

7-15-2020

Foot Measurement Based on Deep Alignment Network

Shi Min

1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China;;

Hanqin Yao

1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China;;

Chunpeng Li

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;;

Liangchen Chen

3. School of Applied Technology, China University of Labor Relations, Beijing 100048, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Foot Measurement Based on Deep Alignment Network

Abstract

Abstract: Foot measurement plays an important role in many areas. Limited by equipment and algorithms, three-dimensional foot measurement cannot make a convenient and quick foot measurement. *A method is proposed by combining the image measurement with the deep neural network.* Based on the physiological structure analysis of foot, key points are extracted and the measurement parameters are defined. *During the key point detection of foot, the activation function and loss function of DAN (Deep Alignment Network) model is optimized, and a data acquisition method is defined based on the handheld camera.* Foot key points are detected, and main parameters are measured. Experimental results show that collecting data based on handheld camera can conveniently measure foot parameters and the precision is high.

Keywords

anthropometrics, foot measurement, target detection, deep network, DAN model

Recommended Citation

Shi Min, Yao Hanqin, Li Chunpeng, Chen Liangchen. Foot Measurement Based on Deep Alignment Network[J]. Journal of System Simulation, 2020, 32(7): 1267-1278.

基于深度 Alignment 网络的足部测量

石敏¹, 姚瀚钦¹, 李淳芃², 陈良臣³

(1. 华北电力大学控制与计算机工程学院, 北京 102206; 2. 中国科学院计算技术研究所, 北京 100190;
3. 中国劳动关系学院应用技术学院, 北京 100048)

摘要: 足部测量在很多领域有重要作用。三维足部测量受到设备和算法的限制, 难以方便快捷地测量足部。结合图像测量和深度神经网络, 提出一种方便快捷的足部测量方法。基于足部生理结构分析, 提取足部关键点并基于关键点定义了测量参数; 针对足部关键点检测, 优化了 DAN(Deep Alignment Network)模型的激活函数和损失函数, 并定义了一种基于手持相机的数据采集方式; 检测足部关键点, 测量足部主要参数。结果显示, 该方法基于手持相机采集数据能够便捷地测量足部参数, 并具有较高的精度。

关键词: 人体测量; 足部测量; 目标检测; 深度网络; DAN 模型。

中图分类号: TP391.9 文献标识码: A 文章编号: 1004-731X(2020)07-1267-12

DOI: 10.16182/j.issn1004731x.joss.19-VR0467

Foot Measurement Based on Deep Alignment Network

Shi Min¹, Yao Hanqin¹, Li Chunpeng², Chen Liangchen³

(1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China;
2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
3. School of Applied Technology, China University of Labor Relations, Beijing 100048, China)

Abstract: Foot measurement plays an important role in many areas. Limited by equipment and algorithms, three-dimensional foot measurement cannot make a convenient and quick foot measurement. A method is proposed by combining the image measurement with the deep neural network. Based on the physiological structure analysis of foot, key points are extracted and the measurement parameters are defined. During the key point detection of foot, the activation function and loss function of DAN (Deep Alignment Network) model is optimized, and a data acquisition method is defined based on the handheld camera. Foot key points are detected, and main parameters are measured. Experimental results show that collecting data based on handheld camera can conveniently measure foot parameters and the precision is high.

Keywords: anthropometrics; foot measurement; target detection; deep network; DAN model

引言

人体测量学是人类学的一个分支学科, 通过



收稿日期: 2019-08-30 修回日期: 2019-11-04;
基金项目: 国家自然科学基金面上项目(61972379),
国家十三五重点研发计划(2017YFC0804900);
作者简介: 石敏(1975-), 女, 山西, 博士, 副教授,
研究方向为虚拟现实; 姚瀚钦(1994-), 男, 广东, 硕
士生, 研究方向为计算机视觉。

测量人体各部分尺寸可以研究、探讨人体的形态特征、类型、个体与群体的差异、发展规律等。近年来, 人体测量学被广泛用于医学诊断、犯罪医学和生物医学应用^[1]。足部测量是人体测量学的一个分支, 在许多领域起到重要的作用。

在网上购鞋时, 消费者需要测量自身的足部参数来购买合适的鞋子; 在鞋产品个人定制中, 设计师需要测量顾客的足部参数, 以此设计出合

适美观的鞋子；制鞋工艺中，需要对全国的足型数据进行统计和分析后设计鞋码；在医学领域中，医生需要测量患者的足部尺寸来判断疾病或者进行矫正；在犯罪医学中，足部的尺寸特征可以帮助警方查找犯人。

传统测量足部的方法为手工测量。手工测量具有简单易操作的优点，但也存在着效率低、劳动强度大的缺点。随着计算机技术和硬件的快速发展，非接触式测量得到了快速发展。该方法利用计算机视觉和图像处理技术恢复了足部的三维形态，用点云、网格来描述足部模型，能够很好的展现足部的形态，对足部进行精确的测量。三维重建的方法可以分为主动式重建和被动式重建。主动式重建能够重建出高精度的三维模型，但是需要昂贵的设备支持。被动式的方法利用相机获取图像，然后通过特定算法计算得到物体的立体空间信息。被动式的方法获取数据方便，但其算法复杂度高，重建时间长，鲁棒性也较差，最后重建的模型精度也比较低。

虽然三维重建的精度高，但具有设备昂贵、重建时间长等问题，并且部分测量任务并不需要重建物体的三维模型，因此一些学者基于图像对物体进行测量。基于图像的测量是指结合图像和图像处理技术对物体进行测量，具有高效、高精度、非损伤测量等特点，在工业、农业、医学等领域得到了很大的关注。但图像测量在足部测量方面并没有得到广泛的应用。

随着机器学习的发展，许多研究者将机器学习与自身领域结合，取得了不错的成果。部分研究者将测量和机器学习相结合，提高测量精度。本文在此基础上结合图像测量和深度学习，提出一种方便快捷地足部测量方法。该方法不同于先重建再测量的三维测量方法，首先通过对足部的生理结构进行研究，分析得到足部的关键尺寸和相关的特征点，再通过改进的深度网络检测图像中足部关键点的位置，最后利用图像中参照物与关键点的信息，实现对足部的测量。

为了提高模型的性能，本文设计了一种数据采集方式，该方式利用手持相机采集数据，具有便捷的优点，采集的数据用于模型的训练和测试。基于采集的数据集，进行了实验，并对结果进行了量化分析。

1 相关工作

1.1 基于三维重建的测量方法

近年来，随着计算机技术和硬件的发展，三维测量在许多领域中得到广泛的应用，例如工业零件测量、服装设计、农业等。三维测量可以分为接触式三维测量和非接触式三维测量。典型的接触式测量工具是三坐标测量机。

由于接触式测量具有受测量目标的限制，易破坏测量目标、速度慢等缺点，更多人把目光放到了非接触式测量，非接触式测量得到迅速的发展。Song 等^[2]通过深度相机 Kinect 重建人体，测量了人体的尺寸。Zhong 等^[3]改进了相位测量轮廓术，并开发了一个测量铁路货车适配器的系统。王俊青等^[4]通过结构光重建足部表面，再用足底扫描仪得到足部信息，通过拼接两种信息得到完整足部形态，并对足部的常见参数进行测量。毋戈等^[5]通过多个深度相机来重建足部，并提出了一种自动测量足部模型的方法。

在足部的三维测量中，首先需要重建出足部模型，然后根据模型的轮廓特征找到足部的关键点，再利用关键点的空间坐标信息完成测量。三维重建的测量方法虽然精度高，但由于设备的限制，该方法不够轻量，设备仪器比较笨重，需要到具有设备的场景下完成测量。

1.2 基于图像的测量

图像处理技术在 20 世纪 80 年代就已经有学者开始研究，如今在许多领域取得了良好的效果。对于一些测量任务而言，其实并不需要重建出物体的三维形态，部分研究者基于以上考虑，利用图像处理技术对物体进行测量。

刘长青等^[6]通过相机拍摄玉米果穗的图片, 利用图像处理技术对玉米果穗长度、果穗宽度、穗行数等参数进行测量。Momin 等^[7]拍摄大豆、豆荚等收获物, 利用图像处理技术检测出大豆、豆荚, 并对大豆、豆荚进行质量评估。柳革命等^[8]采集弹壳的正、侧面图像, 利用边缘检测和 Hough 变换直线等技术设计了一个测量弹壳的尺寸系统。Hao 等^[9]通过图像处理技术测量鱼类的尺寸。Kang 等^[10]开发了一种织物自动评价系统, 通过 CCD 相机拍摄织物, 对织物的结构参数进行自动测量。

图像测量技术相较于三维测量比较轻量, 只需要拍摄被测物体的图片就可以进行测量, 测量过程方便、快捷。但该方法在足部测量方面的研究并不是很多。

1.3 机器学习

近年来, 机器学习发展飞速, 并在人脸关键点检测中获得了不错的应用。Dollar 等^[11]提出了一种级联姿势回归(CPR, Cascaded Pose Regression)的方法检测人脸关键点。Sun 等^[12]将卷积神经网络用于人脸关键点的定位, 该网络能对人脸的 5 个关键点进行定位。Zhang 等^[13]提出了一种级联神经网络处理多任务, 该网络可以同时处理人脸检测和人脸关键点定位问题。这些定位人脸关键点的方法给本文检测足部关键点提供了思路。

1.4 足部测量参数的定义

在足部测量中, 首先需要确定测量的参数。不同的研究者会根据问题去定义不同的测量参数, 并没有固定的测量参数。Liu 等^[14]将脚后跟中点和第一跖跗关节突点的连线作为足部轴线, 并定义了 26 个足部关键点用于测量足部参数。Zhao 等^[15]基于鞋类产品个人定制, 测量了三维足部模型的 6 个围长。朱晓兰等^[16]定义了 12 个足部关键点和 18 个测量参数用于田径运动员的足型。因此本文需要先对足部的生理结构和受力情况进行分析, 然后得到足部的测量参数以及相关关键点。

目前主要的足部测量方法是手工测量和非接触测量。非接触测量中的三维测量由于设备和方法的限制, 没办法做到便捷、快速的测量。对于某些测量任务而言, 基于图像的测量方法更加轻量, 但目前在足部测量方面的研究比较少。机器学习在许多领域中取得了很好的效果, 但与图像测量相结合的研究比较少。本文基于以上分析, 提出了一种基于机器学习的图像足部测量方法。

2 本文方法思路

本文结合图像和深度网络, 提出了一种基于图像的足部测量方法, 利用足部关键点的信息完成足部测量。本文研究方法如图 1 所示。

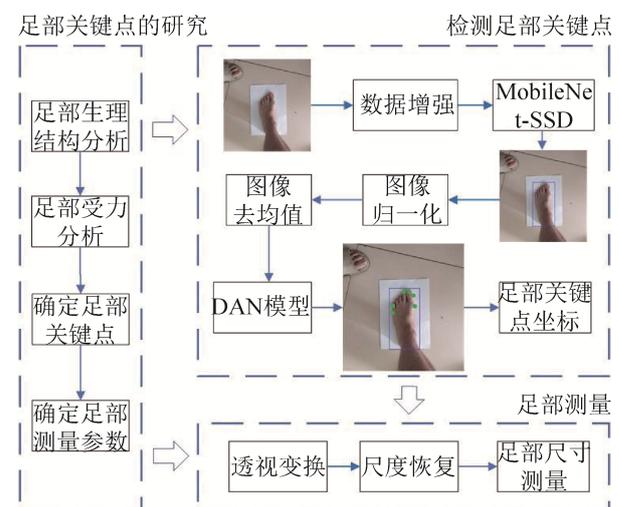


图 1 本文方法研究图

Fig. 1 Method research diagram

该方法不需要对足部进行重建, 数据获取方便, 能够便捷、快速且准确的测量足部参数。本文方法可以分为以下 3 个部分: 足部关键点的研究、检测足部关键点以及足部测量。具体如下:

(1) 足部关键点的研究: 通过分析足部生理结构以及其受力模型, 确定足部的关键部位。然后设计一套基于足部关键点表示的测量参数。

(2) 检测足部关键点: 本文基于手持相机定义了一种数据采集方式, 用于模型的训练与测试。利用 MobileNet-SSD 网络对足部图片进行足部区域检测。然后围绕足部关键点检测这一问题, 对 DAN

(Deep Alignment Network)模型的不足进行优化,修改了 DAN 模型的激活函数和损失函数,并通过优化后的 DAN 对足部关键点进行检测。

(3) 足部测量:通过透视变换对图片进行校正,并利用参照物恢复尺度,再通过得到的足部关键点位置求出足部尺寸。

3 足部测量参数分析

3.1 足部生理结构分析

为了设计合理的足部测量参数,首先需要了解足部的生理结构特点以及在站立、行走等状态下的受力情况。

足部是具有复杂结构的人体器官,包含 26 块骨骼,33 个关节以及大量肌肉与韧带,这些部分的紧密联系和相互配合才使得足部能够完成行走、站立等复杂功能。以下为足部结构介绍:

(1) 足部骨骼:足部骨骼主要分为 3 个部分,跗骨、跖骨以及趾骨,如图 2 左图所示。跗骨位于足部后部,由 7 块骨骼组成;跖骨处于脚的中部,包含 5 块骨头,由内向外分别为第一、二、三、四、五跖骨;趾骨位于脚的前端,除拇趾只有 2 节外,其余 4 趾分为近节、中节和远节趾骨。

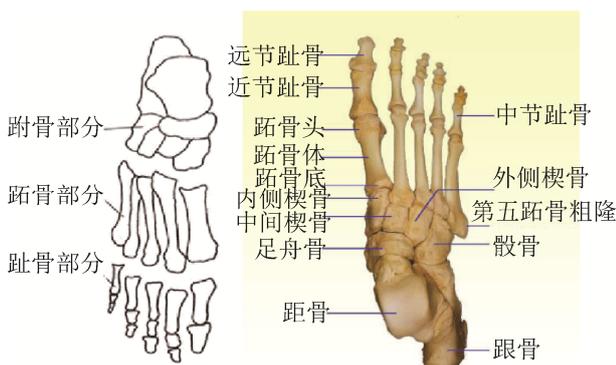


图 2 足部骨骼结构

Fig. 2 Structure of foot bone

(2) 足部关节:关节是骨骼之间连接的部分,足部的主要关节包括踝关节、跗骨间关节、跗跖关节、跖骨间关节、跖趾关节和趾关节。足部关节在人类运动过程中起着重要的作用,它能

够为骨骼提供活动度,让人类完成各种动作。

(3) 足部肌肉:足部上的肌肉属于骨骼肌,主要的肌肉为足背肌和足底肌,是用于保持人体站立、走、跑、跳的肌肉。

(4) 足弓:足弓包含两个纵弓和两个横弓,整体呈拱型,能够吸收足部在行走、跑跳等时的冲击力,保护人体关节和内脏不受损伤。

研究表明,当人体静止站立时,体重的 3/6 由脚跟支撑点承担,其余的 2/6 和 1/6 由第一跖骨和第五跖骨承担^[17]。这表明在静止时足部的主要支撑点为跖骨部分和跟骨部分。当人处于动态运动时,支撑点会在足跟、足弓以及跖、趾骨之间移动,但一个步态周期中前足部分的承重比例会高于后足部分^[17]。这是因为在行走或者跑跳过程中,为了完成动作前足需要弯曲蹬地发力,这涉及到跖骨、趾骨、跖趾关节以及趾骨关节的协同作用。

3.2 关键特征点提取

基于 3.1 节的分析,可得出跖趾关节和趾骨关节是对于人体运动很重要的部位,本文将跖趾关节和趾骨关节的特征位置作为足部的关键特征点,其中包括拇趾关节凸点、小趾关节凸点、第一跖骨关节凸点以及第五跖骨关节凸点。同时足长也是描述足部的重要特征,在购鞋、医学诊断、犯罪医学等领域都需要用到足长这一特征。为了确定足长特征,我们需要定义其相关的足部特征点。一般地,足长的测量涉及到脚后跟以及脚后跟在足长方向上的最远点,这一般为拇趾端点或者第二脚趾端点。本文定义的足部关键点如图 3 所示。

3.3 测量参数

根据 3.2 节的足部关键点,可以定义需要测量的足部参数,分别为拇指长、中轴长、足宽、拇趾里宽、小趾外宽、第一跖趾宽度以及第五跖趾宽度。定义的测量参数能够比较完整地刻画了足部形态。其中足长和足宽是足部的重要特征,能够反映足部的基本情况;其余的参数是足部在站立行走时重要部位的相关参数。



图 3 足部关键特征点
Fig. 3 Key points of foot

为了进行测量, 需要建立坐标系, 坐标系建立如下: 过 O 点做足后跟的切线 L, 过 B 点做直线 L 的垂线, 交点为 B', 以 BB' 作为 Y 轴, 在 BB' 直线上任取一点 O' 做 BB' 的垂线作为 X 轴, 为了更好的描述测量参数, 分别过 C、D、E、F 作 BB' 的垂线, 交点分别为 C'、D'、E' 和 F', 则上述的足部参数表达可以用图 4 表示。

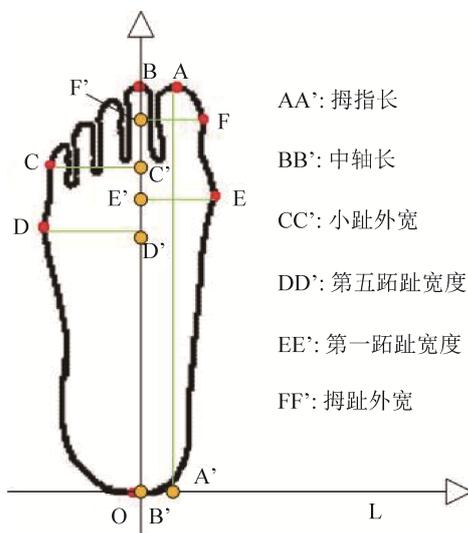


图 4 测量尺寸示意图
Fig. 4 Schematic diagram of foot size to be measured

基于建立的坐标系, 足部各测量尺寸可以用坐标系中的点进行描述, 如表 1 所示。表中的符号代表关键点的坐标, 例如 A_y 代表 A 点的纵坐标, E_x 代表 E 点的横坐标。

表 1 足部尺寸的数学表达式
Tab. 1 Mathematical expression of foot size

足部尺寸	数学表达式
拇指长	$A_y - O_y$
中轴长	$B_y - O_y$
足宽	$E_x - D_x$
拇趾里宽	$F_x - B_x$
小趾外宽	$B_x - C_x$
第一跖趾宽度	$E_x - B_x$
第五跖趾宽度	$B_x - D_x$

4 足部关键点的检测

4.1 足部区域检测

关键点的检测需要先通过目标检测技术检测出相关区域, 再实现关键点的检测。因此在构建数据集之前, 需要对采集的图像进行足部区域的提取。

MobileNet^[18]是一个基于深度可分离卷积的轻量级网络, 它与迁移学习^[19], dropout^[20]等优化方法不同, 其优化的目标是卷积操作的计算量。当卷积核大小为 3, 深度可分离卷积相较于标准卷积少了 8~9 倍的计算量, 然而只有极小的准确率的下。基于 MobileNet 的优点, 有研究者将 MobileNet 和 SSD 相结合组成 MobileNet-SSD, 用于目标检测。本文用 MobileNet-SSD 对足部区域进行检测。MobileNet-SSD 的网络结构图 5 所示。该网络在 MobileNet 后面添加了 8 个卷积层, 并抽离其中几层的结果进行检测, 最后通过 NMS (Non-maximum Suppression) 挑选最合适的区域。

4.2 数据集的构造与预处理

在采集实验数据时我们发现, 由于足部是一个复杂的几何体, 很难找到一个合适的角度拍摄到所有关键特征点。为了拍摄到脚后跟的位置, 被测者需要手持手机从足部两侧进行拍摄, 经测试, 从足部的两侧拍摄会导致另一侧关键点的缺失, 并且被测者为了能够拍摄到脚后跟, 足部可能会处于不自然的站立状态。从足部的上方进行拍摄能够比较容易的观测到尽可能多的足部关键点, 并且足部处于自然的站立状态, 但是这也意味着脚后跟的缺失。

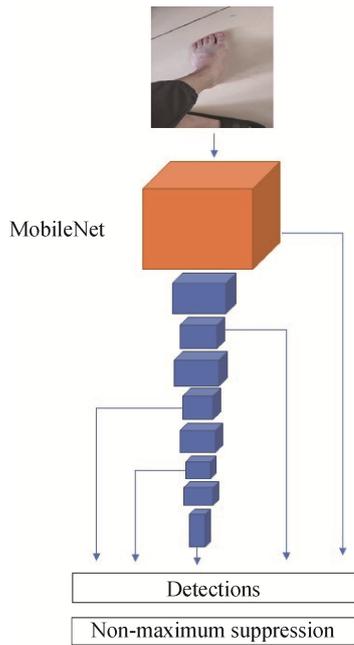


图 5 MobileNet-SSD 网络结构图
Fig. 5 MobileNet-SSD network structure

基于以上分析,为了保证拍摄的图片能够尽可能观测到所有的足部关键点,我们定义以下的图片采取方式:在随机场景下,被测者脚踩在 A4 纸上,并将脚后跟踩在 A4 纸窄边边缘,使得脚与 A4 纸窄边尽量保持垂直。然后被测者手持手机从上方拍摄足部图片,拍摄角度不大于 30° 。该做法能够尽可能从一张图像上观测到尽可能多的足部关键点,并且通过该种方式,缺失的脚后跟位置可以通过 A4 纸的边缘获取。利用该种方式,我们找到了 20 位志愿者,并在随机场景下共采集了 196 张有效的足部图片,部分数据如图 6 所示。



图 6 部分数据集展示
Fig. 6 Partial data set display

具体预处理操作:将所有右脚图片进行垂直翻转,所有图片统一成左脚姿态;将图片分辨率改成网络输入所需分辨率,本文所用网络的输入

112×112 ,因此将图片下采样为 112×112 。为了避免下采样过程中图片长宽比失真,先将原图填充为长宽比为 1:1 的图片,再进行下采样,如图 7 所示;对下采样后的图片进行归一化和去均值;对处理后的图片进行数据标注,标注出图片中的足部关键点与数据增强,通过对图片进行随机平移、缩放和旋转,最后生成 8 800 张数据集。



图 7 下采样后的足部图片
Fig. 7 Foot picture after downsampling

4.3 改进的 DAN

DAN (Deep Alignment Network)是 Kowalski 等^[21]提出用于人脸关键点定位的深度网络。该网络与其他人脸关键点定位的网络相比加入了关键点热点图的信息,这种改进使得该方法能够利用面部的整个图像而不是局部图像,并避免陷入局部最小值。本文基于改进的 DAN 模型对足部关键点进行检测。

DAN 的网络结构图 8 所示。DAN 具有多个阶段,除阶段 1 外其余阶段的结构都是相同的。DAN 的每个阶段包含一个前馈神经网络以及一个转换层。前馈神经网络的输出为 68 个人脸关键点,但本文需要检测的足部关键点为 6 个,为了使其能用于足部关键点的定位,仅将其前馈卷积神经网络的 Fc2 层的输出进行修改,修改后的网络结构如表 2 所示。除了表 2 中展示的卷积层和池化层外,网络中每一层都使用了批量归一化(Batch Normalization),并且在第一个全连接层(Fc1)前添加了 dropout 层,网络的最终输出为 ΔS ,用于更新本阶段的关键点位置。

转换层的结构如图 9 所示,其作用是生成下一阶段的输入,输出为矫正图、热点图以及特征图。令 $x_i \in R^2$ 是第 i 个标记点在图片 I 上的坐标,

图片 I 上 p 个标记点可以表示为向量 $\mathbf{S} = (x_1^T, x_2^T, \dots, x_p^T)^T \in R^{2p}$, 则 \mathbf{S} 为该图片的关键特征点的集合, 也称为形状。

在 DAN 中, 首先确定一个 \mathbf{S}_0 作为规范形状, \mathbf{S}_0 通常在训练集中随机挑选, \mathbf{S}_0 选取完后不再改变。转换层首先通过普氏分析法 (Procrustes Analysis) 计算检测的关键点 \mathbf{S}_1 与规范 \mathbf{S}_0 的变换 T_{t+1} 。变换参数 T_{t+1} 包含了 \mathbf{S}_1 与 \mathbf{S}_0 之间的旋转、平移信息, 可以通过参数 T_{t+1} 对图片进行变换, 生成矫正图 $T_{t+1}(I)$ 。然后转换层将变换后的关键点

$T_{t+1}(\mathbf{S}_1)$ 输入到热点图生成模块, 生成热点图 H_{t+1} 。热点图是包含关键点信息的图片, 其特点是关键点位置的像素强度最高, 越远离关键点像素值越低。特征图生成模块将上一阶段的 Fc1 层输入到一个具有 3 136 单元、激活函数为 ReLu (Rectified Linear Unit Rectified Linear Unit) 的全连接层中, 然后把全连接层的输出变成一个 56×56 的图片, 最后再将该图片上采样成 112×112 的特征图 F_{t+1} , 由于 F_{t+1} 是通过上一阶段 Fc1 层生成的, 该特征图能够为下一阶段的网络提供上一阶段的信息。

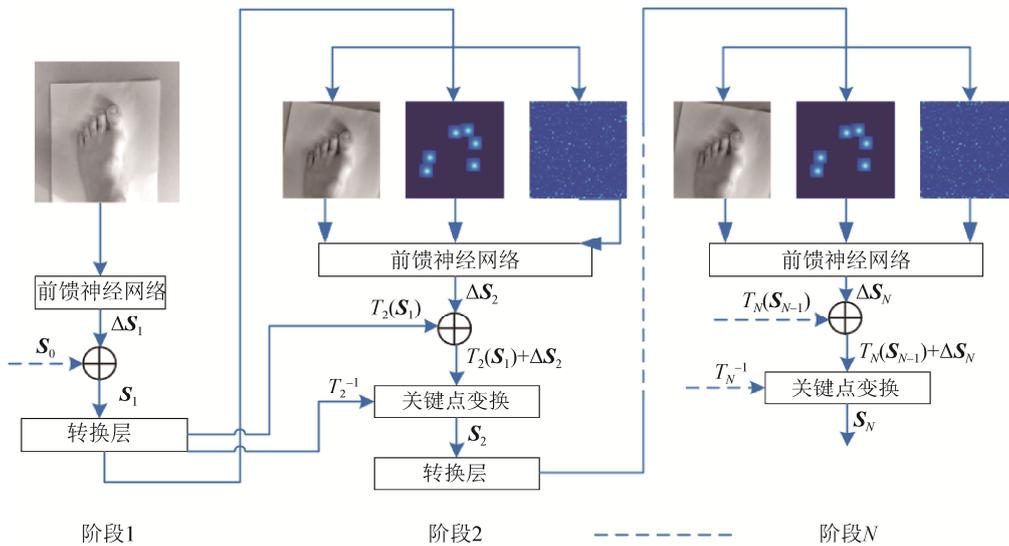


图 8 DAN 的网络结构图
Fig. 8 Network structure of DAN

表 2 DAN 的前馈卷积神经网络

Tab. 2 Feedforward convolutional neural network of DAN

网络层名字	输入	输出	卷积核, 步长
Conv1a	112×112×1	112×112×64	3×3×1,1
Conv1b	112×112×64	112×112×64	3×3×64,1
Pool1	112×112×64	56×56×64	2×2×1,2
Conv2a	56×56×64	56×56×128	3×3×64,1
Conv2b	56×56×128	56×56×128	3×3×128,1
Pool2	56×56×128	28×28×128	2×2×1,2
Conv3a	28×28×128	28×28×256	3×3×128,1
Conv3b	28×28×256	28×28×256	3×3×256,1
Pool3	28×28×256	14×14×256	2×2×1,2
Conv4a	14×14×256	14×14×512	3×3×256,1
Conv4b	14×14×512	14×14×512	3×3×512,1
Pool4	14×14×512	7×7×512	2×2×1,2
Fc1	7×7×512	1×1×256	—
Fc2	1×1×256	1×1×12	—

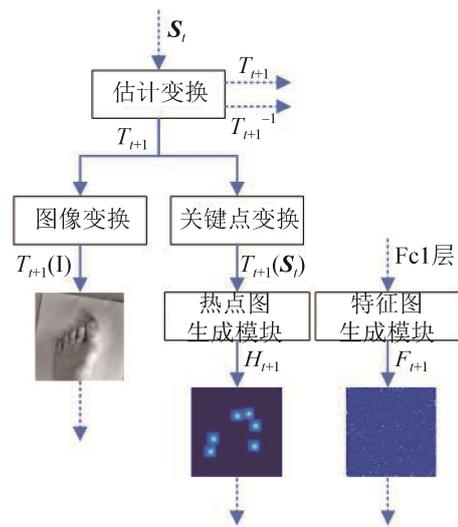


图 9 转换层结构图
Fig. 9 Connection layer structure

转换层的输入输出如图 10 所示, 图中从左到右分为网络输入图 I 、矫正图 $T_{t+1}(I)$ 、热点图 H_{t+1} 以及特征图 F_{t+1} 。

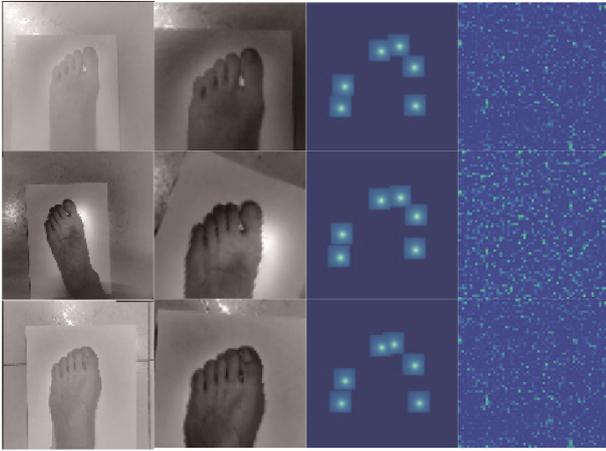


图 10 转换层结果展示图
Fig. 10 Connection layer results display

t 阶段 DAN 网络的输出如式(1)所示:

$$\mathbf{S}_t = T_t^{-1}(T_t(\mathbf{S}_{t-1}) + \Delta \mathbf{S}_t) \quad (1)$$

式中: \mathbf{S}_{t-1} 为 $t-1$ 阶段的形状; $T_t(\mathbf{S}_{t-1})$ 为 \mathbf{S}_{t-1} 通过变换参数 T_t 变幻后的形状; $\Delta \mathbf{S}_t$ 是前馈神经网络的输出。

本文的 DAN 网络采用两阶段的结构, 在本文实验过程中还尝试过更多阶段的结构, 但与两阶段的网络结果差别不大, 因此最终选择两阶段的结构。

原文网络中所用的激活函数为 ReLU, 在训练的过程中发现有部分的神经元不更新权值, 出现“死亡”现象。这是因为 ReLU 在遇到异常大的梯度时会产生很大的权值更新, 其余样本通过该神经元后输出变成负数, 而 ReLU 会将负数更新为 0, 导致该神经元不再更新。为了避免该问题, 本文采用如下激活函数 Leaky ReLU。

网络中损失函数为 L2, 其函数体如式(2)所示。

$$L2(x) = \frac{1}{2}x^2 \quad (2)$$

其导数为

$$dL2(x)/dx = x \quad (3)$$

从 L2 损失函数的导数可以发现, 当 x 的值很大时, 其导数值也很大, 整个损失函数将被这个

点主导, 网络向着大误差的方向更新, 导致网络不容易收敛, 这种情况通常发生在训练初期。

Wing 损失函数是 Feng 等^[22]针对人脸关键点定位研究的损失函数, 该损失函数是一个分段函数, 能够分别处理大误差和小误差时的网络走向, 解决 L2 损失函数对离群点敏感的缺点。该损失函数的函数体如式(4)所示。

$$wing(x) = \begin{cases} \omega \ln(1 + |x| / \varepsilon), & |x| < \omega \\ x - |C|, & otherwise \end{cases} \quad (4)$$

式中: ω 为一个非负数, 将非线性部分约束在 $(-\omega, \omega)$ 区间; ε 为限制曲线的弯曲程度; C 为一个常量, 用于连接损失函数的线性和非线性部分, 其表达式为 $C = \omega - \omega \ln(1 + \omega / \varepsilon)$ 。从函数中可以看出, 当 x 较大时, wing 损失函数产生的函数值不会太大, 网络的训练走向不会被异常点主导; 当 x 较小时, 由于 $\ln x$ 的性质, 则会产生更大的梯度, 网络会被小误差主导, 使得网络更快收敛。本文将损失函数更改为 wing loss, 其中 ω 为 10, ε 为 2。

4.4 足部测量

本文基于透视变换(Perspective Transformation)的原理进行测量。透视变换可以让拍摄的图片投影到一个与地面平行的视平面, 从而完成对图片的校正。通用的透视变换公式如下:

$$[x', y', w'] = [u, v, w] \mathbf{P} \quad (5)$$

式中: $[x', y', w']$ 为变换后的齐次坐标; $[u, v, w]$ 为变换前的齐次坐标; \mathbf{P} 为透视变换矩阵, 大小为 3×3 。通过 4.2 小节中采集的图片可以利用 A4 纸 4 个角的坐标完成透视变换, 透视变换后的图片可以利用足部关键点的位置求解得到足部测量参数在图像中的尺寸。最后通过 A4 纸恢复尺度, 将足部尺寸在图像中的尺度还原到现实世界的尺度。

假设 A4 纸真实的长宽为 L 和 W , 在图像中长宽的像素长度为 L_1 和 W_1 , 并以 A4 纸左下角的点建立 XOY 坐标系。设图中某点的坐标为 (x', y') , 则该点在现实 A4 纸中的坐标 (x, y) 为:

$$x = x' \cdot \frac{L}{L_1} \quad (6)$$

$$y = (L_1 - y') \cdot \frac{L}{L_1} \quad (7)$$

5 实验结果

基于上述方法, 本文进行了基于图片的足部测量实验。本文实验所用计算机的操作系统为 windows10, CPU 为 i7-7700HQ@2.80GHz, 内存为 16G, 显卡为 1050ti, 所用的 IDE 为 PyCharm, 代码框架为 TensorFlow。

5.1 足部关键点实验结果

在实验之前, 先利用手机在随机场景下共拍摄 303 张带有足部的图片。这部分的图片用于训练 MobileNet-SSD 网络, 在训练之前需要将图片的分辨率改成网络输入所需要的分辨率。MobileNet-SSD 网络的输入为 300×300 。下采样后的图片进行标注后, 并将数据集以 7:2:1 的比例划分为训练集, 验证集、测试集进行训练。图 11 为部分数据展示, 从上往下分别为训练集, 验证集和测试集。足部区域模型的检测结果如图 12 所示。

本文方法与 Kazemi 等^[23]的级联回归树方法(以下简称 ERT)以及原始 DAN 网络进行对比。ERT 算法的中心思想是集成学习, 是使用一系列学习器进行学习, 并使用某种规则把各个学习结

果进行整合从而获得比单个学习器更好的学习效果的一种机器学习方法。



图 11 部分实验数据展示

Fig. 11 Part of the experimental data display

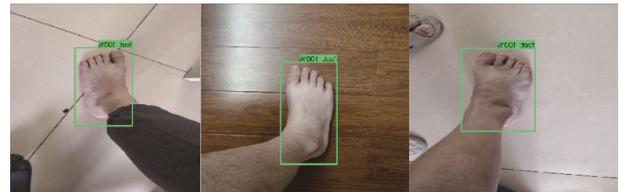


图 12 足部区域检测结果

Fig. 12 Results of the foot area detection

图 13 是部分结果展示, 从上往下分别为 ERT、DAN 与改进后 DAN 方法, 图中绿点是算法检测的关键点。

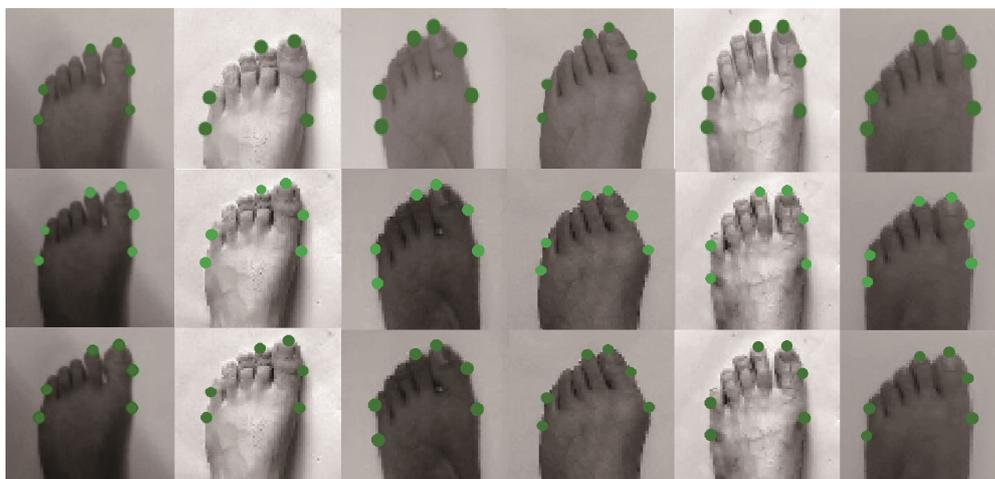


图 13 足部关键点检测部分实验结果展示

Fig. 13 Part of the experimental results of the key point detection

为了比较 ERT 算法和 DAN 两种方法的结果, 本文选用归一化均值误差 MNE (the mean normalized error) 标准评价 3 种方法。MNE 的计算公式如下:

$$\text{MNE} = \frac{\sum_i^n (x - \bar{x})}{n \cdot d_{FD}} \quad (8)$$

式中: i 为关键点序号; n 为关键点个数; x 为真实坐标; \bar{x} 为检测的坐标; d_{FD} 为足宽。MNE 能够衡量检测的坐标和真实坐标的差距, MNE 的值越小代表该模型检测的坐标越接近真实坐标, 也意味着模型的性能更好。本文将 4.2 节得到 8 800 张图片分为训练集和测试集, 其中训练集为 6 850 张图片, 测试集为 1 950 张图片, 3 种方法的结果如表 3 所示。

表 3 3 种方法的 MNE 比较
Tab. 3 MNE comparison of three methods

方法名	MNE
ERT	0.026 6
DAN	0.020 2
改进后的 DAN	0.018 0

从表 3 可以看出, 在测试集中本文方法的 MNE 值小于 ERT 和 DAN, 这意味着本文方法的性能优于 ERT 和 DAN, 本文模型能够更好的检测出足部关键点。

表 4 手工测量数据概况

Tab. 4 Overview of manual measurement data

足部尺寸	拇指长	中轴长'	足宽	拇趾里宽	小趾外宽	第一跖趾宽度	第五跖趾宽度
最大值	266	260	104	48	55	50	56
最小值	257	246	98	42.5	51	44	53
平均值	261.3	254	100.6	45.2	52.3	46	54.6

表 5 部分手工测量数据的数值

Tab. 5 Part of manual measurement data values

足部尺寸	拇指长	中轴长'	足宽	拇趾里宽	小趾外宽	第一跖趾宽度	第五跖趾宽度
1	265	260	98	42.5	53	44	54
2	257	247	103	48	51	50	53
3	266	259	100	44	55	44	56
4	257	246	104	47	52	49	55
5	260	255	100	46	51	44	56
6	265	257	101	47	55	46	55

5.2 足部尺寸测量实验结果

为了比较算法的优劣性, 本文选取了 20 名志愿者进行测试。首先手工测量 20 名志愿者的足部尺寸以作参考, 然后志愿者按照 4.2 中描述的数据采集方式对每只脚拍摄 5 张图片形成测试数据集, 最后用本文所述的算法求出足部尺寸并与手工测量的足部尺寸进行对比。表 4 列出了手工测量数据的概况, 其中包括足部各尺寸的最大值, 最小值以及平均值, 单位为毫米(mm)。表 5 列出了部分手工测量数据的数值, 单位为毫米(mm)。

表 6 给出了 ERT、DAN 与改进后的 DAN 方法得到的足部尺寸与手工测量尺寸的误差, 单位为毫米(mm)。通过分析, 造成测量误差的主要原因是检测的关键点与真实坐标的误差。

从表 6 中可以看出, 虽然 ERT 在测试集中的 MNE 只有 0.026, 但基于 ERT 方法得到的结果整体误差比较大, 主要体现在对关键点的检测不准确, 这也意味着 ERT 方法的鲁棒性不够好。改进后的 DAN 在各关键点上的检测误差都优于 ERT, 且大部分关键点的检测误差优于 DAN, 这表示本文改进后的 DAN 方法在足部测量上具有更好的性能, 且具有鲁棒性。

表 6 实验结果与误差
Tab. 6 Experimental results and errors

方法	误差	拇指长	中轴长	足宽	拇趾里宽	小趾外宽	第一跖趾宽度	第五跖趾宽度	平均值
ERT	最大误差	10	5	10	10	9	13	14	10.1
	最小误差	0	0	0	0	0	0	1	0.14
	平均误差	2.9	2.8	3.1	4.3	4.9	5.7	2.5	3.7
DAN	最大误差	9.01	10.03	7.77	3.38	3.39	7.76	7.03	7
	最小误差	0.26	0.3	0.08	0.1	0.21	0.08	0.35	0.23
	平均误差	3	2.7	3.9	1.5	1.6	3.3	3	2.7
改进后的 DAN	最大误差	9.14	7.18	5.74	3.63	2.46	6.16	7.15	5.96
	最小误差	0.24	0.15	0.04	0.12	0.02	0.15	0.26	0.14
	平均误差	2.6	2.4	1.9	1	3.2	3	2.3	2.3

6 结论

本文提出了一种基于 DAN 模型的足部测量方法。通过实验发现, 该方法与传统手工测量相比, 能够节省劳动力, 提高测量效率; 与三维测量相比能够不受设备的限制, 做到方便、快捷地测量, 并且具有良好的精度。

本文采集的数据主要为青年人群, 实验的结果表明本文方法适用于青年人群, 可以在青年人群中推广, 但本文方法仍存在一定的不足, 以下是问题的总结和改进的方向: (1) 在检测足部关键点时, 脚后跟的位置是通过直线约束来间接获取的, 增加了被测者的操作复杂度。足部经过漫长的进化, 各部分之间应该存在某种联系, 如何利用足部的相关联系间接得到脚后跟的位置还需要进一步的研究; (2) 本文方法在采集的数据集上有良好的表现, 但数据集中并没有老人、小孩、白种人、黑种人等不同年龄段、不同人种的数据。为了证明该方法具有良好的鲁棒性, 需要进一步搜集各类数据, 进行大量实验。

参考文献:

- [1] Pheasant S, Haslegrave C M. Bodyspace: Anthropometry, ergonomics and the design of work[M]. Boca Raton, FL, USA: CRC Press, 2005.
- [2] Song S C, Yu S P, Xu W J. Study on 3 D body scanning, reconstruction and measurement techniques based on Kinect[J]. Tianjin Gongye Daxue Xuebao (S1671-024X), 2012, 31(5): 34-37.
- [3] Zhong K, Li Z, Zhou X, et al. Enhanced phase measurement profilometry for industrial 3D inspection automation[J]. The International Journal of Advanced Manufacturing Technology (S1433-3015), 2015, 76(9-12): 1563-1574.
- [4] 王俊青, 周旭, 姚志明, 等. 基于计算机视觉的脚型参数测量系统的设计与实现[J]. 仪表技术, 2012, (7): 36-40.
Wang Junqing, Zhou Xu, Yao Zhiming, et al. Design and Implementation of the Foot Parameter Measurement System Based on Computer Vision[J]. Instrumentation Technology, 2012, (7): 36-40
- [5] 毋戈. 基于多个深度相机的足部三维重建及形态分析[D]. 上海: 东华大学, 2017.
Wu Ge. Three-dimensional Reconstruction and Shape Analysis of Human Feet Based on Multiple Depth Cameras[D]. Shanghai: Donghua University, 2017.
- [6] 刘长青, 陈兵旗. 基于机器视觉的玉米果穗参数的图像测量方法[J]. 农业工程学报, 2014, 30(6): 131-138.
Liu Changqing, Chen Bingqi. Method of image detection for ear of corn based on computer vision[J]. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(6): 131-138.
- [7] Momin M A, Yamamoto K, Miyamoto M, et al. Machine vision based soybean quality evaluation[J]. Computers and Electronics in Agriculture (S0168-1699), 2017, 140: 452-460.
- [8] 柳革命, 刘鹏. 基于机器视觉的弹壳多尺寸测量[J]. 工业控制计算机, 2016, 29(4): 60-61.
Liu Geming, Liu Peng. Sizes Measurement of Cartridge Case Based on Machine Vision[J]. Industrial Control Computer, 2016, 29(4): 60-61.
- [9] Hao M, Yu H, Li D. The measurement of fish size by machine vision-a review[C]// International Conference on

- Computer and Computing Technologies in Agriculture. Berlin, German: Springer, Cham, 2015: 15-32.
- [10] Kang T J, Choi S H, Kim S M, et al. Automatic structure analysis and objective evaluation of woven fabric using image analysis[J]. *Textile Research Journal* (S0040-5175), 2001, 71(3): 261-270.
- [11] Dollár P, Welinder P, Perona P. Cascaded pose regression[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2010: 1078-1085.
- [12] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 3476-3483.
- [13] Zhang K, Zhang Z, Li Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. *IEEE Signal Processing Letters* (S1070-9908), 2016, 23(10): 1499-1503.
- [14] Liu W. Accuracy and reliability of a technique for quantifying foot shape, dimensions and structural characteristics[J]. *Ergonomics* (S0014-0139), 1999, 42(2): 346-358.
- [15] Zhao J, Xiong S, Bu Y, et al. Computerized girth determination for custom footwear manufacture[J]. *Computers & Industrial Engineering* (S0360-8352), 2008, 54(3): 359-373.
- [16] 朱晓兰, 吴壮志. 三维足型参数自动测量系统设计与实现[J]. *体育科学*, 2010, 30(9): 44-48, 81.
Zhu Xiaolan, Wu Zhuangzhi. Design and Implementation of 3D Foot Size Measurement System[J]. *China Sport Science*, 2010, 30(9): 40-48, 81.
- [17] 王明鑫, 俞光荣, 陈雁西, 等. 正常中国成年人足底压力分析[J]. *中国矫形外科杂志*, 2008(9): 687-690.
Wang Mingxin, Yu Guangrong, Chen Yanxi, et al. Analysis of the plantar pressure distribution of the normal Chinese adult[J]. *Orthopedic Journal of China*, 2008(9): 687-690.
- [18] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[DB/OL]. [2019-10-05] <https://arxiv.org/abs/1704.04861>.
- [19] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks?[C]// *Advances in neural information processing systems*. New York: Curran Associates, 2014: 3320-3328.
- [20] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *The journal of machine learning research* (S1532-4435), 2014, 15(1): 1929-1958.
- [21] Kowalski M, Naruniec J, Trzcinski T. Deep alignment network: A convolutional neural network for robust face alignment[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ: IEEE, 2017: 88-97.
- [22] Feng Z H, Kittler J, Awais M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 2235-2245.
- [23] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 1867-1874.