arXiv:2601.08234v1 [cs.GR] 12 Jan 2026

面向图形应用的统计混合变形计算与分析

Shuxian Li*
联想研究院

Tianyue Wang
联想研究院

Chris Twombly†

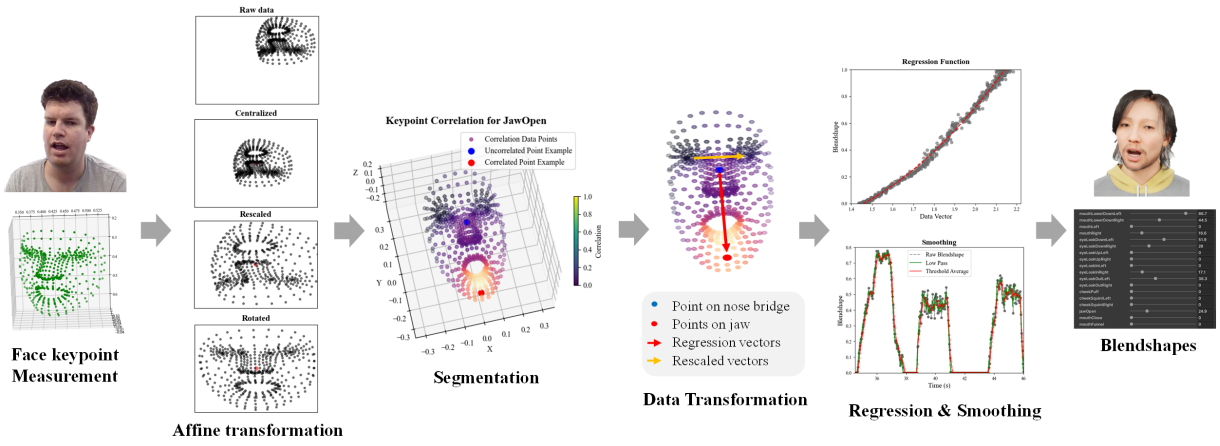


图 1：混合变形估计方法的逻辑流程

摘要

随着虚拟化和人工智能的发展，实时面部虚拟形象动画已广泛应用于娱乐、办公、商业等领域。在此背景下，混合形状因相对简单且易于解释，已成为常见的行业动画解决方案。为了实现实时性能并降低对计算资源的依赖，我们自主开发了一套面向低功耗虚拟现实应用、使用标准网络摄像头的高准确度混合变形预测系统。首先，通过仿射变换和分割提取混合变形特征向量。进一步经过变换和回归分析后，我们为大多数混合形状确定了具有显著预测能力的模型。我们还通过后处理进一步提升响应稳定性，包括采用平滑滤波和非线性变换来尽量减小误差。实验表明，该系统的准确度与 ARKit 6 相当。我们的模型对传感器和硬件的要求较低，且响应实时，能够提供稳定、准确、流畅的视觉体验。

索引词： 混合形状、角色动画、回归

1 引言

随着人工智能和虚拟现实领域的快速发展，数字人作为其中的重要组成部分，在社交媒体、商业和娱乐领域具有广泛的应用潜力。虚拟人涉及许多关键技术，包括虚拟形象创建、动画、绑定、交互和应用管理。其中，面部捕捉和驱动技术是实现虚拟数字人应用所需的最重要组成部分之一。这些技术使用户能够实时动态地控制虚拟形象的表情。

*电子邮箱：lisx14@lenovo.com
†电子邮箱：ct593@cornell.edu

人类虚拟形象通常使用虚拟网格模型来表示。为了让图形程序能够渲染角色，必须完整指定三维网格顶点位置。角色动画的主要类型包括蒙皮网格动画、关节动画和变形动画。存储每个视频帧的所有网格顶点位置虽然准确度很高，但需要大量计算资源，不适用于实时场景。关节动画有时用于控制眼球运动，简单方便，但不适用于精细的面部表情控制。可变形网格动画兼具两者的优点，在降低计算成本的同时保持较高的准确度

可变形模型方法通过由动态参数控制的小幅变形，确定角色网格模型的变化。在众多可变形方法中，混合形状因计算相对简单、具有物理可解释性且便于使用而广受欢迎。动画师提供目标形状，然后在中性形状网格与目标表情网格之间进行插值，以生成动画。这种方法广泛用于为虚拟角色制作动态表情动画。该方法能够实时运行，并可依据特定的表情定义准确映射面部状态。

虚拟形象动画最常用的方法包括完整面部重建、纯图像变换方法、基于关键点的方法和音频驱动技术。这些方法大多需要大型人脸数据集进行训练，而且计算资源需求较高，因此难以应用于性能较低的设备或实时动画。混合形状为解决这些问题提供了一种方案；本文提出一种使用网络摄像头的实时面部动画方法。我们利用统计模型将检测到的关键点转换为用于三维角色动画的混合变形系数。我们独立开发的模型准确率更高、硬件要求更低、所需训练数据更少，并适用于低功耗计算场景。我们的贡献如下：

- 独立开发统计模型，使用网络摄像头实时将关键点转换为混合形状。
- 一套旨在优化动画效果的算法。

- 进行了广泛的统计检验，包括准确率比较和硬件利用率结果；结果表明，该方法在硬件要求较低的情况下，性能接近当前最先进水平。

2 背景

2.1 网格到混合形状

网格模型是表示三维人脸的常见方式，广泛应用于电子游戏、动画、工程、设计等领域。网格由顶点、边和面组成，能够高效地表示复杂的三维形状。每个顶点在三维虚拟空间中都有坐标。通过这些顶点及其连接的边和面，可以准确描述和渲染三维物体的表面；移动顶点的位置，则可以表示不同的面部表情。

精细且准确的面部表情建模需要详细计算所有网格顶点的位置，以及所需的动态变换以进行更新，但完整重建往往需要大量计算资源。三维可变形模型（3DMM）通过降低计算负荷简化了建模过程。1999年，《用于合成三维人脸的可变形模型》[3] 使3DMM人体模型广泛应用于人脸重建和角色动画。作者随后使用主成分分析，识别人脸的常见特征。识别出的特征符合多维正态分布，然后利用与拟合分布的统计距离定义一组拟合参数和基向量。他们能够通过对这些特征向量进行线性组合来构建新的人脸。此后还开发了其他线性和非线性人脸模型及数据集，包括 Basel[33, 15, 14], Surrey[20], Facewarehouse[6], 大规模人脸模型[4], 和四维人脸模型[44]。所有这些方法都独立开发了动态参数，可用于泛化面部，使其符合特定特征。

混合形状与 3DMM 相关，是对通用特征概念的拓展，用于描述动态面部运动，同时具有简单易懂的解释和更明确的定义。混合形状是一组预定义的面部状态，可构成面部运动的语义参数化：

$$f = w_o f_o + w_1 f_1 + \dots \tag{1}$$

其中 w_o,... 为权重系数，f_o,...为一组面部表情状态。权重系数的取值范围为 0 到 1。每一项都代表一个特定的子表情，用于定义表情空间中的一组特征。

混合形状模型是虚拟现实动画中的常用工具，因为它们相对简单、物理定义清晰且易于解释，并且能够将表情限制在高度合理的运动范围内。它们已成为电影[29] 和电子游戏[23] 行业的主要动画方法。至少从 20 世纪 70 年代[32], 起，混合形状式的特征控制就已应用于角色动画，并于 20 世纪 80 年代末[30]在商业软件中实现。近期的研究主要致力于利用计算机视觉和人工智能方法估计人类面部的混合形状状态，再使用混合形状驱动人类虚拟形象进行动画。

传统上，动画师会为每个人类模型单独创建混合形状。角色所需的混合形状数量通常在 50 到 900+ 之间，具体取决于[29]准确度和期望的表情范围。对于大多数电子游戏应用，52 个混合形状已成为事实上的通用标准，虚幻引擎 5[13] 和 Unity[42]等最常用的游戏开发图形引擎均采用这一标准。这一标准基于苹果增强现实套件（ARKit[1]）。ARKit 混合形状采用主成分分析分解和聚类分析创建，用于描述常见的表情模式。ARKit 混合形状使用 iPhone 10 或更高版本设备中的高分辨率摄像头、激光雷达和红外传感器系统。这与标准网络摄像头等性能较弱的系统并不匹配。本研究将混合形状扩展至基于网络摄像头的人工智能地标回归模型。我们对我们的算法与 ARKit 6 进行了直接比较分析。

2.2 地标回归

面部关键点检测方法大致可分为以下几类：传统方法，如主动形状模型 [9] 和主动外观模型 [12]；基于级联形状回归的方法，如级联姿态回归[10]；基于深度学习的方法，如深度卷积神经网络[40], 、多任务卷积神经网络[46] 和深度对齐网络[28]。全身姿态估计算法有很多，其中包括面部关键点检测方案，如 RT MPose[24], 、DWPose[45] 和 OpenPose[7]。我们的算法使用地标检测的输出，具体而言，采用实时性能较高且状态位置信息充足的方法来识别清晰的动作。上述大多数方法缺乏实时性能，或提供的面部关键点不足。Google MediaPipe Holistic Model [31] 满足面部几何方案的这些要求：仅需输入单目摄像头图像，即可实时估计 478 个三维面部/眼部关键点。

MediaPipe 使用两步计算流程进行人脸检测。模型首先使用 BlazeFace[2],进行人脸分割；BlazeFace 基于 MobileNet [19, 36] ，但针对速度进行了进一步优化。这些优化使推理时间相比 MobileNetV2 缩短了 3.5 倍，但回归误差增加了 40%，抖动[2]也增加了 47%。抖动对敏感的混合形状计算尤其明显，因此需要实现平滑算法来补偿这一影响。完成人脸检测和分割后，第二个流程使用深度学习模型[27]进行地标回归，并通过普罗克鲁斯特斯分析[16] 将回归得到的关键点映射到三维度量空间。返回的 z 坐标表示相对于穿过网格质心的参考平面[27]的深度。与其他关键点回归方法相比，MediaPipe 在低功耗计算设备上能够以实时推理速度检测出更多面部关键点。

本研究采用了 MediaPipe Holistic Model 0.9.0.1[31], ，但也可扩展至其他实现。考虑到 MediaPipe 的局限性，仍需开展更多工作以解决这些问题，包括旋转、锚定以及某些非线性面部变形。

3 方法

我们的模型使用标准网络摄像头，实时将面部关键点转换为混合形状。为此，模型执行以下任务：

$$f_b = F(R(T_d(S(T_a(X))))) \tag{2}$$

其中，X 表示回归得到的面部关键点，f_b 表示混合形状估计值。方程中的各项如下：

- T_a：仿射变换
- S：分割
- T_d：数据变换
- R：回归
- F：平滑

方程中的每一项均经过统计检验支持。计算过程见图 1。

3.1 仿射变换

为最小化回归误差，我们执行一系列仿射变换。仿射变换保持点集之间的共线性、直线之间的平行性、集合之间的凸性、向量模的比值以及几何质心不变。我们选择变换，使关键点显示在一个统一的、相对旋转和失真不变的基中。这适用于进一步的数据变换和回归。完整的仿射变换为：

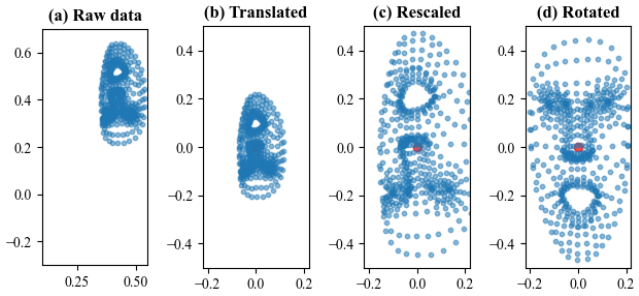


图 2：仿射变换矩阵对（a）原始数据的影响，其中（b）R_1，（c）R_2 （d）R_3 依次应用

$$T_a = R_3 \cdot R_2 \cdot R_1 \cdot X \tag{3}$$

其中，X 是原始关键点位置估计值，R_1, R_2, R_3 是用于将观测到的人脸映射到所需回归基的变换矩阵，而T_a 是变换后的输出。每个矩阵表示一组特定的旋转、变换和重新缩放操作，用于将输入地标点变换为用于回归的眼-眼-鼻屏幕朝向基，如图 2所示。

3.2 分割

混合形状处理的下一步是地标点分割。分割有助于降低数据的维度，同时通过统计分析识别特定的相关特征，从而保持较高的准确度。这样可以将内存和计算需求降低 95-99%。根据训练数据的相对稀疏性，采用相关性分析、特征回归和支持向量回归来识别特征。这足以识别大多数混合形状的特征。一个显著的例外是吐舌。这从直觉上说得通，因为没有使用 MediaPipe Holistic 关键点来描述舌头运动。

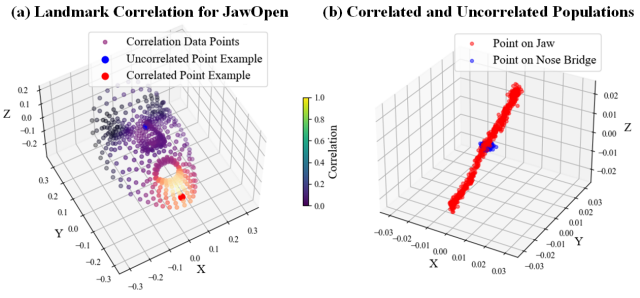


图 3：张嘴的相关性。（a）张嘴对应的面部所有点的线性相关幅度。黄色点表示高度相关，黑色表示不相关。（b）高度相关（蓝色）和不相关（红色）采样点的行为。这一比较展示了明显不同的行为，说明在回归建模之前进行分割是合理的。

我们通过图 3展示了对张嘴进行的更全面分析。使用合成测试数据，我们观察到高度相关与不相关的关键点之间存在明显的行为差异，如图 3(a)-(b) 所示。相关性分析确定了相关性最高的前 2 个百分位数的关键点。我们采用了多种交叉验证方法来检验所选地标的相对重要性，包括 F 回归和使用径向基核的支持向量回归器。我们发现不同的地标点选择方法之间具有良好的一致性——这一点如图 4(a)-(d) 所示。

基于这项分析，我们为每个混合形状模型开发了分割矩阵。该操作通过一组重采样矩阵表示，这些矩阵用于选择特定的关键点子集：

$$S = D \cdot T_a \tag{4}$$

其中，D 是分割矩阵，T_a 是变换后的地标估计值，S 是分割输出。我们在统计建模中使用了以眼鼻为中心的基。

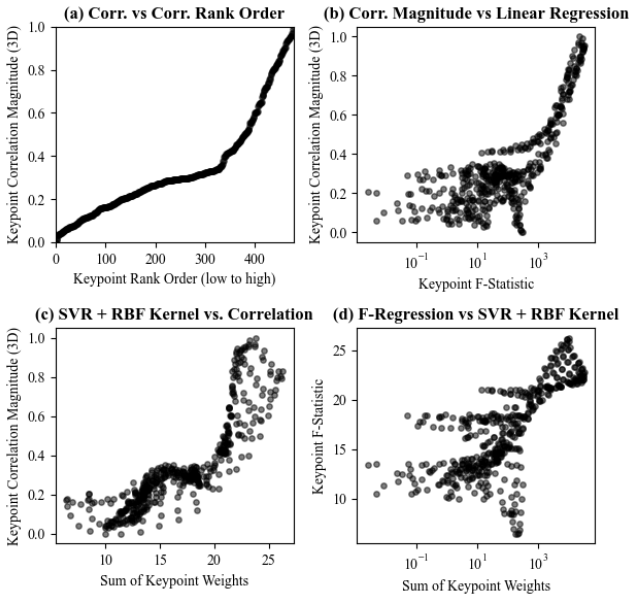


图 4：关键点选择与分割的统计学依据。(a)-(d) 展示了进一步的统计分析。(a) 地标点排名顺序，(b) 三维相关性与一维线性回归，(c) 支持向量回归 + 径向基函数核与三维相关性，(d) F 回归与支持向量回归 + 径向基函数核。

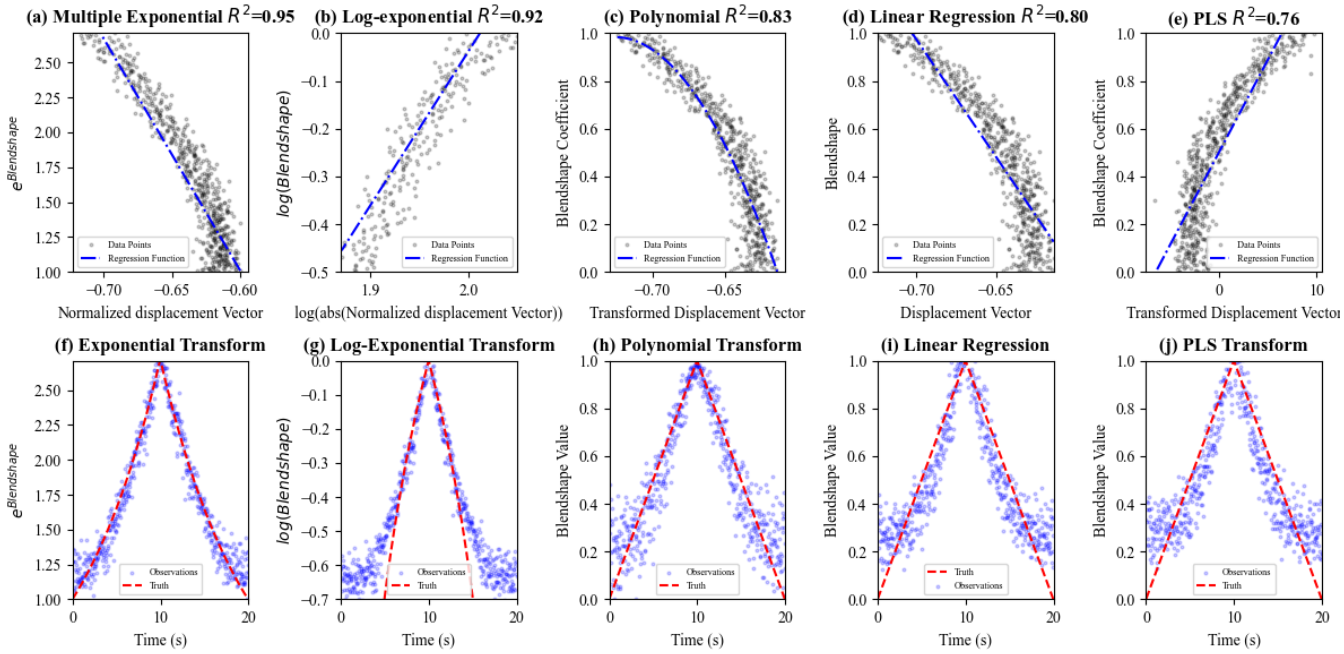
3.3 变换

数据变换对于使特征数据更适合回归是必要的，尤其是在难以观察到自变量与因变量之间所需的线性关系时。有时可以对自变量进行变换以恢复线性，从而使其适用于标准回归。数据变换可以降低预测误差，也可以由统计检验提供依据。标准回归方法的常见要求包括同方差性（方差齐性）和多元正态性。我们根据各模型的具体要求以及观察到的地标数据统计特性，为每个模型量身定制了最后一步数据变换。

例如，对于某些混合形状，我们使用主成分分解[18, 35] (PCD)作为一种变换方法。该方法会沿着方差最大的方向，将标志点位移转换到一个新的坐标系中。随后可以应用方差稳定化及其他变换，以生成更易于分辨、适用于回归的特征。我们使用的变换作用于关键点估计值：

$$T_d = T(S) \tag{5}$$

其中，T 是一个对分段标志点估计 S 应用变换的函数，而T_d 是变换的结果。或者基于



检验/指标	指数	对数指数	多项式	线性	PLS
均方误差[25]	0.0119	0.016	0.0134	0.0163	0.0196
杜宾-沃森检验[25]	1.360	1.261	1.140	0.656	0.680
布劳施-帕甘检验[5]	21.5 p=$2.41 \cdot 10^{-4}$	13.4 p=$9.15 \cdot 10^{-3}$	49.1 p=$2.12 \cdot 10^{-11}$	17.6 p=$2.68 \cdot 10^{-5}$	13.3 p=$2.57 \cdot 10^{-4}$
F统计量[43]	2856	675.3	1556	2463	1946
相关（残差-训练）[35]	0.975	0.967	0.915	0.897	0.874
ξ[8]	0.778	0.679	0.658	0.658	0.623

图 5：CheekPuff 的数据变换和样本回归，以及由此得到的决定系数（R^2）。**(a)** 展示多重指数模型，**(b)** 多重对数指数模型，**(c)** 多项式（次数=2），**(d)** 标准普通最小二乘线性回归 **(e)** 偏最小二乘投影（PLS，次数=1），**(e)**。下方分别是与真实值的比较：**(f)** 指数模型、**(g)** 对数指数模型、**(h)** 多项式、**(i)** 线性模型、**(j)** PLS。性能统计指标包括均方误差[25], 杜宾-瓦特森统计量[11], 布鲁施-佩根检验[5], Fisher 统计量[43],皮尔逊相关[34] 以及 ξ 相关[8]。

用于开发各个回归模型的因变量分布：

$$\varepsilon = T_d(Y - mx) \tag{6}$$

将变换操作T_d 应用于用于创建回归模型$mX + \varepsilon$的训练数据分布。两种类型的变换均有使用。

示例见图5，其中比较了不同变换方法对 CheekPuff 普通最小二乘回归准确度的影响。图5(a)-(e) 显示数据拟合情况，(f)-(j) 显示与真实值数据的准确率比较。表格展示了用于量化模型准确度的不同统计检验结果，包括决定系数（R^2）[25], 均方误差[25], 杜宾-瓦特森统计量[11],布鲁施-佩根检验[5], 费舍尔统计量[43], 真实值与响应之间的皮尔逊线性相关[34] 和 ξ 相关[8]。杜宾-瓦特森检验检测到了异方差性，变换操作在一定程度上对其进行了补偿。

3.4 回归

回归是一种统计过程，用于估计一组检测变量与响应变量之间的关系。我们的混合变形估计系统使用了几种不同的回归模型，具体取决于检测到的回归特征的相对重要性和实用性，并采用 图 5 中所强调的方法。大多数混合形状都可以使用相对简单的方法进行准确估计，例如多重线性普通最小二乘回归。统计分析为此提供了依据。这些模型要获得准确预测，必须满足多元正态性和同方差性。最终数据转换后，约有 10 个混合形状需要更复杂的回归方法。CheekPuff 便是其中的典型例子，其具有显著的异方差性，无法通过更简单的回归方法加以处理。

满足或基本满足线性回归要求的混合形状，可以使用标准技术成功建模。例如，图 6 展示了多元线性回归、多项式回归、主成分回归和偏最小二乘回归。我们还测试了更先进的回归方法，包括采用不同核函数的高斯过程回归、采用不同核函数的支持向量回归，以及不同的集成方法（直方图梯度提升方法、自适应提升回归、梯度回归和K近邻回归），如 图 6所示。我们通过统计检验来确定模型的选择，使用 R^2[25]、 ξ[8]均方误差[25], 杜宾-沃森[38] 自相关检验[11] （DW）、残差误差[25],夏皮罗-威尔克检验[37] （SW）以及内存利用率进行评估，结果如

表 1：混合形状模型的示例性统计行为

模型	R^2[25]	ξ[8]	均方误差[25]	杜宾-沃森统计量[11]	夏皮罗-威尔克斯值[17]	内存（字节）
标准混合形状 (SmileLeft)						
机器学习-普通最小二乘法	0.9919	0.8849	0.0010	1.7679	0.9976	1553
Poly	0.9882	0.8648	0.0014	1.6839	0.9931	20856
PCA	0.9918	0.8501	0.0018	1.7380	0.9964	4192
PLS	0.9899	0.8659	0.0015	1.6906	0.9975	5358
高斯过程回归基函数	0.9886	0.8645	0.0014	1.6274	0.9931	2911867
支持向量回归径向基函数	0.9871	0.8680	0.0015	1.6768	0.9976	1602
梯度提升直方图/径向基函数	0.9905	0.9112	0.0011	1.8433	0.9855	2707469
复杂异方差 (CheekPuff)						
机器学习-普通最小二乘法	0.9365	0.7812	0.0120	0.5719	0.9858	1548
Poly	0.8458	0.6577	0.0134	1.1404	0.9663	30488
PCA	0.7560	0.6235	0.0203	0.6800	0.9811	5486
PLS	0.7649	0.6237	0.0196	0.6800	0.9795	14062
机器学习-指数	0.9509	0.7784	0.0119	1.3601	0.9392	1548
高斯过程回归基函数	0.8405	0.6587	0.0133	0.9960	0.9710	2897439
支持向量回归/多项式	0.0873	0.6430	0.0763	0.0005	0.9539	55957
支持向量回归径向基函数	0.8038	0.6605	0.0164	0.5063	0.96674	32721
梯度提升直方图/平方指数	0.8715	0.7597	0.0107	1.1232	0.9766	26837469
ADA	0.8517	0.9310	0.0123	0.8351	0.9901	8028
Grad. Bost.	0.9043	0.7943	0.0080	1.171	0.9788	5529

表 1。

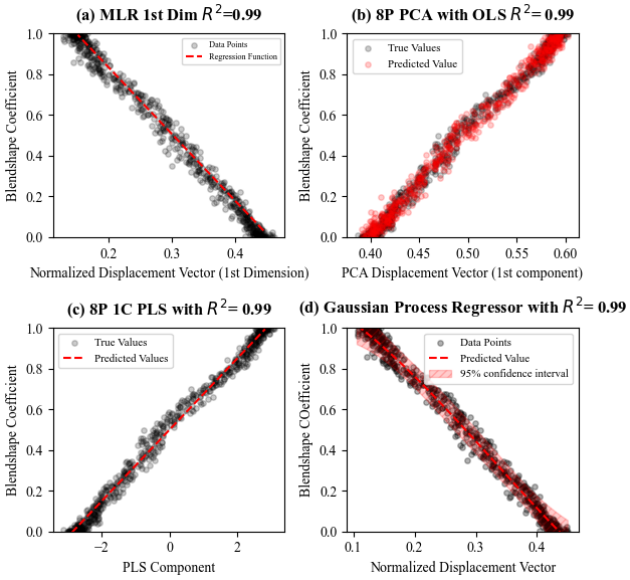


图 6：简单混合形状类别的回归模型。(a) 多元线性回归模型的单个元素。(b) 使用 8 个不同数据向量的 1 个主成分主成分分析模型。(c) 使用 8 个关键点位移向量创建的偏最小二乘模型。(g) 高斯过程回归模型，显示概率分布的 95% 置信区间。

较简单的回归方法通常对微小的观测误差或边缘情况更具鲁棒性。较复杂的回归方法对底层数据分布的假设较少，但会带来额外的计算成本，并降低解释的清晰度。我们观察到不同模型的性能存在显著差异，而且某些混合形状（包括 SmileLeft 和 CheekPuff）存在较强的自相关。这些回归模型能够准确预测特定人脸模型的混合形状系数。将这些模型应用于不同面部时，则会带来额外的复杂性。

这会引入显著偏差，因此促使人们在回归模型中加入额外项：

$$R_t(f) = (1 + \frac{g_t(f)}{\eta_t} - \gamma_t)(w \cdot f + b + \beta_t) \tag{7}$$

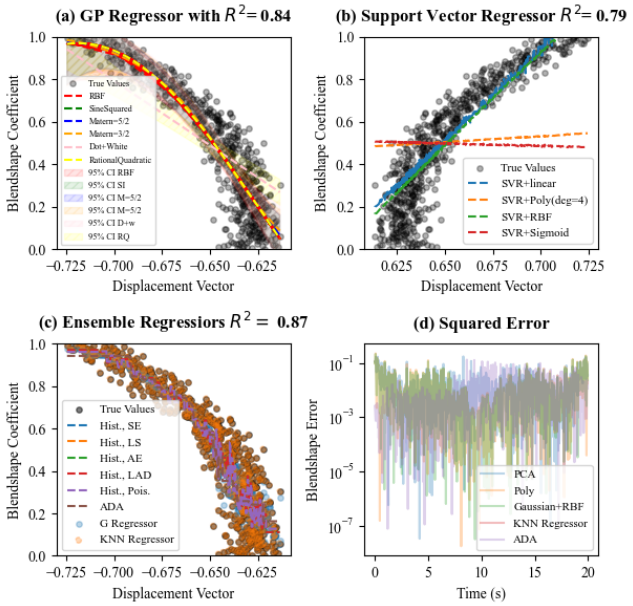


图 7：用于异方差混合形状估计的非标准回归模型。(a) 使用不同核函数的高斯过程回归器。(b) 使用不同核函数的支持向量回归器。(c) 使用不同核函数的集成回归器。(d) 不同回归模型的平方误差比较。某些混合形状需要更复杂的回归流程。

其中，$wf + b$ 描述线性回归模型的标准组成部分、加权系数和初始偏差；$g(f)$表示响应函数，η 是阻尼项，γ 是时间加权偏差校正项，β 是额外的偏差校正项。该 β 项的计算方式为：

$$\beta_{t,0} = \begin{cases} f_t & f_t > f_{t_o} + e^{k\cdot(t_o - t)} \\ \beta_0 & f_t \le f_{t_o} + e^{k\cdot(t_o - t)} \end{cases} \tag{8}$$

其中 k 是可调节的时间尺度，F_t 是当前混合形状估计值，B_t 是偏移项，而 $t_t - to$ 是前一个最小值与当前帧之间的时间差。γ，η 项可以通过经验推导，也可以通过对地标和回归混合形状进行统计检验得出，并针对每种混合形状进行定制。

非线性校正项对 Blin- kLeft 的影响如图 8所示。对于大多数混合形状，在真实人脸中尤其难以观察到动态范围最后 15% 的变化。图 8(a) 展示了 β 项在方程7 和8中的作用。这对图 8(c) 所示的混合形状自相关产生了显著影响。自相关变化最大的是由β 偏移引起的，而后续的加权函数、阻尼和偏移项并未明显改变混合形状行为。两个通过经验推导的权重函数如图 8(b) 所示。这些函数旨在校正回归偏差，尤其是在混合形状范围的上边缘附近。图 8(d) 展示了测试期间所有校正项对混合形状估计值的累积影响。

3.5 平滑

我们观察到显著的关键点抖动，对混合变形估计的稳定性造成了负面影响。在部分遮挡或头部几乎完全旋转等边缘条件下，这一现象尤为明显。

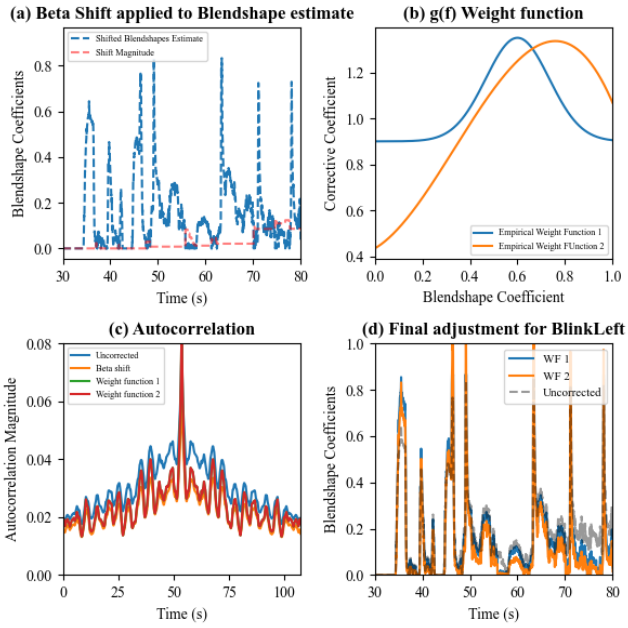


图 8：BlinkLeft 的偏差校正和自相关。(a) 展示了 β 偏移对原始混合形状估计值的影响。(b) 用于偏差校正的两个经验加权函数。(d) Beta 偏移会对混合形状观测值之间的自相关产生较大影响；而后续的加权函数、阻尼和偏移项不会改变自相关行为。(d) 测试期间的初始混合形状估计值和校正后的混合形状估计值。

这促使我们开发平滑算法来补偿这些影响，同时提供额外的偏差校正，并且比标准插值方法具有更快的响应时间。不同混合形状所需的响应行为存在显著差异，因此我们实现了几种不同的平滑算法，包括：

- 卡尔曼滤波器[26]
- 移动平均[22, 17]
- 时间加权移动平均[21]
- 低通滤波
- 门控移动平均

门控移动平均旨在当响应行为出现突然的大幅变化时放宽滤波。这有助于减少平滑滞后，同时在正常响应期间保留平均处理的平滑效果。其可用以下方程表示：

$$f_{avg} = \frac{1}{N}\sum_{i=1}^{N} f_i \tag{9}$$

$$f_{var} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(f_i - f_{avg})^2} \tag{10}$$

$$f_t = \begin{cases} \frac{f_{t-i}+N\cdot f_{avg}}{N+1} & f_{avg} - f_{var} \le f_i \le f_{avg} + f_{var} \\ f_i & f_i \le f_{avg} - f_{var} \\ f_i & f_i \ge f_{avg} + f_{var} \end{cases} \tag{11}$$

其中，f_{avg} 是对前N 个先前的单独状态f_{t-i}求和后得到的平均混合形状估计值，f_t 是当前正在

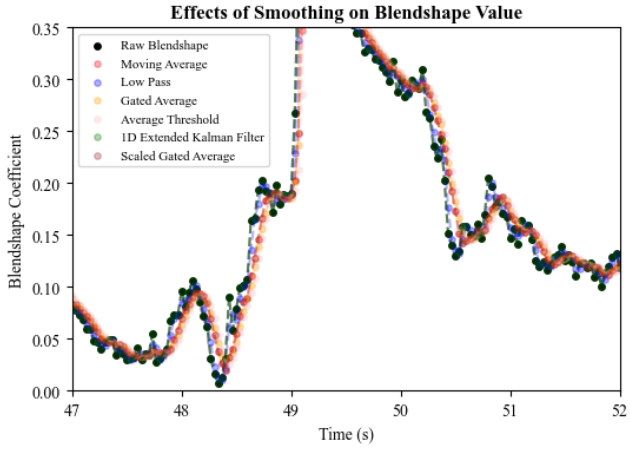


图 9：在更窄的时间范围内，不同平滑算法对混合形状估计值的演示。

考虑进行平滑的状态，f_{var} 是根据平均值计算的 N 个混合形状状态之间的方差，而f_{smooth} 是新的平滑混合形状状态。通过更改用于计算平均值的先前混合形状估计值数量，可以调整平滑效果的强度以及由此产生的滤波滞后。这种平滑器对于具有较大突发动态变化的混合形状特别有用，例如眨眼或张嘴。

4 实验与结果

4.1 数据集与实现

训练使用了两种不同的数据源。使用 iPhone 12 Pro 拍摄的 131 张原始图像用于粗略分割和特征识别，但不用于训练。使用一个经过配置、可接收 LiveLink 混合形状命令的虚幻引擎 4 MetaHuman 虚拟形象，为每个混合形状创建合成训练数据；通常每个混合形状包含 200–600 张图像，分辨率与标准网络摄像头相匹配（400x600 像素）。训练需要四个不同步骤：特征提取/分割、回归、加权函数和平滑。通过相关和协方差分析以及其他统计方法，识别用于回归的相关关键点。回归分析用于创建基础回归模型，并针对准确度、内存使用量和复杂度进行优化。非线性加权函数根据与一系列自相关测试和视频分析的比较结果进行优化。最后，对平滑过程进行优化，以减少帧间抖动。

测试使用了从 21 个视频中提取的 18209 帧，这些视频由 iPhone 14 拍摄，分辨率为 1920x1440 像素。ARKit 可使用传感器套件中的红外摄像头和激光雷达传感器，并用于算法性能比较。由于 ARKit 的要求，需要使用更高的分辨率。6 名不同的测试者（4 名男性和 2 名女性）在各种环境和光照条件下进行录制。每名测试者均包含 6 种不同的复合表情和 9 种头部运动。测试者接受了训练，根据每种特定表情对应的示例虚拟形象来做出面部表情。测试者被要求模仿虚拟形象的表情，并记录其响应。人工标注由一名受过训练的观察员完成，该观察员能够观察同一模型虚拟形象的表情。

4.2 回归性能

在确认标志点数据适用于回归后，我们开发了方法部分所述的完整系统。我们开发了一种仿射变换，将人脸变换到用于回归分析的公共基底。我们投入了大量精力

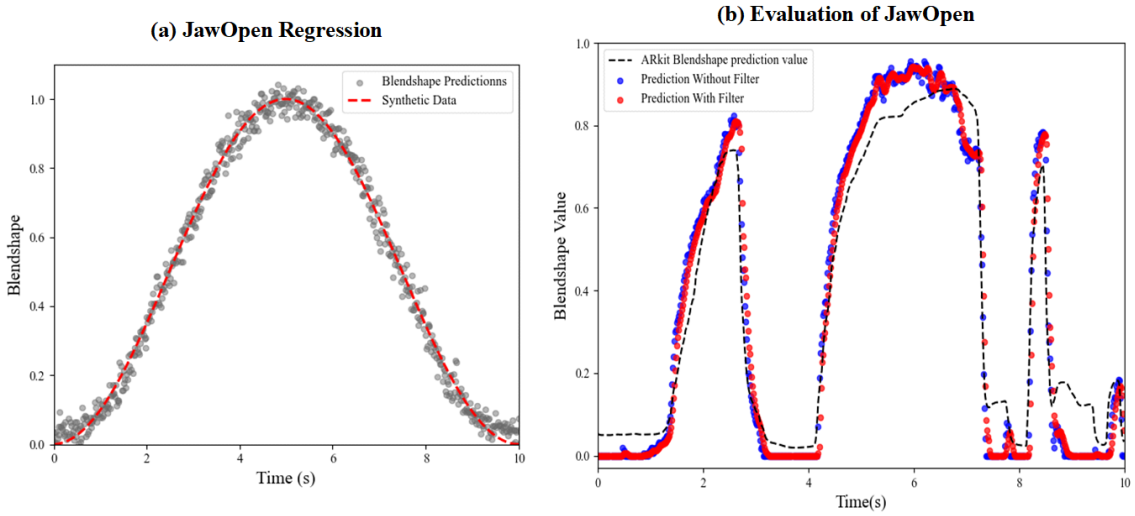


图 10：(a) 显示了 JawOpen 混合形状预测（灰色点）与合成测试数据（红色虚线）的对比。(b) 展示了苹果 ARKit 6[1]（虚线）与未进行滤波的新模型（蓝色）和经过滤波的新模型（红色）之间的高度相似性。

用于找出最适合回归分析和建模的特征。接下来，对数据进行变换，以生成回归所需的位移向量。我们首先使用主成分分析来确定方差最大的方向。随后，根据相关锚点估算位移，并将其转换为锚点的纵横比。针对特定混合形状，有时还需要进一步变换数据。这通常是由异方差性或多重共线性造成的。我们针对每种混合形状，使用变换后的地标向量进行回归分析。每种回归方法都有各自的优缺点，因此我们选择了能够最大限度提高准确度和性能的方法。异方差性和偏差校正促使我们在混合形状模型中添加了额外的回归项。加入这些校正项后，混合形状的动态范围得以恢复，从而能够在真实世界条件下完全闭合人眼。我们针对每种混合形状，分别优化了所用的具体回归模型和权重函数。

在真实环境测试中，我们观察到显著的偏差和噪声。因此，我们开发了平滑算法，并通过统计检验量化了这些平滑算法的效果。使用来自真实环境测试的 3200 次检测构成的常见数据集，对 BlinkLeft 进行了测试。例如，门控平均平滑器使用 5 个而非 10 个先前状态时，一阶自相关最高。采用 5 个状态的门控平均平滑器会带来平均 59 毫秒的延迟，之后才能观察到具有相同值的混合形状预测。与标准插值算法相比，这一延迟要短得多。每种平滑器所需的内存都非常少，因此适用于低功耗计算应用。

我们观察到回归模型与训练数据之间高度一致。进一步的测试结果见 图 10(a)，其中使用正弦波生成函数控制虚幻引擎 5 人类化身模型中的 LiveLink 混合形状命令，红色虚线表示该函数。随后，我们使用完整的回归系统估计混合形状值，结果以灰色圆点表示。源数据与估计值之间产生差异的可能原因，是 Mediapipe Holisitic 0.9.0.1 地标回归过程中引入的回归误差，尤其是 BlazeFace[2] 分割步骤。即使在动态范围极限处（此时地标检测模型表现出明显的不稳定性），回归模型仍能准确匹配生成函数。

我们对进一步开展统计分析，以比较我们的混合形状估计方法 nd-

与苹果 ARKit 的方法 6[1]。首先，使用参考视频片段对单个混合形状进行直接比较，如图 10（b）所示。展示了 ARKit 和我们的回归模型估计的张嘴值，以及采用和未采用门控平均平滑算法时的结果。我们的算法与 ARKit 的响应高度一致，在测试视频中的平均偏差为 -0.0005，均方偏差为 0.007。我们将运动周期结束时偏差增大的区域解释为 ARKit 张嘴模型内置的插值或平滑算法伪影，并已确认在所研究的特定时刻口腔/下颌处于闭合状态。我们的混合形状模型在运动周期开始时始终响应更快，我们进一步认为这是采用响应更灵敏的平滑方法所致。在动态范围上限处（> 0.8），估计的混合形状值存在明显抖动，这是由 MediaPipe Holistic 0.9.0.1 模型在这些时刻不稳定造成的，门控平均平滑器在一定程度上对此进行了补偿。我们的模型能够检测微表情中细微的实时变化，而 ARKit 似乎无法观测到这些变化。

4.3 比较分析

接下来，我们希望将模型性能与现有系统，尤其是 ARKit，进行比较。由于缺乏客观真实值数据，且完整的模型细节尚未公开，这一比较较为复杂。混合形状表情定义与特定的网格运动相关，不同混合形状之间的交互用于创建更加逼真的面部。我们只能使用一组任意定义的混合形状所对应的表情，而这些表情可能与面部真实运动不相符。因此，我们设计了几种统计方法来比较响应准确度。这些方法包括不同类型的相关性分析，以及针对特定检测事件的改进型 $F1$ 分数。结果见表2。

我们的相关性分析重点关注 ARKit 6 与我们的混合形状回归方法之间的响应相似性。我们采用了多种不同类型的相关系数，包括皮尔逊乘积相关[35]，斯皮尔曼秩相关[39, 38]，和 ξ 相关[8]。使用多种相关性测量指标有助于提高实验结果的可信度。具体而言，皮尔逊相关用于衡量变量之间线性关系的强度，而斯皮尔曼和 ξ 相关用于衡量单调关系的强度，这种关系不一定是线性的，

表 2：ARKit 与我们的方法之间的比较，以及各方法之间的响应相似度

混合形状	ARKit[1]			我们的模型			交叉比较					
	Pr	Re	F1[41]	Pr	Re	F1[41]	皮尔逊相关系数[35]	斯皮尔曼相关系数[39]	ξ[8]	准确度	MSD	偏差
左眼眨眼	1.00	1.00	1.00	1.00	1.00	1.00	0.576	0.572	0.305	0.752	0.061	0.216
左眼向下看	0.80	1.00	0.88	0.61	1.00	0.76	0.657	0.678	0.393	0.876	0.015	0.074
左眼向内看	0.80	1.00	0.88	0.50	1.00	0.66	0.489	0.504	0.402	0.773	0.051	0.224
右眼眨眼	1.00	1.00	1.00	1.00	1.00	1.00	0.674	0.675	0.410	0.794	0.042	0.282
右眼向下看	0.80	1.00	0.88	0.61	1.00	0.76	0.658	0.679	0.385	0.876	0.015	0.074
右眼向内看	0.83	1.00	0.90	0.80	0.80	0.80	0.567	0.606	0.400	0.914	0.007	0.133
张嘴	1.00	1.00	1.00	1.00	1.00	1.00	0.921	0.886	0.682	0.965	0.001	0.039
嘴向右	0.88	1.00	0.94	0.80	1.00	0.88	0.636	0.537	0.349	0.961	0.001	0.025
嘴向左	0.88	1.00	0.94	1.00	0.62	0.76	0.610	0.616	0.425	0.945	0.002	0.042
左侧嘴角微笑	0.88	1.00	0.94	1.00	0.62	0.76	0.932	0.953	0.740	0.867	0.017	0.073
右侧嘴角微笑	0.88	1.00	0.94	0.80	1.00	0.88	0.907	0.943	0.702	0.861	0.019	0.067
左侧上唇上提	0.88	1.00	0.94	1.00	0.62	0.76	0.474	0.467	0.407	0.925	0.005	0.078
右侧上唇上提	0.88	1.00	0.94	0.80	1.00	0.88	0.496	0.504	0.387	0.927	0.005	0.080
左眉下压	0.77	1.00	0.87	0.63	1.00	0.77	0.699	0.544	0.404	0.698	0.091	0.337
右眉下压	0.77	1.00	0.87	0.75	0.85	0.80	0.717	0.518	0.373	0.756	0.059	0.235
平均值	0.87	1.00	0.93	0.82	0.90	0.83	0.667	0.645	0.451	0.859	0.019	0.132

斯皮尔曼和 ξ 也对离群值不太敏感，因为进行了额外的数据预处理。

为了进行更客观的比较，我们采用了一种基于测试数据集标注的修订版$F1$评分标准。这些标注记录了何时可以观察到某种表情，但未记录具体的表情强度。$F1$ 分数的计算方法如下：

$$F1 = \frac{(2 \times P \times R)}{(P+R)} \tag{12}$$

其中，P 表示精确率，R 表示召回率。精确率衡量避免预测出假阳性的能力，而召回率表示漏掉正值的概率。精确率和召回率的计算方式如下：

$$Precision = \frac{n_{TP}}{n_{TP}+n_{FP}} \tag{13}$$

$$Recall = \frac{n_{TP}}{n_{TP}+n_{FN}} \tag{14}$$

其中，n_{TP} 为真阳性（TP）观测数，n_{FP}为假阳性（FP）观测数，n_{FN} 为假阴性（FN）观测数。每种不同表情事件的判定标准，是根据各混合形状对应的表情事件所定义的灵敏度和时间阈值，判断其是否处于这些阈值范围内：

$$\delta_{f_b,t_k} = \begin{cases} 1 \text{ if } f_{min} \le f_b;\ t_{min} \le t_k \le t_{max} \\ 0 \text{ if } f_{min} \ge f_b;\ t_k < t_{min}; t_k \ge t_{max} \end{cases} \tag{15}$$

$$n_{TP} = \sum_{t=0}^{t_N} \delta_{f_b,t_k} \tag{16}$$

$$n_{FN} = n_{algo} - n_{TP} \tag{17}$$

$$n_{FN} = n_{ref} - n_{TP} \tag{18}$$

其中，δ 表示对每个检测到的表情事件的判定，f_{min} 是根据人类观察者的判定优化得到的最小灵敏度阈值，t_{min} 和 t_{max} 定义了独特表情事件周围的时间窗口，t_k 表示特定的视频帧编号，f_b 表示混合形状值。对检测到的表情事件的判定进行汇总，可确定 TP、FP 和 FN 观测值的数量。n^{ref} 指已标注的表情事件总数，而n^{algo} 是算法判定正确的数量。时间窗口定义如下：

$$t_{min} = t_{anno} - \Delta_k \tag{19}$$

$$t_{max} = t_{anno} + \Delta_k \tag{20}$$

其中t_{anno} 是特定表情事件的帧编号，Δ_k 是可调整的准确度阈值。例如，该准确度阈值约为 15 帧，但对于帧率更高的视频，需要对其进行调整。



图11：原始图像（左列）、ARKit（中列）与我们的算法（右列）之间的响应对比

我们在表2中展示了我们的算法与 ARKit 之间的相关性，以及 F1 分数的准确率指标。从结果中

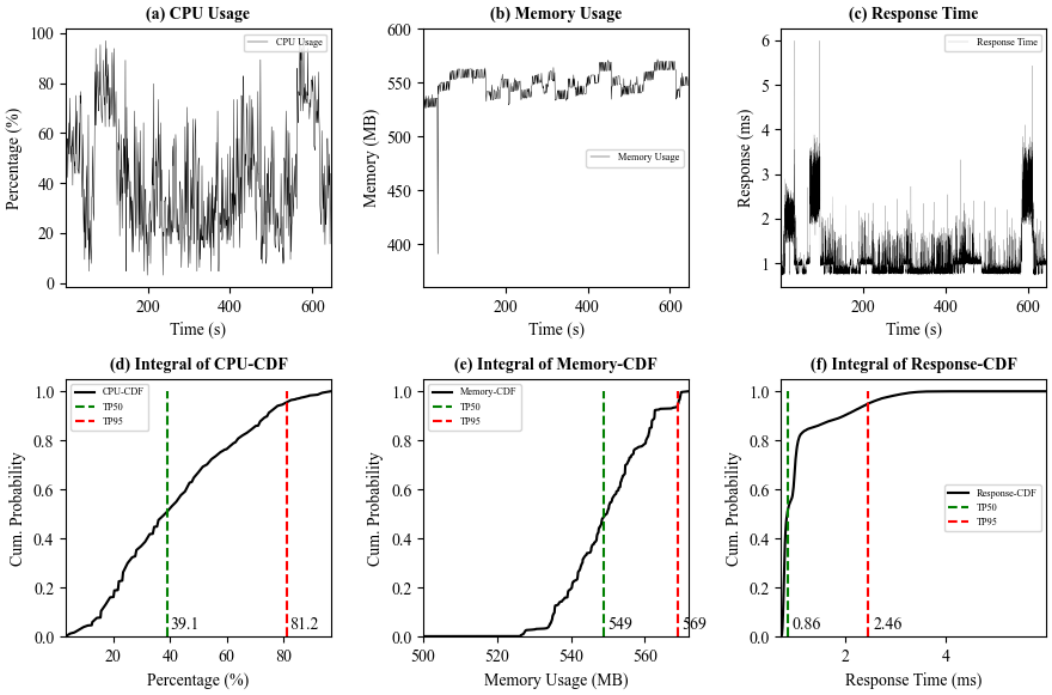


图 12：3D 虚拟形象算法的性能验证数据。(a)、(b)、(c) CPU 使用率、内存使用量和响应时间随时间的变化。(d)、(e)、(f) 分别为 (a)、(b)、(c) 的累积分布函数图，显示了 TP50/TP95 值的位置。

可以看出，我们计算的大多数混合形状都与 ARKit 有良好的相关性，且在主观观察中表现尚可。对于左眼眨眼、右眼眨眼和张嘴，我们的精确率、召回率和$F1$ 分数均达到 100%，这表明其主观观感与 ARKit 相同。然而，左侧其他混合形状的估算准确度不如右侧，这可以作为进一步优化的方向。至于整体 F1 分数，与 ARKit 相比差异较小（9.53%），这表明其在现实世界中的应用也可以接受。

视觉响应对比如 图 11所示。这凸显了我们的算法与真实人脸表情之间高度相似的响应。测试结果表明，与 ARKit 6 相比，我们的算法在眼部和嘴部混合形状的响应准确度方面略有提升。ARKit 表现出显著的鲁棒性，我们仍需进一步改进，才能使系统达到同等水平。

4.4 硬件性能

我们开展了硬件性能分析，以评估我们的计算过程是否适用于实际应用。测试使用配备 14 个核心的第 12 代 Intel(R) Core(TM) i7-12700H CPU 进行。我们测量了整个测试数据集的 CPU 使用率、内存使用量和响应时间。全部 21 个测试视频均输入系统，并按顺序处理，同时记录性能数据，如图 12(a)-(c)所示。我们对性能测量结果进行了统计分析，并通过累积概率图展示，如(d)-(f)所示。单个核心 CPU 使用率的第 50 和第 95 百分位数介于 39.1% 和 81.2% 之间，而总内存使用量在测试期间介于 549MB 和 569MB 之间，表明系统处于稳定的工作状态。超过 90% 的已用内存由 MediaPipe Holistic Model 占用。包括关键点检测过程在内的所有处理步骤均在 CPU 上执行。算法响应时间的第 50 和第 95 百分位数分别介于 0.86ms 和 2.46ms 之间。我们开发了两个版本的算法，分别使用 Python 或 C++。本次实验测试的是 Python 版本，该版本能够执行实时混合形状计算，适用于智能手机和平板电脑应用。

5 结论

我们成功开发了一种统计模型，可将关键点估计值转换为用于 3D 人类虚拟形象动画的混合形状。该模型涉及仿射变换、分割、数据变换、自相关回归和平滑。每个处理环节的开发均以大量统计分析为支撑。我们成功复现了虚幻引擎 4 MetaHuman 虚拟形象和 Apple ARKit 6 所采用的混合形状响应行为，并实现了相当的整体准确度。与 ARKit 6 相比，我们实现的平均均方偏差为 0.019。我们的系统无需大量训练数据即可获得最佳性能。对于缺少 GPU 或类似计算能力、且传感器能力有限的低功耗计算设备，该算法能够提升其虚拟现实能力；这些设备可使用标准网络摄像头，而不必采用 iPhone 所用的集成式激光雷达-红外-深度摄像头系统。我们的算法对硬件的依赖性较低，可为无法部署 ARKit 的 Windows 和 Linux 系统提升虚拟现实能力。

未来的工作可以着重于开发经过优化的加权函数，并以我们自己的关键点回归方法取代 MediaPipe Holistic Model。优化回归的加权函数可以显著改善回归模型的结果，同时在一定程度上弥补数据限制。用更稳健的关键点检测方法取代 MediaPipe 全身模型，可以消除我们在开发过程中观察到的一些模型变形效应。这可以显著提升混合形状估计模型在真实世界条件下的稳健性。总体而言，我们成功地利用规模相对较小的训练数据集开发出了一种实用的混合形状转换算法。

参考文献

[1] 6 2, 7, 8 ARKit - 增强现实 - Apple Developer。V. Bazarevsky、Y. Kartynnik、A. Vakunov、K. Raveendran 和 M. Grundmann。BlazeFace：移动GPU上的亚毫秒级神经网络人脸检测，2019年7月。arXiv:1907.05047 [cs]。doi: 10.48550/arXiv.1907.05047 2, 7[3] V. Blanz 和 T. Vetter。用于三维人脸合成的可变形模型。载于第26届计算机图形学与交互技术年度会议论文集，SIGGRAPH '99，第187–194页。ACM Press/Addison-Wesley Publishing Co.，美国，1999年7月。doi: 10 .1145/311535.311556 2[4] J. Booth、A. Roussos、A. Ponniah、D. Dunaway 和 S. Zafeiriou。大规模三维可变形模型。《国际计算机视觉杂志》，126(2):233–254，2018年4月。doi: 10.1007/s11263-017 -1009-7 2[5] T. S. Breusch 和 A. R. Pagan。异方差和随机系数变异的简单检验。《Econometrica》，47(5):1287–1294，1979年。出版方：[威利、Econometric Society]。doi: 10.2307/ 1911963 4[6] C. Cao、Y. Weng、S. Zhou、Y. Tong 和 K. Zhou。FaceWarehouse：用于可视计算的三维面部表情数据库。《IEEE可视化与计算机图形学汇刊》，20(3):413–425，2014年3月。会议名称：IEEE可视化与计算机图形学汇刊。doi: 10 1109 2013 249 2 . /TVCG. . Z. Cao、G. Hidalgo、T. Simon、S.-E. Wei 和 Y. Sheikh。Openpose：利用部件亲和场进行实时多人二维姿态估计，2019年。2[8]S. Chatterjee。一种新的相关系数。《美国统计学会杂志》，116(536):2009–2022，2021年10月。出版方：泰勒弗朗西斯出版集团；电子版：https://doi.org/10.1080/01621459.2020.1758115。doi: 10.1080/ 01621459 2020 1758115 4, 5, 7, 8 . . T. Cootes、C. Taylor、D. Cooper 和 J. Graham。主动形状模型——其训练与应用。《计算机视觉与图像理解》，61 1 38 59, 1995 10 1006 1995 1004 2 ( ): – 。doi: . /cviu. . P. Dollar、P. Welinder 和 P. Perona。级联姿态回归。载于2010年IEEE计算机学会计算机视觉与模式识别会议论文集，第1078–1085页，2010年。doi: 10.1109/CVPR.2010 .5540094 2[11] J. Durbin 和 G. S. Watson。最小二乘回归中的序列相关检验：I。《Biometrika》，37(3/4):409–428，1950年。出版方：[牛津大学出版社、Biometrika Trust]。doi: 10.2307/2332391 4, 5[12] G. J. Edwards、T. F. Cootes 和 C. J. Taylor。使用主动外观模型进行人脸识别。载于 H. Burkhardt 和 B. Neumann 编，《计算机视觉——ECCV'98》，第581–595页。施普林格柏林海德堡出版社，柏林，海德堡，1998 2 。Epic Games。虚幻引擎，11月。2023 2 。T. Gerig、A. Morel-Forster、C. Blumer、B. Egger、M. Luthi、S. Schoenborn 和 T. Vetter。可变形人脸模型——开放框架。载于2018年第13届IEEE自动人脸与手势识别国际会议（FG 2018）论文集，第75–82页。IEEE，西安，2018年5月。doi: 10.1109/FG.2018.00021 2[15] T. Gerig、A. Morel-Forster、C. Blumer、B. Egger、M. L¨uthi、S. Sch¨onborn 和 T. Vetter。可变形人脸模型——开放框架，2017年9月。arXiv:1709.08398 [cs]。doi: 10.48550/arXiv .1709.08398 2[16] J. C. Gower。广义普氏分析。《Psychometrika》，40 1 33 51, 1975 10 1007 02291478 2 ( ): – 3月。doi: . /BF A. C. HARVEY 和 G. D. A. PHILLIPS。含自回归移动平均扰动项的回归模型的极大似然估计。《Biometrika》，66(1):49–58，1979年4月。doi: 10.1093/biomet/ 66.1.49 6[18] H. Hotelling。两组变量之间的关系。《Biometrika》，28(3/4):321–377，1936年。出版方： [牛津大学出版社、Biometrika Trust] 10 2307 2333955 3 。doi: . / A. G. Howard、M. Zhu、B. Chen、D. Kalenichenko、W. Wang、T. Weyand、M. Andreetto 和 H. Adam。MobileNets：面向移动视觉应用的高效卷积神经网络，2018年4月。2017 1704 04861 [ ] 10 48550 1704 04861 2 . arXiv: . cs . doi: . / arXiv. . P. Huber、G. Hu、R. Tena、P. Mortazavian、W. P. Koppen、W. J. Christ- mas、M. R¨atsch 和 J. Kittler。《多分辨率三维可变形人脸模型及拟合框架》。载于第 11 届计算机视觉、成像与计算机图形学理论及应用联合会议（VISIGRAPP 2016）论文集——第 4 卷：VISAPP，第79–86 页。SciTePress，2016 年。备用出版商：INSTICC ISSN：2184-4321。doi: 10.5220/0005669500790086 2[21] J. S. Hunter。《指数加权移动平均》。Journal of Quality Technology，18(4)：203–210，1986 年 10 月。出版商：泰勒弗朗西斯出版集团 eprint: https://doi.org/10.1080/00224065.1986.11979014。doi: 10.1080/00224065.1986.11979014 6[22] E. Isufi、A. Loukas、A. Simonetto 和 G. Leus。《自回归移动平均图滤波》。IEEE Transactions on Signal Processing，65(2)：274–288，2017 年 1 月。会议名称：IEEE Transactions on Signal Processing。doi: 10 1109 2016 2614793 6 . /TSP. . A. Jandric、T. Skala、M. Mari´cevic 和 L. Horvat。《用于电子游戏开发的风格化和照片级真实感计算机生成三维角色的渲染过程与创建方法及其比较》。Acta graphica：znanstveni ˇcasopis za tiskarstvo i grafi ˇcke komunikacije, 31 2 45 55, 2023 2 ( ): –。出版商：Acta Graphica doo。T. Jiang、P. Lu、L. Zhang、N. Ma、R. Han、C. Lyu、Y. Li 和 K. Chen。《Rtmpose：基于 mmpose 的实时多人姿态估计》，2023 年。2[25] R. A. Johnson 和 D. W. Wichern。《应用多元统计分析》。Prentice Hall，新泽西州 Upper Saddle River，5 2002 4, 5 。版，。R. E. Kalman。《线性滤波与预测问题的新方法》。Transactions of the ASME–Journal of Basic Engineering，82 35 45, 1960 6 (D 系列)：–。Y. Kartynnik、A. Ablavatski、I. Grishchenko 和 M. Grundmann。《利用移动 GPU 从单目视频实时获取面部表面几何》，2019 年 7 月。arXiv: 1907.06724 [cs]。doi：10.48550/arXiv. 1907.06724 2[28] M. Kowalski、J. Naruniec 和 T. Trzcinski。《深度对齐网络：用于鲁棒面部对齐的卷积神经网络》。载于 2017 IEEE计算机视觉与模式识别会议研讨会，第 2034 2043, 2017 10 1109 2017 254 2 – 页。doi: . /CVPRW. . J. P. Lewis、K. Anjyo、T. Rhee、M. Zhang、F. Pighin 和 Z. Deng。《混合形状面部模型的实践与理论》。载于 S. Lefebvre 和 M. Spagnuolo 编，《Eurographics 2014——前沿综述》。Eurographics Association，2014 年。ISSN：1017-4656。doi：10.2312/egst . 20141042 2[30] J. P. Lewis、J. Mooser、Z. Deng 和 U. Neumann。《通过选择性运动衰减减少混合形状干扰》。载于 2005 年交互式三维图形与游戏研讨会论文集，I3D '05，第 25–29 页。计算机协会，纽约州纽约市，美国，2005 10 1145 1053427 1053431 2 年 4 月。doi: . /。C. Lugaresi 、 J. Tang 、 H. Nash 、 C. McClanahan 、 E. Uboweja 、 M. Hays 、 F. Zhang、C.-L. Chang、M. Yong、J. Lee、W.-T. Chang、W. Hua、M. Georg 和 M. Grundmann。《MediaPipe：感知和处理现实的框架》。载于 2019 年 IEEE计算机视觉与模式识别会议（CVPR）增强现实/虚拟现实计算机视觉第三届研讨会，2019 年。2[32] F. I. Parke。《人脸参数模型》 |Computer Science Archive。计算机科学技术报告，犹他大学，1974 年。2[33] P. Paysan、R. Knothe、B. Amberg、S. Romdhani 和 T. Vetter。《用于姿态和光照不变人脸识别的三维人脸模型》。IEEE，意大利热那亚，2009 年。出版物名称：第六届 IEEE 国际高级视频与基于信号的监控会议（AVSS）论文集：智能环境中的安防、安全与监控。2[34] K. Pearson。X.《关于相关变量系统中某一给定偏差系统偏离期望值的判据：该系统可合理地被认为源于随机抽样》。The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science，50(302)：157–175，1900 年 7 月。出版商：泰勒弗朗西斯出版集团 eprint: https://doi.org/10.1080/14786440009463897。doi：10.1080/14786440009463897 4[35] K. Pearson。LIII.《关于空间中点系的最佳拟合直线和平面》。The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science，2(11)：559–572，1901 年 11 月。出版商：泰勒弗朗西斯出版集团 eprint：https://doi.org/10.1080/14786440109462720。doi: 10.1080/14786440109462720 3, 4, 7, 8 2017 1704 04861 [ ] 10 48550 1704 04861 2 M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen。《MobileNetV2：倒置残差与线性瓶颈》。载于 2018 IEEE/CVF计算机视觉与模式识别会议，第 4510–4520 页，2018 年 6 月。ISSN：2575-7075。doi：10.1109/CVPR. 2018.00474 2˙S. S. Shapiro 和 M. B. Wilk。《正态性方差分析检验（完整样本）》。Biometrika，52(3/4)：591–611，1965 年。出版商：[21] 牛津大学出版社，Biometrika Trust[22]。doi：10.2307/ 2333709 4, 5 10 1109 2016 2614793 6 C. Spearman。《两事物间关联的证明与测量》。The American Journal of Psychology，15 101，1904 1 72 10 2307( ): –。出版商：伊利诺伊大学出版社。doi：. / 31 2 45 55, 2023 2 C. Spearman。《根据错误数据计算相关》。British journal of psychology，3(3)：271，1910年。出版商：剑桥大学出版社。7, 8[25] Y. Sun、X. Wang 和 X. Tang。《用于面部关键点检测的深度卷积网络级联》。载于 2013 IEEE计算机视觉与模式识别会议，第 3476–3483 页，2013 年。doi：10.1109/ CVPR.2013.446 2 5 2002 4, 5 A. A. Taha 和 A. Hanbury。《三维医学图像分割的评估指标：分析、选择与工具》。BMC医学影像，82 35 45, 1960 6 : –。出版商：施普林格。U. Technologies。《Unity——手册：实时捕捉》。06724 [B. S. Weir 和 C. C. Cockerham。《用于群体结构分析的 F统计量估计》。Evolution，38(6)：1358–1370，1984 年。出版商：]进化研究学会，威利[28]。doi：10.2307/ 2408641 4 2034 2043, 2017 10 1109 2017 254 2 H. Wu、J. Jia、J. Xing、H. Xu、X. Wang 和 J. Wang。《MMFace4D：用于音频驱动的三维人脸动画的大规模多模态四维人脸数据集》，2023 年 12 月。arXiv：2303.[30] cs 2005 10 1145 1053427 1053431 2 。doi：10.48550/arXiv. 2303.09797 2[32] Z. Yang、A. Zeng、C. Yuan 和 Y. Li。《采用两阶段蒸馏的有效全身姿态估计》， |。K. Zhang、Z. Zhang、Z. Li 和 Y. Qiao。《使用多任务级联卷积网络联合进行人脸检测与对齐》。IEEE信号处理快报，23(10)：1499–1503，2016 年。doi：10.1109/LSP.2016 .2603342 2

tems of points in space。《空间中的点集》。The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11)：559–572，1901 年 11 月。出版商：泰勒弗朗西斯出版集团 eprint：https://doi.org/10.1080/14786440109462720。doi：10.1080/14786440109462720 3, 4, 7, 8[36] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen。《MobileNetV2：倒置残差与线性瓶颈》。载于 2018 IEEE/CVF计算机视觉与模式识别会议，第 4510–4520 页，2018 年 6 月。ISSN：2575-7075。doi：10.1109/CVPR. 2018.00474 2[37] S. S. Shapiro 和 M. B. Wilk。《正态性方差分析检验（完整样本）》。Biometrika，52(3/4)：591–611，1965 年。出版商：[牛津大学出版社，Biometrika Trust]。doi：10.2307/ 2333709 4, 5[38] C. Spearman。《两事物间关联的证明与测量》。The American Journal of Psychology，15 101，1904 1 72 10 2307( )：–。出版商：伊利诺伊大学出版社。doi：./ 1412159 7 [39] C. Spearman。《根据错误数据计算相关》。British jour- nal of psychology, 3(3)：271，1910 年。出版商：剑桥大学出版社。7, 8[40] Y. Sun、X. Wang 和 X. Tang。《用于面部关键点检测的深度卷积网络级联》。载于 2013 IEEE计算机视觉与模式识别会议，第 3476–3483 页，2013 年。doi：10.1109/ CVPR.2013.446 2[41] A. A. Taha 和 A. Hanbury。《三维医学图像分割的评估指标：分析、选择与工具》。BMC医学影像，15 1 28, 2015 8：–。出版商：施普林格。U. Technologies。《Unity——手册：实时捕捉》。2 [42] B. S. Weir 和 C. C. Cockerham。《用于群体结构分析的 F统计量估计》。Evolution，38(6)：1358–1370，1984 年。出版商：[进化研究学会，威利]。doi：10.2307/2408641 4[43] H. Wu、J. Jia、J. Xing、H. Xu、X. Wang 和 J. Wang。《MMFace4D：用于音频驱动的三维人脸动画的大规模多模态四维人脸数据集》，2023 年 12 月。arXiv：2303.09797 [cs]。doi：10.48550/arXiv. 2303.09797 2[44] Z. Yang、A. Zeng、C. Yuan 和 Y. Li。《采用两阶段蒸馏的有效全身姿态估计》，2023 2。K. Zhang、Z. Zhang、Z. Li 和 Y. Qiao。《使用多任务级联卷积网络联合进行人脸检测与对齐》。IEEE信号处理快报，23(10)：1499–1503，2016 年。doi：10.1109/LSP.2016 .2603342 2 [45] [46]