

DOI: 10.13382/j.jemi.B2306967

基于关键帧定位的人体异常行为识别*

刘雨萌 桑海峰

(沈阳工业大学信息科学与工程学院 沈阳 110870)

摘要:近年来,基于视频的人体异常行为识别算法取得了一定的研究成果,但由于监控视频中存储的数据量庞大且视频时间跨度较长,在进行长视频或多行人异常动作检测与识别时,现有的识别方法并不适用。为此,提出了一种基于关键帧定位的人体异常行为识别模型,首先,通过基于标准化流和注意力增强时空图卷积的关键帧定位网络学习正常帧的概率分布,筛选和提取出长视频中的异常帧(关键帧)序列,并将其作为后续网络模型的输入。然后,为了更好地捕捉人体姿势的运动特征和异常情况,提出一种融合注意力和增强残差的时空图卷积异常行为识别算法,将关键帧序列输入到该模型网络中以实现监控视频中的人体异常行为的高效准确识别。使用公开数据集和自建数据集对该方法的有效性进行验证,实验结果表明,在公开数据集 ShanghaiTech Campus 上人体异常行为识别的 TOP-1 准确率达到 82.86%,TOP-5 准确率达到 98.10%,该方法可以更好的完成监控视频中的人体异常行为识别。

关键词: 异常行为识别;关键帧定位;标准化流;时空图卷积;注意力机制

中图分类号: TP391;TN98 **文献标识码:** A **国家标准学科分类代码:** 520.20

Human abnormal behavior recognition based on keyframes localization

Liu Yumeng Sang Haifeng

(School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: In recent years, video based human abnormal behavior recognition algorithms have achieved certain research results. However, due to the large amount of data stored in surveillance videos and the long time span of videos, existing recognition methods are not suitable for detecting and recognizing abnormal actions in long videos or multiple pedestrians. To this end, a human abnormal behavior recognition model based on keyframe localization is proposed. Firstly, a keyframe localization network based on standardized flow and attention enhanced spatial temporal graph convolution is used to learn the probability distribution of normal frames, filter and extract sequences of abnormal frames (keyframes) in long videos, and use them as inputs for subsequent network models. Then, in order to better capture the motion characteristics and abnormal situations of human posture, a spatial temporal graph convolutional abnormal behavior recognition algorithm that integrates attention and enhanced residuals is proposed. The keyframe sequence is input into the model network to achieve efficient and accurate recognition of human abnormal behavior in surveillance videos. Validate the effectiveness of this method using publicly available and self built datasets. The experimental results show that the TOP-1 accuracy of human abnormal behavior recognition on the publicly available dataset ShanghaiTech Campus reaches 82.86%, and the TOP-5 accuracy reaches 98.10%. This method can better complete the recognition of human abnormal behavior in surveillance videos.

Keywords: identification of abnormal behavior; keyframes localization; standardized flow; spatial temporal graph convolution; attention mechanism

0 引言

传统的视频监控模式存在一些问题,包括数据存储量大、依赖人工监控,并容易受到环境干扰等。然而,深度学习在此方面发挥了重要作用,许多研究人员已经将深度学习算法应用于人体异常行为识别任务中^[1]。根据使用的网络模型不同,基于深度学习的人体异常行为识别方法主要分为基于卷积神经网络的方法^[2]、基于循环神经网络的方法^[3]和基于图卷积神经网络的方法^[4-5]。这些方法旨在提高视频监控系统的智能化水平,以便更好地识别异常行为。由于人体骨架不受光照变化和杂乱背景等因素的影响,不少研究人员着重研究人体骨架。文献[6]首次提出基于骨架的动作识别网络模型 STGCN (spatial temporal graph convolutional network)。同时为了提升模型的网络性能,文献[7]在时空图卷积网络中引入注意力机制、文献[8]使用了多流自适应图卷积网络、文献[9]使用了通道级拓扑细化图卷积、文献[10]提出了一种解耦统一图卷积、文献[11]采用多分支时域结构等方法,提升异常动作识别的准确性以及泛化能力。然而,由于监控长视频中存在大量冗余的正常帧,处理每帧会增加模型计算量,影响推理速度。所以在进行异常行为识别前应该对视频中的异常部分着重处理。

目前,关键帧提取方法主要可以分为3类。第1类是基于视频帧聚类的方法^[12-13],通过将内容相似的视频帧自适应地进行分类,成功减少了冗余帧的数量。然而,在聚类过程中会导致时间信息的错乱。第2类是基于运动分析的关键帧提取方法^[14],利用关键关节点的运动特性来提取关键帧。这种方法对于四肢活动变化不明显的异常活动难以区分,并且容易受到遮挡等因素的影响。第3类方法是基于对视频帧评分的关键帧提取方法^[15],通过对每个帧进行评分来选择具有重要信息的关键帧。总体而言,这种方法效果较好,但整体模型的计算量较大。

综上所述,本文提出了一种基于关键帧定位的人体异常行为识别框架。通过对长视频中的关键帧进行高效定位,能够有效剔除冗余帧,从而获得异常帧序列。然后将这些异常帧输入到异常行为识别网络后可以高效迅速且准确地检测和识别异常行为。

本文的主要贡献如下:

1) 为了快速、高效的定位出长视频的异常部分,本文提出了一种基于标准化流和注意力增强时空图卷积的关键帧定位网络 ES-NF (enhanced spatial temporal graph convolutional network-normalizing flow),该网络直接学习正常样本的人体姿态图,得到正常样本的概率分布。通过对输入帧进行评分,利用阈值筛选出视频中的异常帧。

2) 提出了一种融合注意力和增强残差的时空图卷积异常行为识别算法 EA-STGCN (enhanced attention spatial temporal graph convolutional network),能更有效的提取行为的深层特征,更好的捕捉人体姿势的时空关系以及异常情况,从而可以在背景较为杂乱且人物不固定的监控场景下,实现对异常行为的精确识别。

1 基于关键帧定位的人体异常行为识别模型

1.1 整体模型结构

本文提出了一种基于关键帧定位的人体异常行为识别模型,可以实现对监控视频精准定位并检测出异常行为。如图1所示,该模型首先通过 Alphapose^[16]提取每帧视频中所有人体姿态序列。然后,利用本文提出的基于标准化流和注意力增强时空图卷积的关键帧定位网络 ES-NF,对每个人的姿态序列进行评分,再使用阈值筛选出异常姿态序列。最后使用本文提出的融合注意力和增强残差的时空图卷积异常行为识别网络 EA-STGCN,对异常姿态序列进行动作分类,从而实现对长视频中人体异常行为的定位与识别。

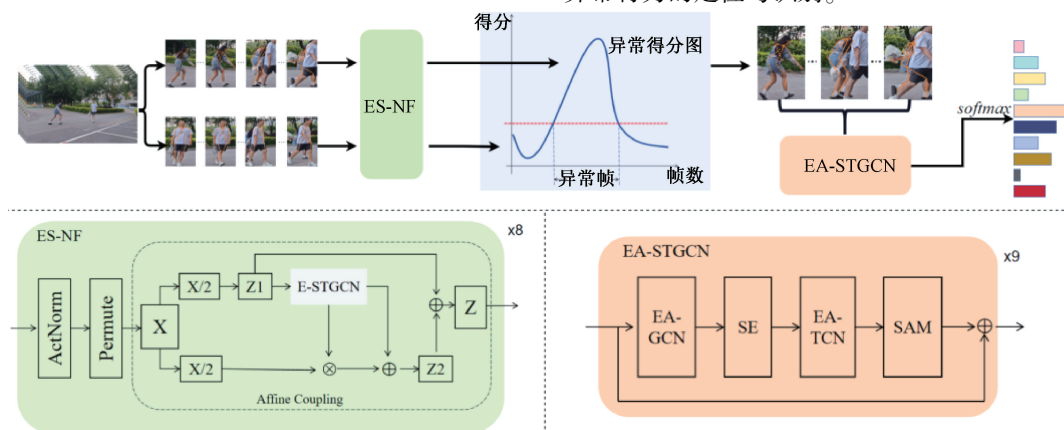


图1 基于关键帧定位的人体异常行为识别模型图

Fig. 1 Model diagram of human abnormal behavior recognition based on keyframe localization

1.2 基于标准化流和注意力增强时空图卷积的关键帧定位网络

标准化流 (normalizing flow)^[17] 是一种用于概率密度估计和生成模型的方法, 可以将一个复杂的后验分布 X 映射到一个简单的先验分布 Z , 从而实现对数据分布的建模。表达式如式(1)所示。

$$Z = F(X) \quad (1)$$

在本文中, 上式的 X 表示所有的骨架序列, Z 代表映射后的异常曲线得分, 目的在于将复杂的行为序列映射到异常得分。 F 为一系列变换函数 f_i 嵌套构成, 其中 i 表示标准化流的层数 (本网络模型采用 8 层标准化流), 具体的表达式如式(2)所示。

$$F = f_1 \times f_2 \times f_3 \times \cdots f_i \quad (2)$$

即标准化流变化过程如式(3)所示。

$$X \xleftrightarrow{f_1} h_1 \xleftrightarrow{f_2} h_2 \xleftrightarrow{f_3} h_3 \cdots \xleftrightarrow{f_i} Z \quad (3)$$

式(3)表示 X 和 Z 之间经过一系列变化是可以相互转换的, 其中 h_i 为输入 X 经过 i 次变换函数 f_i 变换后的结果。具体来说 h_i 和 f_i 的关系如式(4)所示。

$$h_{i+1} = f_{i+1} h_i \quad (4)$$

标准化流应用在关键帧提取中, 能够对正常帧的分布进行建模, 可以根据样本帧在原始数据空间中的概率密度判断其是否为异常帧, 能够有效地提取出最具代表性的关键帧, 并排除掉正常帧的干扰。

本文提出的基于标准化流和注意力增强时空图卷积的关键帧定位网络 ES-NF 可以将复杂的人体姿态映射到一个简单的异常行为得分。若此得分大于所设阈值, 即认为该人体姿态所处的视频帧为异常帧即关键帧。ES-NF 的详细结构图如图 2 所示。

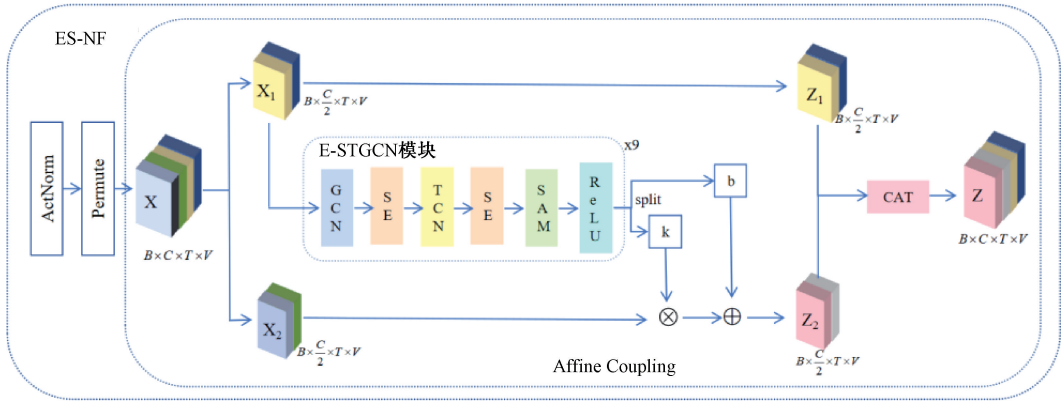


图 2 ES-NF 网络结构图
Fig. 2 ES-NF network structure diagram

与普通标准化流类似, 本文提出的 ES-NF 同样有归一化层 (ActNorm)、置换层 (Permute) 和耦合层 (Affine Coupling)。其中归一化层用于初始化权重和偏置, 并对输入进行归一化; 置换层使用可逆的 1×1 卷积对输入通道的顺序进行置换; 如图 2 所示, 耦合层首先将输入 X 按照通道维度拆分, 然后将其中一个送入到本文提出的融合注意力机制的时空图卷积网络 E-STGCN (enhanced spatial temporal graph convolutional network) 进行人体姿态特征提取, 将结果按照奇偶索引分割得到权重 k 和偏置 b , 从而对另一个输入进行仿射变换。具体实现如式(5)~(8):

$$X_1 = X_2 = X/2 \quad (5)$$

$$Z_1 = X_1 \quad (6)$$

$$Z_2 = k \times X_1 + b \quad (7)$$

$$Z = Z_1 + Z_2 \quad (8)$$

其中, X_1, X_2 是将输入特征 X 按照通道维度平均拆分成的两个特征, Z_1 和 Z_2 是输入 X_1 和 X_2 经过仿射变

换的输出, Z 将 Z_1 和 Z_2 按照维度拼接起来。

其中 E-STGCN 在 STGCN 的基础上添加了通道注意力 SE (squeeze-and-excitation)^[18] 和空间注意力 SAM (spatial attention module), 可以充分利用人体骨架的姿态信息, 使 STGCN 网络模型在时空数据建模中对时间和空间维度的关注更加准确和灵活, 能更好的区分正常和异常的人体姿态特征, 从而可以在 ES-NF 网络中区分正常和异常姿态序列。

标准化流中有一个可解的概率密度函数 pdf , 其中具体表达式如式(9)所示。

$$pdf(X) = pdf(Z) \times |det(dZ/dX)| \quad (9)$$

其中 $|det(dZ/dX)|$ 表示 Z 关于 X 的雅可比矩阵行列式的绝对值。在训练阶段采用最大化对数似然函数作为 ES-NF 网络的损失函数来优化参数, 对上式取对数函数且由式(1)~(4)可得到式(10)。

$$\log pdf(x) = \log pdf(z) + \sum_{i=1}^I \log |det(dh_i/dh_{i-1})| \quad (10)$$

1.3 融合注意力和增强残差的时空图卷积异常行为识别算法

通过 ES-NF 网络对所有人体姿态序列评分并进行阈值筛选,即可以得到视频中异常的人体姿态序列。再使用本文提出的融合注意力和增强残差的时空图卷积 EA-STGCN 对关键帧姿态序列进行异常行为识别。

如图 1 所示,本文提出的 EA-STGCN 由融合残差注意力增强的空间卷积网络 EA-GCN (enhanced attention-GCN)、通道注意力机制 SE、融合注意力的多分支时间卷积网络 EA-TCN (enhanced attention-TCN) 和空间注意力机制 SAM 组成。其中 EA-GCN 结构如图 3 所示,首先在空间卷积中引入动态图卷积神经网络,使邻接矩阵 A 能够自动学习节点与边的变化模式,并根据变化动态调整网络结构,从而更好地适应和捕捉数据中的动态特征。同时,在 GCN 内部引入注意力增强的残差结构,通过对通道进行压缩和激励,可以自适应的调整通道特征的权重,提高网络对重要特征的关注度,增强网络的表达能力,从而提高模型的异常行为识别准确性和泛化能力。

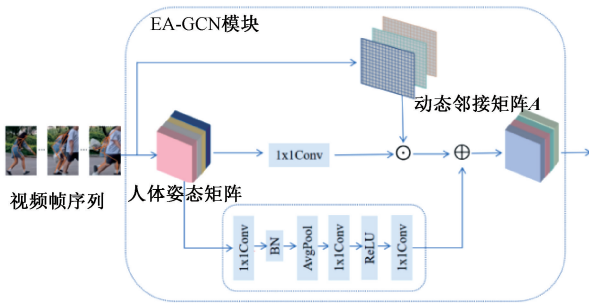


图 3 EA-GCN 网络模型

Fig. 3 EA-GCN network model

图 1 中 EA-TCN 结构如图 4 所示,为了在时间卷积 TCN 中提高对时序信息的建模能力,本文使用多分支卷积来替代原 TCN 网络中的 1D 卷积。本文使用 1×1 卷积将输入特征按通道 C 分成 6 个不同的分支,以通道数为 64 为例,经过 6 个不同的 1×1 卷积后通道数分别变为 2、4、8、16、16 和 18,再分别对每个分支应用不同的 3×1 卷积操作 ($d=1, 2, 3, 4$)、 3×1 最大池化,以及不进行操作。为了增强每个分支对特征的建模能力,在每个分支的卷积操作后添加了通道注意力模块 SE 来增强对于重要特征的表达。在进行维度转换时,本文采用两个卷积层和两个非线性激活函数进行维度转换,可以提取更多的特征信息,增强了网络的表达能力,有助于捕捉更复杂的特征。

通过 9 个累积的 EA-STGCN 模块对输入的异常姿态序列进行动作特征提取,网络前三层的输出通道为 64,中间三层的输出通道为 128,最后三层的输出通道为 256。最后使用 Softmax 分类器进行动作分类。

本阶段使用交叉熵 (Cross Entropy) 作为 EA-STGCN 网络的损失函数,通过计算样本的正常结果 y 和经过网络预测后的结果 y' 来表征两个概率分布之间的距离,对多标签分类任务中的一类任务来看具体的表达式如式 (11) 所示,网络的总的交叉熵为多标签分类任务中每一类的交叉熵之和。通过最小化交叉损失函数来训练模型。

$$\text{Loss} = -y \log y' - (1 - y) \log (1 - y') \quad (11)$$

本文提出的整体模型在训练过程中采用两阶段方式训练,关键帧网络和异常行为识别网络单独分开训练,采用各自的损失函数去训练网络。在测试阶段是采用整体模型直接进行测试。

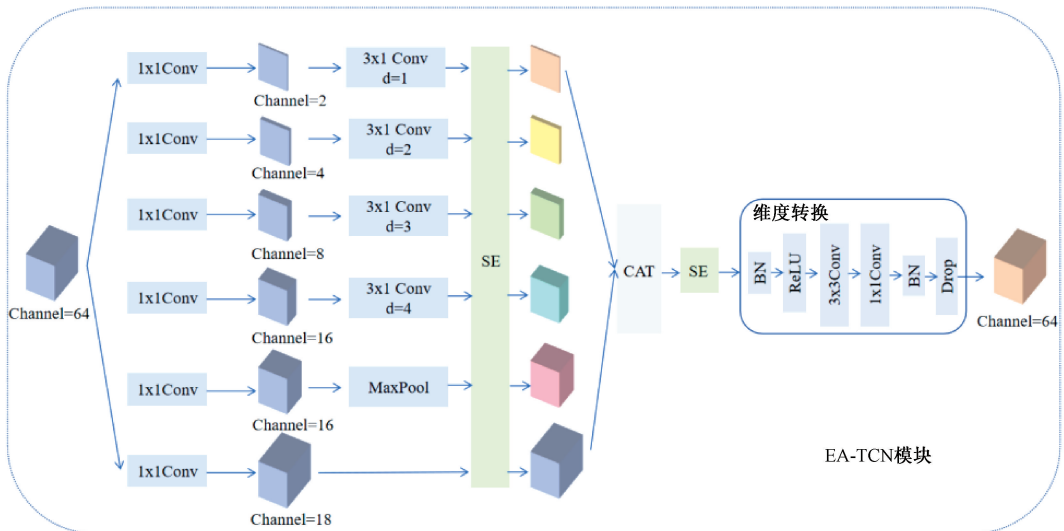


图 4 EA-TCN 网络模型

Fig. 4 EA-TCN network model

2 实验结果与分析

2.1 数据集与评价指标

Shanghaitech Campus 数据集^[19]是目前最全面、最真实的异常行为检测数据集。数据集中包含 330 段训练样本和 107 段测试样本。训练样本仅由正常行为视频样本组成,测试样本由正常行为视频和异常行为视频样本组成,同时对异常行为样本进行帧级标注。视频分辨率为 856×480,包含上海科技大学校园内的 13 个背景,有 130 个异常事件。如图 5 所示,为 Shanghaitech Campus 数据集样例。



图 5 Shanghaitech Campus 数据集样例
Fig. 5 Sample Shanghai Campus dataset

ShanghaiTech Campus 数据集的训练集中无异常行为,所以无法直接在人体异常行为识别网络中训练,为了解决这个问题,本文对 107 个测试视频中的异常帧进行重组分类,分成跑、翻越栅栏、高空抛物、追逐推搡、立定跳远、拿着武器、原地徘徊、抢东西、摔倒以及滑行骑行共 10 个类别,共计 459 段异常视频样本。将训练数据与测试数据按照 2 : 1 的比例进行分配,共有 306 个训练视频和 153 个测试视频样本。由于上述异常动作类别中追逐推搡、立定跳远、抢东西以及摔倒的训练样本数据有限,容易导致模型过拟合。因此,本文对上述 4 种异常行为进行数据扩充,从 NTU-RGB+D 数据集^[20]中为每种异常行为扩充了 300 个视频作为训练样本。同时还从自建数据集中提取了其他异常行为样本进行训练,最终训练样本数据为 1 660 个视频,测试视频为 153 个。将上述修改后的数据集称为 ShanghaiTech Campus & NTU 数据集。

此外,本文模拟 Shanghaitech Campus 拍摄环境建立自建数据集视频,自建数据集中异常行为包括跑、推倒、追逐推搡、高空抛物、原地徘徊、立定跳远和抢东西 7 种异常行为。同时,每段视频的时长不限,如图 6 为本文的部分自建数据集样例:

本文主要使用了 Top-1 准确率(即测试样本中真实标签即为第一个预测结果的概率)和 Top-5 准确率(即测试样本中真实标签包含在前 5 个预测结果之内的概率)这两个指标来评估模型的性能。

2.2 实验设置

本文实验环境为 window11,处理器为 12th Gen Intel (R) Core(TM) i5-12600K 3.70 GHz,内存 16 G,显卡为



图 6 自建数据集样例
Fig. 6 Sample self built dataset

NVIDIA GeForce RTX 3070 Ti。编程语言为 Python,开发工具为 PyCharm,使用 Pytorch 框架进行算法网络搭建、训练和测试,采用随机梯度下降优化策略。在关键帧定位网络 ES-NF 中,初始学习率设置为 0.000 5,权重衰减率设置为 0.000 05,训练轮数 epoch = 20,其中 batchsize 大小为 16,在人体异常行为识别网络 EA-STGCN 中初始学习率设置为 0.1,权重衰减率设置为 0.000 5。并在上述数据集和实验环境中进行了 200 个 epoch 的训练,其中 batchsize 大小设置成 64。

2.3 结果与分析

1) 整体结果

为验证 EA-STGCN 网络模型的收敛能力,在 ShanghaiTech Campus & NTU 数据集上与其他图卷积的基线模型进行对比,图 7 给出了模型损失函数值的变化曲线,纵坐标为损失函数值,横坐标为迭代次数。从图 7 中可以看到,本文模型较其他基线模型具有更快的收敛速度以及稳定性。

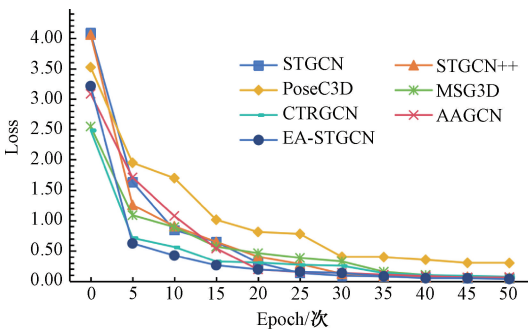


图 7 损失值变化曲线

Fig. 7 Loss value variation curve

同时,在本文重新标注过异常动作类别的

ShanghaiTech Campus 数据集上,将本文异常动作识别方法与其他基于骨架或图卷积的基线模型进行对比,记录并保存了最佳结果,结果如表 1 所示。本文提出的 EA-STGCN 模型的 Top-1 准确率达到 64.10%,Top-5 的准确率达到 93.16%,可能是因为 ShanghaiTech Campus 数据集中人员不固定,导致难以提取出重要特征,导致准确率较低,但较其他基线模型均有较大提升。同时将本文提出的关键帧提取模型 ES-NF 应用于其他骨架或图卷积的基线模型中,与本文所提出的基于关键帧定位的人体异常行为识别网络模型进行对比,从表 1 中可以看出,无论使用何种模型方法,在使用关键帧(ES-NF)提取方法后,其 Top-1 和 Top-5 的准确率均明显高于未使用关键帧提取方法,这说明使用关键帧(ES-NF)提取方法对于异常行为识别具有重要意义。这些结果表明,本文的方法在背景杂乱且人物不固定的监控场景下,能够实现对

异常行为的精确识别。同时本文方法在使用关键帧之后 Top-1 的准确率均高于其他方法,证明了本文整体的优越性。

表 1 本文方法与其他基线模型对比结果
Table 1 Comparison results between this method and other baseline models

模型方法	未使用关键帧(ES-NF)		使用关键帧(ES-NF)	
	提取方法		提取方法	
	TOP-1/%	TOP-5/%	TOP-1/%	TOP-5/%
PoseC3D	52.99	70.94	58.10	80.00
STGCN	54.70	88.03	78.10	93.33
MSG3D	54.70	90.60	72.38	95.24
STGCN++	59.83	92.31	80.95	99.05
CTRGCN	54.70	88.89	77.14	95.24
AAGCN	53.85	84.62	69.52	94.29
EA-STGCN	64.10	93.16	82.86	98.10

表 2 EA-STGCN 消融实验
Table 2 EA-STGCN ablation experiment

模型	Modules					ShanghaiTech		ShanghaiTech & NTU	
	GCN	EA-GCN	TCN	EA-TCN	SE+SAM	TOP-1/%	TOP-5/%	TOP-1/%	TOP-5/%
本文 1	√		√			59.83	92.31	69.59	96.62
本文 2		√	√			61.19	90.60	70.35	97.97
本文 3	√			√		62.39	92.31	70.92	95.27
本文 4		√		√		63.68	94.02	72.30	94.59
本文		√		√	√	64.10	93.16	73.65	95.95

2) 消融实验

为了验证本文所提出模块的有效性,设计了一系列的消融实验,如表 2 所示,对本文提出的 EA-STGCN 内部各模块进行消融实验,验证原始的 GCN、TCN、和改进后如图 1 所示的 EA-GCN、EA-TCN 以及 SE+SAM 的有效性。由表 2 可知,同时使用 EA-GCN、EA-TCN 和 SE+SAM 时可达到最优效果。

同时为了验证本文提出的关键帧定位网络(ES-NF)的有效性,使用 ShanghaiTech Campus 数据集,评估指标为 AUC(ROC 曲线下的面积)。ROC 曲线的横坐标为伪阳性率 FPR,纵坐标为真阳性率 TPR,具体关系表达式如式(12)、(13)所示。

$$FPR = FP / (FP + TP) \tag{12}$$

$$TPR = TP / (TP + FN) \tag{13}$$

其中,TP 为正确判断为阳性;FP 为错误判断为阳性;FN 为错误判断为阴性;TN 为正确判断为阴性。AUC 的值越大,表示模型性能越好,ES-NF 的 ROC 曲线如图 8 所示,本文提出的 ES-NF 网络的 AUC 可以达到 85.31%,且 ES-NF 的参数量为 1.448 k,实现了模型的轻量化。

当 ES-NF 网络预测分数超过阈值时,作为 FP(异常帧),否则作为 TP(正常帧);随着阈值降低,FP 会增加,模型更保守,从图 8 中可以看到,有多组不同阈值下的

FPR 和 TPR,本文整体模型是对人体异常行为进行识别,所以关键帧定位网络的模型可以较为保守,提取出稍多的关键帧输入到后续网络中。通过对 ShanghaiTech Campus 数据集训练,可得到不同阈值下识别准确的样例个数及识别错误的样例个数,结合 ROC 曲线,设定 threshold 为 1.197 79。

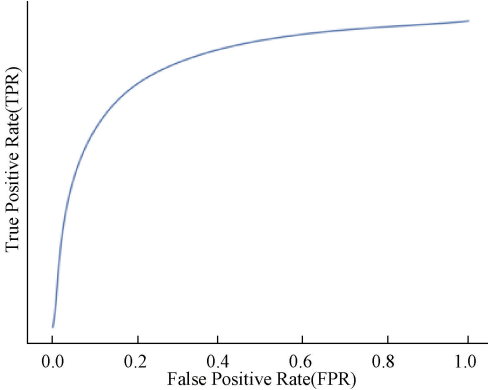


图 8 ES-NF 网络的 ROC 曲线
Fig. 8 ROC curve of ES-NF network

3) 自建数据集上异常行为识别实验

本文还在自建数据集中进行测试,其中每个类别有

50 个视频片段,使用本文方法检测与识别异常行为,具体的检测结果如表 3 所示。其在自建数据集中的平均准确率达到 92.00%,然而,追逐推搡和抢东西两个类别的准确率相对较低,可能是这两种动作都伴随着胳膊和肘部的运动,导致他们之间非常相似,很容易造成误检,进而影响到识别的准确性。

表 3 自建数据集识别结果

Table 3 Self built dataset recognition results			
异常动作类型	识别正确个数	未识别正确个数	TOP-1/%
跑	46	4	92.00
推倒	49	1	98.00
追逐推搡	42	8	84.00
高空抛物	48	2	96.00
原地徘徊	45	5	90.00
立定跳远	48	2	96.00
抢东西	44	6	88.00

3 结 论

本文提出了一种基于关键帧定位的人体异常行为识别网络模型用于对监控视频等较长视频进行快速的异常动作定位与识别。该模型包括基于标准化流和注意力增强时空图卷积的关键帧定位网络 ES-NF,通过对每个人体姿态序列进行异常评分,再通过阈值定位出所有关键帧序列。之后使用本文提出的一种融合注意力机制和增强残差单元的时空图卷积网络的人体异常行为识别算法 EA-STGCN 进行异常动作的分类与识别。该方法适用于背景较为杂乱且人物不固定的监控场景,能实现对异常行为的精确识别。实验证明,整体模型在 ShanghaiTech Campus 数据集上的识别准确率 (Top-1 和 Top-5) 明显优于其他骨架识别方法。且在自建数据集中 Top-1 的准确率可达 92%,然而,在相似异常行为的识别准确率和实时性方面仍有改进的空间,需要进一步研究和扩展。

参考文献

[1] 朱红蕾,卫鹏娟,徐志刚. 基于骨架的人体异常行为识别与检测研究进展[J/OL]. 控制与决策, 2023;1-18. <https://doi.org/10.13195/j.kzyjc.2023.0451>.
ZHU H L, WEI P J, XU Z G. Research progress on skeletal based recognition and detection of human abnormal behavior [J/OL]. Control and decision-making, 2023; 1-18.. <https://doi.org/10.13195/j.kzyjc.2023.0451>.
[2] DUAN H, ZHAO Y, CHEN K, et al. Revisiting skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition,2022: 2969-2978.
[3] DUA N, SINGH S N, SEMWAL V B. Multi-input CNN-GRU based human activity recognition using wearable sensors[J]. Computing, 2021, 103: 1461-1478.
[4] 建中华,南静,刘鑫,等. 基于时空张量融合的人体骨架行为自适应识别方法[J]. 仪器仪表学报,2023, 44(6):74-85.
JIAN ZH H, NAN J, LIU X, et al. An adaptive recognition method for human skeleton behavior based on spatial temporal tensor fusion [J]. Chinese Journal of Scientific Instrument, 2023,44 (6): 74-85.
[5] 陶峰,李燕苹,王瑞. 基于图卷积网络改进的人体动作识别模型[J]. 电子测量技术,2023,46(8):59-64.
TAO F, LI Y P, WANG R. An improved human motion recognition model based on graph convolutional networks [J]. Electronic Measurement Technology, 2023, 46 (8): 59-64.
[6] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the AAAI Conference on Artificial Intelligence,2018.
[7] 张蔚澜,齐华,李胜. 时空图卷积网络在人体异常行为识别中的应用[J]. 计算机工程与应用,2022,58(12): 122-131.
ZHANG W L, QI H, LI S. Application of spatial temporal graph convolutional networks in human abnormal behavior recognition [J]. Computer Engineering and Applications, 2022,58 (12): 122-131.
[8] SHI L, ZHANG Y, CHENG J, et al. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks[J]. IEEE Transactions on Image Processing, 2020, 29: 9532-9545.
[9] CHEN Y, ZHANG Z, YUAN C, et al. Channel-wise topology refinement graph convolution for skeleton-based action recognition [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 13359-13368.
[10] LIU Z, ZHANG H, CHEN Z, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 143-152.
[11] DUAN H, WANG J, CHEN K, et al. Pyskl: Towards good practices for skeleton action recognition [C]. Proceedings of the 30th ACM International Conference on Multimedia, 2022: 7351-7354.
[12] 陶永辉,王勇. 基于改进的 K-means 风电机异常数据检测[J]. 国外电子测量技术,2023,42(4):141-148.

- TAO Y H, WANG Y Detection of abnormal data in wind turbines based on improved K-means [J]. Foreign Electronic Measurement Technology, 2023, 42 (4): 141-148.
- [13] TANG H, LIU H, XIAO W, et al. Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion [J]. Neurocomputing, 2019, 331: 424-433.
- [14] 刘嘉宇, 陈平. 基于关键帧定位和时空图卷积的异常行为识别[J]. 机械与电子, 2022, 40(1): 48-53, 58.
LIU J Y, CHEN P. Anomalous behavior recognition based on keyframe localization and spatial temporal map convolution [J]. Journal of Mechanics and Electronics, 2022, 40 (1): 48-53, 58.
- [15] NIKPOUR B, ARMANFARD N. Spatio-temporal hard attention learning for skeleton-based activity recognition[J]. Pattern Recognition, 2023, 139: 109428.
- [16] FANG H S, XIE S, TAI Y W, et al. RMPE: Regional multi-person pose estimation [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334-2343.
- [17] 杜梓辉. 基于标准化流的点云处理学习网络的算法研究[D]. 广州: 华南理工大学, 2022.
DU Z H Research on algorithms for point cloud processing learning networks based on standardized streams [D]. Guangzhou: South China University of Technology, 2022.
- [18] 张海超, 张闯. 融合注意力的轻量级行为识别网络研究[J]. 电子测量与仪器学报, 2022, 36(5): 173-179.
ZHANG H C, ZHANG C. Research on lightweight behavior recognition networks integrating attention [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (5): 173-179.
- [19] LIU W, LUO W X, LIAN D, et al. Future frame prediction for anomaly detection-a new baseline[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018. DOI: 10. 1109/ CVPR. 2018. 00684.
- [20] SHAHROUDY A, LIU J, NG T T, et al. Ntu rgb+ d: A large scale dataset for 3d human activity analysis [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

作者简介



刘雨萌, 2021 年于沈阳工业大学获得学士学位, 现为沈阳工业大学硕士研究生, 主要研究方向为异常动作识别、深度学习与图像处理。

E-mail: lym19990424@163.com

Liu Yumeng received her B. Sc. degree from Shenyang University of Technology in 2021. Now she is a M. Sc. candidate at Shenyang University of Technology. Her main research interests include recognition of abnormal human behavior, deep learning and image processing.



桑海峰(通信作者), 2000 年于东北师范大学获得学士学位, 2003 年于东北师范大学获得硕士学位, 2006 年于东北大学获得博士学位, 现为沈阳工业大学教授, 主要研究方向为智能视频分析、机器视觉检测与图像识别、无人驾驶之环境感知技术、深度学习技术。

E-mail: sanghaif@163.com

Sang Haifeng (Corresponding author) received his B. Sc. degree from Northeast Normal University in 2000, M. Sc. degree from Northeast Normal University in 2003 and Ph. D. degree from Northeastern University in 2006, respectively. Now he is a professor in Shenyang University of Technology. His main research interests include intelligent video analysis, machine vision detection and image recognition, autonomous driving environment perception technology, and deep learning technology.