arXiv:2601.03319v1 [cs.GR] 6 Jan 2026

CaricatureGS：利用高斯曲率夸张化三维高斯泼溅面部

EldadMatmon Amit Bracha Noam Rotstein Ron Kimmel

以色列海法，以色列理工学院



图 1. 我们的方法生成的照片级真实感 3D 夸张肖像化虚拟形象。

摘要

本文介绍一种面向面部的照片级真实感、可控三维漫画化框架。我们首先采用一种基于高斯曲率的内在表面夸张技术，但该技术与纹理结合时，往往会产生过度平滑的渲染结果。为解决这一问题，我们转而采用近期已被证明能够生成逼真自由视点虚拟形象的三维高斯泼溅（*3DGS*）。给定一个多视角序列，我们提取 *FLAME* 网格，求解曲率加权泊松方程，并获得其夸张化形式。然而，直接形变高斯会产生较差的结果，因此我们通过局部仿射变换，将每一帧扭曲至其夸张化的二维表示，以合成伪真值漫画图像。随后，我们设计了一种交替使用真实与合成监督的训练方案，使单一高斯集合能够同时表示自然与夸张虚拟形象。该方案提高了保真度，支持局部编辑，并能连续控制漫画化程度。为实现实时变形，我们提出一种在原始表面与夸张表面之间进行高效插值的方法。我们还对此进行了分析，并证明其与闭式解之间的偏差有界。在定量与定性评估中，我们的方法优于以往工作，能够生成照片级真实感、由几何结构控制的漫画化虚拟形象。项目主页：*https://c4ricaturegs.github.io*

1. 引言

面部夸张肖像化是指在保留身份的同时，夸张突出的面部特征。尽管这种方法有望打造栩栩如生、身临其境的虚拟化身，但如何以可控的方式在三维空间中生成照片级真实感的夸张效果，仍是一项尚未解决的挑战。成功的基于网格的方法采用基于曲率的方法进行几何变形，例如尺度感知泊松框架 [31]。通过传统的以网格为中心的流程（如纹理映射）渲染这些形变后的表面时，结果往往显得不自然 [31]。近年来，三维高斯泼溅（3D高斯泼溅） [18]逐渐成为一种潜在的多视角表示，它通过直接从不同方向拍摄的一组图像中优化高斯基元，实现了先进的实时照片级真实感。

这引出了以下问题。

我们能否将基于曲率的几何保真度与*3D*高斯泼溅相结合，生成照片级真实感的漫画化人像？

为解决这一问题，我们从主体的多视角视频及其提取出的 FLAME 网格 [23]开始。通过求解加权泊松方程，我们得到形变后的漫画化网格。我们将高斯绑定到原始未变形表面上，并按照此前针对面部表情提出的框架对其进行训练 [21]。随后，在推理阶段，我们根据漫画化网格对原始网格及其绑定的高斯进行形变，对它们执行拉伸、剪切和旋转。然而，如果仅将这些形变建模为额外的表情，并且只使用在输入序列上优化的高斯，

就会导致较低的保真度（见图 5），这揭示了一个领域差距：漫画化人像处于自然表情动态分布之外。

为弥合这一差距，并在缺乏真实漫画化训练数据的情况下，我们通过将每个输入帧与由原始网格到其曲率夸张化对应网格的对应关系所诱导的局部仿射变换（LAT）进行形变，合成伪真实值（GT*），从而生成照片级真实感监督（见第 3.2 节）。训练期间，我们随机交替使用真实视图和 GT*视图，使单个高斯集合能够共同建模自然形变与漫画化变形，让高斯既能利用真实图像，又能适应 GT*。为减轻遮挡相关伪影并保护精细结构（例如头发和网格边界），我们应用空间掩码，在 GT*步骤中冻结受影响的高斯（图 7）。这些高斯仅根据真实帧进行更新，使其属性中能够逐渐积累一致的外观。

尽管优化后的模型仅通过两组视图进行训练，但在推理时仍具备更大的灵活性和可控性。首先，它能够泛化到连续范围内的漫画化强度，并通过高效的线性插值控制夸张程度，近似求解加权泊松方程；我们从理论和实证两方面展示了这一特性。此外，这种表示对全局和局部形变都具有鲁棒性，因此能够进行可控的局部编辑，例如夸大鼻子大小，同时保持无关区域不变。

据我们所知，新的 3DGS 可动画表示首次实现了照片级真实感的漫画化渲染，同时在漫画化形变下忠实保留身份。我们将其与当前最先进的动态面部重建模型 [21], 进行比较；在图像保真度、结构一致性和身份保持指标方面，该模型始终取得更高的分数和更好的定性结果。

我们的贡献包括：

• 一种新颖的 3DGS 训练方案，使用通过局部仿射变换生成的 GT*，以表示真实头像和漫画化头像。

• 结合绑定骨骼的 3DGS，采用曲率加权变形，实现身份保持的照片级真实感漫画化人像。

• 支持可变夸张程度并能精细控制面部特征的实时虚拟形象。

2. 相关工作

2.1. 3D 头部虚拟形象表示

神经隐式表示已成为构建高保真 3D 头部虚拟形象的主流方法，能够基于稀疏多视图观测实现逼真的视图合成。

IMAvatar [45] 结合 3D 可变形模型参数、神经混合形状和蒙皮场来控制姿态与表情，从而生成可动画化的头部虚拟形象。ImFace [43]通过将两个变形场应用于有符号距离函数（SDF）模板，实现身份与表情的解耦。ImFace++ [44] 在此基础上采用两阶段细化框架，进一步提升细节保留效果。

NeRF [24]将空间坐标和视角方向映射为辐射度与密度，并通过体积积分渲染图像。针对头部虚拟形象，Wang 等人 [37] 将稀疏视图编码为具有三维结构感知能力的动画代码网格，并通过多层感知机进行细化。Gafni 等人 [9] 将低维可变形面部模型与神经场景表示相结合，从单目视频中生成照片级真实感、可控的虚拟形象。Gao 等人 [11]采用带有低维表情系数的多层体素场，以捕捉网格混合形状之外的元素（例如头发和配饰）。INSTA [47] 通过将动态 NeRF 嵌入曲面表示中来加速其运行，从而根据较短的单目视频生成可动画的虚拟形象；Avatar-MAV [39] 则通过运动感知神经体素网格将外观与运动解耦。

三维高斯泼溅 [18]将三维场景表示为各向异性的高斯基元，并通过可微泼溅进行渲染。在头部虚拟形象领域，Rig3DGS [28] 在规范高斯空间中重建场景，并学习由 3DMM 引导的形变，以实现高效且照片级真实感的动画；HeadGaS [7] 则通过引入可混合高斯扩展了该表示，其属性会根据表情系数进行调整。MeGA [35] 提出了网格与高斯相结合的混合设计，将泼溅与网格几何结构结合，实现高保真渲染和可编辑头部化身。GaussianAvatars [27]采用绑定继承策略，将可形变的三维高斯绑定到参数化面部网格；SurFhead [21] 则以二维高斯面元 [16], 取代三维高斯，并应用雅可比混合蒙皮和极分解，在动态头部重建方面取得了最先进的结果。

2.2. 网格变形与夸张

经典的基于网格的方法利用几何处理实现形变，例如泊松/拉普拉斯编辑及相关的曲率驱动变形 [19, 32, 33, 42]。针对面部，研究者通过几何驱动和数据驱动两类方法探索了基于网格的形变与漫画化，相关技术从早期的参数化人脸模型发展到现代神经形变网络。Blanz和Vetter [2] 的早期工作提出了三维可变形人脸模型（3DMM），将形状和纹理表示为示例人脸的线性组合，从而能够操控身份和表情。在漫画化领域，Brennan [3]开发了一个交互式系统，通过夸大人物特征与平均脸之间的向量差异来生成线描漫画人像——

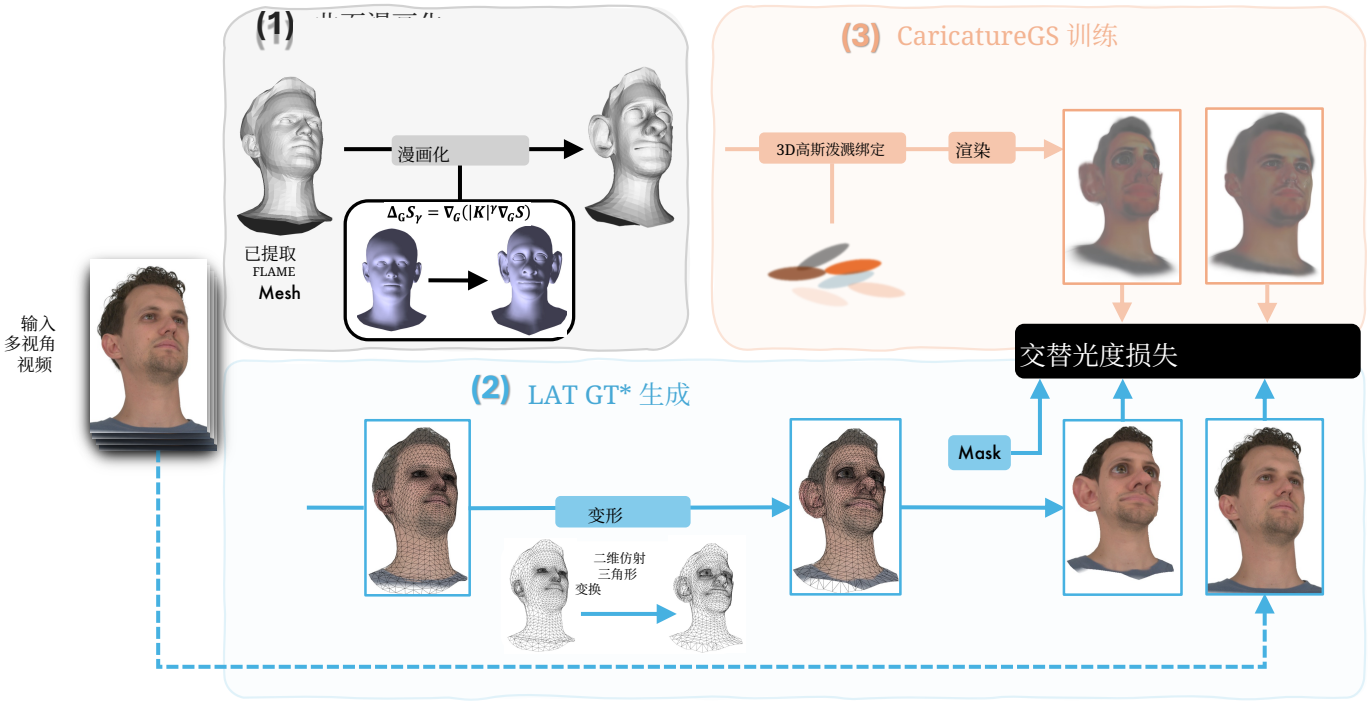


图 2. **CaricatureGS 生成框架。(1)** 我们从某个主体的多视角视频中提取 FLAME 网格，并基于该网格计算曲率驱动的漫画化。结合该主体特有的 FLAME参数，即可得到该主体的漫画化网格。**(2)** 逐三角形的二维仿射变换将中性网格的投影映射至其漫画化后的对应投影，从而对每一帧进行变形，生成伪真值图像对。 **(3)** 各向异性三维高斯基元绑定到原始网格，并通过相应的三维三角形变换映射到漫画化网格。在联合优化过程中，交替渲染中性视图与漫画化视图，并将它们与对应的伪真值进行比较。

接续上段：该系统通过夸大人物特征与平均脸之间的向量差异来生成线描漫画人像。Eigen-satz [8]利用曲率图增强、平滑并迁移特征，同时保留全局结构。后来，Sela等人 [31]提出了一种具有尺度感知能力的基于泊松方程的曲率框架，用于曲面漫画化；该方法在夸大几何特征的同时保持时空连贯性。

数据驱动方法使网格夸张化效果更加丰富，也更加自动化。Wu等人 [38] 从艺术家创作的示例中学习变形模式，在保留身份特征的同时，根据单张二维肖像生成三维漫画化人像。Han等人 [12] 提出了*DeepSketch2Face*，利用卷积神经网络从二维草图推断并细化三维人脸或漫画化网格；其后续工作*Carica-tureShop* [13] 则将逐顶点拉普拉斯缩放与深度学习相结合，根据重建的三维人脸生成照片级真实感的个性化二维漫画化人像。Jung等人 [17] 进一步发展了这一思路，使用多层感知机将潜在编码映射为三维位移，从而实现可控且多样的夸张效果。近期方法则着重于风格适配和更广泛的对应关系。Yan等人 [40] 提出了一个对齐感知的三维人脸变形框架，通过基于控制器的映射实现跨物种对应关系。Olivier等人 [25]探索了从扫描数据到漫画化人像的基于生成对抗网络的风格迁移。Yoon等人 [41] 提出了*LeGO*，这是一种单样本方法，通过微调曲面形变网络来复现目标风格。另一条可应用于面部夸张化的研究路线是生成式方法，例如基于扩散模型和生成对抗网络的3D高斯泼溅编辑器 [6, 22, 36], ，这类编辑器主要作用于外观，同时保持底层几何结构不变。

3. 方法

在此，我们提出一种利用3D高斯泼溅创建可控的、逼真人脸漫画化结果的方法。如图2所示，我们的处理流程始于一段对象的多视角视频，并从中提取拟合FLAME的网格。在第 3.1 节中，我们介绍如何变形几何结构以获得漫画化网格。为监督3D高斯泼溅训练，我们使用二维变形方案生成伪真实漫画图像（GT*）（第 3.2 节）。随后，将高斯基元绑定到中性网格和漫画化网格上，并通过最小化其渲染结果、原始帧与相应GT* 图像之间交替的光度损失进行优化（第 3.3 节）。最后，我们展示了这一共享的高斯集合，尽管仅在这两个图像区域上进行训练，

仍支持通过曲面插值在连续的夸张程度范围内进行实时渲染，并支持区域特定编辑（第 3.4 节）。

3.1. 曲面漫画化

从通过拟合特征点 [34], 获得的、具有时间一致性的 FLAME 网格出发，我们应用曲率驱动变形来夸张面部几何结构。由于该网格在各帧之间保持顶点对应关系一致，这些形变能够保留时间连贯性。为实现这一形变，我们将其表述为曲面上的加权泊松方程。

设 $S \in \mathbb{R}^3$ 为度量张量为 G 、高斯曲率为 $K(p)$的曲面，其中 $p \in S$。对于 $\gamma \in [0, \gamma_f]$，我们定义加权泊松方程

$$\Delta_G S_\gamma \quad = \quad \nabla_G \cdot \big(w(\gamma) \nabla_G S\big). \tag{1}$$

我们采用 [30], 提出的曲率驱动变形模型，其权重由 $w(\gamma) = |K|^\gamma$给出。对于每个 γ，可得到以下一组泊松方程：

$$\Delta_G S_\gamma \quad = \quad \nabla_G \cdot \big(|K|^\gamma \nabla_G S\big). \tag{2}$$

为推导变形后的曲面，我们通过以下最小二乘法求解偏微分方程：

$$\min_{\tilde{x}} \|L\tilde{x} - b\|_A^2. \tag{3}$$

L是离散拉普拉斯–贝尔特拉米算子，定义为 $L = A^{-1}W$， A 是对角面积矩阵， W 是经典的余切权重矩阵，且 $b = \nabla_G \cdot \big(|K|^\gamma \nabla_G(x)\big)$。加权范数定义为 $\|F\|_A^2 = \text{trace}(F^T AF)$。我们将 S_γ 记作方程 2 中加权泊松方程的解。

为适应开放曲面（其中高斯曲率可能在 ∂S 上未定义），或如第 3.4 节所述实现精确的用户控制夸张效果，我们对选定的顶点施加边界条件，即：

$$\min_{\tilde{x} \in \mathbb{R}^n} \|L\tilde{x} - b\|_A^2 \quad \text{s.t.} \quad B\tilde{x} = x^*, \tag{4}$$

其中， $B \in \{0,1\}^{m \times n}$ 选择对应于该顶点集合的行， x^* 是预设的边界位置。对于 y 和 z 坐标，分别独立求解相同的约束系统。

所得网格变形的示例如图 2 的第 **(1)** 部分所示。

3.2. 通过局部仿射变换生成 GT*

通过这些变形后的表面，虚拟化身的几何结构以漫画化形式呈现。为了进行照片级真实感渲染，我们采用网格绑定的 3DGS，详见第 3.3 节。由于未经漫画化优化的 3DGS 效果不佳（第 4.2 节），训练需要真实值监督图像。由于不存在真实的漫画图像，我们生成伪真实值（GT*）：在保留身份的同时确保多视角一致性的照片级真实感漫画图像。

获得此类监督的一种可能方式是单次风格化（例如 Zhou 等人 [46]），通过单张示例图像缩小自然面部与漫画化面部之间的差距。然而，这种方法无法将风格与姿态和身份解耦，通常会将二者一并迁移，而非仅迁移风格（见补充材料）。因此，我们提出一种替代方法：局部仿射变换（LAT），如图 2 的第 **(2)** 部分所示。

LAT 利用了中性网格与变形网格共享的连接关系，这意味着二者具有逐三角形对应关系。考虑对应的三维三角形 $X = \{X_1, X_2, X_3\} \in \mathbb{R}^3$ 和 $Y = \{Y_1, Y_2, Y_3\} \in \mathbb{R}^3$。令 $\pi: \mathbb{R}^3 \to \mathbb{R}^2$表示图像平面投影，其中 $x_i = \pi(X_i)$ 和$y_i = \pi(Y_i) \in \mathbb{R}^2$。假设 $\{x_1, x_2, x_3\}$ 不共线，则存在唯一的仿射映射，

$$\Phi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}, \qquad A \in \mathbb{R}^{2\times 2},\ \mathbf{b} \in \mathbb{R}^2, \tag{5}$$

使得 $\Phi(x) = y$。随后，我们使用这些逐三角形二维仿射变换，将颜色从原始图像映射到变形网格的二维投影上。实际操作中，我们从每个目标像素向原始图像进行逆向变形，并使用双线性插值，以避免产生空白区域。

漫画变形可能会显露中性姿态下先前被自身遮挡的区域，也可能遮挡原本可见的区域，从而使真实图像*中的某些像素缺少有效对应关系。为解决这一问题，我们生成二维三角形级别的遮挡区域掩码。此外，由于散乱发丝位于网格范围之外，无法可靠地进行变形，我们将头发边界也纳入掩码。最终输出为伪真实值（GT*）：高质量漫画化图像，既保留身份，又确保多视角一致性，并为 3D高斯泼溅提供有效监督；同时还提供指示逐像素有效性的掩码（更多详情请参阅附录）。

3.3. CaricatureGS 训练

我们使用三维高斯泼溅框架 [18]，以照片级真实感的方式对虚拟化身的外观进行建模。每个高斯 g_i 都存储局部属性：位置 μ_i、尺度 s_i、旋转 r_i、不透明度 σ_i 和视角相关颜色 c_i。在每个时间帧 $k \in [0, N]$，FLAME 网格 $\mathcal{M} \subset \mathbb{R}^3$由三角形 $\{T_j[k]\}_{j=1}^M$表示，其中 M 是网格面的数量。为确保时空一致性，每个高斯 G_i 都通过绑定索引 b_i与一个特定三角形 T_j 相关联 [27]，从而将其局部属性转换到世界空间。

基于这种绑定式高斯设置，SurFhead [21]采用了二维高斯面元 [16], ，以定向平面高斯圆盘表示表面，并将线性混合蒙皮（LBS）替换为用于高斯形变的雅可比混合蒙皮（JBS），即:

$$\Sigma_i^{1/2} = \mathbf{J_b} r_i s_i, \quad \mu_i' = \mathbf{J_b}\mu_i + T_j^x$$
$$\text{where } \mathbf{J_b} = \exp\left(\sum_{i\in adj} v_i \log(U_i)\right) \cdot \sum_{i\in adj} v_i P_i, \quad (6)$$

其中， v_i 是学习得到的权重， T_j^x 是三角形的重心。 U_i 和 P_i 分别是通过对雅可比梯度 $\mathbf{J}$ 进行极分解得到的旋转和拉伸。极分解将旋转与拉伸分离，从而确保几何上准确的高斯变形（更多细节请参见 [21]）。

我们表明，原本为自然面部表情设计的设置可以通过使用变形后的漫画化网格进行高斯变形，并使用 GT*进行 3DGS 优化，从而适用于漫画化建模。然而，仅在 GT*上训练会引入由遮挡导致的伪影，并将模型限制在单一的表情程度。为克服这些局限性，我们提出了一种联合优化过程，在真实视频帧及其漫画化 GT*对应帧之间随机交替进行监督，同时维护一个共享高斯集合，其绑定机制确保两个监督域中的运动学保持一致。第 3.2 节中引入的掩码会阻止对对应于漫画化 GT*像素的高斯进行监督，因为这些像素无法可靠地进行变形。该联合优化方案使漫画化 3DGS 能够学习超越 GT*的内容，并利用原始帧中的监督同时填补由遮挡造成的孔洞。正如第 5.2 节进一步展示的那样，尽管头发像素被排除在直接 GT*监督之外，该策略仍能有效捕获漫画化头像的头发细节。此外，正如第 3.4 节所述，该方法还支持在推理时生成任意程度的中间漫画化形象，而无需额外捕获。

3.4. CaricatureGS 特性

联合优化不仅利用原始帧中存在但 GT*中缺失的信息，补充了漫画化高斯，还在推理过程中提供了可控性优势。

控制漫画化程度。在目标夸张程度 γ_f下进行联合训练后，我们通过实验观察到，单一绑定的高斯集合具有出色的泛化能力，能够无缝渲染由针对任意 $\gamma \in [0, \gamma_f]$变形的网格生成的头像，而无需额外优化。然而，为每个 γ 获取变形网格都需要求解一个曲率加权泊松问题，这会造成运行时瓶颈——

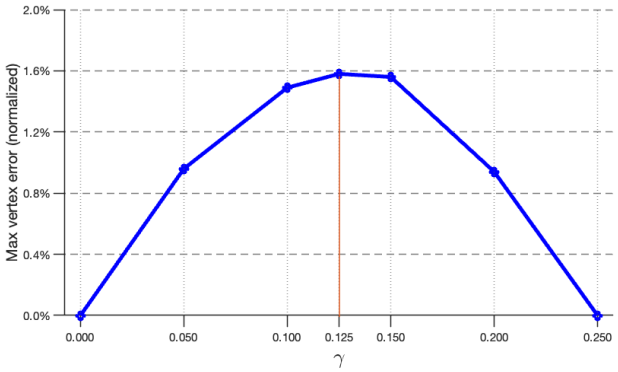


图 3. 误差相对于 γ的参数化趋势。误差经网格包围盒对角线归一化后，从 γ的两端开始增大，并在$\frac{\gamma_f}{2}$ 处达到可忽略的最大值，此时$\gamma_f = 0.25$。

使交互式控制漫画化程度变得不切实际。因此，我们需要一种能够从原始网格 S_0 和预计算的漫画化网格 S_{γ_f}高效生成的表示。我们将这种表示定义为逐顶点混合:

$$S_{\text{blend}}(\gamma) = (1-\alpha)\, S_0 + \alpha\, S_{\gamma_f}, \quad \alpha \equiv \frac{\gamma}{\gamma_f}. \quad (7)$$

我们将近似$S_{\text{blend}}(\gamma)$与精确解 $S(\gamma)$之间的残差定义为

$$\delta S(\gamma) = S_{\text{blend}}(\gamma) - S(\gamma). \quad (8)$$

在补充材料中，我们证明了该残差的 L^2 能量可以结合庞加莱不等式与拉克斯-米尔格拉姆定理进行界定，具体为

$$\|\delta S(\gamma)\|_{L^2} \lesssim \tilde{C}\,\gamma(\gamma_f - \gamma)\,\|\nabla_G S_0\|_{L^2},$$
$$\tilde{C} = C_P\,(\ln|K|)^2\, e^{\max\{0, \gamma_f \ln|K|\}}, \quad (9)$$

其中 C_P 为常数。

该界限在端点 $\gamma = 0$、γ_f处为零，这意味着不存在误差，正如式（7）所预期的那样；在 $\gamma = \frac{\gamma_f}{2}$ 附近达到最大值，而在实践中误差仍然很小。经验上，我们在不同的 γ 和不同对象上评估 S_{blend}（γ）与 S_γ 之间的最大变形误差，并以网格包围盒对角线进行归一化。如图 3 所示，最坏情况下的偏差微乎其微，这证明了插值的保真度，并确认它如预测所示，接近夸张效果的理论中点。这意味着，采用这种近似，在推断新的 γ 值时无需额外求解泊松方程，从而实现对漫画化程度的全面交互式控制。在图 5 中，我们展示了这种插值方案如何让单个高斯集合在 γ 的整个范围内平滑地表示形状变形。

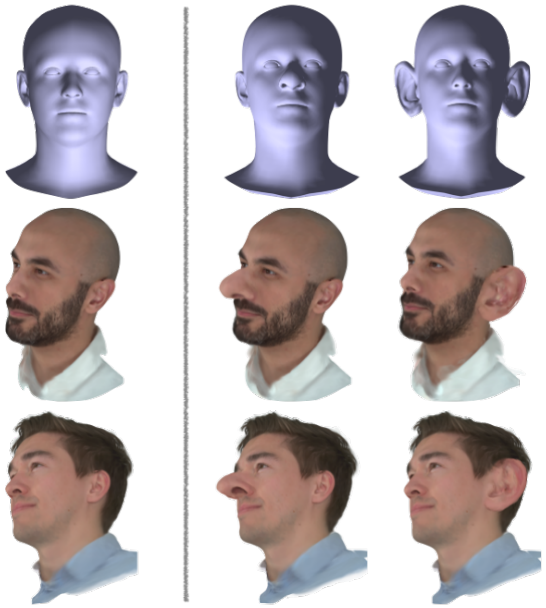


图4. 局部化、语义可控的人脸夸张效果可视化。

局部漫画化控制。 我们的曲率加权模型利用局部曲率 K ，通过求解无约束泊松方程生成整体一致的漫画化效果。为了针对特定区域，我们求解式（4）中的约束最小二乘系统，使得只有选定的感兴趣区域发生曲率形变，从而产生平滑且局部化的夸张效果，并与面部其他部分和谐融合。结合第 3.3 节中的训练方案，与网格绑定的 3D高斯泼溅能够忠实地跟踪这些形变，因此同一高斯集合便可实现语义可控的夸张效果，同时保留身份和整体形状（见图 4）。

4. 实验

我们从两个主要方面评估卡通化虚拟化身：（i）照片级真实感渲染；（ii）身份保持。所有实验均在NeRSemble数据集 [20]上进行，并与近期最先进的4D数字人重建方法 SurFhead [21]进行比较。除非另有说明，我们采用无约束的夸张处理$\gamma_f = 0.25$。

4.1. 数据集

NeRSemble数据集 [20]提供了一个多视角面部表演数据集，该数据集由16台空间排列且同步的高分辨率摄像机捕获。它包含10个脚本序列，其中包括4个情绪驱动（EMO）序列和6个表情驱动（EXP）序列，此外还包含一个额外的自由自我重演序列。为确保公平比较，我们采用与 [21] 相同的训练／验证／测试划分，并进行120,000次训练迭代。更多实现细节请参见补充材料。

方法	CLIP-I ↑	CLIP-D ↑	CLIP-C ↑	DINO ↑	SD ↑
SurFhead	0.67	0.0006	0.944	0.757	0.460
Ours	0.73	0.014	0.945	0.888	0.539

表1。漫画化头像的定量比较。对于所有报告的指标，数值越高越好。

4.2. 基线

据我们所知，目前尚无明确的方法能够根据输入多视角视频构建动态3D照片级真实感模型。为此，我们使用作者的官方实现，将SurFhead [21] 作为对比方法。SurFhead在头部重建和重演方面达到了最先进的性能，并且原则上能够通过JBS处理网格变形，因此是最适合作为比较对象的基线。我们在原始输入序列上训练SurFhead，并在推理阶段使用 γ_f夸张底层网格，具体方法详见第2.2节，从而驱动高斯表示一个漫画化数字人。

4.3. 指标

由于漫画化模型本身具有欠约束性，且缺乏真实图像，对其进行定量评估本质上颇具挑战。我们使用以下指标进行评估：

• **CLIP-I** （图像–提示相似度） [15]: 渲染图像与 CLIP 空间中文本之间的余弦相似度。

•**CLIP-D** （方向相似度） [10]: 衡量源图像与编辑后图像之间的变化，与源提示词和编辑后提示词之间的变化是否一致。

• **CLIP-C** （空间一致性）：按照 [14], 的方法，我们报告沿新视角轨迹的相邻新视图之间的 CLIP图像对齐度（基于图像嵌入）。

• **DINO**（身份/结构一致性）：按照 [46],的方法，我们从渲染结果和对应的原始测试帧中提取 DINO [5]特征，并计算嵌入的余弦相似度。

• **SD**（分数蒸馏）：受 DreamFusion [26],启发，我们将无参考度量定义为：

$$\mathrm{SD} = 1 - \frac{1}{BTN} \sum_{b,t,n}^{B,T,N} \frac{\left\| \epsilon_\theta\left(x_t^{(b,t,n)},\, t\right) - \epsilon_{b,t,n} \right\|_2^2}{\left\| \epsilon_{b,t,n} \right\|_2^2}. \quad (10)$$

其中， $\epsilon_\theta(x_t, t)$是扩散模型 [29] 在时间步 t预测的噪声， ϵ 是真实噪声； B, T, N分别表示图像数量、时间步和种子数。较高的 SD 值表示渲染图像与扩散模型的训练分布更加一致，而该分布旨在近似自然图像分布。

文本提示见附录。这些指标共同评估：(i) 渲染结果对漫画化意图的反映程度（CLIP-I、CLIP-D、SD），

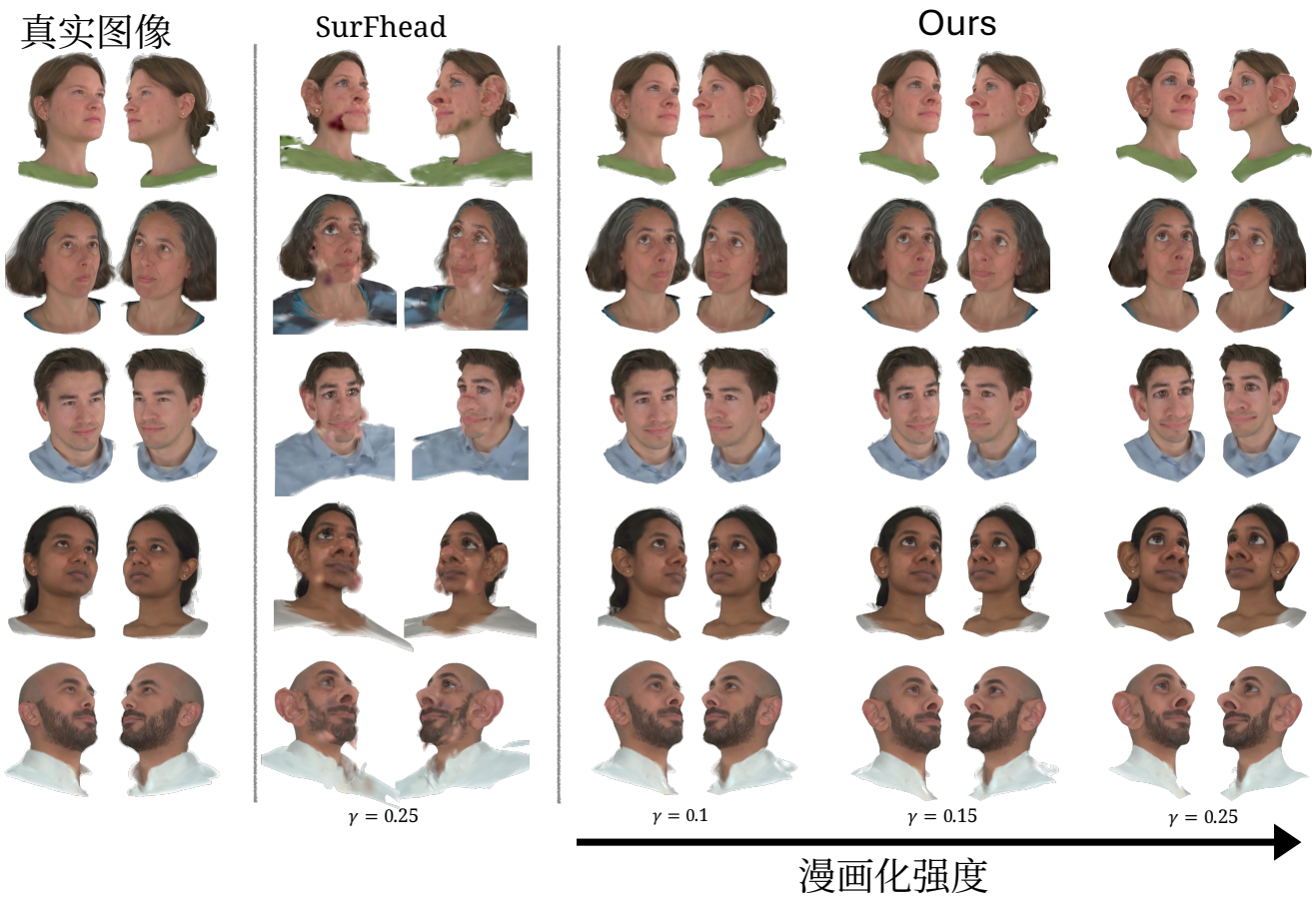


图 5. 我们处理流程的渲染结果 [21]。**SURFHEAD：** 漫画化生成：首先使用最先进的 SURFHEAD 模型 [21], 重建虚拟化身，然后进行网格夸张化。**我们的方法：** 不同漫画化强度下的渲染结果。我们的基于近似的控制沿漫画化强度轴平滑插值，同时保持视觉保真度。

(ii) 身份保持情况，以及夸张效果在多大程度上仍局限于漫画化（DINO、CLIP-D）；以及 (iii) 沿新轨迹生成的视图之间的一致性（CLIP-C）。

4.4. 结果

图 5 并排展示了我们的方法和基线在目标夸张程度 γ_f 下的渲染结果。我们的方法在保持主体身份的同时，生成自然且赏心悦目的夸张效果，并在不同视图间保持一致，同时减少了基线中可见的失真。该图还展示了漫画化程度的可控性：通过将 γ 从 0 调至 γ_f，实现连续控制，并表明第 3.4 节中的近似能够有效支持中间夸张程度。

为进行定量评估，我们使用第 4.3 节中的指标进行了全面比较。如表 1 所示，我们的方法在所有指标上均持续优于基线，表明学习得到的编辑忠实地呈现了预期的漫画化效果，同时保留了身份和视图一致性。

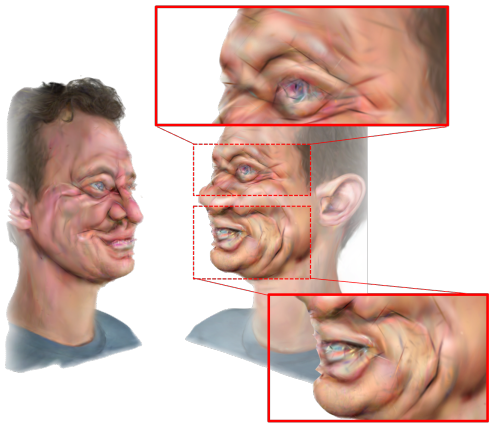
4.5. 基于扩散的编辑

作为另一种基线，我们改造了一种由扩散模型驱动、文本引导且无需网格的 3D高斯泼溅编辑器 [6]，用于漫画化。使用作者提供的实现，我们在主体的多视图图像上，针对每条提示词运行 5,000 步优化，并由 ControlNet-Pix2Pix 提供引导。图 6a 展示了全局编辑，图 6b 则展示了局部编辑，分别对面部和鼻子进行了手动掩码。尽管各个视图中的编辑效果在视觉上看似合理，但显然，与我们的方法不同，该基线存在以下问题：(i) 几何结构漂移；(ii) 不稳定且依赖视图的镜面高光；(iii) 多视图连贯性较差。

5. 消融实验

5.1. 交替训练

在本小节中，我们展示了使用 LAT 生成的 GT*进行训练，对于控制漫画化程度至关重要。正如第 4.4 节所讨论的，仅使用输入图像训练无法实现泛化：使用漫画化网格进行渲染会产生严重失真的输出。在补



(a) 编辑指令：“把他变成一幅写实的漫画肖像。”结果出现了肤色偏移和镜面反射退化。



(b) 编辑指令：“把他的鼻子变大。” 几何结构出现崩坏，且不同视图之间出现颜色不一致。

图 6. GaussianEditor [36] 漫画化尝试。（**a**）全局编辑。（**b**）局部语义编辑。两者都显示出几何结构和外观保真度下降，尤其是在新视图中。

充材料中，我们展示了仅使用 GT*训练同样会失败：中性渲染结果显得不真实，高斯结构也发生了扭曲。这些互补的失败情况凸显了交替使用这两种监督方式对于有效控制漫画化的必要性。

5.2. 掩码

由于 GT*生成的特性，某些精细细节，尤其是头发，通常会在漫画化阶段被错误呈现。为了解决这一问题，我们识别网格中的头发区域，并在 GT*监督迭代期间使用适当的掩码冻结相应的高斯参数，从而防止在渲染漫画化结果时更新这些区域（见第 3.2 节）。图 7 展示了这一效果：左侧对头发区域进行了掩蔽并保持冻结，而右侧则未进行冻结，允许其自由训练，导致头发呈现出不自然的塑料质感。

6. 局限性

尽管我们的方法为逼真的 3D 漫画化提供了强大的框架，但仍存在若干局限性。

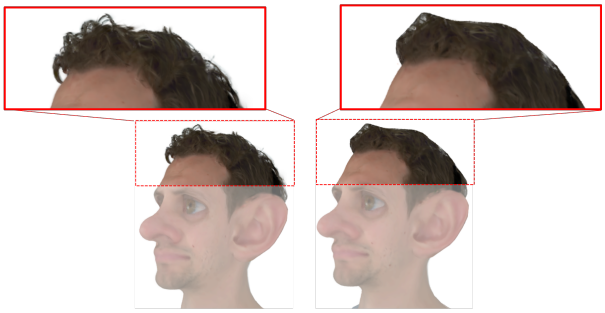


图 7。头发掩码的消融实验。不进行掩蔽时，GT*会在头发区域引入明显的伪影。在 GT*监督期间，对与头发相关的高斯进行掩蔽和冻结，可以有效防止这些伪影。

尽管我们的方法优于基线，但仍存在镜面反射伪影，而在 LAT 中过度拉伸会放大眼睑细微的不准确之处，使其在视觉上更加明显。这种现象也出现在头发上：仅使用输入视角监督训练 Caricature 3DGS 头发（不使用 GT*）可大幅缓解这一问题。不过，在某些情况下，我们观察到头发略有过度平滑。补充材料中提供了这些现象的定性示例。最后，变形后的 FLAME 网格无法完全覆盖所有面部表情。比如，漫画化结果中的眼睑闭合并不完美：在某些表情下本应完全闭合的眼睛往往仍微微张开，导致最终漫画化结果中的眼睑几何结构失真。

7. 讨论

本研究展示了，曲率驱动的几何形变与网格绑定的 3D 高斯泼溅（3DGS）能够结合为一个可控的虚拟形象模型，即使在大幅夸张效果下仍能保持照片级真实感。关键在于一种训练方案：在真实视图与生成的伪真实标签夸张视图之间交替进行监督；后者通过带有可靠性掩码的逐三角形局部仿射变换（LAT）生成。单个高斯集合便能联合学习自然外观与夸张化外观，同时保留身份与表情。先前研究表明，有意的形状夸张可以强化有助于识别的判别性几何线索 [30]。展望未来，我们假设，将可控的夸张效果作为插件式增强集成到人脸识别流程中，或可提高系统对姿态和表情变化的鲁棒性。最后，将我们基于几何的形变与基于扩散模型的编辑器结合，或可实现语义引导编辑；这类编辑既具有照片级真实感，又不局限于仅外观变化，而是能够联合控制形状与外观。

参考文献

[1] Kendall E. Atkinson 和 Weimin Han。理论数值分析：泛函分析框架。Springer，第 3 版，2009年。[2] Volker Blanz 和 Thomas Vetter。《用于合成三维人脸的可变形模型》。[3] Susan E. Brennan。漫画生成器：通过计算机动态夸张面部。*Leonardo*，18 178，1985 3 170( ): – 。出版方：麻省理工学院出版社。[4] Richard L. Burden 和 J. Douglas Faires。数值分析。Brooks/Cole，第 10 版，2010年。[5] Mathieu Caron、Hugo Touvron、Ishan Misra、Hervé Jégou、Julien Mairal、Piotr Bojanowski 和 Armand Joulin。自监督视觉变换器中涌现的特性。载于*IEEE/CVF*国际计算机视觉大会（*ICCV*）论文集，2021年。[6] Yiwen Chen、Zilong Chen、Chi Zhang、Feng Wang、Xiaofeng Yang、Yikai Wang、Zhongang Cai、Lei Yang、Huaping Liu 和 Guosheng Lin。GaussianEditor：利用高斯泼溅实现快速、可控的三维编辑。载于*2024*年*IEEE/CVF*计算机视觉与模式识别大会（*CVPR*），第 21476–21485 页，美国华盛顿州西雅图，2024年。IEEE。[7] Helisa Dhamo、Yinyu Nie、Arthur Moreau、Jifei Song、Richard Shaw、Yiren Zhou 和 Eduardo Pérez-Pellitero。HeadGaS：基于三维高斯泼溅的实时可动画头部虚拟形象。*arXiv* 电子预印本，文章编号 arXiv:2312 02902，2023年。[8]Michael Eigensatz、Robert W. Sumner 和 Mark Pauly。曲率域形状处理。计算机图形论坛，27(2):241–250，2008年。_电子版：https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2008.01121.x。[9] Guy Gafni、Justus Thies、Michael Zollhöfer 和 Matthias Nießner。用于单目 4D 面部虚拟形象重建的动态神经辐射场。*arXiv* 电子预印本，文章编号 arXiv:2012.03065，2020年。[10] Rinon Gal、Or Patashnik、Haggai Maron、Gal Chechik 和 Daniel Cohen-Or。StyleGAN-NADA：由 CLIP 引导的图像生成器区域适配。*arXiv* 预印本 *arXiv:2108.00946*，2021年。[11] Xuan Gao、Chenglai Zhong、Jun Xiang、Yang Hong、Yudong Guo 和 Juyong Zhang。根据单目视频重建个性化语义 NeRF 人脸模型。*arXiv* 电子预印本，文章编号 arXiv:2210.06108，2022年。[12] Xiaoguang Han、Chang Gao 和 Yizhou Yu。DeepSketch2Face：一种基于深度学习的素描系统，用于三维人脸与漫画化建模。*ACM*图形学汇刊，36(4):1–12，2017年。[13] Xiaoguang Han、Kangcheng Hou、Dong Du、Yuda Qiu、Yizhou Yu、Kun Zhou 和 Shuguang Cui。CaricatureShop：个性化、照片级真实感的漫画素描，2018年。arXiv:1807.09064 [cs]. [14] Ayaan Haque、Matthew Tancik、Alexei A. Efros、Aleksander Holynski 和 Angjoo Kanazawa。Instruct-NeRF2NeRF：通过指令编辑三维场景。*arXiv* 电子预印本，文章编号 arXiv:2303.12789，2023年。[15] Jack Hessel、Ari Holtzman、Maxwell Forbes、Ronan Le Bras 和 Yejin Choi。CLIPScore：一种无需参考的图像描述评估指标。*arXiv* 预印本，文章编号 arXiv:2104.08718，2021年。[16] Binbin Huang、Zehao Yu、Anpei Chen、Andreas Geiger 和 Shenghua Gao。面向几何精确辐射场的二维高斯泼溅。载于计算机图形学与交互技术特别兴趣小组会议论文集 ’ *24*， 1 11, 2024 2403 17888 页码 – 。[ ] arXiv：。cs。[17] Yucheol Jung、Wonjong Jang、Soongjin Kim、Jiaolong Yang、Xin Tong 和 Seungyong Lee。通过学习式形状控制生成深度可变形三维漫画。载于计算机图形学与交互技术特别兴趣小组会议论文集，第1–9页，2022年。arXiv:2207.14593 [cs]。[18] Bernhard Kerbl、Georgios Kopanas、Thomas Leimkuehler 和 George Drettakis。用于实时辐射场渲染的三维高斯泼溅。《*ACM*图形学汇刊》，42(4):1–14，2023年。[19] ByungMoon Kim 和 Jarek Rossignac。Geofilter：网格滤波参数的几何选择。计算机图形论坛，24:295–302，2005年。[20] Tobias Kirschstein、Shenhan Qian、Simon Giebenhain、Tim Walter 和 Matthias Nießner。NeRSemble：人头的多视角辐射场重建。《*ACM*图形学汇刊》， 42 4 1 14, 2023 2305 03027 ( ): – 。[ ] arXiv：。cs。[21] Jaeseong Lee、Taewoong Kang、Marcel C. Bühler、Min-Jung Kim、Sungwon Hwang、Junha Hyung、Hyojin Jang 和 Jaegul Choo。SurFhead：用于几何精确二维高斯表面元头部化身的仿射绑定混合，2024年。arXiv:2410.11682 版本：1。[22] Guohao Li、Hongyu Yang、Yifang Men、Di Huang、Weixin Li、Ruijie Yang 和 Yunhong Wang。使用三维高斯生成对抗网络生成可编辑头部虚拟形象，2024年。arXiv:2412.19149 [cs]。[23] Tianye Li、Timo Bolkart、Michael. J. Black、Hao Li 和 Javier Romero。从四维扫描数据中学习面部形状与表情模型。《*ACM*图形学汇刊》（*SIGGRAPH Asia*论文集），36(6):194:1–194:17，2017年。[24]Ben Mildenhall、Pratul P. Srinivasan、Matthew Tancik、Jonathan T. Barron、Ravi Ramamoorthi 和 Ren Ng。NeRF：将场景表示为神经辐射场以进行视图合成。*CoRR*，abs/2003.08934，2020年。[25] Nicolas Olivier、Glenn Kerbiriou、Ferran Argelaguet Sanz、Quentin Avril、Fabien Danieau、Philippe Guillotel、Ludovic Hoyet 和 Franck Multon。自动三维面部漫画化研究：从规则到深度学习。虚拟现实前沿，2:1–15，2022年。[26] Ben Poole、Ajay Jain、Jonathan T. Barron 和 Ben Mildenhall。DreamFusion：使用二维扩散实现文本到三维，2022年。arXiv:2209.14988 [cs]。[27] Shenhan Qian、Tobias Kirschstein、Liam Schoneveld、Davide Davoli、Simon Giebenhain 和 Matthias Nießner。GaussianAvatars：采用绑定三维高斯的照片级头部虚拟形象，2024年。arXiv:2312.02069 [cs]。[28] Alfredo Rivero、ShahRukh Athar、Zhixin Shu 和 Dim-itris Samaras。Rig3DGS：创建可控肖像

来自随意单目视频。*arXiv* 预印本，文章编号 arXiv:2402.03723，2024年。[29]Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。使用潜在扩散模型进行高分辨率图像合成。载于 *IEEE/CVF*计算机视觉与模式识别会议论文集，第10684–10695页，2022年。[30] Matan Sela、Yonathan Aflalo 和 Ron Kimmel。曲面的计算漫画化。计算机视觉与图像理解，141:1–17，2015年。[31] Matan Sela、Yonathan Aflalo 和 Ron Kimmel。曲面的计算漫画化。计算机视觉与图像理解，141:1–17，2015年。[32] Olga Sorkine 和 Marc Alexa。尽可能刚性的曲面建模。载于欧洲图形学*/ACM SIGGRAPH*几何处理研讨会论文集，第109–116页，2007年。[33] Olga Sorkine、Daniel Cohen-Or、Yaron Lipman、Marc Alexa、Christian Rössl 和 Hans-Peter Seidel。拉普拉斯曲面编辑。载于欧洲图形学 / *ACM SIGGRAPH*几何处理研讨会论文集，第179–188页。ACM出版社，2004年。[34] Justus Thies、Michael Zollhöfer、Marc Stamminger、Christian Theobalt 和 Matthias Nießner。Face2Face：RGB视频的实时面部捕捉与表情重演。*arXiv* 预印本，文章编号 arXiv:2007.14808，2020年。[35] Cong Wang、Di Kang、He-Yi Sun、Shen-Han Qian、Zi-Xuan Wang、Linchao Bao 和 Song-Hai Zhang。MeGA：用于高保真渲染与头部编辑的混合网格-高斯头部虚拟形象。*arXiv* 预印本，文章编号 arXiv:2404 19026，2024年。。[36] Junjie Wang、Jiemin Fang、Xiaopeng Zhang、Lingxi Xie 和 Qi Tian。GaussianEditor：使用文本指令精细编辑三维高斯，2024年。arXiv:2311.16037 [cs]。[37] Ziyan Wang、Timur Bagautdinov、Stephen Lombardi、Tomas Simon、Jason Saragih、Jessica Hodgins 和 Michael Zoll-höfer。学习动态人头的组合式辐射场。*arXiv* 预印本，文章编号 arXiv:2012.09955，2020年。[38] Qianyi Wu、Juyong Zhang、Yu-Kun Lai、Jianmin Zheng 和 Jianfei Cai。从二维到三维的动态漫画化，2018年。arXiv:1803.06802 [cs]。[39] Yuelang Xu、Lizhen Wang、Xiaochen Zhao、Hongwen Zhang 和 Yebin Liu。AvatarMAV：使用运动感知神经体素快速进行三维头部虚拟形象重建。*arXiv* 预印本，文章编号 arXiv:2211.13206，2022年。[40] Xirui Yan、Zhenbo Yu、Bingbing Ni 和 Hang Wang。通过对齐感知控制器实现跨物种三维面部变形。人工智能促进会会议论文集，36(3):3018–3026，2022年。[41] Soyeon Yoon、Kwan Yun、Kwanggyoon Seo、Sihun Cha、Jung Eun Yoo 和 Junyong Noh。LeGO：利用曲面变形网络，仅凭单个示例生成可动画化的风格化人脸，2024年。arXiv:2403.15227 [cs] 版本：1。[42] Yizhou Yu、Kun Zhou、Dong Xu、Xiaohan Shi、Hujun Bao、Baining Guo 和 Heung-Yeung Shum。使用基于泊松方程的梯度场操控进行网格编辑。载于*ACM SIGGRAPH 2004*论文集，第644–651页。2004年。

[43] Mingwu Zheng、Hongyu Yang、Di Huang 和 Liming Chen。《ImFace：一种基于隐式神经表示的非线性三维可变形人脸模型》。*arXiv* 电子预印本，文章编号：arXiv:2203.14510，2022年。[44] Mingwu Zheng、Haiyu Zhang、Hongyu Yang、Liming Chen 和 Di Huang。《ImFace++：一种精巧的非线性三维可变形人脸模型，基于隐式神经表-示》。*arXiv* 电子预印本，文章编号：arXiv:2312.04028，2023年。[45] Yufeng Zheng、Victoria Fernández Abrevaya、Marcel C. Bühler、Xu Chen、Michael J. Black 和 Otmar Hilliges。《I M Avatar：基于视频的隐式可变形头部数字人》。*arXiv* 电子预印本，文章编号：arXiv:2112.07471，2021年。[46] Yang Zhou、Zichong Chen 和 Hui Huang。《通过 DINO语义引导实现可变形单次人脸风格化》，2024年。arXiv:2403.00459 [cs].[47] Wojciech Zielonka、Timo Bolkart 和 Justus Thies。《即时体积头部虚拟形象》。*arXiv* 电子预印本，文章编号：arXiv:2211.12499，2022年。

CaricatureGS：利用高斯曲率夸张化三维高斯泼溅面部

补充材料

8. 实现注意事项

除非另有说明，我们按照 SurFhead 的训练实验方案和评估划分 [21]，对每个对象的三维高斯泼溅模型进行 120，000 次迭代优化。所有实验均在单台 NVIDIA RTX 3090（24GB 显存）上运行。每个对象的优化耗时为 ≈ 4小时（这是离线训练时间，并非渲染运行时间。）

我们使用了 NeRSemble 数据集 [34]，其中包含 10 个对象、4 种情绪（EMO）和 6 种表情（EXP）。表情 EXP2 留作测试，摄像机8 则作为训练期间的验证视角。

训练开始时，通过求解无约束泊松方程进行一次漫画化处理，使用 $\gamma = 0.25$ （≈1 分钟）对 FLAME 基础模板进行变形。

由于 FLAME 在不同对象间共用同一模板，因此会保存变形后的曲面，并将其用于所有对象。除非另有说明，我们报告渲染测试序列中 256 帧的指标，并对所有摄像机视角进行汇总。

CLIP 配置。 在文本与图像对齐方面，我们使用以 ViT-B/32 为主干网络的 OpenAI CLIP，并采用其默认预处理流程。

提示词如下：**源文本：** “自然光照下，一张写实的中性面部图像。” **编辑文本：** “自然光照下，一张照片级真实感的漫画化面部图像，鼻子被大幅夸张，耳朵也很大。”

继承默认设置。 除非另有说明，优化器、学习率调度器、球谐函数的阶数以及高斯体的增密与剪枝均沿用 SurFhead [21] 配置。

9. 线性模型与误差分析

符号说明。 设 $S(u, v)$ 为一个参数曲面，其中 $(u, v) \in \mathbb{R}^2$，度量张量为 G， K 表示曲面上各点的高斯曲率 S，并且

$$w(\gamma) \;=\; |K|^{\gamma} \;=\; e^{\gamma L}, \qquad L \;\equiv\; \ln|K|. \quad (11)$$

对于 $\gamma \in [0, \gamma_f]$，记 S_γ 为满足狄利克雷边界条件 x^*（定义于∂S上）的加权泊松问题的解。

为避免在 $K = 0$ 处出现退化，我们使用 ϵ 来稳定其幅度。请注意，为方便起见，我们将其称为 $|K|_\epsilon = \sqrt{K^2 + \epsilon^2}$ 固定 $\epsilon > 0$ 的。为简洁起见，我们用 $|K|$ 表示这些稳定后的量。

1) 带割线权重的泊松方程。原始族定义为

$$\Delta_G S_\gamma \;=\; \nabla_G \cdot \big(w(\gamma)\, \nabla_G S\big). \quad (12)$$

注意， S_0 和 S_{γ_f} 分别指 $\gamma = 0$ 和 $\gamma = \gamma_f$。定义顶点混合，

$$S_{\mathrm{blend}}(\gamma) \;=\; (1-\alpha)\, S_0 \;+\; \alpha\, S_{\gamma_f}, \quad \alpha \;\equiv\; \frac{\gamma}{\gamma_f}. \quad (13)$$

根据 Δ_G 的线性性质和方程 (13)

$$\begin{aligned} \Delta_G S_{\mathrm{blend}}(\gamma) \;&=\; (1-\alpha)\, \Delta_G S_0 \;+\; \alpha\, \Delta_G S_{\gamma_f} \\ &=\; \nabla_G \cdot \Big(w_{\mathrm{sec}}(\gamma)\, \nabla_G S\Big), \end{aligned} \quad (14)$$

其中割线权重 为

$$w_{\mathrm{sec}}(\gamma) = 1 + \frac{\gamma}{\gamma_f}\big(|K|^{\gamma_f} - 1\big). \quad (15)$$

因此， $S_{\mathrm{blend}}(\gamma)$ 在层级 γ处满足精确泊松方程，其中 $w(\gamma)$ 被 $w_{\mathrm{sec}}(\gamma)$ 替代，且 $S_{\mathrm{interp}}|_{\partial S} = x^*$ （ x^*见 (4)）。

2) 余项与性质割线 w_{sec} 是 $[0, \gamma_f]$中 w 的线性插值。根据闭区间上 C^2 函数的经典插值余项定理（例如，[4, 定理 3.1], [1, §3.3]），对于每个 $\gamma \in [0, \gamma_f]$，存在$\xi(\gamma) \in (0, \gamma_f)$，使得

$$w_{\mathrm{sec}}(\gamma) - w(\gamma) \;=\; \frac{w''(\xi)}{2}\, \gamma(\gamma_f - \gamma). \quad (16)$$

由于 $w''(\gamma) = L^2 e^{\gamma L}$，我们得到

$$w_{\mathrm{sec}}(\gamma) - w(\gamma) \;=\; \frac{L^2}{2}\, e^{\xi L}\, \gamma(\gamma_f - \gamma). \quad (17)$$

割线模型在两个端点处都是精确的（其中 $\alpha = 0$和 $\alpha = 1$，从而在 $[0, \gamma_f]$ 中得到一个保留凸性所导致的非负性的解析表达式。

由于 $w'' \geq 0$， $\gamma \mapsto w(\gamma)$ 是凸的，因此 $w_{\mathrm{sec}} - w$ 在 $[0, \gamma_f]$ 上为非负，并在端点处为零。

特别是在 $\gamma = \gamma_f/2$,

$$\Big|w_{\mathrm{sec}}\big(\tfrac{\gamma_f}{2}\big) - w\big(\tfrac{\gamma_f}{2}\big)\Big| \;\leq\; \frac{\gamma_f^2}{8}\, L^2\, \max\big(1, e^{\gamma_f L}\big). \quad (18)$$

这个上界 在 $\gamma_f/2$ 处达到最大值，因为$\gamma(\gamma_f-\gamma)$在此处取最大值。

3）用于残差界的庞加莱不等式与拉克斯–米尔格拉姆定理。

在全文中，我们用割线 $w_{\mathrm{sec}}(\gamma)$ 近似依赖于 γ 的权重 $w(\gamma)=|K|^{\gamma}$，以便采用低成本的顶点混合，而不必针对每个 γ 都求解一个新的泊松问题。为了证明这种替代方案的合理性，我们应当量化权重误差如何传播到几何残差 $\delta S(\gamma)\equiv S(\gamma)-S_{\mathrm{blend}}(\gamma)$。我们的目标是推导 δS 的范数界，该界仅取决于：(i) 区域的椭圆性常数和庞加莱常数，(ii) $\nabla_G S_0$ 的大小，以及 (iii) 附录公式（18）中的标量割线余项。这为该混合提供了一个与网格和度量无关的误差预算。

设定（冻结算子）。设 (S,G) 是一个具有利普希茨边界 ∂S 的紧黎曼曲面。我们施加狄利克雷条件 $u\big|_{\partial S}=0$。

我们固定曲面 S上的微分算子，即相对于度量张量 G的梯度和散度。

令 $V\equiv H_0^1(S)$，并定义

$$\begin{aligned} a(u,v) &= \int_S \langle \nabla_G u, \nabla_G v\rangle_G \, dA_G \\ \|u\|_V &\equiv \|\nabla_G u\|_{L^2(S)}. \end{aligned} \tag{19}$$

我们还将对偶范数定义为

$$\|F\|_{V'} \equiv \sup_{v\in V\setminus\{0\}} \frac{|F(v)|}{\|v\|_V}. \tag{20}$$

利用庞加莱不等式，存在 $C_P>0$，使得对所有 $u\in H_0^1(S)$，

$$\|u\|_{L^2(S)} \le C_P \|\nabla_G u\|_{L^2(S)} = C_P \|u\|_V. \tag{21}$$

因此，$\|u\|_V$ 是 $H_0^1(S)$上的真正范数，并且等价于 $H_0^1(S)$上的标准 H^1-范数。

根据柯西–施瓦茨不等式，

$$\begin{aligned} |a(u,v)| &\le \|u\|_V \|v\|_V \quad \text{(boundedness)}, \\ a(v,v) &= \|v\|_V^2 \qquad \text{(coercivity with } \alpha=1) \end{aligned} \tag{22}$$

其中强制性意味着存在 $\alpha>0$，使得

$$a(v,v) \ge \alpha \|v\|_V^2 \quad \forall\, v\in V.$$

拉克斯–米尔格拉姆。若 a在希尔伯特空间 V上有界且具有强制性，且 $F\in V'$ 有界，则存在唯一解 $u\in V$，满足对所有 $v\in V$，$a(u,\ v)=F(v)$，并且有如下估计

$$\|u\|_V \le \frac{1}{\alpha}\|F\|_{V'} \overset{(22)}{=} \|F\|_{V'}. \tag{23}$$

对于每个 γ，我们求解如下加权泊松偏微分方程：

$$\Delta_G S_\gamma = \nabla_{\mathrm{G}}\big(w(\gamma)\,\nabla_G S\big), \qquad S_\gamma\big|_{\partial S}=x^*. \tag{24}$$

令 $S_{\mathrm{blend}}(\gamma)=(1-\alpha)S_0+\alpha S_{\gamma_f}$，其中 $\alpha=\gamma/\gamma_f$，并定义

$$\begin{aligned} \psi(\gamma) &\equiv w_{\mathrm{sec}}(\gamma)-w(\gamma) \\ \mathcal{R}_\Delta(\gamma) &\equiv \nabla_{\mathrm{G}}\big(\psi\,\nabla_G S\big). \end{aligned} \tag{25}$$

定义 $F\in V'$（弱残差泛函）为

$$\begin{aligned} F(v) &= \langle \mathcal{R}_\Delta, v\rangle \\ &= \int_S \big(\nabla_{\mathrm{G}}(\psi\,\nabla_G S)\big)\, v\, dA_G \\ &= -\int_S \psi\,\langle \nabla_G S, \nabla_G v\rangle_G \, dA_G, \end{aligned} \tag{26}$$

其中 $v\big|_{\partial S}=0$。

利用对偶范数，并根据柯西–施瓦茨不等式和$\|\psi\|_{L^\infty}$界，可直接得到

$$\begin{aligned} |F(v)| &\le \|\psi\|_{L^\infty(S)} \|\nabla_G S\|_{L^2(S)} \|\nabla_G v\|_{L^2(S)} \\ &= \|\psi\|_{L^\infty} \|\nabla_G S\|_{L^2(S)} \|v\|_V, \end{aligned} \tag{27}$$

利用 (20)，我们得到

$$\|F\|_{V'} \le \|\psi\|_{L^\infty} \|\nabla_G S\|_{L^2(S)}. \tag{28}$$

令 $\delta S\equiv S_{\mathrm{blend}}-S_\gamma$。将 S_{blend}和 S_γ 的弱形式相减，得到

$$\begin{aligned} a(\delta S, v) &= a(S_{\mathrm{blend}}, v) - a(S_\gamma, v) \\ &= \int_S w_{\mathrm{sec}}\,\langle \nabla_G S, \nabla_G v\rangle_G \, dA_G \\ &\quad - \int_S w(\gamma)\,\langle \nabla_G S, \nabla_G v\rangle_G \, dA_G \\ &= \int_S \psi\,\langle \nabla_G S, \nabla_G v\rangle_G \, dA_G \\ &= -\int_S \nabla_{\mathrm{G}}\big(\psi\,\nabla_G S\big)\, v\, dA_G \quad (*) \\ &\equiv -F(v). \end{aligned} \tag{29}$$

在 (*) 中，我们对 ∂S 使用分部积分和狄利克雷边界条件。

取 $v=\delta S$ 作为测试函数，并利用强制性和对偶性，

$$\begin{aligned} \|\delta S\|_V^2 &= a(\delta S, \delta S) \\ &= -F(\delta S) \le \|F\|_{V'}\,\|\delta S\|_V \\ \Rightarrow \|\delta S\|_V &\le \|F\|_{V'}. \end{aligned} \tag{30}$$

结合 $\|F\|_{V'}$ 的界可得能量 估计

$$\begin{aligned} \|\delta S\|_V &\le \|\psi\|_{L^\infty(S)} \|\nabla_G S\|_{L^2(S)} \\ \|\delta S\|_V &\le \|w_{\mathrm{sec}}-w\|_{L^\infty} \|\nabla_G S\|_{L^2(S)}. \end{aligned} \tag{31}$$

可选的 L^2 界。根据庞加莱不等式， $H_0^1(S)$，

$$\begin{aligned}\|\delta S\|_{L^2(S)} &\leq C_P \|\delta S\|_V \\ &\leq C_P \|w_{\text{sec}} - w\|_{L^\infty} \|\nabla_G S\|_{L^2(S)}.\end{aligned} \tag{32}$$

总之，割线误差界给出了残差 δS 的能量界：

$$\begin{aligned}\|\delta S(\gamma)\|_{L^2} \lesssim\ & C_P(\ln\|K\|)^2 e^{\max(0,\gamma_f \ln\|K\|)} \\ & \times\gamma(\gamma_f - \gamma)\,\|\nabla_G S\|_{L^2(S)}.\end{aligned} \tag{33}$$

该能量界取决于区域的几何常数（C_P）。式（33）中的曲率取其全局最大值

$$\|K\| = K_\infty = \max_{s\in S} |K(s)| \tag{34}$$

我们注意到 $S_0 = S$（根据定义，对于 $\gamma = 0$，由于没有对 S进行形变），因此式（33）可以用任一术语来表述。

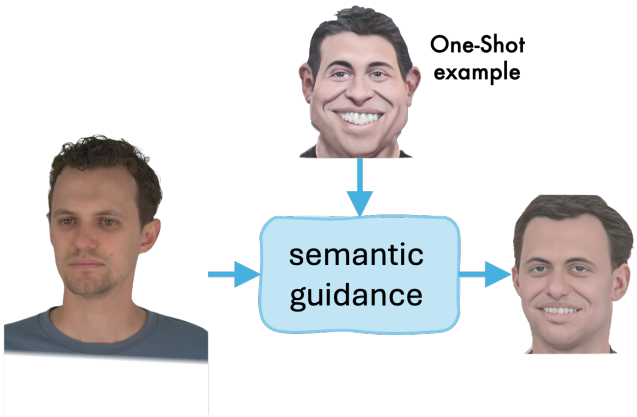
10. 通过单次风格化实现漫画化 GT*

如第3节所述，单次风格化方法（例如可变形 Style GAN [46]）通过对齐 DINO 特征，并将预训练 GAN 适配到单个漫画范例，以弥合自然图像与漫画图像的域差距。给定目标风格图像（图 8a），这些方法可以为任意输入合成风格化输出。实际观察中，我们发现身份与表情之间存在明显的纠缠，这会同时降低身份保真度和表情准确度（图 8）。此外，这些输出在不同视角或表情下并不一致：视角发生变化，或从源图像迁移表情时，该方法会出现结构漂移，并趋向于坍缩到参考风格（图 8b 和 8c），因此不适用于我们的 3DGS 重建场景。

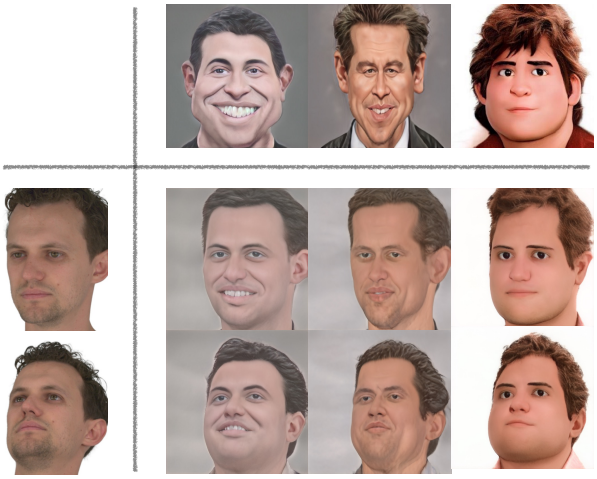
实验方案。 我们使用官方实现进行了 [46] 实验，并采用 Style1、Style2 和 Style3 作为目标风格范例，采用 E MO3、EMO4 作为表情提示。

11. 掩蔽和 GT*

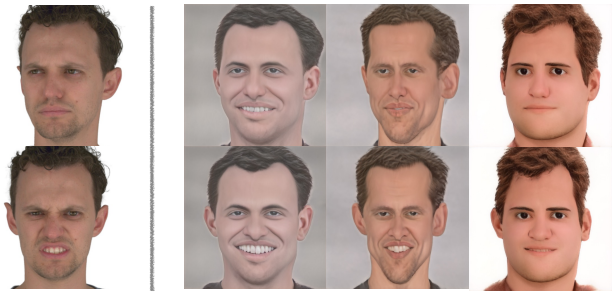
如第 3.2 节所述，GT* 监督是通过将拟合到每个原始帧的 FLAME 网格投影到图像上构建的。因此，GT* 的质量会受到网格与图像错位的影响。实际中，在 γ=0处微不足道的细微拟合误差会随着漫画化强度增加而被放大，在眼睑和眼球等精细几何结构周围尤为明显；见图 9。此外，形变可能会显露原始投影中被遮挡的三角形（例如沿着眼睑褶皱的三角形），从而产生缺乏可靠光度支持的像素。



（a）可变形 StyleGAN [46]：以目标风格范例为条件进行风格化。



(b) 视角变化会导致身份漂移和结构伪影（例如颈部几何结构）。



(c) 表情无法保留，输出会偏向风格示例（例如持续微笑、直视）。

图 8. 漫画化单次风格化的局限性。身份与表情相互纠缠，且缺乏视角/表情一致性，这阻碍了 3DGS监督。

为避免这些失效模式，我们构建了一个具有可见性感知能力的 GT* 掩码。(i) 对于相较于原始投影在非零 γ 处新变为可见的三角形，我们抑制对其进行监督；(ii) 对容易因对齐误差放大而受到影响的解剖脆弱区域（眼睑、耳尖）进行掩码处理。

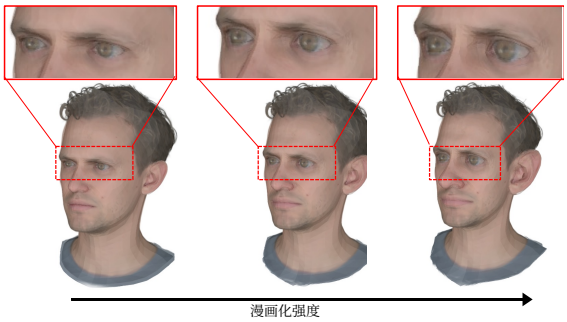


图9. 随着漫画化强度增加时的 FLAME–图像配准误差 γ。投影漂移集中在细小、高曲率的结构上（眼睑/虹膜边缘），并随 γ 增大；若未经滤除便用于监督，将引入错误监督。

这种筛选会在不一致的标签传递到锚定于这些区域的高斯之前将其移除，从而在训练期间产生更干净的梯度，并使外观/几何结构更加稳定。由此得到的 GT* 保留了形变感知监督的优势，同时避免了投影漂移和遮挡引入的伪影。

12. 消融：交替监督

设置。 如第 5.1 节所述，我们希望得到一个单一的 3D高斯泼溅模型，既能渲染原始虚拟化身（γ=0），也能渲染其漫画化版本（γ=γ_f）。我们在相同训练预算下比较三种训练方案：(i) 仅原始数据：仅使用原始帧进行监督。(ii) 真实标注* 仅：仅使用漫画化（真实标注*）帧进行监督。(iii) 交替训练（我们的方法）：交替使用来自两种来源的小批次。我们将目标夸张程度设为 γ_f=0.25，并沿插值路径 $\gamma \in \{0, 0.10, 0.15, 0.20, 0.25\}$进行评估。

结果。 仅原始数据 (i) 能很好地拟合未变形场景，但无法推广到漫画化几何结构（图 10），在非零 γ下会产生明显的失真。相反，仅 GT* (ii) 能表示漫画化虚拟化身，但在 γ=0 处会显著退化。此外，仅 GT* 会在头发及其他超出跟踪网格支撑区域的结构周围产生系统性伪影（例如孔洞或覆盖不足），因为在变形域中，这些像素从未受到直接监督，见图 11。

我们的交替训练方案（iii）在两个端点都保持了高保真度，并在 γ 之间生成平滑的插值（见图 12），避免了方案（ii）中出现的头发/遮挡失效。实际上，交替训练充当了一种简单的多域正则化器，因为它在保留网格外部的外观（来自原始帧）的同时，学习了 GT* 所需的夸张几何结构和视角相关效应。

结论。 交替监督对于获得一个在 γ=0 和 γ=γ_f 处都忠实、并且沿插值路径保持稳定的单一3D高斯泼溅是必要的，而仅在任一区域上训练会导致域特定过拟合和典型的失效模式。



图10. 仅在原始帧上训练



图11. 仅使用 GT*帧进行训练。

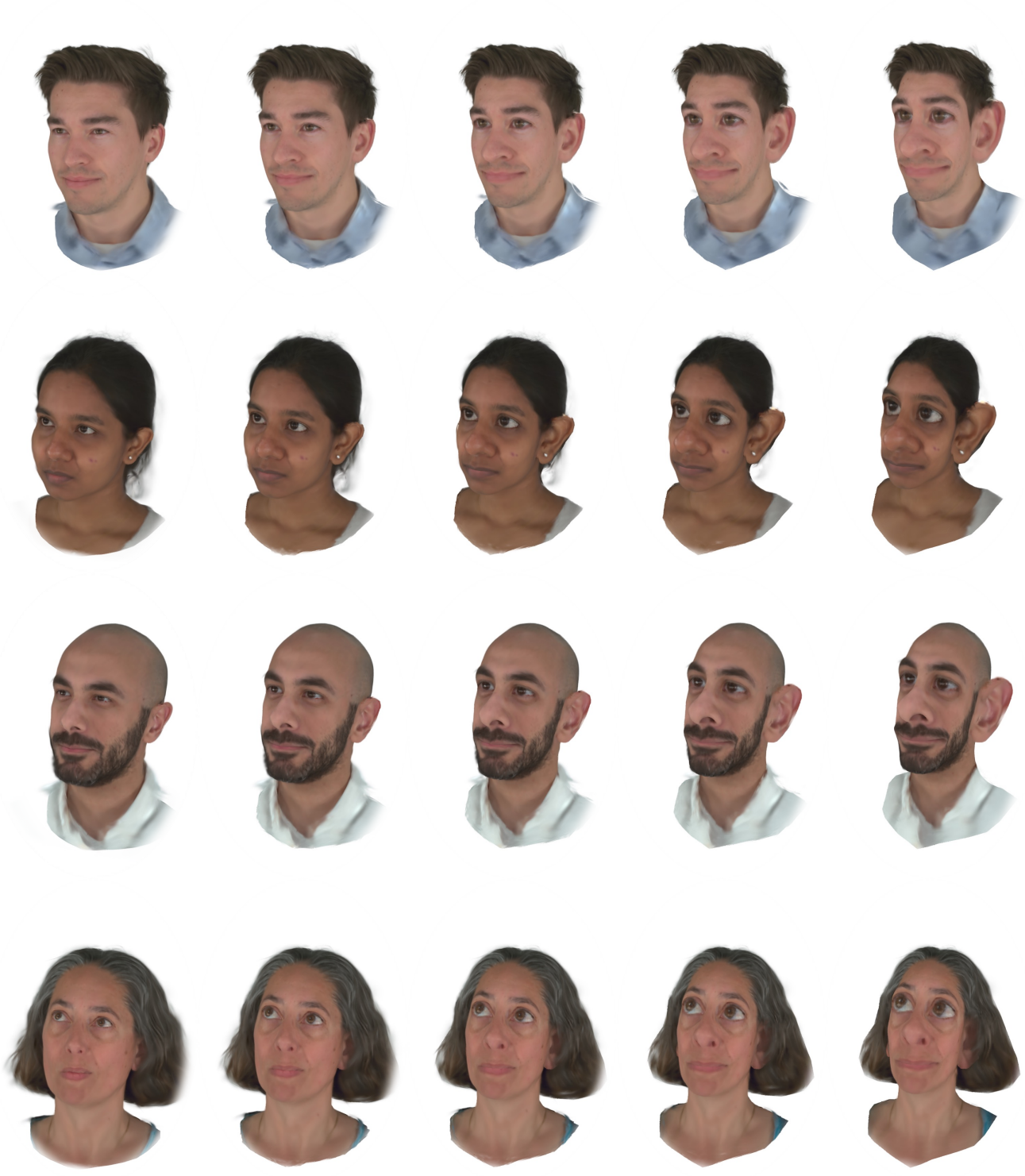


图12。在原始帧和 GT*帧之间交替进行训练