

基于长短期记忆网络的红外人体步态识别方法研究

梅建华¹, 云利军^{1,2*}, 朱小鹏¹

¹云南师范大学信息学院, 云南 昆明 650500;

²云南省光电信息技术重点实验室, 云南 昆明 650500

摘要 步态识别作为一种非接触式远距离的生物特征识别技术,其任务是根据远距离行人的行走模式来实现身份识别。针对使用卷积神经网络(CNN)构建模型进行红外人体步态识别存在识别效果不佳的问题,利用长短期记忆网络(LSTM),根据人体身高比例遮挡穿着部分后的图像,让网络集中注意力提取腿部的变化特征以及每个红外人体步态周期的时间维度特征,从而构建了一种新的步态识别模型。在中国科学院提供的CASIA C红外步态数据库中,对穿着服饰部分进行预遮挡处理后的数据进行了实验测试,所提模型的识别准确率优于卷积神经网络模型。实验结果表明,在部分特征缺失情况下,利用长短期记忆网络进行步态识别能有效提高识别准确率。

关键词 成像系统; 步态识别; 卷积神经网络; 长短期记忆网络; 红外人体图像

中图分类号 TP181

文献标志码 A

doi: 10.3788/LOP202259.0811005

Infrared Human Gait Recognition Method Based on Long and Short Term Memory Network

Mei Jianhua¹, Yun Lijun^{1,2*}, Zhu Xiaopeng¹

¹College of Information, Yunnan Normal University, Kunming, Yunnan 650500, China;

²Yunnan Provincial Key Laboratory of Optoelectronic Information Technology, Kunming, Yunnan 650500, China

Abstract Gait recognition is a type of noncontact remote biometric recognition technology used for identity recognition based on the walking patterns of distant pedestrians. Aiming at the problem of poor recognition effect in infrared human gait recognition using convolutional neural network (CNN), the long and short term memory network (LSTM) is used to cover the image after wearing according to the proportion of human height, so that the network can focus on extracting the change characteristics of legs and the time dimension characteristics of each infrared human gait cycle. Therefore, a new gait recognition model is developed. In the CASIA C infrared gait database provided by the Chinese Academy of Sciences, the experimental test was carried out on the data after the preocclusion processing of the wearing part, and recognition accuracy of the proposed model was higher than that of convolutional neural network model. The experimental results indicated that using LSTM for gait recognition considerably enhanced the recognition accuracy when some features were unavailable.

Key words imaging systems; gait recognition; convolutional neural network; long and short term memory network; infrared body image

1 引言

步态识别的主要任务^[1-2]是通过视频图像实现

对远距离人群的身份识别,由于其在智能视觉安全和监控系统中应用的潜力,引起了研究者的兴趣。步态识别^[3]作为一种生物特征识别技术,与人

收稿日期: 2021-03-22; 修回日期: 2021-04-19; 录用日期: 2021-04-28

基金项目: 云南省应用基础研究计划重点项目(2018FA033)、云南师范大学研究生科研创新基金(YJSJJ21-B77)

通信作者: *yunlijun@ynnu.edu.cn

脸、指纹、虹膜等生物特征识别技术相比较具有独特的优势,因为步态识别可以在无需受试者配合情况下从低分辨率的视频序列中识别受试者。在实际场景中,服装^[4]、行走速度^[5-6]、携带条件^[7]和摄像机视点^[8]等因素会导致采集的步态图像发生显著变化,从而影响步态识别的准确率。

深度学习^[9]提取和学习抽象信息的能力突出,可以完全自动化地构建多层特征,由低层次特征来构建高层次特征。近些年,卷积神经网络(CNN)作为深度学习的一种,与传统的机器学习方法相比,使用CNN对图像进行特征提取和分类可以构造一种端到端的学习方式,特征提取简单高效、分类效果较好。因此,基于图像识别方式的步态识别领域逐渐引入CNN。与传统的神经网络相比,CNN通过共享权值以及池化操作减少了网络的参数量,通过卷积实现特征提取,通过分类器完成分类任务。Shiraga等^[10]提出了GEINet,GEINet是一个8层的CNN,其中前6层是卷积层、池化层和归一化层的2个连续重复,最后2层是2个全连接层,GEINet以步态能量图作为网络的输入,在OUISIR数据集上的平均rank 1准确率达到91.6%。虽然GEINet获得了较高的识别准确率,但是使用的数据集都是在较为理想情况下采集的。在实际应用中人体步态受穿着因素的影响较大,利用该网络在同一个人变换穿着的情况下,对单张图像进行识别的错误率较高。

为了减小背包以及服饰对人体步态识别的影

响,本文根据人体各个部分与身高的比例^[11]对图像进行部分遮挡处理,使用CNN对处理前后的步态图像进行特征提取和分类,对比实验结果表明,遮挡处理后的识别准确率比处理前有所降低。由此可知,在使用遮挡处理方式降低背包和服饰对步态识别影响的同时,丢失了部分空间信息,导致CNN提取的步态特征的类型差异有所降低。为此,本文提出了一种使用长短期记忆网络(LSTM)对经遮挡处理后的步态周期图像进行步态时间特征提取及分类识别的方法,并在预处理后的中国科学院CASIA C红外步态数据集进行了实验,所提方法的识别准确率较高。实验结果表明,所提方法降低了遮挡处理对识别准确率的影响,使得识别准确率可以与CNN相媲美甚至略优于CNN。

2 数据处理

2.1 红外步态图像预处理

实验所用数据集是中国科学院的CASIA C红外步态数据集,该数据集是用红外热感摄像机在夜间拍摄的大规模数据集,包含153个个体类别,并且每个类别包含正常行走、快走、慢走及背包行走4种形态。首先使用背景减除法^[12]提取运动中的人体轮廓,接着使用二值化和形态学方法对人体轮廓图像进行处理,最后将其剪裁缩放至128 pixel×128 pixel尺寸的步态序列。原始图像和二值化剪裁缩放处理后的图像如图1所示,其中[图1(a)]为原始图像,[图1(b)]为二值化后的图像,[图1(c)]为居中剪裁后的图像。



图1 原始红外图像和二值化后的人体轮廓图像。(a)原始图像;(b)二值化后的图像;(c)居中剪裁后的图像

Fig. 1 Original infrared image and human contour image after binarization. (a) Original image; (b) binarized image; (c) centered and clipped image

为了减小穿着对步态识别准确率的影响,所提方法对预处理好的红外步态图像根据人体各部分比例进行遮挡处理。人体各部分比例如图2所示,其中 H 为人的身高。从图中可以看出,人体各部分比例,其中颈部距离头顶 $0.182H$,腹部大约距离头顶 $0.47H$,根据该比例在 $[0.18H, 0.47H]$ 范围内对人体图像进行遮挡处理。

具体的处理步骤如下:首先,对归一化并缩放至宽高比为1的人体图像,将其宽度和高度分别记为 W 和 H ;然后将二值化步态图像在 $[0.18H, 0.47H]$ 区间内像素点的值设置为0。

$$I(x, y) = \begin{cases} O(x, y), & O(x, y) = 0 \\ 0, & O(x, y) \neq 0 \end{cases}, 0 \leq x \leq W, \\ 0.18H \leq y \leq 0.47H, \quad (1)$$

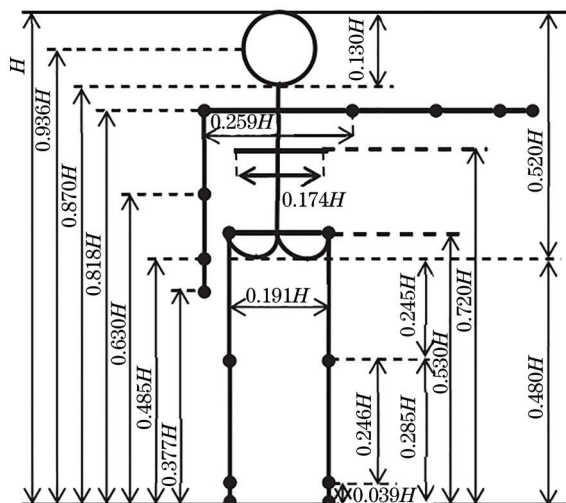


图 2 人体各部分比例

Fig. 2 Proportion of each part of human body

式中: $I(x, y)$ 是经过遮挡处理后的步态图像; $O(x, y)$ 是归一化处理后的步态图像; (x, y) 代表图像中的横纵坐标。经过遮挡处理后的步态图像序列如图 3 所示。

2.2 步态周期的估算

基于卷积神经网络方法的步态识别研究, 在无

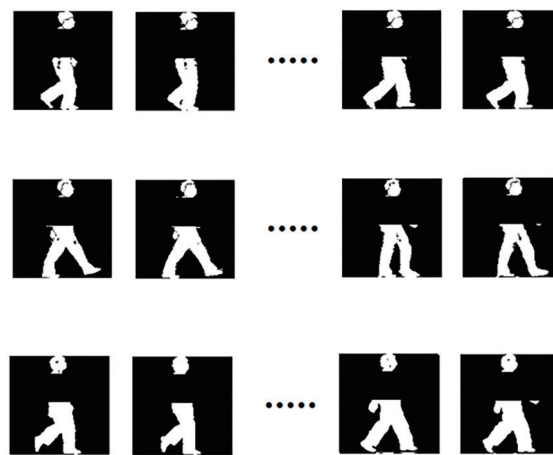


图 3 遮挡处理后的步态图像序列

Fig. 3 Gait image sequence after occlusion processing

序的图像数据集下即可取得较高的识别准确率, 但是利用长短期记忆网络方法进行步态识别是针对有序图像进行特征提取的, 因此需要计算数据集中的步态周期。步态周期是指同一只脚经过相同位置的时间间隔, 由于归一化自相关(NAC)函数可以体现不同图像间的相似度, 因此所提方法采用 NAC 估算步态周期^[13], 该函数的表达式为

$$C_{NAC}(N) = \frac{\sum_{x,y} \sum_{i=0}^{T(N)} g(x,y,i)g(x,y,i+N)}{\sqrt{\left[\sum_{x,y} \sum_{i=0}^{T(N)} g(x,y,i)^2\right] \cdot \left[\sum_{x,y} \sum_{i=0}^{T(N)} g(x,y,i+N)^2\right]}}, \quad (2)$$

式中: $C_{NAC}(N)$ 为与移动 N 幅步态图像的自相关程度; $T(N) = N_{total} - N - 1$, N_{total} 为总的步态图像数量;

$g(x,y,i)$ 表示第 i 帧步态图像中 (x,y) 处的像素值。通过 NAC 函数估算得到的每个个体不同行走状态下的

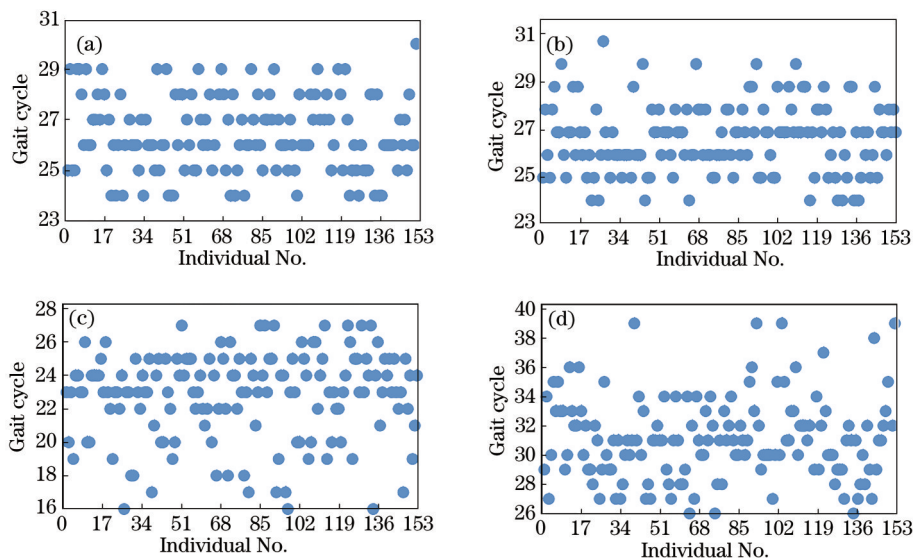


图 4 每个个体不同行走状态下的步态周期。(a) 背包行走; (b) 正常行走; (c) 快速行走; (d) 慢速行走

Fig. 4 Gait cycle of each individual in different walking states. (a) Backpack walking; (b) normal walking; (c) fast walking; (d) slow walking

步态周期如图 4 所示。

3 神经网络模型

3.1 卷积神经网络

卷积神经网络^[14]是一种包含卷积运算的前馈深度神经网络,由于通过感受野和权值共享减少了神经网络需要训练的参数数量,在图像识别领域应用广泛。卷积神经网络主要由输入层、隐藏层和输出层组成。输入层接收经过标准化处理后的数据。隐藏层包含卷积层、池化层及激励层,其中卷积层主要负责进行卷积运算驱动反向传播来调整参数权重;池化层主要负责特征降维,减少参数数量的同时提高网络的容错性;激励层主要负责对卷积层进行一次非线性映射。

输出层根据学习的图像特征输出图像分类结果。

借鉴 ResNet^[15]的设计思想,本实验组设计了一个多分支的卷积神经分类网络,其结构如图 5 所示。在卷积神经网络中,卷积次数越多,所保留的细节信息就越少,同时所提取的特征就越抽象。步态识别更加注重图像中的细节信息,因此所设计的卷积神经网络包含 3 个分支,每个分支提取的特征抽象程度不同,最后对每个分支不同抽象程度的特征进行融合,这样在保留细节信息的同时也能够提取到高级特征。该网络采用的激活函数是 relu,通过对对应特征值相加的方式进行特征融合,图 5 中 pool 2 表示使用 (2, 2) 的窗口进行池化,conv 64×3×3 表示使用 64 个 3×3 大小的卷积核进行卷积运算,dropout 的值为 0.5。

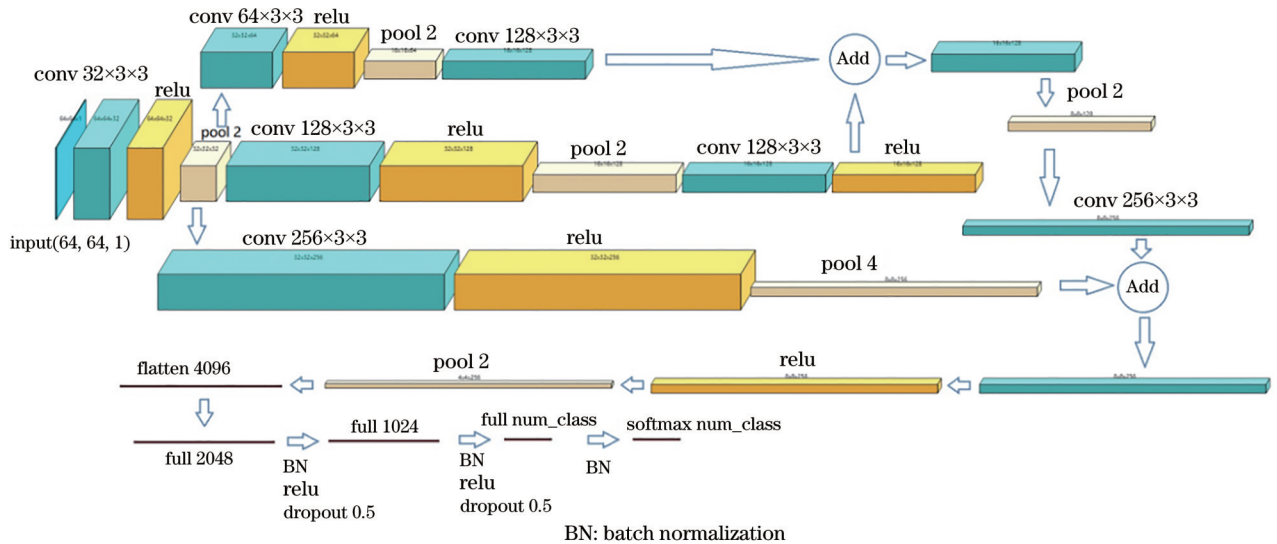


图 5 多分支卷积神经网络

Fig. 5 Multi-branch convolutional neural network

3.2 长短期记忆网络

循环神经网络(RNN)^[16]相对于传统的神经网络,增加了一种循环结构,加入了时间特性。循环神经网络中的每一个当前神经元的输入与前一个神经元的输出有关,网络会对前面的运算信息进行记忆并加入到当前神经元的运算中,因此循环神经网络主要用来处理序列型数据,标准循环神经网络结构如图 6 所示,其中 x 是输入, h 是隐藏层神经元, o 为隐藏层神经元输出, L 为损失函数, y 为训练集标签, U 、 V 、 W 是网络的权值。因为 RNN 中的权值是共享的,所以图 6 中的 W 、 U 、 V 都相同。

标准循环神经网络在前向传播的过程中,输入 x 与输出 y 之间的关系为

$$h^t = \varphi(Ux^t + Wh^{t-1} + b), \quad (3)$$

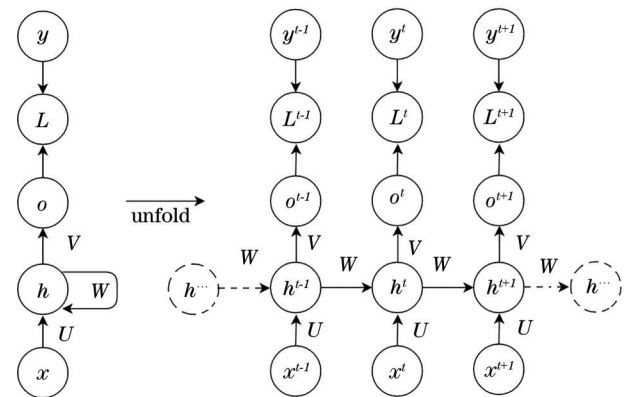


图 6 标准循环神经网络结构

Fig. 6 Standard cyclic neural network structure

$$o^t = Vh^t + C, \quad (4)$$

$$y^t = \sigma(o^t), \quad (5)$$

式中: φ 是激活函数, 在 RNN 中一般使用 tanh 函数作为激活函数; b 是隐藏层计算的偏置量; C 为输出神经元的偏置量; σ 是用于处理输出单元的激活函数, 一般使用 softmax 函数。从(3)式可知, 在 t 时刻隐藏层神经元 h^t 不仅与 t 时刻的 x 输入有关, 而且还与上一时刻隐藏层神经元 h^{t-1} 有关, 因此隐藏层输出神经元 o 与输入 x 之间存在时间相关性。

LSTM 是一种改进后的循环神经网络, 标准循环神经网络隐藏层的重复结构一般只有一个 tanh 激活层, 但是 LSTM 的重复模块中引入了输入门、遗忘门、输出门及一个细胞状态, 用 3 个门实现对信息的选择性记忆, 加入不断改变的细胞状态, 确保每一个隐藏层神经元的输出可以受到之前比较久远的输入的影响, 因此 LSTM 相比于标准循环神经网络有着更长的记忆时间。

所提基于 LSTM 构建的步态识别模型如图 7 所示, 该模型主要是由 2 个 LSTM 结构线性堆叠而成的, 第 1 个 LSTM 结构中设置了 512 个记忆单元, 为了防止模型过拟合在第 1 个 LSTM 结构中加入了 dropout 结构, 失活比率为 0.5; 第 2 个 LSTM 结构设置了 256 个记忆单元, 同样加入失活比率为 0.5 的

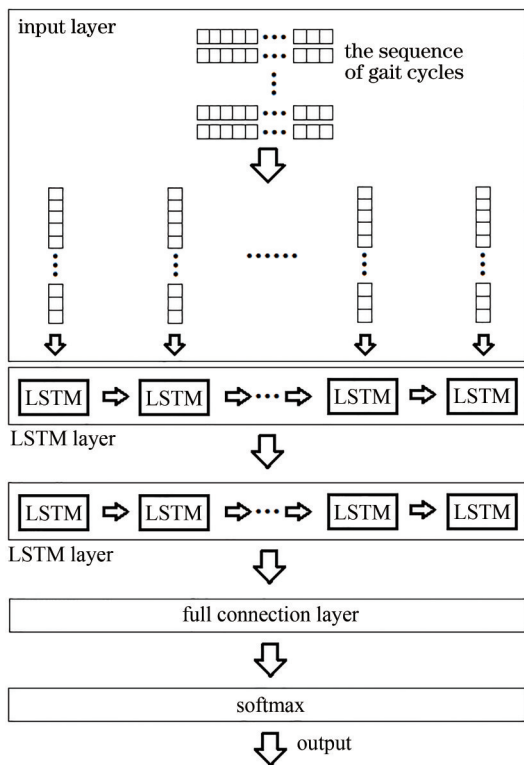


图7 LSTM步态识别模型

Fig. 7 LSTM gait recognition model

dropout 结构。将预处理好的 1 个步态周期中的每张图片重构为 1 个一维向量, 然后将这些重构后的一维向量作为二维矩阵中的 1 行组成 1 个二维矩阵, 该二维矩阵中每 1 行都是某一时刻的步态图像的像素值, 因此各行像素值之间具有时间先后关系, 最后将该二维矩阵作为模型的输入。设置 LSTM 结构中的时间步为 1 个步态周期, 输入数据经过线性堆叠的 LSTM 结构进行特征提取, 为了加快模型收敛, 将提取后的特征序列经过 BN 层进行归一化处理, 归一化后的特征序列经全连接层处理和输出类别等长的特征向量, 最后经过 softmax 激活函数层得到与对应类别的相似度概率序列, 取该概率序列的最大值对应的类别为最终分类结果。

4 实验设计与结果分析

4.1 实验设计

以 TensorFlow 作为 Backbone 的 Keras 深度学习框架搭建实验环境, 所用数据集来自中国科学院自动化研究所提供的 CASIA C 步态数据集。该数据集包含 153 个研究对象, 在与摄像机呈 90° 视角下进行采集, 每个研究对象包含背包行走、正常行走、快走和慢走 4 种行走状态, 由于原始数据集中的研究对象不在拍摄图片中央, 因此, 在图像预处理阶段需要将剪裁后的研究对象放置到图像中央, 接着根据人体各部分比例对数据集进行部分遮挡处理, 让网络注意力集中在腿部, 最后把图片处理成 128×128 的大小。针对 CNN, 对每个研究对象的 4 种不同状态下的步态图像进行混合乱序处理, 并根据常见数据集划分方法之一的留出法将整个数据集以 8:2 的比例划分成训练集和测试集, 然后缩放至 64×64 大小后送入到 CNN 中进行分类识别。针对 LSTM 模型, 对每个研究对象不同行走状态的图像序列, 以估算得出的各个不同行走状态下的步态周期帧数中的最大步态周期帧数作为滑动窗口大小, 连续向前滑动获取 1 个步态周期序列, 然后将步态周期中的每 1 帧像素值处理成 1 个向量, 并将这些向量拼接成 1 个二维矩阵, 再对处理后的这些二维矩阵进行打乱处理; 最后同样对整个数据集采用留出法的划分方式以 8:2 的比例划分为训练集和测试集, 并缩放至 32×32 的大小后送入到 LSTM 模型中进行分类识别。

CNN 的网络结构如图 5 所示, 初始学习率设置为 1×10^{-3} , 卷积核大小设置为 3×3 , 最大池化层的

卷积核为 2×2 和 4×4 两种,选择 categorical_crossentropy 多分类交叉熵损失函数作为损失函数,使用 Adam 优化器,设置 EarlyStop 机制,当 val_loss 持续 6 次不下降就停止训练,训练过程中动态调整学习率,当学习率 3 次不下降就自动降低学习率继续训练, batch size 设置为 32, 训练迭代的 epoch 设为 10。LSTM 模型中初始学习率设置为 1×10^{-3} , 由于存在 4 种不同的行走状态,估算出来的步态周期并不相同,而最大的步态周期帧数必然包含了所有行走状态的步态周期,因此以 4 种行走状态中最大的步态周期为 LSTM 网络中的时间步长, input size 设置为 1024, 同样使用 Adam 优化器,以 categorical_crossentropy 多分类损失函数作为损失函数, batch size 设置为 32。

4.2 结果分析

分类结果主要从模型分类准确率方面进行评价。CNN 和 LSTM 的步态识别准确率如图 8 所示。

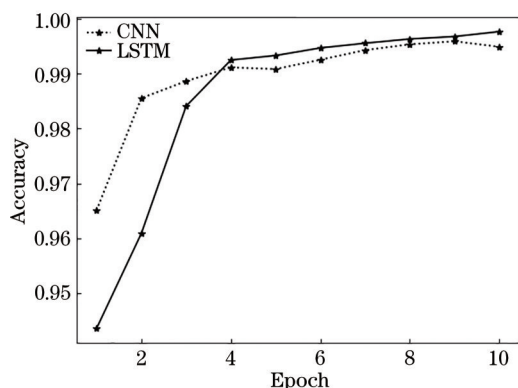


图 8 CNN 和 LSTM 识别准确率对比

Fig. 8 Comparison of recognition accuracy between CNN and LSTM

从图 8 可以看出, CNN 在迭代训练 10 个 epoch 后, 分类准确率达到 99.48%, LSTM 在迭代训练同样 epoch 数量后, 分类准确率达到 99.76%, 表明 LSTM 模型对遮挡处理后的步态图像序列进行分类识别的准确率略高于 CNN 模型。由于训练 LSTM 模型时是从 1 个步态周期中提取步态特征的, 相比于训练 CNN 模型时从 1 个单帧图像中提取步态特征而言需要更长的学习时间, 因此在训练迭代次数不够时, LSTM 的识别准确率低于 CNN 模型。步态图像序列遮挡部分穿着后丢失了部分姿态特征, CNN 模型注意力集中在对腿部空间特征的提取, 而 LSTM 模型相比于 CNN 模型, 不仅可以提

取 1 个步态周期内序列图像的腿部空间特征, 而且还可以提取腿部行走过程中的时间特征, 因此 LSTM 模型的分类性能优于 CNN 模型。

5 结 语

将 LSTM 应用于红外步态识别领域, 并在中国科学院自动化所提供的 CASIA C 数据库上进行了仿真实验, 所提方法取得了较好的识别效果。在实验过程中, 与 CNN 进行了对比分析, 从分类准确率方面验证了所提 LSTM 方法在只关注腿部行走特征的情况下, 在步态识别领域有更好的应用效果。实际应用场景中环境复杂多变, 如行人穿着、背包等都会导致摄像机拍摄的图像发生较大改变从而影响识别准确率, 人群聚集产生遮挡导致人体轮廓提取困难等问题, 步态识别领域以现有的技术手段很难直接应用在现实场景。但是, 作为可应用于远距离的一种生物特征识别技术, 在人脸识别技术受限情况下, 步态识别技术可以作为人脸识别技术的补充, 因此对步态识别领域的研究具有重大意义。步态识别未来的研究思路主要是: 尽可能地脱离实验室理想化的环境, 将研究环境不断向现实场景接近, 将 CNN 和 LSTM 相结合, 提取更加稳定的人体步态特征, 让多种因素对步态识别的影响降到最低, 从而得到能够很好解决步态识别在复杂场景下各种困难的识别模型, 并将取得的成果应用于安防领域, 维系社会安定。

参 考 文 献

- [1] Wang L, Tan T N, Ning H Z, et al. Silhouette analysis-based gait recognition for human identification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1505-1518.
- [2] Kong K, Tomizuka M. A gait monitoring system based on air pressure sensors embedded in a shoe [J]. IEEE/ASME Transactions on Mechatronics, 2009, 14(3): 358-370.
- [3] Mori A, Makihara Y, Yagi Y. Gait recognition using period-based phase synchronization for low frame-rate videos [C] // 2010 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, Turkey. New York: IEEE Press, 2010: 2194-2197.
- [4] Rokanujjaman M, Islam M S, Hossain M A, et al. Effective part-based gait identification using frequency-domain gait entropy features [J]. Multimedia Tools and

- Applications, 2015, 74(9): 3099-3120.
- [5] Mansur A, Makihara Y, Aqmar R, et al. Gait recognition under speed transition[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2521-2528.
- [6] Tsuji A, Makihara Y, Yagi Y. Silhouette transformation based on walking speed for gait identification[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 717-722.
- [7] Tao D C, Li X L, Wu X D, et al. General tensor discriminant analysis and Gabor features for gait recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1700-1715.
- [8] He Y W, Zhang J P. Deep learning for gait recognition: a survey[J]. Pattern Recognition and Artificial Intelligence, 2018, 31(5): 442-452.
何逸炜,张军平.步态识别的深度学习:综述[J].模式识别与人工智能,2018,31(05):442-452.
- [9] Lukoševičius M, Jaeger H. Reservoir computing approaches to recurrent neural network training[J]. Computer Science Review, 2009, 3(3): 127-149.
- [10] Shiraga K, Makihara Y, Muramatsu D, et al. GEINet: view-invariant gait recognition using a convolutional neural network[C]//2016 International Conference on Biometrics (ICB), June 13-16, 2016, Halmstad, Sweden. New York: IEEE Press, 2016: 16252579.
- [11] Rida I, Bouridane A, Marcialis G L, et al. Improved human gait recognition[M]//Murino V, Puppo E. Image analysis and processing-ICIAP 2015. Lecture notes in computer science. Cham: Springer, 2015, 9280: 119-129.
- [12] Liu Z M, He S J, Hu W J, et al. Moving object detection based on background subtraction for video sequence[J]. Journal of Computer Applications, 2017, 37(6): 1777-1781.
刘仲民,何胜皎,胡文瑾,等.基于背景减除法的视频序列运动目标检测[J].计算机应用,2017,37(6):1777-1781.
- [13] Iwama H, Okumura M, Makihara Y, et al. The OUISIR gait database comprising the large population dataset and performance evaluation of gait recognition [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(5): 1511-1521.
- [14] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 15523970.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [16] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[EB/OL]. (2014-09-08) [2021-02-01]. <https://arxiv.org/abs/1409.2329>.