第 xx 卷（200y），第 z 期，第 1–15 页

基于深度学习的局部图块面部重定向

Yeonsoo Choi[2, *], Inyup Lee[1, *], Sihun Cha[1], Seonghyeon Kim[3], Sunjin Jung[1], Junyong Noh[1]

[1]韩国科学技术院视觉媒体实验室，大韩民国[2]
Netmarble F&C，大韩民国[3]Anigma，大韩民国

arXiv:2601.08429v1 [cs.CV] 13 Jan 2026

摘要

在数字动画时代，为虚拟角色制作逼真的面部动画这一追求，推动了各种重定向方法的发展。对于形状相似的模型之间的面部运动重定向，现有方法已取得显著成效；然而，当重定向对象是与人类面部结构差异显著的风格化或夸张三维角色时，便会面临挑战。在这种情况下，必须考虑目标角色的面部结构及其可能的运动范围，以便在重定向后保留原始面部动作所承载的语义。为此，我们提出一种基于局部图块的重定向方法，可将源表演视频中捕捉到的面部动画迁移至目标风格化三维角色。我们的方法由三个模块组成。自动图像块提取模块从源视频帧中提取局部图块。随后，重演模块处理这些图像块，生成相应的目标局部图块。权重估计模块则为每一帧计算目标角色的动画参数，从而生成完整的面部动画序列。大量实验表明，我们的方法能够成功地将源面部表情的语义含义迁移至面部特征比例差异显著的风格化角色。

CCS 概念

•计算方法 → 动画；动作捕捉；动作处理；

1. 引言

自首部三维长篇动画上映以来， [Las95],为三维角色创作在感知上逼真的面部动画，一直是计算机图形学中的重要课题。面部表情涉及深层肌肉细微而复杂的运动。这些细微运动尤其难以在3D模型上复现。因此，人们尝试了多种技术来克服这一难题，创作出极为逼真的面部动画。

关键帧动画一直是制作逼真面部动画的热门选择，但即使对技艺娴熟的动画师而言，也需要付出大量手工劳动。因此，人们提出了基于动作捕捉的方法，作为获取逼真面部运动并将其重现于目标模型上的可行替代方案。由于表演者与目标模型通常存在形状差异，动作捕捉的应用推动了面部动画重定向的发展 [CLK01；DCFN06；BBA*07；SLS*12]。一些方法将这一过程自动化重定向过程，假设人类表演者与目标面部模型的外观彼此相似 [WLVP09；WBLP11；CWLZ13；CHZ14]。由于这一假设不适用于风格化三维角色，因此，将动作重定向到风格化三维角色时，通常需要在预处理步骤中明确指定两个模型对应点之间的变形关系 [NN01；SP04；SCSN11；SLS*12；Sai13；SML16；OTUT19]。

深度学习技术的进步在面部动画重定向方面取得了重大突破，使得能够使用深度神经网络估计三维角色的混合形状或绑定参数 [CPM14；ACC*18；ACF*19；CVW19]。近期研究采用基于图像的重演过程，只需提供人类表演者的 RGB 视频 [KJS*21；MCW*21]，，无需标记点，也无需在源与目标之间进行任何传统的对应配置。这种方法不受特定绑定设置的限制，无论是基于混合形状还是基于关节。这些基于深度学习的方法还假设源与目标的面部形状足够相似，从而能够成功应用基于图像的重演。二者在尺寸或位置上的任何差异

† *这些作者对本研究的贡献相同。

已提交至《计算机图形论坛》（2026年1月）。

源与目标之间的面部元素会导致图像空间中的对齐偏差，从而在重演过程中产生误差。该误差会传递到动画参数计算步骤，导致不合理的重定向结果，如图 7 和图 8所示。为解决这一问题，我们提出了一种基于局部图块的重定向方法，通过在图像空间中对齐各个面部元素，将源人类表演者的面部动画重定向到目标风格化三维角色。

面部动作编码系统（FACS） [EF78] 将面部表情分解为众多动作单元（AUs）。由于动作单元是相互独立的局部模式，组合后可生成面部表情 [DN08],我们将人类表演者的面部和风格化三维角色的面部分解为局部图块。随后，通过源与目标之间语义对应的局部图块，将面部动画从源重定向至目标。基于图块的重定向能够从人类表演者处局部捕捉面部表情的语义含义。通过关注每个面部元素周围观察到的局部表情，我们的方法能够在目标风格化三维角色上忠实地重现源视频中人类表演者的面部表情。

我们的方法由三个模块组成：自动图像块提取模块（APEM）、重演模块（RM）和权重估计模块（WEM）。APEM 是一种新颖的局部图块提取技术，采用基于图块的方法，即使面部元素存在显著的形状差异，也能确保源面部与目标面部之间精确对齐。在训练期间，APEM 从源视频帧和渲染后的目标动画帧中提取局部图块。在推理阶段，它仅从源视频帧中提取局部图块。提取图块后，RM 将每个源图块中的面部表情传递至对应的目标图块，从而生成匹配的重演目标局部图块。随后，WEM 根据重演后的目标局部图块回归目标角色的动画参数。在源视频的每一帧执行这一过程，即可生成目标角色的完整面部动画。

所提出方法的贡献可总结如下：

- 将面部动画重定向到风格化三维角色，这些角色的面部元素在形状和大小上与人类表演者的面部元素显著不同。
- 一种利用自动提取的局部补丁，将面部表情从源表演视频转移到目标风格化三维角色模型的方法。
- 在传递后有效保留源面部表情的语义含义，充分利用目标角色的运动范围。

2. 相关工作

2.1. 面部运动重定向

面部运动重定向是将从源模型捕捉的面部表演传递到目标模型，同时保留面部表情语义含义的过程 [PL06]。其目标是创建具有视觉吸引力且在感知上合理的动画，准确反映源表演所表达的预期表情，并将其呈现到目- 标模型上。实现这一目标的一种方式是采用并行参数化。该技术使两个面部能够共享语义上相同的表情参数空间 [RZL*17]。Noh 和 Neumann [NN01] 提出了一种称为表情克隆的方法，该方法利用径向基函数（RBF）将运动向量从源面部映射到目标面部。Sumner 和 Popović [SP04]提出了一种变形传递方法，将源网格呈现的变形复制到具有不同网格结构的目标网格上。Xu 等人 [XCLT14] 同样采用了变形传递，并且在假设源网格与目标网格具有相同网格结构的前提下，同时支持大尺度和细尺度编辑。随后，Gao 等人 [GYQ*18], 使用基于网格的自编码器结构，以无监督方式解决了变形传递问题。

其他方法进一步探索了基于径向基函数的回归技术，使用手工表情对 [DCFN06；SCSN11] 或手动构建的对应点 [SLS*12；RZL*17]。增强现实框架的最新进展也影响了面部重定向方法。苹果公司的增强现实框架 ARKit [App17], 使 iOS 设备能够在不使用显式标记的情况下捕捉面部表情，仅通过设备内置的摄像机和传感器实时跟踪面部动作。与以往的面部重定向方法不同，我们的方法减少了人工工作量：只需在目标模型上选取少量关键点，使其对应于源视频中表演者自动检测到的面部标志点。这比在源和目标的整个面部上选取全面对应点的任务简单得多。具体而言，不同于需要手动进行数据准备或建立混合形状对应关系的 ARKit，我们的方法提供了一种与骨骼绑定和网格均无关的风格化角色动画重定向方法，可自动学习源表示与目标表示之间的对应关系。

2.2. 基于视频的面部重定向

随着深度学习的发展，面部再现和面部交换通常统称为深度伪造，由于能够生成具有操纵后面部表情的逼真视频，受到了广泛关注。在面部再现中，源视频中的表演会被迁移到另一段视频中 [NKH19；TZS*20；KGT*18]。另一方面，面部交换会将目标视频中的面部替换为源视频中的面部[PGC*21；

NHSW20]。尽管这些任务是在二维视频域中进行的，面部重定向与它们密切相关，因为它涉及将面部表情从源映射到目标。已有多种方法探索面部重定向 [KJS*21；MCW*21] 至三维几何域。这些方法先在二维图像中重演面部，随后回归三维动画参数，以实现角色间重定向 [KJS*21]，或将人类表演者的表演重定向至特定表演者／替代角色 [MCW*21]。借助基于图像的方法，这些方法无需依赖配对数据集或几何先验即可实现重定向过程。近期，已有研究提出了 LivePortrait 等基于关键点的方法 [GZL*24] ，用于生成高质量的二维人像视频。不同于专注于真实人脸的二维方法，我们的方法为三维角色提供明确的动画参数，在忠实呈现细腻表情的同时简化动画过程，无需处理绑定控制或预定义对应关系。我们的工作拓展了这些基于二维图像的技术，使其能够重定向至风格化三维角色，并通过使用局部补丁专注于单个人脸元素。

2.3. 风格化角色的面部重定向

多项研究致力于将面部动画重定向到风格化角色。Aneja 等人 [ACF*16] 为人类角色和风格化角色分别训练了卷积神经网络，并使用感知模型检索二维角色的面部表情图像。此方法进一步扩展为利用神经网络，根据人类图像生成风格化三维角色的表情，并优化表情清晰度 [ACF*19]。与这些需要人工标注图像和表情标签数据集的方法不同，我们的方法不需要成对数据集。Zhang 等人 [ZCZ20] 分别为人类角色和风格化角色训练变分自编码器，从而在二者之间建立共享潜在空间。我们的方法也采用了类似思路，在源模型和目标模型的每个面部元素之间建立潜在空间 [BP14]。

2.4. 基于局部补丁的方法

利用局部图块的方法已被探索用于解决各种面部操控任务。Tena 等人 [TDM11] 提出了一种用于面部重建的局部模型。由于采用了局部变形空间，该模型比整体方法更加灵活。Wu 等人 [WBGB16] 通过约束补丁运动遵循解剖学限制，进一步提升了该方法的鲁棒性。为构建高保真模型，Tran 等人 [TLL19]提出了一种在全局模型和局部模型之间取得平衡的双路径网络。该方法能够根据图像重建精细三维人脸模型。类似地，Huang 等人 [HZLH17] 提出了一种双路径网络，可同时感知全局结构和局部细节，以合成逼真的正面视图。

Chandran 等人 [CBGB20]在标志点检测任务中采用基于局部图块的模型。他们提取面部的语义区域，利用高分辨率图像训练区域网络。与这些基于局部图块的方法类似，我们从面部提取局部图块并将其对齐，以构建由独立训练的网络组成的模型，各网络专注于特定的局部图块。通过关注单个人脸元素，我们的模型提升了与这些面部元素具体相关任务的性能。

一些方法采用局部重定向的概念，以实现与我们工作相似的目标。Liu 等人 [LMC*11] 使用自动生成的分割掩码将面部划分为多个区域，并为每个局部图块优化最佳混合变形权重。生成区域的数量会随阈值而变化，该阈值决定每个区域包含的顶点数量。Chandran 等人 [CCGB22] 先进行逐图块重定向，再将结果应用于解剖模型。只要源模型与目标模型的图块布局相同，这种基于图块的方法就能在网格结构不同的两个几何模型之间进行重定向。我们直接从图像中提取图像块，无需为源对象建立几何模型，也不需要进行复杂设置。

3. 方法

本节将介绍所提出方法的具体细节以及所需的数据输入。图 2(a) 展示了我们的方法在推理阶段的整体流程。该流程始于 APEM，它会自动提取源对象的局部补丁。提取完成后，重演模块（RM）会为目标角色生成对应的重演图像块。随后，这些目标图块由 WEM 处理，生成主成分分析空间（PCA）权重，并将其应用于 PCA 混合形状，从而生成目标三维角色的面部表情。该流程逐帧执行，最终生成完整的面部动画。APEM、RM 和 WEM 的训练细节分别见第 3.2 节、第 3.3 节和第 3.4 节。

3.1. 数据准备

为了训练处理流程中的网络，需要经过预处理的人类表演者视频帧、目标风格化角色三维动画的渲染帧，以及与之对应的动画参数。我们方法的一个关键特点是能够在没有直接监督的情况下进行训练，因此源视频帧与目标角色的动画帧可以不成对。在训练阶段，我们使用源视频和目标角色渲染的三维动画中的图像帧及其对应的二维标志点。此外，还会使用目标角色动画的 PCA 权重。在推理阶段，只需要图像帧

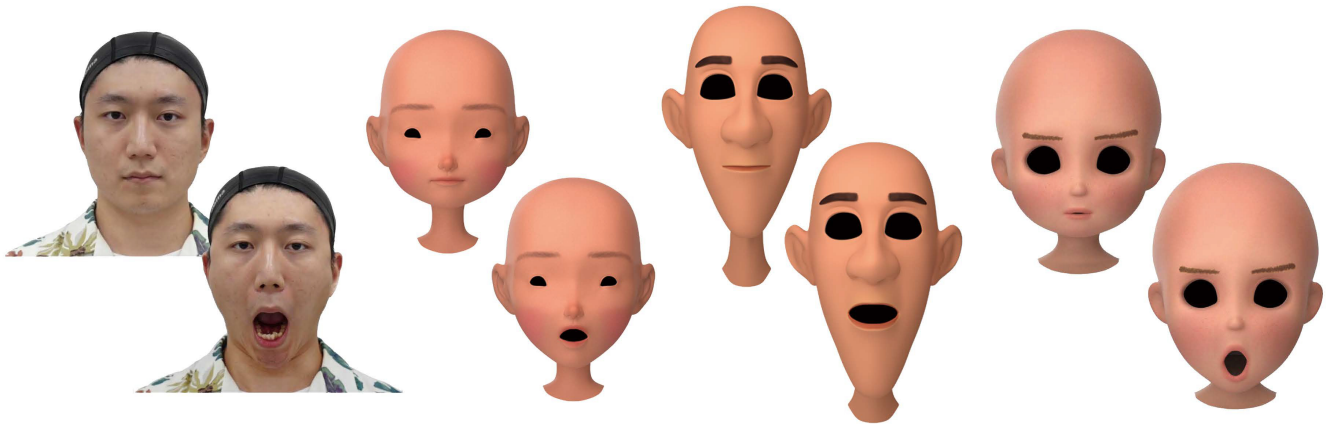

图1： 我们的方法使用局部图块，将源表演视频中的面部动画重定向到风格化目标三维角色上。

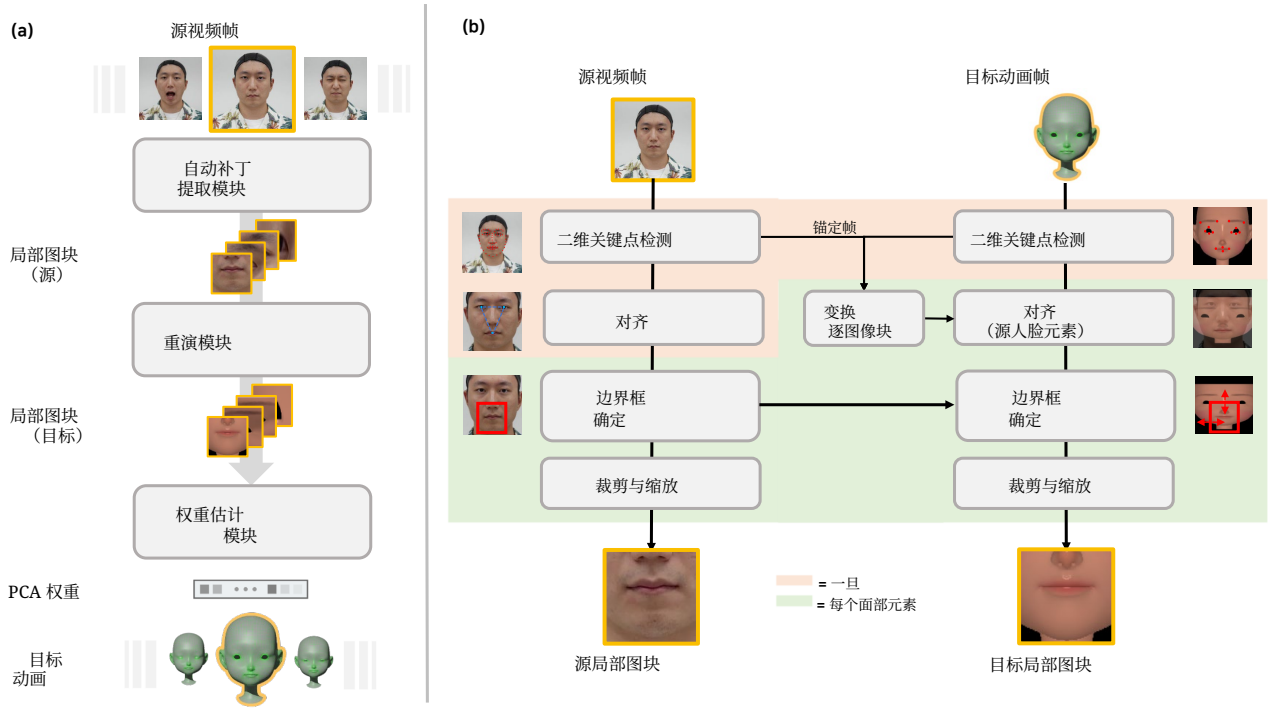

图 2： 所提方法在推理时的概览 *(a)*，以及自动提取图块的示意图 *(b)*，其中以嘴唇周围的局部图块提取为例。

来自源视频即可。接下来的章节将详细介绍这一数据准备过程。

3.1.1. 视频数据

作为源数据，我们拍摄了两位人类表演者（一位女性、一位男性）展示丰富多样面部动作的视频。从中性表情开始，按照面部动作编码系统（FACS）激活的表情依次使用一台经济型摄像机以 60 帧/秒录制，总计 33，618 帧：第一位表演者贡献了 12，896 帧，第二位贡献了 20， 722 帧秒。为进行测试，我们分别从第一位和第二位表演者的数据中保留 2, 122 和 3,482帧，其余帧用于训练。此外，我们还从第二位表演者采集了 1,353 帧，并在不同光照条件下进行拍摄，以评估第 4.6 节所述的光照条件鲁棒性。我们采用 Phong光照模型 [Pho75] 渲染目标动画，并使用正交投影；摄像机的朝向与人类表演视频素材的拍摄角度高度一致。

3.1.2. 3D动画数据

我们的 3D动画数据主要来自两个来源：通过 ARKit 重定向的动画，以及由艺术家手工制作的动画。这些动画经主成分分析分解。在所有实验中，主成分参数向量的维度均为 20，基向量能够捕获总方差的 99.9% 。因此，PCA 基底构成了所捕获数据的线性混合形状系统。对于每个动画帧，我们都会存储 PCA 权重及渲染的局部面片。这些数据用于训练 WEM。关于训练所用 3D动画数据具体内容的更多细节，请参见第 4.1 节。

3.1.3. 图像增强

训练过程中，会对局部图块进行颜色增强和几何增强。对于颜色增强，我们进行颜色抖动，亮度范围为 (0.5, 2 2 2)，饱和度范围为 (0.5,)，对比度范围为 (0.5,)，色调范围为 $(-0.2, 0.2)$。对于几何增强，我们采用 Moser 等人提出的随机仿射变换。 [MCW*21]：旋转范围为 $(0°, 10°)$，平移范围为 (0%, 5%)，缩放范围为 (0%, 5%)。

3.2. 自动图像块提取模块

APEM 在对齐源角色与目标角色的面部元素方面发挥着重要作用，这是在形状和大小差异显著的面部之间传递动作的关键环节。该模块包含四个阶段：二维关键点检测、对齐、边界框确定和裁剪，每个阶段均针对源视频帧和渲染后的目标动画帧进行设计。请注意，训练阶段会同时提取源局部面片和目标局部图块，而推理阶段仅提取源局部面片。

对于源视频帧，我们采用现成的面部标志点检测方法 HRNet [WSC*20], 来估计二维标志点。这些标志点在图2（b）中以红点表示。为对齐每一帧，模块通过计算围绕这三个面部元素的标志点的平均坐标，确定每只眼睛及嘴巴的中心点。该过程仅适用于源阶段；无论用户选择哪些局部图块，都只选择眼睛和嘴巴作为对齐元素。这三个中心点在图中以蓝点表示

2（b）随后用于计算源面部的最终中心点，以便将每一帧对齐到标准位置和朝向。对齐后，首先进行裁剪操作，从视频帧中分离出面部区域，去除不必要的信息。接下来，为每个面部元素定义一个边界框，使其包围当前关注的面部元素（此处为嘴巴）的完整运动范围（ROM），如图2 （b）所示。随后，通过裁剪并将边界框调整尺寸为 128×128 像素，提取局部图像块。

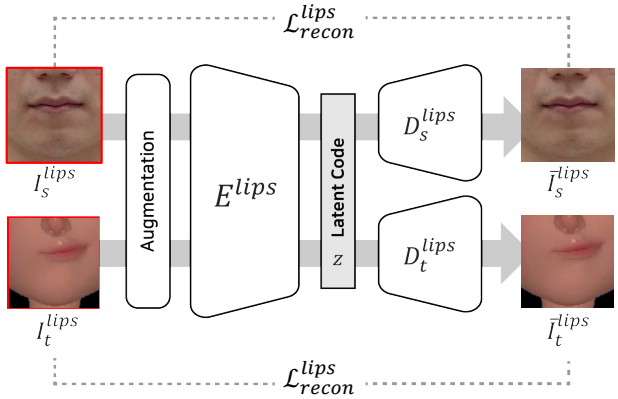

图3： 重演模块（*RM*）中局部图块的自编码器训练过程。（以从嘴唇提取的局部图块为例。）应用图像重建损失 $\mathcal{L}_{recon}^{lips}$ 来训练解码器 D_s^{lips}和 D_t^{lips} ，分别重建源图块 $\bar{I}_s^{lips}$ 和目标图块 $\bar{I}_t^{lips}$ 。

对于目标三维角色，首先提取所有目标帧的二维标志点。为此，用户在目标网格上选择三维顶点，其正交投影将确定二维标志点的位置。接下来，在对齐步骤中，为 N 个面部元素分别计算一个变换矩阵，使目标面部元素形状与源面部元素形状对齐。此计算基于目标锚定帧中每个面部元素的二维标志点（该帧的绑定参数设为零），以及源视频第一帧中对应元素的标志点；源视频的第一帧显示中性表情。随后，使用源的边界框参数初始化对齐后的目标帧。然后根据宽度和高度进一步调整这些参数，使其充分包围目标的运动范围。具体而言，根据目标运动范围与初始化边界框之间的尺寸差异，增大或减小宽度和高度。

逐帧关键点预测结果仅用于根据源数据确定 ROM，而非实现精确的逐帧对齐。因此，关键点预测中的轻微抖动不会影响整体对齐过程，因为主要目的是估计面部动作的最大范围。最后，根据这些元素专属边界框，对对齐后的 N 个目标面部元素分别进行裁剪和调整大小。由此，每个面部元素都得到一个 128x128 大小的局部图块。图 2 (b) 展示了这一过程在每个面部元素上的具体应用，以嘴巴为例。局部图块的选取依据是动画数据中观察到的最显著动作。通常重点关注眼睛和嘴唇周围的区域；如果额头、下巴和鼻子等其他区域也表现出显著运动，同样可以将其纳入考虑。

3.3. 表情重演模块

重演模块（RM）对给定的源局部图块应用图像到图像转换，以生成重演后的目标局部图块。该模块可以包含任意数量的自编码器，数量与所使用的局部图块数量相对应。这些自编码器采用相同的架构，与 Moser 等人提出的架构一致。 [MCW*21]。图 3 展示了其中一个自编码器的训练概览。每个自编码器由一个编码器 E^L 和两个解码器 D_s^L 和 D_t^L 组成，分别用于源和目标。请注意， L是重定向过程中使用的面部元素的局部图块（$L \in$ {左眼、右眼、嘴唇等}）

源 I_s^L 和目标 I_t^L的局部图块通过共享编码器 E^L编码到公共潜在空间 Z 中。随后，将潜代码 $z \in Z$ 输入解码器 D_s^L 和 D_t^L，分别生成 $\bar{I}_s^L$和 $\bar{I}_t^L$。这些自编码器以无监督方式训练，仅需输入数据即可重构源和目标，无需配对的源图像和目标图像数据。网络通过以下两个损失函数进行训练。

L_1 针对源和目标，分别计算真实值与生成图像块之间的逐像素距离。损失函数表示如下：

$$\mathcal{L}_{L_1}^L = \left\| I_s^L - \bar{I}_s^L \right\|_1 + \left\| I_t^L - \bar{I}_t^L \right\|_1, \tag{1}$$

其中， I_s^L 是源局部图像块， I_t^L 是目标局部图像块， $\bar{I}_s^L$ 是为源生成的局部图像块， $\bar{I}_t^L$ 是为目标生成的局部图像块。

此外，还分别计算源和目标的真实值与生成面片之间的结构相似性。损失函数表示如下：

$$\mathcal{L}_{\text{SSIM}}^L = \text{SSIM}\left(I_s^L, \bar{I}_s^L\right) + \text{SSIM}\left(I_t^L, \bar{I}_t^L\right), \tag{2}$$

其中 $SSIM(\cdot)$ 表示 Wang等人提出的结构相似性指数度量 [WBSS04]。

自编码器的总损失函数表示如下：

$$\mathcal{L}_{recon}^L = \mathcal{L}_{L_1}^L + \mathcal{L}_{\text{SSIM}}^L. \tag{3}$$

RM中的自编码器经过联合训练。训练完成后，在推理步骤中使用每个局部图块对应的编码器和目标解码器。

3.4. 权重估计模块

给定重演后的目标局部图块，WEM估计三维面部动画的 PCA 参数。WEM由一个编码器和一个多层感知机构成：编码器将给定的局部图块编码为潜向量，多层感知机则根据编码后的潜向量估计PCA系数的标量权重。对于网络的输入表示，我们尝试了两种设置：1）使用RM预测的图像；2）直接使用RM学习到的潜在空间。如图所示

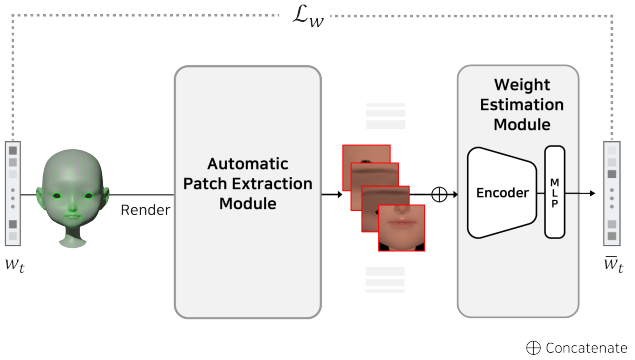

图 4： *WEM* 的训练过程。*WEM* 根据目标角色的局部补丁估计 *PCA* 权重 $\bar{w}_t$，并将估计得到的 $\bar{w}_t$ 与用于构建目标角色动画序列的 w_t 进行比较。 $\mathcal{L}_w$

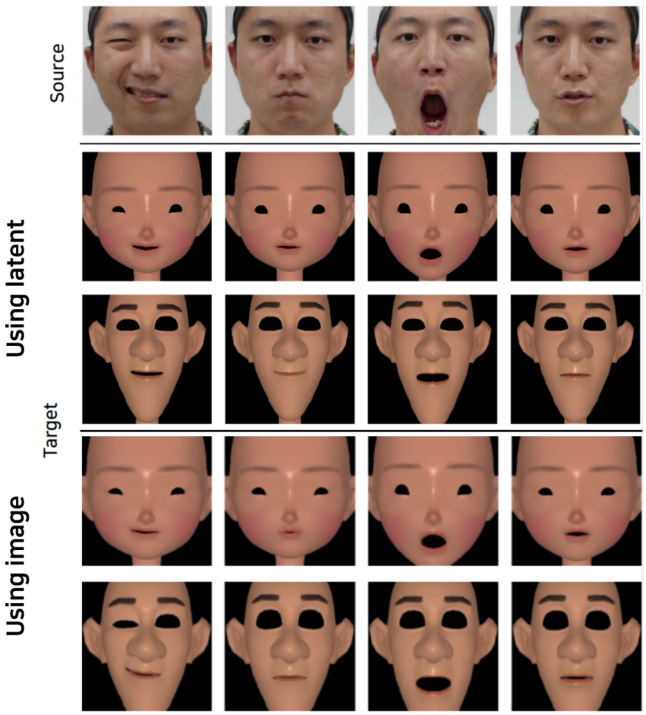

图 5： 比较直接使用潜在空间与使用 *RM* 预测的图像对 *WEM* 进行训练的效果。

图 5显示，使用预测图像生成的重定向表情比直接使用潜在空间生成的表情更加准确。输入局部图像块使 WEM 能够仅根据目标数据估计参数，而使用潜在空间则需要根据编码后的源信息和目标信息估计参数，从而增加了复杂性。利用学习到的潜在表情空间还会延长训练时间，因为 WEM 需要在 RM 完全训练完成后再进行训练，从而引入了顺序依赖关系。

因此，与 Moser 等人不同，我们选择在图像空间中输入局部图像块，而不是直接使用编码后的潜向量。

来自 RM 的局部补丁在馈入编码器之前，沿通道维度进行拼接。鉴于输入已拼接，我们为 WEM 采用单一的全局网络架构，因为它估计的是从整个动画网格序列中导出的 PCA 参数。由于每个 PCA 基都按降序表示显著的全局面部变化，而非语义解耦的表情，因此无法在局部处理 PCA 参数。随后，单一编码器将拼接后的全局输入投影到其学习得到的潜在空间中，以生成维度为 2048 的潜向量，方式与 Kim 等人类似。 [KJS*21]。该向量被输入多层感知机，以生成 PCA 权重。为了训练网络，我们使用一组配对数据：PCA 权重 w_t以及第 I_t^L 节中得到的目标风格化三维角色的渲染局部图块 3.1。训练过程如图 4所示。我们使用 L_1 损失，通过以下公式计算真实权重 w_t 与估计权重 $\bar{w}_t$之间的误差：

$$\mathcal{L}_w = \|w_t - \bar{w}_t\|_1 . \tag{4}$$

3.5. 推理

网络训练完成后，会自动从源视频的每一帧中提取局部图块，并依次通过重演模块（RM）和 WEM 进行处理，从而生成每一帧的 PCA 权重，如图2 (a)所示。在推理步骤中，编码器 E^L 和D_t^L重演模块（RM）中的目标解码器用于表情重演过程。将 WEM 估计出的 PCA 权重与预先计算的 PCA 基相乘，即可生成目标角色的面部动画。

我们的方法估计出的逐帧动画可以无缝导出到 Blender 或 Maya 等主流动画平台，从而便于集成到现有的动画工作流程中。为便于编辑并适配不同的骨骼绑定系统，可通过外部后处理将 PCA 参数转换为其他绑定值。此过程可能需要使用求解器，使估计的动画与混合变形和校正形状生成的可行形状相匹配，如 Moser 等人 [MCW*21]所述。这种适应性提升了该方法在各种动画制作环境中的实用性。

4. 实验

本节将介绍实现细节，并报告所提出的方法与先前方法之间的比较结果。我们还进行了实验，以验证设计选择。

4.1. 实现细节

我们使用了五个不同的风格化三维角色来收集动画数据：Mery（© meryproject.com）、Malcolm（© 2023 AnimSchool）、Child、Girl 和 Piers。该数据集包含使用 ARkit 为 Mery、Malcolm 和 Child 获取的 18，811 帧，以及由艺术家为 Girl 和 Piers 手工制作的另外 16，000 帧。训练重演模块（RM）时，帧率调整为 24 帧/秒，并减少了中性表情帧的数量，以构建表情分布均衡的数据集。Mery、Malcolm、Child 和 Girl 的面部动作主要集中在眼睛、眉毛和嘴巴周围，因此这些模型使用了三个图像块。相比之下，Piers 的运动范围较为独特，尤其是在眼睛和眉毛区域，因此需要使用五个独立图像块来有效捕捉这些细微差异。对于写实角色“Metahuman”（© Metahuman），艺术家使用 Metahuman Creator 创建了该角色，并为其采用了与 Mery 及其他角色相同的一组图像块。

表1： 实验所用三维角色的详情。各列分别展示每个角色的顶点数、用于训练重演模块和 *WEM* 的帧数，以及所用图像块的数量。

Model	Vertices	REM Training Frames	WEM Training Frames	Number of Patches
Mery	5065	5840	18811	3
Malcolm	2228	10542	18811	3
Girl	1402	2900	16000	3
Child	4689	5420	18811	3
Piers	1307	2900	16000	5
Metahuman	29835	3482	3482	3

表2： 训练与推理时间比较 *s.*

方法	训练（m）	推理（ms）
Kim 等人	47	2.69
Moser 等人	3904	6.31
Ours	774	24.11

表1 详细列出了这些规格。RM 经过 50,000 次迭代训练，批量大小为 16，使用 Adam 优化器，学习率为 $5e^{-5}$。WEM 也经过 50,000 次迭代训练，批量大小为 64，采用相同的优化器设置。两个模块均使用分辨率为$128 \times 128 \times 3$的局部补丁进行训练，训练设备为一台面向消费者的图形处理器 NVIDIA GeForce RTX 3090。我们使用 PyTorch 实现该架构。表2 比较了三种不同方法的总训练时间和推理时间。训练时间（分钟）按整个流程估算，包括所有方法的表情重演和权重估计。我们的方法需要 774 分钟训练，耗时较长

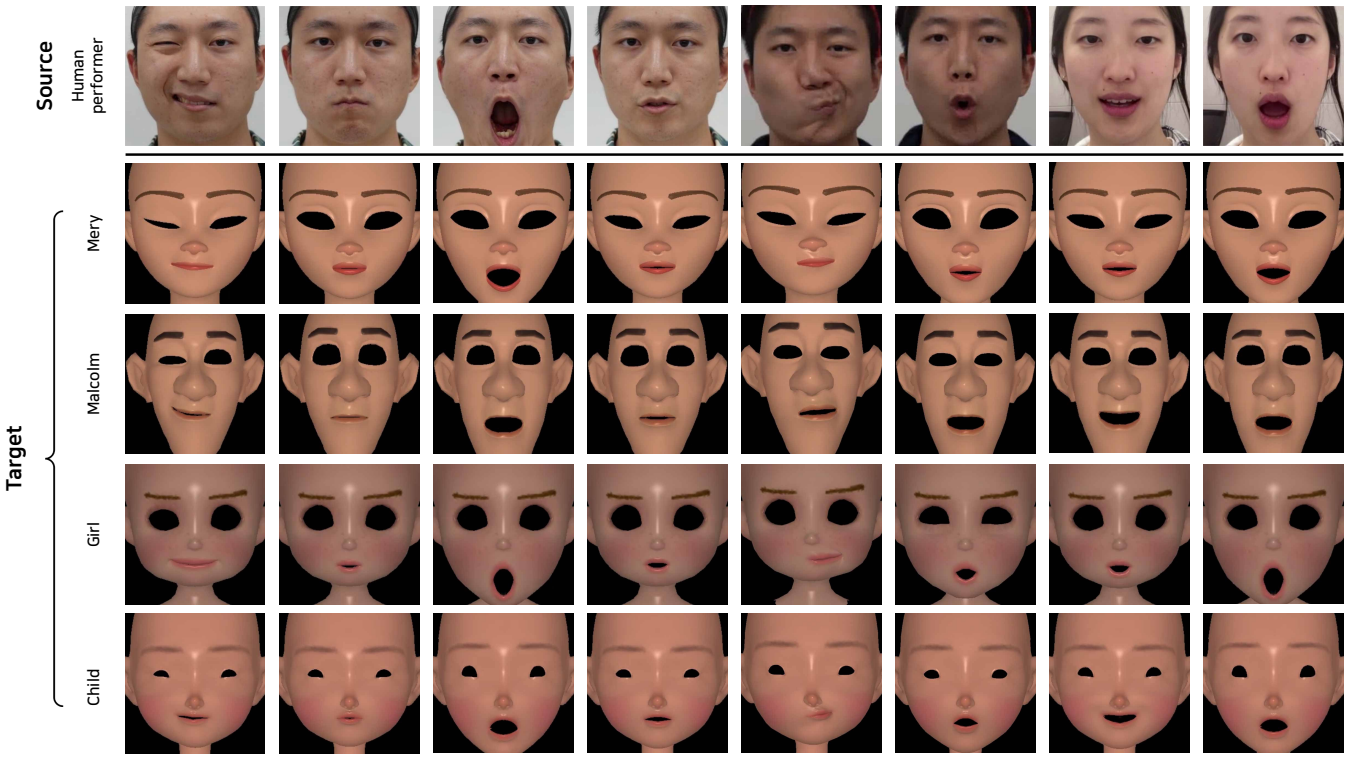

图6： 将人类表演视频重定向到形状和风格各异的多种三维角色模型后的结果。如表 *1* 所示，以上所有角色均使用了三个局部补丁。第*1*至第*6*列展示了男性表演者的结果，第*7*至第*8*列展示了女性表演者使用未见过的表情时的结果。第*5*列和第*6*列中的源视频由同一位男性表演者拍摄，但拍摄环境与训练数据所用环境完全不同。 © 面部绑定：*meryproject.com*，*2023 AnimSchool*

比 Kim 等人（47 分钟）更长，但显著短于 Moser 等人（3904 分钟，因为他们对每个模块都进行了 50,000 次迭代训练）。不过，就推理时间（毫秒）而言，我们的方法略慢于其他方法。尽管如此，我们的方法推理时间为 24.11 毫秒，已接近许多实际应用所需的实时性能（13 毫秒）。

4.2. 重定向结果

视频中捕捉的人类表演重定向到目标风格化三维角色后的结果如图 1 和图6所示。这些结果表明，我们的方法能够有效地将源表演者的表情准确传达至不同的风格化角色模型。值得注意的是，即使眼睛和嘴巴等面部特征存在显著的比例差异，这些表情仍能在目标模型上得到忠实重现。例如，如第二行和第三行所示，Malcolm 和 Girl 的面部与表演者相比，分别更加纤细和宽阔，但不同的嘴型仍得到了很好的重现。在第六列中，Mery、Girl 和 Child 上准确重现的噘起的嘴唇体现了我们方法的精确性。关于我们结果的更全面了解，请参阅补充视频。

4.3. 比较

本节中，我们将所提出的方法与两种先前方法进行比较：Kim 等人 [KJS*21] 和 Moser 等人 [MCW*21]，如图 7、8、9、10 及补充视频所示。我们使用了 Kim 等人提供的原始实现 [KJS*21]。由于无法获得 Moser 等人 [MCW*21] 的实现，我们根据原论文中的描述忠实地重新实现了该方法。各方法的所有超参数均按照相应论文中的说明进行设置。

首先，我们在与训练时相同的光照设置下进行了比较，如图7所示。图7 展示了若干结果示例，表明该方法能够在眼睛远大于表演者眼睛的角色（如 Mery 和 Girl）上成功重现面部动作。与先前方法相比，我们的方法生成的结果在语义上更接近源表情，例如亲吻表情（第三列）和眨眼表情（第六列）。

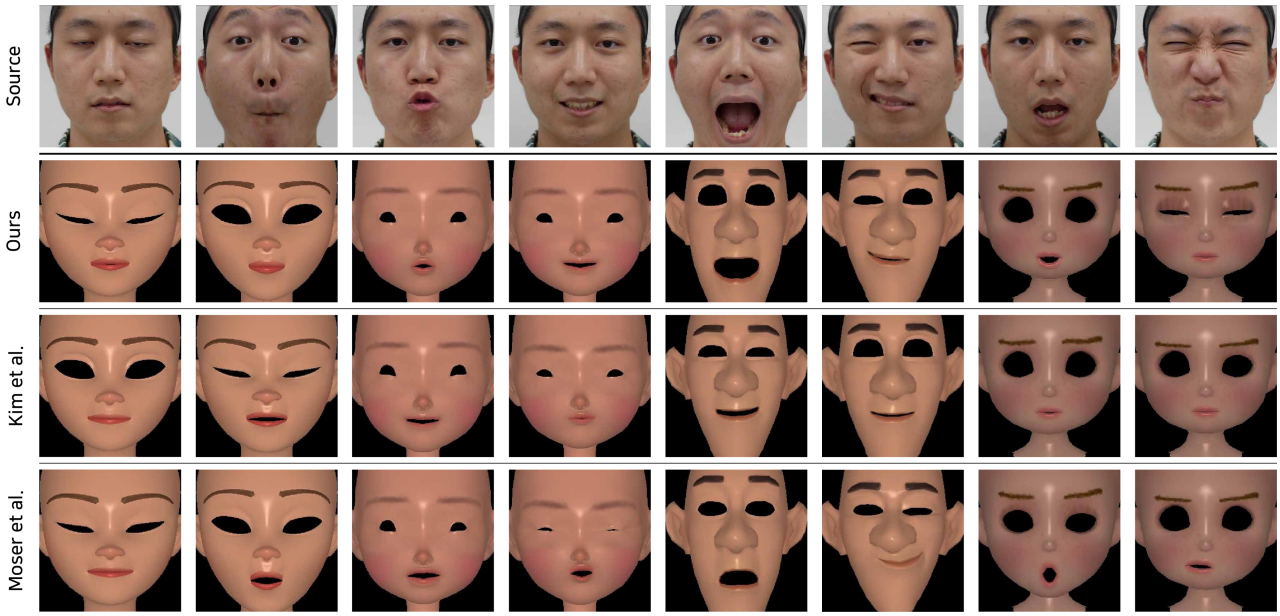

图7： 在与训练所用光照条件相同的情况下拍摄的视频中，将重定向结果与先前方法进行比较。每个角色（*Mery*、*Child*、*Malcolm*、*Girl*）的结果从左到右分布在两列中。© 面部绑定：*meryproject.com*，*2023 AnimSchool*

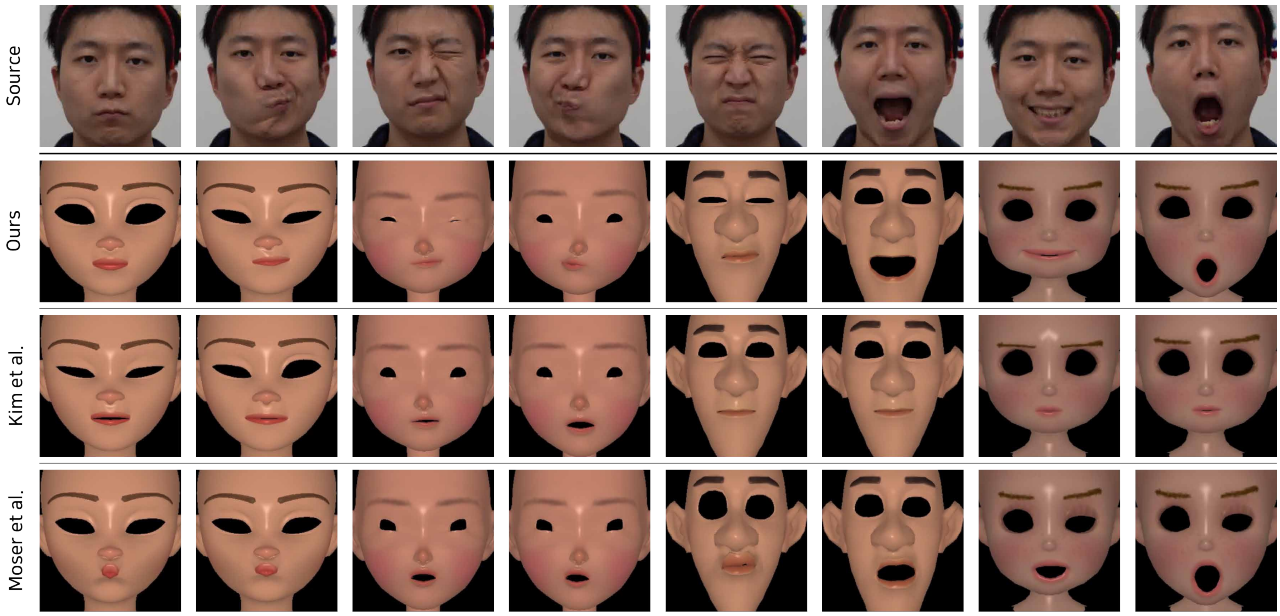

图8： 在与训练所用光照条件不同的情况下拍摄的视频中，将重定向结果与先前方法进行比较。每个角色（*Mery*、*Child*、*Malcolm*、*Girl*）的结果从左到右分布在两列中。© 面部绑定：*meryproject.com*，*2023 AnimSchool*

为了评估我们的方法在实际场景中的鲁棒性——这些场景中源表演者进行面部捕捉时的光照条件会发生变化——我们在不同于训练期间所用光照设置的条件下进行了额外比较。尽管模型仅在单一光照设置下接受训练，但仍表现出卓越的适应能力对光照变化的适应性。该适应性在图8中得到了明显体现：在较暗的光照条件下，我们的模型始终能够生成高质量结果，在性能上超越其他方法。这展示了我们方法的多功能性。

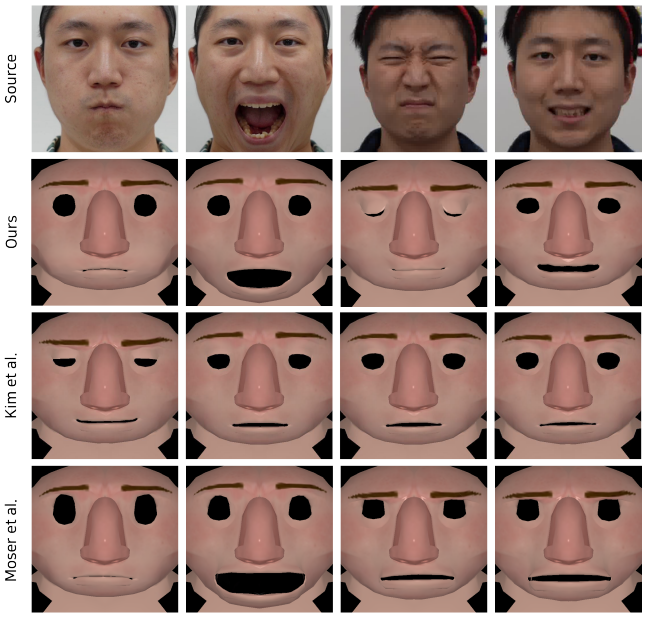

图9： 比较了使用五个图像块的 *Piers* 重定向结果与先前方法的结果。视频是在与训练时相同的光照条件下（前两列）或不同的光照条件下（后两列）拍摄的。

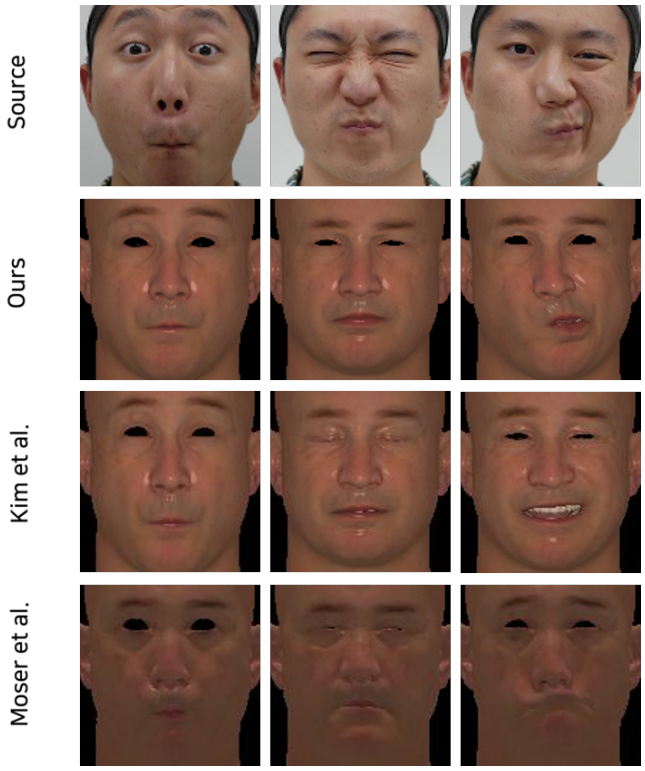

图10： 在逼真人类角色上的重定向结果比较。© *Metahuman*

图 9 展示了我们在“Piers”模型上的重定向结果。该模型使用了五个图像块，同时还列出了先前方法的结果。这表明，我们的方法能够将复杂表情（如鼓腮、张嘴、眯眼和微笑）重定向到面部形状和大小明显不同的模型上。显然，先前方法被设计为处理整个面部，因此在应对由面部形状和大小的巨大差异所带来的复杂重定向挑战时表现不佳。图 10展示了逼真人类角色的对比结果。结果表明，在将复杂且多样的面部表情（如噘嘴、皱眉或将嘴巴歪向一侧）重定向到逼真人类角色时，我们的方法优于其他方法。

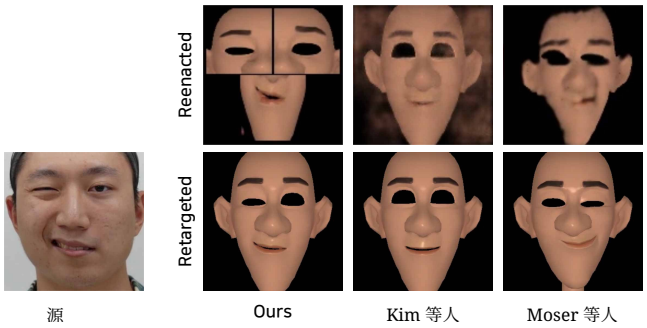

图11： 表情重演结果比较。与我们的方法不同，先前方法中不准确的表情重演导致了错误的重定向结果。© *Facerigs*：*2023 AnimSchool*

图11展示了使用图6 中采用的第一个表情时，我们的方法和先前方法对 Malcolm 进行表情重演的结果。11图11 中的结果清楚地表明，我们的方法在重定向性能上优于先前方法。由于先前方法无法处理源表演者与目标角色之间较大的结构差异，其表情重演结果已包含严重伪影，生成的面部表情与源表情并不匹配。通过关注局部图块内面部元素的运动，我们的方法能够在面部结构存在差异的情况下，重演源表情。

为了定量评估我们方法的重定向质量

方法中，由于源人类表演者没有真实值动画参数，我们进行了角色间重定向实验。不同于最初的配置，我们使用角色的渲染图像作为源，并将该角色的动画重定向到目标角色。如表 1所示，该实验使用了共享同一 WEM动画的 Mery 和 Malcolm。我们将使用 Mery 的 WEM 训练数据制作的 Mery 渲染图像作为输入，并将 Mery 的面部表情重定向到 Malcolm。随后，将重定向后的动画与

表 3： 角色间重定向结果的定量评估。

方法	顶点误差（$\times 10^{-3}$ 毫米）	
	Mery→Malcolm	Malcolm→Mery
Kim 等人	1.4644	1.1305
Moser 等人	2.6543	6.3122
Ours	**1.4094**	**1.0587**

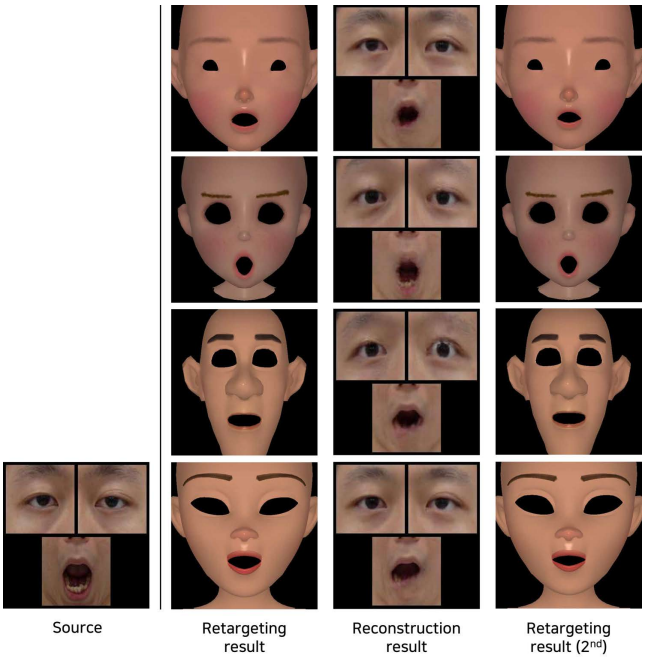

图 12： 循环重定向结果。将源表情重定向到目标角色模型。随后，将重定向后的图像重演给源表演者，再次将其重定向到目标角色。© 面部绑定：*meryproject.com*，*2023 AnimSchool*

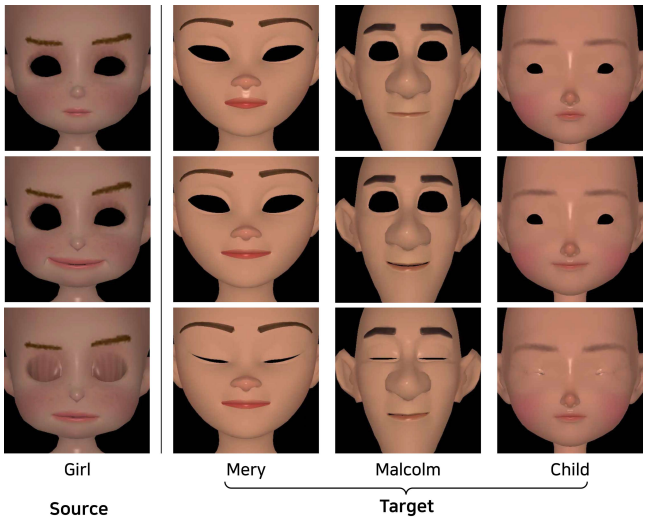

图 13： 角色间重定向结果。© 面部绑定：*meryproject.com*，*2023 AnimSchool*

Malcolm 的 WEM 训练数据作为真实值进行比较，因为两个角色共享相同的动画。我们还进行了这一反向过程，将 Malcolm 的表情重定向到 Mery。我们计算了重定向后目标角色的顶点与相应真实值之间的平均绝对误差（MAE）。如表 3所示，与其他方法相比，我们的方法产生的误差更低 [KJS*21] [MCW*21], ，表明其在准确重定向面部表情方面表现更优。

4.4. 循环重定向结果

为了检验我们的处理流程是否会造成信息损失，我们进行了循环重定向。我们首先将从人类表演者身上捕捉到的面部动画重定向到四个采用三个图像块的角色模型。随后，我们渲染重定向结果，并将其作为输入传给重演模块，以使用源解码器重建原始源人物补丁，

源解码器， D_s^L。接下来，我们再次将重建的图像块作为输入，将面部动画重定向回角色模型。循环重定向的结果如图12 和补充视频所示。理想情况下，重建的源图块应与输入图块相同，第二轮循环的结果也应与初次重定向的结果相同。重建的源图块中的细节（如头发和皱纹）略显模糊，但局部图块中捕捉到的面部元素动作得到了很好的保留，因此循环重定向后的结果仍与原始面部表情十分接近。这一结果验证了我们的方法能够在源与目标之间对面部表情进行语义匹配，并将其解码为所需的局部图块，同时在整个循环过程中不会造成太多信息损失。

4.5. 角色间重定向结果

我们进行了额外的角色间重定向实验，以进一步评估方法的泛化能力。基于图像的方法不仅可以使用人类表演者视频作为输入，也可以使用三维角色动画的渲染结果。我们将三维角色动画的渲染结果作为输入，将一个三维角色的面部表情重定向到另一个角色上。由于我们在渲染后的二维图像空间中执行动作迁移，因此源和目标三维模型的网格结构不必完全相同，从而实现与网格无关的重定向。我们将 Girl 动画的渲染结果设为源视频，并将其重定向到其他角色。结果如图 13 和补充视频所示。Girl 的眼部动作在 Mery 和 Child 身上都得到了恰当的再现。这些结果验证了我们的方法具有足够的通用性，可以

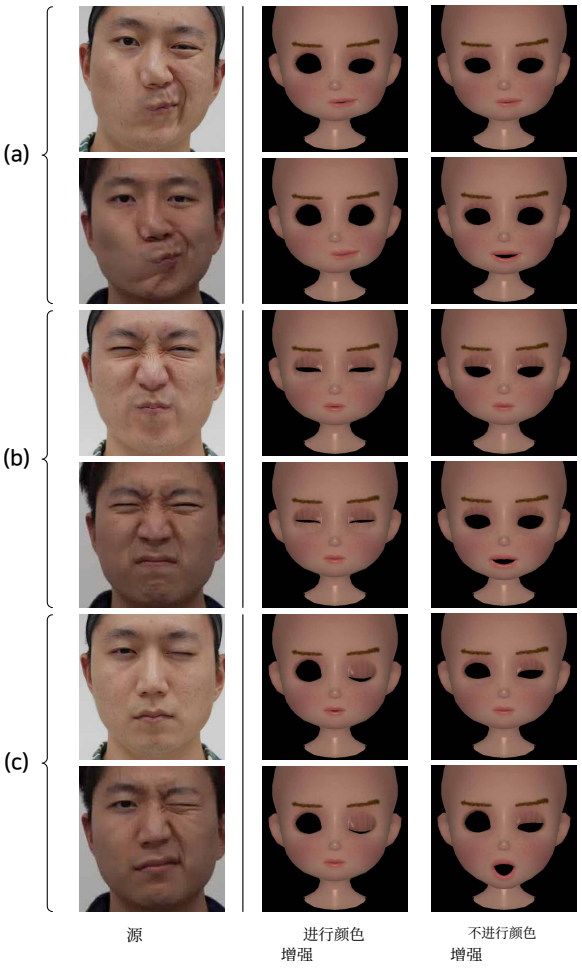

图 14： 对比使用和未使用颜色增强生成的重定向结果。通过颜色增强，在不同光照条件下，源表情在两个相似帧中均得到一致的再现。

使用不同类型的驱动源：视频或三维动画。

4.6. 颜色增强的效果

颜色增强旨在确保模型在色调或光照条件变化时仍能稳健运行。为评估其效果，我们分别使用和不使用颜色增强训练 RM，并采用由同一位表演者拍摄的两类测试视频：一类是在与训练数据相同的光照条件下拍摄的，另一类则是在不同的光照条件下拍摄的。图 14 对比了两段视频中出现的三个相似面部表情。（b）和（c）中观察到的源视频面部表情几乎相同，而（a）中的面部表情在右眼闭合程度上略有差异。经颜色增强训练的 RM 所生成的结果，在不同光照条件下的相似帧中呈现出语义一致的面部表情

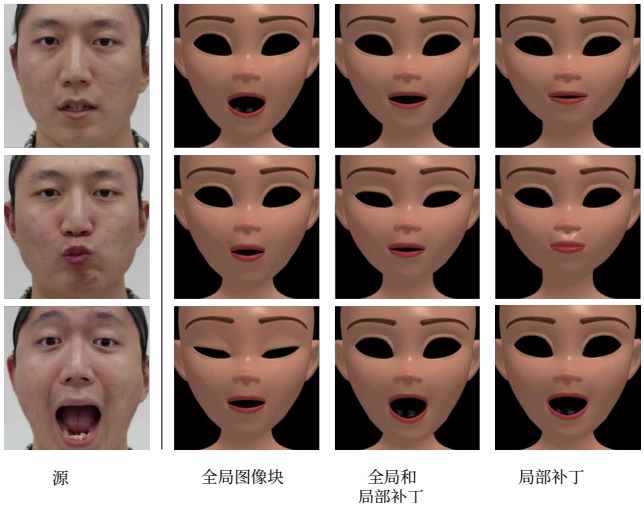

图 15： 使用局部和全局图像块生成的重定向结果。仅使用局部图像块时，效果最佳。 © 面部绑定：*meryproject.com*

。相比之下，未经颜色增强训练的模块在这三种表情中，嘴巴闭合状态均发生了变化。

4.7. 使用局部图块的效果

我们进行了实验，以评估使用局部补丁的效果。实验分为三种情况：使用包含整个面部的全局图像块、同时使用全局图像块和局部补丁，以及仅使用局部补丁。对于全局图像块和局部补丁的设置，我们将所有单独的图像块调整大小至 128x128 分辨率，然后在空间上拼接成一张合成图像。如图 15所示，使用全局图像块得到的结果与先前基于图像的方法完全相同 [KJS*21； MCW*21]。同时使用全局图像块和局部补丁时，由于重演图像块之间存在冲突信息，重定向结果仍不理想。我们发现，最佳方案是仅使用局部补丁，如图 15 所示。

5. 讨论

5.1. 局限性

尽管我们改进了将人类表演者的面部动画重定向到风格化三维角色的效果，但我们也承认该方法存在一些局限性。我们的方法对于具有夸张面部特征的类人角色十分有效，但面对具有独特解剖特征的非人类角色时，会出现某些局限性。例如，象鼻或昆虫触角等特征可能极具挑战性。这些特征超出了我们方法的适用范围，不过我们仍然考虑了源与目标之间大多数的形状和尺寸差异。我们的方法

在非配对数据设置下学习源身份与目标身份之间的共享表征，并使用特定于身份的解码器对这些信息进行解码。为了确保训练成功，使用包含充分运动范围（ROM）的动画数据非常重要。如果数据未能全面表示运动范围，网络可能无法建立适当的共享表征。这些局限性为未来的改进提供了方向，也凸显了严格进行数据收集与准备对于提升我们的面部动画重定向方法性能的重要性。同样，我们的方法本质上依赖于目标角色的绑定设置。皱纹等精细细节只有在目标的绑定设置能够表现这些特征时，才能被传递。要实现这些复杂细节的传递，就需要更加 sophisticated 的绑定设置。

通过分析，如图 16 所示，我们发现，由于基于图像的方法本身存在局限性，完整的闭眼动作未能完全转移。具体而言，当人类表演者的眼睛完全闭合时（如前两行所示），RM 错误地将睫毛对齐并识别为空洞的黑色背景的一部分，而非完全闭合的眼睛。因此，如第三行所示，RM 将黑色睫毛区域解读为眼睛略微睁开，导致最终的 WEM 预测结果中眼睛未能完全闭合。可以通过先进的目标建模来解决这一问题，例如在角色网格中添加睫毛。此外，我们的方法未能正确转移偏航和俯仰方向上的动作，但由于我们通过随机应用平面内仿射变换进行了数据增强，因此成功转移了滚转运动。

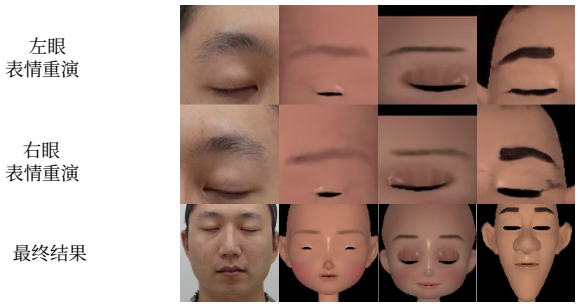

图 16： 角色 *Child*、*Girl* 和 *Malcolm* 的闭眼动作重演过程及最终重定向结果。

5.2. 未来工作

所提出的 APEM 要求用户在目标 3D 角色模型上定义顶点，以获取用于局部图块提取的二维标志点。一个值得探索的未来研究方向是，为各种类型的风格化三维角色开发自动标志点检测技术，以便通过分析给定的面部动画，帮助选出动作幅度较大的面部区域。这项扩展还将开辟一个可能的研究方向：根据目标角色的运动范围，自动识别最佳局部补丁。在基于图像的表情重演方法中，对齐相应的局部补丁至关重要，任何偏差都可能导致输出质量显著下降。因此，另一项可能的未来工作是探索将图像对齐到规范坐标的方法的应用 [PZZ*22；YJNK22；OGKD23]。

我们的方法目前尚未考虑头部姿态变化的重定向，这为未来研究提供了另一个潜在方向。将头部姿态作为附加条件纳入方法中，或可提升其真实感和适用性。尽管所提出的模块在不同光照条件下具有足够的泛化能力，但我们的主要目标是成功地将人类表演者的面部表情迁移到单个风格化角色上。因此，这些模块经过设计，使每个目标角色的每个训练模型仅处理一名表演者。拓展模型能力的一种方式是使用预训练 2D 先验，例如 StyleGAN [KLA19]，，从不同身份中提取神经表情特征。利用这些特征可以显著提升模型在多名表演者之间的泛化能力。

6. 结论

本文提出了一种新颖的基于局部图块的重定向方法，用于将人类表演者的面部动画迁移到风格化三维角色，同时适应面部元素在形状和大小上的差异。我们方法的核心是在表情迁移过程中关注单个面部元素的运动。为此，我们使用 APEM，在训练期间于图像空间中高效提取并对齐源和目标的局部图块，从而即使存在结构差异，也能在推理阶段实现表情重定向。我们的方法包含三个主要模块：前述用于图块提取的 APEM；根据源图块重现目标局部图块的 RM；以及负责从重演后的目标图块计算主成分分析权重的 WEM。这些模块协同工作，使目标三维角色能够生成面部动画，将源视频中的表情逐个图块、逐帧进行迁移。需要指出的是，我们主要关注新颖的“局部”图块提取方法。借助基于图块的架构，即使面部元素存在较大的形状差异，也能确保源面部与目标面部之间实现精确对齐。仅根据用户选定的顶点，源图块与目标图块之间的对齐即可通过新提出的 APEM 过程自动完成。实验结果表明，该方法能够有效地将现场表演者的面部动画迁移到各种风格化三维角色上。值得注意的是，该方法充分利用了目标角色面部运动的完整范围，从而确保动态且真实的动画体验。

参考文献

[ACC*18] Aneja, Deepali、Chaudhuri, Bindita、Colburn, Alex 等。“学习从人类生成三维风格化角色表情”。2018年IEEE冬季计算机视觉应用会议（WACV）。IEEE。2018年，第160–169页 1。[ACF*16] Aneja, Deepali、Colburn, Alex、Faigin, Gary 等。“通过深度学习建模风格化角色表情”。亚洲计算机视觉会议。Springer。2016年，第136–153页 3。[ACF*19] Aneja, Deepali、Colburn, Alex、Faigin, Gary 等。“从人类学习风格化角色表情”。arXiv预印本 arXiv:1911.08591（2019）1, 3。[App17] 苹果公司。ARKit。版本 1.0。2017年。网址：https://developer.apple.com/arkit/ 2。[BBA*07] Bickel, Bernd、Botsch, Mario、Angst, Roland 等。“面部几何与运动的多尺度捕获”。ACM图形学汇刊（TOG）26.3（2007），33–es 1。[BP14] Bouaziz, Sofien 和 Pauly, Mark。“半监督面部动画重定向”。（2014），6 3。[CBGB20] Chandran, Prashanth、Bradley, Derek、Gross, Markus 和 Beeler, Thabo。“面向超高分辨率面部标志点检测的注意力驱动裁剪”。IEEE/CVF计算机视觉与模式识别会议论文集。2020年，第5861–5870页 3。[CCGB22] Chandran, Prashanth、Ciccone, Loïc、Gross, Markus 和 Bradley, Derek。“局部解剖学约束的面部表演重定向”。ACM图形学汇刊（TOG）41.4（2022），第1–14页 3。[CHZ14]Cao, Chen、Hou, Qiming 和 Zhou, Kun。“用于实时面部跟踪与动画的位移动态表情回归”。ACM图形学汇刊（TOG）33.4（2014），第1–10页 1。[CLK01] Choe, Byoungwon、Lee, Hanook 和 Ko, Hyeong-Seok。“表演驱动的基于肌肉的面部动画”。《可视化与计算机动画学报》12.2（2001），第67–79页 1。[CPM14] Costigan, Timothy、Prasad, Mukta 和 McDonnell, Rachel。“使用神经网络进行面部重定向”。第7届国际游戏运动会议论文集。2014年，第31–38页 1。[CVW19] Chaudhuri, Bindita、Vesdapunt, Noranart 和 Wang, Baoyuan。“多张人脸的联合人脸检测与面部运动重定向”。IEEE/CVF计算机视觉与模式识别会议论文集。2019年，第9719–9728页 1。[CWLZ13] Cao, Chen、Weng, Yanlin、Lin, Stephen 和 Zhou, Kun。“用于实时面部动画的三维形状回归”。ACM图形学汇刊（TOG）32.4（2013），第1–10页 1。[DCFN06] Deng, Zhigang、Chiang, Pei-Ying、Fox, Pamela 和 Neumann, Ulrich。“通过交叉映射动作捕捉数据为混合形状面部制作动画”。2006年交互式三维图形与游戏研讨会论文集。2006年，第43页 1, 2。[DN08] Deng, Zhigang 和 Noh, Junyong。“计算机面部动画：综述”。数据驱动的三维面部动画（2008），第1–28页 2。[EF78] Ekman, P. 和 Friesen, W.V.《面部动作编码系统：面部运动测量技术》。帕洛阿尔托，加利福尼亚州：Consulting Psychologists Press，1978年 2。[GYQ*18] Gao, Lin、Yang, Jie、Qiao, Yi-Ling 等。“自动非配对形状变形传递”。ACM图形学汇刊（TOG）37.6（2018），第1–15页 2。

[GZL*24] Guo, Jianzhu、Zhang, Dingyun、Liu, Xiaoqiang 等人。LivePortrait：通过拼接和重定向控制实现高效肖像动画。2024年。arXiv：2407 . 03168 [cs.CV]。网址：https://arxiv.org/abs/2407.03168 3.[HZLH 17] Huang, Rui、Zhang, Shu、Li, Tianyu 和 He, Ran。“超越面部旋转：用于照片级真实感和身份保持正面视图合成的全局与局部感知生成对抗网络”。IEEE 国际计算机视觉会议论文集。2017年，2439–2448 3.[KGT*18] Kim, Hyeongwoo、Garrido, Pablo、Tewari, Ayush 等人。《深度视频肖像》。2018 1805 11714 。arXiv：。[ ] 2 cs.CV。

[KJS*21] Kim, Seonghyeon, Jung, Sunjin, Seo, Kwanggy- oon等。《基于深度学习的无监督人脸重定向》。《计算机图形学论坛》。第40卷，第7期。Wiley 在线图书馆。2021年，45–55页 1，3，7，8，11，12。

[KLA19] Karras, Tero、Laine, Samuli 和 Aila, Timo . 一种用于生成对抗网络的基于风格的生成器架构。2019年。arXiv：1812.04948 [cs.NE] 13。

[Las95] Lasseter, John。《玩具总动员》。电影。由华特迪士尼影业发行。皮克斯动画工作室，1995年1。

[LMC*11] Liu, Ko-Yun, Ma, Wan-Chun, Chang, Chun-Fa等。“一种用于局部重定向和渲染面部表演的框架”。计算机动画与虚拟世界22.2-3（2011），159–167页 3。

[MCW*21] Moser, Lucio、Chien, Chinyu、Williams, Mark 等。“用于制作的半监督视频驱动面部动画迁移”。ACM图形学汇刊（TOG）40.6（2021年），1–18 1, 3, 5–8, 11, 12.[NHSW20] Naruniec, Jacek、Helminger, Leonhard、Schroers, Christopher 和 Weber, Romann M.。“用于视觉特效的高分辨率神经人脸交换”。《计算机图形学论坛》。第39卷。第4期。Wiley Online Library。2020年，173–184 3.[NKH19] Nirkin, Yuval、Keller, Yosi 和 Hassner, Tal。FSGAN：主体无关人脸交换与表情重演。2019年。arXiv：1908.05932 [cs.CV] 2.[NN01] Noh, Jun-yong 和 Neumann, Ulrich。“表情克隆”。第28届计算机图形学与交互技术年会论文集。2001年，277–288 1, 2.[OGKD23] Ofri-Amar, Dolev、Geyer, Michal、Kasten, Yoni 和 Dekel, Tali。“神经凝聚：将图像对齐至联合语义图谱”。arXiv预印本 arXiv:2302.03956（2023年）13.[OTUT19] Onizuka, Hayato、Thomas, Diego、Uchiyama, Hideaki 和 Taniguchi, Rin-ichiro。“基于地标的模板面部表情变形传递，用于自动生成虚拟形象混合变形目标”。IEEE/CVF国际计算机视觉会议研讨会论文集。2019年 1.[PGC*21] Perov, Ivan、Gao, Daiheng、Chervoniy, Nikolay 等。DeepFaceLab：集成式、灵活且可扩展的面部交换框架。2021年。arXiv：2005.05535 [cs.CV] 2.[Pho75] Phong, Bui Tuong。“计算机生成图像的光照”。《美国计算机协会通讯》18.6（1975年），311–317 4.[PL06] Pighin, Frederic 和 Lewis, John P.。“面部运动重定向”。ACM SIGGRAPH 2006课程。2006年，2–es 2.[PZZ*22] Peebles, William、Zhu, Jun-Yan、Zhang, Richard 等。“GAN监督的稠密视觉对齐”。IEEE/CVF计算机视觉与模式识别会议论文集。2022年，13470–13481 13。

[RZL*17] Ribera, Roger Blanco I、Zell, Eduard、Lewis, John P 等。“通过自动运动范围对齐进行面部重定向”。ACM图形学汇刊（TOG）36.4（2017年），1–12 2.[Sai13] Saito, Jun。“平滑的接触感知面部混合变形传递”。数字制作研讨会论文集。2013年，7–12 1.[SCSN11] Song, Jaewon、Choi, Byungkuk、Seol, Yeongho 和 Noh, Junyong。“特征面部重定向”。《计算机动画与虚拟世界》22 194 1, 2 2 3 2011 187 . - ( ), – .[SLS*12] Seol, Yeongho、Lewis, John P、Seo, Jaewoo 等。“混合变形的时空表情克隆”。ACM图形学汇刊（TOG）31.2（2012年），1–12 1, 2.[SML16] Seol, Yeongho、Ma, Wan-Chun 和 Lewis, John P.。“根据表演捕捉创建特定演员的面部绑定”。2016年数字制作研讨会论文集。2016年，13–17 1.[SP04] Sumner, Robert W 和 Popović, Jovan。“三角网格的变形传递”。ACM图形学汇刊（TOG）23.3（2004年），399–405 1, 2.[TDM11] Tena, J Rafael、De la Torre, Fernando 和 Matthews, Iain。“交互式基于区域的线性三维人脸模型”。ACM SIGGRAPH 2011论文集。2011年，1–10 3.[TLL19] Tran, Luan、Liu, Feng 和 Liu, Xiaoming。“迈向高保真非线性三维人脸可变形模型”。IEEE/CVF计算机视觉与模式识别会议论文集。2019年，1126–1135 3.[TZS*20] Thies, Justus、Zollhöfer, Michael、Stamminger, Marc 等。Face2Face：RGB视频的实时人脸捕捉与表情重演。2020年。arXiv：2007.14808 [cs.CV] 2.[WBGB16] Wu, Chenglei、Bradley, Derek、Gross, Markus 和 Beeler, Thabo。“用于单目人脸捕捉的解剖学约束局部变形模型”。ACM图形学汇刊（TOG）35.4（2016年），1–12 3.[WBLP11] Weise, Thibaut、Bouaziz, Sofien、Li, Hao 和 Pauly, Mark。“实时的基于表演的面部动画”。ACM图形学汇刊（TOG）30.4（2011年），1–10 1.[WBSS04] Wang, Zhou、Bovik, Alan C、Sheikh, Hamid R 和 Simoncelli, Eero P.。“图像质量评估：从误差可见性到结构相似性”。IEEE图像处理汇刊13.4（2004年），600–612 6.[WLVP09] Weise, Thibaut、Li, Hao、Van Gool, Luc 和 Pauly, Mark。“Face/off：实时面部木偶操控 ”。2009 年 ACM SIGGRAPH/Eurographics计算机动画研讨会论文集。2009年，7–16 1.[WSC*20] Wang, Jingdong、Sun, Ke、Cheng, Tianheng 等。“用于视觉识别的深度高分辨率表征学习”。IEEE模式分析与机器智能汇刊43.10（2020年），3349–3364 5.[XCLT14] Xu, Feng、Chai, Jinxiang、Liu, Yilong 和 Tong, Xin。“可控高保真面部表演迁移”。ACM图形学汇刊33.4（2014年）。ISSN：0730-0301 2.[YJNK22] Yang, Sejong、Jeon, Subin、Nam, Seonghyeon 和 Kim, Seon J oo。“稠密跨物种面部嵌入”。神经信息处理系统进展35（2022年），33275–33288 13.[ZCZ20] Zhang, Juyong、Chen, Keyu 和 Zheng, Jianmin。“轻松实现从人类到虚拟形象的面部表情重定向”。IEEE可视化与计算机图形学汇刊28.2（2020年），1274–1287 3。

数据可用性声明

支持本研究结果的数据可在以下存储库中公开获取：

- **Mery** - 可从 Mery 项目获取，网址为 https://www.meryproject.com。
- **Malcolm-** 可从 AnimSchool 获取，网址为 https://www.animschool.com。
- **Child** - 此数据集由我们的研究团队内部创建。
- **Girl** - 此数据集从 cgtrader 免费下载，网址为 https://www.cgtrader.com/free-3d-models/character/child/cartoon-girl-rigged-578fe010-4c89-4724-96f3-de378cf64817
- **Piers** - 此数据集从 cgtrader 免费下载，网址为 https://www.cgtrader.com/free-3d-models/character/man/maya-character-rig-piers-3d-rig
- **Metahuman-** 此数据集由我们的研究团队使用 Metahuman Creator 内部制作。 https://www.unrealengine.com/en-US/metahuman

利益冲突声明

作者声明，本文发表不存在任何利益冲突。