计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(11):215−222 [doi: 10.15888/j.cnki.csa.008773]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于骨架序列提取的异常行为识别①

吴　晨[1], 孙　强[2], 倪宏宇[3], 颜文旭[1]

[1](江南大学 物联网工程学院, 无锡 214122)
[2](哈工大机器人 (合肥) 国际创新研究院, 合肥 230601)
[3](国网浙江省绍兴供电公司, 绍兴 312000)
通信作者: 颜文旭, E-mail: ywx01@jiangnan.edu.cn

摘　要: 视频监控系统中的人员异常行为识别研究具有重要意义. 针对传统算法检测实时性和准确性差, 易受环境影响的问题, 提出一种基于骨架序列提取的异常行为识别算法. 首先, 改进 YOLOv3 网络用以对目标进行检测、结合 RT-MDNet 算法进行跟踪, 得到目标的运动轨迹; 然后, 利用 OpenPose 模型提取轨迹中目标的骨架序列; 最后通过时空图卷积网络结合聚类对目标进行异常行为识别. 实验结果表明, 在存在光照变化的复杂环境下, 算法识别准确率达 94%, 处理速度达 18.25 fps, 能够实时、准确地识别多种目标的异常行为.

关键词: 异常行为识别; 人体骨架序列; 卷积神经网络; 深度学习; 姿态估计

引用格式: 吴晨,孙强,倪宏宇,颜文旭.基于骨架序列提取的异常行为识别.计算机系统应用,2022,31(11):215–222. http://www.c-s-a.org.cn/1003-3254/8773.html

Recognition of Abnormal Behavior Based on Skeleton Sequence Extraction

WU Chen[1], SUN Qiang[2], NI Hong-Yu[3], YAN Wen-Xu[1]

[1](School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)
[2](HRG International Institute (Hefei) of Research and Innovation, Hefei 230601, China)
[3](State Grid Shaoxing Power Supply Company, Shaoxing 312000, China)

Abstract: The research on the recognition of abnormal human behavior in video surveillance systems is of great significance. As traditional algorithms are easily affected by the environment and have poor timeliness and accuracy, an abnormal behavior recognition algorithm based on skeleton sequence extraction is proposed. Firstly, the improved YOLOv3 network is used to detect targets and is combined with the RT-MDNet algorithm to track them for target trajectories. Then, the OpenPose model is employed to extract the skeleton sequence of targets in the trajectories. Finally, the spatiotemporal graph convolutional network combined with clustering is applied to recognize the abnormal behavior of the targets. The experimental results indicate that the proposed algorithm has a processing speed of 18.25 fps and recognition accuracy of 94% under a complex background of light changes, which can accurately identify the abnormal behavior of various targets in real time.

Key words: abnormal behavior recognition; human skeleton sequence; convolutional neural network (CNN); deep learning; pose estimation

视频监控系统已成为车站、商场等公共场所识别安全威胁、预防犯罪行为的重要方式. 传统的监控系统主要依靠人力来分析海量视频信息, 不但会消耗大量人力财力, 而且不能保证识别的准确性[1,2]. 因此,

① 基金项目: 国网浙江省电力有限公司科技项目 (5211SX220003)
收稿时间: 2022-02-21; 修改时间: 2022-03-21, 2022-03-28; 采用时间: 2022-03-29; csa 在线出版时间: 2022-07-07

异常行为智能识别在视频监控领域具有广阔的发展前景.

早期, 主要基于视觉图像进行行为识别, 包括模板匹配法、状态空间法两种方法[3]. 前者将图像的颜色、纹理等特征与特征模板相匹配, 分为视频帧特征匹配及视频片段特征匹配两种, 常用方法有 STIP (space-time interest points)、DT (decision tree)、iDT (improved dense trajectories) 等[4,5]; 后者将运动序列看作不同状态间的一次遍历利用各类动作数据训练其模型参数[6], 常用方法有最大熵马尔可夫模型、隐马尔可夫模型、条件随机场等. 但以上方法易受背景抖动、光照变化及图像噪声等影响.

近年来, 深度学习持续发展, 应用在行为识别领域也成效显著. 相比传统方法, 深度图像具有视图不变性、光照不变性, 只依赖关键参数, 而不依赖人体的时间和空间信息, 因此更适合于行为的实时识别[7–9]. 骨骼不会随着视点或外观的变化而改变, 遭受的类内差异更小; 骨骼也不会受到无关信号的干扰, 可以大大简化动作识别的学习过程. 因此, 相比其他动作模式, 在行为识别任务中, 人体骨架往往能传达更多的信息[10]. 并且随着 Kinect 等设备的发展, 骨架信息的提取也越来越便捷.

基于人体骨架进行行为识别, 常用循环神经网络、卷积神经网络及图卷积网络等方法. 前两者需要运用人体运动专业知识来建模, 较为复杂[11,12]. 文献 [13] 设计了一种带信任门的时空 long short term memory (LSTM), 用于联合学习骨架序列的时空信息并自动去除噪声关节, 但构建深度 LSTM 来提取高级特征较为困难[14,15]. Convolutional neural networks (CNN) 可以提取高级特征, 但无法对视频的长期时间依赖性进行有效建模[16]. 文献 [17] 运用深度渐进强化学习方法提取关键帧, 提高了训练效率, 但识别准确率一般. 文献 [18] 设计了一种注意力和时间双通道的伪图卷积网络, 能提取更多关键信息, 但计算速度较慢. 因此, 进行实时、准确的检测是异常行为识别的难点.

为对异常行为进行实时准确的识别, 本文提出了一种利用人体骨骼轨迹的识别方法: 根据识别目标种类改进 YOLOv3 网络的损失函数后, 将其用以检测目标, 并用 RT-MDNet (real-time multi-domain convolutional neural network) 算法跟踪目标; 构建出目标的运动轨迹后, 通过 OpenPose 模型得到轨迹中目标的骨架序列, 最后运用时空图卷积神经网络结合聚类, 充分利用人体关节的时间、空间关系对目标进行异常行为识别.

1 目标运动轨迹构建

识别异常行为时, 利用行为序列比利用视觉图像识别准确率更高、鲁棒性更强, 因此, 本文进行了目标运动轨迹构建, 提取目标特征后, 对比视频帧序列的前后帧, 判断候选样本是否为跟踪目标. 轨迹构建过程中, 主要存在 3 方面误差: 目标快速运动或者目标与镜头距离发生变化时, 目标在视频中的尺寸会发生改变;光照变化、或目标被遮挡时, 跟踪目标会丢失; 随着跟踪时间的变长, 跟踪性能随之下降, 会偏离甚至丢失跟踪目标. 针对以上, 本文改进 YOLOv3 网络对目标进行检测后, 再通过 RT-MDNet 跟踪算法对多目标在复杂场景下进行长时间的跟踪.

1.1 基于改进 YOLOv3 网络的目标检测

准确检测目标是构建其运动轨迹的第 1 步. YOLO 网络基于端到端学习直接在网络中提取特征来预测边界框位置及其类别, 流程简单、检测速度快[19,20]. 基于目标检测实时性与准确性的要求, 利用其对不同尺度目标适应性强的识别优势, 本文采用 YOLOv3 网络作为主体网络开展对视频中目标的检测, 并依据检测对象的特征优化其参数.

YOLOv3 网络以 Darknet-53 为主体, 与 YOLOv2 网络的 Darknet-19 相比, 该网络结构更深, 可有效防止过拟合、提高检测准确率.

将输入标准化到 416×416 的统一尺寸后, 通过特征提取网络进行 5 次下采样后得到 13×13 的特征图. 将其划分为 13×13 的栅格, 每个栅格预测 3 个边界框, 预测的输出特征图包含预测框的宽高、中心坐标及置信度. 置信度 *confidence* 表示如下:

$$confidence = Pr(Object) \times IoU_{\text{pred}}^{\text{truth}} \tag{1}$$

其中, $IoU_{\text{pred}}^{\text{truth}}$为预测边界框与实际边界框的重合度, $Pr(Object)$为栅格预测人体目标的置信概率. 目标的中心坐标落在哪个栅格内, 就由其对目标进行预测. 设置置信度阈值筛除得分低的预测框后, 对剩余预测框执行非极大值抑制, 即可输出最终检测目标的位置.

原始 YOLOv3 网络损失函数中的预测框宽高损失函数, 采用均方误差函数的方法, 没有考虑到重合面积和位置的影响, 因此, 本文作出以下改进:

$$L_{\text{loc}} = 1 - IoU_{\text{pred}}^{\text{truth}} + \frac{b^2}{c^2} \tag{2}$$

其中, b 为预测框与实际框中心点间的距离, c 为包围预测框与实际框的最小矩形的对角线距离.

本文令识别目标种类仅为人体, 可以直接去掉分类损失函数, 其余函数不变, 经过上述改进后的总损失函数为:

$$Loss = L_{\text{obj}} + L_{\text{loc}} + L_{\text{conf}} \tag{3}$$

其中, L_{obj}为预测框中心坐标损失函数, L_{conf}为置信度损失函数, 检测流程图如图 1 所示.

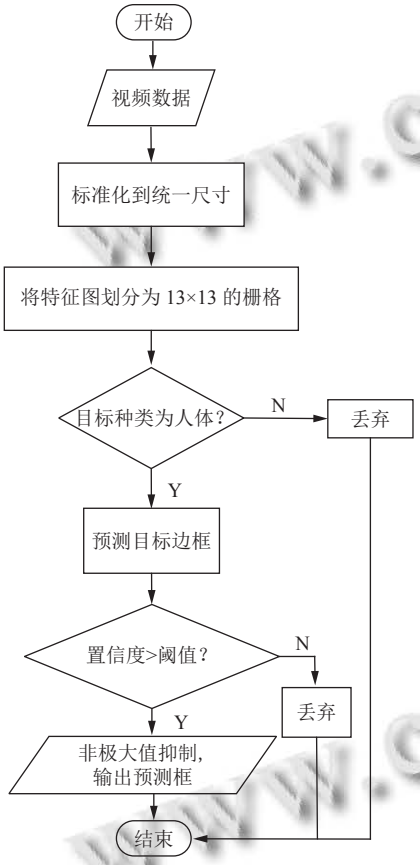


图 1 目标检测算法流程图

1.2 基于 RT-MDNet 算法的目标跟踪

目标丢失后需要再次搜索并定位时, YOLOv3 网络将其当作一个回归问题来处理, 存在难以准确定位异常比例目标的问题. 使图像与输入边界框候选区域间的重合度最大化能够更精确地定位[21], 因此, 本文在 YOLOv3 网络检测的基础上引入 RT-MDNet 跟踪算法, 以纠正模型的错误更新.

RT-MDNet 跟踪算法将每个视频看作一个独立的域, 外观模型f^d定义为:

$$f^d = \left[\varphi^1\left(x^d;R\right),\cdots,\varphi^D\left(x^d;R\right)\right] \in \Re^{2\times D} \tag{4}$$

其中, D 为训练集总数, 第 d 域的图片x^d及边界框 R 为输入. 通过φ^d函数得出 d 域最后一层全连接层的背景、前景的二分类得分后, 当前帧目标边界框的预测值即为得分最高的前景.

令 YOLOv3 网络、RT-MDNet 跟踪算法的边界框分别为:

$$R_1^d = (x_1, y_1, \omega_1, h_1) \tag{5}$$

$$R_2^d = (x_2, y_2, \omega_2, h_2) \tag{6}$$

两个边界框的重合度为:

$$IoU = \frac{R_1^d \cap R_2^d}{R_1^d \cup R_2^d} = \frac{col\_I \times row\_I}{\omega_1 \times h_1 + \omega_2 \times h_2} \tag{7}$$

其中, 重合部分宽为:

$$col\_I = \left|\min(x_1+\omega_1, x_2+\omega_2) - \max(x_1, x_2)\right|$$

重合部分高为:

$$row\_I = \left|\min(y_1+h_1, y_2+h_2) - \max(y_1, y_2)\right|$$

则期望目标边界框R^*为: $R^* = \begin{cases} R_1^d, IoU < Y \\ R_2^d, IoU \geqslant Y \end{cases}$.

运用 YOLOv3 网络开始检测时, 同步初始化 RT-MDNet 跟踪算法并计算 IoU. IoU 大于或等于阈值 Y 时, 选择 RT-MDNet 算法的边界框R_2^d为模型f^d的输入; IoU 小于阈值 Y 时, 当前帧目标跟踪结果可信度低, 为找回目标, 选择 YOLOv3 网络的边界框R_1^d为模型f^d的输入.

2 目标骨架提取

人体骨架由关节点、骨骼组合而成, 两个人体关节点相连得到一段骨骼. 人体关节点位置会随人体动作而改变, 骨骼角度也会随之发生明显变化. 相对于时空信息、光流信息、时间特征等特征, 人体骨架更能充分表达图像中的人体信息. 根据时间顺序组合人体骨架后得到的人体骨架序列也更符合行为的动态特性, 能准确地描述行为.

骨架提取分为自底向上 (bottom-up)、自顶向下 (top-down) 两种, 前者先在全局进行关键点检测, 再对关键点进行聚类, 这类方法实时性较好; 后者先在图像中检测到目标, 再分别对每个目标进行关节点定位, 这

类方法近距离检测效果不佳, 并且随着目标数量的增加, 耗时会呈线性增长. 基于目标检测实时性的要求, 本文采用自底向上的 OpenPose 模型进行骨架提取.

OpenPose 模型通过卷积神经网络直接从视频中提取人体骨骼关节点坐标, 核心是 VGG19 和部分亲和场 PAFs 的应用. 整体流程可概括为: 经 VGG19 网络获得目标的图像特征后, 将图像特征分为两支, 分别预测每个关节点的置信度、亲和度[22,23].

人体骨架中用于行为识别的人体关节点有 14 个, 为头、颈、左肩、右肩、左肘、右肘、左腕、右腕、左髋、右髋、左膝盖、右膝盖、左脚踝、右脚踝. OpenPose 原理图如图 2 所示, 其中, F 是经过 10 层 VGG19 后提取到的目标特征, 置信度向量S可以表示为:

$$S^1=\rho^1(F) \tag{8}$$

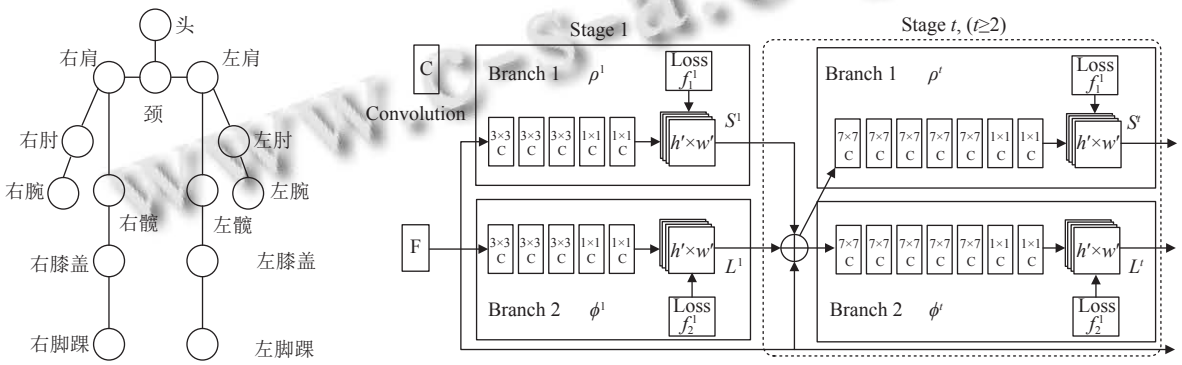
$$S^t=\rho^t\left(F,S^{t-1},L^{t-1}\right),\ \forall t\geqslant 2 \tag{9}$$

其中, ρ^t为置信度向量第t次迭代的卷积神经网络, 亲和度向量L可以表示为:

$$L^1=\phi^1(F) \tag{10}$$

$$L^t=\phi^t\left(F,S^{t-1},L^{t-1}\right),\ \forall t\geqslant 2 \tag{11}$$

其中, ϕ^t为亲和度向量第t次迭代的卷积神经网络.



(a) 人体骨架模型　　(b) OpenPose 原理图

图 2　骨架提取原理

置信度向量的数量与关节点数量一致, 置信度反映了关节点出现在每个像素位置的可能性; 亲和度向量的数量与骨骼段数量一致, 亲和度反映了每一对人体部位的相关性, 即它们是否属于同一个人. 将人体两两关节的最优连接问题转化为最大权值二分图匹配问题, 骨架关节作为二分图中的节点、PAFs 作为二分图中边的权值, 利用贪心算法和匈牙利匹配算法, 确定人体各个关节的位置, 使骨架关节和骨骼相连得到完整的人体骨架特征.

OpenPose 对多人骨架的检测效果图如图 3 所示, 轨迹中目标的人体骨架序列为$\{J_1,J_2,\cdots,J_m\}$, 其中$J_i,1\leqslant i\leqslant m$为按时间顺序组合得到的第 i 个骨架.

3　目标异常行为识别

早期基于骨架特征进行行为识别只是在各个时间步骤对其进行时序分析, 忽视了人体关节的空间关系, 导致识别准确率有限. 为提高识别准确率, 本文将图卷积拓展为时空图卷积, 设计用于行为识别的骨骼序列通用表示. 图中包含符合关节连接的空间边、在连续时间中连接相同关节的时间边, 在此基础上构建多层时空图卷积, 沿空间和时间两个维度进行信息整合.

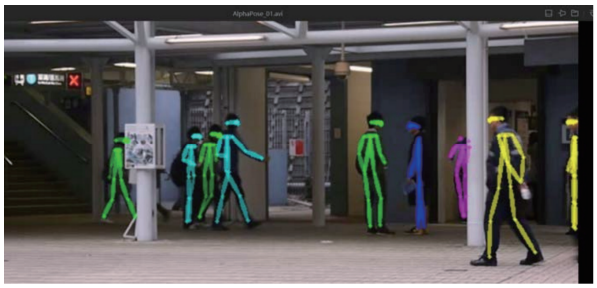


图 3　骨架检测效果图

3.1　时空图卷积

在一个 T 帧的有 N 个关节的骨架序列上构建时空图$G=(V,E)$. 节点集$V=\{v_{ti}|t=1,\cdots,T,\ i=1,\cdots,N\}$包含了骨架序列中的所有关节; E 由两个子集组成, 空间

边$E_S=\{v_{ti}v_{tj}|(i,j)\in H\}$描述了每一帧的内部骨架连接, H 是一组自然连接的人体关节, 时间边$E_F=\{v_{ti}v_{(t+1)i}\}$则在连续帧中连接相同的关节.

给定核大小为$K\times K$的卷积算子和通道数为 c 的输入特征映射f_{in}, 则输出特征图上, 位置 x 的卷积输出$f_{\text{out}}(x)$可以表示为:

$$f_{\text{out}}(x)=\sum_{h=1}^{K}\sum_{\omega=1}^{K}f_{\text{in}}(p(x,h,\omega))\cdot w(h,\omega) \tag{12}$$

其中, 采样函数$p(x,h,\omega)$枚举了位置 x 的所有邻域, 权函数$w(h,\omega)$为 c 维实空间中的权值向量, 用于计算与 c 维采样输入特征向量的内积.

将式 (12) 中的输入特征图用一个空间图V_t替换, 即可拓展为空间图卷积, V_t中每个节点特征图映射$f_{\text{in}}^t:V_t\to\Re^c$都对应着一个 c 维向量. 对于每个节点, 其邻域为$B(v_{ti})=\left\{v_{tj}\middle|d\left(v_{tj},v_{ti}\right)\leqslant L\right\}$, 其中, $d\left(v_{tj},v_{ti}\right)$表示从$v_{tj}$到$v_{ti}$任意路径的最短长度, 文中令 L=1, 即只有与之相邻的节点才会被采样. 在邻域上定义采样函数, 则$p\left(v_{ti},v_{tj}\right)=v_{tj}$. 将邻域$B(v_{ti})$划分为固定数量的 K 个子集, 则权函数$w\left(v_{ti},v_{tj}\right)=w'\left(l_{ti}\left(v_{tj}\right)\right)$、映射$l_{ti}:B(v_{ti})\to\{0,\cdots,K-1\}$.

将空间图卷积拓展到时间域中, 即得时空图卷积. 邻域为$B(v_{ti})=\left\{v_{qj}\middle|d\left(v_{tj},v_{ti}\right)\leqslant K,\ |q-t|\leqslant\lfloor\Gamma/2\rfloor\right\}$, 标签映射为$l_{ST}\left(v_{qj}\right)=l_{ti}\left(v_{tj}\right)+(q-t+\lfloor\Gamma/2\rfloor)\times K$, 其中, q 是对时间域的拓展, Γ是控制时间域的卷积核大小、$l_{ti}\left(v_{tj}\right)$是$v_{ti}$处单帧情况的标签映射.

本文将节点 i 的 1 邻域划分为 3 个子集. 第 1 子集为节点 i 本身, 第 2 子集为比节点 i 更靠近骨架中心的邻节点集合, 第 3 子集为比节点 i 更远离骨架中心的邻节点集合, 分别表示了静止、向心运动和离心运动的运动特征. 则: $l_{ti}\left(v_{tj}\right)=\begin{cases}0, \text{ if } r_j=r_i\\ 1, \text{ if } r_j<r_i\\ 2, \text{ if } r_j>r_i\end{cases}$, r_i为节点 i 到人体中心的距离.

时空图卷积网络包含 9 层时空图卷积, 前 3 层输出 64 通道、中间 3 层输出 128 通道、后 3 层输出 256 通道. 4、7 层的时间卷积层设为池化层, 每次经过 ST-GCN 结构后, 将一半的神经元进行 dropout 处理. 输入的骨架序列数据首先进行正则化, 其次经过全局池化后得到 256 维特征向量, 最后经 Softmax 分类后, 可得行为的发生概率.

3.2 聚类

设 t 时刻第 m 个人的骨架序列在经过时空图卷积网络后决策出来的行为为$A_t(m)$. 实际应用中, 少数帧的骨架提取会因外部环境的干扰而存在噪声, 对行为分类的准确率产生影响, 因此, 不能直接将$A_t(m)$作为最终决策的行为输出. 故本文使用深嵌入集群的方法, 将训练集样本联合嵌入到一个潜在空间, 使用基于 Dirichlet 过程的混合来确定行为是否正常.

对于输入样本 x_i, 用 z_i 表示编码器的潜在嵌入, 用Θ表示聚类层的参数, 用 y_i 表示利用聚类层计算的软聚类分配. 第 i 个样本被分配到第 u 聚类的概率p_{iu}为:

$$p_{iu}=Pr(y_i=u|z_i,\Theta)=\frac{\exp\left(\theta_u^{\mathrm{T}}z_i\right)}{\sum_{u=1}^{U}\exp\left(\theta_u^{\mathrm{T}}z_i\right)} \tag{13}$$

聚类的目标是最小化当前模型聚类预测 P 与目标分布 Q 之间的 KL 散度, 即:

$$L_{\text{cluster}}=KL(Q\|P)=\sum_i\sum_u q_{iu}\log\frac{q_{iu}}{p_{iu}} \tag{14}$$

目标分布的目的是通过规范化和将每个值逼近 0 或 1 来加强当前的集群分配, 将 P 变换为 Q 的函数的递归应用最终能得到一个硬赋值向量. 目标分布的每个成员都使用式 (15) 计算:

$$q_{iu}=\frac{p_{iu}\Big/\left(\sum_i p_{iu}\right)^{\frac{1}{2}}}{\sum_u p_{iu}\Big/\left(\sum_i p_{iu}\right)^{\frac{1}{2}}} \tag{15}$$

用为编码训练集计算的 K-means 质心来初始化聚类层, 优化是以期望最大化 (EM) 的方式完成的.

Dirichlet 过程混合模型 (DPMM) 是评价比例数据分布的一种有效方法, 能够在拟合阶段评估一组分布参数、在推理阶段使用拟合模型为每个嵌入样本提供一个评分. 该模型是多模式的, 每个模式代表与一个正常行为相对应的比例, 在测试时, 使用混合模式对每个样本进行对数概率评分.

模型的训练包括两个阶段: 保持网络聚类分支不变的自编码器预训练阶段, 以及对嵌入和聚类进行优化的微调阶段.

在预训练阶段, 模型通过最小化重构损失L_{rec}来学习对序列进行编码和重构, L_{rec}是原始的时间位姿图与 ST-GCN 重构的时间位姿图之间的L_2损失.

在微调阶段, 模型优化了由重构损失和聚类损失组成的组合损失函数. 根据L_{cluster}对聚类层进行优化,

根据L_{rec}对译码器进行优化, 根据$L_{combined}$对编码器进行优化. 当编码器被优化为两种损耗时, 译码器被保留并作为正则化器来保持编码器的嵌入质量. 这一阶段的综合损失为:

$$L_{combined} = L_{rec} + L_{cluster} \tag{16}$$

4 实验结果与分析

本文实验平台, 硬件配置为 i7-6700 3.4 GHz CPU、NVIDIA GTX 1080 GPU、64 GB RAM, 软件运行环境为 Windows 10 64 位, 平台为 Python 3.6+OpenCV 3.3.1 开源视觉库+TensorFlow 1.8.0 开源机器学习框架.

本文选用 CUHK 数据集进行实验验证. CUHK 是一个大规模的人员搜索基准, 由 18 184 张图像、8 432 个身份组成, 每个图像都存在多个查询人和多个背景人. 图像跨越数百个场景, 并且包括视点、照明、分辨率及遮挡的变化. 该数据集的训练集包含 11 206 幅图像、5 532 个查询人, 测试集包含 6 978 幅图像、2 900 个查询人.

4.1 异常行为识别实验

本节测试算法对不同环境的鲁棒性, 分别涉及强光、弱光、稀疏和拥挤, 部分实验场景如图 4 所示, 使用文献 [24] 中的精确率、召回率作为性能指标, 本文算法在验证集上的检测结果如表 1 所示.

由表 1 可知, 强光、稀疏条件下, 本文算法精确率最高; 弱光、拥挤条件下, 本文算法精确率最低. 由精确率降幅可知, 其他条件不变时, 拥挤比弱光造成了更多误检; 而不同情况下召回率差别不大, 表明弱光和拥挤不会造成更多漏检. 原因在于, 弱光或拥挤会导致骨架提取不完整, 拥挤导致遮挡后更易造成骨架提取不完整, 引起较多误检; 正常行为骨架会被误检为异常行为骨架, 但异常行为骨架不会被检测为正常行为骨架, 漏检情况较少. 总体上来说, 本文算法在不同环境下各项性能指标差距不大, 精确率均能达到 90% 以上, 说明算法具有较好的鲁棒性.

表 1 异常行为识别性能 (%)

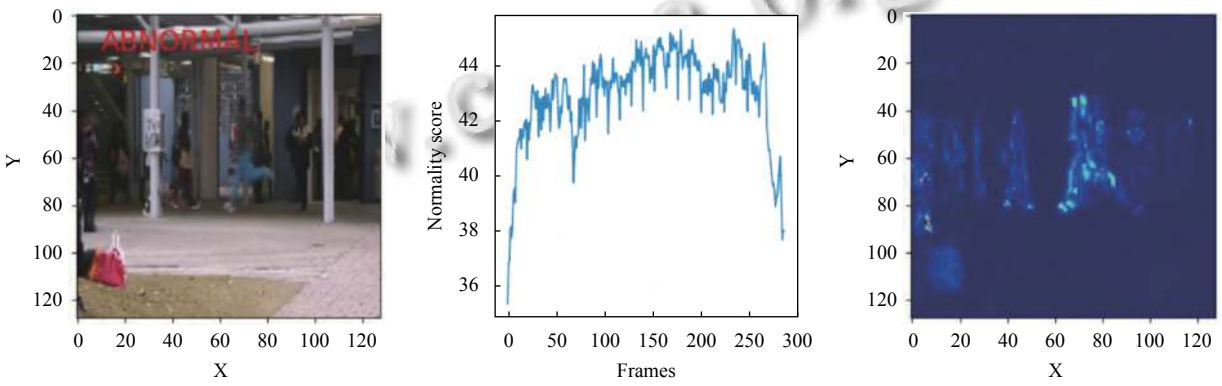
影响因素	精确率	召回率
弱光、拥挤	90.9	96.9
弱光、稀疏	95.6	96.5
强光、拥挤	92.6	96.7
强光、稀疏	97.1	96.4
总计	94	96.6

4.2 对比实验

利用光流法[5]、Motion Instability[25]、支持向量机[26] 和马可科夫随机场[27] 等方法进行异常行为识别, 并与本文算法的识别结果相比较, 表 2 为实验结果.

表 2 异常行为识别方法性能比较

方法	精确率 (%)	处理速度 (fps)
光流法[5]	90.5	9.36
Motion Instability[25]	92.7	11.23
支持向量机[26]	80.3	16.84
马尔科夫随机场[27]	83.5	15.58
本文算法	94	18.25



(a) Original video (b) Normality curve (c) Diff video

图 4 部分实验场景

由表 2 可知, 进行异常行为识别时, 本文算法性能最优. 原因在于, Motion Instability 基于轨迹计算运动不稳定性来判别异常行为, 遮挡严重时, 难以准确跟踪; 光流法忽略了运动目标在时间上的动态关联信息, 采用的是迭代的求解计算方法, 需要的计算时间比较长, 且受噪声影响较大; 支持向量机法在有遮挡的情况下

特征选取比较困难, 且不适用于多人突发事件的环境; 马尔科夫随机场只能反映目标在方向、速度大小上的简单变化, 在复杂运动场景中, 识别精确率不高. 实验结果表明, 本文算法具有较好的异常行为识别性能.

5 结束语

为解决现有算法易受环境影响, 不能实时、准确地识别异常行为的问题, 文中提出了一种基于骨架序列提取的异常行为识别算法. 用改进损失函数后的 YOLOv3 网络检测目标, 并通过 RT-MDNet 算法对其跟踪, 得到目标的运动轨迹后, 利用 OpenPose 提取目标的骨架序列, 最后应用时空图卷积神经网络结合聚类对异常行为进行识别. 实验结果表明, 本文算法能端到端地识别人员的异常行为, 识别精确率达 94%、处理速度达 18.25 fps, 满足实时性、准确性和鲁棒性的要求.

参考文献

1 Li H, Lu MJ, Hsu SC, *et al*. Proactive behavior-based safety management for construction safety improvement. Safety Science, 2015, 75: 107–117. [doi: 10.1016/j.ssci.2015.01.013]

2 Guo HL, Yu YT, Skitmore M. Visualization technology-based construction safety management: A review. Automation in Construction, 2017, 73(3): 135–144. [doi: 10.1016/j.autcon.2016.10.004]

3 安妙, 孔英会, 沈辉, 等. 基于深度学习的行为识别及在电力系统的应用. 电力科学与工程, 2019, 35(3): 59–65. [doi: 10.3969/j.ISSN.1672-0792.2019.03.009]

4 朱红岷, 戴道清, 李静正. 基于图像处理的变电站视频智能分析研究. 计算机工程与应用, 2018, 54(7): 264–270. [doi: 10.3778/j.issn.1002-8331.1611-0212]

5 Rao AS, Gubbi J, Marusic S, *et al*. Crowd event detection on optical flow manifolds. IEEE Transactions on Cybernetics, 2016, 46(7): 1524–1537. [doi: 10.1109/TCYB.2015.2451136]

6 Fuse T, Kamiya K. Statistical anomaly detection in human dynamics monitoring using a hierarchical Dirichlet process hidden Markov model. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(11): 3083–3092. [doi: 10.1109/TITS.2017.2674684]

7 杨天明, 陈志, 岳文静. 基于视频深度学习的时空双流人物动作识别模型. 计算机应用, 2018, 38(3): 895–899, 915. [doi: 10.11772/j.issn.1001-9081.2017071740]

8 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

9 Qiu ZF, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3D residual networks. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 5534–5542. [doi: 10.1109/ICCV.2017.590]

10 Zhang SY, Yang Y, Xiao J, *et al*. Fusing geometric features for skeleton-based action recognition using multilayer LSTM networks. IEEE Transactions on Multimedia, 2018, 20(9): 2330–2343. [doi: 10.1109/TMM.2018.2802648]

11 刘庭煜, 陆增, 孙毅锋, 等. 基于三维深度卷积神经网络的车间生产行为识别. 计算机集成制造系统, 2020, 26(8): 2143–2156. [doi: 10.13196/j.cims.2020.08.015]

12 Cao CQ, Lan CL, Zhang YF, *et al*. Skeleton-based action recognition with gated convolutional neural networks. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(11): 3247–3257. [doi: 10.1109/TCSVT.2018.2879913]

13 Liu J, Shahroudy A, Xu D, *et al*. Spatio-temporal LSTM with trust gates for 3D human action recognition. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 816–833. [doi: 10.1007/978-3-319-46487-9_50]

14 Sainath TN, Vinyals O, Senior A, *et al*. Convolutional, long short-term memory, fully connected deep neural networks. Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. South Brisbane: IEEE, 2015. 4580–4584. [doi: 10.1109/ICASSP.2015.7178838]

15 Liu J, Wang G, Duan LY, *et al*. Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Transactions on Image Processing, 2018, 27(4): 1586–1599. [doi: 10.1109/TIP.2017.2785279]

16 Wang LM, Xiong YJ, Wang Z, *et al*. Temporal segment networks: Towards good practices for deep action recognition. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 20–36. [doi: 10.1007/978-3-319-46484-8_2]

17 Tang YS, Tian Y, Lu JW, *et al*. Deep progressive reinforcement learning for skeleton-based action recognition. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5323–5332. [doi: 10.1109/CVPR.2018.00558]

18 Yang HY, Gu YZ, Zhu JC, *et al*. PGCN-TCA: Pseudo graph convolutional network with temporal and channel-wise

attention for skeleton-based action recognition. IEEE Access, 2020, 8: 10040–10047. [doi: 10.1109/ACCESS.2020.2964115]

19 Qian HF, Zhou X, Zheng MM. Detection and recognition of abnormal behavior based on multi-level residual network. Proceedings of 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference. Chengdu: IEEE, 2019. 2572–2579. [doi: 10.1109/IAEAC47372.2019.8997756]

20 包志强, 邢瑜, 吕少卿, 等. 改进 YOLO V2 的 6D 目标姿态估计算法. 计算机工程与应用, 2021, 57(9): 148–153. [doi: 10.3778/j.issn.1002-8331.2001-0367]

21 牟清萍, 张莹, 张东波, 等. 目标丢失判别机制的视觉跟踪算法及应用研究. 计算机工程与应用, 2021, 57(9): 140–147. [doi: 10.3778/j.issn.1002-8331.2001-0346]

22 Markovitz A, Sharir G, Friedman I, *et al*. Graph embedded pose clustering for anomaly detection. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10536–10544. [doi: 10.1109/CVPR42600.2020.01055]

23 Yan SJ, Xiong YJ, Lin DH. Spatial temporal graph convolutional networks for skeleton-based action recognition. Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 7444–7452.

24 Powers D. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2011, 2(1): 37–63.

25 Xie SY, Guan YP. Motion instability based unsupervised online abnormal behaviors detection. Multimedia Tools and Applications, 2016, 75(12): 7423–7444. [doi: 10.1007/s11042-015-2664-8]

26 杨志芳, 李乾. 基于骨骼关键点的异常行为识别及异构平台部署. 自动化与仪表, 2021, 36(11): 49–52. [doi: 10.19557/j.cnki.1001-9944.2021.11.011]

27 Xing ZY. Driver's intention recognition algorithm based on recessive Markoff model. Journal of Intelligent & Fuzzy Systems, 2020, 38(2): 1603–1614. [doi: 10.3233/JIFS-179524]

(校对责编: 牛欣悦)