

基于空间特征的无线体域网人体姿态识别算法

金驰, 李志军, 孙大洋, 胡封晔

(吉林大学, 吉林 长春 130012)

摘要: 针对传统基于图像视频的姿态识别算法中所存在的计算成本高、摄像盲区多、隐私易泄露等问题, 提出了一种基于手机加速度与陀螺仪数据的卷积—卷积长短时记忆—注意力 (CCLA, convolution-convolutional long short-term memory-attention) 人体姿态识别算法。使用卷积神经网络对姿态数据进行空间特征提取, 采用卷积长短时记忆网络挖掘数据中隐含的时序信息, 模拟人脑选择注意力机制构建 Attention (注意力) 编码器进行更高层次的时空特征提取, 以实现姿态的精准分类。在加州大学欧文分校提出的基于智能手机的人体活动与转换姿态识别数据集上对 CCLA 算法进行了测试, 实现了对 12 元姿态的分类识别, 识别准确率达 93.27%。

关键词: 神经网络; 姿态识别; 注意力机制; 无线体域网

中图分类号: TP183

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2019.00121

Human activity recognition algorithm based on the spatial feature for WBAN

JIN Chi, LI Zhijun, SUN Dayang, HU Fengye

Jilin University, Changchun 130012, China

Abstract: Traditional image-based activity recognition algorithms have some problems, such as high computational cost, numerous blind spots and easy privacy leakage. To solve the problem above, the CCLA (convolution-convolutional long short-term memory-attention) activity recognition algorithm based on the acceleration and gyroscope data was proposed. The convolutional neural network was used to extract spatial features of activity data and got the hidden time series information from the convolutional long short-term memory network. Simulating human brain selecting attention mechanism, attention-encoder was constructed to extract the spatial and temporal features at a higher level. The CCLA algorithm was tested on UCI-HAPT (university of California Irvine-smartphone-based recognition of human activities and postural transitions) public data set, and realized the classification of 12 types of activity with the accuracy of 93.27%.

Key words: neural network, activity recognition, attention mechanism, wireless body area network (WBAN)

1 引言

随着深度学习技术的不断发展, 深度学习被越来越多地用于解决传统问题。早期姿态识别算法主要基于计算机视觉技术。从提出至今, 基于视频图像信息的人体姿态识别算法已得到广泛研究^[1-3]。但是该技术大多局限于研究层面, 在实际应用中, 仍然存在着较明显的缺陷, 如计算成本高、摄像盲区多等。与之相反的是, 近年来移动设备和传感器在

尺寸、成本和功耗方面都取得了显著进步, 这为研究人们的日常活动和姿态识别提供了新的数据来源, 基于陀螺仪、加速度计等多种信息采集方案被提出。人体姿态识别的实现方式也从对经典机器学习算法进行改进与应用这一层面逐渐发展, 逐步走向多模型融合^[4-5]。

为了解决基于视频图像信息的人体姿态识别算法中存在的问题, 实现低算力、高实时性的姿态识别, 本文构建了一种基于陀螺仪、加速度计数据

的卷积—卷积长短时记忆—注意力（CCLA，convolution-convolutional long short-term memory-attention）算法，并在数据集上对其进行评估。实验结果表明，该方法在保证准确率的同时，可实现毫秒级的识别时延。同时，通过与当前所提出的各项方法对比，验证了该算法的有效性。

2 CCLA 算法模型概述

CCLA 算法模型按照数据预处理、特征提取、预测输出的流程进行设计，CCLA 算法流程如图 1 所示，各层次设计介绍如下。

2.1 数据预处理层

数据预处理层分为标准化单元与数据分帧单元^[6]。标准化单元对数据进行缩放，去除其量纲，考虑加速度与角速度数据特性，算法模型中采用了 min-max 标准化。分帧单元对数据进行帧长包括 100 组数据、重叠率为 90% 的分帧^[7]。在 50 Hz 回传条件下，帧长包括 100 组数据的数据帧包含 2 s 的活动数据信息，可有效实现对当前活动的判别，且两次识别之间的时间间隔为 0.2 s，可基本满足应用场景的实时要求。

2.2 空间特征提取层

空间特征提取层采用卷积神经网络对数据进行空间特征提取^[8-11]。数据经标准化与分帧处理后，形成尺寸为 $N_s \times \text{Samplepoint} \times \text{Channel}$ 的“数据图”流。因此，卷积神经网络可按照标准的图像处理架构对其进行处理。本文模型中采用“Conv-Pool-Conv-Pool-Conv”结构，网络共有 5 层。数据填充选择“SAME”方式对边界进行零值填充。在卷积层之间加入池化层，池化方式采用最大值池化法。经空间特征提取层处理后，输出数据尺寸为 $N_s \times N_f \times \text{Samplepoint} \times \text{Feature}$ ，这表示从 N_f 个方面对数据进行了空间特征提取。

2.3 衔接层

经空间特征提取层处理后，需将时间序列进行分割，并按时序输入至 ConvLSTM 单元，以供下一步时间特征提取。衔接层将上层输出数据重构为尺寸是 $N_s \times N_f \times \text{Timestep} \times \text{Batch} \times \text{Feature}$ 的 5 维数据流，其中， $\text{Timestep} \times \text{Batch} = \text{Samplepoint}$ 。Conv 层与 ConvLSTM 层之间通过衔接层进行衔接，可直接将上一层中提取的 N_f 维空间特征传递至下一层，尽量保留特征。

2.4 时间特征提取层

在时间特征提取层中采用 ConvLSTM 网络^[12-14]。由于循环神经网络自身具有一定的深度，且计算复杂度较高^[12,15]，故经测试后选择双层 ConvLSTM 网络进行时间特征提取。经该层处理后，输出数据尺寸为 $N_s \times N_f \times \text{Timestep} \times \text{Batch} \times \text{Feature}$ 。此外，为了与 Attention 编码层进行衔接，还需在该层最后加入衔接层将数据流进行重塑，最终输出数据尺寸为 $N, T, F = (N_s, \text{Timestep} \times \text{Batch}, N_f \times \text{Feature})$ 。

2.5 Attention 编码层

Attention 编码层的主体是 Attention 单元^[16-17]，但在数据进入 Attention 单元之前，需要先经过全连接单元进行处理。已知输入数据尺寸为 (N, T, F) ，全连接单元的作用是将 F 个特征映射至 F^* 个特征， F^* 个特征为原本 F 个特征的线性组合。此过程可作为提高下一步 Attention 编码分辨率的特征预处理。Attention 编码单元所实现的是以每个时间点的特征为基础，对时间点各自进行编码，获取特征的更高层次表达。编码原则^[18]为

$$\begin{cases} e_{ij} = a(s_{t-1}, h_j) \\ \alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^F \exp(e_{ik})} \\ C_t = \sum_{j=1}^F \alpha_{ij} h_j \end{cases} \quad (1)$$

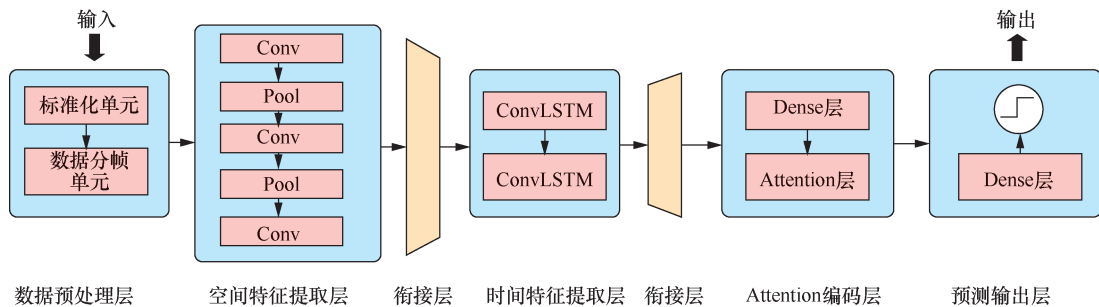


图1 CCLA 算法流程

其中, s_t 为 t 时刻编码器状态, h_{ij} 代表 t 时刻第 j 个特征, a 为转换函数, e_{ij} 为 t 时刻第 j 个特征的影响力因子, α_{ij} 为编码权值, C_t 为 t 时刻的编码结果。

2.6 预测输出层

预测输出层主要是对上一层的编码结果进行识别分类, 由于 Attention 编码中已包含时空信息, 因此, 可直接通过 Dense+Softmax 结构进行预测输出, 实现多元姿态识别。

综上所述, CCLA 算法模型架构如图 2 所示。

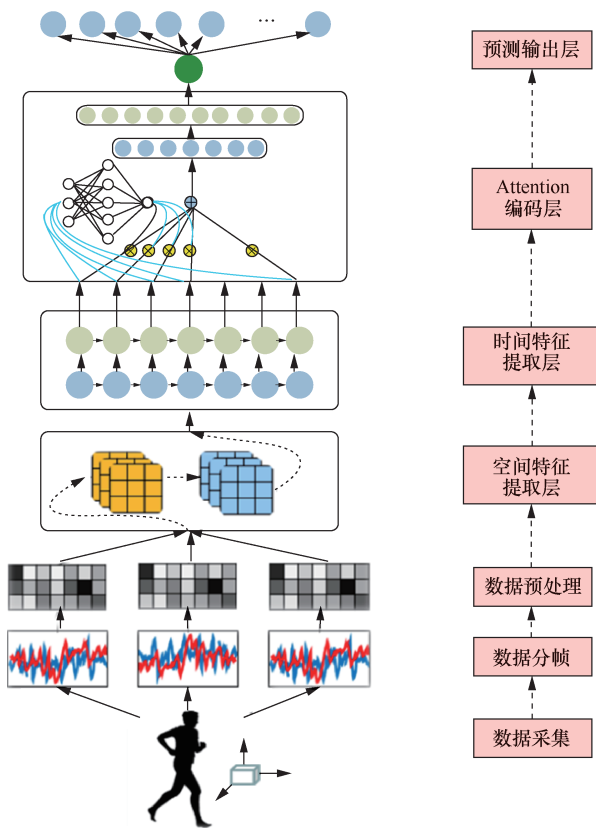


图 2 CCLA 算法模型架构

3 实验与结果分析

3.1 姿态识别样本库

模型的训练与测试采用加州大学欧文分校提出的基于智能手机的人体活动与转换姿态识别 (UCI-HAPT, university of California Irvine-smartphone-based recognition of human activities and postural transitions) 数据集^[19], 数据集中包含 12 类姿态数据, 其中, 包括 3 类静态动作 (站姿、坐姿、躺姿)、3 类动态动作 (行走、下楼、上楼) 及 6 类转换动作 (站—坐、坐—站、坐—躺、躺—坐、站—躺、躺—站)。在对无标签数据进行过滤后, 获得有效

数据共 815 614 条。经分帧获得样本 81 552 个, 数据集样本分布如图 3 所示, 其中, 选取 70 000 个样本作为训练集, 11 552 个样本作为测试集, 测试集数量占整体样本数量的 14%。图 3 中蓝色部分为训练集各类样本分布, 黄色部分为测试集样本分布, 右侧数字为该样本总数。

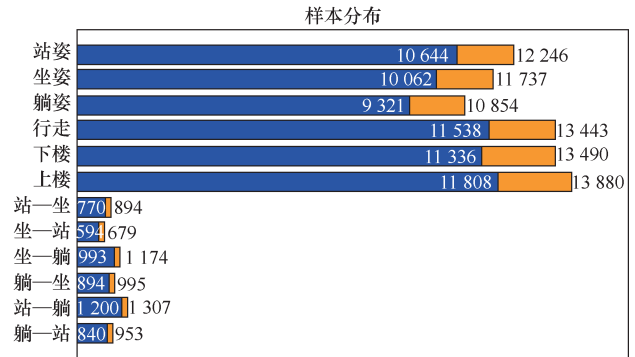


图 3 数据集样本分布

3.2 模型参数与测试环境

CCLA 网络结构细节如表 1 所示, 网络整体为 16 层, 除了进行尺寸变换的 Permute 层与 Reshape 层, 网络实际深度为 11 层, 参数总量为 27 658 个。测试选用 Intel Core i3-3110M 处理器, 在 Windows 10 64 位操作系统下进行, 模型运行在 Python 3.6 内核的 Anaconda 环境中。

表 1 CCLA 网络结构细节

层数	层类型	输出尺寸	参数量/个
1	Conv	(None,8,100,6)	136
2	MaxPooling	(None,8,100,6)	0
3	Conv	(None,8,100,6)	1 032
4	MaxPooling	(None,8,100,6)	0
5	Conv	(None,10,100,6)	1 290
6	Reshape	(None,10,10,10,6)	0
7	Permute	(None,10,10,10,6)	0
8	ConvLSTM	(None,10,8,10,6)	5 216
9	ConvLSTM	(None,10,10,10,6)	6 520
10	Permute	(None,10,10,10,6)	0
11	Reshape	(None,100,60)	0
12	Dense	(None,100,196)	11 956
13	Permute	(None,196,100)	0
14	Dropout	(None,196,100)	0
15	Attention	(None,100)	296
16	Dense	(None,12)	1 212
总参数			27 658

3.3 评估度量指标

为了评价 CCLA 算法的性能，本文在 UCI-HAPT 数据集上进行了一组实验，并使用准确率、精准率和召回率 3 个指标对模型进行评价，其定义分别为

$$\begin{cases} \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Precision} = \frac{TP}{TP+FP} \\ \text{Recall} = \frac{TP}{TP+FN} \end{cases} \quad (2)$$

其中，FN (false negative) 表示被判为负样本的正样本，FP (false positive) 表示被判为正样本的负样本，TN (true negative) 表示被判为负样本的负样本，TP (true positive) 表示被判为正样本的正样本。

3.4 测试结果与分析

通过训练与测试，算法在训练集和测试集分别取得了 95.28%和 93.27%的准确率，算法在训练集的精准率、召回率分布矩阵和在测试集的精准率、召回率分布矩阵分别如图 4 和图 5 所示，其中，粗体数字表示分类正确的样本数目。

根据图 4 和图 5 可知，算法在站姿、躺姿、

坐姿、行走、下楼、上楼 6 类标准动作上表现出良好的识别性能，仅行走与下楼之间混淆次数较多；在其余 6 类转换动作上的性能较前 6 类有所下降，较严重的在于站—躺（躺—站）与坐—躺（躺—坐）之间的混淆，但这可以通过前一标准状态进行算法优化。通过前文介绍的数据集样本分布可知，相较于标准动作，转换动作的样本数量很少，这有可能是模型在此 6 类样本分类中性能下降的主要原因。

3.5 模型性能对比评估

为了对模型进行进一步评估，本文引用了 UCI-HAPT 官网中的一篇相关文献^[5]与本文所提出的 CCLA 模型进行对比，对比范围为 6 类标准动作，指标为准确率，模型性能对比如表 2 所示。通过比较可以看出，在 6 类动作的识别上，CCLA 算法模型在 4 类姿态（坐姿、行走、下楼、上楼）上均取得最好效果，比其他方法的最优性能准确率提高了 2%~21%，在两类姿态（躺姿、站姿）上虽并未达到最好效果，但与最优性能准确率相差在 3%以内。同时，CCLA 每个动作识别耗时仅为 6 ms。总之，存在转换动作干扰条件下，CCLA 模型的整体性能表现仍然占优势。

姿态	站姿	坐姿	躺姿	行走	下楼	上楼	站—坐	坐—站	坐—躺	躺—坐	站—躺	躺—站	精准率
站姿	10 498	26	93	0	24	0	4	2	0	1	1	16	98.43%
坐姿	28	9 903	10	2	66	0	57	38	0	6	18	13	96.79%
躺姿	90	99	9 115	2	5	0	4	0	0	0	9	6	97.70%
行走	2	2	0	11 028	559	0	43	23	34	29	31	0	93.85%
下楼	3	12	5	424	10 628	0	33	24	2	3	20	0	95.28%
上楼	0	0	0	0	0	11 732	0	0	47	15	52	14	98.92%
站—坐	1	2	1	46	23	3	604	7	5	4	25	3	83.43%
坐—站	3	2	0	11	14	0	12	486	3	9	2	4	89.01%
坐—躺	0	0	0	4	2	8	4	0	552	0	195	1	72.06%
躺—坐	1	5	0	12	1	15	3	1	10	675	6	147	77.05%
站—躺	1	9	4	4	14	36	6	5	340	3	841	1	66.53%
躺—站	17	2	3	5	0	14	0	8	0	149	0	635	76.23%
召回率	98.63%	98.42%	97.79%	95.58%	93.75%	99.36%	78.44%	81.81%	55.59%	75.50%	70.08%	75.60%	95.28%

图 4 算法在训练集的精准率、召回率分布矩阵

姿态	站姿	坐姿	躺姿	行走	下楼	上楼	站—坐	坐—站	坐—躺	躺—坐	站—躺	躺—站	精准率
站姿	1 568	20	52	0	4	0	0	0	0	0	0	0	95.37%
坐姿	13	1 620	22	0	3	0	8	1	0	0	0	4	96.95%
躺姿	14	23	1 456	0	2	0	1	0	0	0	1	0	97.26%
行走	0	0	1	1 780	185	1	14	10	6	5	0	0	88.91%
下楼	0	6	1	105	1 849	1	4	3	5	0	4	0	93.48%
上楼	0	0	0	2	1	2 065	0	0	29	1	4	3	98.10%
站—坐	0	0	1	9	2	0	85	0	5	1	5	0	78.70%
坐—站	3	5	0	4	2	0	1	71	12	2	0	9	65.14%
坐—躺	0	0	0	0	1	1	0	0	50	0	20	0	69.44%
躺—坐	0	0	0	4	2	2	0	0	0	89	0	28	71.20%
站—躺	0	0	0	1	3	2	11	0	74	0	73	0	45.12%
躺—站	4	1	0	0	0	0	0	0	0	3	0	69	89.61%
召回率	97.88%	96.72%	94.98%	93.44%	90.02%	99.66%	68.55%	83.53%	27.62%	88.12%	68.22%	61.06%	93.27%

图5 算法在测试集的精准率、召回率分布矩阵

表2 模型性能对比

分类器	姿态						准确率
	坐姿	站姿	躺姿	行走	下楼	上楼	
MC-SVM	95.6	97.1	100	87.2	72.6	87.2	89.3%
MCHF-SVM	94.7	97.8	100	87.2	74.0	81.7	89.0%
CCLA	97.0	95.4	97.3	88.9	93.5	98.1	90.2%

4 结束语

针对传统的基于图像视频的姿态识别算法中存在的计算成本高、摄像盲区多、隐私易泄露问题，以卷积神经网络、卷积长短时记忆网络与 Attention 模型为基础，提出了一种基于手机加速度与陀螺仪数据的人体姿态识别算法。针对姿态识别技术进行了大量研究并对提出的 CCLA 模型算法进行了实验与评估，验证了 CCLA 算法性能表现更优。但在研究过程中也发现，基于可穿戴传感器的姿态识别技术仍处于起步阶段。当前所提出的各项方法中，诸多细节在研究中被忽略，导致其在实际应用中的性能表现仍有待提高，这将是下一步重点研究的方向。

参考文献:

- [1] VAROL G, LAPTEV I, SCHMID C. Long-term temporal convolutions for action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1510-1517.
- [2] HU J F, ZHENG W S, LAI J, et al. Jointly learning heterogeneous features for RGB-D activity recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2186-2200.
- [3] DONAHUE J, HENDRICKS L A, ROHRBACH M, et al. Long-term recurrent convolutional networks for visual recognition and description[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 39(4): 677-691
- [4] ZHANG L, WU X, LUO D. Human activity recognition with HMM-DNN model[C]//2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI&CC). IEEE, 2015: 192-197.
- [5] ANGUITA D, GHIO A, ONETO L, et al. Human activity recognition

- on smartphones using a multiclass hardware-friendly support vector machine[C]//International Workshop on Ambient Assisted Living. Springer, 2012: 216-223.
- [6] 李云鹤. 智能穿戴设备基于动态模板匹配算法的 3D 手势识别[J]. 物联网学报, 2019, 3(1): 97-105.
LI Y H. 3D gesture recognition based on dynamic template matching algorithm for intelligent wearable devices[J]. Chinese Journal on Internet of Things, 2019, 3(1): 97-105.
- [7] ALSHEIKH M A, SELIM A, NIYATO D, et al. Deep activity recognition models with triaxial accelerometers[C]//Proceedings of the 23rd ACM International Conference on Multimedia. ACM, 2015: 689-692.
- [8] ZENG M, YU T, WANG X, et al. Semi-supervised convolutional neural networks for human activity recognition[C]//2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017: 522-529.
- [9] LEE S M, YOON S M, CHO H. Human activity recognition from accelerometer data using convolutional neural network[C]//2017 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2017: 131-134.
- [10] JIANG W, YIN Z. Human activity recognition using wearable sensors by deep convolutional neural networks[C]//Proceedings of the 23rd ACM International Conference on Multimedia. ACM, 2015: 1307-1310.
- [11] YANG J, NGUYEN M N, SAN P P, et al. Deep convolutional neural networks on multichannel time series for human activity recognition[C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence. 2015: 3995-4001.
- [12] HAMMERLA N Y, HALLORAN S, PLOETZ T. Deep, convolutional, and recurrent models for human activity recognition using wearables[J]. Journal of Scientific Computing, 2016, 61(2): 454-476.
- [13] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting[C]//Advances in Neural Information Processing Systems. 2015: 802-810.
- [14] TAO D, WEN Y, HONG R. Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition[J]. IEEE Internet of Things Journal. 2016, 3(6): 1124-1134.
- [15] ORDÓÑEZ F, ROGGEN D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition[J]. Sensors, 2016, 16(1): 115-140.
- [16] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 379-389.
- [17] LAI J L, YI Y. Key frame extraction based on visual attention model[J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 114-125.
- [18] RAFFEL C, ELLIS D P W. Feed-forward networks with attention can solve some long-term memory problems[J]. arXiv, 2016: 1-6.
- [19] REYES-ORTIZ J L, ONETO L, SAMÀ A, et al. Transition-aware human activity recognition using smartphones[J]. Neurocomputing, 2016, 171: 754-767.

[作者简介]



金驰(1997-), 男, 辽宁沈阳人, 吉林大学硕士生, 主要研究方向为无线体域网和人机交互技术。



李志军(1971-), 男, 吉林长春人, 吉林大学高级工程师, 主要研究方向为无线资源管理技术研究。



孙大洋(1979-), 男, 吉林长春人, 吉林大学副教授, 主要研究方向为无线传感器网络、物联网通信与节能技术、智能终端软件开发。



胡封晔(1974-), 男, 河南原阳人, 吉林大学教授、博士生导师, 主要研究方向为无线体域网、认知无线电、无线能量和信息同传技术、空时通信技术。