

基于Alphapose骨骼点和GRU的摔倒检测技术

李金泳, 孙琛琛, 陈庆涛, 刘述煌, 李金字, 徐婉瀛

天津理工大学, 天津

Email: 834855032@qq.com

收稿日期: 2021年3月14日; 录用日期: 2021年4月8日; 发布日期: 2021年4月15日

摘 要

基于现有的一些人体摔倒检测模型实现复杂, 适用性差等缺点, 本文提出了一种新的更加简便, 适用性更强的人体摔倒检测模型。该模型是一种基于GRU神经网络和人体骨骼关键点的人体摔倒检测模型。该模型中, 先通过Alphapose对每一帧图像进行人体骨骼关键点识别与检测, 然后将得到的骨骼关键点数据进行归一化数据处理, 再分组输入到GRU神经网络中进行时序特征提取, 最后将GRU模型中隐含层的输出向量输入到全连接层进行处理并得出检测结果。本文使用的是UR Fall Detection Dataset热舒夫大学摔倒数据集进行测试实验, 并与多种检测模型的实验性能进行了横向对比。实验结果表明本文的模型在多场景, 多视角和多种摔倒姿势等情况下较其他模型都有较高的检测精度, 且实现难度较其他模型而言要低。

关键词

摔倒检测, 骨骼关键点, GRU神经网络

Fall Detection Technology Based on Alphapose Bone Key Points and GRU Neural Network

Jinyong Li, Chenchen Sun, Qingtao Chen, Shuhuang Liu, Jinyu Li, Wanying Xu

Tianjin University of Technology, Tianjin

Email: 834855032@qq.com

Received: Mar. 14th, 2021; accepted: Apr. 8th, 2021; published: Apr. 15th, 2021

文章引用: 李金泳, 孙琛琛, 陈庆涛, 刘述煌, 李金字, 徐婉瀛. 基于 Alphapose 骨骼点和 GRU 的摔倒检测技术[J]. 计算机科学与应用, 2021, 11(4): 840-848. DOI: 10.12677/csa.2021.114086

Abstract

Based on the shortcomings of some existing human fall detection models, such as complex implementation and poor applicability, this paper proposes a new simpler and more applicable human fall detection model. This model is a human fall detection model based on GRU neural network and key points of human bones. In this model, the key points of human bones are identified and detected through Alphapose for each frame of image, Then the obtained bone key point data is normalized data processing, and then grouped and input into the GRU neural network for time series feature extraction, finally, the output vector of the hidden layer in the GRU model is input to the fully connected layer for processing and the detection result is obtained. This article uses the UR Fall Detection Dataset Rzeszow University fall data set for test experiments, and horizontal comparison with the experimental performance of a variety of detection models. Experimental results show that the model in this paper has higher detection accuracy than other models in multiple scenes, multiple perspectives, and multiple falling postures, and the difficulty of implementation is lower than other models.

Keywords

Fall Detection, Bone Key Points, GRU Neural Network

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

我国老龄化现象不断加剧, 据国家统计局公布数据显示[1], 2019 年末全国大陆人口 140005 万人, 我国 60 周岁及以上人口 2538 万人, 占总人口的 18.1%。其中 65 周岁及以上人口 17604 万人, 占总人口的 12.6%。由于老年人身体机能下降, 摔倒后容易引起严重后果, 而当前对于老人的服务资源不能满足社会要求, 独生子女外出工作, 摔倒的老人往往错过最佳救治时间。当老年人独自在家发生摔倒时, 及时的检测能尽可能保障老人的安全。

目前, 人体摔倒检测主要有基于穿戴式传感器的摔倒检测和基于计算机视觉的摔倒检测。基于穿戴式传感器的摔倒检测通常将设备放在人的身体某个部位上, 由外部传感设备采集数据, 提取相关数据后利用机器学习模型进行分组, 检测的准确性十分依赖于提取特征值的好坏, 对于设备的要求较高[2] [3]。基于计算机视觉的摔倒检测是将摄像头放置在用户日常活动范围的固定位置, 对获取的图像进行处理, 通过机器学习、深度学习等方法实现对监控视频进行摔倒检测[4]。杨碧瑶[5]使用 PCANet 提取图像特征, 依次通过两个 SVM 分类器来判断是否摔倒。徐世文[6]采用 CenterNet 目标检测网络, 利用建立的红外图像摔倒数据集进行网络的训练与测试。Wen-Nung Lie [7]采用卷积神经网络(CNN)为每个输入帧提取 2D 骨架, 然后使用 LSTM 来对摔倒进行预测。

当前两种主流的摔倒检测方法都存在一些问题。基于穿戴式传感器的摔倒检测设备成本要求高, 检测精度低, 用户需要随身佩戴传感器设备导致生活不便。基于计算机视觉的摔倒检测法存在场景适应性差, 不同摔倒姿势误检率高等问题[8]。针对以上问题, 本文提出一种基于人体骨骼关键点的摔倒检测模型。通过使用 Alphapose 实现对人体骨骼关键点特征的提取, 将得到关键点坐标归一化后根据时间序列

输入 GRU 神经网络进行分类, 实现人体摔倒检测, 避免了场景中环境的改变导致图像特征提取效果差, 并在公开数据集上进行测试, 表明本模型拥有良好的性能。

2. 系统结构

检测算法与 GRU 神经网络模型结构如图 1 所示, 首先将视频分割成帧, 帧图像输入 Alphapose 得到每帧图片人体的 17 个坐标点 (x, y) 。这 17 个关键点包括人的头部, 上半身, 下半身几乎所有关节的关键点, 可以有效的代表人体。人体在运动时, 身体骨骼点 x, y 的位置存在明显变化, 将得到的坐标点进行归一化处理。检测多帧图像后可以得到 (x, y) 的序列, 当序列长度满足 GRU 设置的 size 后, 将序列输入 GRU 神经网络提取特征, 每当有一个新的帧加入后删除原序列的最旧坐标值, 保持坐标序列为设置的窗口长度。最后将特征信息通过全连接层进行分类, 根据输出结果判断是否摔倒。

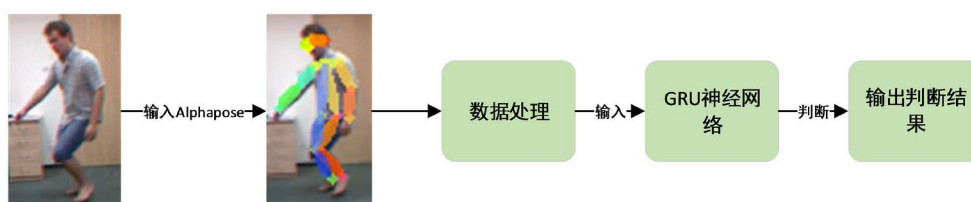


Figure 1. Fall detection flow chart
图 1. 摔倒检测流程图

3. 摔倒检测方法

3.1. Alphapose 骨骼点检测模型

Alphapose 是由上海交通大学卢策吾团队提出的一种自顶向下的姿态估计技术[9], 通过区域姿态估计 (Regional Multi-Person Pose Estimation, RMPE) 框架来实现姿态估计, 能有效解决目前骨骼模型检测框定位错误和检测框冗余问题。

RMPE 框架主要包括 Symmetric Spatial Transformer Network (SSTN), Parametric Pose Non-Maximum-Suppression (P-NMS) 和 Pose-Guided Proposals Generator (PGPG) 三个模块。对称空间变换网络 SSTN 模块先进行姿态估计, 将估计结果映射到原图, 以此来调整原本的识别框, 它能使目标检测的框变得更精准。姿态非极大值抑制器 P-NMS 通过姿态距离和空间距离消除冗余的检测框, 当姿态距离和空间距离的权重和低于某一阈值时删除检测框, 它能有效的消除冗余的检测框。姿态引导样本生成器 PGPG 通过 Kmeans 算法得到原子姿势, 根据密集采样生成偏移量产生增强的训练建议, 增加了样本的数量, 它使样本数量增多, 模型的精准度得到提高。通过这三个模块 Alphapose 有效的提升了骨骼检测的性能。Alphapose 骨骼点检测模型效果如图 2。

卢策吾团队在 COCO 数据集上与 Openpose, Mask-RCNN 进行骨骼检测模型对比试验, Alphapose 在准确率和平均速度上均有优势, 是目前精确度最高的骨骼检测模型。本文摔倒检测模型采用 Alphapose 得到每帧图像的 17 个骨骼关键点坐标 (x, y) 。

3.2. GRU 网络模型

GRU 网络是 LSTM 网络的一种变体, 它是一种特殊的 RNN (循环卷积网络) 类型。摔倒动作并不是瞬时发生而是由一连串的状态组成的, 与传统的卷积神经网络相比, RNN 能记录并使用先前的状态信息。

传统的 RNN 结构如图 3 所示, 是一种能将以往的信息利用的网络模型, 但是这种模型存在许多的



Figure 2. Alphaspose skeleton point detection model

图 2. Alphaspose 骨骼点检测模型

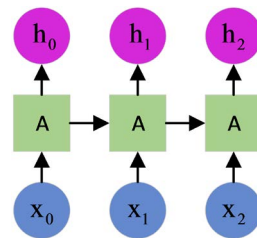


Figure 3. RNN structure diagram

图 3. RNN 结构图

弊端。当预测所需的信息距离比较远时，因为结果可以拆分成中间梯度的乘积，如果中间权重小于 1，经过多次自乘后，梯度会出现指数级缩小，这称为梯度消失，与梯度消失相对，如果中间权重大于 1，经过多次自乘后会出现梯度爆炸。**LSTM** (长短期记忆网络)可以解决这一问题，其模型的核心思想是“细胞状态”，它通过叫做“门”的结构来通过或去除信息到细胞状态。**LSTM** 有三种“门”，在第一步通过遗忘门来决定忘记的信息，第二步通过输入门来确定细胞存放的新信息，第三通过输出门来更新旧细胞的状态并作为结果输出。结构如图 4。

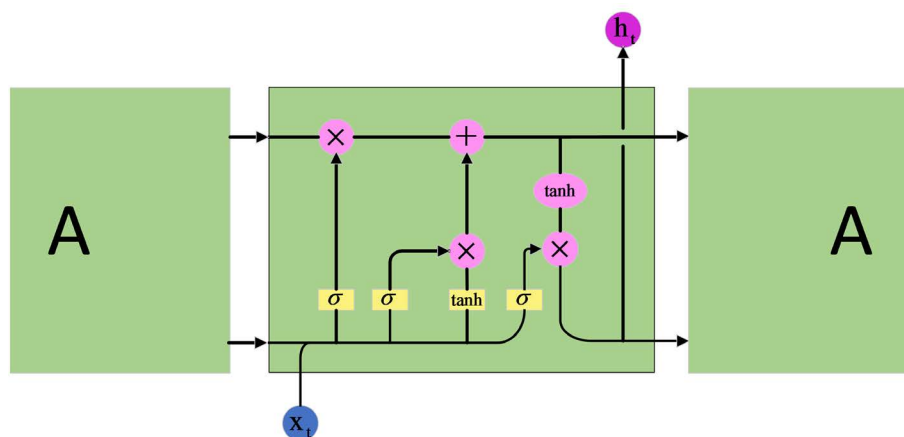


Figure 4. LSTM structure diagram

图 4. LSTM 结构图

GRU 作为 LSTM 的一种变体，他只有更新门和重置门，具体结构如图 5 所示。

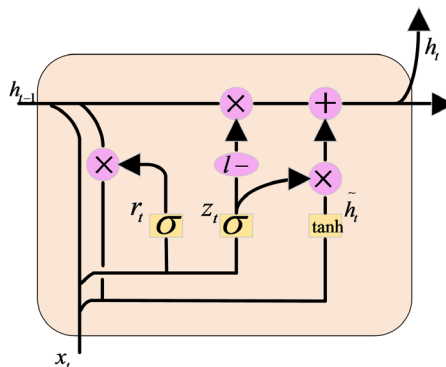


Figure 5. GRU cell structure

图 5. GRU 细胞结构

更新门相当于将 LSTM 的输入门和遗忘门融合，并加入少许改动，控制前一细胞状态信息带入当前状态的程度。重置门控制前一状态有多少信息写入到候选集上。计算过程为公式(1)到公式(4)。

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

GRU 能够利用之前的信息，由于 GRU 的结构比 LSTM 更简单，在相同数据集的情况下 GRU 拥有明显的速度优势。本文摔倒模型将得到的骨骼坐标点进行处理后，按照时间的顺序输入并得到摔倒结果分类。

3.3. 人体摔倒检测模型

本文基于 Alphapose 和 GRU 神经网络提出了摔倒检测模型。Alphapose 骨骼关键点检测是目前已知技术中准确度和速度最高的模型，GRU 神经网络与其他处理时间序列中包含的信息拥有优势的模型 LSTM 和 RNN 相比在摔倒检测中拥有较高的准确率，实验将说明这一点。因为 GRU 拥有比 LSTM 更简单的细胞结构，所以神经网络的处理速度更快[10]。本文首先使用 Alphapose 对运动的每一帧图像进行人体骨骼点提取得到 17 个(x,y)坐标，对得到的坐标进行处理后将骨骼点坐标输入 GRU 神经网络提取特征，每有新的一帧图像加入序列，旧的一帧图像被抛弃，保证序列中始终存储最新的一系列帧图像。将 GRU 提取的隐层特征经过全连接层分为两类：摔倒，未摔倒，来实现人体摔倒检测判断。如公式(5)所示

$$y = f(h_t) = \begin{cases} 1, & \text{摔倒} \\ 0, & \text{未摔倒} \end{cases} \quad (5)$$

4. 数据集与数据处理

UR Fall Detection Dataset 热舒夫大学摔倒数据集包含 30 个跌倒与 40 个日常生活活动摔倒的视频与加速度等数据。通过 Alphapose 对视频进行处理，它通过将视频流转换成若干帧图像，再对每一帧图像进行骨骼点坐标识别，得到坐标 json 文件。

由于每个图像中人体在图像中的位置不同，摔倒检测只需要关注人体的动作，同时增强模型通用性，所以需要数据归一化。确定每一个图像坐标的 X_MIN, X_MAX, Y_MIN, Y_MAX，通过下列等

式对数据进行归一化处理。计算过程为公式(6) (7)。

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

$$y = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (7)$$

GRU 数据样本通过滑动窗口法获取。首先设置窗口大小 s ，如图 x 所示，当前帧为 n ，其标签 $L(n)$ 为该样本标签，其中 $Z(n)$ 包含 17 个骨骼点 (x, y) 的坐标，即 $Z(n) = \{(x_1, y_1), (x_2, y_2), \dots, (x_{17}, y_{17})\}$ ，窗口不断向右滑动来获取新样本，当有新数据添加进窗口时，删除窗口最左侧数据帧。据此来获取测试数据样本。

5. 实验

本文实验环境为：Win10 操作系统，编程环境为 Anaconda 下 Python3.6，开发工具为 Pycharm 2020.1，处理器为 Intel(R) Core(TM) i7-8750H CPU @2.20 Ghz 2.21 GHz，内存为 16 G，显卡为 GTX1050，搭建 GRU 网络使用 Keras2.2.4。

5.1. 评价标准

实验结果的样本主要被分为四类[11]，其中真正例 TP(True Position)表示摔倒样本被正确分类的样本；假正例 FP(False Position)表示未摔倒样本被模型错误分类为摔倒样本；假反例 FN(False Negative)表示摔倒样本被错误分类为未摔倒样本；真反例(True Negative)表示未摔倒样本被正确分类为未摔倒样本。

本文选取摔倒检测常用的准确率(Accuracy)和漏警率(MA)表示，准确率表示模型分类正确的样本占所有样本的比重，可以评价一个模型的整体性能，如公式(8)所示。漏警率(MA)表示将摔倒模型错分为非摔倒样本占实际摔倒样本总数的比重，因为错把摔倒样本当作未摔倒样本造成比把未摔倒样本错分为摔倒样本的后果要严重的多，所以要关注模型的漏警率，如公式(9)所示，其中 P 表示正例的样本数量， N 表示反例的样本数量， TP ， TN ， FN ， P ， N 通过实验统计获取。

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (8)$$

$$\text{MA} = \frac{FN}{TP + FN} \quad (9)$$

5.2. 实验结果分析

因为目前循环卷积神经网络包括传统意义上的 RNN (循环卷积网络)，LSTM (长短时记忆)，GRU (LSTM 的变体)以及其他类型的循环卷积神经网络，本文主要选取 RNN，LSTM，GRU 三种来进行横向对比实验，利用控制变量法实验中除去神经网络的搭建的其他步骤均保持一致。实验数据集选取热舒夫大学摔倒数据集 UR Fall Detection Dataset。

在训练过程中，选取摔倒过程和躺下的状态为摔倒训练样本，其他行为(坐下，起立，行走)为正常训练样本，窗口大小为 40 帧进行训练。本文采用交叉验证法进行实验，通过随机均匀选取 70 个视频中 14 个视频作为训练集进行循环训练达到训练预期，使用剩余数据作为训练样本对模型进行评估，其中摔倒样本包括 24 个，未摔倒样本包括 32 个，共划分为 5 组 P1，P2，P3，P4，P5。实验样本的分类结果如表 1~表 3。

Table 1. RNN experimental sample classification results
表 1. RNN 实验样本分类结果

	P1	P2	P3	P4	P5	average
TP	22	19	18	24	22	21
FP	5	1	1	6	2	3
FN	2	5	6	0	2	3
TN	27	31	31	26	30	29

Table 2. LSTM experimental sample classification results
表 2. LSTM 实验样本分类结果

	P1	P2	P3	P4	P5	average
TP	21	21	22	21	23	21.6
FP	2	2	2	0	2	1.6
FN	3	3	2	3	1	2.4
TN	30	30	30	32	30	30.4

Table 3. GRU experimental sample classification results
表 3. GRU 实验样本分类结果

	P1	P2	P3	P4	P5	average
TP	24	21	21	22	24	22.4
FP	2	0	0	0	2	0.8
FN	0	3	3	2	0	1.6
TN	30	32	32	32	30	31.2

对比 3 个不同模型的 TP, FP, FN, TN 可以发现 GRU 不管在摔倒样本检测还是在未摔倒样本检测的准确率高于其他模型。使用 3 个模型 average 的数值来计算不同模型的 Accuracy 和 MA, 结果如表 4 所示。从表中可以看出, 本文提出的 GRU 摔倒检测模型在准确率和漏警率上明显优于 RNN 和 LSTM 模型。

为了进一步得到更直观的比较, 图 6, 图 7 分别为不同实验样本得到的 3 个模型准确率 Accuracy 和漏警率 MA 对比折线图。从图中可以看出, GRU 模型在不同的实验样本中均表现出优于其他模型的性能。

Table 4. Comparison results of different models
表 4. 不同模型对比结果

	RNN	LSTM	GRU
Accuracy	89.3%	92.9%	95.7%
MA	12.5%	10.0%	6.6%

本文 GRU 模型通过与 RNN, LSTM 模型进行对比得到分类为 TP, FP, FN, TN 的实验样本, 得到的实验数据包括利用交叉验证法的五组测试样本和平均值。实验数据表明 GRU 相比于 RNN 准确率提高 6.4%, 相比于 LSTM 准确率提高 2.8%。为了尽量避免把摔倒的样本分类为未摔倒的样本, MA 越小性能越有优势。在 3 种模型中, GRU 拥有最低的漏警率 MA。

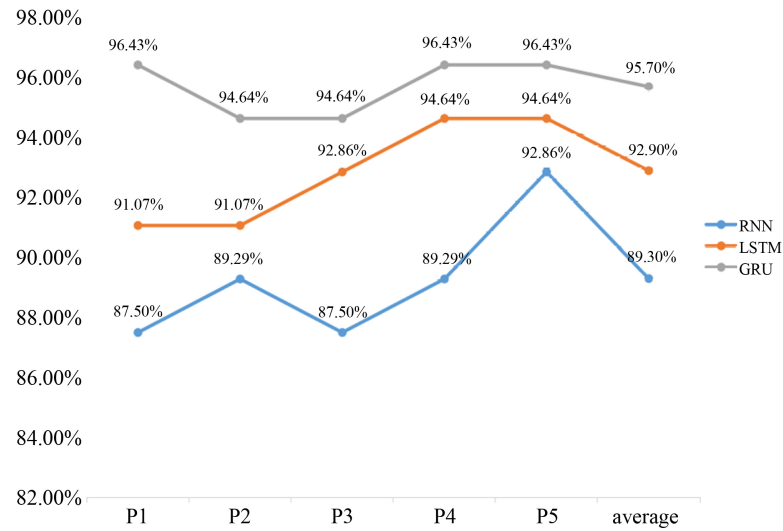


Figure 6. Accuracy comparison line chart

图 6. Accuracy 对比折线图

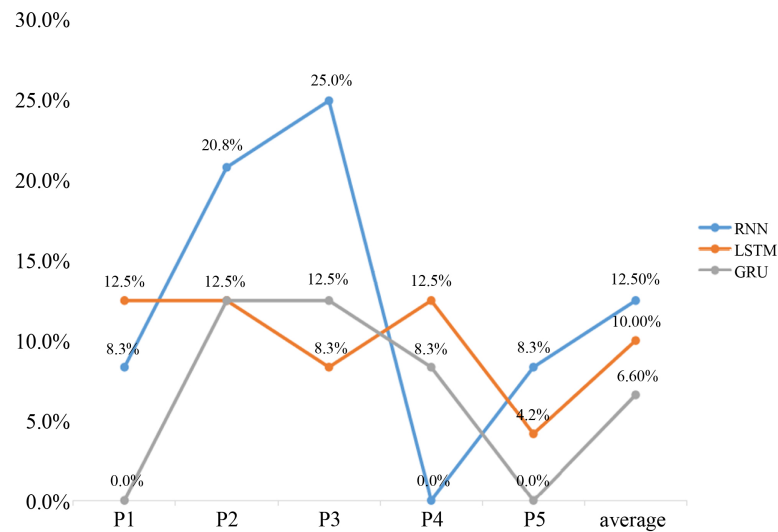


Figure 7. MA comparison line chart

图 7. MA 对比折线图

6. 总结

本文比较了主流的两种摔倒检测方法，针对可穿戴式的设备成本要求高，检测精度低和当前计算机视觉环境适应性差，误检率高等问题，提出了一种基于 Alphapose 和 GRU 的摔倒检测模型，该模型通过 Alphapose 检测人体骨骼关键点，通过对关键点坐标进行归一化处理后，利用滑动窗口法将时序信息输入 GRU 网络提取特征实现摔倒检测分类与判断。通过 UR Fall Detection Dataset 热舒夫大学摔倒数据集实验表明本模型拥有较高的精度。

基金项目

大学生创新创业训练计划项目(202010060160)。

参考文献

- [1] 国家统计局. 中华人民共和国 2019 年国民经济和社会发展统计公报[EB/OL]. http://www.stats.gov.cn/tjsj/zxfb/202002/t20200228_1728913.html, 2020-02-28.
- [2] 王亚玲. 基于 NB-IoT 技术的可穿戴式老人摔倒监测系统的设计与开发[D]: [硕士学位论文]. 南昌: 江西师范大学, 2019.
- [3] 贺佚南. 基于可穿戴传感器的摔倒检测研究[D]: [硕士学位论文]. 广州: 华南理工大学, 2018.
- [4] Miao Yu, Liyun Gong, and Stefanos Kollias.(2017).Computer vision based fall detection by a convolutional neural network. Proceedings of the 19th ACM International Conference on Multimodal Interaction,416-420. doi: <https://doi.org/10.1145/3136755.3136802>
- [5] 杨碧瑶. 基于计算机视觉的独居老人摔倒检测方法研究[D]: [硕士学位论文]. 西安: 陕西科技大学, 2020.
- [6] 徐世文. 基于红外图像特征的人体摔倒检测方法[D]: [硕士学位论文]. 绵阳: 西南科技大学, 2020.
- [7] W. Lie, A. T. Le and G. Lin.(2018).Human fall-down event detection based on 2D skeletons and deep learning approach.2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369778.
- [8] 陈永彬, 何汉武, 王国桢, 王桂棠. 基于机器视觉的老年人摔倒检测系统[J]. 自动化与信息工程, 2019, 40(5): 37-41.
- [9] Fang, H., Xie, S., Tai, Y. and Lu, C. (2017) RMPE: Regional Multi-Person Pose Estimation. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2353-2362. <https://doi.org/10.1109/ICCV.2017.256>
- [10] Fu, R., Zhang, Z. and Li, L. (2017) Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, 11-13 November 2016, 324-328. <https://doi.org/10.1109/YAC.2016.7804912>
- [11] Noury, N., *et al.* (2007) Fall Detection-Principles and Methods. 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 22-26 August 2007, 1663-1666. <https://doi.org/10.1109/IEMBS.2007.4352627>