DOI:10.3969/j.issn.1003-5060.2024.03.010

基于单视图的带纹理三维人体网格参数化重建

邢 燕[1], 徐 冬[1], 洪沛霖[2], 檀结庆[1]
(1.合肥工业大学 数学学院,安徽 合肥 230601; 2.安徽中医药大学 医药信息工程学院,安徽 合肥 230012)

摘 要:针对计算机视觉中的三维人体重建问题,文章提出一种端到端的的网络框架,在三维和二维混合监督下,从单幅彩色图像重建带纹理信息的精准三维人体网格。使用 4 个编码器分别提取形状姿态特征、纹理特征、光照参数和像机参数,得到的图像特征被送入三维回归模块,迭代推断出三维人体参数;纹理参数送入纹理解码器网络得到纹理图;学习到的人体参数可转化为三维人体网格;对于损失函数的设置,预测的人体网格顶点与真实顶点的差值用来进行三维监督;通过预测的像机参数、光照参数和纹理计算二维渲染损失;通过三维关节投射得到的二维关节与图像上的二维关节真值计算二维关节重投影损失;生成对抗网络的鉴别器使得渲染图像更加真实。该文方法与现有的三维人体重建方法相比具有竞争力,而且重建的三维人体网格带有纹理信息。

关键词:三维人体重建;深度学习;蒙皮多人线性(SMPL)模型;形状姿态;纹理
中图分类号:TP391.41 **文献标志码**:A **文章编号**:1003-5060(2024)03-0347-07

Parameter reconstruction of 3D textured human mesh based on single image

XING Yan[1], XU Dong[1], HONG Peilin[2], TAN Jieqing[1]
(1. School of Mathematics, Hefei University of Technology, Hefei 230601, China; 2. School of Medical Information Engineering, Anhui University of Chinese Medicine, Hefei 230012, China)

Abstract: Aiming at the problem of 3D human reconstruction in computer vision, an end-to-end network framework is proposed to reconstruct accurate 3D human mesh with texture from a single color image under the hybrid supervision of 3D and 2D. In this paper, four encoders are used to extract shape and pose features, texture features, illumination parameters and camera parameters respectively. And the features obtained are sent to the 3D regression module to iteratively infer the parameters of the 3D human model. Texture parameters are fed into the texture decoder network to obtain texture maps. The learned human model parameters can be transformed into the 3D human mesh. For the setting of loss function, the difference between the predicted human mesh vertex and the ground truth is used for 3D supervision. The 2D rendering loss is calculated by predicted camera parameters, illumination parameters and mapped texture. The 2D joint reprojection loss is calculated by projecting the 3D joints to the 2D joints and then comparing with the ground truth. The discriminator of generative adversarial network(GAN) is used to make the rendered images more realistic. The qualitative and quantitative experimental results show that the proposed method achieves comparable performance with some state-of-the-art 3D human reconstruction methods. Moreover, the reconstructed 3D human mesh possesses the corresponding texture map according to the input human image.

Key words: 3D human reconstruction; deep learning; skinned multi-person linear (SMPL) model;

收稿日期:2023-02-20;**修回日期**:2023-03-07
基金项目:国家自然科学基金资助项目(62172135);合肥工业大学校级教研资助项目(KCSZ2022034);安徽中医药大学教研重点资助项目(2020xjjy_zd005)
作者简介:邢 燕(1977—),女,安徽合肥人,博士,合肥工业大学副教授,硕士生导师;
洪沛霖(1977—),男,安徽合肥人,安徽中医药大学讲师,通信作者,E-mail:hongpeilin@ahtcm.edu.cn;
檀结庆(1962—),男,安徽望江人,博士,合肥工业大学教授,博士生导师.

shape and pose; texture

0 引 言

三维建模在动画、服装设计、游戏、虚拟现实等领域有着广泛的应用。传统上,基于优化的方法[1-2]为由图像恢复姿态和形状提供了可行的解决方案,然而运行时间慢、依赖良好初始化、不正确的局部最小值会导致不精确、不稳定、鲁棒性差等。最近的研究重点转移到基于学习的方法上,研究用体素、点云、网格、隐式函数等表示重建三维对象的方法。由于从单幅二维图像获取三维形状的歧义性和关节动物的复杂性,从单幅二维图像重建三维人体模型是一个巨大的挑战。

三维人体重建可以分为参数化重建和非参数化重建 2 种方法。传统的非参数化方法一般需要借助激光扫描仪、深度相机等特殊的数据收集设备,而且容易受到噪声的影响。文献[3]使用深度相机从 Kinect 传感器获得的数据中恢复出三维人体;文献[4]根据 2 台 Kinect 传感器同时扫描获得的数据,提出一种基于全局配准的三维人体重建方法。深度神经网络的出现为三维人体重建提供了新的思路。一般对网格顶点直接变形,虽然重建的人体具有个性化和灵活性,但顶点变形不受约束,会出现变形异常、不符合人体测量学的情况,例如出现人体上不可能的关节角度或极瘦的身体等不合理现象。文献[5]使用基于图像的卷积神经网络,通过一系列的图卷积层对变形网格的三维顶点坐标进行回归;文献[6]通过在图卷积神经网络的损失函数中引入拉普拉斯先验和部分分割损失实现三维人体重建。

参数化人体重建方法仅需 1 组低维向量参数即可描述人体形状,其中常见的人体参数化模型有智能分类器和姿态估计(smart classifier and pose estimator,SCAPE)模型[7]和蒙皮多人线性(skinned multi-person linear,SMPL)模型[8]。文献[9]使用 SMPL 参数化人体模型,在网络中使用基于距离场的碰撞损失和深度排序感知损失估计三维人体姿态和形状;文献[10]提出一种采用 SMPL 模型的多级拓扑构建的图卷积神经网络重建三维着装人体;文献[11]以人体的二值化图像为输入,以人体形状参数误差、正/侧面轮廓误差为损失函数实现三维人体重建;文献[12]采用深度神经网络作为编码器、SMPL 模型作为解码器,提出去噪自编码器模块,从结构误差中恢复人体。

文献[1-2]采用多阶段方法恢复三维人体网格:首先估计二维关节位置,然后根据这些信息估计三维模型参数。这种分阶段的方法通常不是最优的,本文提出一个端到端的解决方案,直接学习从图像像素到人体模型参数的映射。

文献[13]提出生成对抗网络(generative adversarial network,GAN),GAN 在医学、自动驾驶、地理、图像处理等领域取得了令人瞩目的成就;文献[14]使用判别器鉴别生成的渲染图像和真实输入图像,重建鸟、牛、摩托车等对象;文献[15]使用判别器判断人体形状和姿势参数是否与真值相同,重建三维人体。这些研究都给本文提供了新的思路。本文采用 GAN 鉴别器作为弱监督,判断图像是来自真实输入图像还是渲染图像,从而使重建的纹理效果更加精确。

本文的主要工作如下:

1) 提出一个端到端的框架,与多阶段方法不同,可直接从单张二维图像恢复三维人体网格。

2) 大部分三维人体重建结果都是无纹理的,而本文不仅重建出三维人体的形状和姿态,还从二维 RGB 图像中学习到纹理图。

3) 设计三维损失、二维重投影损失、二维渲染损失和生成对抗损失的组合损失,提高了三维人体形状和姿态重建精度并生成较合理的纹理图。

1 方 法

本文以单视图为输入,利用编码器、解码器网络回归 SMPL 模型参数,借助 GAN 网络重建带纹理的三维人体网格。

1.1 模型管线图

本文的网络结构框架如图 1 所示。从输入的单幅 RGB 图像中提取特征,采用相机、纹理、光照以及形状编码器提取相机参数、纹理特征、光照参数和形状特征。形状编码器采用预训练的 ResNet-50[16]对图像进行编码,得到的图像特征被送入三维回归模块,推断出三维人体 SMPL 模型参数;再用 SMPL 模型重建三维人体网格。相机、光照、纹理编码器都使用了二维卷积块、平均池化层以及线性层的结构,得到相应的属性参数。其中,纹理参数再被送入纹理解码器网络,得到纹理图。

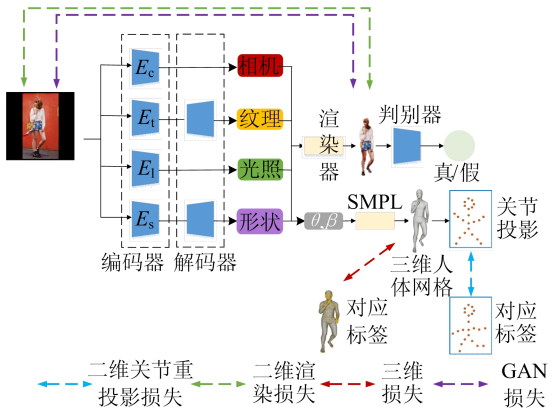


图 1　网络结构的框架

本文将二维和三维约束相结合,使用三维损失、二维损失、GAN 损失,二维损失使用关节点重投影损失,图像渲染损失。从这些约束条件出发,本文的网络结构可以精准重建出带纹理的三维人体形状和姿态。此外,本文模型在预测和训练阶段皆不需要真实的相机外参和光照系数,就能从输入的 RGB 图像中回归出相机参数和光线参数,这使得本文方法应用更加广泛。

1.2　属性编码器

设 $\boldsymbol{I}\in\mathbf{R}^{H\times W\times 3}$ 表示高为 H、宽为 W 的彩色输入图像,$O(\boldsymbol{S},\boldsymbol{T})$ 表示三维人体网格,$\boldsymbol{S}=(\boldsymbol{\theta},\boldsymbol{\beta})$ 为形状属性,其中 $\boldsymbol{\theta}\in\mathbf{R}^{3K}$ 和 $\boldsymbol{\beta}\in\mathbf{R}^{10}$ 分别表示位姿参数和形状参数。纹理属性 $\boldsymbol{T}\in\mathbf{R}^{H\times W\times 3}$ 表示分辨率为 $H\times W$ 的 UV 图。$\boldsymbol{C}=(a,e,d)$ 为相机属性,其中 $a\in[0°,360°]$、$e\in[-90°,90°]$、$d\in[0,+\infty]$ 分别表示方位角、仰角和距离参数。光照属性 $\boldsymbol{L}\in\mathbf{R}^{l}$ 由球谐函数建模,它由一个不同的角频率的球面基组成,l 是球谐函数系数的维数。

本文通过 4 个子编码器独立预测属性,相机编码器 E_c 预测 1 个由 (a_x,a_y,e,d) 组成的 4 维向量,其中:e 和 d 分别表示相机的仰角和距离参数;a_x 和 a_y 表示方位角的笛卡尔坐标,$a=\mathrm{atan2}(a_x,a_y)$。用这种方式计算方位角参数,可以避免在定义域[0°,360°]中出现不连续的回归问题。用 ResNet-50 作为形状编码器 E_s 的网络,得到的图像特征被送入三维回归模块,迭代推断出三维人体 SMPL 模型参数 $\boldsymbol{\theta}\in\mathbf{R}^{3K}$ 和 $\boldsymbol{\beta}\in\mathbf{R}^{10}$,其中 $K=23$ 个关节点。对于纹理编码器 E_t,本文不是通过编解码器模型直接输出纹理 UV 图,而是先预测一个二维流图,然后应用空间变换生成纹理 UV 图 $\boldsymbol{T}$。对于光照属性,子编码器 E_l 直接编码一个 l 维向量,作为球谐函数模型系数。

输入图像 $\boldsymbol{I}_i\in\mathbf{R}^{H\times W\times 3}$,$i=1,2,\cdots,N$,其中 N 为训练样本的数量。三维人体网格重建是训练编码器 E_η,从单幅图像预测三维人体网格属性:

$$\boldsymbol{A}_i=(\boldsymbol{C}_i,\boldsymbol{L}_i,\boldsymbol{S}_i,\boldsymbol{T}_i)=E_\eta(\boldsymbol{I}_i)\tag{1}$$

其中:η 为编码器的可训练参数;E_η 为 4 个子编码器的并集。

1.3　可微渲染

给定三维属性 $\boldsymbol{A}=(\boldsymbol{C},\boldsymbol{L},\boldsymbol{S},\boldsymbol{T})$,在相机视图 $\boldsymbol{C}$ 和光照环境 $\boldsymbol{L}$ 下,三维人体网格 $O(\boldsymbol{S},\boldsymbol{T})$ 可以渲染为二维图像。三维人体的渲染过程可表示为:

$$\boldsymbol{I}_r=R(\boldsymbol{A})=R(\boldsymbol{C},\boldsymbol{L},\boldsymbol{S},\boldsymbol{T})\tag{2}$$

其中:$\boldsymbol{I}_r$ 为输出的渲染图像;R 为可微渲染器,不包含任何可训练参数。

1.4　回归的 SMPL 模型参数

SMPL 模型是一种参数化人体模型,可以进行任意的人体建模和动画驱动。SMPL 模型又是一种生成模型,通过参数可对人体的形状和姿势进行调整,如人体的身高、体重、身体比例以及三维表面随关节的变形。形状参数 $\boldsymbol{\beta}\in\mathbf{R}^{10}$ 由形状的主成分分析空间的前 10 个系数组成。位姿参数 $\boldsymbol{\theta}\in\mathbf{R}^{3K}$ 由 $K=23$ 个关节以轴角表示的相对三维旋转组成。SMPL 模型由 $N=6\ 980$ 个顶点的三角网格 $M(\boldsymbol{\theta},\boldsymbol{\beta})\in\mathbf{R}^{3N}$ 组成,通过调整参数 $\boldsymbol{\beta}$、$\boldsymbol{\theta}$,根据关节旋转角度 $\boldsymbol{\theta}$ 进行正运动学拼接,再用线性混合蒙皮对曲面进行变形获得。

1.5　损失

本文整体的训练损失 L 包括二维损失 L_{2D}、三维损失 L_{3D}、生成对抗损失 L_{gan} 3 个部分,即

$$L=\lambda_1L_{2D}+\lambda_2L_{3D}+\lambda_3L_{gan}\tag{3}$$

其中,λ_1、λ_2、λ_3 为权值常数。

1.5.1　二维损失

二维损失 L_{2D} 由关节重投影损失 L_{2D}^{joints} 和渲染损失 $L_{2D}^{rendering}$ 组成,即

$$L_{2D}=\mu_1L_{2D}^{joints}+\mu_2L_{2D}^{rendering}\tag{4}$$

三维关节点 $X(\boldsymbol{\theta},\boldsymbol{\beta})\in\mathbf{R}^{3K}$ 重投影的二维关节点为:

$$\hat{\boldsymbol{x}}=\Pi(X(\boldsymbol{\theta},\boldsymbol{\beta}))\tag{5}$$

其中,Π 为投影运算。

为了使关节重投影误差最小,人体位姿更准确,设置了关节重投影损失,即

$$L_{2D}^{joints}=\frac{1}{K}\sum_{i}^{K}\|v_i(\boldsymbol{x}_i-\hat{\boldsymbol{x}}_i)\|_1\tag{6}$$

其中:$\boldsymbol{x}_i\in\mathbf{R}^{2K}$ 是第 i 个真实的二维关节点;$v_i\in\{0,1\}$ 表示每个模型中 K 个关节的可见性,若可见,则为 1,否则为 0。

本文希望渲染的图像和输入的图像是接近

的,因此定义二维渲染损失:

$$L_{2D}^{\text{rendering}}=\frac{1}{N}\sum_{i}^{N}\|R(E_{\eta}(\boldsymbol{I}_i))-\boldsymbol{I}_i\|_1 \quad (7)$$

其中:$\boldsymbol{I}_i$ 为第 i 幅图像;E_{η} 为 E_c、E_l、E_t 和 E_s 的并集;$R(\cdot)$为可微渲染器,它将二维图像空间与三维属性空间连接起来。

1.5.2 三维损失

重投影损失使得神经网络生成一个三维人体,解释二维关节位置;然而人体测量学上不合理的三维物体或具有自交的物体可能使重投影损失最小化。因此本文引入三维损失:

$$L_{3D}=\frac{1}{N}\sum_{i}^{N}\|\boldsymbol{V}_i-\hat{\boldsymbol{V}}_i\|_2 \quad (8)$$

其中:$\boldsymbol{V}\in\mathbf{R}^{3N}$为为输入图像对应的三维人体的真实顶点;$\hat{\boldsymbol{V}}\in\mathbf{R}^{3N}$为预测的三维网格顶点。

1.5.3 生成对抗损失

本文使用生成对抗性损失训练模型中4个子网络,使得重建的三维人体渲染图更接近输入图像。本文模型的生成对抗损失定义为:

$$L_{\text{gan}}=L_G+L_D \quad (9)$$

其中

$$L_G=E_{P_z}[(1-D(R(E_{\eta}(\boldsymbol{I}))))^2] \quad (10)$$

$$L_D=E_{P_z}[D(R(E_{\eta}(\boldsymbol{I})))^2]+E_{P_{\text{data}}}[(1-D(\boldsymbol{I}))^2] \quad (11)$$

其中:$G(\cdot)$生成器;$D(\cdot)$为判别器;P_{data}、P_z 分别为真实数据分布和生成数据分析。

2 实 验

2.1 数据集和评估指标

2.1.1 数据集

本文使用UP-3D数据集[2]和THuman数据集[17]进行单幅二维图像的三维人体重建实验。在UP-3D数据集中,使用了8 000多个数据项,每项包含1张RGB图像,对应的关节点和真实的SMPL参数。THuman数据集大约有7 000个数据项,剔除了多人图像,在训练中使用了其中4 000多个数据项。大约按3∶1划分训练集和测试集。

2.1.2 评估指标

在定量评估时,使用逐顶点误差平均值(mean per vertex error, mPVE)作为第一评价标准,计算真实顶点与预测顶点之间的欧氏距离;倒角(Chamfer)距离作为第二评价标准,用来测量预测点集和真实点集(P_1、P_2)间的相似性,定义如下:

$$d_{\text{Chamfer}}(P_1,P_2)=\frac{1}{|P_1|}\sum_{\boldsymbol{p}_1\in P_1}\min_{\boldsymbol{p}_2\in P_2}\|\boldsymbol{p}_1-\boldsymbol{p}_2\|_2+\frac{1}{|P_2|}\sum_{\boldsymbol{p}_2\in P_2}\min_{\boldsymbol{p}_1\in P_1}\|\boldsymbol{p}_2-\boldsymbol{p}_1\|_2 \quad (12)$$

此外,还用二维图像的评估指标FID(fréchet inception distance)[18]评估渲染图像。FID计算真实的输入图像和生成的渲染图像在特征空间的距离,FID值越低意味着图像的质量越高。本文从0°~360°每30°一个间隔的视角计算FID的平均值。这3个评估指标数值越小表示结果越好。

如文献[19]所述,常用的度量不能完全反映几何重建的质量,如表面的平滑度和连续性。因此本文认为视觉效果也很重要。

2.2 实验设置

本文的网络是使用Pytorch实现的,形状子网络的图像编码器使用在ImageNet[20]上预训练的ResNet-50模型。在训练中,人形模板网格有$N=6\,980$个顶点和13 776个面,输入图像的分辨率为128×128。本文使用$\beta_1=0.5$、$\beta_2=0.999$的Adam优化器,批处理大小设置为16,学习率初始化为1×10^{-4},学习率衰减策略使用余弦退火学习率。整个训练过程一共500个epoch,每20个epoch进行一次测试。本文网络框架中的一些超参数设置为:$\lambda_1=1$,$\lambda_2=1$,$\lambda_3=0.000\,1$,$\mu_1=1$,$\mu_2=0.1$。

2.3 实验结果及比较

2.3.1 定量结果

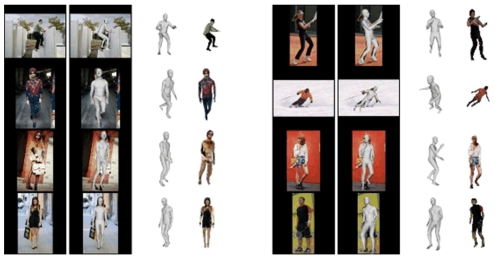
本文定量比较的结果见表1所列。本文使用mPVE和Chamfer作为度量标准,卷积网格回归(convolutional mesh regression,CMR)方法[5]的结果是最好的,本文结果是次好的,与CMR的数据非常接近;接着是人体网格重建(human mesh recovery,HMR)方法[15]和SMPLify-X[21];最后是分层网格变形(hierarchical mesh deformation,HMD)方法[22]。CMR方法是非参数方法,它利用图卷积对所有顶点进行变形,定量结果相对参数方法更好一些,但顶点变形使人体表面不太光滑,也可能会出现奇异点,视觉效果不太理想。而使用SMPL模型参数是为了在任何情况下重建视觉上有说服力的人体几何结构。从定量结果来看,本文方法优于HMR、SMPLify-X、HMD 3种参数化方法的。对比的4种研究结果取自于文献[5,15,21-22]。

表 1　与先进方法的定量比较

方法	mPVE	Chamfer 距离
HMD	0.349	0.336
SMPLify-X	0.269	0.178
HMR	0.242	0.211
CMR	0.173	0.155
本文方法	0.179	0.167

2.3.2　定性结果

数值指标不能完全反映重建质量,因为一些具有良好数值指标的重建结果在视觉上并不令人满意,所以在与其他方法比较时,不仅要定量比较数值结果,而且要注意视觉效果的比较。在 UP-3D 数据集上,本文方法重建的三维人体模型如图 2所示。图 2a、图 2b 中:第 1 列表示输入图像;第 2 列表示图像视角观察到的重建的三维人体网格;第 3 列表示其他视角观察到的重建的三维人体网格;第 4 列表示带纹理的三维人体。从图 2 可以看出,本文方法不仅能从 RGB 图像重建三维人体的形状和姿态,而且能根据输入的 RGB 图像重建三维人体的纹理,而作为比较的其他方法都没有给出人体的纹理。参数化方法重建的人体模型更加光滑和完整。



(a)　　(b)

图 2　本文方法的可视化结果

HMR 方法、CMR 方法、CMR 延伸的参数化方法和本文方法的定性比较如图 3 所示,可以从视觉上直观看出三维人体重建结果的差异。图 3 中:第 1 行是输入图像;第 2、第 3 行是 HMR 方法重建的相应视角及其他视角结果;第 4、第 5 行是 CMR 方法重建的三维结果;第 6、第 7 行是 CMR 延伸的参数化方法重建的三维结果;第 8～第 10 行是本文方法重建的三维结果和二维纹理展示。由图 3 的第 1、第 2 列可以看出:CMR 非参数化方法的一个缺点,即当人物在图像中占比很小时,CMR 方法重建的结果可能连基本人体形状都没有;虽然 HMR 方法和 CMR 延伸的参数化方法可以重建三维人体,但姿态不正确;而本文方法不仅重建正确的三维人体形状和姿态,还学习到人体纹理。由图 3 的第 3 列可以看出,本文方法重建的人体双臂是自然下垂的,其他 3 种方法重建的胳膊姿态都不正确,其他视角也可以佐证。图 3 第 4 列的 HMR 方法的人体腿部没有像输入图像那样交叉站立。由图 3 第 5、第 6 列 CMR 方法的其他视角图上的人体腿部可观察到非参数方法的另一个缺点,即重建的身体不是很光滑。

总体来看,虽然 CMR 非参数化方法的数值结果最好,但视觉效果还有很多不足。与这些方法相比,本文方法重建小尺寸人物优于其他方法,重建中大尺寸人物也具有竞争力,此外还给出人体的纹理。

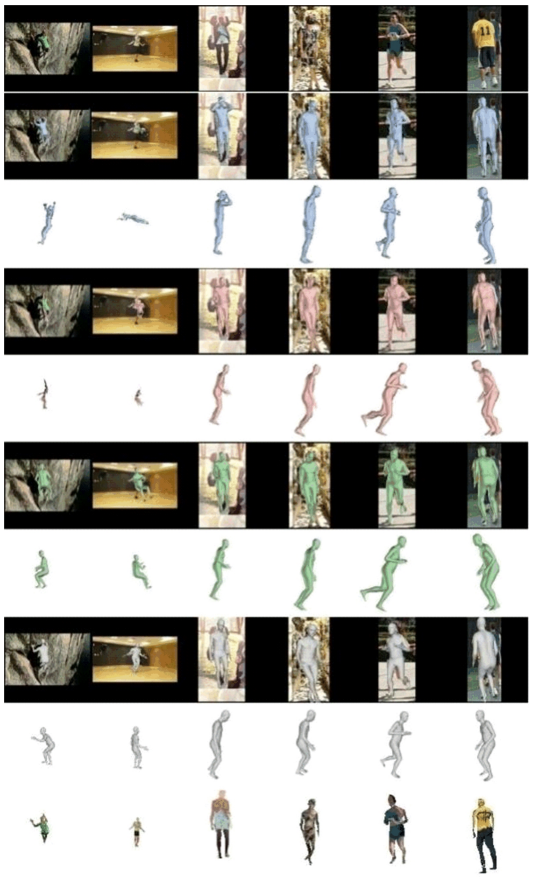


图 3　本文与其他方法的定性比较

2.4　消融实验

本节描述消融实验,说明不同损失项(二维渲染损失二 $L_{2D}^{rendering}$、二维关节重投影损失 、三维损失 $L_{3D}^{L_2}$和 GAN 损失 L_{gan})在本文网络结构中的重要性。消融实验的设置见表 2 所列。表 2 中的每行都代表不同损失项的组合,“√”表示该组合选中的上方损失项。根据表 2 的损失设置,在测试

集上不同损失组合的 mPVE、Chamfer 距离和 FID 的结果见表 3 所列。消融实验的可视化结果如图 4 所示,图 4 中第 1 行表示无背景的输入图像,第 2～第 7 行对应表 2 和表 3 中组合 1～组合 6 不同损失组合的可视化结果。

表 2 消融实验设置

组合	$L_{2D}^{rendering}$	L_{2D}^{joints}	$L_{3D}^{L_2}$	$L_{3D}^{Chamfer}$	L_{gan}
1			√		
2	√		√		
3	√			√	
4	√				√
5	√	√	√		
6	√	√	√		√

表 3 不同损失组合的消融研究

损失组合	mPVE	Chamfer 距离	FID
1	0.190	0.233	161.8
2	0.199	0.237	56.9
3	0.766	0.113	137.0
4	0.476	0.279	86.4
5	0.181	0.184	55.9
6	0.179	0.167	54.3

在消融实验中,首先仅考虑三维监督,用顶点误差的 2-范数作为损失函数,这时三维损失较小但渲染损失很大(见表 3 损失组合 1);视觉效果见图 4 第 2 行,只用三维监督,纹理不正确,相机视角也没有被正确估计。接着加上了二维渲染损失,表明渲染质量的 FID 指标有明显降低,但是三维质量的误差稍微变大(见表 3 组合 2);图 4 第 3 行也表明这时重建出较正确的三维形状、姿态和纹理。为了对比 2-范数和 Chamfer 距离在三维损失中的效果,表 3 组合 3 把三维监督中的 2-范数换成 Chamfer 距离,Chamfer 距离是找重建点集和真实点集之间每个顶点和最近顶点的双向距离的平均值,这时 Chamfer 距离明显降低,而 mPVE 和 FID 都显著增大;对应的图 4 第 4 行中人体的头部肩部会出现相连的情况,这是因为在特征显著或细节丰富而顶点不够密集的区域,Chamfer 距离作为损失过分追求损失最小化造成人体形状和姿势的扭曲。再考虑不用三维监督的情况,只用二维渲染损失和一个生成对抗损失。这时,因为缺少三维损失,三维指标和图像指标与组合 2 相比都变差(见表 3 组合 4);由图 4 第 5 行也可以看出重建结果缺少人体胳膊纹理。在组合 1～组合 4 中,组合 2 综合指标最优,因此本文最终选择了二维和三维混合监督的方法,并且三维损失中使用 2-范数。但由于形状和姿态与输入图像(图 4 第 1 行)还不是特别一致,尝试加上二维关节重投影损失(见表 3 组合 5),所有指标都进一步改善。最后添加生成对抗损失,所有指标达到最优(见表 3 组合 6),视觉效果也有所改善(见图 4 第 7 行)。

图 4 具有不同损失的消融研究结果的可视化结果

3 结 论

本文提出一种端到端的网络框架,在三维损失、二维渲染损失、二维关节重投影损失和 GAN 损失的混合监督下,利用 4 个子网络,在允许没有真实相机参数和照明系数的情况下,可以重建带

纹理信息的精准三维人体网格。在后续工作中,将探索在没有三维监督的情况下重建具有纹理且形状姿态准确的三维人体网格。

[参　考　文　献]

[1] BOGO F, KANAZAWA A, LASSNER C, et al. Keep it SMPL: automatic estimation of 3D human pose and shape from a single image[C]//European Conference on Computer Vision. Switzerland: Springer, 2016: 561-578.

[2] LASSNER C, ROMERO J, KIEFEL M, et al. Unite the people: closing the loop between 3D and 2D human representations[C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2017: 4704-4713.

[3] 周瑾,潘建江,童晶,等. 使用 Kinect 快速重建三维人体[J]. 计算机辅助设计与图形学学报,2013,25(6):873-879.

[4] 孙瑜亮,缪永伟,鲍陈,等. 基于全局配准累积误差极小的人体 RGB-D 数据三维重建[J]. 计算机辅助设计与图形学学报,2019,31(9):1467-1476.

[5] KOLOTOUROS N, PAVLAKOS G, DANIILIDIS K. Convolutional mesh regression for single-image human shape reconstruction[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4496-4505.

[6] LIN K, WANG L, JIN Y, et al. Learning nonparametric human mesh reconstruction from a single image without ground truth meshes[C]//IEEE International Conference on Image Processing. Anchorage: IEEE, 2021: 964-968.

[7] ANGUELOV D, SRINIVASAN P, KOLLER D, et al. SCAPE: shape completion and animation of people[J]. ACM Transactions on Graphics, 2005, 24(3): 408-416.

[8] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: a skinned multi-person linear model[J]. ACM Transactions on Graphic, 2015, 34(6): 248.

[9] JIANG W, KOLOTOUROS N, PAVLAKOS G, et al. Coherent reconstruction of multiple humans from a single image[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5578-5587.

[10] 毛爱华,禚冠军,禤骏. 采用多级拓扑图卷积网络的可变形三维着装人体重建[J]. 计算机辅助设计与图形学学报,2022,34(12):1899-1910.

[11] 许豪灿,李基拓,陆国栋. 由 LeNet-5 从单张着装图像重建三维人体[J]. 浙江大学学报(工学版),2021,55(1):153-161.

[12] MADADI M, BERTICHE H, ESCALERA S. SMPLR: deep learning based SMPL reverse for 3D human pose and shape recovery[J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2020(106): 106.

[13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//27th International Conference on Neural Information Processing Systems. Cambridge: ACM, 2014: 2672-2680.

[14] HU T, WANG L, XU X, et al. Self-supervised 3D mesh reconstruction from single images[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 5998-6007.

[15] KANAZAWA A, BLACK M, JACOBS D, et al. End-to-end recovery of human shape and pose[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7122-7131.

[16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[17] ZHENG Z, TAO Y, WEI Y, et al. DeepHuman: 3D human reconstruction from a single image[C]//IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 7738-7748.

[18] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[C]//31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6629-6640.

[19] WANG N, ZHANG Y, LI Z, et al. Pixel2Mesh: 3D mesh model generation via image guided deformation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43: 3600-3613.

[20] JIA D, WEI D, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255.

[21] PAVLAKOS G, CHOUTAS V, GHORBANI N, et al. Expressive body capture: 3D hands, face, and body from a single image[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 10967-10977.

[22] ZHU H, ZUO X, WANG S, et al. Detailed human shape estimation from a single image by hierarchical mesh deformation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4491-4500.

(责任编辑　朱晓临)