3.267	25.849	0.8325	0.1737
未压缩 32 位浮点数	–	7273.588	0.000	24.711	0.8203	0.1731
SPZ v4	–	1725.248	0.000	24.728	0.8201	0.1753
PlayCanvas 压缩 PLY	–	2115.379	0.000	24.707	0.8199	0.1748
PlayCanvas SOG	WebP 压缩等级 0	1827.164	0.000	24.673	0.8196	0.1745
PlayCanvas SOG	WebP 压缩等级 4	1689.433	0.000	24.673	0.8196	0.1745
PlayCanvas SOG	WebP 压缩等级 9	1687.178	0.000	24.673	0.8196	0.1745
ksplat	第 0 级	5720.082	0.000	24.662	0.8194	0.1746
ksplat	第 1 级	3128.467	0.000	24.662	0.8194	0.1746
ksplat	第 2 级	3128.467	0.000	24.662	0.8194	0.1746
Draco	8 位	1298.538	0.000	24.678	0.8167	0.1850
Draco	12 位	2207.859	0.000	24.712	0.8203	0.1730
Draco	16 位	3117.056	0.000	24.711	0.8203	0.1730
FCGS [7]	λ=1.6×10^{-3}	1315.263	0.000	24.140	0.8101	0.2020
FCGS [7]	λ=8×10^{-4}	1377.575	0.000	24.358	0.8139	0.1919
FCGS [7]	λ=4×10^{-4}	1486.304	0.000	24.376	0.8148	0.1879
FCGS [7]	λ=2×10^{-4}	1575.558	0.000	24.457	0.8159	0.1836
FCGS [7]	λ=10^{-4}	1675.071	0.000	24.519	0.8163	0.1819
均匀 + Zstd	4 位	1409.117	0.000	24.186	0.7847	0.2284
均匀 + Zstd	6 位	1847.243	0.000	24.731	0.8198	0.1752
均匀 + Zstd	8 位	2160.051	0.000	24.713	0.8203	0.1733
FP16 + Zstd	第 1 级	3393.563	0.000	24.713	0.8203	0.1731
FP16 + Zstd	第 5 级	3242.388	0.000	24.713	0.8203	0.1731
FP16 + Zstd	第 9 级	3208.188	0.000	24.713	0.8203	0.1731
Draco	位置 11 / 属性 8 位	1441.924	0.000	24.713	0.8203	0.1735
PlayCanvas SOG	默认	1659.841	0.000	24.673	0.8196	0.1745

表 S36。**在 ActorsHQ 上，每个已部署层级在未校准高斯上的峰值信噪比和结构相似性指数均优于所有格式设置，且数据包大小更小**。每个会话的第一帧、中间帧和最后一帧评估帧，涵盖所有留出相机。数据包千字节不包括会话状态；会话千字节/帧将每个解码器的 PCA 基、超网络更新和 NSC 更新分摊到其对应会话的评估帧中。各格式对未校准高斯进行编码。

S6. 神经会话校准细节

NSC 在每个会话和层级上仅拟合一次，所用数据为一秒预热采集，即实时流媒体开始前八个输入视角的第0–29帧（图 S11）；随后从第 30 帧开始实时采集，并使用冻结的会话状态。每次拟合执行 300 步 AdamW，学习率为 10^{-3}，其中包含 50 个预热步数，权重衰减为 10^{-4}，并进行 0.5 的梯度裁剪。在 10^{-4} 下进行拟合时，三个层级的全部 48 个 DNA-Rendering 会话的损失都会高出 5–17%，因此每个会话（包括未见过的设备）都使用 10^{-3}。超网络更新和 NSC 共同恢复了通用超网络模糊掉的服装颜色和印刷细节（图 S12）。

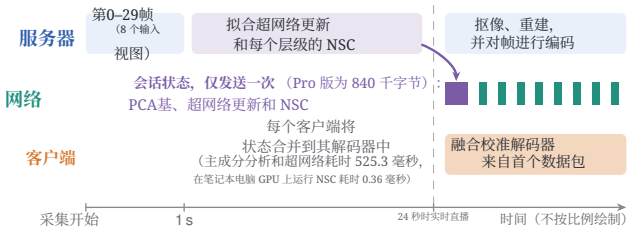

图 S11. **神经会话校准是一次性设置步骤，不会增加任何每帧开销**。与摄像机标定类似，服务器会在实时流媒体开始前，利用采集的第一秒数据拟合超网络并生成 NSC 更新，然后将它们与第 0 帧的 PCA基一并传输一次；每个客户端都会将它们合并到自己的融合解码器中，并解码共享令牌流。服务器设置时间为使用四块 RTX 5090 GPU 的 Pro 配置下的数据。

	HiFi4G			ActorsHQ		
配置	Lite	Pro	Ultra	Lite	Pro	Ultra
未压缩，不剪枝	22.821			24.438		
未压缩，采用相同的剪枝	22.808			24.439		
未压缩，采用相同的剪枝 + NSC	24.271			25.904		
普通	22.396	22.524	22.670	23.683	23.898	24.195
超网络	22.771	22.851	22.832	24.316	24.370	24.411
NSC，普通解码	23.861	24.059	24.184	25.343	25.632	25.798
超网络 + NSC（已部署）	24.148	24.270	24.306	25.721	25.806	25.854

表 S37。**在未见过的设备上，编解码器带来的开销很小；超网络和会话校准则带来了显著提升**。一秒预热后所有帧在留出相机上的峰值信噪比。“相同剪枝”表示不使用编解码器，采用我们的令牌剪枝和客户端剪枝；“不剪枝”表示两者均不采用。NSC 针对每个层级分别在输入视图上拟合；未压缩的 +NSC 行则不使用编解码器。

未见过的设备上的每个层级。 表 S37 将图 4扩展至所有层级和两种解码器，而表 S38则给出了每个序列及其自助法区间。

笔记本客户端在 525.3 毫秒内合并 PCA 基和超网络更新，并在 0.36 毫秒内完成 NSC。

预热预算。表 S40 调整了预热帧 F和步数 S，用于普通解码时的 NSC。单帧配合 300 步，便能分别恢复 HiFi4G 和 ActorsHQ 完整增益的 80% 和 93%。部署采用 F=30 和 S=300作为通用的零初始化协议，无需设备的先验数据。

校准目标。 在 Pro 上恢复的令牌（表 S39），更新三个逐槽位多层感知机层中的全部 1.31 M 个参数，会导致模型过拟合 30 帧校准窗口，使留出视图的峰值信噪比降低 0.13 dB，并使后续帧降低 0.35 dB。从校准损失中移除学习感知图像块相似度项，会使学习感知图像块相似度下降 0.012。在 HiFi4G 和 ActorsHQ 上，包含 12 个参数的颜色仿射变换使峰值信噪比分别变化 +0.076 和 +0.119 dB；而我们的秩为 5 的逐槽位多层感知机更新（44.5 千字节）达到了同等大小逐帧特征适配器增益的 95.7% 和 88.9%，且不产生逐帧开销。将八个视图分配到四个 RTX 5090 GPU 后，Pro 在采集开始 24 秒后即可上线；此时间包括 1 秒采集、零帧主成分分析、超网络更新和 NSC；一台

训练域。在 16 个 DNA-Rendering 测试场景中，NSC 为 +1.068/+1.154/+1.178 dB 的增益加到 Lite/Pro/Ultra 层级的超网络上，在所有层级的全部 16 个场景中提升了峰值信噪比，同时也提升了结构相似性指数和学习感知图像块相似度（表 S41），因此 DNA-Rendering 会话也部署了它。

S7. 实现细节

超网络。表S42展示了 DNA-Rendering 上三个层级的普通主成分分析解码、预训练通用超网络及其一次性超网络更新。仅使用通用权重即可恢复



图 S12. **超网络更新和 NSC 可恢复留出相机视角下的服装颜色和印花细节**。 从留出相机视角观察的两个 DNA-Rendering 测试序列中的第 30 帧，由预训练通用超网络从同一 Pro 数据包解码而来，分别未使用和使用基于第0–29帧拟合的会话状态（超网络更新和 NSC）；旁边为真实值。插图放大了框选区域，标签给出了逐视图 PSNR。

序列	Tier	PSNR	Δ_{unc}	Δ_{hyper}	n_f^+/n_f	n_c^+/n_c
问候	Lite	23.667	$+0.875\,[+0.713, +1.054]$	$+1.037\,[+0.857, +1.212]$	120/120	8/8
问候	Pro	23.845	$+1.053\,[+0.897, +1.209]$	$+1.014\,[+0.871, +1.152]$	120/120	8/8
问候	Ultra	23.908	$+1.115\,[+0.979, +1.264]$	$+1.111\,[+0.955, +1.266]$	120/120	8/8
雨伞	Lite	24.101	$+2.060\,[+1.665, +2.417]$	$+2.101\,[+1.673, +2.492]$	120/120	8/8
雨伞	Pro	24.246	$+2.205\,[+1.820, +2.549]$	$+2.187\,[+1.795, +2.538]$	120/120	8/8
雨伞	Ultra	24.282	$+2.241\,[+1.830, +2.611]$	$+2.248\,[+1.831, +2.624]$	120/120	8/8
换衣服	Lite	24.675	$+1.083\,[+0.853, +1.336]$	$+0.993\,[+0.833, +1.181]$	120/120	8/8
换衣服	Pro	24.720	$+1.128\,[+0.900, +1.382]$	$+1.058\,[+0.867, +1.284]$	120/120	8/8
换衣服	Ultra	24.728	$+1.137\,[+0.905, +1.393]$	$+1.064\,[+0.856, +1.305]$	120/120	8/8
Actor01	Lite	26.000	$+1.545\,[+1.084, +1.992]$	$+1.685\,[+1.285, +2.077]$	119/119	8/8
Actor01	Pro	26.121	$+1.666\,[+1.198, +2.123]$	$+1.758\,[+1.311, +2.187]$	119/119	8/8
Actor01	Ultra	26.180	$+1.725\,[+1.276, +2.161]$	$+1.750\,[+1.323, +2.167]$	119/119	8/8
Actor03	Lite	25.441	$+1.018\,[+0.669, +1.367]$	$+1.125\,[+0.788, +1.456]$	112/112	8/8
Actor03	Pro	25.491	$+1.068\,[+0.742, +1.388]$	$+1.115\,[+0.795, +1.420]$	112/112	8/8
Actor03	Ultra	25.527	$+1.105\,[+0.787, +1.412]$	$+1.136\,[+0.818, +1.442]$	112/112	8/8

表 S38。**NSC 在每个未见过的设备序列和层级上都能提高峰值信噪比**。已部署层级的峰值信噪比，以及它们与我们的未压缩重建 (Δ_{unc}) 和不含 NSC 的超网络 (Δ_{hyper}) 之间的差异；数值为通过交叉自助法计算的 95% 区间，该方法独立地对帧和摄像机进行重采样（10,000 次抽样）； n_f^+/n_f 和 n_c^+/n_c 分别统计 NSC 提高峰值信噪比的帧数和留出相机数。

在轻量版/Pro/Ultra 层级分别恢复 31.9/38.7/24.1% 的编码损失；即使对每个解码器都拟合 NSC，超网络更新仍可将这一比例提升至 66.7/80.4/68.5%，并分别增加 +0.296/+0.193/+0.098 dB。针对 16 个测试场景的受控消融实验支持空间上下文和头部度量监督的作用。添加 8 个最近邻上下文后，轻量版/Pro 的峰值信噪比分别提升 +0.107/+0.024 dB；随机打乱邻居对应关系则会抵消这一增益（比空间邻域低 0.109/0.019 dB）。在头部度量中进行监督 $W^\top W$ 比未加权令牌均方误差多解释了 2.0/2.6 个百分点的头部空间残差能量，并将学习感知图像块相似度提升了 0.0024/0.0015。通用超网络使用 AdamW 训练 30,000 步，峰值学习率为 3×10^{-4}（预热 1000 步，余弦衰减至 3×10^{-5}，权重衰减 0.01），每批处理来自四个场景的 4096 个令牌；在 160 个训练场景上涵盖十种秩和位深设置，并在三个留出主体上进行验证。超网络更新以 10^{-3} 的学习率，在第0–29帧上训练 1000 步，每层级耗时

校准目标 / 损失	参数	kB	留出视图 PSNR	留出视图 学习感知图像块相似度	后续帧 PSNR	后续帧 学习感知图像块相似度
逐槽位多层感知机， r=5 （已部署）	20,480	41.0	25.944	0.0916	24.827	0.1022
逐槽位多层感知机, r=2	8,192	16.4	25.945	0.0909	24.865	0.1007
逐槽位多层感知机, r=8	32,768	65.5	25.937	0.0916	24.775	0.1027
逐槽位多层感知机, r=32	131,072	262.1	25.907	0.0930	24.636	0.1059
完整逐槽位多层感知机	1.31 M	2625.5	25.816	0.0963	24.475	0.1115
仅输出层， r=5	25,670	51.3	25.829	0.0931	24.912	0.0991
逐槽位多层感知机 r=5 +所有偏置	38,926	77.9	25.940	0.0915	24.794	0.1025
逐槽位多层感知机 + 输出， r=5	46,150	92.3	25.938	0.0917	24.806	0.1023
L_1 仅使用损失（不含结构相似性指数和学习感知图像块相似度）	20,480	41.0	26.008	0.1070	24.771	0.1210
不含学习感知图像块相似度（$0.8L_1 + 0.2$（1−结构相似性指数））	20,480	41.0	26.052	0.1034	24.869	0.1166

表 S39。**对于 Pro令牌，在感知监督下对三个逐槽位多层感知机层进行秩5更新，可避免全多层感知机过拟合和感知退化**。在八个输入视角的第0–29帧上进行固定的300步拟合；千字节数表示 FP16参数字节数。留出视角得分是在六个场景的第52帧，对 0012 09、0022 10 30 40 89 0166 04 个非输入相机的评分；后续帧得分是在帧 – 至 和 上，对八个留出相机的评分。 – –

预热帧 F	S=30	S=100	S=300
1	0.87 / 0.82	1.16 / 1.47	1.20 / 1.53
3	0.87 / 0.86	1.14 / 1.50	1.27 / 1.61
10	0.90 / 0.88	1.19 / 1.50	1.36 / 1.65
30	0.98 / 1.05	1.36 / 1.46	1.50 / 1.63

表 S40。**仅一帧预热即可恢复完整校准增益的 80% 和 93%**。在首个 F 帧上拟合 S 步的秩为 5 的 NSC 时，Pro 相较普通流的 PSNR 增益（dB，HiFi4G / ActorsHQ）；本预算研究对所有三个层级的普通解码流拟合一个 NSC，而已部署的会话则为每个层级分别拟合一个。完整增益为一秒预热的增益（F=30, S=300 ）。

2.5–4.2 秒，运行于一台 RTX 5090 上。在笔记本电脑上使用 FP16 时，超网络在全部 1,488 个解码帧中保留了与 FP32 完全一致的 20% 剪枝后高斯元选择，并且恢复的编码损失占比变化不超过 8×10^{-6}。

计时。 BiRefNet Lite 基于带掩码标注的 DNA-Rendering 数据，在官方 Lite 检查点上以 512 分辨率进行微调，并以批大小为四运行骨干网络 CUDA 图。阿尔法量化为 8 位，并在重建前乘以按面积平均的有效区域。每个图形处理器上的三个生产进程通过专用 CUDA MPS 上下文共享该模型；所有多 GPU 服务器均采用此设置。帧从内存中回放，且已完成去畸变。稳定吞吐量统计源帧 36–131 为$96/(t_{131} - t_{35})$，其中 t 表示数据包就绪的时间；每个场景取三次运行中的最佳结果，再对所有场景求平均。吞吐量不受层级影响：抠图和重建是每帧的主要耗时，而秩为 48（Pro）和 秩为 80 的编码在相同图形处理器和场景上的吞吐量比为 1.002。客户端解码、超网络、头部和剪枝的计时范围为从内存中的数据包到传递给渲染器的高斯，使用 3 个测试序列的真实数据包（每个 Pro 数据包平均含 5,628 个令牌组）；超网络采用 FP16，头部采用 BF16。在配备融合 Triton 8 近邻核的笔记本电脑 RTX 3050 Ti 上，轻量版/Pro/Ultra 的所有计时通道中位数分别为 21.7/22.1/22.8 毫秒（Pro 中位帧为 0.76/5.55/15.80 毫秒）；在 RTX 5090 上，启动时选择 Torch 体素搜索，Pro 的耗时为 3.9 毫秒。

实时广播。在 30 帧预热后，以 30 帧每秒进行的 64 秒（1,920 帧）实时广播中，五个 RTX 5090 GPU 通过 Wi-Fi 向笔记本电脑 RTX 3050 Ti 传输 Pro 数据包，且无漂移。三段序列的采集到渲染帧端到端延迟平均为 204–304 毫秒：服务器端从采集到传输为 171–261 毫秒，Wi-Fi 传输为 10–14 毫秒，客户端解码为高斯耗时 21–33 毫秒，gsplat [86] 光栅化耗时 1.8–2.0 毫秒。将 BiRefNet Lite 替换为视频抠图网络 RVM [43] 并基于其遮罩重新拟合会话状态，会使 Pro 在全部 16 个测试序列上的 PSNR 降低，平均降低 1.84 dB。

重建器选择。 我们基于 PointSplat [21]，因为它是一种专为多视图人体采集设计的前馈式重建器，其每令牌高斯头提供了一个自然的拆分点。它的 Point-Image Transformer 将所有输入视图融合到表演者粗略点代理周围的一个共享令牌集中，每个令牌驱动 16 个高斯槽位，因此无论融合多少个视图，每个空间令牌都只需编码一次；不同于像素对齐重建器，后者会为每个输入视图生成一个高斯图。在八个视图下，它还

Tier	解码器	PSNR	SSIM	学习感知图像块相似度
Lite	普通	24.950	0.8198	0.1809
Lite	超网络	25.772	0.8491	0.1239
Lite	超网络 + NSC	26.840	0.8700	0.1043
Pro	普通	25.582	0.8419	0.1459
Pro	超网络	26.065	0.8598	0.1121
Pro	超网络 + NSC	27.219	0.8802	0.0941
Ultra	普通	25.904	0.8528	0.1256
Ultra	超网络	26.095	0.8614	0.1097
Ultra	超网络 + NSC	27.273	0.8818	0.0923
Pro	NSC，普通解码	26.758	0.8697	0.1091

表 S41。**NSC 在轻量版/Pro/Ultra 层级下，在超网络基础上分别提升 +1.068/+1.154/+1.178 dB**。在 16 个测试场景中使用相同数据包；针对每个序列和层级拟合的 NSC 更新基于八个输入相机的第0–29帧；每帧字节数保持不变，且每次更新均为一次性下载，大小为 44.5 千字节。

Tier	kB	普通通用	+ 超-网络更新	Gain (dB)	已恢复 (%)
Lite	94.8	24.950	25.343	25.772 +0.822	66.7
Pro	208.2	25.582	25.815	26.065 +0.483	80.4
Ultra	321.9	25.904	25.971	26.095 +0.191	68.5

表 S42。**预训练通用超网络和超网络更新都能在每帧字节数为零时恢复编码损失**。在 16 个测试场景上，对以普通方式解码的相同数据包、仅使用通用超网络以及使用其一次性超网络更新时的峰值信噪比进行比较，其中 Recovered 表示恢复的编码损失份额（经过匹配剪枝的未压缩结果减去普通流）

在表格中的前馈重建器中表现最佳。S32和 S43为 26.17 dB，而 CodecSplat（表 S32）为 23.14 dB，GPS-Gaussian、DA3 和 WorldMirror 2.0 则为 18.38–20.47 dB。

无状态流式传输。对真实数据包进行离线回放，揭示了无状态编码的四项流式传输特性。(i) 在第 30、60 或 89 帧加入的观看者，在 Pro 层级下下载的数据量恒定为 1.05–1.06 兆字节（0.84 兆字节会话状态加一个数据包）。有状态流会累积此前的每次更新，在 16 个场景中，3DGStream、HiCoM 和 QUEEN-l 的累计数据量分别达到 11.7/251.5/483.3 兆字节、22.9/114.4/204.0 兆字节和 127.8/1710.6/3237.8 兆字节。(ii) 在 0012 09 序列中丢失第 45 帧，对 Pro 仅影响该帧；到第 46 帧时，渲染便恢复为逐位精确。3DGStream 的相对神经变换缓存会使错误传播到全部 25 个后续帧（第 70 帧比无损流低 2.94 dB）。(iii) 当三个合并的分层解码器常驻客户端（总计 137 兆字节）时，在 0012 09 和 0022 10 序列的第 30–89 帧中，每 10 帧切换一次层级或逐帧随机切换层级，解码出的高斯均与各层级的固定流逐位一致：对所有解码后的均值、尺度、旋转、不透明度和颜色进行逐元素比较，发现没有任何帧存在差异，最大绝对差为 0；此外，10 帧训练计划的渲染峰值信噪比和学习感知图像块相似度也与固定流完全相同。(iv) 在由轨迹驱动、时长 60 秒的 FIFO 链路模拟中，以 30 帧每秒运行并设定 200 毫秒截止期限（阶跃、方波和随机游走带宽），基于因果关系的逐帧层级自适应将阶跃轨迹中错过截止期限的帧占比从 74.9% 降至 5.0%。当最早的未发送帧超过截止期限时，清除排队的未发送帧并发送最新帧，可将跳过或延迟帧的占比降至 0.85%（跳过 0.67%，交付 29.8 帧每秒）。三种逐序列基线方法均错过了 100% 的截止期限即使带宽恒定为 100 兆比特/秒。

S8. 前馈重建器

在场景或其他采集装置上训练的通用重建器，在八个视图下会丢失表演者的细节（表 S43），并导致面部和四肢模糊、重影或破碎（图 S13）：表 S43中加入的三个重建器里，表现最好的 WorldMirror 2.0 达到 20.47 dB，比我们最轻量的层级低 6.37 dB，后者为 22。7× 兆字节的 Pro。我们在八个输入视图、软掩码和表 S4的评分设置上运行另外三个前馈重建器的已发布模型：Depth Anything 3 [42] 及其高斯头、WorldMirror 2。0 [68] 和 GPS-Gaussian [93]。DA3 和 WorldMirror 2.0 按各自约定接收标定相机，并通过八个输入相机对齐到采集坐标系；WorldMirror 2.0 使用这些相机，通过其已发布的`gsdepth+gtcamera` 路径反投影其预测深度。GPS-Gaussian 根据标定几何关系，从八个输入相机组成互不重叠的立体对，在其原生 1024^2 分辨率下进行校正，使标定光轴的参考点主点处于零视差位置，然后合并其高斯。与 CodecSplat 一样，DA3 和 WorldMirror 2.0 分别使用掩码和未掩码输入运行，并报告场景平均峰值信噪比较高的变体；GPS-Gaussian 则像其发布版本一样使用掩码输入。每个重建器都将高斯限制在输入阿尔法范围内。补充视频中的前馈片段（0811 06 和 0438 02）标注了各方法在该序列评估中的每帧大小（千字节/帧）和峰值信噪比，结果来自表S23、S32和S43；GPS-Gaussian、DA3、WorldMirror 2.0 和 Pro 还标注了表S43中的帧每秒。PointSplat512 是我们服务器运行的重建器，因此其片段标注了我们的四图形处理器速率 32.2 帧每秒；CodecSplat 则需要其发送端和接收端共同生成

方法	高斯	kB ↓	兆比特/秒 ↓	帧每秒 ↑	峰值信噪比 ↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
GPS-Gaussian [93]	533k	29,860.3	7,166.5	34.6	18.38	0.602	0.297
DA3 [42]	176k	26,714.3	6,411.4	6.6	18.98	0.581	0.342
WorldMirror 2.0 [68]	84k	4,731.8	1,135.6	6.9	20.47	0.631	0.302
PointSplat512 [21]	85k	4,753.1	1,140.8	–	26.17	0.865	0.103
我们的轻量版	**65k**	**94.8**	**22.8**	–	26.84	0.870	0.104
我们的专业版	**65k**	208.2	50.0	32.2	27.22	0.880	0.094
我们的 Ultra 版	**65k**	321.9	77.3	–	**27.27**	**0.882**	**0.092**

表 S43。 **HoloSplat 的每个层级在峰值信噪比和结构相似性指数方面均优于 GPS-Gaussian、DA3、WorldMirror 2.0 和 PointSplat512，且至少少用 14。7× 字节。** DNA-Rendering，16 个测试场景，第 30 帧，八个输入视角及表 S4中的评分；对于重建器，千字节数按每帧高斯以未压缩 32 位浮点数计算；对于我们的层级，则按数据包计算。兆比特/秒为 30 帧/秒时的有效载荷速率；帧每秒显示四个 RTX 5090 GPU 的性能， $4/t$ 根据重建器使用单个图形处理器的推理时间计算， t 输入和阿尔法已位于图形处理器上，并显示我们的服务器吞吐量。



图 S13. **从八个视图来看，HoloSplat 保持了面部清晰、四肢完整，而 CodecSplat 和 DA3 的面部模糊，WorldMirror 2.0 的手臂出现重影，GPS-Gaussian 则使身体支离破碎。**第 30 帧的留出视图及逐视图 PSNR，其中我们的专业版领先第二名 PointSplat512 1.93 和 2.42dB；插图放大显示 2× 面部。

高斯，因此标注为“< 9帧每秒”，因为在一台 RTX 5090 上，其发送端和接收端耗时 276 + 199 毫秒，四除以该耗时可得 8.4 帧每秒（表 S31）。其他片段标注的数值（每帧大小（千字节/帧）、0811 06 / 0438 02 上的峰值信噪比，单位为 dB）如下：PointSplat512：5,949，28.18 / 4,462，28.31；CodecSplat：46.5，23.74 / 46.7，24.68；DA3：32,908，19.65 / 24,100，21.03；WorldMirror 2.0：5,868，19.82 / 4,591，21.92；GPS-Gaussian：33,018，16.79 / 23,592，20.52。

输入来自同一个相机环上的一个摄像机。GPS-Gaussian 通过立体视觉匹配邻近视图，Tele360 [70] 则使用从同一个相机环均匀采样的输入。因此，我们也以 45° 间隔从同一个相机环选取八个输入，让 GPS-Gaussian 合并八组循环相邻对，并在相同的 16 个场景上重新运行 GPS-Gaussian、PointSplat512 和全部三个层级，同时保留各自的会话状态。表S44显示，HoloSplat 依然领先：Pro 达到 27.51 dB，比 26.5× 少用字节的 PointSplat512 高 0.93 dB，比 GPS-Gaussian 高 10.99 dB。

方法	千字节 ↓	峰值信噪比↑	结构相似性指数 ↑	学习感知图像块相似度 ↓
GPS-Gaussian [93]	95,757.8	16.52	0.493	0.375
PointSplat512 [21]	5,495.7	26.58	0.870	0.099
我们的轻量版	**94.6**	27.12	0.876	0.099
我们的专业版	207.4	27.51	0.886	0.089
我们的 Ultra 版	320.4	**27.59**	**0.888**	**0.087**

表 S44。 **以相机环上的 8 个输入为条件时，HoloSplat 的每个层级在峰值信噪比和结构相似性指数方面仍优于 GPS-Gaussian 和 PointSplat512**。 DNA-Rendering，16 个测试场景，第 30 帧，以摄像机 1、7、……、43 为输入；评分和千字节单位同表 S43。