• 计算机工程与应用 •

基于堆叠沙漏网络的量体特征点定位

邹 昆[1,2]，王伟灿[2]，董 帅[1*]，李文生[1,2]

(1. 电子科技大学中山学院计算机学院 广东 中山 528402；2. 电子科技大学计算机科学与工程学院 成都 611731)

【摘要】为提高复杂背景和任意着装情况下的量体特征点定位精度，将堆叠沙漏网络 (SHN) 引入人体图像量体特征点定位中，并针对 SHN 模型输出特征图分辨率过低导致定位精度不足的问题，构建了一种 Deconv-SHN 模型。一方面用多个反卷积层代替初始模型的输出层以提高输出特征图的分辨率，另一方面基于 Smooth L1 和局部响应对目标函数进行了优化。在自建的 6 700 幅正面人体图像数据集上对 Deconv-SHN 模型、SHN 模型以及传统算法进行实验的结果表明，Deconv-SHN 模型在复杂背景和任意着装情况下的特征点定位精度较传统算法有显著提升，也明显优于 SHN 模型，基本满足人体参数测量应用的要求。

关 键 词 量体特征点定位； 深度学习； 非接触式人体参数测量； 堆叠沙漏网络

中图分类号 TP391.41 文献标志码 A doi:10.12178/1001-0548.2019122

Anthropometric Feature Points Localization Based on Stacked Hourglass Network

ZOU Kun[1,2], WANG Wei-can[2], DONG Shuai[1*], and LI Wen-sheng[1,2]

(1. School of Computer Engineering, Zhongshan Institute, University of Electronic Science and Technology of China Zhongshan Guangdong 528402;
2. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract In order to improve the accuracy of anthropometric feature point localization in complex background and arbitrary dress cases, the stacked hourglass network (SHN) is introduced into the localization of anthropometric feature points in body images. However, the resolution of the SHN model's output feature map is too low to obtain high accurate feature points. So, a Deconv-SHN model is proposed to address this problem. On the one hand, the output layer of the initial model is replaced by several deconvolution layers to improve the resolution of the output feature map. On the other hand, the objective function is optimized based on Smooth L1 and local response. According to the experimental results on the self-built dataset consisting of 6 700 human body images, the localization precision of the Deconv-SHN model in complex background and arbitrary dress cases is significantly higher than that of the traditional algorithm, which is also obviously superior to the SHN model, and basically meets the requirements of anthropometric applications.

Key words anthropometric feature points localization; deep learning; non-contacting anthropometry; stacked hourglass network (SHN)

人体参数测量是服装定制、虚拟试衣、人体建模等应用中的一个重要环节，而基于正交图像的非接触式人体参数测量方法由于其简便易行、适合在网络环境下应用等优点得到了广泛关注。此类方法以人体的正、侧面图像作为输入，利用图像处理算法定位量体特征点并结合辅助信息 (如身高) 计算二维量体数据，最后通过围度拟合获得人体围度数据[1]，其精度在很大程度上依赖于特征点定位的准确性。近年来，国内外学者对量体特征点定位算法做了许多研究，大致可分为两类：基于图像分割的特征点定位和基于统计学习模型的特征点定位。

基于图像分割的特征点定位算法通常先提取整体或局部人体轮廓，然后利用人体形态先验知识进行特征点定位。文献 [2] 在图像差分并二值化后利用标准人体形态特征进行特征点定位，但对人体形态的标准程度要求较高，无法适用于所有体形的人体。文献 [3] 利用颜色信息和 Canny 算子检测人体轮廓，然后利用 Freeman 8 连通链码，通过考虑轮

收稿日期：2019－05－20；修回日期：2020－06－16
基金项目：国家自然科学基金 (61502088)；广东省自然科学基金 (2016A030313018)
作者简介：邹昆 (1980－)，男，博士，教授，主要从事人工智能与计算机视觉方面的研究.
通信作者：董帅，E-mail：dongshuai@zsc.edu.cn

廓上相邻点的方向变化来确定特征点，但当前特征点的检测与上一特征点有依赖关系，易造成检测的不稳定。文献 [4] 使用阈值分割和边缘检测算法提取人体完整轮廓线，然后利用 Harris 角点检测算法进行特征点定位，适应了多变的人体形态，但由于图像上角点过多需手工选取所需特征点。文献 [5] 通过在色调通道进行阈值分割来提取人体区域，并利用形态学方法得到单像素宽的人体轮廓，进而将轮廓划分为不同的分段，将各分段视为一维信号，利用其极小、极大值点来定位特征点。上述算法都要求背景单一且人体着装与背景有显著差异，虽然文献 [5] 也尝试在固定的真实背景下进行图像采集，利用高斯混合模型对背景进行建模，但效果并不理想。文献 [6] 在进行人脸检测后根据先验知识确定特征点所在区域，然后利用带形状约束的非闭合 Snake 模型在提取局部轮廓线的同时定位特征点，减少了图像背景和人体着装带来的干扰，但该算法对初始轮廓的设置有一定依赖性，且在部分复杂背景环境和着装情况下仍然会出现较大误差。

基于统计学习模型的特征点提取算法适用于对柔性体特征点 (如人脸特征点) 的定位，常用的模型有主动形状模型[7](active shape model, ASM) 和主动表现模型[8](active appearance model, AAM) 等。近年来，已开始有学者将这些模型应用到人体特征点定位中。文献 [9] 利用改进的 ASM 模型对人体特征点进行搜索，提高了特征点定位的精度，但其研究是在实验室环境下获取的图像上进行，图像背景单一，干扰较少。此类算法利用的统计模型都存在其自身的缺陷，ASM 只利用了形状信息，AAM 加入了纹理信息对其进行改进，但两种模型都对光照和姿态的变化比较敏感，且在初始值不理想的情况下，都容易陷入局部极值从而使定位精度下降。

以上两类特征点定位算法的精度都依赖于特征工程的构建，而在数据量充足的情况下，相对于手动构建特征工程，深度学习可以提取到更好的特征表达。在计算机视觉领域中，基于深度学习的目标检测、图像分类等的准确率较传统方法有大幅提升，但深度学习在基于正交图像的人体参数测量中的应用则十分少见。文献 [10] 提出了一种基于深度学习的复杂背景和多姿态情况下的人体参数测量方法，利用 deeplabv3 对人体进行语义分割，得到人体轮廓，然后利用 openpose 提取关键点，用于对轮廓进行分割，通过局部轮廓匹配找到数据库中的适配人体模型，将模型的尺寸作为结果返回。该方法需要大规模的人体模型数据库支持，而其提取的关键点也并非量体特征点。由于姿态识别中的人体关节点定位[11-13] 以及人脸分析中的人脸关键点定位[14-16] 与量体特征点定位有许多相似之处，因此本文将用于人体关节点定位的深度神经网络模型引入量体特征点定位中，并对其进行改进，旨在解决传统算法难以在复杂背景和任意着装情况下准确定位特征点的问题，从而能够满足远程服装定制等应用对高精度人体参数测量的要求。

本文采用文献 [12] 提出的用于人体关节点定位的堆叠沙漏网络 (stacked hourglass networks, SHN) 作为实验的基础网络。该网络采用残差模块作为基础模块，利用其构成可以提取不同尺度特征的沙漏网络，此外为了更好地捕获特征点间的空间关系，对多个沙漏网络进行了堆叠。在复杂背景和任意着装情况下，SHN 定位的特征点基本分布在人工标记附近，但距离高精度人体参数测量应用的要求还有一定差距。所以本文在 SHN 基础上利用反卷积层替代初始模型的输出层并修改了原始目标函数，构建了反卷积堆叠沙漏网络 (deconvolutional stacked hourglass networks, Deconv-SHN)。修改后的网络在仅增加少量计算的情况下提高了特征点定位的精度，基本能够满足服装定制等应用对人体参数测量的要求。

1 堆叠沙漏网络

在人体关键点检测中，堆叠沙漏网络[12] 在定位精度上取得了优异的成绩而且经常被应用到其他检测模型中作为提取特征的基础网络。

1.1 残差模块

堆叠沙漏网络的基础模块为残差模块，该模块可在提取高层特征的同时保留低层的信息，其结构如图 1 所示。

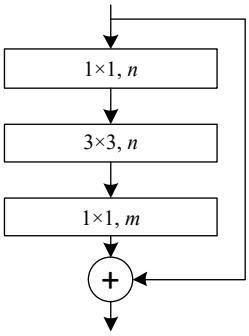

图 1 残差模块结构

该模块首先通过 n 个 1×1 的卷积核将特征降至 n 维，然后利用 n 个 3×3 卷积核进行特征提取，最后利用 m 个 1×1 卷积核将特征升至 m 维 (其中 $n<m$)，这种瓶颈式结构在有效提取特征的同时缩减了计算量及内存使用量。

1.2　沙漏网络

沙漏网络是堆叠沙漏网络的主要组成部件，其结构如图 2 所示。图中浅绿色模块为图 1 所示的残差模块，模块中第 1 行数值表示输入模块的通道数，C_IN 表示输入沙漏网络的通道数，第 2 行数值表示通过模块后输出的通道数；红色模块为下采样层；灰色模块为上采样层；虚线框框出的位置用来更改网络的阶数，如果将框中的内容替换成一个一阶的沙漏子网络，则完成了二阶沙漏网络的构建，依次类推可以构建更高阶的沙漏网络。

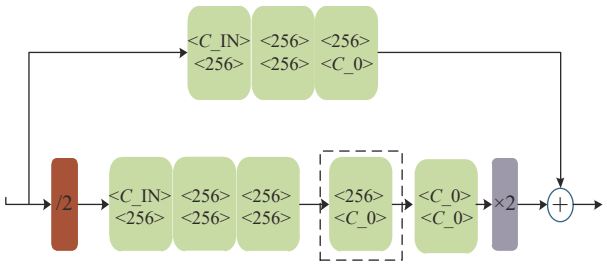

图 2　沙漏网络结构

1.3　堆叠沙漏网络

堆叠沙漏网络则是将沙漏网络进行串行的堆叠。为了解决由于网络加深导致的底层参数难以训练更新的问题，堆叠沙漏网络采用了中继监督策略对底层损失进行监督训练。图 3 展示了包含两个沙漏子网络的二级堆叠沙漏网络。

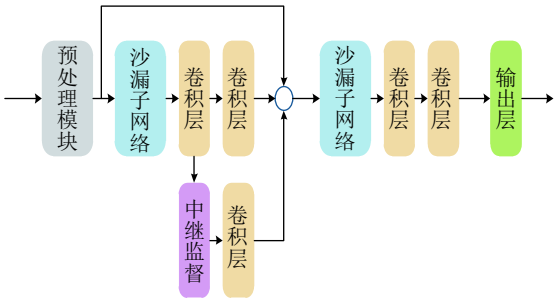

图 3　二级堆叠沙漏网络结构图

2　数据集构建

虽然目前存在许多公开的深度学习人体数据集，但这些数据集中人体姿态各异，不适用于测量人体参数信息。考虑到数据集对实验结果的重要性，本文自建了人体测量数据集，并在此数据集上进行后续的实验。

本文对采集数据时的拍摄条件和人体站姿提出如下要求：尽可能在自然背景和任意着装情况下进行拍摄；拍摄设备位于被拍摄者正前方 3~5 m 且拍摄方向与地面垂直；拍摄人体正面图像时人体基本站姿为：昂首挺胸、掌心向前、双臂张开、双脚脚后跟并拢、前脚掌分开一定角度 (也可以接受自然站立下双脚脚后跟未并拢的情况)；拍摄人体侧面图像时人体站姿为：成立正姿势，手臂自然下垂贴于身体两侧，站姿可参考图 4。

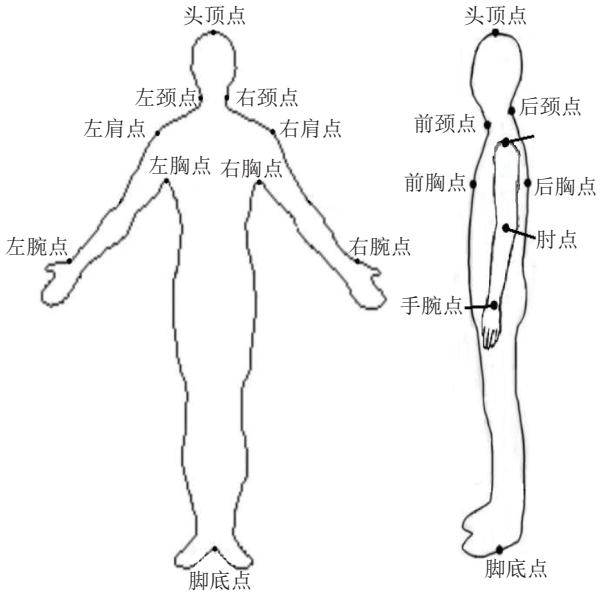

图 4　人体站姿示意图

对每位采集对象拍摄了 1～3 幅正面图像以及 1 幅侧面图像，其中拍摄多幅正面图像时，手臂张开的幅度有一定差异。共采集了 6 700 幅正面图像及 3 300 幅侧面图像。

在对数据集进行标注时参考了国标 GB/T 16160-2017《服装用人体测量的尺寸定义与方法》[17] 中规定的人体特征点，详细标注点名称及位置如图 4 所示。由 10 名标注人员对每幅图像进行标注，取平均值作为最终标注结果。

3　反卷积堆叠沙漏网络

本文通过利用自建数据集中的 5 700 幅图像及文献 [12] 的参数对 SHN 重新训练，然后用 1 000 幅图像对其定位效果进行评估发现，该模型具有较好的普适性，在复杂背景和任意着装情况下仍能得到较为精确的定位结果，然而其精度距离服装定制

等应用对人体参数测量的要求还有一定差距。本文还发现 SHN 中的堆叠次数在 3 级及以上时算法的准确率基本没有提升，所以为了减少网络过拟合的可能性，本文中的实验均在二级堆叠沙漏网络上进行。

对模型的误差来源分析发现，在 SHN 训练过程中，要将高分辨率图像上的特征点位置缩小到低分辨率 (64×64) 的网络输出特征图上，而该变换过程的不可逆性导致无法将网络输出预测值准确地还原到高分辨率图像上，从而导致精度丢失。虽然直接提升模型输入图像的分辨率可以增大输出分辨率，从而减小对真实标记的缩放倍数，但会导致计算量过大，因此本文构建了 Deconv-SHN 模型。

3.1 网络结构

构建的 Deconv-SHN 模型结构如图 5 所示，其中虚线框中的为新增模块，因为输入信息的多少决定了能够还原多少信息，所以增加反卷积的层数需要由网络输入大小来定，当网络输入为 256×256 时，对应增加的反卷积层数为 2。

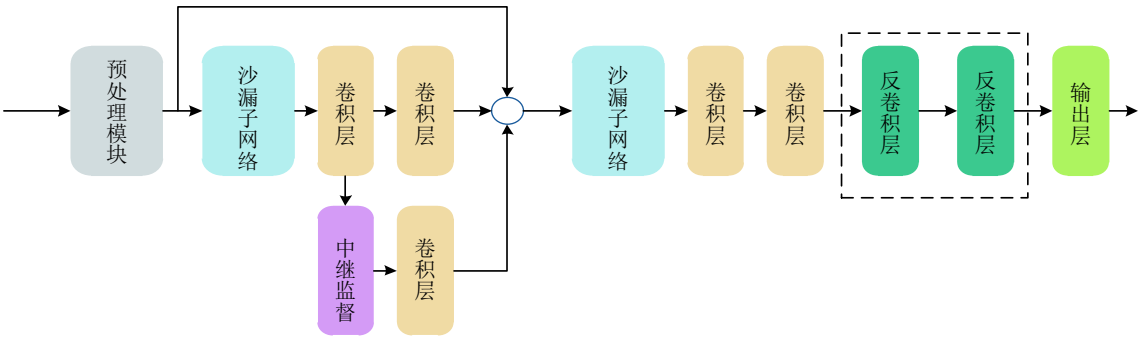

图 5 反卷积堆叠沙漏网络结构

从图 3 和图 5 可见，Deconv-SHN 的结构与 SHN 基本保持一致，而加入的反卷积层可有效减少对真实标记的缩放。

3.2 目标函数优化

3.2.1 基于 Smooth L1 的目标函数优化

SHN 采用了在回归问题中常用的均方误差损失 (也被称为 L2 损失) 作为损失函数对网络进行训练，计算公式为：

$$L_{\mathrm{L2}}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

式中，y_i表示网络的预测值；$\hat{y}_i$表示对应的目标值；n 为样本数目。

L2 损失的平方操作使模型在噪点处产生较大的损失，这相当于给噪点赋予了较大的权重，当模型向减小噪点处损失的方向进行优化时会使模型的整体性能变差，所以本文利用 Smooth L1 损失[18]优化原损失函数，减小网络对噪声点的敏感度，让网络具有更好的泛化能力，Smooth L1 损失函数为：

$$L_{\mathrm{smooth\_L1}}=\frac{1}{n}\sum_{i=1}^{n}d_i$$

$$d_i=\begin{cases}\lambda(y_i-\hat{y}_i)^2 & |y_i-\hat{y}_i|<\delta\\ \delta|y_i-\hat{y}_i|-(1-\lambda)\delta^2 & \text{其他}\end{cases}$$

式中，y_i表示网络的预测值；$\hat{y}_i$表示对应的目标值；λ为损失调节系数，其值通常取 0.5；δ为平方损失和绝对值损失的临界点。

3.2.2 基于局部响应的目标函数优化

在人体参数测量数据集的图像中目标人物只占整幅图像的一部分，如图 6 所示，而预测的关键点应分布在此局部区域内，基于此考虑加入相应的损失可一定程度削弱图像背景区域的干扰，并帮助网络训练。因此，本节基于由人体边界上的特征点确定的外接矩形框 (图 6 中的红框)，设计了基于人体边界框信息的局部响应损失函数。

考虑到有特征点不存在 (不可见) 的情况，该损失函数由两部分构成：一部分为预测的特征点分布在人体边界框中的数目与要预测的特征点数目之间的差值带来的损失；另一部分为出现在人体边界框之外的特征点数目带来的损失。加上之前堆叠沙漏网络中关于点位置预测的 Smooth L1 损失，最后网络的目标函数由这 3 部分损失以加权求和的方式构成：

$$L_{\mathrm{total}}=L_{\mathrm{smooth\_L1}}+\lambda_{\mathrm{inbox}}(n_{\mathrm{pred\_in}}-n_{\mathrm{true\_in}})^2+\lambda_{\mathrm{outbox}}{n_{\mathrm{pred\_out}}}^2$$

式中，$L_{\mathrm{smooth_L1}}$为特征点位置损失；$n_{\mathrm{pred_in}}$为在人体边界框内预测到的特征点数目；$n_{\mathrm{true_in}}$为在人体边界框内真实存在的特征点数目；$n_{\mathrm{pred_out}}$为预测在边界框外的特征点数目；因为各部分损失不在同一

个数量级上，所以加入了权重系数λ_{inbox}和λ_{outbox}用于调节各部分损失在目标函数中所占的比例。

图 6　人体边界框示意图

3.3　模型及训练参数

网络的输入为 256×256 的 3 通道彩色图像，沙漏网络的堆叠次数为 2，训练时 batch_size 设置为 20，网络优化器选用 RMSprop[19]，学习率的初始值为 2.5×10^{-4}，训练过程中步数每增加 50 000 步学习率下降到原来的 10%。为了减少网络过拟合的可能性，在将训练图像输入到网络模型之前采取了随机裁剪、随机修改亮度和对比度等数据增强方法，同时在模型中加入了批归一化[20] 操作。

在后续实验中，如无特殊说明，目标函数中的参数λ和参数δ分别设置为 0.5 和 2，参数λ_{inbox}和λ_{outbox}分别设置为 100 和 50。

4　实验结果与分析

本文以提取人体正面图像中的量体特征点为例进行对比实验。

4.1　特征点定位

文献 [6] 算法在传统算法中特征点定位精度较高，且在一定程度上弱化了对拍摄背景的要求，因此选择该算法作为传统算法的代表与基于深度学习的算法进行对比，而在深度学习算法方面则选择了 SHN 以及本文提出的对 SHN 的 3 种优化方法进行实验对比。实验中 4 种深度学习算法均采用 5 700 幅图像作为训练集，训练时采用 3.3 节中超参数的设置，然后在同样的 1 000 幅图像的测试集上对文献 [6] 算法和 4 种深度学习算法进行评估，测试集和训练集中不存在相同人员。

考虑到数据集中图像分辨率存在差异以及人体在图像中所占比例不一等原因，计算定位特征点与人工标记特征点的归一化距离能更客观地反应定位精度。参考了文献 [12] 中的方法，利用头部在图像中高度的 2/3 对误差进行归一化处理，归一化距离为：

$$D_{\text{norm}}=\frac{D_{\text{img}}}{\frac{2}{3}H_{\text{head}}}$$

式中，D_{img}为定位的特征点与人工标记特征点在图像中的像素距离；H_{head}为图像中的头部高度。一般情况下，一个成年人头高的 2/3 在 20 cm 左右，这样归一化距离在 0.1 以下的特征点定位误差在 2 cm 以内，而服装定制对大部分人体参数的精度要求即是在 2 cm 以内，由此可将 0.1 作为可接受的归一化误差阈值。

图 7 为 5 种算法在不同归一化距离内的特征点检出率曲线图，由于特征点较多，只选取了部分特征点进行展示，其中 Deconv 表示仅做了结构优化的网络，Deconv-S-L1 表示在 Deconv 基础上加了基于 Smooth L1 的目标函数优化后的网络，Deconv-BBox 表示基于局部响应的目标函数优化后的网络。而表 1 则给出了 3 种优化方法及 SHN 在归一化距离小于 0.1 内的各特征点的检出率 (后面简称为 0.1-检出率)。

从图 7 和表 1 可以看到，在特征点定位精度方面，基于深度学习的特征点定位算法比文献 [6] 算法表现出极大优势，所以将深度学习应用到量体特征点定位中是可行的。此外也可看出，网络结构的修改使得检测效果得到大幅度提升，可见模型精度与输出特征图的分辨率有很大关系，而反卷积在只增加相对较少计算量的情况下便可获得较大分辨率的特征图，所以用反卷积修改网络存在其优越性。从 Deconv-S-L1 的检测结果来看，虽然网络在一些特征点定位的精度上没有得到较大的提升但是也没有产生消极的影响，而且在理论上该损失函数可以减小过拟合的风险，所以利用该方法修改目标函数是可取的。从 Deconv-BBox 的检测结果来看，利用该方法修改目标函数后定位效果整体上取得了一定的提升，而且在训练过程中发现，Deconv-BBox 收敛到该效果所需要的迭代次数要比其他网络模型少许多，所以利用该方法修改目标函数是可取的。

图 8 给出了人工标记 (红色十字)、SHN(黄色十字) 以及 Deconv-BBox(绿色十字) 在光线较暗、

背景较为复杂、前后背景差异不明显、光线较亮且光照不均匀的情况下的定位效果，从对比结果可见，优化后的网络模型在对绝大多数特征点的检测中更加接近人工标记的位置。

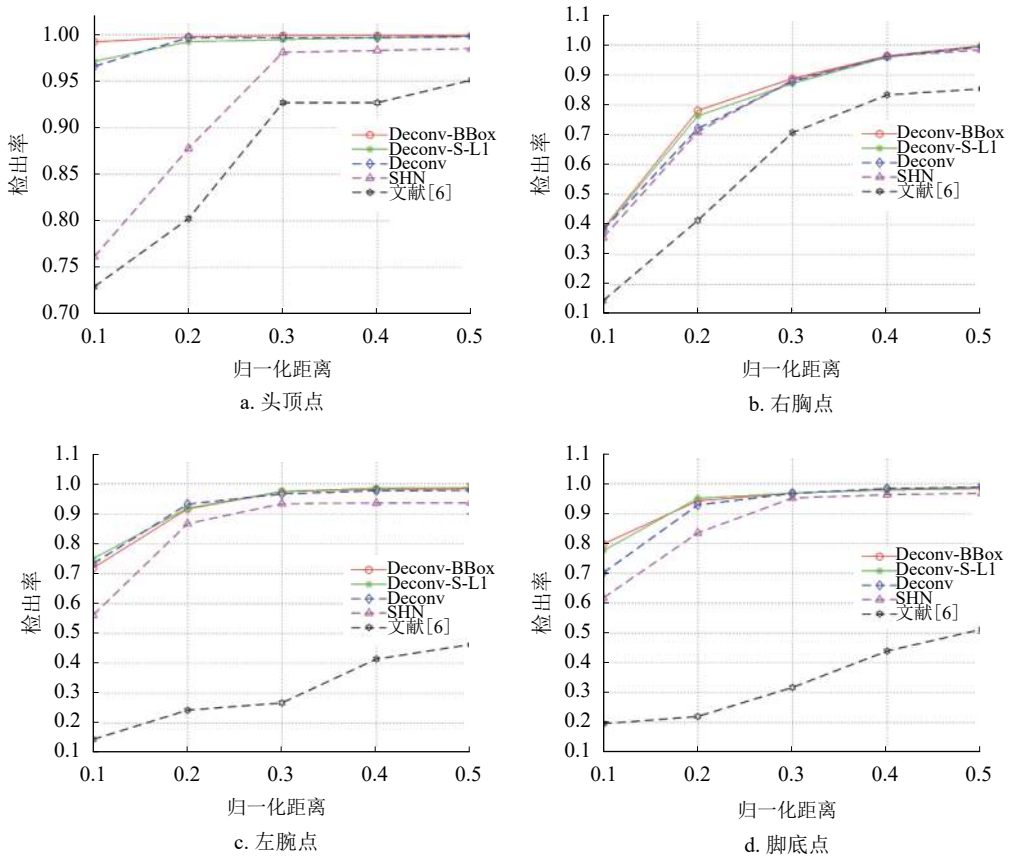

图 7　特征点检出率

表 1　归一化距离小于 0.1 的特征点检出率

%

提取算法	头顶点	左颈点	右颈点	左肩点	右肩点	左胸点	右胸点	左腕点	右腕点	脚底点
文献[6]	72.93	65.86	66.15	26.57	23.72	13.92	14.63	14.64	14.13	19.51
SHN	76.10	72.00	71.90	53.90	60.00	32.10	35.70	56.10	52.50	61.70
Deconv	96.50	92.90	93.80	68.90	72.90	33.90	38.20	73.30	69.70	70.10
Deconv-S-L1	97.10	94.80	93.60	70.80	70.70	35.90	38.20	75.00	72.30	77.50
Deconv-BBox	99.60	95.90	96.10	75.10	76.70	37.00	38.60	71.80	69.50	79.80

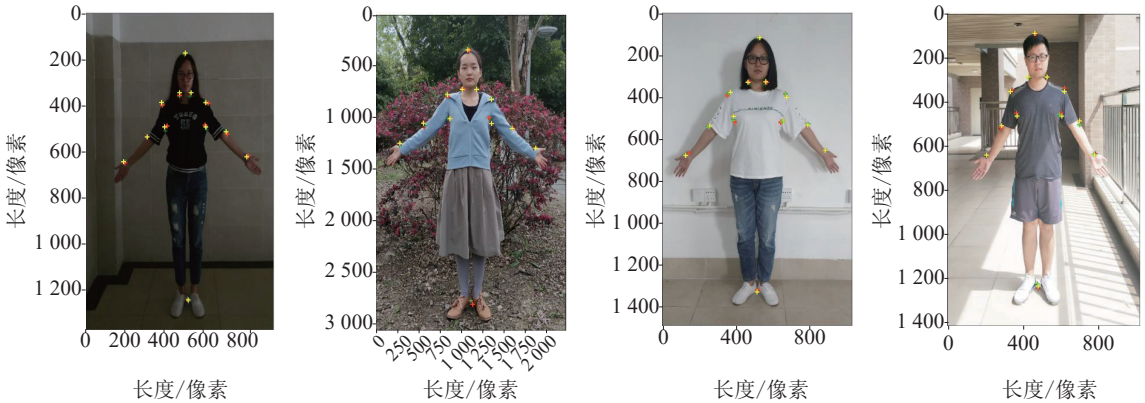

图 8　特征点定位效果对比

图像的训练标签是由人工标记得到的，虽然遵循了国标[17]中的量体特征点定义，但不同人员做的标记还是会有一定差异，一些衣着也会影响标记人员对部分特征点的把握 (如左右胸点 (腋下点))。所以为了观察人工标记的误差并给网络模型的定位效果一个对比标准，进行了以下实验：选取 30 个具有标记经验的人对 1 000 幅图像进行特征点标记，然后将 30 份标记结果的平均值作为 1 000 张图像的真实标记值，最后在此基础上分别计算人工标记的 0.1-检出率均值以及之前实验中表现最好的 Deconv-BBox 模型的 0.1-检出率，实验结果如表 2 所示。

表 2　人工标记与网络提取 0.1-检出率对比

%

特征点	人工标记	网络模型
头顶点	99	95
左颈点	98	92
右颈点	98	93
左肩点	78	76
右肩点	83	86
左胸点	68	69
右胸点	69	70
左肘点	71	72
右肘点	72	74
左腕点	73	83
右腕点	71	86
脚底点	90	82
平均	80.8	81.5

从表 2 可以看到，人工标记的效果和网络预测的效果在误差的分布上较为类似，对较容易定位的头顶点、左右颈点和脚底点都取得了较为理想的结果，其次是肩点，而受衣着影响较大的胸点、肘点和腕点的定位效果相对较差。对较容易定位的点，人工标记的结果相对较好，但对于较难定位的点，网络模型的预测结果反而较好。分析其原因可能是，对于较容易定位的特征点，人工标记时产生误差的可能性较小，而且其工作是在原始图像上进行的，分辨率较高，而模型在预测过程中存在对图像的压缩，丢失了部分图像的细节信息；对于较难分辨的特征点的位置，不同的标记人员会得到不同的估计值，即使同一个标记人员也可能在长期的标记工作中，对特征点位置的估计也会产生变化，而网络模型在学习的过程中，为了获得更小的误差可能会偏向于学习一种平均水平，从而使得网络模型在较难分辨的特征点上取得了更好的定位效果。从平均水平来看，网络模型对特征点的定位效果略优于人工标记的结果。

从表 2 中人工标记一列可以看出，人工标记也会产生一定的误差，而在实验中是以人工标记作为真实值对网络进行训练，所以标记的质量也会对网络模型预测的准确率产生影响，如果能够获得更合规的图像并在上面进行更精确的人工标记，网络模型的准确率应该还能够得到进一步的提升。

此外，本文也对 Deconv-BBox 网络损失函数的参数设置进行了两组实验。实验 1 将参数λ_{inbox}和λ_{outbox}分别固定为 100 和 50，调整 Smooth L1 损失中的λ和δ进行训练，得到 10 个特征点的 0.1-检出率均值 (称其为平均检出率)，如表 3 所示。可以看到，λ的取值对结果影响不明显，多数情况下取 0.5 更优，而δ取 1、2、3 时差异不大，但其值显著增大后结果有较明显的变差。实验 2 将λ和δ分别固定为 0.5 和 2，然后调节参数λ_{inbox}和λ_{outbox}，这两个参数主要用于调节 Smooth L1 损失与局部响应损失之间的权重，而局部响应损失的两部分的数量级相同，故先进行了不同数量级的测试，将二者值同设为 1、100 和 1 000，得到的平均检出率分别为 72.63%、73.92% 和 73.05%，比设为 100 时明显更优。进而在该数量级下，进一步调节这两个参数的值，得到结果如表 4 所示。当λ_{inbox}和λ_{outbox}有一个设为 0 时，网络表现有较明显的变差，说明这两部分对结果都是有贡献的，二者均非 0 时结果差异不算大，当分别取 100 和 50 时效果最佳。

表 3　损失函数参数设置实验 1 结果

平均检出率/%		δ					
		1	2	3	10	50	100
λ	0.2	73.21	73.72	73.39	72.15	72.21	71.25
	0.5	73.93	74.01	73.94	72.53	72.38	71.23
	0.8	73.13	73.71	73.18	72.59	72.30	71.03

表 4　损失函数参数设置实验 2 结果

平均检出率/%		λ_{outbox}			
		0	50	100	150
λ_{inbox}	0	72.59	73.09	72.98	72.95
	50	73.11	73.88	73.75	73.72
	100	73.09	74.01	73.92	73.69
	150	73.05	73.83	73.77	73.63

4.2　实际尺寸预测

为了将深度学习方法应用到人体参数测量中，需要将图像中的像素距离转换为实际距离。本文沿用文献 [6] 中的距离换算方法，利用头顶点和脚底

点之间的距离和被拍摄者的身高信息确定像素距离与真实尺寸的比例尺S：

$$S = \frac{H}{\sqrt{(P_{\mathrm{h}}x - P_{\mathrm{f}}x)^2 + (P_{\mathrm{h}}y - P_{\mathrm{f}}y)^2}} \tag{1}$$

式中，P_{h}表示提取的头顶点；P_{f}表示脚底点；H表示被测者实际身高。

为了单纯考量特征点定位带来的误差，本文仅对肩宽、臂长、胸宽等正面尺寸信息进行估计，同时选择利用人工标记特征点计算得到的尺寸信息作为标准结果。由于文献 [6] 算法无法准确定位特征点，实际尺寸估计不再与其进行对比实验。表 5 给出了在 150 幅有真实身高数据的人体正面图像上，对 SHN 和 Deconv-SHN 模型进行真实尺寸预测的误差对比，可以看到，无论是平均误差、最大误差还是误差小于 2 cm 占比，Deconv-SHN 模型均明显优于 SHN 模型，且误差小于 2 cm 的样本占比最低也在 80% 左右。由于这部分测试集中人体都穿着较紧身的衣服，所以对胸点的定位准确了许多，胸宽的预测也较精准；对于预测精度表现相对较低的左臂长，可通过选取左右臂长的均值作为最终预测结果来一定程度提高预测精度。

表 5　真实尺寸预测误差

测量项目	SHN			Deconv-SHN		
	平均误差/cm	最大误差/cm	<2 cm占比/%	平均误差/cm	最大误差/cm	<2 cm占比/%
颈宽	1.32	3.78	76.6	0.46	2.03	97.3
肩宽	2.18	7.55	50.0	1.30	4.36	80.2
胸宽	1.28	5.49	81.3	0.95	2.74	90.6
左臂长	1.92	7.27	65.3	1.41	3.86	79.6
右臂长	2.02	7.10	58.6	1.21	4.13	84.0

5 结 束 语

为了解决在复杂背景和任意着装情况下传统量体特征点定位算法精度不够的问题，本文提出将 SHN 应用到量体特征点定位中，并针对其不足，构建了 Deconv-SHN。实验结果表明：在复杂背景和任意着装情况下，深度学习方法的定位效果明显优于传统算法；且与 SHN 相比，Deconv-SHN 定位精度更高，预测的实际尺寸误差能够基本满足服装定制等应用对人体参数测量的要求。

本文的研究工作得到了广东省中山市社会公益重大专项 (2017B1014) 的资助，在此表示感谢！

参 考 文 献

[1] 骆顺华, 王建萍. 基于二维图像非接触式人体测量方法探析[J]. 纺织学报, 2013, 34(8): 151-155.
LUO Shun-hua, WANG Jian-ping. Research on 2D image-based non-contact anthropometric technology[J]. Journal of Textile Research, 2013, 34(8): 151-155.

[2] 邓卫燕, 陆国栋, 王进, 等. 基于图像的三维人体特征参数提取方法[J]. 浙江大学学报 (工学版), 2010, 44(5): 837-840.
DENG Wei-yan, LU Guo-dong, WANG Jin, et al. Extraction of feature parameters of three-dimensional human body based on image[J]. Journal of Zhejiang University (Engineering Science), 2010, 44(5): 837-840.

[3] LIN Y L, WANG M J J. Automated body feature extraction from 2D images[J]. Expert Systems with Applications, 2011, 38: 2585-2591.

[4] 丁中娟. 基于平面投影轮廓获取人体尺寸的研究[D]. 上海: 东华大学, 2017.
DING Zhong-juan. Research on acquisition of body dimension basing on the outline of plane projection[D]. Shanghai: Donghua University, 2017.

[5] ASLAM M, RAJBDAD F, KHATTAK S, et a1. Automatic measurement of anthropometric dimensions using frontal and lateral silhouettes[J]. IET Computer Vision, 2017, 11(6): 434-447.

[6] 邹昆, 马黎, 傅瑜, 等. 面向人体参数测量的非闭合 Snake 模型局部轮廓提取[J]. 计算机辅助设计与图形学学报, 2018, 30(1): 147-154.
ZOU Kun, MA Li, FU Yu, et al. Anthropometry oriented local contour extraction based on unclosed Snake model[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(1): 147-154.

[7] COOTES T F, TAYLOR C J, COOPER D H, et a1. Active shape models—their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.

[8] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[C]//The European Conference on Computer Vision. Freiburg: Springer, 1998: 484-498.

[9] 朱欣娟, 熊小亚. 基于改进 ASM 模型的人体特征点定位和建模方法[J]. 系统仿真学报, 2015, 27(2): 286-294.
ZHU Xin-juan, XIONG Xiao-ya. Feature point positioning and modeling approach for human body based on improved ASM[J]. Journal of System Simulation, 2015, 27(2): 286-294.

[10] CHEN Y, WANG Y. An anthropometric dimensions measurement method using multi-pose human images with

complex background[C]//3rd International Conference on Computer Graphics and Digital Image Processing. Rome: IOP Publishing, 2019, 1335: 1-7.

[11] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//The IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7291-7299.

[12] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]//The European Conference on Computer Vision. Amsterdam: Springer, 2016: 483-499.

[13] CHU X, YANG W, OUYANG W, et al. Multi-context attention for human pose estimation[C]//The IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 5669-5678.

[14] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection[C]//The IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 3476-3483.

[15] RANJAN R, PATEL V M, CHELLAPPA R. HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2019, 41(1): 121-135.

[16] ZHANG J, SHAN S, KAN M, et al. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment[C]//The European Conference on Computer Vision. Zurich: Springer, 2014: 1-16.

[17] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 服装用人体测量的尺寸定义与方法, GB/T 16160-2017[S]. 北京: 中国标准出版社, 2017.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration. Anthropometric definitions and methods for garment, GB/T 16160-2017[S]. Beijing: Chinese Standard Press, 2017.

[18] GIRSHICK R B. Fast R-CNN[C]//The IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.

[19] TIELEMAN T, HINTON G. RMSprop: Divide the gradient by a running average of its recent magnitude [EB/OL]. [2018-07-15]. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[20] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. Lille: IMLS, 2015: 448-456.

编　辑　漆　蓉