



# 混合 CNN-HMM 的人体动作识别方法

张 振<sup>1,2</sup>, 张师榕<sup>1</sup>, 赵转哲<sup>1,2\*</sup>, 刘永明<sup>1,2</sup>, 阚延鹏<sup>1</sup>, 涂志健<sup>2</sup>

(1. 安徽工程大学机械工程学院 安徽 芜湖 241000; 2. 芜湖赛宝机器人产业技术研究院有限公司 安徽 芜湖 241000)

**【摘要】**针对当前人体动作识别算法检测精度不佳和实验场景多样性的问题,提出了一种混合卷积神经网络-隐马尔可夫模型(CNN-HMM)的人体动作识别方法。建立了抬腿、深蹲和仰卧臀桥 3 组分别包含 1 个标准动作姿态和 5 个非标准动作姿态的人体康复训练动作模型库,结合可穿戴式惯性动作捕捉系统 PN2.0 获取实验数据。最后从准确率、灵敏度和特异性 3 个方面进行性能评估。实验结果表明,该方法能够以较高识别率将 6 种不同动作姿态区分开,其平均识别准确率为 97.00%,相较于单一 CNN 方法提高了 5.78%。

**关键词** 卷积神经网络; 隐马尔可夫模型; 人体动作识别; 模式识别与智能系统; 感知神经元  
中图分类号 TP391.4 文献标志码 A doi:10.12178/1001-0548.2021326

## Human Motion Recognition Method Using Hybrid CNN-HMM

ZHANG Zhen<sup>1,2</sup>, ZHANG Shirong<sup>1</sup>, ZHAO Zhuanzhe<sup>1,2\*</sup>, LIU Yongming<sup>1,2</sup>, KAN Yanpeng<sup>1</sup>, and TU Zhijian<sup>2</sup>

(1. School of Mechanical Engineering, Anhui Polytechnic University Wuhu Anhui 241000;

2. Wuhu Ceprei Robot Technology Research Co., Ltd. Wuhu Anhui 241000)

**Abstract** Aiming at the problems of poor detection accuracy of current human motion recognition algorithms and the diversity of experimental scenes, a new human motion recognition method based on hybrid convolutional neural network-hidden Markov model (CNN-HMM) is proposed. In order to verify the effectiveness of the method, we establish three sets of human rehabilitation training motion models including one standard motion posture and five non-standard motion postures for leg-lifting, squat and hip bridge, respectively. The experimental data are obtained by the wearable inertial motion capture system, Perception Neuron 2.0 (PN2.0). Finally, the performance of the proposed method is evaluated in terms of accuracy, sensitivity and specificity. Three groups of the experimental results show that the proposed method can distinguish the six different motion gestures with a high average recognition rate of 97.00%, which is 5.78% higher than the single CNN method.

**Key words** convolutional neural networks; hidden Markov model; human motion recognition; pattern recognition and intelligent system; perception neuron

人体动作识别技术作为人工智能及计算机视觉领域的一个重要研究方向,成为国内外研究人员相继探讨的热点话题,目前已成功应用于人机交互、视频监控、体育训练和康复医疗等诸多领域<sup>[1-4]</sup>。

近年来,深度学习因其出色的解决问题的能力而备受关注<sup>[5-6]</sup>。研究人员将深度学习应用到人体动作识别领域,在人体动作识别的研究方法和建模算法方面取得了切实有效的进步。如文献[7]提出了一种新的跌倒识别方法,该方法基于 OpenPose 深度卷积网络获取人体倾斜姿态动态特征,根据支持

向量机完成跌倒行为二分类,以人体下降姿态动态特征的阈值判断排除混淆性较大的非跌倒行为,在人体动作数据集上取得了较高的准确率。文献[8]针对人体动作识别精度不高的问题,在克服光照影响方面,对人体关节数据进行预处理;为了解决动作特征描述的问题,对人体动作数据进行针对性编码。文献[9]结合离散余弦变换(discrete cosine transform, DCT)和主成分分析法(principal component analysis, PCA),利用改进优化后的支持向量机对 5 组不同类型的人体小样本动作数据进行识别。文

收稿日期: 2021-11-08; 修回日期: 2021-12-03

基金项目: 安徽省重点研究与开发计划(202004b11020006); 安徽省自然科学基金(2108085QF278); 安徽省留学人员创新项目择优资助计划(2020LCX013)

作者简介: 张振(1991-),男,博士,主要从事机器学习与模式识别方面的研究。

\*通信作者: 赵转哲, Email: zhuanzhe727@ahpu.edu.cn

献 [10] 在深度神经网络 (deep neural network, DNN) 的基础上, 利用 VGG19 提取 DNN 特征, 根据水平和垂直梯度以及垂直方向信息计算特征, 利用参数选择最佳特征进行识别。文献 [11] 在信道状态信息 (channel state information, CSI) 的基础上进行延展, 基于室内 wifi 信号提出了一种包含离线和在线两个阶段的 CSI-HC 人体动作识别方案。文献 [12] 提出了一种新的双层条件随机场 (DL-CRFs) 的人体动作识别模型, 并在两个 RGB-D 视频序列数据集上对其有效性进行验证。文献 [13] 基于视觉数据集开发了羽毛球动作自动识别系统, 利用 AlexNet 网络模型在 fc8 层提取训练数据集特征, 得到了 82.00% 的识别精度。

上述文献提出的人体动作识别方法虽然取得了较好的成效并应用于多个场景, 但多是针对视频和图像进行处理, 亦或是采用光学式动作捕捉系统, 运动过程中易受到强光、遮挡等环境影响, 导致数据缺失, 将其应用于人体动作识别领域仍然具有一定的局限性<sup>[14-16]</sup>。与此同时, 可穿戴式惯性传感器体积逐渐变小, 重量逐渐减轻, 测量精度逐渐提高, 相较于光学式动作捕捉系统, 其价格也相对较低<sup>[17-19]</sup>。因此, 将惯性动作捕捉系统应用于人体动作识别领域, 具有较高的研究价值。

为此, 本文综合卷积神经网络 (convolutional neural network, CNN) 在特征提取和数据分类领域优异表现<sup>[20-21]</sup> 以及隐马尔可夫模型 (hidden Markov model, HMM) 在时间建模和动态分类领域精准有效<sup>[22-23]</sup> 的特点, 提出了一种混合卷积神经网络-隐马尔可夫模型 (CNN-HMM) 的深度学习算法, 结合可穿戴式惯性动作捕捉系统 Perception Neuron 2.0 (PN2.0) 对人体动作进行识别。通过 3 组人体动作识别实验对本文方法进行了验证。

## 1 混合 CNN-HMM

本文混合 CNN-HMM 方案如图 1 所示, 该方法将 CNN 的输出概率矩阵嵌入 HMM 的观测概率矩阵中建模。

将输入数据流表示为  $X = X_1, X_2, \dots, X_T$ , 选择 CNN 输出的最大后验概率  $P(S|X)$ , 根据贝叶斯决策规则得到模型  $X$  的最佳拟合序列  $\tilde{S}$ :

$$\tilde{S} = \arg \max_S P(S|X) = \arg \max_S \frac{P(X|S)P(S)}{P(X)} \quad (1)$$

式中,  $P(X|S)$  为类条件概率;  $P(S)$  为类先验概率;

$P(X)$  为证据因子。对于样本  $X$ ,  $P(X)$  与  $S$  无关, 于是后验概率  $P(S|X)$  表示为类先验概率  $P(S)$  和类条件概率密度  $P(X|S)$  的乘积, 可得:

$$\tilde{S} = \arg \max_S P(X|S)P(S) \quad (2)$$

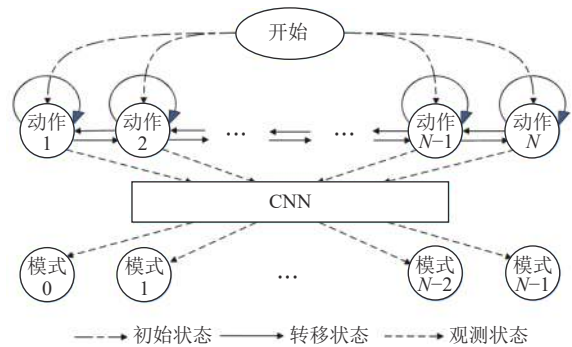


图 1 混合 CNN-HMM 结构图

根据输入的时间变化, 使用 HMM 建模, 采用一阶马尔可夫假设, 将  $P(X|S)$  最大化, 可得:

$$P(X|S) \approx \max_{S_1 \dots S_T} \prod_{t=1}^T P(s_t|s_{t-1})P(x_t|s_t) \quad (3)$$

式中,  $P(s_t|s_{t-1})$  为 HMM 中不同动作之间的转移概率;  $P(x_t|s_t)$  为类条件概率函数。根据贝叶斯规则将  $P(x_t|s_t)$  转换为似然度, 可得:

$$P(x_t|s_t) = \frac{P(s_t|x_t)P(x_t)}{P(s_t)} \quad (4)$$

删除常量  $P(x_t)$ , 根据式 (1)~式 (4) 可得混合 CNN-HMM 方法的最终动作预测序列  $\tilde{S}$ :

$$\tilde{S} = \arg \max_S \left\{ \max_{S_1 \dots S_T} \prod_{t=1}^T P(s_t|s_{t-1}) \frac{P(s_t|x_t)}{P(s_t)} P(S) \right\} \quad (5)$$

## 2 系统概述

该混合 CNN-HMM 人体动作识别方法分为训练阶段和测试阶段两部分, 其目的是在训练数据集上建立模型以及在测试数据集上评估性能, 从而得到一个具有泛化能力的人体动作识别算法模型。

### 2.1 训练阶段

为了学习如何从输入中预测输出, 在训练阶段, 使用带有标注的训练数据集寻找最优参数, 建立最优训练模型并预测输出, 其流程如图 2 所示, 具体步骤如下。

1) 数据预处理: 在预处理过程中先对训练数据

集进行数据清洗, 随后将其按照一定比例缩放在特定区域, 根据实验动作将训练数据集分割为连续的数据段 (称为窗口)。

2) 训练 CNN: 选择 CNN 作为基线分类算法, 将分割后的数据段带入混合 CNN-HMM 算法中进行训练。

3) 生成 HMM 概率: 为了确保时间序列的平滑性, 将 CNN 分类结果与 HMM 相结合, 并在此步骤中生成 HMM 的初始概率和转移概率。

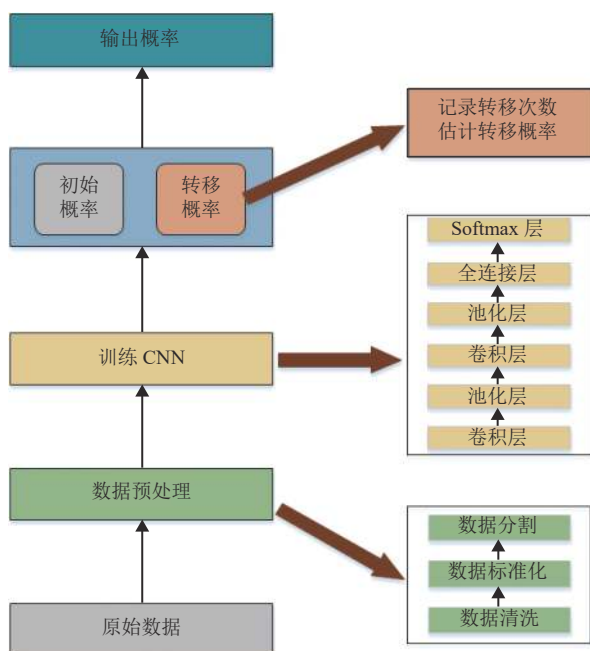


图 2 混合 CNN-HMM 训练阶段

## 2.2 测试阶段

为了评估模型性能的好坏以及测试最终的泛化能力, 在系统的测试阶段, 将从未使用过的测试数据集应用于训练阶段建立的预测模型, 从而完成动作序列的预测, 其流程如图 3 所示, 具体步骤如下。

1) 数据预处理: 实验数据被分割为数据段, 称为数据窗口, 这些数据标记为测试数据。

2) CNN 分类: 利用 CNN 模型对预处理后的数据进行分类, 输出总和为 1 的后验概率分布, 即每个动作模式对应的概率。

3) 使用 HMM 重新分类: 为了充分利用实验中的时间信息, 将步骤 2) 中 CNN 输出的后验概率分布结果嵌入 HMM 的观测概率矩阵, 使用 Viterbi 算法对其重新分类, 从而得到最佳的动作序列。

4) 性能评估: 根据不同的评价指标测试算法性能。

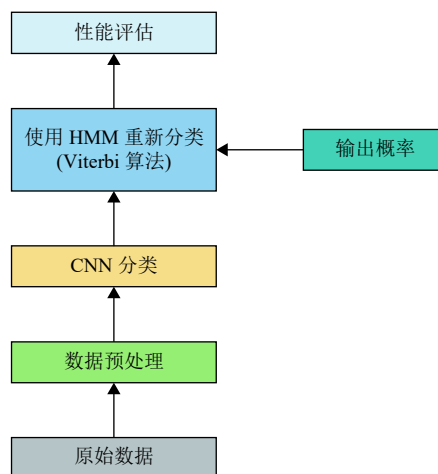


图 3 混合 CNN-HMM 测试阶段

## 3 实验

为了验证所提方法的可行性, 本文制定了一种包含抬腿 (实验 I)、深蹲 (实验 II) 和仰卧臀桥 (实验 III) 这 3 种常见的人体康复训练动作识别方案。在每组实验中, 分别建立一个包含 6 种不同动作姿态的动作模型库, 通过诺亦腾公司自主研发的可穿戴式惯性动作捕捉系统 PN2.0 收集实验数据, 使用预处理后的数据集训练模型, 最后对算法性能做出评价分析。图 4 为获取实验数据的可穿戴式惯性动作捕捉系统, 分别通过 USB 接口和 HUB 连接计算机以及发送由感知神经元节点生成的实验数据。在每组实验中, 实验人员将 6 种不同动作分别重复 50 次, 共收集 135 000 帧的训练数据, 同时, 使用摄像机记录整个实验过程。



图 4 可穿戴式惯性动作捕捉系统 Perception Neuron 2.0

### 3.1 建立动作模型库

在每组实验中, 6 种不同动作模型分别记为 PAT0、PAT1、PAT2、PAT3、PAT4 和 PAT5, 其中, PAT0 为标准动作模型, PAT1~PAT5 为非标

准动作模型。每个动作模型的详细设置情况如表 1 所示, 其对应的具体动作如图 5 所示。

表 1 每组实验中动作模式的设定

模式	动作模型
实验 I	PAT0 左右不摆动的情况下向前抬腿45°
	PAT1 左右不摆动的情况下向前抬腿20°
	PAT2 向左偏离10°
	PAT3 向左偏离20°
	PAT4 向右偏离10°
实验 II	PAT0 双脚平行, 与肩同宽, 蹲下时双腿平行于地面
	PAT1 下蹲不充分, 大腿与地面不平行
	PAT2 膝盖内扣
	PAT3 双腿分开
	PAT4 膝盖超过脚尖
实验 III	PAT0 双脚与髌同宽, 肩、髌和膝呈直线, 小腿与地面垂直
	PAT1 脚跟站位靠前, 损失运动范围
	PAT2 伸腕幅度不足
	PAT3 膝盖内敛
	PAT4 双腿分开
PAT5 双脚前后不平行	

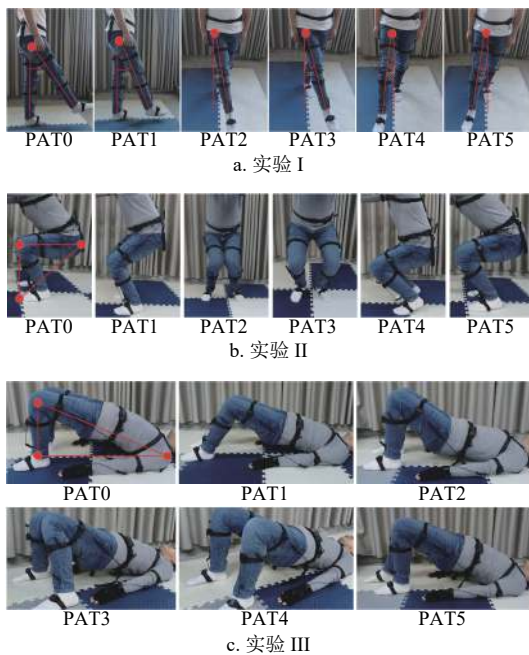


图 5 动作模型库的建立

### 3.2 实验数据采集

实验中, 通过感知神经元产生并由 HUB 实时输出 Calc 类型数据, 该类型包括全局坐标系中的位移、速度、加速度以及陀螺仪数据。其数据类型分别为, 位置坐标: P-x、P-y、P-z; 速度坐标: V-x、V-y、V-z; 加速度坐标: A-x、A-y、A-z; 陀螺

仪坐标: G-x、G-y、G-z。

图 6 表示实验 I 在标准动作模式 (PAT0) 下一个感知神经元生成的部分时间序列数据样本示意图。

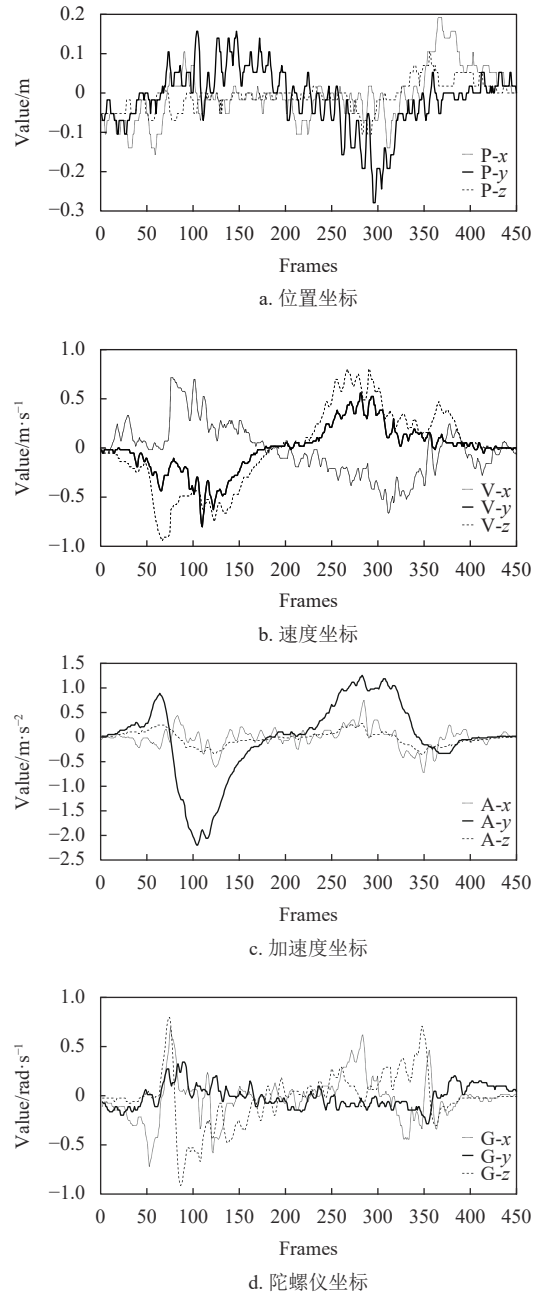


图 6 标准动作的时间序列数据样本

### 3.3 数据预处理

为了进一步提高动作识别效果, 消除训练数据集中的异常值及噪音, 先对实验数据进行清洗, 再根据 Z-score 方法对其标准化:

$$Z = \frac{X - \mu}{\sigma} \quad (6)$$

式中,  $X$  为数据流中单个特征值;  $\mu$  为  $X$  的平均



值;  $\sigma$  为  $X$  的标准差。

实验数据的分割大小取决于实验人员一次运动产生的数据集规模的大小, 图7展示了一个滑动窗口沿着输入数据将其分割为若干相同大小的数据段的过程, 从而将输入数据划分为若干带有标签的运动模式。在每组实验中, 一次运动大约需要3~4 s, PN2.0设备以120 fps的速度向计算机实时输出数据, 因此将滑动窗口大小设为450。

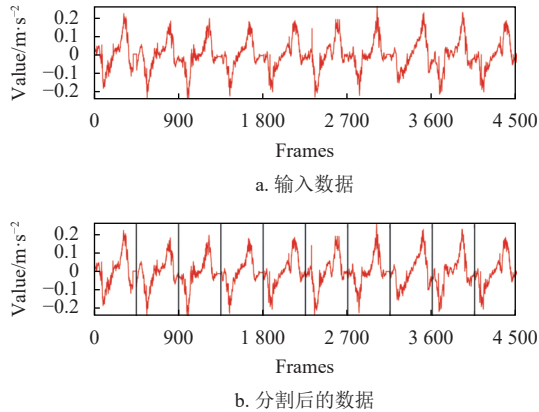


图7 将输入数据分割为固定大小的数据段

### 3.4 建立 CNN-HMM 模型

混合 CNN-HMM 方法共分为 CNN 分类和 HMM 预测两个部分。在 CNN 分类部分, 将预处理后的实验数据集输入在包含卷积层、池化层等结构的 CNN 框架中训练。为了有效避免过拟合现象, 使用 Dropout 方法随机删除神经元, 另外在 CNN 最后增加 Softmax 函数, 将 CNN 的输出表示为总和为 1 的后验概率, 以便 CNN 和 HMM 混合方法模型的建立。表2列出了 Adam 算法下本文 CNN 的具体参数设置。

表2 CNN 结构参数

参数	类型
数据输入	96×135 000
窗口尺寸	96×450
激活函数	Relu
滤波器类型	Valid
隐含层节点数	1000
卷积层 I 内核	12
池化层 I 内核	10
卷积层 II 内核	6
池化层 II 内核	3
Dropout 概率/%	50
优化算法	Adam 算法

在 HMM 预测部分, 将 HMM 表示为  $\lambda = \{\boldsymbol{\pi}, \mathbf{A}$ ,

$\mathbf{B}\}$ , 其中,  $\boldsymbol{\pi} = \{\pi_1, \pi_2, \dots, \pi_q\}$  表示初始概率向量, 由于在每组实验中共建立 6 种不同的动作姿态, 且每次动作随机产生, 因此初始状态向量中每个动作产生概率均为 1/6;  $\mathbf{A}$  为每组实验中 6 种不同动作间的状态转移概率矩阵, 通过式 (7) 计算矩阵  $\mathbf{A}$  中不同元素的转移概率:

$$a_{ij} = k_{ij} / \sum_{j=1}^N k_{ij} \quad (7)$$

式中,  $k_{ij}$  为状态  $i$  到状态  $j$  的转移次数;  $N$  为不同状态的数量总和。 $\mathbf{A}$  的计算结果如表3所示, 其中, 行表示当前动作, 列表示下一动作, 如在实验 II 中, 在  $t$  时刻状态序列为 PAT0, 在时刻  $t+1$  为 PAT2 的概率为 16.00%;  $\mathbf{B}$  表示观测概率矩阵, 由 CNN 分类器得到的后验概率嵌入。

表3 HMM 的转移概率矩阵  $\mathbf{A}$

		PAT0	PAT1	PAT2	PAT3	PAT4	PAT5	%
实验 I	PAT0	74.00	4.00	12.00	4.00	0	6.00	
	PAT1	4.00	74.00	0	6.00	8.00	8.00	
	PAT2	4.00	6.00	74.00	6.00	6.00	4.00	
	PAT3	8.00	10.00	2.00	70.00	4.00	6.00	
	PAT4	4.08	6.12	8.16	10.20	71.43	0	
	PAT5	6.00	0	4.00	2.00	12.00	76.00	
实验 II	PAT0	62.00	6.00	16.00	6.00	6.00	4.00	
	PAT1	6.00	72.00	2.00	10.00	4.00	6.00	
	PAT2	4.08	2.04	75.51	6.12	4.08	8.16	
	PAT3	6.00	6.00	4.00	72.00	4.00	8.00	
	PAT4	14.00	8.00	2.00	0	72.00	4.00	
	PAT5	8.00	6.00	2.00	6.00	8.00	70.00	
实验 III	PAT0	74.79	2.99	10.99	4.12	2.99	4.12	
	PAT1	3.74	70.60	1.88	7.74	8.30	7.74	
	PAT2	3.74	5.46	73.80	2.01	6.63	8.36	
	PAT3	1.70	10.24	4.52	74.50	4.52	4.52	
	PAT4	5.04	12.02	12.02	2.70	68.20	0.02	
	PAT5	4.73	6.22	9.20	3.23	7.71	68.91	

## 4 实验结果分析

采用准确率 (Accuracy)、灵敏度 (Sensitivity) 和特异性 (Specificity) 3 种评价指标验证算法性能, 计算公式分别为:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (8)$$

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

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (10)$$

式中, TP、FP、FN 和 TN 分别表示真阳性、假阳性、假阴性和真阴性。在本研究中, TP 表示一个特定模式在数据集中被正确识别的数量; FP 表示在数据集中将其他模式识别为该特定模式的数量;

FN 表示在数据集中将该特定模式错误识别为其他模式的数量; TN 表示在数据集中将其他模式正确识别的数量。表 4 为单一 CNN 方法和混合 CNN-HMM 方法在 3 种评价指标上的计算结果。

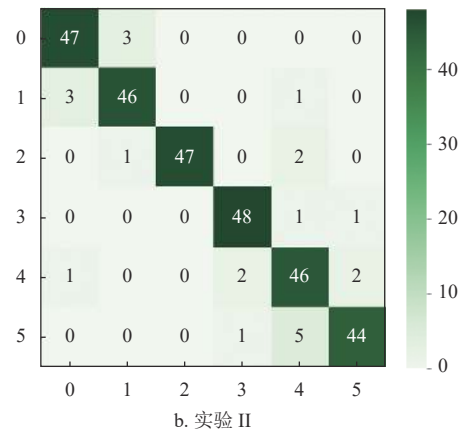
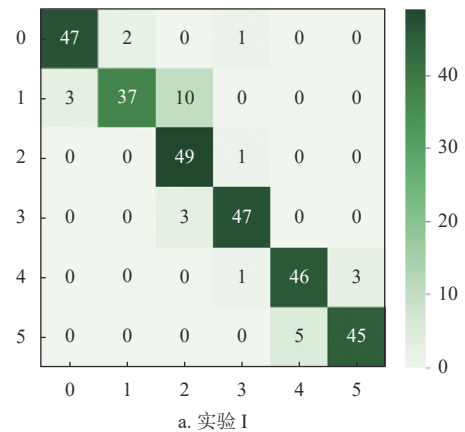
表 4 单一 CNN 与混合 CNN-HMM 方法性能对比

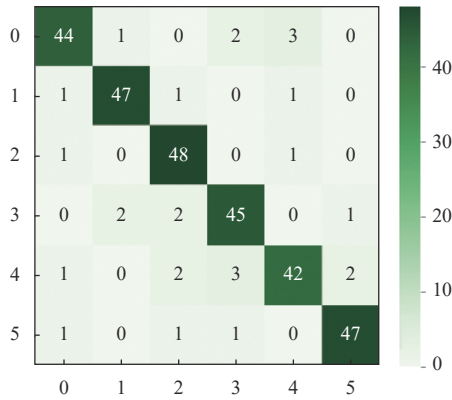
%

		Accuracy		Sensitivity		Specificity	
		CNN	混合CNN-HMM	CNN	混合CNN-HMM	CNN	混合CNN-HMM
实验 I	PAT0	98.00	100.00	94.00	100.00	98.80	100.00
	PAT1	95.00	96.67	74.00	80.00	99.20	100.00
	PAT2	95.33	96.67	98.00	100.00	94.80	96.00
	PAT3	98.00	99.67	94.00	100.00	98.80	99.60
	PAT4	97.00	99.33	92.00	96.00	98.00	100.00
	PAT5	97.33	99.67	90.00	100.00	98.80	99.60
实验 II	PAT0	97.67	99.33	94.00	98.00	98.40	99.60
	PAT1	97.33	99.00	92.00	96.08	98.40	99.20
	PAT2	99.00	100.00	94.00	100.00	100.00	100.00
	PAT3	98.33	99.67	96.00	98.04	98.80	99.60
	PAT4	95.33	98.33	90.00	95.92	96.40	99.20
	PAT5	97.00	99.00	88.00	97.96	98.80	99.60
实验 III	PAT0	96.67	99.20	88.00	94.00	98.40	99.20
	PAT1	98.00	99.60	94.00	98.00	98.80	99.20
	PAT2	97.33	99.60	96.00	98.00	97.60	99.60
	PAT3	96.33	99.60	90.00	96.00	97.60	99.60
	PAT4	95.67	99.60	84.00	98.00	98.00	99.60
	PAT5	98.00	99.60	94.00	100.00	98.80	99.60

由表 4 可以看出, 单一 CNN 和混合 CNN-HMM 两种方法在 3 种评价指标上均取得了较好的实验结果, 这表明两种方法均可成功运用到人体动作识别领域。另外, 从表 4 中不难看出, 在 3 组人体动作识别实验中, 相较于单一 CNN 方法, 使用混合 CNN-HMM 方法在多个评价指标上均有不同程度地提高, 证明了本文方法的优越性。

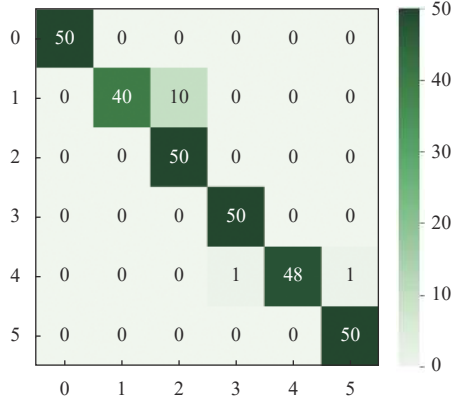
图 8 和图 9 分别为使用单一 CNN 和混合 CNN-HMM 两种方法在 3 组实验中产生的混淆矩阵, 其中, 矩阵的行表示真实动作序列, 列表示算法识别得到的动作序列。由混淆矩阵可以得出, 使用单一 CNN 方法在 3 组实验中对 6 种动作分别成功识别 271 次、277 次和 273 次, 其识别准确率分别为 90.33%、92.33% 和 91.00%; 使用混合 CNN-HMM 方法在 3 组实验中对 6 种动作分别成功识别 288 次、293 次和 292 次, 其识别准确率分别为 96.00%、97.67% 和 97.33%, 相较于单一 CNN 方法, 其平均识别准确率提高了 5.78%。特别地, 实验 I 中, 对于 PAT1, CNN 成功识别 37 次, 混合 CNN-HMM 成功识别 40 次, 与实际序列差距较大, 造成这种现象的原因可能是数据集间差异较小, 算法难以将其正确识别。



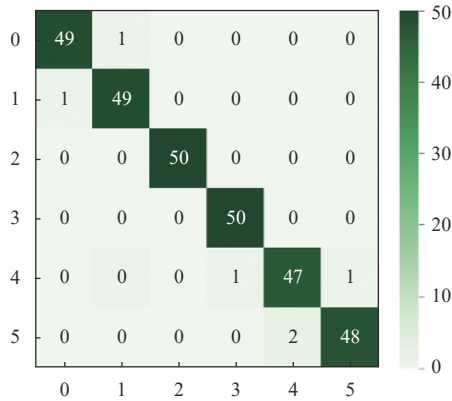


c. 实验 III

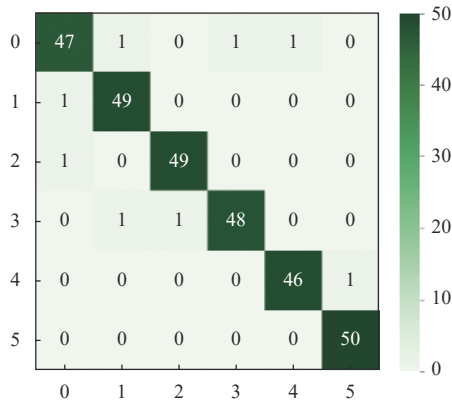
图 8 单一 CNN 方法性能混淆矩阵



a. 实验 I



b. 实验 II



c. 实验 III

图 9 混合 CNN-HMM 方法性能混淆矩阵

为了更加直观地观察输入动作序列的分布情况以及两种方法对于 3 组动作识别实验的结果差异, 图 10 展示了 3 组实验中实际输入动作序列与使用单一 CNN 和混合 CNN-HMM 两种方法识别得到的动作序列对比图。由图 10 可以看出, 相较于单一 CNN 方法, 使用混合 CNN-HMM 方法在 3 组实验中的识别结果与实际输入的动作序列差距均更为微小, 识别精度更高, 从另一个角度验证了本文方法在人体动作识别方案中的可行性及优越性。

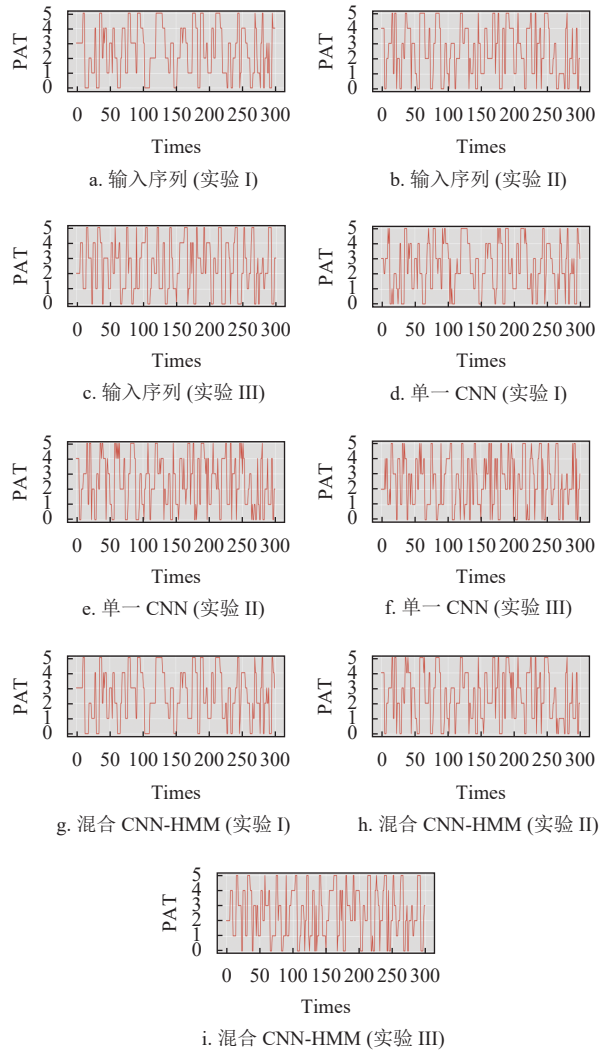


图 10 输入序列与使用单一 CNN、混合 CNN-HMM 方法识别结果对比

## 5 结束语

本文提出了一种混合卷积神经网络-隐马尔可夫模型 (CNN-HMM) 的深度学习的方法, 并成功应用于人体动作识别领域。实验结果表明了本文方法能以较高准确率将 6 种不同动作姿态区分开, 相较于单一 CNN 方法, 平均识别准确率提高了 5.78%。本文方法为深度学习与人体动作识别领域提供了新

的研究思路, 也为康复运动与体育训练领域提供了新的研究参考。但在实验过程中, 感知神经元可能会遇到磁干扰现象, 因此本实验应当在开阔无辐射的场地中进行。在今后的研究中, 将考虑设计人机交互界面从而得到更佳的体验效果, 同时也可尝试将其应用于其他相关领域。

### 参 考 文 献

- [1] LI N, CHENG X, GUO H, et al. Recognizing human interactions by genetic algorithm-based random forest spatio-temporal correlation[J]. *Pattern Analysis & Applications*, 2016, 19(1): 267-282.
- [2] ZHANG X. Application of human motion recognition utilizing deep learning and smart wearable device in sports[J]. *International Journal of System Assurance Engineering and Management*, 2021, 12(4): 835-843.
- [3] PAN Z G, LI C. Robust basketball sports recognition by leveraging motion block estimation[J]. *Signal Processing: Image Communication*, 2020, 83: 15784.
- [4] 邢蒙蒙, 魏国辉, 刘静, 等. 多模态人体动作表示识别及其正骨康复训练应用综述[J]. *生物医学工程学杂志*, 2020, 37(1): 180-184,190.  
XING M M, WEI G H, LIU J, et al. A review on multi-modal human motion representation recognition and its application in orthopedic rehabilitation training[J]. *Journal of Biomedical Engineering*, 2020, 37(1): 180-184,190.
- [5] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(1): 221-231.
- [6] HABIB M, FARIS M, QADDOURA R, et al. Toward an automatic quality assessment of voice-based telemedicine consultations: A deep learning approach[J]. *Sensors*, 2021, 21(9): 1-26.
- [7] 韩锐, 黄泽帆. 基于人体姿态动态特征的跌倒行为识别方法[J]. *湖南大学学报(自然科学版)*, 2020, 47(12): 69-76.  
HAN K, HUANG Z F. Falling behavior recognition method based on dynamic characteristics of human body posture[J]. *Journal of Hunan University (Natural Sciences)*, 2020, 47(12): 69-76.
- [8] 胡青松, 张亮, 丁娟, 等. 人体动作数据编码与 CNN 精确识别[J]. *电子科技大学学报*, 2020, 49(3): 473-480.  
HU Q S, ZHANG L, DING J, et al. Data encoding and CNN accurate recognition of human body motion[J]. *Journal of University of Electronic Science and Technology of China*, 2020, 49(3): 473-480.
- [9] 蒋留兵, 周小龙, 车俐. 基于无载波超宽带雷达的小样本人体动作识别[J]. *电子学报*, 2020, 48(3): 602-615.  
JIANG L B, ZHOU X L, CHE L. Few-shot learning for human motion recognition based on carrier-free UWB radar[J]. *Acta Electronica Sinica*, 2020, 48(3): 602-615.
- [10] KHAN M A, JAVED K, SABA T. Human action recognition using fusion of multiview and deep features: An application to video surveillance[J]. *Multimedia Tools and Applications*, 2020, DOI: 10.1007/s11042-020-08806-9.
- [11] HAO Z J, DUAN Y, DANG X C, et al. CSI-HC: A wifi-based indoor complex human motion recognition method[J]. *Mobile Information Systems*, 2020, DOI: 10.1155/2020/3185416.
- [12] LIU T L, DONG X D, WANG Y Z, et al. Double-layer conditional random fields model for human action recognition[J]. *Signal Processing: Image Communication*, 2020, DOI: 10.1016/j.image.2019.115672.
- [13] RAHMAD N A, ASARI M A. The new convolutional neural network (CNN) local feature extractor for automated badminton action recognition on vision based data[J]. *Journal of Physics Conference Series*, 2020, 1529(2): 022021.
- [14] SULTANI W, SHAN M. Human action recognition in drone videos using a few aerial training examples[J]. *Computer Vision and Image Understanding*, 2021, 206: 103186.
- [15] LI G, LI C Y. Learning skeleton information for human action analysis using kinect[J]. *Signal Processing: Image Communication*, 2020, 84: 155814.
- [16] ZHANG J X, HU H F, LU X L. Moving foreground-aware visual attention and key volume mining for human action recognition[J]. *ACM Transactions on Multimedia Computing Communications and Applications*, 2019, 15(3): 1-16.
- [17] SERS R, FORRESTER S, MOSS E, et al. Validity of the perception neuron inertial motion capture system for upper body motion analysis[J]. *Measurement*, 2020, 149: 1-9.
- [18] 张鏊豪, 何百岳, 杨旭升, 等. 基于可穿戴式惯性传感器的人体运动跟踪方法综述[J]. *自动化学报*, 2019, 45(8): 1439-1454.  
ZHANG J H, HE B Y, YANG X S, et al. A review on wearable inertial sensor based human motion tracking[J]. *Acta Automatica Sinica*, 2019, 45(8): 1439-1454.
- [19] SENA J, BARRETO J, CAETANO C, et al. Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble[J]. *Neurocomputing*, 2021, 444: 226-243.
- [20] ISKANDAR D N F A, LATIF G, ALGHAZO J, et al. Brain MR image classification for glioma tumor detection using deep convolutional neural network features[J]. *Current Medical Imaging*, 2021, 17(1): 56-63.
- [21] 张祥东, 王腾军, 朱劲俊, 等. 基于扩张卷积注意力神经网络的高光谱图像分类[J]. *光学学报*, 2021, 41(3): 49-59.  
ZHANG X D, WANG T J, ZHU S J, et al. Hyperspectral image classification based on dilated convolutional attention neural network[J]. *Acta Optica Sinica*, 2021, 41(3): 49-59.
- [22] 周昱, 张杰, 沈安文. 基于改进 3D 马尔可夫模型的动态车辆检测[J]. *华中科技大学学报(自然科学版)*, 2011, 39(9): 48-52.  
ZHOU L, ZHANG J, SHEN A W. Dynamic identification of vehicles using improved 3D-Markov model[J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2011, 39(9): 48-52.
- [23] WEN R, WANG Q, LI Z B. Human hand movement recognition using infinite hidden markov model based sEMG classification[J]. *Biomedical Signal Processing and Control*, 2021, 68: 102592.