

复杂场景下的人体轮廓提取及尺寸测量

吴泽斌¹, 张东亮¹, 李基拓², 麻菁¹, 信玉峰³

(1. 浙江大学计算机科学与技术学院, 浙江 杭州 310027;

2. 浙江大学机械工程学院, 浙江 杭州 310027;

3. 三明学院艺术与设计的学院, 福建 三明 365004)

摘 要: 自然拍摄的人体照片由于背景图案较为复杂, 采用传统基于图片色彩空间或能量梯度的图像处理难以准确地识别人体的轮廓。采用神经网络的方法, 可以提高识别的精度。但是, 一般的神经网络方法由于计算量与参数规模较大, 难以在移动终端部署。因此, 提出了一种轻量级的神经网络策略以提取人体轮廓。该网络采用 MobileNet V2 与 U-Net 框架, 通过构建特定姿态的人体数据集进行训练, 识别相应的人体轮廓形状。人体轮廓经过提取关键点、拟合回归分析等后续处理, 可估算人体的尺寸。该方法可应用在移动终端上, 通过拍摄的人体照片的方法测量人体的尺寸。实验表明, 该方法能准确地提取复杂背景照片中的人体轮廓并测量尺寸, 在速度与存储占用方面较一般神经网络有一定优势。

关 键 词: 图像处理; 轮廓提取; 人体尺寸测量; 轻量级神经网络; 深度学习

中图分类号: TP 391

DOI: 10.11996/JGJ.2095-302X.2020050740

文献标识码: A

文章编号: 2095-302X(2020)05-0740-10

Contour recognition and information extraction of human bodies in complex scenes

WU Ze-bin¹, ZHANG Dong-liang¹, LI Ji-tuo², MA Jing¹, XIN Yu-feng³

(1. College of Computer Science and Technology, Zhejiang University, Hangzhou Zhejiang 310027, China;

2. School of Mechanical Engineering, Zhejiang University, Hangzhou Zhejiang 310027, China;

3. School of Art and Design, Sanming University, Sanming Fujian 365004, China)

Abstract: Due to the complexity of background, it is difficult to accurately recognize the contour of human body by traditional image-processing methods based on color space or energy gradient. Neural network can improve the accuracy of recognition. However, due to the large scale of computations and parameters, it is difficult to deploy the general neural network methods in mobile devices. Therefore, we proposed a lightweight neural network to extract human body contours. This network utilized MobileNet V2 and U-Net framework to recognize the contours of human bodies by building a human-body dataset with specific poses for training. The contours of human bodies can be used to

收稿日期: 2020-04-08; 定稿日期: 2020-05-28

Received: 8 April, 2020; Finalized: 28 May, 2020

基金项目: 国家自然科学基金项目(61732015, 61972340); 浙江省重点研发计划项目(2018C01090)

Foundation items: National Natural Science Foundation of China (61732015, 61972340); Key R&D Program Projects in Zhejiang Province (2018C01090)

第一作者: 吴泽斌(1995-), 男, 浙江杭州人, 硕士研究生。主要研究方向为图形图像处理、计算机视觉等。E-mail: wuzb1995@zju.edu.cn

First author: WU Ze-bin (1995-), male, master student. His main research interests cover image processing, computer vision, etc.

E-mail: wuzb1995@zju.edu.cn

通信作者: 张东亮(1971-), 男, 浙江温州人, 教授, 博士, 博士生导师。主要研究方向为计算机辅助设计、计算机图形学、交互设计等。

E-mail: dzhang@zju.edu.cn

Corresponding author: ZHANG Dong-liang (1971-), male, professor, Ph.D. His main research interests cover computer-aided design, computer graphics, interactive design, etc. E-mail: dzhang@zju.edu.cn

measure the sizes of human bodies after the subsequent processes, such as the extraction of key points and analysis of fitting regression. This method can be applied to mobile terminals to measure the body sizes by taking pictures. Experiments show that this method can accurately extract the contours of human bodies in photos with complex backgrounds and measure the body sizes, and that it possesses some advantages over the general neural network in terms of speed and storage.

Keywords: image processing; contour extraction; body measurement; lightweight neural network; deep learning

远程人体尺寸的测量一直是服装定制领域和虚拟试衣技术的关键。随着网上支付、移动支付的兴起,网购服装时尺码不准的问题也随之产生。由于尺码问题,网上服装交易常伴随着大量退换货,造成了买卖双方时间与金钱的损失。远程人体尺寸测量的发展可以有效减少这类纠纷。基于手机等移动设备的远程测量由于所需设备小巧,易于使用,更适合用户操作。测量尺寸数据后,结合移动端的虚拟试衣技术,用户能随时查看自己的穿衣效果。真实尺寸搭建的虚拟人体还能用于医学假体制造,刑事侦察等领域。

在自然场景下,受光线条件影响,传统图像识别方法很难从照片中准确测量人体尺寸。以往使用三维扫描方法^[1]或通过 kinect^[2]结合额外的照片深度数据的方法进行远程非接触式人体尺寸测量。由于操作设备限制,上述方法应用部署较难。

近年来随着神经网络不断发展,复杂图片语义分割准确性的提高,使得精确提取人体轮廓测量尺寸成为可能。文献[3]首次提出全卷积网络(fully convolutional network, FCN),应用于语义分割任务,获得了良好的效果。U-Net^[4]利用编码器-解码器结构的特点,通过特征的拼接融合技术,在医学影像上取得了不错的成绩。BADRINARAYANAN 等^[5]基于编码器-解码器结构提出的 SegNet 网络,将语义分割的性能进一步提高。有了准确的轮廓,对照片进行轮廓及关键点的精确识别,即可基于照片进行人体尺寸测量。

然而,为达到精确度的提升,算法的空间占用与耗时也显著增加,对于移动端程序的实际应用造成了一些困难。因此,MobileNet^[6], SqueezeNet^[7]等轻量化网络被提出。本文使用 MobileNet V2^[8]与 U-Net^[4]网络相结合的方法,构建了存储空间要求小、运算速度快的神经网络架构,以此提取人体轮廓。通过该轮廓提取方法,本文利用正面与侧面的 2 张特定姿态全身照,结合用户的身高数

据,完成了人体三围等关键尺寸的测量。实验表明,这种测量方法在速度、存储容量上有一定优势,适合部署在移动设备中。

1 相关工作

基于轮廓的人体尺寸测量方法在便携性上有很多优势。先前的研究者在轮廓提取方向提出了很多优秀的操作方法,有较强的实用性。

1986年,CANNY^[9]提出了边缘检测算子 canny 算子,借此可在相对简单的背景环境中提取部分轮廓信息。之后,ROTHER^[10]提出了使用高斯混合模型建模的 GrabCut 算法,结合交互信息实现图像分割,得到特定对象的轮廓。HOSHIAI 等^[11]提出了通过主动形状模型(active shape model, ASM)和主动外观模型(active appearance model, AAM)提取人体上半身轮廓的方法,该方法可用于基于有特定特征的人体尺寸测量。

随着神经网络的发展,图像语义分割技术与人体轮廓提取有了长足进步。神经网络解决了前人方法中需要交互操作,对复杂背景鲁棒性不强等问题。HE 等^[12]提出了 Mask RCNN 网络,实现了在鲁棒性、精度、速度上都较好的实例分割;SHEN 等^[13-14]提出了基于 FCN 和姿态模板的 PortraitFCN+网络,用于提取半身人像轮廓。

受限于输入样本的数据格式,分割模型有一定的局限性,边缘很难准确分割。图像融合类的自然图像抠图(image matting)任务通过预测图像中每个像素上前景与背景的比例,能获得更精确的轮廓。抠图作为一种“软分割”技术,使得边缘信息被更好的保留。XU 等^[15]建立了可用的自然图像抠图数据集,结合编码器-解码器结构,使提取的边缘更准确,细节得以更多地保留,让神经网络能够处理自然图像抠图任务;CHEN 等^[16]通过结合分割模型和抠图模型,获得了一种全自动的人像抠图方法;LUTZ 等^[17]引入生成对抗网络提供了抠

图任务新的解决方案; ZHANG 等^[18]通过分支网络进一步提升了结果的连通性与准确性。

本文结合自然抠图任务的特点, 使用规模小、精度高的 MobileNet V2 结合 U-Net 高效的编码器-解码器结构, 预测图片中每个像素的人体前景比例, 再通过图像二值化获取特定姿势人体的轮廓信息。轮廓信息通过 HU^[19]不变矩筛选, 活动轮廓模型(snake)^[20]纠正可获得人体尺寸关键点。结合人体尺寸关键点, 本文实现了适用于移动设备部署的人体尺寸测量方法。

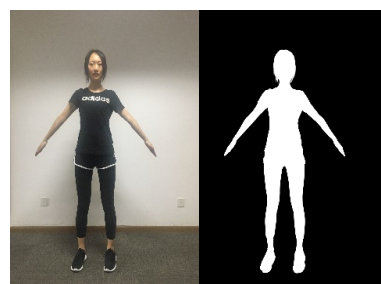
2 人体轮廓提取

为获得快速且小巧的轮廓提取算法, 本文使用 U-Net^[4]作为基础的编码器-解码器结构框架, 利用 MobileNet V2^[8]中的反向残差模块, 在尽可能保留模型精度的前提下缩减卷积操作的运算规模。本文以抠图任务作为训练目标, 使用 RGB 图像作为输入, 输出能够区分该图像人物与背景的 alpha 图层。通过对输出进行简单的轮廓检测, 即可获得所需的轮廓信息。为了提高算法的准确性, 可对人物正面样本和人物侧面样本分别进行训练。

2.1 数据集

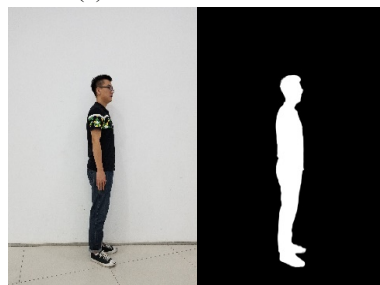
由于特定姿势的人物抠图任务数据集样本数量有限, 本文构建了特定姿势人物数据集。数据集初始数据样本为 41 位身着不同衣服的被试拍摄的 471 张特定姿态不同背景的照片, 其中正面照 234 张, 侧面照 237 张。初始数据样本进行手动前景与背景分离后, 可拆分为初始训练集与初始测试集, 拆分保证两者的被试不重叠(图 1)。为了扩大数据集规模, 使网络得到更大的泛化能力, 本文通过 XU 等^[15]的方法扩充实验数据。初始训练集的前景与 MSCOCO 数据集^[21]中标注不为人物的图片背景合成, 得到经过扩充的训练集。初始测试集的前景与实地拍摄的多张室内外无人照片合成, 与初始测试集合并后作为测试集进行测试。训练中所有样本的背景均不相同。

背景的替换无法代替现实场景下光照的影响。本文受 LUTZ 等^[17]的数据集生成方法启发, 将拍摄获得的原始数据通过 $[-10^\circ, 10^\circ]$ 的随机小角度旋转, 获得了与原始数据有细微差别的扩充的数据集, 并入训练样本, 获得实际的光影效果(图 2)。所有训练集样本通过小范围随机地调整亮度、饱和度、对比度, 扩充色彩范围, 进一步降低光线的影响。数据集配置见表 1。



(a) 女性正面

(a) Front view of female

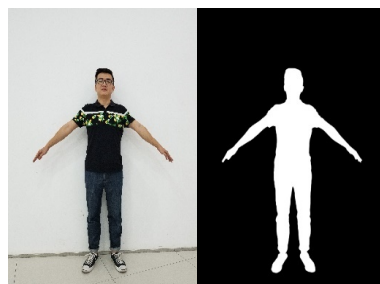


(b) 男性侧面

(b) Side view of male

图 1 不同性别被试侧面与正面原始拍摄数据

Fig 1. Side and front shooting data of different gender subjects



(a) 原始拍摄

(a) Original image



(b) 色彩角度增强

(b) Color and angle enhancement sample



(c) 背景替换增强

(c) Background replacement enhancement sample

图 2 数据集图像样例

Fig. 2 Dataset image sample

表1 人体轮廓提取数据集配置(张)
Table 1 Human contour extraction dataset configuration (sheet)

数据类型	训练集		测试集	
	前景	样本	前景	样本
合成样本		17 340		240
正面样本	原图变换样本	204	3 060	30
	总量	20 400		270
侧面样本	原图变换样本	207	3 105	30
	总量	20 700		270
样本总量	411	41 100	60	540

对不同姿态样本训练不同的神经网络模型,可以有效提高模型的精度。特定姿势人物数据集包含人物全身信息,其中人物姿态便于人体尺寸测量的下一步开展。使用该数据集能够对本文提出的神经网络进行有效地训练。

2.2 网络模型

2.2.1 倒置残差

经典计算机视觉网络训练过程中,标准卷积核参数多,卷积过程所需计算开销极大。为降低计算开销与存储空间大小。MobileNet V2^[8]中结合深度可分离卷积的倒置残差方法,显著降低了卷积操作的参数数量和计算开销,同时保证深度网络精度的可靠性。

如图3所示,对于长度为 I_h 、宽度为 I_w 、通道数为 M 的输入图片,若要得到通道数为 N 的输出,标准卷积核参数数量为 $D_k \times D_k \times M \times N$ 。一次标准卷积所需计算量($Cost$)为

$$Cost = D_k \times D_k \times M \times N \times I_h \times I_w \quad (1)$$

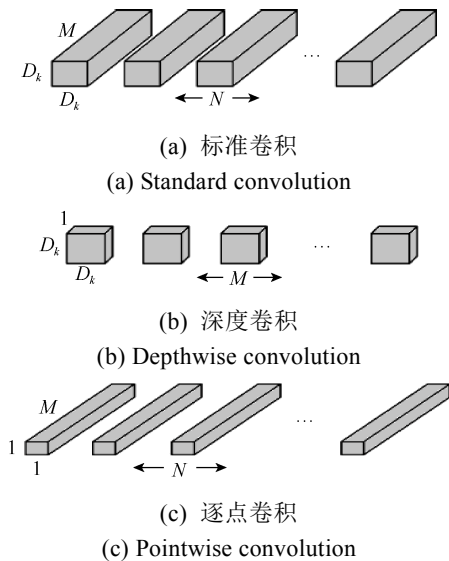


图3 标准卷积与深度可分离卷积的比较

Fig. 3 Comparison of standard convolution and depthwise separable convolution

深度可分离卷积将标准卷积核拆分为 $D_k \times D_k \times M$ 的深度卷积核与包含 $M \times N$ 个参数的逐点卷积核。深度卷积对每个通道提取相应的空间特征。逐点卷积将各通道的特征通过线性组合融合。逐点卷积后输出与标准卷积规格相同,参数数量与计算开销均减少了一个数量级。深度可分离卷积将卷积参数减少至 $D_k \times D_k \times M + M \times N$,计算开销为

$$Cost = (D_k \times D_k \times M + M \times N) \times I_h \times I_w \quad (2)$$

倒置残差方法在应用深度可分离卷积的同时,通过跨越连接,增强了算法精度,其结构如图4所示。深度卷积只能提取每个通道相应的特征,对输入的通道数较敏感。倒置残差先扩张输入的通道数,提取特征后再压缩,以此在步长为1的卷积中使用残差网络跨越连接的技巧。低维度下,激活函数ReLU容易造成信息的丢失,使用线性激活函数输出可以保留更多信息。

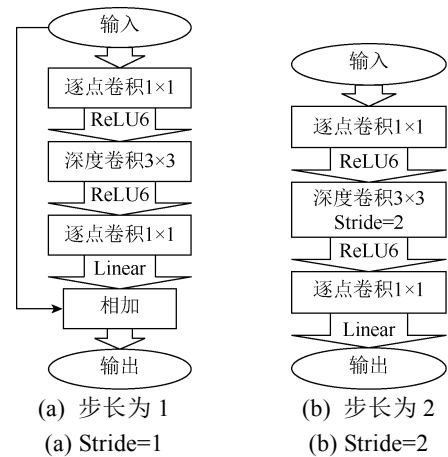


图4 倒置残差结构

Fig. 4 Inverted residual structures

本文采用 MobileNet V2 作为特征提取器,将倒置残差的通道扩张倍数设置为6倍,使用步长为2的倒置残差模块代替池化层,获得了较好的特征提取效果。

2.2.2 网络结构

网络结构以 U-Net 的编码器-解码器结构为基础。编码器即为特征提取器,使用 MobileNet V2 搭建,输入为三通道图像,输出的长宽是原始输入的1/32,通道数为320的特征张量。解码器结构使用双线性插值算法进行上采样,经过卷积操作后将提取的特征张量逐层还原,获得16通道输出张量。16通道输入进行标准卷积滤波和双线性插值后,获得与原始输入长宽一致的单通道 alpha 图像。

整体网络结构如图 5 所示。编码器网络的超参数与 MobileNet V2 网络基本相同, 具有 1 个卷积层和 7 个线性瓶颈模块, 由于只作为特征提取器, 为保留更多信息, 去除最后的标准卷积层与平均池化层。解码器网络具有与之对应的 9 个卷积层和 4 个拼接融合层。通过 4 个拼接层, U-Net 将编码器与解码器中规格相同的特征张量拼接起来, 再经过卷积操作融合, 以获得更多特征, 增加精确度。相对 U-Net 中使用转置卷积进行上采样的方法, 使用双线性插值的上采样方法不需要训练额外的参数, 计算开销更少。

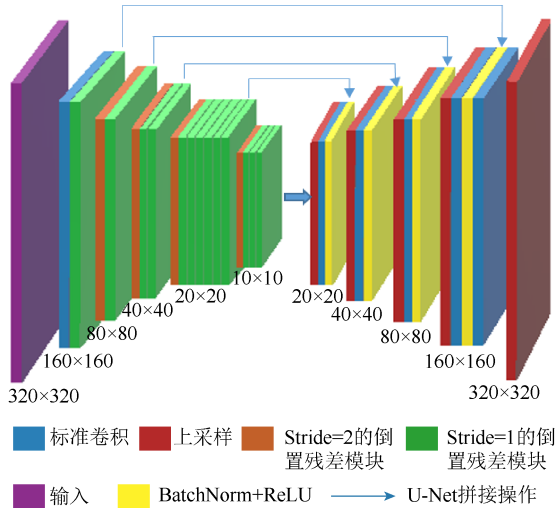


图 5 轮廓提取网络结构

(以 3 通道 320×320 RGB 图片作为输入)

Fig. 5 The structure of contour extraction network (Take 3-channel 320×320 RGB picture as input)

2.2.3 损失函数

人体尺寸测量对于轮廓点的精确度要求较高, 文献[15]针对图像融合采用的损失函数可以有效的保证轮廓细节。

在文献[15]损失函数的基础上, 本文调整 α 预测损失和合成损失的线性组合比例, 使用新的合成图像方法, 可以定义整个网络的损失函数。 α 预测损失定义为 α 通道真实值 α_g 和输出的 α 通道值 α_p 之间的 L_1 损失(绝对值之差)。合成损失被定义为灰度化的真实输入图像 c_g 和预测的合成图像值 c_p 之间的 L_1 损失。单个像素 i 的损失函数为

$$L^i = \gamma \|\alpha_p^i - \alpha_g^i\|_1 + (1 - \gamma) \|c_p^i - c_g^i\|_1 \quad (3)$$

其中, $\alpha_p^i, \alpha_g^i, c_p^i, c_g^i$ 分别对应图像矩阵 $\alpha_p, \alpha_g, c_p, c_g$ 在像素 i 处的值; γ 为比例系数。

由于训练集中包含的一部分未经背景替换的训练样本, 可使用伪背景与预测的前景合成 c_p 。

真实输入 c_g 和与其对应的 α 通道真实值 α_g 经过处理可以得到前景为黑的伪背景。由于伪背景与真实背景存在差异, 在本文实验中 γ 设定为 0.6。与文献[15]方法相比, 本文对整体计算了损失, 而不是仅针对边界区域。

3 人体尺寸测量

在人体轮廓中获取尺寸数据分为 2 步: 提取人体特征点和使用特征点拟合尺寸。对人体轮廓特定区域进行扫描可以找到与人体尺寸数据相关的特征点相对位置。以身高输入为比例尺, 利用特征点相对位置和性别信息, 能完成对人体肩宽、三围等数据的拟合, 以此可以测量人体尺寸, 分析体型。

3.1 人体特征点

对于正面照片, 从上至下分别提取了头顶、肩点、指尖、腋窝、三围高度点和脚底的特征点位置; 对于侧面照片, 则提取头顶、脚底及三围高度点位置。正面与侧面照片需按指定姿势拍摄(图 6)。

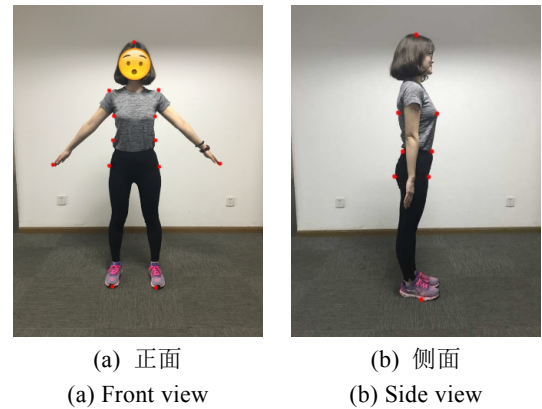


图 6 拍照姿态及特征点位置
Fig. 6 Posture and feature point position

根据“服装用人体测量的尺寸定义与方法”^[22], 通过人体三围高度与身高的比例数据, 结合人体侧面曲线的凹凸特性, 可以扫描局部轮廓提取各个特征点信息。首先, 头顶、脚底以及正面的指尖位置能够快速确定; 其次, 头顶到指尖的轮廓变化程度能确定肩点, 指尖到脚底的轮廓能确定腋窝点作为正面胸围高度的上线; 最后, 正面与侧面三围点位置由三围身高比确定范围, 通过侧面曲线及腋窝点位置进行调整。

3.2 尺寸拟合

三围信息等数据无法直接从图像提取, 需要

利用二维的宽度、厚度信息间接拟合得到。二维宽度与厚度信息由特征点相对位置和身高确定。常用的拟合手段有基于截面形状的曲线拟合法^[23]与基于统计数据的回归分析法。本文采用回归策略进行拟合。先前的研究者使用多种函数回归方式对人体尺寸数据进行实验, 包括线性回归、二次函数、三次函数、指数函数、对数函数^[24]以及运用神经网络的尺寸拟合方法^[25], 并进行了比较。

本文对比了先前研究者的回归策略, 平衡精确度与效率后, 参考了黄秀丽^[26]和刘国联等^[27]的回归方法分别完善了青年男性和女性的拟合模型, 用于拟合人体肩宽和三围。具体函数及参数见表 2。

表 2 人体尺寸拟合公式		
Table 2 Body size fitting formulas		
类别	拟合公式	
	女性	男性
胸围	$1.853 \times \text{胸宽} + 1.748 \times$	
	$\text{胸厚} + 0.004 \times \text{胸宽} \times$	$4.188 \times \text{胸宽} - 0.058 \times$
	$\text{胸宽} + 0.005 \times \text{胸厚} \times$	$\text{胸宽} \times \text{胸宽} + 0.052 \times$
	$\text{胸厚} - 0.01 \times \text{胸宽} \times$	$\text{胸宽} \times \text{胸厚} - 21.187$
	$\text{胸厚} - 7.6779$	
腰围	$1.7566 \times \text{腰宽} + 1.122 \times$	$1.822 \times \text{腰宽} + 1.550 \times$
	$\text{腰厚} + 0.009 \times \text{腰厚} \times$	$\text{腰厚} - 0.003 \times \text{腰宽} \times$
	$\text{腰厚} - 0.002 \times \text{腰宽} \times$	$\text{腰厚} - 4.026$
	$\text{腰厚} + 1.9201$	
臀围		$4.351 \times \text{臀宽} - 1.519 \times$
	$2.465 \times \text{臀宽} - 0.025 \times$	$\text{臀厚} + 0.106 \times \text{臀宽} \times$
	$\text{臀宽} \times \text{臀宽} + 0.043 \times$	$\text{臀厚} - 0.073 \times \text{臀宽} \times$
	$\text{臀宽} \times \text{臀厚} + 3.384$	$\text{臀宽} - 0.018 \times \text{臀厚} \times$
		$\text{臀厚} - 9.681$

3.3 筛选与细化

为提高人体尺寸测量的鲁棒性, 测量方法中增加了筛选与细化操作。

神经网络输出得到的 α 图由 $HU^{[19]}$ 不变矩进行筛选, 判断是否为特定姿势人体全身图像。经测试, 由于 $HU^{[19]}$ 不变矩对形状的敏感性, 该方法能有效判断神经网络输入照片是否符合规范。使用该方法进行初步筛选后再进行特征点提取。

因为神经网络输入经过压缩, 获得的人体关键点信息进一步细化后再拟合人体尺寸能获得更好的精度。初步轮廓图与特征点位置一起经过缩放, 得到与原图长宽一致的轮廓图 and 对应的初步特征点位置。初步特征点位置可以通过活动轮廓模型 (snake)^[20] 细化, 如图 7 所示。该模型以局部区域放大后的轮廓线为初始值, 对关键点所在的局部区域迭代, 可以获得更准确的关键点位置信息。

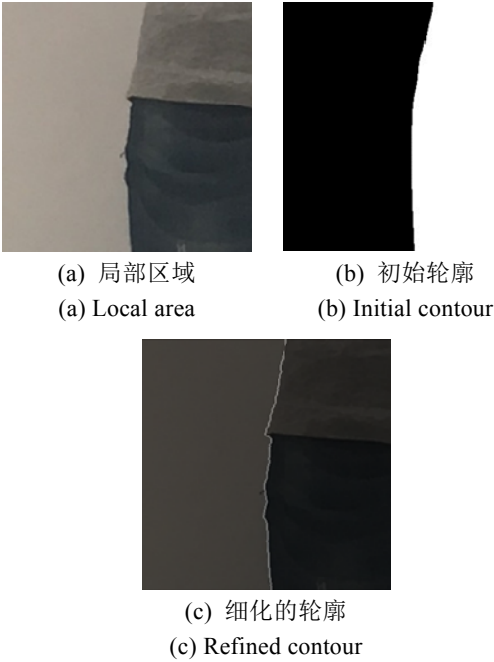


图 7 局部区域细化效果
Fig. 7 Local area refinement effect

为保证精确度, 拍照设备需经过校准。相机镜头拍摄人体时横向会产生畸变。本文对采用的红米 4X 手机进行了畸变矫正。将手机竖直水平面固定于离地 1.5 m 的位置, 拍摄距手机 3 m 的宽度为 95 cm 的门框。经分析得到以下照片不同比例位置的横向畸变数据(表 3), 并以此进行矫正。

表 3 照片横向畸变系数表		
Table 3 Lateral distortion coefficients of photos		
相对高度(0~1)	相对中心位置的横向畸变系数	
0 (顶部)	1.008 6	
0.20	1.004 2	
0.35	0.997 8	
0.50	1.000 0	
0.65	1.001 0	
0.80	1.005 3	
1.00 (底部)	1.010 7	

经分析, 红米 4X 手机镜头大致以枕型畸变为主, 且失真不明显。通过分段线性函数可获得与相对高度相关的连续畸变系数曲线。相应位置的人体尺寸通过该曲线矫正, 代替计算量较大的照片整体单应性矩阵变换。镜头畸变大致人体尺寸测量分析时手机位置及拍摄距离与设备调教时保持一致。

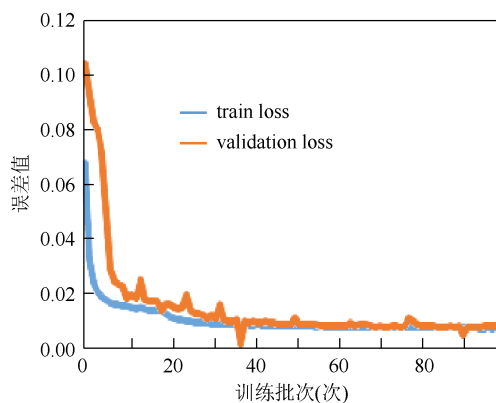
4 实验结果与分析

实验采用 PyTorch 框架训练轮廓提取神经网络, 应用时将特定姿态(图 5)正、侧面人体彩色全

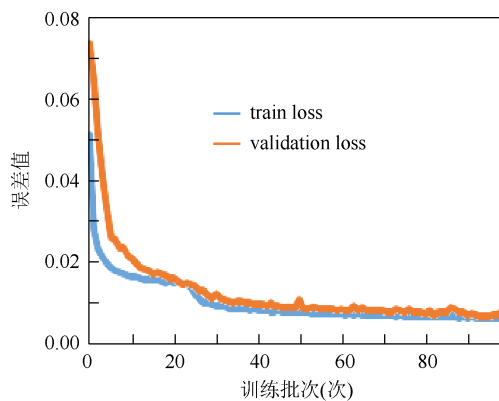
身照缩放至 320×320 分别输入, 提高运行速度。模型转化为 TorchScript 形式, 通过 LibTorch 函数库实现在 C++ 中调用, 便于部署。模型训练使用 320×320 大小图片作为输入, 经过预先训练, 然后进行微调。使用 Adam 作为优化器, 并应用了交叉验证方法。模型学习率设置为 10^{-4} , 每批训练样本数为 8。

4.1 可靠性与局限性

本文通过采用训练样本与验证样本 9:1 的方式进行 K 折交叉验证(即 $K=10$), 当验证集 Loss 下降停滞时结束训练。采用 Adam 作为优化器, 可以降低过拟合的几率。训练集与验证集 Loss 变换曲线如图 8 所示。



(a) 正面模型曲线
(a) Curve of front model



(b) 侧面模型曲线
(b) Curve of side model

图 8 训练过程 Loss 变换曲线

Fig. 8 Loss function transformation curve of training process

由图 8 可知, 训练集与验证集的 Loss 均随训练批次的增加而降低, 且随后趋于稳定。正面模型验证集 Loss 扰动较多。整体未表现出明显的过拟合现象。

由于人体尺寸测量聚焦于特定姿势的图像, 本文训练、验证及测试所用数据集也采用特定姿势。因此, 数据集本身量不大, 获得的训练模型对测量所需特定姿势仍具有较好的效果。但也由于对特定姿势的训练, 经测试, 模型对其他姿势的人像照片表现一般。

4.2 时间与空间

本文采用 torchstat 对多个神经网络结构在输入为 320×320 的三通道 RGB 图片的情况下进行了参数规模、浮点计算量与内存消耗的计算(表 4)。

表 4 不同神经网络计算资源占用情况

Table 4 Computing resource occupancy of different neural

网络	参数规模(M)	GFLOPs
ResNet18	11.600	3.720
PSPNet	4.930	9.320
SqueezeNet	1.240	1.740
MobileNet V2	0.185	0.320
本文方法	0.240	1.480

与图像分割网络 PSPNet^[28]、残差网络 ResNet18^[29]、轻量级神经网络 SqueezeNet^[7]相对比, 本文方法在参数规模和浮点计算量上都有一定优势。

表 5 不同神经网络实际模型规模与处理时间

Table 5 Actual model scale and processing time of different neural networks

网络	每张图片处理时间	模型文件规模
Mask-RCNN	2 min	500 MB
PGN	5 min	1 GB
本文方法	10 s	9 MB

使用 CPU (i3 双核 3.7 G)模拟移动处理器运行 Mask-RCNN^[12]和 PGN^[30]网络, 与本文方法进行对比实验(表 5)。结果表明, 算力有限时, 本文方法图像处理速度较快, 占用空间也较少。

4.3 精度分析

本文使用 3 个指标用于评估预测的 alpha 图的质量^[31]: 平均交并比(mIoU)、均方误差(mean square error, MSE)和绝对误差和(sum of absolute differences, SAD)。梯度误差和连通性误差反映人类观察者的感知视觉质量, 在尺寸测量领域作用有限。实验使用本文构建的测试集测试, 在指标计算时将预测的 alpha 值和真值归一化为 0 到 1, 且由像素数平均。测试采用输入均为 320×320 的三通道 RGB 图片。

PGN 可以对人体各部分进行有效的实例分割, 这里测试 Mask-RCNN 与 PGN 对人体整体分割的二分类效果, 输出为二分类概率(alpha 通道图)。通过对特定姿势进行训练(表 6), 本文方法能够有效地保证获得 alpha 图的精确度。

使用 Mask-RCNN, PGN 和本文方法分别进行实际尺寸测量, 定性分析如图 9 和图 10 所示。可以发现, 本文方法在腋窝、脚尖等细节处有一定

优势, 使得对人体尺寸的测量更精确。

表 6 各神经网络精确度分析

Table 6 Accuracy analysis of each neural network

网络	mIoU (%)		MSE (10-3)		SAD (10-3)	
	正面	侧面	正面	侧面	正面	侧面
Mask-RCNN	87.8	92.3	17.10	7.23	19.70	8.84
PGN	94.2	94.4	5.23	3.29	8.55	5.20
本文方法	95.6	96.0	2.89	1.95	7.26	4.33

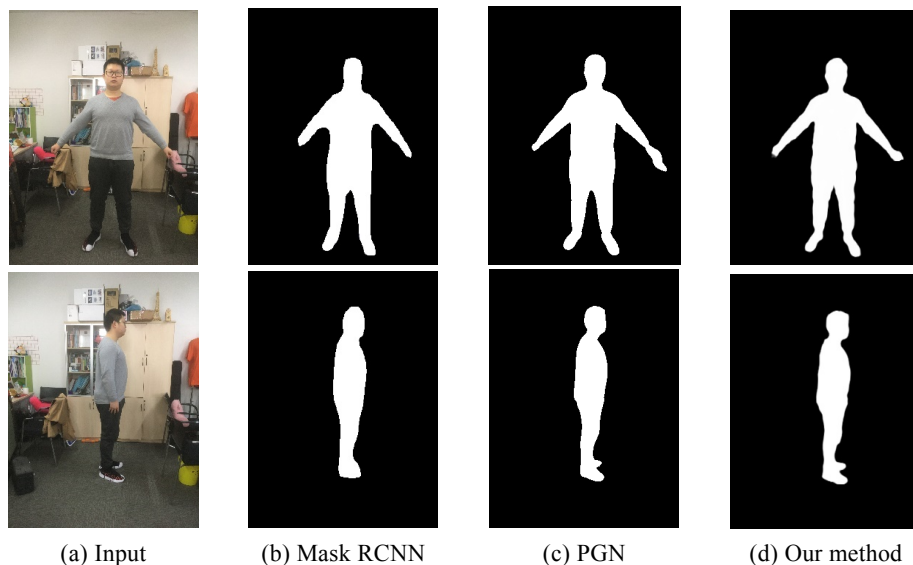


图 9 不同方法测试中人像分割效果对比

Fig. 9 Comparison of image segmentation effects in different methods

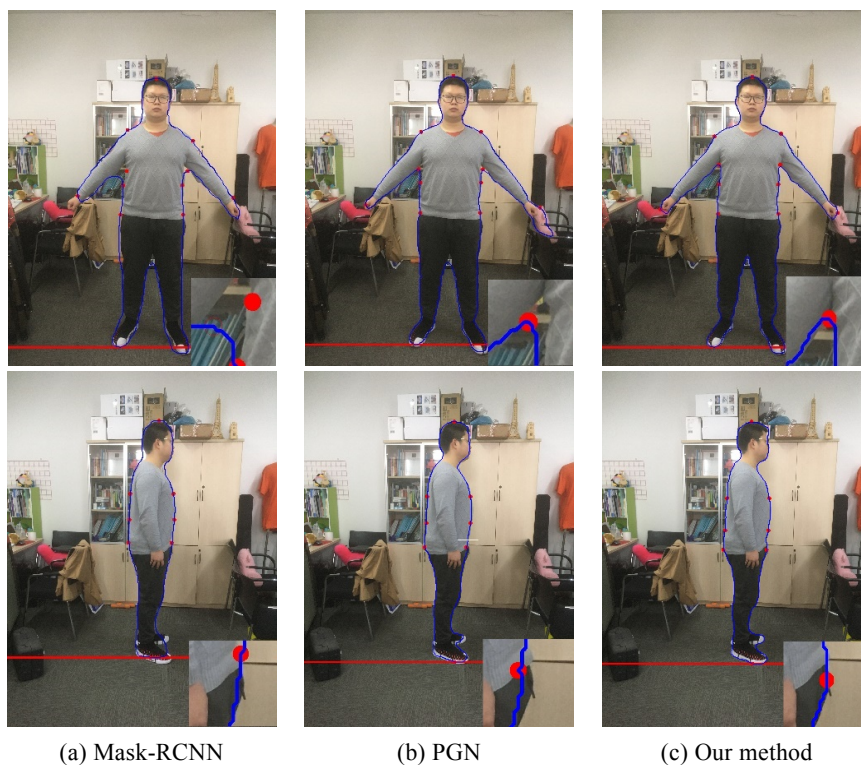


图 10 不同方法测试中人体轮廓(蓝线)及特征点(红点)提取效果对比

Fig. 10 Comparison of human contour (blue line) and feature point (red point) extraction in different methods

4.4 测量误差分析

使用本文的测量方法对 8 位不同性别、穿着贴身的被试进行了人体尺寸拟合测试, 结果见表 7。由于照片中穿着衣物影响和拟合式的误差, 由人体曲线和比例获得的三围位置会有一定偏差, 造成测量的围度有误差。本文方法相对误差保证在 3% 内, 基本达到与手工测量、围度拟合相当的精度, 且误差水平恒定。

表 7 尺寸测量平均误差

Table 7 Average errors of human body size measurement

性别	绝对误差(cm)				相对误差(%)			
	肩宽	胸围	腰围	臀围	肩宽	胸围	腰围	臀围
女	0.57	1.19	1.93	1.32	1.51	1.44	2.23	1.34
男	0.76	1.55	1.91	1.29	1.65	1.69	2.38	1.45

文献[26]对三围及其对应的宽度、厚度做了相关分析与偏相关分析, 并对三围二阶最小二乘回归分析获得的拟合式进行了拟合程度检验。其中, 臀围与臀宽、臀厚的相关性较胸围腰围略小一些。同时臀部模型的复合系数 R 值为 0.851, 略小于胸围的 0.902 与腰围的 0.922, 这可能是臀围误差的成因之一。对被试手动测量辅助数据进行拟合并分析, 得到的结果验证了这一点。

对比手工测量获得的辅助数据与由本文方法从照片得到的辅助数据后, 可发现, 衣物的贴身程度会影响特征点定位, 从而对测量误差产生影响。由于衣物无法完全贴身, 腋窝点的定位会偏下, 胸宽数据也可能偏大。其中男性侧面曲线模糊, 胸围高度受腋窝控制, 误差更明显。同时, 由于本文测量算法对照片中腰部高度的定位主要依赖高度比例, 而不像胸部、臀部定位可以参考侧边的突起曲线, 数据误差也较大。图像分辨率与镜头畸变也可能是影响结果的因素之一。

5 总结与展望

本文提出了一种基于轻量化神经网络的轮廓提取方案, 使用 U-Net 框架, 引入 MobileNet V2 作为特征提取器, 在保证较高精度的前提下, 降低了网络的存储空间和卷积操作的运算量, 以便在移动端部署。本文创建了一组特定动作的人体全身图像数据集, 用于训练该网络。基于这项轮廓提取方案, 本文实现了一种小巧快速的人体尺寸测量解决方案。

本文方法也有一些待优化之处, 如数据集的规模较小、人体尺寸测量方法仍需改进等; 模型开销的进一步缩减也将是下一步深入研究的内容。在训练样本中去除衣物影响, 或采用端到端的神经网络拟合人体尺寸等方法值得下一步尝试。轮廓提取方案亦可以应用于移动端的三维人体模型生成, 为虚拟试衣技术提供帮助。

参考文献

References

- [1] 郭云昕, 张微, 刘咏梅, 等. 三维人体测量技术的现状和比较[J]. 国际纺织导报, 2016, 44(8): 38-40, 42-44, 46.
- [2] GUO Y X, ZHANG W, LIU Y M, et al. The situation and comparison of three-dimensional body measurement technology[J]. Melliand, 2016, 44(8): 38-40, 42-44, 46 (in Chinese).
- [3] DAO N L, DENG T, CAI J F. Fast and automatic body circular measurement based on a single kinect[C]//Signal and Information Processing Association Annual Summit and Conference (APSIPA) 2014 Asia-Pacific. New York: IEEE Press, 2014: 1-4.
- [4] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241.
- [6] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [7] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2019-12-30]. <http://arxiv.org/abs/1704.04861>.
- [8] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[C]//International Conference on Learning Representations (ICLR). New York: IEEE Press, 2017: 1-13.
- [9] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 4510-4520.
- [10] CANNY J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.
- [10] ROTHER C. GrabCut: interactive foreground extraction using iterated graph cuts[J]. ACM Transactions

- Graphics, 2004, 23(3): 309-314.
- [11] HOSHIAI K, FUJIE S, KOBAYASHI T. Upper-body contour extraction and tracking using face and body shape variance information[C]//Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots. New York: IEEE Press, 2008: 391-398.
- [12] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE Press, 2017: 2980-2988.
- [13] SHEN X Y, TAO X, GAO H Y, et al. Deep automatic portrait matting[M]//Computer Vision – ECCV 2016. Heidelberg: Springer, 2016: 92-107.
- [14] SHEN X Y, HERTZMANN A, JIA J Y, et al. Automatic portrait segmentation for image stylization[J]. Computer Graphics Forum, 2016, 35(2): 93-102.
- [15] XU N, PRICE B, COHEN S, et al. Deep image matting[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2017: 311-320.
- [16] CHEN Q, GE T Z, XU Y Y, et al. Semantic human matting[C]//ACM Multimedia Conference. New York: ACM Press, 2018: 618-626.
- [17] LUTZ S, AMPLIANITIS K, SMOLIC A. AlphaGAN: generative adversarial networks for natural image matting[C]//British Machine Vision Conference (BMVC). Heidelberg: Springer, 2018: 259-276.
- [18] ZHANG Y, FAN L, HUANG Q, et al. A late fusion CNN for digital matting[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019: 7461-7470.
- [19] HU M K. Visual pattern recognition by moment invariants[J]. IEEE Transactions on Information Theory, 1962, 8(2): 179-187.
- [20] KASS M, WITKIN A, TERZOPOULOS D. Snakes: active contour models[J]. International Journal of Computer Vision, 1988, 1(4): 321-331.
- [21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[M]//Computer Vision – ECCV 2014. Heidelberg: Springer, 2014: 740-755.
- [22] 中国纺织工业联合会. GB/T 16160-2017 服装用人体测量的尺寸定义与方法[S]. 北京: 中国标准出版社, 2017.
- China National Textile and Apparel Council. GB/T 16160-2017 Anthropometric definitions and methods for garment[S]. Standards Press of China, Beijing, China. 2017 (in Chinese).
- [23] 王玉秀. 非接触式人体测量系统的研究: 图像处理、人体围度曲线分析与计算[D]. 天津: 天津工业大学, 2003.
- WANG Y X. Research on non-contact anthropometric system -- image processing, analysis and calculation of human body circumference curve[D]. Tianjin: Tiangong University, 2003 (in Chinese).
- [24] 尚笑梅, 卢晨, 王辉. 基于自动测体的围度曲线拟合模型[J]. 纺织学报, 2010, 31(4): 128-131, 138.
- SHANG X M, LU C, WANG H. Human body circumference curve fitting model based on automatic measurement[J]. Journal of Textile Research, 31(4): 128-131, 138 (in Chinese).
- [25] 齐啸, 计忠平. 基于三维人体模型的尺寸测量和拟合[J]. 现代计算机: 专业版, 2019(7): 74-81.
- QI X, JI Z p. Size measurement and fitting based on 3D Human body model[J]. Modern Computer: Professional Edition, 2019(7): 74-81 (in Chinese).
- [26] 黄秀丽. 基于数字图像的青年女体测量系统研究[D]. 苏州: 苏州大学, 2009.
- HUANG X L. Research on measurement of young female's bodies based on digital images[D]. Suzhou: Soochow University, 2009 (in Chinese).
- [27] 刘国联, 季开宸, 金蓉, 等. 青年男体非接触式测量系统围度拟合的完善[J]. 丝绸, 2014, 51(10): 20-25.
- LIU L G, JI K C, JIN R, et al. Perfection of girth fitting in non-contact body measurement system of young male[J]. Journal of Silk, 2014, 51(10): 20-25 (in Chinese).
- [28] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2017: 6230-6239.
- [29] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016: 770-778.
- [30] GONG K, LIANG X D, LI Y C, et al. Instance-level human parsing via part grouping network[M]//Computer Vision – ECCV 2018. Heidelberg: Springer, 2018: 805-822.
- [31] RHEMANN C, ROTHER C, WANG J, et al. A perceptually motivated online benchmark for image matting[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2009: 1826-1833.