

人体模型建模中全局映射智能估计人体形状的方法

季勇^{1,2}, 仇明慧¹, 魏佳¹, 施静^{*1}, 张华³

(1. 江苏工程职业技术学院 江苏省先进纺织工程技术中心, 江苏 南通 226006; 2. 南通大学 安全防护用特种纤维复合材料研发国家地方联合工程研究中心, 江苏 南通 226019; 3. 南通三兆制衣科技有限公司, 江苏 南通 226017)

摘要:针对人体模型建模中形状估计存在的性能低、适用性不强等问题,提出一种利用卷积神经网络进行人体形状估计的智能方法,并利用卷积神经网络层次化特征提取和多视图融合的方法精确估计人体形状。以人体模型为基础,对人体三维形状进行尺寸估计,实现快速人体形状估计,且使估计过程自动化,提高估计结果的精确性。研究表明,该方法克服了传统人工测量技术无法捕捉人体特征、估计人体形状与姿态的局限性,可用于人体模型尺寸的测量、服装虚拟现实对人体模型的建模等。

关键词:卷积神经网络;服装;人体形状;虚拟现实;深度学习

中图分类号:TP 18;TS 941.26 文献标志码:A 文章编号:2096-1928(2024)03-0223-06

Intelligent Estimation Method of Human Body Shape in Human Body Modeling

Ji Yong^{1,2}, Qiu Minghui¹, Wei Jia¹, Shi Jing^{*1}, Zhang Hua³

(1. Jiangsu Advanced Textile Engineering Technology Center, Jiangsu College of Engineering and Technology, Nantong 226006, China; 2. National and Local Joint Engineering Research Center for Research and Development of Special Fiber Composites for Safety Protection, Nantong University, Nantong 226019, China; 3. Nantong Sanzhao Clothing Technology Co., Ltd., Nantong 226017, China)

Abstract: To solve the problems of low performance and weak applicability in shape estimation in human body modeling, an intelligent method using convolutional neural networks for human body shape estimation was proposed, and the principle of convolutional neural network framework for human body shape estimation was explored. The convolutional neural network was used for hierarchical feature extraction and multi-view fusion to estimate human shape accurately. Based on the human body model, size estimation of the three-dimensional shape of the human body was carried out to achieve fast human body shape estimation, and the estimation process was automated, resulting in relatively accurate estimation results. Research shows that this method overcomes the limitations of traditional manual measurement techniques in capturing human features, estimating human shape and posture. The results can be used for measuring human model dimensions, modeling human models in clothing virtual reality, and more.

Key words: convolution neural network, clothing, body shape, virtual reality, deep learning

卷积神经网络(convolutional neural networks, CNN)已经成为估计人体形状的重要工具^[1]。CNN模型可以从单个或多个视角图像中提取特征,并预

测人体的三维形状和姿态^[2]。近年来,许多研究者从不同角度提出采用CNN模型来预测人体三维形态^[3],并取得一定的进展。这些模型通常包括卷积

收稿日期:2023-12-30; 修订日期 2024-04-08。

基金项目:江苏省高等学校基础科学(自然科学)研究面上项目(22KJB520031);江苏省现代教育技术研究项目(2022R105190);江苏省科技副总项目(FZ20230577);南通市基础科学研究和社会民生科技计划项目(JC12022105);江苏政府留学奖学金。

作者简介:季勇(1987—),男,博士研究生。

*通信作者:施静(1974—),女,教授,博士生导师。主要研究方向为服饰文化与服装设计。Email:790030874@qq.com

层、池化层、全连接层和输出层等组件^[4]。胡凯等^[5]使用注意力机制增强 CNN 模型对关键点的关注度,从而提高人体形状估计的准确性。KIM J E^[6]使用多种技术和方法来改进 CNN 模型的性能,如多视角输入、合并层、注意力机制和数据增强等。然而,目前 CNN 模型通常基于最小化误差拟合项的生成方法,该方法需要获取二维图像并且还需要添加深度信息^[7],使实时应用程序变得非常复杂。目前,大多数 CNN 模型都是基于单个视角图像估计人体形状,限制了其在处理多个视角或复杂场景时的性能。除了 CNN 模型,一些研究者还使用了其他机器学习方法解决人体形状估计问题,如支持向量机、随机森林、神经网络等^[8],这些方法通常需要手工提取特征,并使用监督学习算法进行训练。虽然这些方法在一定程度上可以提高人体形状估计的准确性,但它们通常需要更多的人工干预和专业知识,并且在处理复杂场景时表现不佳。

为了解决这些问题,并进一步提高 CNN 模型在估计人体形状方面的性能和适用性^[9-10],文中使用端到端的全局映射,从 2D 图像到形状参数,区分了两种输入图像的情况:一种是以不同形式出现的正面单视图;另一种是从正面和侧面同时出现的两种图像,将多个视角的信息整合到 CNN 模型中,以提高其性能和鲁棒性。

1 单视图图像估计人体形状

1.1 单视图图像形式

单视图图像形式多样,最常见的情况是将一个

以中立位姿势站立的人体正面二元轮廓图像缩放到一个固定的高度,但是摄像机参数是未知的,因此需要一个固定的缩放比例。其次,已知摄像机参数,当人站在离摄像机已知距离时,则会输入一幅轮廓和高度相同的固定尺寸影像。由于一个轮廓可以代表多种不同的人体形态,因此以二维输入的轮廓估计真实的三维形状存在不确定性。为解决这些问题,可采用影子额外影像线索来弥补缩放轮廓。文中结合额外影像线索的方法,同时根据兰伯特的光照模型^[11],对图像亮度和对比度进行处理,并考虑光源的位置、物体表面的法向量以及材质表面的反射率等因素来估计人体形状,从而生成具有不同亮度和对比度的图像,产生多样化的训练数据,提高深度学习模型的准确性和鲁棒性。

1.2 单视图图像估计人体形状

针对单视图图像形式,通过提取内在图像可以重建人体形状或场景结构,具体如图 1 所示。单视图输入大小为 128×128 像素,将单通道输入图像与已知的形状参数一起输入到单视图网络框架中,通过学习从输入到形状参数 β 的映射,可以准确地估计人体形状,并在实际应用中高效运行。单视图网络框架针对回归问题进行定制,并考虑各种输入类型、预期应用和可用硬件等因素,如采用 CNN 模型提取输入图像的特征,并使用全连接层将这些特征映射到形状参数 β 。为了适应不同的输入类型,对单视图网络进行了定制,包括网络结构、激活函数、损失函数等。

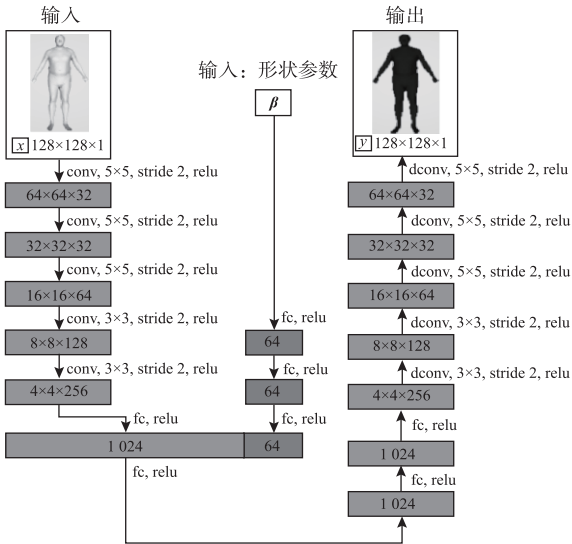


图 1 单视图网络框架

Fig.1 Single view of network framework

单视图网络框架是一个由卷积块和全连接层组成的神经网络,可以学习从输入图像到形状参数

β 的映射。该网络由 5 个卷积块和 3 个全连接层组成,每一层都使用 relu 作为激活函数。全连接层之

间使用 dropout 层,并在第 1、第 2 和第 5 个卷积块之后使用 max pooling 操作。通过卷积运算提取输入图像的特征,并将其映射到形状参数 β ,为防止单视图网络框架的训练过程过度拟合,提高模型的准确性和泛化能力,该过程的卷积运算公式为

$$F(i,j)=X(3\times3)_{ij}\times f(3\times3)\times f_{i,j}(\beta)。(1)$$

式中: F 为经过卷积操作后提取的图像特征,用于映射到形状参数 β ; i,j 为图像中某个像素点的坐标位置; X 为图像中某个像素点及其邻域像素点的矩阵。

在公式(1)中,卷积操作是 CNN 模型中的核心操作之一,其通过滤波器对输入图像进行卷积运算,得到一系列特征图,每个特征图对应一个滤波器,这些特征图可以捕捉到不同尺度和方向的图像特征,如边缘、纹理、形状等。形状参数 β 是用于表示人体三维形状的向量,通常包括身高、体重、肢长等参数。像素点坐标位置表示图像中每个像素点的位置信息,是人体形状估计中重要的输入信息之一,通过对每个像素点及其邻域像素点的矩阵进行内积运算,可以提取该位置的特征。

1.3 估计结果

基于单视图估计人体形状的模型优化算法,可以形成单个图像的学习网络,具体如图 2 所示。然而,这样的网络结果通常存在一些偏差,由于单个图像丢失了许多有关身体形状的信息,因此输出效果较差[见图 2(a)和图 2(b)],由图可知样本模糊且估计不准确;同样,侧面视图的结果也不理想[见图 2(c)和图 2(d)]。为了改善单个视角人体形状估计性能,可以使用多视角图像或其他补充信息来弥补缺失的信息。如使用多个摄像头从不同角度拍摄图像,并将它们组合起来以获得更完整和准确的身体形状;也可以使用其他类型的传感器或数据源,如深度传感器或动作捕捉系统,以获得更准确和详细的身体形状信息。这些改进将为虚拟现实、运动训练和医疗保健等各种应用领域提供更准确和高效的人体形状估计。



(a) 高瘦样本前视图



(b) 矮胖样本前视图



(c) 高瘦样本侧视图



(d) 矮胖样本侧视图

图 2 单视图人体估计

Fig. 2 Single view of human body estimation

2 双视图估计人体形状

2.1 双视图估计人体形状网络框架

为了尽可能准确地估计人体形态和身体各部位尺寸,在单视图网络框架的基础上,提出双视图网络框架(dual view network, DVN),具体如图 3 所示。由图 3 可知,采用两种不同轮廓的人体输入方式,从正面和侧面观察人体,假设摄像机参数是已知的,则能够知道模特和摄像机之间的距离^[12]。在不同的测试图像中,对不同水平的输出滤波器都可

进行可视化,但网络更倾向于学习一些正面视图的过滤器,只有少部分倾向于侧视图。该方法在两组卷积层(每个视图具有共享权重)之后添加了一个

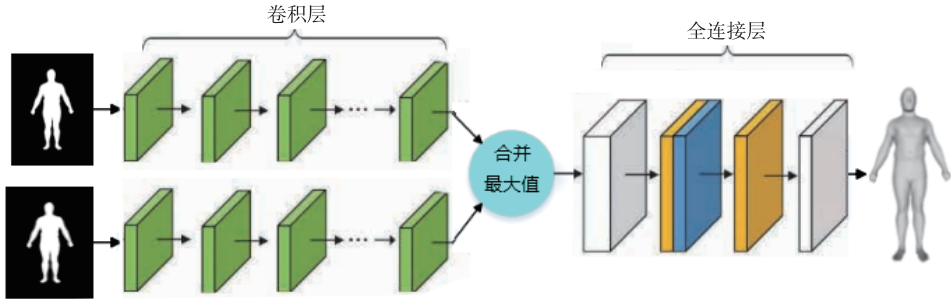


图 3 双视图网络框架
Fig.3 Dual views of network framework

具体而言,在网络中添加一个注意力模块,用于学习每个像素的重要性权重,并将其应用于特征图,这样可以使网络关注到重要的身体部位和形态特征,从而提高估计的准确性。再将形状参数 β 引入双视图网络框架中,进一步优化网络输出结果。在双视图网络框架中,形状参数 β 被加入合并层之前的全连接层中,可以使网络更加关注人体形态和身体各部位的尺寸信息。同时,在训练过程中,将形状参数 β 纳入损失函数,由此合成的注意力模块可以提高网络对重要身体部位和形态特征的关注度。

2.2 双视图估计人体形状模型

在 DVN 方法中,从二值图像中学习回归,因此需要从头开始训练网络。DVN 方法采用了全局映射的运算,将输入特征映射到形状空间,即将二维图像映射到三维形状空间,这个过程需要通过卷积操作提取图像特征,并通过优化损失函数学习形状参数。

基于此,双视图网络框架采用了两种不同轮廓的人体输入方式,从正面和侧面观察人体,以提高估计精度。在 DVN 方法中,优化损失函数是学习形状参数的关键,其采用了基于主成分分析(principal component analysis, PCA)的损失函数,将估计结果投影到学习的 PCA 空间中,从而获得人体形状参数模型的误差下限。通过最小化损失函数学习形状参数,实现对形状参数的准确估计,具体公式为

$$S_{(i,j)} = \sum_{i=1}^n \sum_{j=1}^m K_{m,n} \mathbf{I}_{m,n} \quad (2)$$

式中: S 为卷积网络转换后的特征映射; $\mathbf{I}_{m,n}$ 为输入图像中的像素值组成的二维矩阵坐标, m 为矩阵的行数(样本数), n 为矩阵的列数(特征数); K 为 S 内积运算的权重参数。

完全连接的、与视图底层类似的合并层,从而输入图像都会被送入两个独立的 5 层卷积网络,再使用级联操作将 2 个卷积网络的特征进行合并。

衡量人体模型与给定图像的匹配程度公式为

$$S'_{(i,j)} = \sum_{i=1}^K \sigma_i + \sum_{i=1}^n \sum_{j=1}^m d_{ij} \quad (3)$$

式中: σ_i 为衡量人体部件 i 的外观特征与其相应部件模板的不匹配程度(由外观特征与模板之间的误差得出); d_{ij} 为衡量部件 i 和部件 j 的空间相对位置与相应模板之间相对位置的形变程度(由部件 i 和部件 j 之间的空间关系计算得出,用于衡量两个部件之间的相对位置是否符合其模板中定义的相对位置)。

人体形态的描述公式为

$$L^*(\beta) = \operatorname{argmin}_L \left(\sum_{i=1}^K \sigma_i + \sum_{i=1}^n \sum_{j=1}^m d_{ij} \right) \quad (4)$$

式中: L^* 为最佳的人体形状配置; β 为形状参数; argmin_L 为损失函数(L)最小化; $\sum_{i=1}^K \sigma_i$ 为所有部件; $\sum_{i=1}^n \sum_{j=1}^m d_{ij}$ 为部件间相对位置的形变量。

通过公式(4)的最小化,确保每个部件都与其相应的模板匹配。并且各个部件之间的相对位置也符合模板中定义的相对位置,从而实现对人体形状的准确估计。此外,采用优化损失函数学习形状参数和部位相对位置变换矩阵,可以进一步提高估计精度。

3 实验部分

3.1 软件配置

采用 Microsoft 公司 2017 开发的软件 Visual Studio;电脑配置为:GPU Nvidia Titan X 16GB RAM, CPU Intel Core i7 - 8700。基于 MFC 平台,使用 Python 语言,采用 TensorFlow 技术框架,使用 64 的小批量梯度下降方法和学习率为 $5e-5$ 的 Adam 优化器进行训练。

3.2 方法

在定量实验中,准备两个测试数据集 T_1 和 T_2 。 T_1 由 1 200 个图像组成, T_2 由 1 000 个图像组成,这些图像包括 50% 的男性和 50% 的女性,其人体姿势处于大致相同的静止状态。为了量化 DVN 方法,对人体 7 个部位进行测量,其中对于需要直线测量的部位,计算两个极端顶点之间的欧氏距离即可。每个部位有 7 个测量值,计算真实数据与测试结果之间的平均误差和标准偏差,人体尺寸测量结果见表 1。人体模型的每个动作序列由 100 ~ 200 帧组成,从每个动作序列中采集 20% 的帧作为测试数据,另外 80% 的帧作为训练数据。训练后,将获得的参数输入网络,以此重建序列,并且与真实数据进行比较。

表 1 人体尺寸测量结果比较

Tab.1 Comparisons of anthropometry measurements		
单位:mm		
测量部位	测试结果	真实数据
胸围	27 ± 25	25 ± 24
臀围	19 ± 16	16 ± 14
腰围	17 ± 15	15 ± 13
肩宽	14 ± 8	13 ± 12
大腿围	15 ± 13	12 ± 11
小腿围	12 ± 10	10 ± 9
小腿长	11 ± 9	8 ± 7

注:通过测量方法对互补结果进行比较,误差以平均值 ± 标准差表示,即“正”和“负”。

3.3 结果与讨论

表 1 给出了人体模型的重构误差,包含训练或测试帧的平均顶点欧氏距离误差^[13]。由表 1 可以看出,测试部位的测量误差保持在较低的水平,轮廓误差(测试结果值减去真实数据值)均小于 5 mm。

测试方法比较如图 4 所示。基于数据集 T_1 和 T_2 ,图 4 展示了博韦尔特、随机森林、典型相关分析、地面真实、双视图网络 5 种不同测试方法。其中:博韦尔特方法稳定,但不够精准;随机森林方法强大,但计算成本高;典型相关分析方法能揭示关系,但对数据分布要求严格;地面真实方法最准确,但难以获得完美结果。由图 4 可知,博韦尔特方法的平均误差约为 9.5 mm,随机森林方法约为 8.2 mm,典型相关分析方法约为 6.7 mm,地面真实方法约为 5.4 mm,而双视图网络方法的平均误差最小,约为 4 mm,表明双视图网络方法在减少误差方面表现最为出色。

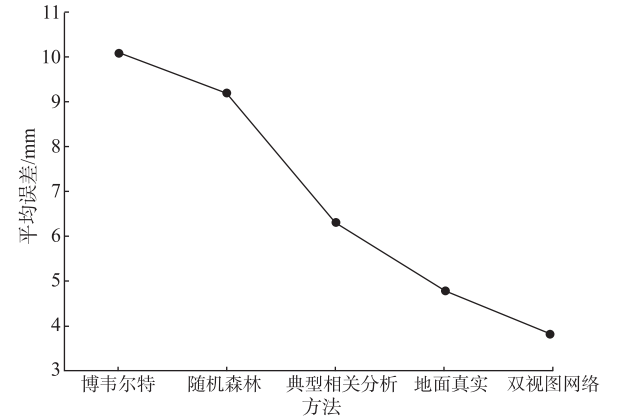


图 4 测试方法比较

Fig.4 Comparisons of testing methods

人体形状估计是计算机视觉领域的一个重要问题,其应用涵盖了人机交互、虚拟现实、运动分析等多个领域。随着深度学习技术的发展,基于卷积神经网络的人体形状估计方法逐渐成为主流。然而,由于人体形态的多样性和复杂性,单一视角的图像往往难以提供足够的信息,从而导致估计精度不高。为了解决这个问题,文中提出了 DVN,该方法采用两种不同轮廓的人体输入方式,从正面和侧面观察人体,以提高估计精度。在人体形状估计的执行过程中,双视图合并层对每个维度执行最大化操作,允许网络从单个图像中分别学习特征,然后通过合并层区分和融合,提高了学习的精确度和后续估计的精确度,同时利用自适应批量归一化^[14]技术进一步提高网络的性能。该方法可对单视图网络进行改进,使得网络能够更好地处理多个视图输入,并从中学习更准确和全面的人体形态和身体各部位的尺寸信息,有效提高人体形状估计的时效性和准确性。

4 结 语

文中提出一种基于卷积神经网络的智能型人体形状估计方法,利用数据集进行人体模型建模,进一步提出人体形状估计与优化算法,并在多个数据集上进行了测试和验证,通过实验对比来验证该方法的有效性和可行性。结果表明,该方法通过巧妙融合两个轮廓的身体深度图,显著提升了基于单个深度图对人体形状的预测精度,增强了网络的准确性和鲁棒性,同时也更好地适应不同的数据分布和场景。基于卷积神经网络智能人体形状估计在人体模型重建、人体形状估计以及服装虚拟现实等领域具有重要意义,应用前景广阔。

参考文献:

- [1] JI Y, JIANG G M, TANG M T, et al. Three-dimensional simulation of warp knitted structures based on geometric unit cell of loop yarns [J]. Textile Research Journal, 2020, 90(23/24) : 2639-2647.
- [2] HASHMI M F, ASHISH B K K, KESKAR A G, et al. FashionFit: analysis of mapping 3D pose and neural body fit for custom virtual try-on[J]. IEEE Access, 2020(8) : 91603-91615.
- [3] 张勇, 吴廷轩, 何滨珂, 等. 3D 人体建模技术探讨 [J]. 计算机时代, 2020(9) : 34-36.
ZHANG Yong, WU Tingxuan, HE Binke, et al. Discussion on 3D human body modeling technology [J]. Computer Era, 2020(9) : 34-36. (in Chinese)
- [4] JI Y, JIANG G M, CONG H L. Sustainable improvements for customized platform effectiveness in garment production [J]. Autex Research Journal, 2019, 19(4) : 355-362.
- [5] 胡凯, 陈旭, 朱俊, 等. 基于多尺度 3D 卷积神经网络的行为识别方法 [J]. 重庆邮电大学学报(自然科学版), 2021, 33(6) : 970-976.
HU Kai, CHEN Xu, ZHU Jun, et al. Multiscale 3D convolutional neural network for action recognition [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2021, 33(6) : 970-976. (in Chinese)
- [6] KIM J E. A study on the application of 3D digital technology for fashion design[J]. Journal of the Korean Society of Costume, 2007(57) : 45-58.
- [7] TAO X Y, CHEN X, ZENG X Y, et al. A customized garment collaborative design process by using virtual reality and sensory evaluation on garment fit [J]. Computers and Industrial Engineering, 2018 (115) : 683-695.
- [8] 金光, 任工昌, 桓源, 等. 基于改进 YOLOv5 的皮革抓取点识别及定位 [J]. 皮革科学与工程, 2024, 34(1) : 32-40.
- JIN Guang, REN Gongchang, HUAN Yuan, et al. Grab point identification and localization of leather based on improved YOLOv5[J]. Leather Science and Engineering, 2024, 34(1) : 32-40. (in Chinese)
- [9] 杨晨晖, 陈辰. 虚拟试衣系统中的模型变形 [J]. 厦门大学学报(自然科学版), 2014, 53(1) : 46-51.
YANG Chenhui, CHEN Chen. Model deformation in virtual dressing system [J]. Journal of Xiamen University (Natural Science), 2014, 53(1) : 46-51. (in Chinese)
- [10] 宋莹, 相思曼, 孙雅致. 基于虚拟仿真技术的服装逆向流程优化设计 [J]. 丝绸, 2022, 59(4) : 59-64.
SONG Ying, XIANG Siman, SUN Yazhi. Reverse flow optimization design of garments based on virtual simulation technology [J]. Journal of Silk, 2022, 59(4) : 59-64. (in Chinese)
- [11] HEIDRICH W, SEIDEL H P. Realistic, hardware-accelerated shading and lighting [J]. Computer graphics, 1999:33.
- [12] DANEREC R, DIBRA E, ÖZTIRELI C, et al. DeepGarment: 3D garment shape estimation from a single image [J]. Computer Graphics Forum, 2017, 36 (2) : 269-280.
- [13] LI H X, WANG Y R. Structural optimization of yoga top based on 3D virtual-reality technology [J]. The Journal of the Textile Institute, 2020, 111(6) : 916-923.
- [14] HUANG L C, YANG R Y. Automatic alignment for virtual fitting using 3D garment stretching and human body relocation [J]. The Visual Computer, 2016, 32(6) : 705-715.

(责任编辑:张 雪)