一种基于关键点的红外图像人体摔倒检测方法

徐世文，王 姮，张 华，庞 杰
（西南科技大学 信息工程学院，四川 绵阳 621000）

摘要：针对已有人体摔倒检测方法在复杂环境场景下易受光照影响、适应性差、误检率高等问题，提出了一种基于关键点估计的红外图像人体摔倒检测方法。该方法采用红外图像，有效避免了光照等因素的影响，经过神经网络找到人体目标中心点，然后回归人体目标属性，如目标尺寸、标签等，从而得到检测结果。使用红外相机采集不同情况下的人体摔倒图像，建立红外图像人体摔倒数据集并使用提出的方法进行检测，识别率达到97%以上。实验结果表明提出的方法在红外图像人体摔倒检测中具有较高的精度与速度。

关键词：红外图像；关键点估计；摔倒检测；神经网络
中图分类号：TP391.4 **文献标识码**：A **文章编号**：1001-8891(2021)10-1003-05

Human Fall Detection Method Based on Key Points in Infrared Images

XU Shiwen，WANG Heng，ZHANG Hua，PANG Jie
(*School of Information Engineering, Southwest University of Science and Technology, Mianyang* 621000, *China*)

Abstract：To address the problems with existing human fall detection methods for complex environments, which are susceptible to light, poor adaptability, and high false detection rates, an infrared image human fall detection method based on key point estimation is proposed. This method uses infrared images, which effectively eliminates the influence of factors such as lighting; first, the center point of the human target is found through a neural network, and second, the human target attributes, such as the target size and label, are regressed to obtain detection results. An infrared camera was used to collect human body fall images in different situations and establish datasets containing infrared images of human falls. The proposed method was used for experiments; the recognition rate exceeded 97%. The experimental results show that the proposed method has a higher accuracy and speed than other two methods in infrared image human fall detection.

Key words：infrared image, key point estimation, fall detection, neural network

0 引言

随着医疗保障的提高，人口老龄化已是当今社会的一个问题。根据相关调查表明，老年人受到意外伤害的主要原因之一就是摔倒。因此，防止老年人摔倒也变得越来越重要。世界卫生组织报告说，每年因跌倒造成的严重伤害超过3730万人次，死亡64.6万人[1]。摔倒是一个重要的公共健康问题，其伤害很大程度上取决于救助响应时间的长短。智能的摔倒检测系统可以全天候工作，及时做出反应，实时保护人们的安全。

现在主流的人体摔倒检测方法根据检测传感器的不同大致分为基于穿戴式的摔倒检测、基于环境式的摔倒检测以及基于计算机视觉的摔倒检测3类。基于穿戴式检测法通常将加速度计以及陀螺仪等传感器佩戴在身体上，收集运动数据[2-4]，使用采集得到的传感器数据训练MLP[5]（multilayer perceptron）、SVM[6]（support vector machines）等机器学习算法进行人体摔倒检测。基于外部传感器的摔倒检测方法需要随身穿戴传感器，存在用户穿戴起来不方便和不自在，容易脱落等问题。基于环境式的摔倒检测是提前在指定的区域布置好诸如压力传感器、声音传感器等，通过传感器采集到的数据进行检测，这种方法存在容易被环境噪声影响[7]，成本高等问题。

收稿日期：2020-02-18；**修订日期**：2020-02-21.
作者简介：徐世文（1994-），男，四川省成都市人，硕士研究生，主要研究方向为计算机视觉与图像处理、深度学习。E-mail：1411761943@qq.com。
通信作者：王姮（1971-），女，硕士，教授，主要研究方向为机器人技术及应用、自动化技术研究。E-mail：wh839@qq.com。

另外，基于计算机视觉的摔倒检测法通常对摄像头拍摄到的图像进行目标的提取，获取其特征，再通过对特征的分析从而得到摔倒检测结果。文献[8]将行人用矩形框表示，通过矩形框的长宽比例来说明行人的姿态，从而进行摔倒检测；文献[9]将目标的轮廓拟合成椭圆，提取其几何与运动两种特征，组成一个新的特征，使用 SVM 进行摔倒判断；文献[10]使用高斯混合模型得到人体目标，提取多帧特征并融合得到基于时间序列的运动特征，使用一个简单的卷积神经网络判断摔倒。文献[11]中提出一种基于人体骨骼关键点和神经网络的人体摔倒检测方法，通过 Alphapose 检测人体骨骼关键点，并用来训练 LSTM（long short term memory）神经网络，实现人体摔倒的检测。

上述人体摔倒检测研究中使用的视频与图像均是可见光图像，然而在生活中应用人体摔倒检测的往往是老人和容易出现情况的病人，这些地方一般都是需要 24 h 监控。可见光图像在夜晚和光照条件不好的场景中不能很好地呈现出图像，在这些情况下不能做到准确地检测摔倒情况。红外图像目标识别技术是指通过对红外图像进行预处理，然后提取目标特征，最后实现目标的定位与识别[12]。与可见光图像相比，红外图像直观反映的是物体的温度，一般而言红外图像中行人的亮度比背景亮度要高，且纹理、颜色和光照对红外图像几乎没什么影响，这使得红外图像在进行人体检测方面具有很大的优势和潜力。

针对上述问题，本文提出了一种基于关键点估计的红外图像人体摔倒检测方法。该方法采用红外相机采集图像，图像经过全卷积网络得到人体目标中心点，并在中心点位置回归出目标位置以及状态属性等，从而实现人体摔倒检测。

1 基于中心点的目标检测方法

目标检测识别往往在图像上将目标以轴对称的框形式框出，大多数成功的目标检测器都是罗列出大量的候选框并对其分类。这样做法浪费时间，并且低效，还需要额外的后处理。本文中提出采用一种不同的方法，构建模型是将目标作为一个点，即目标的中心点。检测器通过热力图寻找中心点然后回归目标的其他属性，比如大小，3D 位置坐标，方向甚至姿态。相比较于基于目标框的检测器，基于中心点的检测网络 centernet[13]是一个端到端的、独特的、简单而快速的目标检测器。

在网络中，仅仅将图像传入全卷积网络，得到高维特征图，然后在特征图上进行卷积操作得到热力图，确定目标中心点、中心点偏移值以及目标尺寸大小。网络整体架构如图 1 所示。

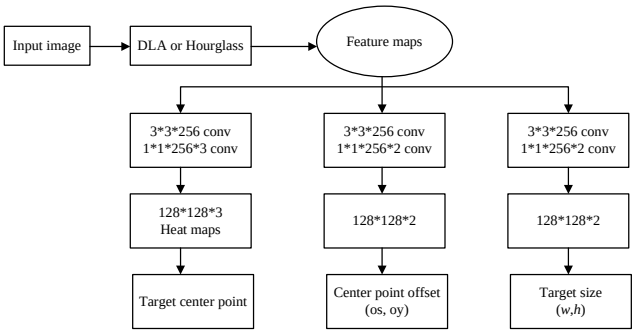

图 1 Centernet 网络整体架构

Fig.1 Overall network architecture of Centernet

1.1 关键点估计

令 $I \in R^{W\times H\times 3}$ 为输入图像，其宽为 W，高为 H。目标是生成关键点热力图 $Y^{\wedge} \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$，其中 R 是输出步长（即尺寸缩放比例），C 是关键点类型数（即输出特征通道数），在本文中由于目标类别有 3 种，因此 C=3。使用 $Y^{\wedge}_{x,y,c}$=1 表示检测到的关键点，$Y^{\wedge}_{x,y,c}$=0 表示前景。其中使用深层聚合网络 DLA[14]（deep layer aggregation）预测得到每张图像 I 的关键点热力图 $Y^{\wedge}$，DLA 是带多级跳跃连接的图像分类网络。

在训练关键点预测网络中，对于 Ground Truth（即 GT）的关键点 c，其位置定义为 $p\in R^2$，计算得到低分辨率上的对应关键点为 $\tilde{p}\in[\frac{p}{R}]$。然后通过使用高斯核 $Y_{xyc}=\exp(-\frac{(x-\tilde{p}_x)^2+(y-\tilde{p}_y)^2}{2\sigma_p^2})$ 将每一个 GT 关键点映射到热力图 $Y \in [0,1]^{\frac{W}{R}\times\frac{H}{R}\times C}$ 上，其中 σ_p 是目标尺度-自适应的标准方差[15]。如果对于同个类 c（同个关键点或是目标类别）有两个高斯函数发生重叠，选择元素级最大的。训练目标函数是一个像素级逻辑回归的焦点损失函数，其公式如下[16]：

$$L_k=\frac{-1}{N}\sum_{xyc}\begin{cases}(1-Y^{\wedge}_{xyc})^{\alpha}\ln(Y^{\wedge}_{xyc}) & \text{if } Y_{xyc}=1\\ (1-Y^{\wedge}_{xyc})^{\beta}\ln(Y^{\wedge}_{xyc})^{\alpha} & \text{otherwise}\\ \ln(1-Y^{\wedge}_{xyc}) & \text{otherwise}\end{cases} \quad (1)$$

式中：α和β是损失函数的超参数；N 是图像 I 中的关键点个数，式中除以 N 是为了将所有焦点损失归一化。在实验中，一般选取α=2，β=3。

由于在图像下采样时，GT 的关键点会因数据是离散的而产生偏差，因此对每个中心点附加预测了一个局部偏移 $O^{\wedge} \in R^{\frac{W}{R}\times\frac{H}{R}\times 2}$，所有类别 c 共享同个偏移预测，这个偏移使用 L1 loss 来训练，公式如下：

$$L_{\text{off}} = \frac{1}{N}\sum_{p}\left|O^{\wedge}_{\tilde{p}} - (\frac{p}{R} - \tilde{p})\right| \tag{2}$$

1.2 目标作为点

将目标作为点来进行目标检测是本文算法的核心思想，具体做法如下。假设在图像中目标 k 的边界框 c_k 为 $(x_1^{(k)}, y_1^{(k)}, x_2^{(k)}, y_2^{(k)})$，那么其中心位置为 $p_k = (\frac{x_1^{(k)} + x_2^{(k)}}{2}, \frac{y_1^{(k)} + y_2^{(k)}}{2})$。通过使用上文中提到的关键点估计器 $Y^{\wedge}$来预测图像中所有目标的中心点。此外，为每个目标回归出目标的尺寸，其公式为 $s_k = (x_2^{(k)} - x_1^{(k)}, y_2^{(k)} - y_1^{(k)})$。为了减少计算负担，对所有对象种类使用单一的尺寸预测，表达式为：$s^{\wedge} \in R^{\frac{W}{R}\times\frac{H}{R}\times 2}$。另外在每个中心点位置添加一个 L1 loss，目标尺寸的损失函数如式(3)所示：

$$L_{\text{size}} = \frac{1}{N}\sum_{k=1}^{N}\left|s^{\wedge}_{p_k} - s_k\right| \tag{3}$$

在此方法中，不用归一化目标尺寸而是直接使用原始像素坐标。为了调节总的 loss 的影响，分别对损失函数中的偏移损失和尺寸损失乘以一个影响系数，整个训练的目标损失函数如式(4)所示：

$$L_{\text{det}} = L_k + \lambda_{\text{size}}L_{\text{size}} + \lambda_{\text{off}}L_{\text{off}} \tag{4}$$

实验中，根据经验影响系数分别设置为 $\lambda_{\text{size}} = 0.1$，$\lambda_{\text{off}} = 1$。在整个过程中，通过使用单个网络来预测出关键点 $Y^{\wedge}$，中心点偏移量 $O^{\wedge}$ 以及目标尺寸 $S^{\wedge}$。整个网络预测结果会在每个位置输出（$C+4$）个值，即关键点类别 c，关键点偏移量 x、y 以及尺寸 W、H，所有的输出共享同一个全卷积网络骨干。对于每一个通道，其主干的特征都要通过一个单独的 3×3 卷积，ReLU 和一个 1×1 卷积。

1.3 从点回归到边界框

在前面提到所提的检测方法是通过检测每个目标的中心点，也就是在热点图中提取每个类别的峰值点。得到峰值点的方法是：将热力图上的所有的响应点与其相邻的 8 个点进行比较，如果该点响应值大于或者等于其他 8 个临近点则保留，否则丢弃。最后保留所有满足之前要求的前 100 个峰值点。令 $p_c^{\wedge}$ 是检测到的 c 类别的 n 个中心点的集合，$p^{\wedge} = \{(x_i^{\wedge}, y_i^{\wedge})\}_{i=1}^{n}$ 表示类别 c 的所有中心点的集合，每个关键点都以整型坐标(x_i,y_i)的形式给出。通过使用一个关键点值 $Y^{\wedge}_{x_i,y_i,c}$ 作为其检测置信度的度量，并且在关键点位置处生成一个边界框，其表达式如下式所示：

$$(x_i^{\wedge} + \delta x_i^{\wedge} - w_i^{\wedge}/2, y_i^{\wedge} + \delta y_i^{\wedge} - h_i^{\wedge}/2$$
$$x_i^{\wedge} + \delta x_i^{\wedge} + w_i^{\wedge}/2, y_i^{\wedge} + \delta y_i^{\wedge} + h_i^{\wedge}/2)$$

式中：$(\delta x_i^{\wedge}, \delta y_i^{\wedge}) = O^{\wedge}_{x_i^{\wedge}, y_i^{\wedge}}$ 是关键点偏移预测结果，$(w_i^{\wedge}, h_i^{\wedge}) = S^{\wedge}_{x_i^{\wedge}, y_i^{\wedge}}$ 是目标边界框的尺度预测结果。

2 数据集采集与处理

目前基本上所有的基于视觉的摔倒检测研究都是在可见光图像上基础上进行的，为了避免复杂光照条件影响以及能在夜晚和白天 24 h 工作，本文研究了基于红外图像下的人体摔倒检测。因为没有公开的红外图像人体摔倒数据集，为此在这对公开的人体摔倒数据集进行分析，了解其数据集中的人体行为，摔倒场景，图像分辨率等内容。在此基础上，搭建人体摔倒场景，设定行为内容，然后使用红外成像设备获取红外数据，制作红外图像人体摔倒数据集。

2.1 公开人体数据集分析

本文主要研究了 MuHAVi-MAS17 和 Le2i 两个公开数据集，这两个摔倒数据集在摔倒检测中使用最多，其数据量大，内容丰富，是目前主流的人体摔倒检测数据库。

MuHAVi-MAS17 是一个行为识别数据集，其中包含了人们在生活中的常做的行为，诸如走路、坐、奔跑以及需要的摔倒动作。此数据集使用 8 个摄像头在不同的方位来录制数据，内容丰富且样本多样化。数据集中每个人都有多个不同视角的摔倒图像，有左摔和右摔姿势，分辨率为 720×576。

Le2i 摔倒数据集是法国学者们使用一个分辨率为 320×240 的相机在一个拟真的场景中录制而来。数据集中有 200 多个视频序列，包含办公室、咖啡室、客厅以及演讲室等不同场景。在各种场景中的人进行了多种日常动作和摔倒行为，日常动作有下蹲、行走以及弯腰等，摔倒姿势有前后摔和左右摔等，内容丰富，数据充足。

2.2 红外图像采集系统

通过对上述摔倒数据集的分析与研究，本文使用红外图像设备自行采集红外图像人体摔倒图像，

建立摔倒数据集。选择在一间场景较为简单的房间为摔倒场景，将红外相机放置在房间的不同角落以获得不同方向的图像。红外相机分辨率为 640×480，输入电压 12 V，是一款高清单目热红外成像仪。红外相机如图 2 所示。

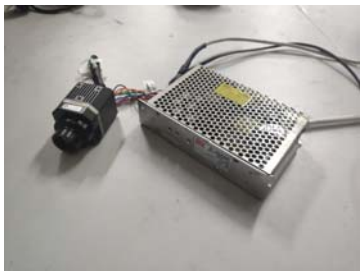

图 2 红外相机以及电源

Fig.2 Infrared camera and power supply

通过电脑读取红外相机获取的原始实时流数据，由于原始红外数据是 14 位的，无法使用电脑显示出来，所以使用 OpenCV 对其进行预处理，转为 8 位数据，并对图像线性拉伸，提高对比度。整套录制场景如图 3 所示。

图 3 红外数据录制场景

Fig.3 Infrared data recording scene

3 实验对比与分析

本文实验平台是一台 Intel Xeon 八核 E5-2620V4（2.1 GHz，QPI 速度 8.0 GT/s），64 G 内存的高性能计算工作站。为了验证本文提出的检测方法，邀请了几位成人模拟室内日常活动以及摔倒行为，图像中包含单人活动、两人活动以及多人（3～4 人）活动，共录制了 4 组数据，共获取到 30000 多张红外图像。对已经获取的红外人体摔倒数据集进行筛选，考虑到获取数据时帧率很高，导致相邻的图像内容变化不大，经过观察选择每 10 张中提取一张作为有效数据，提取大约 3000 张图像作为训练与测试数据，其中训练集 2500 余张，测试集约 260 余张。部分数据集如图 4 所示。

通过使用本文提出的检测方法在制作的红外行人摔倒数据集上进行测试，经过参数的调整，得到比较好的结果，部分实验结果如图 5 所示。

图 4 部分人体摔倒数据集

Fig.4 Partial human fall dataset

图 5 人体摔倒检测效果

Fig.5 Human fall detection effect

在以往的人体摔倒检测中，摔倒的状态往往是平躺或者侧躺的姿势，本文为了提高样本的多样性，更加真实地模拟现实场景中的摔倒，添加了例如趴着、跪着摔倒以及由于摔倒而脚抬升等不同的摔倒姿势。通过图 5 中的实验结果可以看出，算法能够准确地检测出各种摔倒姿态与正常两状态。对于在摔倒时发生前有行人或物体遮挡情况下亦能准确检测，能够满足在一定场景内的行人摔倒检测需求。

为了分析本文提出算法的性能和实时性，通过使用 YOLO v3、Faster RCNN 算法与之做对比实验，测试结果如表 1 所示。

表 1 对比实验结果

Table 1 Comparison of experimental results

Algorithm	Accuracy/(%)	Time/(ms/frame)
Yolo v3	96.9	0.0421
Faster RCNN	95.7	0.441
Ours	98.4	0.0462

从表中可以看出，Yolo v3 与本文算法的运行速度很快，而 Faster RCNN 算法速度较慢。由于本文方法网络的整个输出都是直接从关键点估计得出，因此不需要基于 IOU（intersection over union）的非极大值抑制 NMS（non max suppression）或者其他后续处理，这对整个网络的检测速度有了很大的提升。表中可以看出本文方法在红外图像人体摔倒检测中

准确率达到了 98%以上，比其他两种方法高，对有遮挡和各种不同姿态的摔倒方式等较为复杂的情况下都能有效定位与识别。

为了进一步分析实验结果的可靠性，选择摔倒检测中常用的准确率和召回率来对本文训练的模型进行评判。通过改变识别阈值，得到不同阈值下的准确率与召回率，最后得到 P-R（precision-recall）曲线，如图 6 所示。

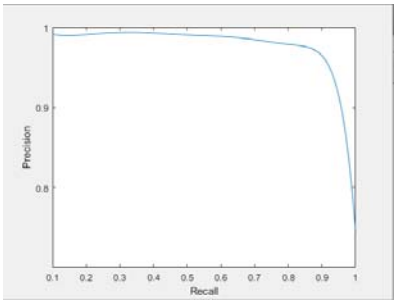

图 6 本文算法的 P-R 曲线

Fig.6 The P-R curve of the algorithm in this paper

对于 P-R 曲线来说，曲线下的面积越大，即 AP（average precision）值越大，证明模型的性能越好。从图 6 曲线中能够看出，在使用的较少的测试数据下本文训练的模型性能优越，能够准确进行红外图像下的人体摔倒检测。

4 结束语

针对人体摔倒检测问题，本文提出了一种基于关键点估计的红外图像人体摔倒检测方法。基于目前人体摔倒检测所使用的数据集特点，搭建红外图像采集系统与环境，建立了自己的红外图像人体摔倒数据集。通过关键点估计来找到人体目标中心点，然后回归人体目标属性，如目标尺寸、标签等，从而得到检测结果。实验结果表明，本文提出的方法在红外图像上能实时地进行人体摔倒检测，有较好的准确性和鲁棒性，具有较高的实际应用价值。在未来的工作中，扩展自建的红外图像人体摔倒数据集，丰富人体摔倒的场景和姿态，进一步研究红外图像下人体摔倒检测问题是未来工作的重点研究内容。

参考文献：

[1] Santos G , Endo P, Monteiro K , et al. Accelerometer-based human fall detection using convolutional neural networks[J]. *Sensors*, 2019, **19**(7): 1644.

[2] Gia T N, Sarker V K, Tcarenko I, et al. Energy efficient wearable sensor node for IoT-based fall detection systems[J]. *Microprocessors and Microsystems*, 2018, **56**: 34-46.

[3] Nadee C, Chamnongthai K. Multi sensor system for automatic fall detection[C]//2015 *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA) of IEEE*, 2015: DOI: 10.1109/APSIPA.2015.7415408.

[4] Tzeng H W, CHEN M Y, CHE J Y. Design of fall detection system with floor pressure and infrared im age[C]//2010 *International Conference on System Science and Engineering of IEEE*, 2010: 131-135.

[5] Kerdegari H, Samsudin K, Rahman Ramli A, et al. Development of wearable human fall detection system using multilayer perceptron neural network[J]. *International Journal of Computational Intelligence Systems*, 2013, **6**(1): 127-136.

[6] LIU Chengyin, JIANG Zhaoshuo, SU Xiangxiang, et al. Detection of human fall using floor vibration and multi-features semi-supervised SVM[J]. *Sensors*, 2019, **19**(17): 3720(doi: 10.3390/s19173720).

[7] Mazurek P, Wagner J, Morawski R Z. Use of kinematic and mel-cepstrum-related features for fall detection based on data from infrared depth sensors[J]. *Biomedical Signal Processing and Control*, 2018, **40**: 102-110.

[8] MIN W, CUI H, RAO H, et al. Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics[J/OL]. *IEEE Access*, 2018, **6**: 9324-9335.

[9] FENG W, LIU R, ZHU M. Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera[J]. *Signal, Image and Video Processing*, 2014, **8**(6): 1129-1138.

[10] 邓志锋, 闵卫东, 邹松. 一种基于 CNN 和人体椭圆轮廓运动特征的摔倒检测方法[J]. **图学学报**, 2018, **39**(6): 30-35.
DENG Zhifeng, MIN Weidong, ZOU Song. A fall detection method based on CNN and human elliptical contour motion features[J]. *Journal of Graphics*, 2018, **39**(6): 30-35.

[11] 卫少洁, 周永霞. 一种结合 Alphapose 和 LSTM 的人体摔倒检测模型[J]. **小型微型计算机系统**, 2019, **40**(9): 1886-1890.
WEI Shaojie, ZHOU Yongxia. A human fall detection model combining alphapose and LSTM[J]. *Minicomputer System*, 2019, **40**(9): 1886-1890.

[12] 赵芹, 周涛, 舒勤. 飞机红外图像的目标识别及姿态判断[J]. **红外技术**, 2007, **29**(3): 167-169.
ZHAO Qin, ZHOU Tao, SHU Qin. Target recognition and attitude judgment of aircraft infrared image[J]. *Infrared Technology*, 2007, **29**(3): 167-169.

[13] ZHOU X, WANG D , Krhenbühl P. Objects as points [J/OL][2019-04-25]. arXiv:1904.07850(https://arxiv.org/ abs/1904.07850).

[14] YU F, WANG D, Shelhamer E, et al. Deep layer aggregation[J/OL] [2019-01-04]. arXiv:1707.06484(https://arxiv.org/abs/1707.06484)

[15] Law H, DENG J . CornerNet: detecting objects as paired keypoints[J]. *International Journal of Computer Vision*, 2020, **128**(3): 642-656.

[16] LIN T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017: DOI**:** 10.1109/ICCV.2017.324.