

基于形状和边判别的三维模型无监督配准*

蒋宇涛,刁俊淇,肖俊[†],王颖

(中国科学院大学人工智能学院, 北京 100049)
(2024年3月24日收稿; 2024年5月23日收修改稿)

Jiang Y T, Diao J Q, Xiao J, et al. Unsupervised 3D registration of articulated shapes by shape and edge regularization [J]. Journal of University of Chinese Academy of Sciences, 2026, 43 (5): 657-666. DOI: 10. 7523/j. ucas. 2024. 056.

摘要 非刚性三维模型的配准在动画驱动、纹理传递、语义理解等不同领域具有广泛的应用,是计算机图形学的重要研究问题。然而基于骨架变形的非刚性物体,如人体、动物等,往往按照关节的运动变化带动整体的形状变形,这种复杂的相互作用给三维配准任务带来了困难和挑战。尽管近期的研究利用深度学习技术在重建具有骨架结构的三维形状时可以取得不错的效果,但这些方法往往严重依赖于大型带标注的数据集。在无监督的情况下,依然容易生成过度拉伸和部分交叉的模型网格。针对三维模型的配准问题,提出一种基于生成式对抗网络的无监督网络框架,利用三维生成器生成拟合扫描点云的模型网格,同时使用2个新型的判别器对变形网格进行正则化约束,该方法在公开数据集上取得了较好的实验效果。

关键词 三维重建; 三维配准; 对抗网络; 非刚性物体

中图分类号: TP183 文献标志码: A DOI: 10. 7523/j. ucas. 2024. 056

Unsupervised 3D registration of articulated shapes by shape and edge regularization

JIANG Yutao, DIAO Junqi, XIAO Jun, WANG Ying

(School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract Registration of non-rigid 3D models has a wide range of applications in different application areas such as animation driving, texture transfer, semantic understanding, etc. and is an important research topic in computer graphics. However, non-rigid objects based on skeletal deformation, such as human bodies and animals, often induce overall shape deformations in accordance with the movements of the joints. The non-rigid deformations and complex interactions of such objects thus pose difficulties and challenges to the 3D alignment task. Although recent studies using deep learning techniques can achieve good results in reconstructing 3D shapes with skeleton

* 国家自然科学基金(U21A20515, 62102393, 62206263, 62271467)、北京市自然科学基金(4242053)、中国博士后科学基金(2022T150639, 2021M703162)、机器人技术与系统国家重点实验室开放研究课题(SKLRIS-2022-KF-11)和中央高校基本科研业务费专项资金资助

[†] 通信作者, E-mail: xiaojun@ucas. ac. cn

structures, these methods tend to rely heavily on large datasets with annotations. In the unsupervised case, they are still prone to generating overstretched and partially crossed model meshes. The paper proposes an unsupervised GAN-based network framework for the registration of 3D models, using a 3D generator to generate model meshes that fit the scanned point clouds, and two novel discriminators to impose regularization constraints on the deformed meshes, which achieves good results on public datasets.

Keywords 3D reconstruction; 3D registration; adversarial networks; non-rigid objects

获取三维形状的密集对应关系在许多多媒体应用中发挥着至关重要的作用,包括三维动画、模型迁移以及从大量扫描数据中学习通用统计模型的工作,如带关节与非刚性形变手部模型(hand model with articulated and non-rigid deformations, MANO)^[1]、带蒙皮多人线性人体模型(skinned multi-person linear model, SMPL)^[2]和 FaceScape^[3]。在这些工作中,首先将模板拟合到不同拓扑的扫描数据的形状,然后利用模板的统一拓扑,通过最近邻的方式提取模型之间的密集对应关系。通过这样的方法,可以让不同模型网格上的对应点共享一致的语义。然而,由于扫描数据中经常存在噪声和肢体缺失,基于模板的配准工作仍然面临许多挑战,尤其是在涉及铰接形状的点云配准任务时,大幅度的非刚性变形和复杂的关节运动,使得此类三维配准任务更加困难。在实际应用中,通常需要使用大量人工选择的模型点对作为先验信息,才能得到符合预期的模型映射关系^[4]。

为了在减轻繁琐劳动的同时,获得高质量的对应关系,研究人员在过去几十年间开发了多种非刚性物体的配准方法。例如最优步长非刚性迭代最近点算法(optimal step nonrigid iterative closest point, NICP)^[5],通常适用于模板和原始数据大致拟合的情景,因其连续帧非常相似而被广泛应用于四维跟踪任务。然而通常情况下,当模板和原始数据之间存在着较大的形状和姿态差异时,这类方法很容易发生错误。另外一些方法利用参数化表示^[6-7]获得了相对更好的拟合效果,然而这种参数表示的获取途径非常困难。最近的一些研究通过深度学习技术提高了铰接形状的建模质量^[8-10],尽管这些方法相比于传统方法在模型重建效果上得到了显著的提升,但往往严重依赖于大型带标注的数据集。

因此,一些后续工作尝试在原始数据无需额外标注的情况下,解决三维模型的配准问题。例

如 Groueix 等^[11]和 Li 等^[12]的工作,提出通过无监督的倒角距离约束,来训练深度神经网络,从而将给定模板拟合到输入的原始点云数据的形状。然而,这些方法可能会在某些区域出现显著的性能下降,例如原始数据中存在肢体缺损等。Wang 等^[13]的工作则提出通过笼来控制模板的变形,以无监督的方式在不同形状之间训练网络生成笼的变形参数,进而控制模板变形,并以此建立对应关系。另外,基于函数映射的方法^[14-16]也被广泛应用在三维形状的配准任务中。例如, Eisenberger 等^[16]通过结合谱空间和坐标空间的方法,消除模糊对应,实现了更准确的三维配准。然而,这些方法大多只关注全局形状的正则化约束,往往会失去对局部几何形状的控制。因此,如图 1 所示,它们的重建结果可以反馈三维人体模型大致的形状信息,但在肢体部位仍经常出现网格翻转、过度拉伸或相互交叉的畸变现象。

为解决上述问题,本文提出一个基于生成式对抗网络(generative adversarial network, GAN)的网络框架,利用未配对的点云和网格作为三维模型正则化的先验信息。该网络框架由变形部分和判别部分组成。其中,变形部分基于深度变形的三维对应^[11](3D correspondences by deep deformation, 3D-CODED),将预定的模板拟合到输入点云上。判别部分包括形状判别器和边判别器,用于对变

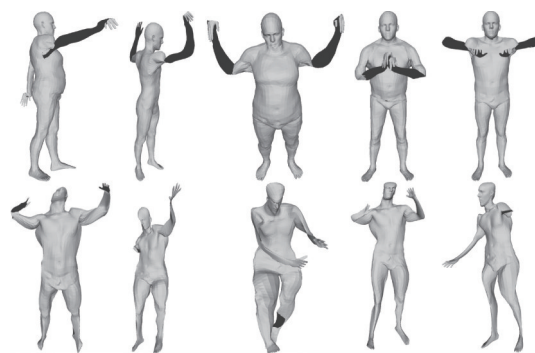


图 1 三维人体模型配准的误差现象

Fig. 1 Errors in 3D human model registration

形网格进行畸变检测,以保证变形网格的几何形状和边缘分布正确。形状判别器获取一对点云局部块并学习它们是否匹配,而边判别器评估点云块上的种子点与该块内所有点之间的距离,以确定生成的网格的有效性。

基于这2个判别器的对抗性损失,可以联合网格变形过程、网格几何形状以及网格边缘分布的优化,从而减轻生成网格中的失真。在SURREAL数据集^[17]、FAUST数据集^[18]和SMAL数据集^[19]上测试本文的方法,结果表明,本文方法在定性和定量结果上都较先前工作表现得更好。本文方法参考3D-CODED^[11]中变形网络的设计,并通过引入2个新型的判别器,大幅减少了生成三维形状中的畸变。总之,本文的工作在以下2个方面做出了贡献:

1)提出一种通用的框架,对无监督的模板拟合方法中出现的网格伪影进行正则化约束。

2)提出一个形状判别器,用于正则化变形网格中过度拉伸的畸变。并基于该形状判别器衍生了一个边判别器,用于约束相互交叉的网格边缘。

1 相关工作

铰接形状的三维重建和配准工作,因其在视觉内容创作和三维动画中的应用,引发了计算机视觉和图形学界的广泛关注。本节将回顾先前的相关工作在铰接形状表示和三维配准任务中的贡献,以及GAN在三维形状分析任务中的应用。

1.1 基于神经网络的铰接形状表示

研究人员已经探索了不同三维形状的特殊表示方法,以便更准确地恢复它们的形状细节。尤其是铰接形状,如三维人体、手部和动物等,均设计了不同的表示方法。早期的工作^[20]中采用了可适配于深度学习的密集三维体素作为形状表示,然而这些方法的计算过程繁琐,模型细节也较为模糊。随后的一系列工作^[11,21-23]基于固定拓扑的网格进行了研究,但让神经网络学习每个网格顶点的位置并不容易,可能会产生形状交叉和畸变。同时,参数化的三维人体表示方法,如SMPL^[2]、通用人体模型(generic human model, GHUM)^[24]等,也被广泛讨论和使用。近期,基于隐式表示的方法^[25-26]取得了令人满意的模型重建效果。然而,这些方法不能保证模型密集对应关系的准确性,因此难以应用于三维动画生成等任

务。本文的工作致力于解决模型网格的过度拉伸问题,在完成三维点云配准任务的同时,得到精美的模型重建结果。

1.2 基于神经网络的三维配准

通常的三维配准方法致力于预测2个输入之间的密集对应关系。而对于铰接形状的配准工作,由于其不同物体部件之间的非刚性变形使得该工作具有挑战性。早期的工作通过人工标注特征,获取不同姿态下三维模型的内在结构信息,从而进行三维形状配准,典型的方法有热核特征^[27]。而另一些基于函数映射的方法^[16,28-29],则将配准问题转化到谱空间解决。Sharma和Ovsjanikov^[30]的工作则在特征空间内,对模型的顶点进行配准。通过学习源数据与目标间的拉普拉斯特征基函数确定最优对,从而由模型局部到整体逐渐提高匹配精度。Ginzburg和Raviv^[15]的工作构建了一个循环映射,通过迭代更新每个点的局部描述符,在谱域内优化三维形状间的软对应关系。

考虑到铰接形状具有拓扑一致性,还有一些工作通过预测固定拓扑的网格来实现形状配准,从而较为轻松地寻找密集对应关系。Hirshberg等^[31]首先检测人体模型头部和脚踝的标志,然后将这些标志作为模板变形的先验。在他们后续的工作^[32]中,在使用模板配准扫描点云的同时对目标模型进行建模,并驱动模板变形以适应新的人物动作。然而,这些方法并不能获得令人满意的效果,可能会遗漏四肢等人体部位。在基于深度学习的工作中,Groueix等^[11]提出通过网络引导预定模板向目标变形的的方法,并给出有监督和无监督2种训练方式。而Li等^[12]则以自监督的方式把深度神经网络与线性形状蒙皮相结合。Wang等^[13]则使用控制笼作为中介来建立不同形状之间的关系,但这种方法难以保证预测姿态的合理性,也不能准确地还原形状细节。尽管上述方法在无监督条件下取得了一定的成果,然而,由于缺乏对局部细节的处理,往往会在肢体部位出现畸变。为此,本文在现有框架的基础上,提出2种基于GAN的局部形状判别器,用以约束重建模型局部块的正确性,从而避免肢体的畸变和扭曲。

1.3 基于GAN的三维形状分析

GAN^[33]网络凭借其无监督学习和逼真的内容合成能力,一度在人工智能多个研究领域掀起

热潮。许多研究工作也探索了 GAN 在三维模型生成和铰接形状分析任务中的作用。Abrevaya 等^[34]将 GAN 网络应用在生成三维头像的任务中, GAN 网络在该工作中展现出很强的灵活性, 可以很好地辨别出人脸部的固有特征以及表情对其产生的影响, 使得最终生成的人物面部表情细节更加真实生动。而 Kanazawa 等^[35]则将 GAN 的判别器结构融合到端到端的网络框架中, 从人物的运动数据中无监督地学习人体姿态的正则化约束, 从而防止三维人体模型出现自渗透。Ma 等^[36]则将区域判别器扩展到三维网格, 提出一种 Mesh-VAE-GAN 的结构, 可以有效地保留目标模型的局部细节。Zhou 等^[10]提出一种基于对抗性高斯过程的网络架构, 这种架构利用高斯过程的非参数特性, 来适应人体模型复杂的几何形状。并使用有监督的对抗性训练方法优化网络, 从而更准确地重建人体模型表面的网格。以上方法大多使用了监督训练, 在无监督的条件下很难产生理想的效果。本文的实验着重关注先前工作在模型重建过程中出现的网格畸变问题, 并提出新的正则化约束以减少此类误差。

2 研究方法

给定一个输入点云和一个铰接对象的模板网格, 本文的目标是对模板进行变形, 使得变形后的三维形状与输入点云相匹配。得到重建模型之后, 可以通过重建模型的固定拓扑, 以及输入点云与重建模型的对应关系, 计算出不同点云之间的密集对应关系, 进而可以有效支持纹理转移和铰

接形状的 3D 动画。本文采用两阶段的过程来完成三维人体模型的配准任务, 整个过程如图 2 所示:

1) 通过变形网络将模板变形, 从而拟合输入的点云;

2) 使用形状判别器和边判别器对生成网格模型中的过度拉伸和穿插进行正则化约束。

2.1 变形网络

本文方法基于 3D-CODED^[11]中提出的变形网络, 该网络以点云 P 和模板网格 T 作为输入。点云 P 首先被编码成代表形状和姿态信息的全局特征 Z_p , 然后将全局特征 Z_p 与模板网格 T 连接起来, 输入到解码器中预测模板变形后的位置, 使得变形后的模板能够拟合输入的点云 P 。

实验部分使用 PointNet^[37]作为编码器, 全局特征 Z_p 包含了 1 024 维来自输入点云的点信息。选择一个简单的多层感知机 ($1\ 027 \times 513 \times 256 \times 3$) 作为解码器, 生成变形后模型顶点的空间位置。

2.2 对抗式的三维形状正则化约束

3D-CODED^[11]容易生成过度拉伸的网格和扭曲的身体部位, 尤其是在无监督训练模式下。如图 1 所示, 在实验中发现这种不规则的畸变经常出现在手部和腿部等部位。产生该问题的根本原因在于 3D-CODED 在无监督损失训练模式下的倒角损失^[11]没有考虑到变形后的模板的局部几何形状和边缘分布的正确性。GAN 网络在图像^[33]和三维形状^[35]的无监督内容合成任务上表现出了优越的性能, 可以作为变形网络的正则化约束。

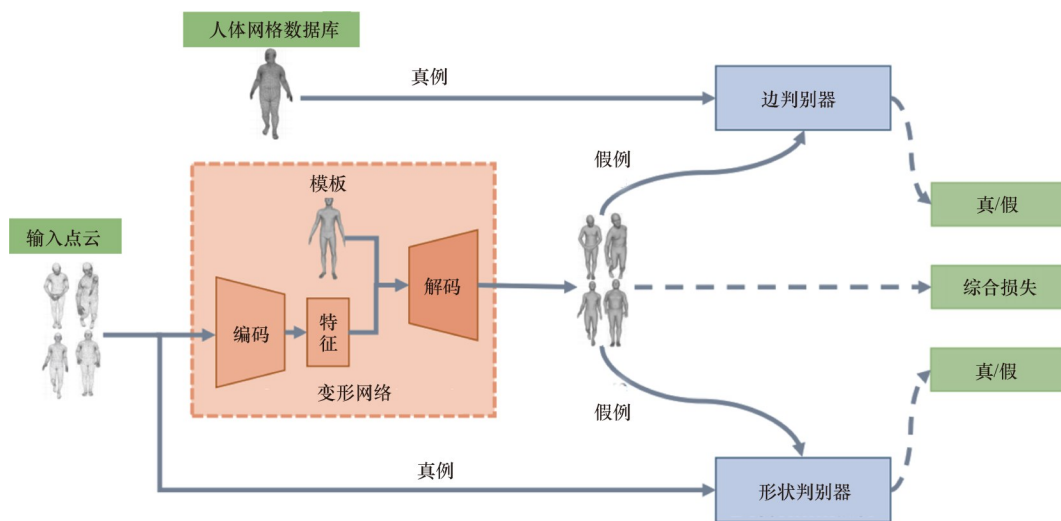


图 2 网络结构的设计细节

Fig. 2 Design details of the proposed method

基于观察结果,大多数情况下具有畸变的网格部位在点位置和边缘分布上呈现出了异常的分布。因此,一个足够强大的网格形状判别器可以很容易地识别这些位置,并将存在异常形变的局部块识别为假例。如图3所示,3D-CODED重建结果在肢体部位发生畸变的情况最为严重。在本文实验中,使用了类似PatchGAN^[38]的基于局部形状的判别器,可以更好地捕捉网格顶点分布及其区域内边的关系。因此,设计了一种形状判别器,用来判断变形网格与输入点云的对应局部块内顶点分布是否一致。然而,这个判别器仅考虑了各个顶点的空间分布,没有考虑局部块内各点的连接情况。为克服这一缺陷,基于形状判别器又设计了一个边判别器,该判别器的输入为局部块内由种子点与其他点连接而成所有边,判断变形网格的边分布是否与输入点云上最近邻局部块中的边分布相类似。基于这样的设计,本文方法在不依赖带标注的点和对固定拓扑网格的前提下,依然可以有效缓解肢体部位的畸变,减少异常模型网格。图4展示了这2个判别器的总体架构。

形状判别器从成对的输入中获取这2个局部区域的几何信息,并输出一个二进制标签。将假例对标记为0,真例对为1。判别器的目标是让来自点云和与其对应的变形网格的局部块尽可能匹配。因此,首先对输入模型进行小范围扰动(添加小范围的噪声和随机的小范围旋转,实验设置为1 cm的位移扰动,以及绕任意轴旋转0.3 rad的旋转扰动);然后,在输入点云上随机选择 N 个种子点($N=64$),将每个种子点及与其最近的 K 个相邻点($K=128$)构成局部块。将输入点云的局部块与

扰动结果的对应位置组成一个局部块对,标记为真例;将输入点云的局部块和来自变形网络生成模型的最近邻局部块组成另一个局部块对,标记为假例。这种处理方式建立在生成模型与输入点云空间位置大致相似,且对应位置的空间结构同样相似的基础上。

图4展示了形状判别器的设计结构。将2个模型的局部块输入到一个共享编码器中,分别学习局部块内部的全局特征。然后,将这2个局部块的特征连接起来,使用解码器来预测该局部块的标签。形状编码器由3个一维卷积层、1个最大池化层和1个全连接层组成,其解码器包含3个一维卷积层和1个最大池化层。另一种避免生成过度拉伸或相互交叉的模型网格的思路是对网格的边缘分布进行正则化约束。与形状判别器的结构类似,本文设计了一个边判别器。区别在于,边判别器分别以输入点云的局部块以及生成模型对应位置的局部块中,每个点到其种子点的欧氏距离作为输入。欧氏距离公式的定义如下

$$d(p, q) = \sqrt{\sum_{i=1}^3 (p_i - q_i)^2}, \quad (1)$$

其中, p 和 q 分别表示局部块内的点和种子点的三维坐标。

与评估顶点空间位置的形状判别器不同,边判别器更关注局部块内部顶点之间的相对位置关系。

2.3 损失函数设计

本文的目标是对整个网络进行无监督的训练,因此3D-CODED^[11]中的无监督损失 L_{un} 可以用于baseline的无监督训练。此外,2个对抗性损失

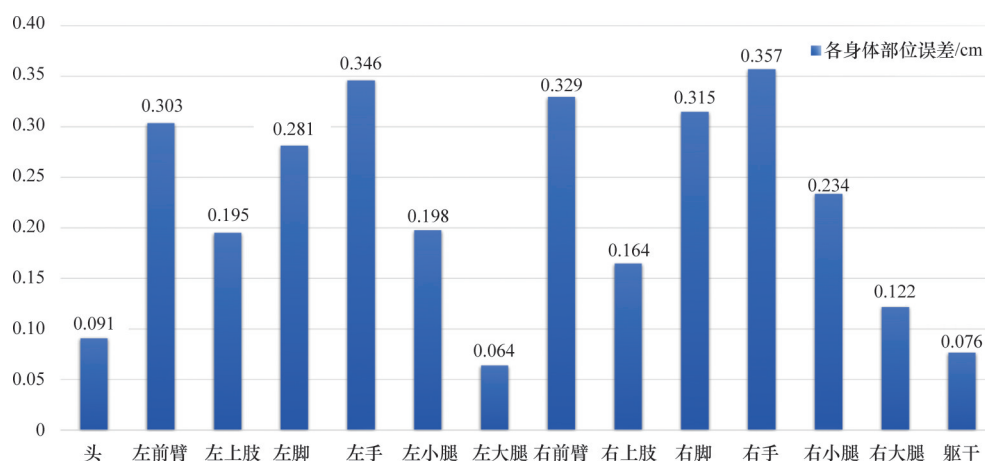


图3 3D-CODED重建结果的误差分布

Fig. 3 Error distribution of 3D-CODED results

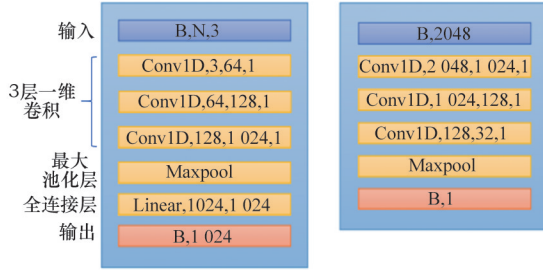


图 4 判别器结构

Fig. 4 Structure of the discriminator

L_{adv-g} 和 L_{adv-t} , 可以作为网格几何形状和边缘分布的正则化约束, 限制生成结果中的畸变。因此, 总损失如下

$$L_{total} = L_{un} + \lambda_1 \cdot L_{adv-g} + \lambda_2 \cdot L_{adv-t}. \quad (2)$$

3D-CODED 中的无监督损失定义为倒角损失 L_{CD} 、拉普拉斯损失 L_{lap} 和边缘损失 L_{Edge} 的总和

$$L_{un} = L_{CD} + \lambda_{lap} \cdot L_{lap} + \lambda_{Edge} \cdot L_{Edge}, \quad (3)$$

其中, 输入点云与目标网格对应点之间的倒角距离定义为

$$L_{CD}(S, P_M) = \sum_{p \in S} \min_{q \in P_M} |p - q|^2 + \sum_{q \in P_M} \min_{p \in S} |p - q|^2, \quad (4)$$

其中, p 和 q 表示三维坐标。拉普拉斯损失 L_{lap} 鼓励在模板与生成模型网格上的拉普拉斯算子保持一致, 而边缘损失 L_{Edge} 则鼓励模板与生成模型对应边长的比率为 1。虽然这 2 种损失是为约束过度拉伸和交叉的模型网格而设计的, 但是应用于无监督训练时结果较差。主要原因是它们以全局的方式惩罚模型的畸变, 而忽略了局部块的畸变, 尤其在肢体部位。

参考 PatchGAN^[38] 和 LS-GAN^[39] 中损失函数的设计, 设置形状对抗损失和边对抗损失, 形状判别器的对抗损失定义如下

$$\begin{aligned} L_{adv-g} &= L_{d-g} + L_{g-g}, \\ L_{d-g} &= 0.5 \cdot E_{x \in P_f^g(x)} \left[\left(D_g(x) - 1 \right)^2 \right] + \\ &\quad 0.5 \cdot E_{x \in P_f^g(x)} \left[\left(D_g(x) \right)^2 \right], \\ L_{g-g} &= 0.5 \cdot E_{x \in P_f^g(x)} \left[\left(D_g(x) - 1 \right)^2 \right], \end{aligned} \quad (5)$$

其中, L_{d-g} 表示形状判别器进行真假判别的损失, L_{g-g} 表示三维模型变形过程的判别损失。用 $P_f^g(x)$ 和 $P_f^g(x)$ 表示真假局部块的点分布, 形状判别器对局部块的判别结果表示为 $D_g(x)$ 。相类似地, 边判别器的对抗损失 L_{adv-t} 可以用不同的判别器 $D_t(x)$ 来表示。与大多数 GAN 网络的训练方法

相同, 本文也按照默认的训练协议来优化网络参数, 将训练过程分为对抗性的 2 个阶段, 分别更新判别器和生成器的参数。

为提高性能, 采用了与 3D-CODED 同样的后处理策略。即在测试过程中, 对于每个输入点云, 根据生成模型与输入点云的倒角距离优化网络参数, 使生成模型与输入点云更贴近。本文在该部分使用 Adam 优化器以 5×10^{-4} 的学习率共进行了 1 000 次迭代优化。

3 实验结果

本节首先介绍实验选取的数据集、评价指标以及实验环境。然后, 将本文方法与三维形状配准领域的先进方法进行比较, 并对该网络框架的性能进行详细分析。

3.1 数据集选取

实验在三维人体以及部分动物模型这 2 类对象上进行了测试。对于三维人体模型, 使用标准的 SURREAL 数据集^[17] 和 FAUST 数据集^[18]。而对于动物模型, 则选取了 SMAL 数据集^[19] 来生成训练集和测试集。SURREAL 数据集是一个大型合成数据集, 包含不同姿势的三维人体形状。实验遵循 3D-CODED 的协议生成训练数据。实验过程中, 使用 20 000 个不同姿势的点云进行训练, 以及 5 000 个点云用于测试。FAUST 数据集是通过扫描真实人体获得的, 每个扫描模型包含约 170 000 个顶点。将这 200 个扫描模型作为测试数据, 并按照 FAUST 的基准, 反馈每个测试主体自身和主体之间对应关系的准确性。数据集扫描模型保留了噪声和孔洞, 尤其是在模型脚和手等肢体部位, 这使得本次实验具有一定的挑战性。

3.2 评价指标

实验使用倒角距离作为重建形状和与之对应的真值形状之间的距离度量, 用来反映重建形状与原始输入的相似程度。此外, 还使用来自 2 个不同点云的对应点之间的平均欧氏距离作为对应度量。在 FAUST 数据集中, 研究人员收集了测试用的点云对, 并标注了它们与真值间的对应关系。因此, 可以通过在线评估的方法获得对应测试的结果。而 SURREAL 数据集和 SMAL 数据集则仅含有点云和与其相应的固定拓扑网格。在本文实验中, 每个测试集被拆分成 2 500 个点云对进行评估。其中, 真值的对应关系是按照 2.2 节中的算

法,基于固定拓扑的网格构建的。

3.3 实验环境

实验的网络框架使用 PyTorch 搭建。使用 150 次迭代训练变形网络,50 次迭代训练 2 个对抗网络,并使用 Adam 优化器优化网络。对于 2 个判别器,初始学习率设置为 10^{-3} 。对于变形网络,初始学习率同样设置为 10^{-3} ,并分别在迭代周期为 80 和 120 时将学习率调整为 10^{-4} 和 10^{-5} ,训练过程中使用的批量大小设置为 64。

3.4 与先前工作的对比结果

本文比较了基于模板的变形方法,即 3D-CODED 和基于控制笼的变形方法,即深度笼^[13](Deep-Cage)。这 2 个方法发布了它们的源代码,本文用相同的数据对它们的模型进行训练,从而进行公平的对比实验。由表 1 的重建与配准评价结果可知,本文方法在人体数据和动物数据上均超越了先前的方法,这证明了引入 2 个对抗判别器的有效性。为评估本文实验在 FAUST 数据集上的密集对应关系,在其官方基准服务器上进行测试实验,并将实验结果展示在表 1 的右侧 2 列(即 FAUST-inter 和 SAML-Reg)。

表 1 中重建(FAUST 和 SMAL)的误差数值表示生成模型与输入点云的倒角距离,配准(FAUST-inter 和 SAML-Reg)的误差数值表示变形后三维形状的顶点和与其对应的真值点之间的平均欧氏距离。与 3D-CODED 相比,本文方法将

FAUST、FAUST-inter、SMAL 和 SMAL-Reg 数据集的配准效果分别提升 0.117(约 12%)、2.52(约 25%)、0.03(约 16%)和 0.57(约 18%)。与 Deep-Cage 相比,性能提升效果更加显著。在 SMAL 数据集上也可以得到类似的实验结果,并且 Deep-Cage 由于缺乏合适的预制笼,导致无法处理动物数据。

本文方法不仅在数值评价指标上表现良好,还拥有更好的三维模型重建效果。图 5 展示了模型重建细节。在图 5(a)和 5(b)中,从左往右依次是输入的扫描点云、3D-CODED、Deep-Cage 以及本文方法的重建结果。如图 5(a)所示,与 3D-CODED 相比,本文方法可以更好地适应不同姿态的三维形状,并保证重建结果的平滑性。虽然 Deep-Cage 更加擅长反映模型的姿态变化,但是在保留形状细节方面的能力明显不足。图 5(b)展示了姿态校正的实验结果。3D-CODED 和 Deep-Cage 无法对某些姿态给出准确的拟合,而本文方法可以修正明显的三维模型畸变。

此外,本文方法还可以扩展到其他铰接形状,如动物,图 6 就展示了该部分的实验结果。图 6 从

表 1 重建与配准结果评价

Table 1 Evaluation of reconstruction and registration results				
方法	FAUST	SMAL	FAUST-inter	SMAL-Reg
3D-CODED	0.961	0.19	10.46	3.23
Deep-Cage	2.160	*	21.65	*
本文方法	0.844	0.16	7.64	2.66

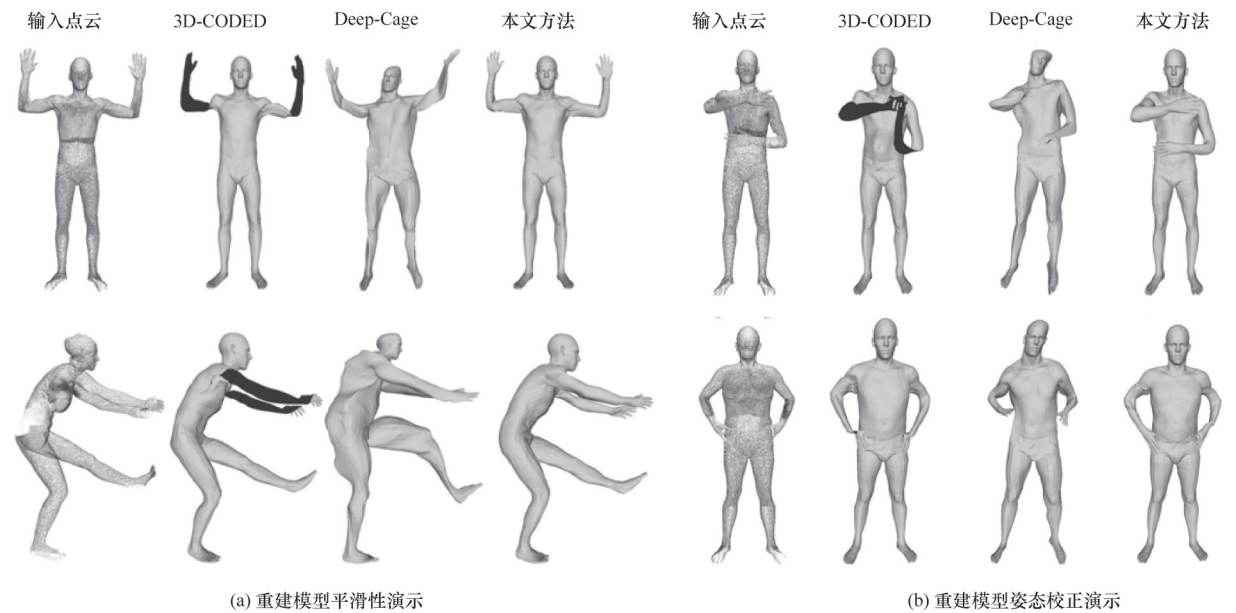


图 5 重建模型结果

Fig. 5 Reconstruction model results

左到右依次展示了来自 SMAL 数据集的扫描点云、3D-CODED 和本文方法的重建结果。

3.5 消融实验

在本小节的实验中,将依次添加 2 个判别器,从而评估它们对三维模型重建结果的影响。

如表 2 所示,由倒角距离和配准误差,可以证明 2 个判别器对最终的生成效果都起到了积极的作用。此外,它们还极大地改善了重建模型的视觉效果,如图 7 所示。

如图 7 所示,从左到右依次是输入的扫描点云、实验的 baseline、加入形状判别器的实验结果以及 2 个判别器共同作用的结果。可以看到,本文方法成功地恢复了输入扫描点云中包含的三维形状细节,并减少了 3D-CODED 在手部形状和姿

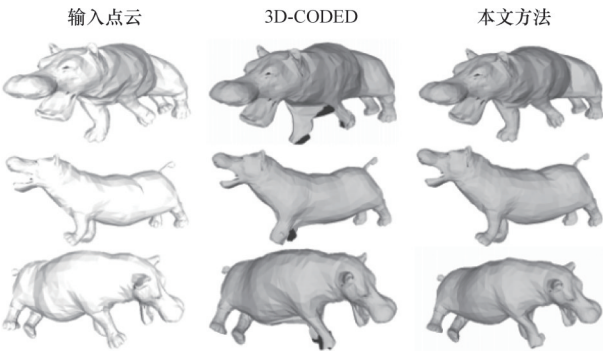


图 6 动物重建结果对比

Fig. 6 3D animal reconstruction

表 2 2 个判别器的消融实验

Table 2 Ablation studies on two discriminators

方法	FAUST-chamfer	FAUST-inter
Baseline	0.961	10.46
仅含形状判别	0.904	8.27
本文方法	0.844	7.64

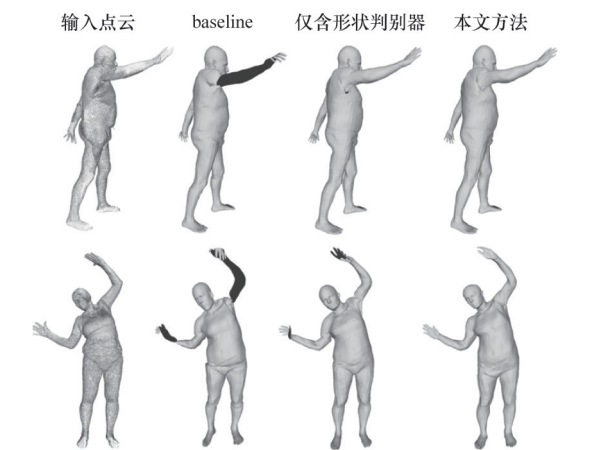


图 7 2 个判别器对重建结果的影响

Fig. 7 Impact of two discriminators on reconstruction results

态细节上的误差。虽然形状判别器基本上可以惩罚模型网格的畸变,但无法纠正手部形状的误差。究其原因在于,人体模型手部的顶点分布差异非常小,难以获取合适的损失。而加入边判别器后,这些关节部位的模型得到修缮,并且获得了更丰富的细节。

4 总结

本文提出一种基于 GAN 的无监督神经网络三维模型配准方法,建立了扫描点云和变形网格间的密集对应关系,并针对铰接形状中容易发生畸变的变形网格,添加了正则化约束。创新性设计 2 个判别器,联合限制模型网络的伪影,明显消除了模型网格扭曲和翻转。与先前的工作相比,本文方法更好地捕捉了三维模型局部形状的信息,并且在没有预先设置真值网格的情况下,依然修正了生成三维形状中的局部异常。此外,还对这 2 个判别器进行了消融实验,验证其必要性。实验结果表明,本文方法在综合性能上优于先前的方法。

本文方法依然难以处理一些特殊情景的三维配准问题,例如关于夸张人体姿态的拟合以及跨类别三维形状的匹配任务,这值得在未来的研究工作中进行更深层次的探索。

参考文献

[1] Romero J, Tzionas D, Black M J. Embodied hands: modeling and capturing hands and bodies together[J]. ACM Trans Graph, 2017, 36(6): 1-17. DOI: 10.1145/3130800.3130883.

[2] Loper M, Mahmood N, Romero J, et al. SMPL: a skinned multi-person linear model [M]//Seminal Graphics Papers: Pushing the Boundaries, Vol 2. 2023: 851-866. DOI: 10.1145/3596711.3596800.

[3] Yang H T, Zhu H, Wang Y R, et al. FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 598-607. DOI: 10.1109/CVPR42600.2020.00068.

[4] Sumner R W, Popović J. Deformation transfer for triangle meshes[J]. ACM Transactions on Graphics, 2004, 23(3): 399-405. DOI: 10.1145/1015706.1015736.

[5] Amberg B, Romdhani S, Vetter T. Optimal step nonrigid ICP algorithms for surface registration [C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA. IEEE, 2007: 1-8. DOI: 10.1109/

- CVPR. 2007. 383165.
- [6] Feng A W, Casas D, Shapiro A. Avatar reshaping and automatic rigging using a deformable model[C]//Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games. 2015: 57-64. DOI:10.1145/2822013.2822017.
 - [7] Zuffi S, Black M J. The stitched puppet: a graphical model of 3d human shape and pose[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3537-3546. DOI:10.1109/cvpr.2015.7298976.
 - [8] Halimi O, Imanuel I, Litany O, et al. The whole is greater than the sum of its nonrigid parts [EB/OL]. arXiv 2020: 2001.09650. (2020-01-27) [2024-05-09]. <http://arxiv.org/abs/2001.09650v1>.
 - [9] Deprelle T, Groueix T, Fisher M, et al. Learning elementary structures for 3D shape generation and matching [J]. Advances in Neural Information Processing Systems 32, 2019: 7433-7443.
 - [10] Zhou B, Franco J S, Bogo F, et al. Reconstructing human body mesh from point clouds by adversarial gp network[C]//Proceedings of the Asian Conference on Computer Vision. 2020. DOI:10.1007/978-3-030-69525-5_8.
 - [11] Groueix T, Fisher M, Kim V G, et al. 3d-coded: 3d correspondences by deep deformation[C]//Proceedings of the european conference on computer vision (ECCV). 2018: 230-246. DOI:10.1007/978-3-030-01216-8_15.
 - [12] Li C L, Simon T, Saragih J, et al. Lbs autoencoder: self-supervised fitting of articulated meshes to point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11967-11976. DOI: 10.1109/cvpr.2019.01224.
 - [13] Wang Y F, Aigerman N, Kim V G, et al. Neural cages for detail-preserving 3D deformations [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 72-80. DOI: 10.1109/CVPR42600.2020.00015.
 - [14] Bhatnagar B L, Sminchisescu C, Theobalt C, et al. Combining implicit function learning and parametric models for 3D human reconstruction [C]//Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 311-329. DOI: 10.1007/978-3-030-58536-5_19.
 - [15] Ginzburg D, Raviv D. Cyclic functional mapping: self-supervised correspondence between non-isometric deformable shapes [C]//Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part V 16. Springer International Publishing, 2020: 36-52. DOI:10.1007/978-3-030-58558-7_3.
 - [16] Eisenberger M, Lahner Z, Cremers D. Smooth shells: multi-scale shape registration with functional maps [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12265-12274. DOI: 10.1109/CVPR42600.2020.01228.
 - [17] Varol G, Romero J, Martin X, et al. Learning from synthetic humans [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 109-117. DOI:10.1109/CVPR.2017.492.
 - [18] Bogo F, Romero J, Loper M, et al. FAUST: dataset and evaluation for 3D mesh registration [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 3794-3801. DOI:10.1109/CVPR.2014.491.
 - [19] Zuffi S, Kanazawa A, Jacobs D W, et al. 3D menagerie: modeling the 3D shape and pose of animals [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 6365-6373. DOI: 10.1109/CVPR.2017.586.
 - [20] Varol G, Ceylan D, Russell B, et al. Bodynet: volumetric inference of 3D human body shapes [C]//Proceedings of the European conference on computer vision (ECCV). 2018: 20-38. DOI:10.1007/978-3-030-01234-2_2.
 - [21] Rakotosaona M J, Ovsjanikov M. Intrinsic point cloud interpolation via dual latent space navigation [C]//European Conference on Computer Vision. Cham: Springer, 2020: 655-672. DOI:10.1007/978-3-030-58536-5_39.
 - [22] Ranjan A, Bolkart T, Sanyal S, et al. Generating 3D faces using convolutional mesh autoencoders [C]//Proceedings of the European conference on computer vision (ECCV). 2018: 704-720. DOI:10.1007/978-3-030-01219-9_43.
 - [23] Kanazawa A, Tulsiani S, Efros A A, et al. Learning category-specific mesh reconstruction from image collections [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 386-402. DOI:10.1007/978-3-030-01267-0_23.
 - [24] Xu H Y, Bazavan E G, Zanfir A, et al. GHUM & GHUML: generative 3D human shape and articulated pose models [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 6183-6192. DOI: 10.1109/CVPR42600.2020.00622.
 - [25] Chibane J, Alldieck T, Pons-Moll G. Implicit functions in feature space for 3D shape reconstruction and completion [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 6970-6981. DOI:10.1109/CVPR42600.2020.00700.
 - [26] Saito S, Huang Z, Natsume R, et al. Pifu: pixel-aligned implicit function for high-resolution clothed human digitization [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 2304-2314. DOI: 10.1109/ICCV.2019.00239.
 - [27] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion [C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2009, 28(5): 1383-1392. DOI:10.1111/J.1467-8659.2009.01515.X.
 - [28] Marin R, Melzi S, Rodolà E, et al. FARM: functional automatic registration method for 3D human bodies [J]. Computer Graphics Forum, 2020, 39(1): 160-173. DOI: 10.1111/cgf.13751.

- [29] Ovsjanikov M, Corman E, Bronstein M, et al. Computing and processing correspondences with functional maps [M]//SIGGRAPH ASIA 2016 Courses. 2016: 1-60. DOI: 10.1145/2988458.2988494.
- [30] Sharma A, Ovsjanikov M. Weakly supervised deep functional maps for shape matching [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems, ACM. 2020: 19264-19275.
- [31] Hirshberg D A, Loper M, Rachlin E, et al. Evaluating the automated alignment of 3D human body scans [C]//2nd International Conference on 3D Body Scanning Technologies. Hometrica Consulting, 2011: 76-85. DOI: 10.15221/11.076.
- [32] Hirshberg D A, Loper M, Rachlin E, et al. Coregistration: simultaneous alignment and modeling of articulated 3D shape [C]//Computer Vision - ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12. Springer Berlin Heidelberg, 2012: 242-255. DOI: 10.1007/978-3-642-33783-3_18.
- [33] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [EB/OL]. arXiv 2014: 1406.2661. (2014-06-10)[2024-05-09]. <http://arxiv.org/abs/1406.2661>.
- [34] Abrevaya V F, Boukhayma A, Wuhler S, et al. A decoupled 3d facial shape model by adversarial training [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9419-9428. DOI: 10.1109/ICCV.2019.00951.
- [35] Kanazawa A, Black M J, Jacobs D W, et al. End-to-end recovery of human shape and pose [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7122-7131. DOI: 10.1109/CVPR.2018.00744.
- [36] Ma Q L, Yang J L, Ranjan A, et al. Learning to dress 3D people in generative clothing [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 6468-6477. DOI: 10.1109/CVPR42600.2020.00650.
- [37] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 77-85. DOI: 10.1109/CVPR.2017.16.
- [38] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 5967-5976. DOI: 10.1109/CVPR.2017.632.
- [39] Mao X D, Li Q, Xie H R, et al. Least squares generative adversarial networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2813-2821. DOI: 10.1109/ICCV.2017.304.