



DOI:10.12404/j.issn.1671-1815.2406479

引用格式:许辰月,王蓉,郭放,等.改进时空图归一化流的异常行为识别方法[J].科学技术与工程,2025,25(18):7693-7699.

Xu Chenyue, Wang Rong, Guo Fang, et al. Abnormal behavior recognition method of improved spatio-temporal graph normalizing flow[J]. Science Technology and Engineering, 2025, 25(18): 7693-7699.

改进时空图归一化流的异常行为识别方法

许辰月,王蓉,郭放,曾昭龙

(中国人民公安大学信息安全学院,北京 100038)

摘要 针对异常行为识别中人体动态骨架特征表达能力不充分的问题,提出了一种基于改进时空图归一化流的无监督异常行为识别方法,利用 Transformer 和卷积块注意力模块,在全局域和时空域中提高模型的特征表达能力,提升异常行为识别算法性能。首先,将 Transformer 模块引入归一化流的仿射层,在全局层面增强动态骨架特征信息的有效性;然后,分别在空间与时间图卷积模块中引入卷积注意力,有效地提升动态骨架特征的空间和时间表达能力;最后,在 ShanghaiTech 数据集和 UB-normal 数据集上进行仿真验证,识别精确度分别达到 86.4% 和 70.2%,证明了方法的有效性。

关键词 异常行为识别;时空图卷积;归一化流;动态骨架特征

中图分类号 TP391.4; 文献标志码 A

Abnormal Behavior Recognition Method of Improved Spatio-temporal Graph Normalizing Flow

XU Chen-yue, WANG Rong, GUO Fang, ZENG Zhao-long

(School of Information Network Security of People's Public Security University of China, Beijing 100038, China)

[Abstract] In order to solve the problem of insufficient feature extraction of human dynamic skeleton features in abnormal behavior recognition, an unsupervised abnormal behavior recognition method based on enhanced spatiotemporal graph normalization flow was proposed. Transformer and convolution block attention module were employed to enhance the feature expression capability of the model and the performance of the abnormal behavior recognition algorithm in the global and spatiotemporal domains. Firstly, the Transformer module was incorporated into the affine layer of the normalized flow to augment the efficacy of dynamic skeleton feature information at the global level. Subsequently, the convolution attention was introduced into the convolution module of space and time graphs respectively to effectively enhance the spatial and temporal representation of dynamic skeleton features. Finally, simulation verification was conducted on the ShanghaiTech and UBnormal datasets, and the recognition accuracy attains 86.4% and 70.2% respectively, thereby demonstrating the effectiveness of the method.

[Keywords] anomalous action recognition; spatio-temporal graph convolution; normalizing flow; dynamic skeleton features

人体异常行为识别是计算机视觉领域的一个重点研究方向,这一研究方向的核心是将包含行人异常动作的视频帧识别出来,并加以标识。异常行为识别技术有广泛的应用空间,如街道安全视频监控、考场视频监控以及养老院视频监控等,能够大大节省人力物力,有极大的实际应用价值。

人体异常行为识别算法多基于监督学习的基础进行研究,这些方法大致可以分为三类。第一类是基于支持向量机的异常行为识别方法,例如,邵宝平等^[1]提出,通过支持向量机机制结合长短期记忆网络,构建的双层识别模型分辨正常和异常行

为。第二类是基于卷积神经网络的异常行为识别算法,神经网络作为分类器可以运用于单分类和多分类问题,在细分异常行为种类方面应用较多,例如,王新文等^[2]提出了一种基于双重残差卷积网络的跌倒识别方法,通过残差网络的嵌套,在避免模型训练时产生梯度消失和过拟合等问题的同时充分融合了深层和千层视觉特征。姬晓飞等^[3]提出了一种基于增强特征的人体检测与异常行为识别融合算法,将卷积神经网络提取到的人体外观特征与人体热力图特征融合得到增强特征,后送入 Transformer 网络进行异常行为识别。第三类是基于

收稿日期:2024-08-29 修订日期:2025-04-01

基金项目:中央高校基本科研业务费专项资金(2024JKF11);中国人民公安大学安全防范工程双一流专项(2023SYL08)

第一作者:许辰月(1999—),女,汉族,安徽淮南人,硕士研究生。研究方向:异常行为识别。E-mail:2697378920@qq.com。

*通信作者:王蓉(1971—),女,汉族,辽宁丹东人,博士,教授。研究方向:模式识别。E-mail:dbdxwangrong@163.com。

统计方法的异常行为识别方法,彭月平等^[4]应用粗糙集理论对原始数据集中信息进行提取和分类,提升异常行为识别及分类算法精确度。

现实中的视频数据量大,且异常行为种类繁多多变,现有异常行为标注数据集数量较少、质量较低,因此基于无监督学习的人体异常行为识别算法研究就显得十分必要。通常基于无监督学习的人体异常行为识别算法多采用聚类算法作为分类器,将异常行为与正常行为进行区分,例如,Nady等^[5]利用梯度时空自相关描述符,提取视频帧中的时空运动特征,接着使用K-medoids聚类技术对提取出的特征进行分类。Cho等^[6]提出了一个隐式的双路径自动编码器,隐式地建模外观和运动特征学习正常的视频模式,基于归一化流的生成模型对提取特征进行正态密度估计,并使用非分布外检测来识别异常。上述方法均以视频帧作为输入对视频中人体进行异常行为识别,在这些方法中环境的变化对算法的鲁棒性有一定的影响,所以另一部分异常行为识别算法以人体姿态估计图作为输入,通过先预处理视频帧数据,提取人体骨架图作为异常行为识别方法的输入,有效规避了背景、光线以及服装不同而产生的噪声。Markovitz等^[7]使用了聚类技术,首先提取了人体姿态图,并映射到空间作为时空特征,后采用深度嵌入式聚类技术对提取出的时空特征进行计算。Sato等^[8]利用了预训练的骨架特征提取器,并将提取骨架特征与用户提示在公共空间域对齐设置异常分数。Zaheer等^[9]提出无监督生成性合作学习(unsupervised generative cooperative learning, GCL),通过构建生成器与判别器之间的交叉监督进行异常行为识别Hirschorn等^[10]首次将归一化流技术运用到人体异常行为识别领域,提出了时空图归一化流模型(special temporal graph normalizing flow, STG-NF),提

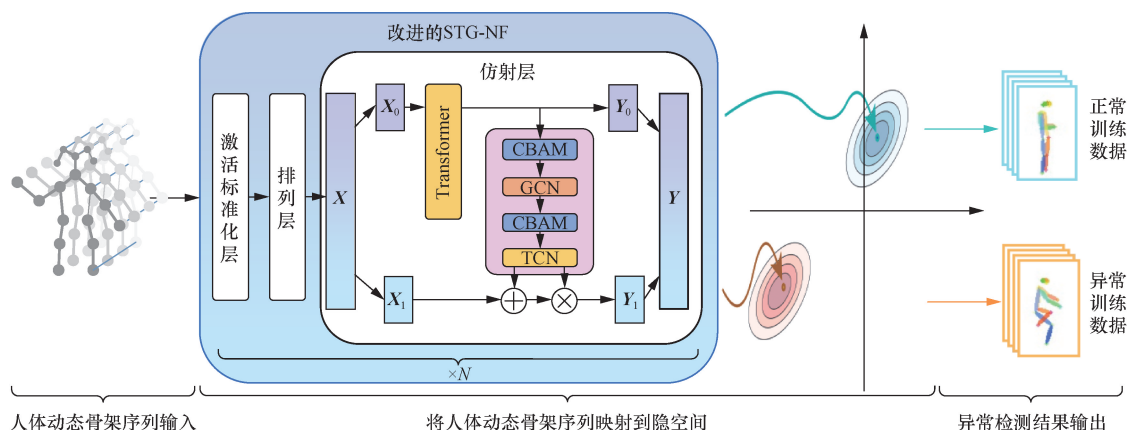
取视频中人体的姿态图作为特征,进行异常行为识别,在识别效果上也有着不俗的表现。

STG-NF实现了计算的简便化和系统设计的轻量化,然而在特征提取的全面性方面仍有优化的空间。因此,在STG-NF模型基础上,现进行改进,提出一种基于多注意力机制与归一化流的人体异常行为识别方法。首先,将Transformer网络引入归一化流,优化全局特征的提取。其次,在时空图卷积网络的空间图卷积网络前添加卷积块注意力模块,增强空间特征的表达性。最后,在时空图卷积网络的时间图卷积网络前添加卷积块注意力模块,增强时间特征的表达性。

1 模型结构

针对STG-NF在提取特征方面进行优化,改进后模型能够从人体姿态图序列中提取到更加完善的动态骨架全局和时空特征,有效提升了异常行为识别的精度。模型具体流程如图1所示。

模型使用Alphapose人体骨架提取模型^[11]提取人体姿态图后送入改进的时空图归一化流模型中。在每一个时空归一化流步长中,从视频帧中提取出的人体姿态图序列输入至激活归一化层,对数据进行预处理,对原始参数进行初始化和批量标准化。然后通过排列层重新排列输入通道顺序,并送入仿射层。在仿射层中,输入的数据按照通道数划分为两部分, X_0 保持不变,而 X_1 则基于 X_0 进行仿射变换。 X_0 首先经过Transformer层进行动态骨架的全局特征提取,增强动态骨架特征的全局信息,然后进入改进后的时空图卷积模块,通过双层的卷积注意力模块(convolutional block attention module, CBAM)^[12],增强动态骨架特征的时空信息,一同提升了特征的表达能力。将提取出的特征向量通过矩



CBAM 为双层的卷积注意力模块;GCN 为图卷积模块;TCN 为时空卷积模块

图1 改进后的STG-NF模型总体结构图

Fig.1 Improved overall structure of STG-NF

阵加法与矩阵乘法与 X_1 结合得到 Y_1 , 将 Y_0 和 Y_1 连接得到最终的特征向量 Y 。通过高斯变换将 Y 映射到隐空间, 归一化为标准分布, 通过设定的阈值与计算出的视频帧分数比较, 进而区分出包含异常行为的视频帧。

1.1 基于 Transformer 的归一化流

将 Transformer 模型^[13]引入归一化流仿射层, 在分离出的特征进入时空图卷积模型之前, 先进入 Transformer 模块进行特征优化, 有效提升了信息提取的有效性, 优化了全局的特征提取, 对异常行为识别模型的精度优化有一定的提升效果。

1.1.1 归一化流

归一化流^[14]是一种无监督的密度估计模型, 定义为可逆映射 $f: X \rightarrow z$, 即从数据空间 X 到隐空间 z 。函数 f 是使用一个由参数 θ 参数化的神经网络来实现的。

$$z = f_{\theta}(x) = g_{\theta}^{-1}(x) \quad (1)$$

即

$$x = g_{\theta}(z) \quad (2)$$

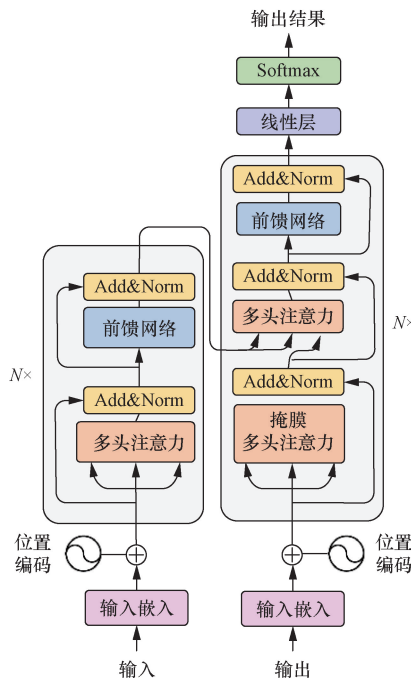
通过 K 个 f 函数相乘, 可以得到可逆的变化序列, 即归一化流模型。

$$f = f_1 \circ f_2 \circ \dots \circ f_K \quad (3)$$

它由一个神经网络参数化, 其结构旨在确保雅可比矩阵行列式的可逆性和有效的计算。

1.1.2 Transformer 注意力机制

Transformer 的整体结构如图 2 所示。对编码器和解码器使用堆叠的自注意力层和点级的全连接层,



Add&Norm 表示层归一化; $N \times$ 表示有 N 个模块组成

图 2 Transformer 结构图

Fig. 2 Structure of Transformer

图的左半部分和右半部分分别展示了 Transformer 模型中编码器和解码器中自注意力层等卷积层的堆叠方式。多头注意力机制是 Transformer 网络中的重要模块, 不同于单头注意力机制对查询序列、键和值做单一注意力卷积操作, 多头注意力机制将查询序列、键和值的计算并行化, 用学习到的不同的线性投影对查询序列、键和值进行映射, 并在每个查询序列、键和值的映射过程中并行地执行注意力卷积操作, 最终输出 d_v 维的输出向量。

多头注意力机制使得模型可以连带地关注到不同表示子空间的不同部位的信息, 而单头注意力机制无法做到这一点。多头注意力机制的具体操作过程用公式为

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_2, \text{head}_h) \mathbf{W}^0 \quad (4)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q} \mathbf{W}_i^Q, \mathbf{K} \mathbf{W}_i^K, \mathbf{V} \mathbf{W}_i^V) \quad (5)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别为查询向量、键向量和值向量; \mathbf{W} 为注意力权值矩阵。

其中参数矩阵代表含义为

$$\mathbf{W}_i^Q \in \mathbf{R}^{d_{\text{model}} \times d_k} \quad (6)$$

$$\mathbf{W}_i^K \in \mathbf{R}^{d_{\text{model}} \times d_k} \quad (7)$$

$$\mathbf{W}_i^V \in \mathbf{R}^{d_{\text{model}} \times d_v} \quad (8)$$

放缩点积注意力是 Transformer 网络中最基础的组成环节, 在放缩点积注意力机制中, 模型的输入是维度为 d_k 的查询序列和键, 以及维度为 d_v 的值组成, 模型对每一组查询序列和键进行点积运算后除以 $\sqrt{d_k}$, 接着使用 Softmax 函数来获取这些值的权重。放缩点积注意力机制放弃了权重矩阵, 直接建立查询序列与键之间的映射关系, 不需要参数, 降低了模型的复杂性, 由使用放缩防止了数据溢出, 在实际应用上提升了模型的计算速度。具体计算过程为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V} \quad (9)$$

1.1.3 基于 Transformer 的归一化流

依据 Transformer 结构能够很好地提取全局特征, 关注到远距离特征之间的相互影响, 本文研究将 Transformer 结构引入归一化流的仿射层, 将分流后的特征送入 Transformer 层进行使用全局注意力, 进行全局动态骨架特征增强, 有效地增加了模型的特征表达能力。具体结构如图 3 所示。

1.2 空间特征提取与优化

1.2.1 改进的空间特征提取模块

利用 CBAM 的特性将其与 ST-GCN 模块结合起来, 提升 ST-GCN 网络的特征提取能力, 改进后的 ST-GCN 网络结构如图 4 所示。

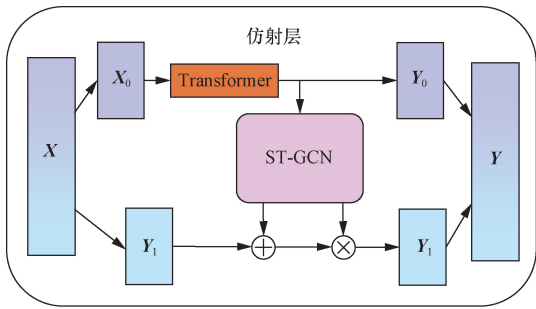


图3 改进后的归一化流结构图

Fig. 3 Structure of improved normalizing flow

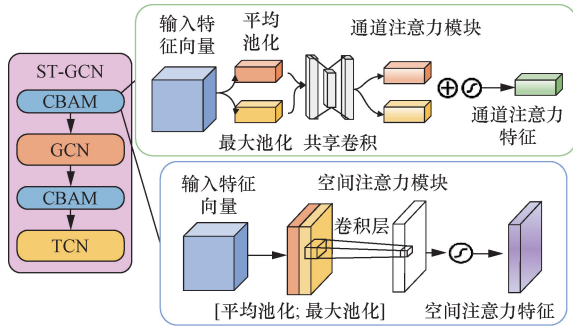


图4 改进后的时空图卷积模块结构图

Fig. 4 Structure of improved spatio-temporal convolution

为了提升动态骨架特征的空间信息表达能力,提出了改进的时空图卷积模块,首先在图卷积(graph convolution network, GCN)加入了一个CBAM卷积注意力模块,使用CBAM模块对动态骨架特征的通道和空间两个维度进行重点特征聚焦,达到提升特征空间信息表达能力的目的。

1.2.2 CBAM卷积注意力

CBAM作为一个轻量级的前馈神经网络自注意力模块,可以灵活地插入各类神经网络结构中,进行端到端的训练,具体结构如图5所示。

CBAM模块同时在通道和空间两个维度进行注意力图推断,然后与输入特征图进行乘法运算,进行自适应的特征细化操作。

在CBAM模块中,输入特征向量 $F \in \mathbf{R}^{b \times c \times t \times v}$,其中 b 为批次大小, c 为通道数, t 和 v 分别为时间维度帧数和空间维度节点数。该特征向量首先进入通

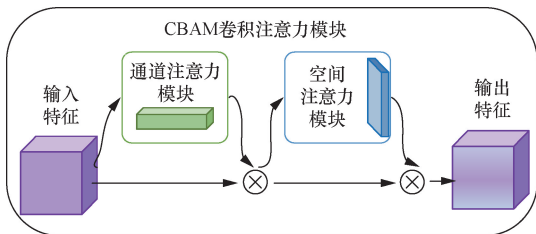


图5 CBAM卷积注意力模块结构图

Fig. 5 Structural diagram of CBAM convolutional attention module

道注意力模块。在通道注意力模块中,输入特征首先并行通过平均池化操作和最大池化操作聚合空间信息,得到的描述符分别为 F_{avg}^c 和 F_{max}^c ,这两个描述符在由多层感知器组成的共享网络中生成通道注意力图 $M_c \in \mathbf{R}^{c \times 1 \times 1}$,在共享网络作用于每一个描述符后,使用求和运算进行向量合并,得到输出特征向量。具体运算步骤的公式为

$$M_c(F) = \sigma \{ \text{MLP} [\text{AvgPool}(F)] \} +$$

$$\text{MLP} (\text{MaxPool}(F)) = \sigma (W_1 \{ W_0 [F_{avg}^c(x)] \}) +$$

$$W_1 \{ W_0 [F_{max}^c(x)] \} \quad (10)$$

式(10)中: σ 为 sigmoid 函数; $W_0 \in \mathbf{R}^{b \times c \times t \times v}$; $W_1 \in \mathbf{R}^{c \times c/r}$; W_0 和 W_1 分别为共享网络的权重; AvgPool 为平均池化操作; MaxPool 为最大池化操作。

通道注意力模块输出的特征向量将进入空间注意力模块进行进一步的特征提取工作,空间注意力模块作为通道注意力模块的补充,更加关注输入特征的空间维度信息提取。首先串行合并了平均池化和最大池化操作用于生成特征描述符,并应用卷积计算生成空间注意力图 $M_s(F) \in \mathbf{R}^{H \times W}$ 对重点部分进行聚焦。具体操作为分别通过两种池化操作生成两张二维的注意力图 $F_{avg}^s \in \mathbf{R}^{1 \times W \times H}$ 和 $F_{max}^s \in \mathbf{R}^{1 \times W \times H}$,分别代表进行平均池化和最大池化操作后所得的特征向量,后通过标准卷积进行连接,最终生成二维的空间注意力图。具体运算步骤的公式为

$$M_s(F) = \sigma (f^{7 \times 7} \{ ([\text{AvgPool}(F); \text{MaxPool}(F)] \}) \}) = \sigma \{ f^{7 \times 7} ([F_{avg}^s; F_{max}^s]) \} \quad (11)$$

式(11)中: σ 为 sigmoid 函数; $f^{7 \times 7}$ 代表滤波器大小为 7×7 的卷积运算。

1.3 时间特征提取与优化

1.3.1 改进的时间特征提取模块

在经过图卷积模块后,在时间卷积模块之前又加入了一个CBAM模块,继续对经过图卷积的特征向量进行通道和空间注意力计算,获取深层的通道和空间信息,丰富提取出的特征向量。

在时间卷积模块之前引入CBAM注意力模块,不仅可以进一步整合完善先前图卷积网络中提取出的特征向量,同时关注重点特征向量通道,对于之后的时间特征提取有着一定的提升与优化作用,使得总体获得的特征向量包含更加丰富的信息,便于之后异常行为的识别工作。此处添加的CBAM模块的具体网络结构与数据流动方式与1.2.2节中对于CBAM模块的介绍一致。

1.3.2 时空动态建模

ST-GCN模型基于人体骨架动力学,区别于以往依赖手工提取或通过遍历规则进行的传统骨骼建模方法,提出了一种从人体姿态图中学习空间和

时间两个维度信息的动态骨架模型,优化了骨架模型的表达能力并使其拥有更好的泛化能力。在对骨架序列内的时空动态进行建模的任务中,通过将时间上相邻的节点添加入邻接节点集合,即

$$B(v_{i_t}) = \{v_{i_q} \mid d(v_{i_t}, v_{i_q}) \leq K, |q - t| \leq \lfloor \frac{\Gamma}{2} \rfloor\} \quad (12)$$

式(12)中: v_{i_t} 为在 t 时间的第 i 个节点; v_{i_q} 为在 q 时间的第 j 个节点参数; Γ 为时间核大小,控制包含在邻居图中的时间范围。

通过该操作扩展了邻域的概念,将空间图卷积神经网络扩展到空间时间域。在采样函数的构造上,该模型对于节点 v_{i_t} 的邻接节点集合 $B(v_{i_t}) = \{v_{i_q} \mid d(v_{i_t}, v_{i_q}) \leq D\}$ 上定义采样函数 $P: B(v_{i_t}) \rightarrow V$, 可表示为

$$P(v_{i_t}, v_{i_q}) = v_{i_q} \quad (13)$$

式(13)中: $d(v_{i_t}, v_{i_q})$ 为从节点 v_{i_t} 到节点 v_{i_q} 的最短路径,该模型中的取值为 1。

在加入采样函数和权重函数后,完整时空图卷积模型为

$$l_{st}(v_{i_q}) = l_{ii}(v_{i_q}) + (q - t + \lfloor \Gamma/2 \rfloor)K \quad (14)$$

式(14)中: $l_{ii}(v_{i_q})$ 为在 v_{i_q} 上的单帧标签映射。

这样,就对所构造的时空图有了一个定义良好的卷积运算。

1.4 异常分数计算

经过全局特征提取、空间特征提取和通道特征提取后的特征可以表示为

$$f_{out}(v_{i_t}) = \sum_{v_{i_m} \in N(v_{i_t})} f_{in}(v_{i_m})w(v_{i_t}, v_{i_m}) \quad (15)$$

式(15)中: $N(v_{i_t})$ 为关节点 v_{i_t} 的邻接点集; $w(v_{i_t}, v_{i_m})$ 为由时空邻接矩阵定义的节点 v_{i_t} 和节点 v_{i_m} 之间的权重。

模型通过学习高斯分布为 $\mathbf{Z} \sim N(\mu_{normal}, I)$ 可逆映射,得到由人体动态骨架序列 p_x 映射在隐空间的高斯分布 p_z , 对每个人体动态骨架序列的异常分数进行估计,该过程可由公式表示为

$$\log_2 p_z(z) = \text{Const} - \frac{1}{2}(z - \mu_{normal}) \quad (16)$$

在模型的训练过程中,通过不断减小训练样本的负对数似然性值 L_{total} 来实现模型的优化,该过程可以由公式表示为

$$L_{total} = -\log_2 p_z[f(x)] - \log_2 \left| \det \left(\frac{df}{dx} \right) \right| \quad (17)$$

2 实验结果与分析

2.1 数据集与评估指标

使用了 ShanghaiTech 数据集和 UBnormal^[15] 数

据集作为异常行为识别模型常用数据集。ShanghaiTech 数据集包含 13 个场景的 330 个训练和 107 个测试视频,其中的测试集已经完成了每一帧的异常行为标记。ShanghaiTech 数据集数据的丰富性,和异常行为种类的多样性,都使得在异常行为识别领域,该数据集在同类数据集的横向比较中占有优势。UBnormal 数据集是针对异常行为识别算法研究的公开监督数据集,拥有 29 个虚拟场景共 236 902 个视频帧。与其他数据集相比,UBnormal 数据集进行了像素级的异常行为标注,并使全监督学习在异常行为识别领域成为可能。

在评估方面,模型对视频的每一帧计算一个分数,数据集分数是用于连接测试集中所有帧分数的 ROC 曲线下的面积,这是最常见的度量标准称为 Micro AUC。

2.2 实验环境

模型在 GPU 型号为 RTX3090, pytorch 版本为 3.8.0 的环境下, epoch 设置为 8, batch size 设置为 256, 学习率为 4×10^{-5} , 实验结果如下。

2.3 消融实验

为减少实验赘述,将插入在 ST-GCN 图卷积前方的 CBAM 模块简称为 CBAMg, 将插入在 ST-GCN 时间卷积前方的 CBAM 模块简称为 CBAMt, 将 Transformer 模块简称为 Trans, 消融实验结果如表 1 所示。

在表 1 中可以看到,使用 ShanghaiTech 数据集进行训练时,在添加 Transformer 模块后,精确度上升了 0.2%; 在图卷积前方添加 CBAM 模块后,精确度上升了 0.3%; 在时间卷积前方添加 CBAM 模块后,精确度上升了 0.1%; 添加 Transformer 模块并在时间卷积前方添加 CBAM 模块后,精确度上升了 0.2%; 添加 Transformer 模块在并图卷积前方添加 CBAM 模块后,精确度上升了 0.4%; 在图卷积前方添加 CBAM 模块并在时间卷积前方添加 CBAM 模块后,精确度上升了 0.4%; 在添加全部 3 个模块后,精确度上升了 0.5%。使用 UBnormal 数据集进行训练时,在添加 Transformer 模块后,精确度上升了 0.3%; 在图卷积前方添加 CBAM 模块后,精确度上升了 0.8%; 在时间卷积前方添加 CBAM 模块后,精确度上升了 0.5%; 添加 Transformer 模块并在时间卷积前方添加 CBAM 模块后,精确度上升了 1.0%; 添加 Transformer 模块在并图卷积前方添加 CBAM 模块后,精确度上升了 0.7%; 在图卷积前方添加 CBAM 模块并在时间卷积前方添加 CBAM 模块后,精确度上升了 1.0%; 在添加全部 3 个模块后,精确度上升了 1.2%。

表 1 消融实验结果

Table 1 Ablation experiment results

使用方法	AUC	
	ShanghaiTech	UBnormal
STG-NF	85.9	69.0
STG-NF + Trans	86.1	69.3
STG-NF + CBAMg	86.2	69.8
STG-NF + CBAMt	86.0	69.5
STG-NF + Trans + CBAMg	86.1	70.0
STG-NF + Trans + CBAMt	86.3	69.7
STG-NF + CBAMg + CBAMt	86.3	70.0
STG-NF + Trans + CBAMg + CBAMt	86.4	70.2

由实验结果分析可知,Transformer 模块能够增强模型提取到的全局人体动态骨架特征的表达能 力,CBAMg 模块有效提升了提取人体动态骨架特征的空间表达能力,CBAMt 模块有效提升了提取人体动态骨架特征的时间表达能力。在添加相应模块后明星性能都有一定程度的提升,因此该消融实验可以证明单一模块与融合模块对于改进的 STG-NF 网络模型的有效性。

2.4 对比实验

在保证配置环境及初始训练参数一致的情况下,本文研究将改进的 STG-NF 网络模型与其他网络模型进行实验来验证改进网络模型的有效性,结果如表 2 所示。

由实验结果分析可知,相较于 Markovitz 等^[7]使用时空图卷积自编码器进行特征编码,本文方法通过引入 Transformer 模块丰富了人体动态骨架特征的表达能力,在模型检测精确度方面提升了 10.3%。相较于 Zaheer 等^[9]将合作学习应用到异常行为识别领域,本文方法通过将人体动态骨架序列映射到隐空间后计算其异常分数,将归一化流应用到异常行为识别领域,在模型检测精确度方面提升了 6.8%。相较于 Barbalau 等^[16]和 Wang 等^[17]提出的自监督方法,本文研究通过引入多个注意力模块,提升所提取特征的丰富程度,在模型检测精确度方面分别提升了 2.6% 和 2.2%。

可以看出,改进后的时空图归一化流网络模型在异常行为识别的精确度方面相较于之前的其他异常

表 2 对比实验结果

Table 2 Comparative Experiment Results

方法	年份	是否监督学习	AUC
文献[7]	2020 年	半监督学习	76.1
文献[9]	2022 年	无监督学习	79.6
文献[16]	2022 年	自监督学习	83.8
文献[17]	2022 年	自监督学习	84.2
STG-NF	2023 年	无监督学习	85.9
改进后的 STG-NF	2024 年	无监督学习	86.4

行为识别模型有所提升,证明了本文模型的先进性。

2.5 可视化展示

模型在 ShanghaiTech 数据集和 UBnormal 数据集上的可视化结果如图 6 ~ 图 8 所示。其中折线图代表了视频每一帧的异常分数,红色部分代表真实值中标注为异常的视频帧。

如图 6 所示,在 ShanghaiTech 数据集中的人行道场景下,将自行车的出现定义为异常行为。在该视频序列的前半段未出现骑自行车的画面,异常分数较低;后半段出现骑自行车的画面,异常分数较高,预测结果与数据集标注值相符。如图 7 所示,在 UBnormal 数据集中的街道场景下,将跌倒定义为异常行为。在该视频序列前半段行人正常行走,异常分数较低;后半段中行人跌倒,异常分数较高,预测结果与数据集标注值相符。如图 8 所示,在 UBnormal 数据集中的路口场景下,打架倒定义为异常行

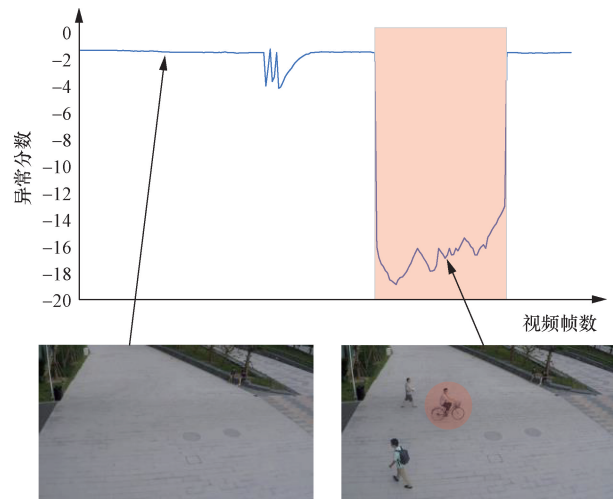


图 6 ShanghaiTech 数据集上可视化结果图
Fig. 6 Visualization on ShanghaiTech dataset

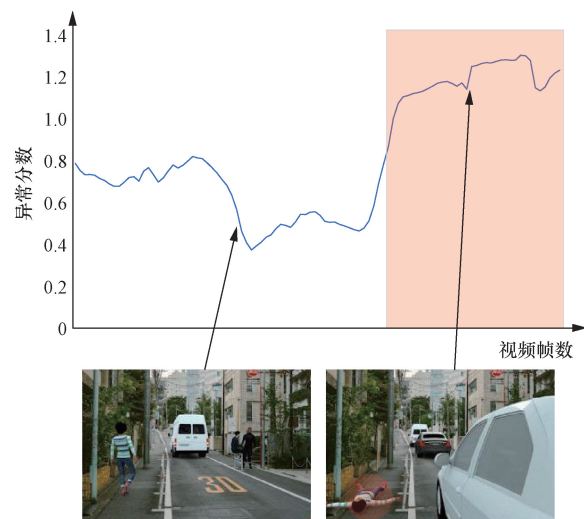


图 7 UBnormal 数据集上可视化结果图 1
Fig. 7 Visualization on UBnormal dataset 1

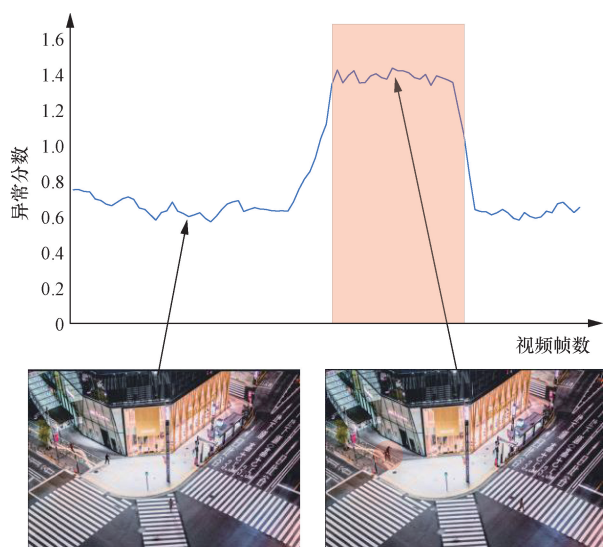


图8 UBnormal数据集上可视化结果图2

Fig. 8 Visualization on UBnormal dataset 2

为。在该视频序列前半段行人正常行走,异常分数较低;后半段中行人打闹,异常分数较高,预测结果与数据集标注值相符。

可视化结果展示模型能够给异常帧打出较高的分数,从而将正常帧与异常帧有效地分别出来。

3 结论

提出了一种基于时空图归一化流改进的异常行为识别方法,经实验得出以下结论。

(1)在原有时空图卷积归一化流模型基础上,在空间卷积与时间卷积前分别引入 CBAM 模块,丰富了人体动态骨架特征的时空特征。

(2)在归一化流的分流衍射阶段加入 Transformer 模块,丰富了人体动态骨架的全局特征,有效提升了模型精度。

(3)最终模型在 ShanghaiTech 数据集上进行仿真验证取得精度为 86.4% 的识别结果,表明了该方法具有先进性。

参 考 文 献

- [1] 邵宝平, 常世新, 赵建东. 数据驱动下交通异常行为的双层识别模型研究[J]. 科学技术与工程, 2023, 23(14): 6257-6263. Shao Baoping, Chang Shixin, Zhao Jiandong. Double layer identification model of traffic abnormal behavior driven by data[J]. Science Technology and Engineering, 2023, 23(14): 6257-6263.
- [2] 王新文, 谢林柏, 彭力. 跌倒异常行为的双重残差网络识别方法[J]. 计算机科学与探索, 2020, 14(9): 1580-1589. Wang Xinwen, Xie Linbo, Peng Li. Double residual network recognition method for falling abnormal behavior[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(9): 1580-1589.
- [3] 姬晓飞, 张东阳. 人体检测与异常行为识别联合算法[J]. 科学技术与工程, 2023, 23(8): 3370-3378.

- Ji Xiaofei, Zhang Dongyang. A joint algorithm of human detection and abnormal behavior recognition[J]. Science Technology and Engineering, 2023, 23(8): 3370-3378.
- [4] 彭月平, 蒋镛圻, 徐蕾. 基于粗糙集的多类人群异常行为识别算法[J]. 科学技术与工程, 2021, 21(11): 4524-4533. Peng Yueping, Jiang Rongqi, Xu Lei. An algorithm for identifying multi-class abnormal behavior of population based on rough set model[J]. Science Technology and Engineering, 2021, 21(11): 4524-4533.
- [5] Nady A, Atia A, Abutabl A. Real-time abnormal event detection in crowded scenes[J]. Journal of Theoretical and Applied Information Technology, 2018, 96(18): 6064-6074.
- [6] Cho M A, Kim T, Kim W J, et al. Unsupervised video anomaly detection via normalizing flows with implicit latent features[J]. Pattern Recognition, 2022, 129: 108703.
- [7] Markovitz A, Sharir G, Friedman I, et al. Graph embedded pose clustering for anomaly detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE/CVF, 2020: 10539-10547.
- [8] Sato F, Hachiuma R, Sekii T. Prompt-guided zero-shot anomaly action recognition using pretrained deep skeleton features [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE/CVF, 2023: 6471-6480.
- [9] Zaheer M Z, Mahmood A, Khan M H, et al. Generative cooperative learning for unsupervised video anomaly detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE/CVF, 2022: 14744-14754.
- [10] Hirschorn O, Avidan S. Normalizing flows for human pose anomaly detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Vancouver: IEEE/CVF, 2023: 13545-13554.
- [11] Fang H S, Li J, Tang H, et al. Alphapose: whole-body regional multi-person pose estimation and tracking in real-time [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6): 7157-7173.
- [12] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: ECCV, 2018: 3-19.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 1-11.
- [14] Dinh L, Krueger D, Bengio Y. Nice: non-linear independent components estimation[J]. arxiv preprint arxiv: 1410. 8516, 2014.
- [15] Acsintoae A, Florescu A, Georgescu M I, et al. Ubnormal: new benchmark for supervised open-set video anomaly detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE/CVF, 2022: 20143-20153.
- [16] Barbalau A, Ionescu R T, Georgescu M I, et al. SSMTL + + : revisiting self-supervised multi-task learning for video anomaly detection [J]. Computer Vision and Image Understanding, 2022, 45(6): 7157-7173.
- [17] Wang G, Wang Y, Qin J, et al. Video anomaly detection by solving decoupled spatio-temporal jigsaw puzzles [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 494-511.