引文格式:张杰,杨雪,龚智龙,等.利用BiP-GAN进行行人视频异常事件自动检测[J].武汉大学学报(信息科学版),2025,50(7):1266-1276.DOI:10.13203/j.whugis20240259

Citation:ZHANG Jie,YANG Xue,GONG Zhilong,et al. BiP-GAN Pedestrian Video Anomaly Event Automatic Detection[J]. Geomatics and Information Science of Wuhan University,2025,50(7):1266-1276.DOI:10.13203/j.whugis20240259

利用BiP-GAN进行行人视频异常事件自动检测

张 杰[1] 杨 雪[1] 龚智龙[1] 关庆锋[1]

1 中国地质大学(武汉)国家地理信息系统工程技术研究中心,湖北 武汉,430074

摘 要:视频监控系统在安全和监督领域扮演着至关重要的角色,如何在不需要人为干预的情况下从视频中自动精准识别具有潜在安全威胁的行人非正常行为或事件,减少对大量视频监控画面进行人工审查的压力,是目前计算机视觉领域的研究热点之一。近年来,人工智能技术的快速发展使得视频异常检测技术得到了大幅提升,但多变、多样环境下异常与正常行为的细微差异区分还存在挑战。构建了一种新的双向预测生成对抗网络(bidirectional prediction generative adversarial network,BiP-GAN)视频行人异常检测模型。该模型主要包括交叉循环注意力(criss-cross attention,CCA)-U-Net生成器和Globle-Patch判别器,利用光流模型在光流变化及图像序列运动特征上的捕获优势,将其用于生成器和判别器的损失函数计算。CCA-U-Net生成器以经典U-Net模块为基础,通过CCA模块增强模型对视频行为关键特征的识别能力。Globle-Patch判别器通过结合Globle判别器和Patch判别器在全局和局部特征的感受优势,提高模型全局及局部的特征感受能力,提高模型的鲁棒性和准确性。BiP-GAN的预训练策略采用前4帧正向预测和后4帧反向预测的双向预测模式,使模型更好地结合图像序列的上下文特征,生成图像质量更好的预测帧。另外,BiP-GAN采用Warm-up与余弦退火学习率函数(cosine annealing function,CAF)相结合的学习率衰减方法,加快模型寻找全局最优解,从而节省计算资源。实验利用公开数据集CUHK Avenue、UCSD ped2和ShanghaiTech对BiP-GAN进行了验证和分析,其曲线下面积的平均值分别为87.3、96.2、73.9,均高于已有经典模型(如Ada-GAN、Con-GAN、Mul-GAN)。消融实验表明了CCA-U-Net生成器、Globle-Patch判别器、双向预测策略以及Warm-up与CAF结合的学习率衰减方法对于模型的有效性。

关键词:生成对抗网络;行人视频异常事件检测;深度学习;人工智能

中图分类号:P208 **文献标识码**:A **收稿日期**:2024-12-03

DOI:10.13203/j.whugis20240259 **文章编号**:1671-8860(2025)07-1266-11

BiP-GAN Pedestrian Video Anomaly Event Automatic Detection

ZHANG Jie[1] *YANG Xue*[1] *GONG Zhilong*[1] *GUAN Qingfeng*[1]

1 National Engineering Research Center of Geographic Information System, China University of Geosciences (Wuhan), Wuhan 430074, China

Abstract: **Objectives**: Video surveillance system plays a vital role in the field of security and supervision. How to automatically and accurately identify abnormal pedestrian behaviors or events with potential security threats from videos without human intervention, and reduce the pressure of manual review of a large number of video surveillance images, is one of the current research hotspots in the field of computer vision. In recent years, the rapid development of artificial intelligence technology has greatly improved video anomaly detection technology. However, there are still challenges in distinguishing subtle differences between abnormal and normal behavior in changing and diverse environments. **Methods**: We construct a new video pedestrian anomaly detection model based on bidirectional prediction generative adversarial network (BiP-GAN). The model mainly includes CCA-U-Net generator and Globle-Patch discriminator. The advantages of op-

基金项目:国家自然科学基金(42271449)。

第一作者:张杰,硕士,主要从事视频图像处理研究。18333981575@163.com

通信作者:杨雪,博士,副教授。yangxue@cug.edu.cn

tical flow model in capturing optical flow changes and image sequence motion characteristics are used to calculate the loss function of the generator and discriminator. Based on the classic U-Net module, the criss-cross attention (CCA)-U-Net generator introduces CCA module to enhance the recognition ability of the model for the key features of video behavior. Globle-patch discriminator combines the global and local feature perception advantages of Globle discriminator and Patch discriminator, improves the global and local feature perception ability of the model, and the robustness and accuracy of the model. The pre-training strategy of BiP-GAN adopts the bidirectional prediction mode of the first 4 frames of forward prediction and the last 4 frames of reverse prediction, so that the model can better combine the context features of the image sequence and generate prediction frames with better image quality. In addition, BiP-GAN uses a learning rate decay method combining Warm-up and cosine annealing function (CAF) to speed up the model to find the global optimal solution, thus saving computing resources. **Results:** BiP-GAN is verified and analyzed by using the public datasets CUHK Avenue, UCSD ped2 and ShanghaiTech. The average area under the curve of BiP-GAN is 87.3, 96.2 and 73.9, respectively. All of them are higher than the existing classic models (such as Ada-GAN, Con-GAN, Mul-GAN). Ablation experiments show the effectiveness of the CCA-U-Net generator, Globle-Patch discriminator, bidirectional prediction strategy, and the learning rate decay method combining Warm-up and CAF for the model. **Conclusions:** The proposed BiP-GAN model effectively enhances the accuracy and robustness of video anomaly detection through bidirectional prediction, attention mechanisms, multi-scale discrimination, and an optimized training strategy. Experimental results demonstrate its superiority over existing models, confirming its potential for practical application in intelligent surveillance systems.

Key words: generative adversarial network; anomalous event detection in pedestrian video; deep learning; artificial intelligence

数字化时代,视频监控已成为现代社会安全架构的核心组成部分。随着高清摄像机的普及和视频数据处理技术的进步,越来越多的公共和私人空间被实时监控,以预防和响应犯罪和其他安全事故。持续监控产生的海量视频数据使得人力实时分析监控成为挑战,从而对视频自动化分析提出了需求。自动异常事件检测作为视频分析的一项关键功能,核心目标是在不需要人为干预的情况下,准确地从视频中识别出可能表示安全威胁的非正常行为或事件,如交通监控中的事故检测、零售环境中的盗窃预防、公共安全的威胁识别,以及疲劳驾驶或违反交通规则等。因此,如何实现视频异常行为准确、及时检测对于防止或最小化风险至关重要。近年来,随着人工智能技术的快速发展,基于深度学习算法的视频异常检测成为研究热点。

相比于传统机器学习方法手动设计提取视频特征问题,利用深度学习方法能够自动从原始数据中学习特征表示,实现异常行为的自动检测。如文献[1]提出了一个自监督多任务学习框架,结合时间箭头、运动异常和外观预测等任务来检测异常事件。文献[2]提出通过不同尺度来学习视频的时空特征,以捕捉细粒度和粗粒度的异常行为。该方法通过多任务自监督策略增强模型的泛化能力,使其在异常检测任务上表现更出色。文献[3]提出了鲁棒时序特征幅度(robust temporal feature magnitude,RTFM)方法,该方法在弱监督环境下利用时间特征的大小差异来区分正常与异常视频片段。通过选择视频中最显著的特征,RTFM 实现了更高的异常检测准确度,同时减少了对精确标注的需求。文献[4-5]使用卷积神经网络(convolutional neural network,CNN)来进行视频异常检测。这些方法的优点是它们能够自动从原始视频中学习复杂的时空特征,而无需人工设计特征,这对于处理大规模的视频数据尤为重要。文献[6]提出的双尺度串行网络方法则关注多尺度特征的提取,试图通过多尺度的视角来捕捉视频中不同行为模式的异常性,从而提高检测的准确性。视频异常事件检测不仅需要提取空间特征(例如每一帧的图像信息),还需要捕捉时间上的动态变化(例如人物行为的演变)。文献[7]通过时空自编码网络来建模时空特征,这为如何处理视频中时空动态提供了一个有效的方向。此类方法通常在处理动作检测、行为分析等任务时具有优势。基于生成对抗网络(generative adversarial network, GAN)模

型的视频异常检测是目前较为广泛的一种方法[8]。该方法通过生成器和判别器的对抗训练,学习视频的正常模式,从而实现异常事件检测[9-11]。文献[11]提出了一种基于GAN的改进模型,采用多尺度判别器和增强型生成器结构,以提升生成帧与真实帧之间的相似度,并有效识别视频中的异常事件。该方法在复杂场景下显示出良好的泛化能力。此外,利用光流模型估计视频之间的光流信息[12],捕捉视频中的运动特征,也是目前大多数研究中采用的一种方式。文献[13]提出了双向预测网络,使用U-Net网络双向预测视频序列的中间帧,通过正向与反向预测帧的峰值信噪比(peak signal-to-noise ratio,PSNR)值判断异常。虽然,现有研究从视频特征获取、训练机制制定等角度提出了相应的解决方案,但是依然存在缺陷。

首先,视频异常事件检测任务的重点是需要结合视频序列的上下文信息,实现视频帧随时间变化特征的学习。但是,目前很多GAN模型忽略了上下文信息融合,使得模型异常检测精度无法得到提升[14]。其次,行人视频拍摄角度多样、异常复杂,典型的公开数据集主要有两类:(1)利用高位设置的监控摄像头拍摄获取,如UCSD(University of California at San Diego) pedestrian1 & pedestrian2(简写为ped2))数据集;(2)通过行人视角拍摄的视频数据,如CUHK(The Chinese University of Hong Kong) Avenue数据集。现有大多模型在验证其有效性时主要对两类数据中的其中一种有较好的提升效果,对另一类的数据集提升效果不明显或检测精度较差,存在泛化性差的问题[15-18]。最后,现有大多GAN模型在训练过程中耗费了大量的计算资源用于最优解求解,存在模型效能差的问题[19-21]。

针对以上问题,本文构建了一种新的双向预测GAN(bidirectional prediction GAN, BiP-GAN)行人视频异常检测模型。该模型主要包括交叉循环注意力(cross cirss attention,CCA)-U-Net生成器和Globle-Patch判别器,利用光流模型在光流变化及图像序列运动特征的捕获优势,将其用于生成器和判别器的损失函数计算。BiP-GAN的预训练策略采用双向预测模式,使模型更好地结合图像序列的上下文特征生成图像质量更好的预测帧。另外,BiP-GAN采用了一种新型的学习率衰减方法,使模型能够更快地找到全局最优解,从而节省计算资源。实验利用公开数据集CUHK Avenue、UCSD ped2和ShanghaiTech对BiP-GAN进行了验证和分析。

1 BiP-GAN

预测模型通常具有结构简单、泛化能力强的特点,可以很好地利用时间和空间信息,对于视频异常检测更有优势。传统帧预测网络通常依赖于目标的前几帧,使得运动特征获取不够全面且容易受噪声干扰。例如,文献[18]使用了GAN双向预测的方法,以经典U-Net网络作为生成器,传统的Patch判别器作为鉴别器。针对传统帧预测网络存在的问题,本文构建了一种基于BiP-GAN的视频异常事件检测模型,如图1所示。该模型以CCA-U-Net网络作为GAN的生成器来预测生成视频帧,由Globle-Patch作为判别器来区分出预测帧与真实帧。此外,BiP-GAN模型添加了Lite-Flownet光流模块,从而增加视频序列的运动特征。

与现有GAN不同的是,BiP-GAN训练策略结合了正向预测和反向预测,并将生成的两个预测帧进行图像特征加权融合,丰富预测帧的上下文信息。首先,将输入帧T_1、T_2、…、T_9分为正向预测输入帧T_1、T_2、T_3、T_4和目标帧T_5以及反向预测输入帧T_9、T_8、T_7、T_6。然后,将两组输入帧分别通过独立的GAN结构进行预测,并且根据双向预测帧的生成图片质量即PSNR进行加权特征融合,从而提高生成的预测帧的质量,减少预测帧和真实帧的差异,为异常事件的检测提供更可靠的输入。通过双向预测的训练模式与加权特征融合,模型可以更好地结合视频序列的上下文信息,生成更接近真实帧的预测帧,从而提高检测精度。

1.1 生成器网络构建

1.1.1 U-Net预测网络

U-Net作为生成器可以有效地学习视频序列中的时空特征。通过编码器-解码器的结构,网络可以同时捕捉低层次的细节特征和高层次的语义信息,从而更好地理解视频序列的时空动态。U-Net通过上采样操作在不同的尺度上进行特征融合和重建。这种多尺度的上采样可以提高网络对不同尺度异常事件的检测能力,使网络在处理多尺度视频时更加灵活。相比于文献[9]中的U-Net预测网络,本文在U-Net中的上采样和下采样部分都引入CCA模块,使用更多的卷积层以提取更深层次的图像特征。在卷积操作之后,每

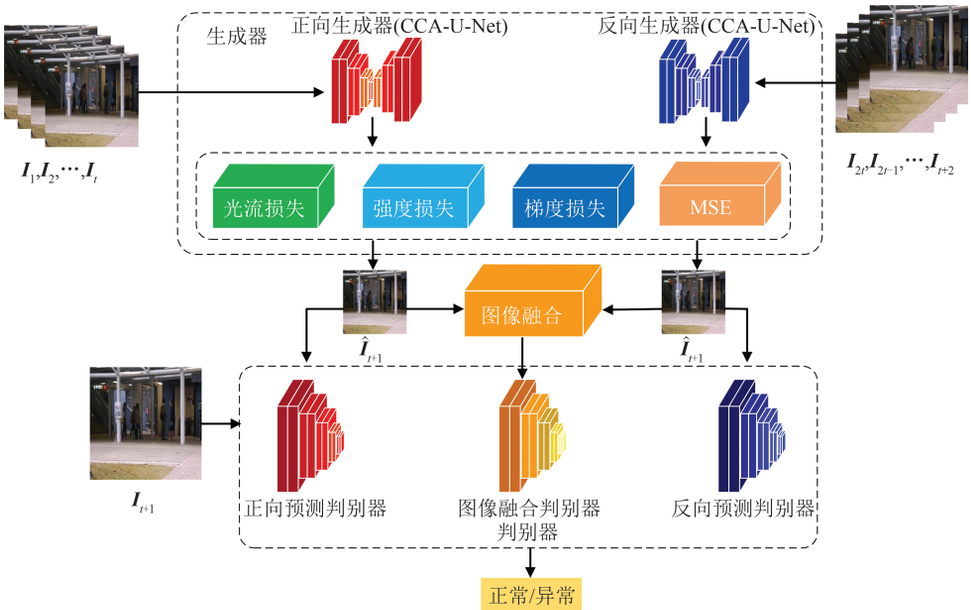

图 1　BiP-GAN网络框架

Fig. 1　BiP-GAN Network Framework

个通道都会经过批标准化和线性整流函数(rectified linear unit, ReLU)激活函数。输出使用了tanh激活函数,提升U-Net作为生成器的整体性能。U-Net预测网络的结构如图2所示。

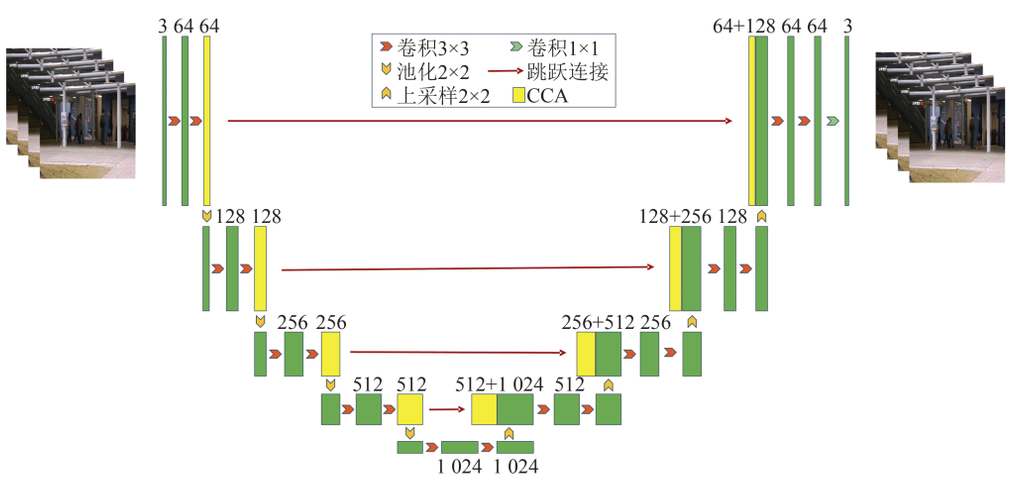

图 2　U-Net预测网络

Fig. 2　U-Net Prediction Network

1.1.2　CCA模块

CCA模块是一种用于增强神经网络对时空关系建模能力的注意力机制[18]。它可以自适应地捕捉输入特征图中的时空依赖关系,并将这些信息应用于网络的下一层操作。该模块的核心思想是通过交叉卷积操作来建模特征图中的时空关系,通过在特征图的行和列维度上执行卷积操作来获取行间和列间的关联信息。CCA模块添加到U-Net网络中可以提升网络对细节特征的感知能力。通过在特定位置上执行行间和列间的卷积操作,网络可以更加细致地分析特征图中的细节信息,从而有助于准确地检测和定位异常事件[19]。CCA模块的结构如图3所示。

1.2　Globle-Patch判别器网络构建

本文提出的Globle-Patch判别器通过融合全局特征和局部细节信息提升对抗训练中判别器的判别能力,使生成器能够生成更接近真实帧的预测帧。Globle判别器在卷积的最后一层通过线性变换获得真假概率值,与传统的Patch判别器最后一层通过卷积计算概率相比,能更好地根据全局特征进行判别。本文对识别全局特征效果更好的Globle判别器与传统的Patch判别器进行加权融合,通过多层卷积和线性层来提取图像特征并进行真假判别,并利用LeakyReLU激活函数

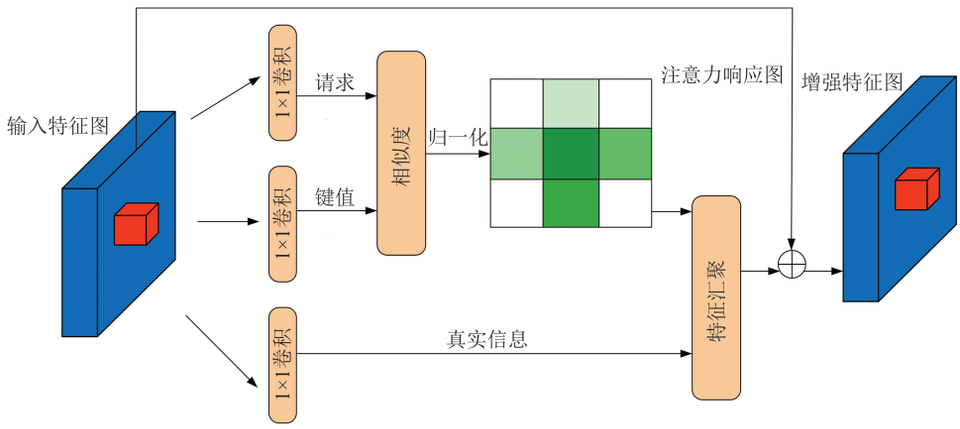

图3 CCA模块

Fig. 3 CCA Module

增强非线性能力。此外,本文所提的Patch判别器采用PatchGAN思想[9],将输入图像划分为多个小块并对每个小块进行判别。这种局部判别的方式可以更加细致地感知图像的细节信息,提高对异常事件的敏感性。相比传统判别器,本文所提的Globle-Patch判别器结构(见图4)在特征提取和判别能力上更具有优势,对图像的全局和局部特征进行综合判别[21]。

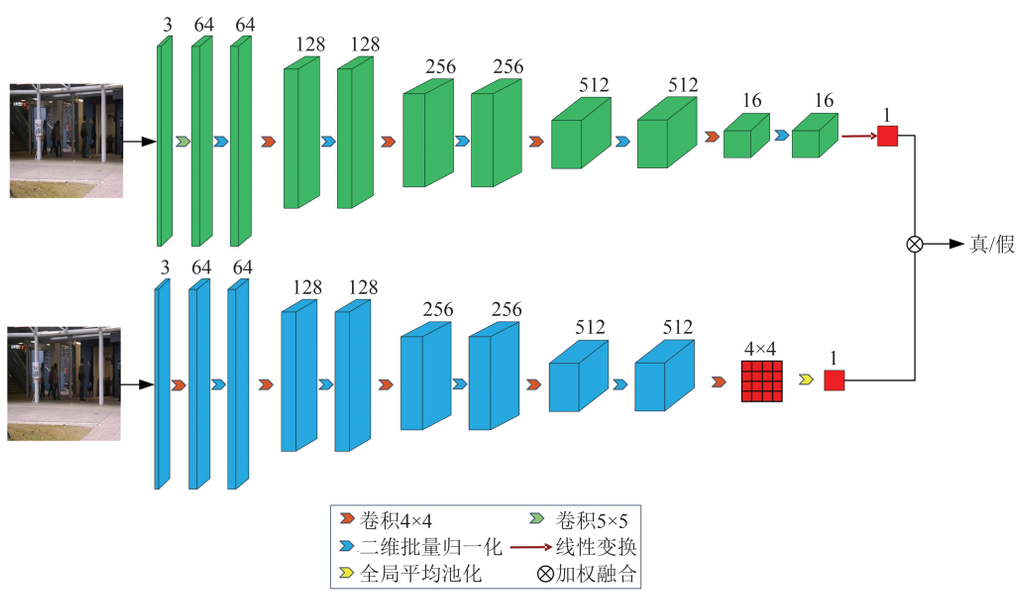

图4 Globle-Patch判别器网络结构

Fig. 4 Network Structure of Globle-Patch Discriminator

1.3 预训练策略设计

学习率衰减在深度学习领域是一种常用的优化策略,用于在训练过程中逐渐降低学习率。GAN在做异常事件检测任务时,如果学习率设置过高,可能导致训练不稳定、生成器和判别器之间的动态失衡。学习率衰减可以在训练后期减小学习率,使训练过程更加稳定,并有助于生成器和判别器达到更好的收敛状态。

余弦退火学习率函数(cosine annealing function,CAF)可以在训练过程中逐渐降低学习率,可以帮助模型在训练的早期阶段更快地收敛,并在后期阶段进行更细致的调整[22]。在初始阶段使用较高的学习率可以使模型快速探索梯度空间,并迅速收敛到局部最优解附近。随着训练的进行,学习率逐渐减小,使得模型能够在更小的学习率下进行更深入的优化。

通常在深度学习训练过程开始阶段使用较高的学习率帮助模型跳出局部最优解,但模型在开始训练时的初始值是随机的,此时的权重值与图片的特征相关性较小,学习率较大的情况下可能会使模型越来越偏离最优解,所以Warm-up学习率衰减策略有效地解决了这一问题。该策略在训练的初始阶段以一个较小的学习率逐渐增大到设置的阈值,这样可以避免初始随机权重值

对训练过程的影响[23]。本文将两种办法的优点进行结合,提出了一种改进后的CAF:

$$r(T)=\begin{cases} r_{\text{int}}\lambda_{\text{vel}}T_{\text{cur}}, s<1\,000 \\ r_{\min}+\dfrac{1}{2}(r_{\max}-r_{\min})\times \\ \quad [1+\cos(\dfrac{T_{\text{cur}}}{T_{\text{total}}}\pi)], s>1\,000 \end{cases} \tag{1}$$

式中,$r(T)$表示改进后的CAF;T_{cur}表示模型当前迭代次数;T_{total}表示模型总迭代次数;r_{int}表示学习率的初始值;$r_{\max}$表示最大学习率;$r_{\min}$表示最小学习率;λ_{vel}是根据模型的迭代次数可调整的权重;s表示训练迭代次数。

Warm-up策略可以在训练的初始阶段逐渐增加学习率,帮助模型快速启动训练,可以加速模型的收敛并提高训练的效率。改进后的CAF可以在后续训练过程中逐渐降低学习率,有助于模型在后期阶段更稳定地收敛到最优解。这可以减少训练过程中的震荡现象,使生成器和判别器能够更好地相互学习和优化。Warm-up和CAF策略的结合可以帮助模型更好地应对噪声和扰动。

1.4　目标函数及异常分数

1.4.1　目标函数

本文将强度、梯度、光流以及均方误差(mean square error,MSE)损失函数进行结合,设计了生成器的目标函数:

$$L_G=\lambda_{\text{int}}L_{\text{int}}(\hat{I}_{t+1},I_{t+1})+\lambda_{\text{gd}}L_{\text{gd}}(\hat{I}_{t+1},I_{t+1})+\lambda_{\text{op}}L_{\text{op}}+\lambda_{\text{adv}}L_{\text{adv}}(\hat{I}_{t+1}) \tag{2}$$

式中,L_G表示生成器的目标函数;λ_{int}、λ_{gd}、λ_{op}、λ_{adv}分别表示强度、梯度、光流以及MSE 4种损失函数的权重;L_{int}、L_{gd}、L_{op}、L_{adv}分别代表强度、梯度、光流以及MSE 4种损失函数;$\hat{I}_{t+1}$表示生成的预测帧;I_{t+1}表示视频序列的真实帧。

考虑到判别器在训练过程中仅仅需要计算预测帧与真实帧的MSE损失,因此构建本文的判别器目标函数为:

$$L_D=L_{\text{adv}}(\hat{I}_{t+1},I_{t+1}) \tag{3}$$

式中,L_D表示判别器的目标函数。

为了使预测帧接近真实帧,根据文献[24]提出的GAN模型,本文也使用了强度和梯度损失。强度惩罚保证了RGB空间中所有像素的相似性,而梯度惩罚可以使生成的图像更加清晰。空间强度中最小化预测帧$\hat{I}$与真实帧I之间的距离的计算公式为:

$$L_{\text{int}}\left(\hat{I}_{t+1},I_{t+1}\right)=\left\|\hat{I}_{t+1}-I_{t+1}\right\|_2^2 \tag{4}$$

梯度损失的计算公式为:

$$L_{\text{gd}}(\hat{I}_{t+1},I_{t+1})=\sum_{i,j}\left(\left|\left|\hat{I}_{i,j}-\hat{I}_{i-1,j}\right|-\left|I_{i,j}-I_{i-1,j}\right|\right|+\left|\left|\hat{I}_{i,j}-\hat{I}_{i,j-1}\right|-\left|I_{i,j}-I_{i,j-1}\right|\right|\right) \tag{5}$$

式中,$\hat{I}_{i,j}$、$I_{i,j}$分别表示预测帧和真实帧的像素点。

现有工作在预测帧生成过程中只考虑了强度和梯度损失[24],无法保证预测帧具有正确的运动特征。为了提高预测帧运动特征准确性,本文在生成器部分引入了光流损失[12]来评估预测帧与真实帧光流特征差异。在实际估算光流信息过程中,本文采用Lite-Flownet进行光流估计,用f表示Lite-Flownet,光流损失函数的计算公式为:

$$L_{\text{op}}=\left\|f(\hat{I}_{t+1},I_t)-f(I_{t+1},I_t)\right\|_1 \tag{6}$$

MSE损失函数用于衡量预测值与真实值之间的差异,计算公式为:

$$L_{\text{adv}}(\hat{I}_{t+1},I_{t+1})=(\hat{I}_{t+1}-I_{t+1})^2 \tag{7}$$

1.4.2　异常分数

对于图像质量评价方法,PSNR是一种常用的图像质量评价指标,它基于根据峰值信号与噪声的比率来衡量图像的质量。计算公式为:

$$P\left(I,\hat{I}\right)=10\lg\frac{(\max\hat{I})^2}{\frac{1}{N}\sum_{i=0}^{N}\left(I_i-\hat{I}_i\right)^2} \tag{8}$$

式中,P表示预测图像与真实图像之间的PSNR,单位为dB;$\hat{I}_{i,j}$、$I_{i,j}$分别表示预测帧和真实帧;max$\hat{I}$表示图像中像素值的最大值;N表示输入视频序列的长度。本文对PSNR进行归一化,并将每个测试视频中的所有帧取到[0,1],计算每帧的异常分数的计算公式为:

$$S(t)=\frac{P(I_t,\hat{I}_t)-\min_t(P(I_t,\hat{I}_t))}{\max_t(P(I_t,\hat{I}_t))-\min_t(P(I_t,\hat{I}_t))} \tag{9}$$

式中,S表示异常分数;max$_t$、min$_t$分别表示PSNR的最大值和最小值。

2　BiP-GAN异常检测实验

2.1　实验配置

采用Python 3.8版本编程语言,本文在x86_64架构的Linux系统中按照PyTorch 1.1.0框架搭建了所提模型,计算资源为Nvidia GeForce RTX 3090 Ti GPU,依赖CUDA11.5和CUDNN7支持。根据目前视频异常检测常规操作,将输入网络前的视频帧缩放至256×256像素

大小,3通道的像素值归一化到[-1,1]。实验共进行20 000次对抗训练迭代,采用Adam优化器以及改进后的CAF函数,批处理块大小设置为4。对于灰度数据集,生成器和判别器的学习率初始值分别设置为0.03和0.003。而对于色阶数据集,生成器和判别器的学习率初始值分别从0.02和0.002。权重因子λ_{int}、λ_{gd}、λ_{op}、λ_{adv}按照已有研究[9]分别设置为1.0、1.0、2.0、0.05。

2.2 数据集

本文采用公开数据集CUHK Avenue、UCSD ped2以及ShanghaiTech,将所提方法与现有经典算法进行了比较。CUHK Avenue数据集是在中国香港街道上通过摄像头实时采集的,包含了不同天气条件、不同时间段和不同场景的街道视频。数据集中的视频场景主要是城市街道,包括交通路口、人行道、商业区等,包含了一些常见的异常事件,如行人闯红灯、行人逆行、交通违规等。UCSD ped2数据集是在美国加利福尼亚大学圣地亚哥分校的校园中通过摄像头进行采集的,视频是在校园内的行人区域拍摄的。数据集中的视频场景主要是校园行人区域,包括人行道、广场、楼宇入口等,包括了行人行为异常,如行人奔跑、跌倒、突然停止等。ShanghaiTech数据集是在中国上海的城市环境中通过摄像头进行实时采集的。数据集中的视频场景是在城市街道和公共场所进行拍摄的,包括城市街道、地铁站、公园、广场等不同的城市环境。3种数据集内异常事件如图5所示。

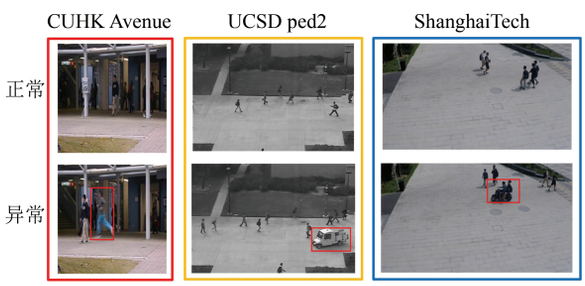

图5 CUHK Avenue、UCSD ped2、ShanghaiTech 3种数据集异常事件

Fig. 5 Abnormal Events of CUHK Avenue, UCSD ped2 and ShanghaiTech Datasets

BiP-GAN模型作为一个无监督检测模型,在训练过程中对正常视频的序列特征进行学习,验证阶段对测试集的视频序列进行双向预测。根据式(8)实时计算预测帧与真实帧的PSNR值,并根据式(9)计算每个预测帧的异常分数。异常分数即为每一预测帧归一化的PSNR值,代表其异常的概率值。

经过多次实验发现,当视频中无异常时,PSNR值普遍在36左右;当PSNR值低于34时,一般会有异常事件出现,图6展示了在实时检测CUHK Avenue数据集时的PSNR曲线以及出现异常情况时的图像帧与光流图。当视频中突然出现跑动的人时,PSNR数值会瞬间降低。完成输入视频序列检测后,根据异常分数和真实标签计算真正率(true positive rate,TPR)、假阳性率(false positive rate,FPR)及阈值并绘制接收者操作特征(receiver operating characteristic,ROC)曲线,计算曲线下的面积(area under the curve,AUC)作为模型的整体检测异常事件的评估指标。

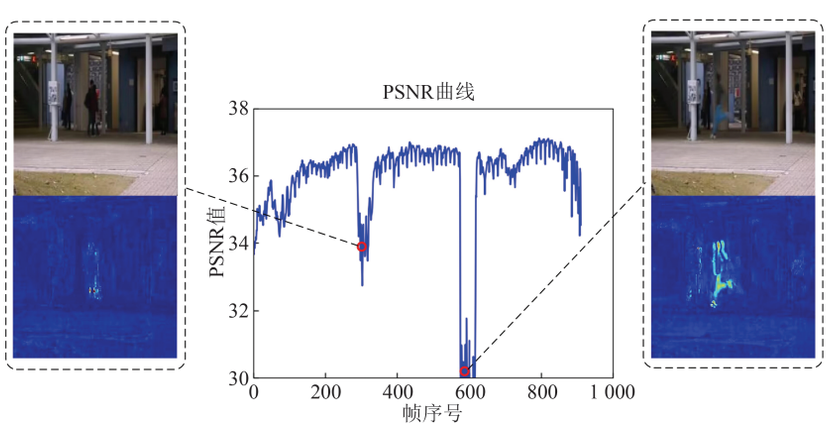

图6 视频序列的正常帧与异常帧对比

Fig. 6 Comparison of Normal Frames and Abnormal Frames of Video Sequence

2.3 输入帧序列超参数设置

根据现有研究发现输入视频帧序列的长度对视频异常检测精度有影响[25]:当输入视频帧序列较小时,模型特征提取会相对容易,但是特征

中包含的上下文时间特征较少;当输入视频帧序列较大时,提取到的特征中包含的上下文时间特征较充分,但模型提取特征时处理的数据量会比较大且复杂。针对现有研究对输入帧设置及硬件条件的限制,本文在CUHK Avenue数据集上将输入帧序列长度 T 分别设置为9和17对模型进行训练。

表1为不同输入帧序列长度即 $T=9$ 和 $T=17$ 时的平均AUC以及模型训练时长。可以看出,当视频帧输入序列长度由9帧增加到17帧后,伴随训练时间的增长,AUC精度并没有相应的提升,所以本文模型最终选择输入序列长度为9帧进行训练。

2.4　消融实验

为了验证各个模块对于本文模型精度结果的影响,在CUHK Avenue数据集上对模型开展消融实验,比较模型在添加不同模块后的性能。具体包括:对于主干网络,探究传统的正向预测与双向预测训练策略对于实验结果的影响。比较光流模块、CCA模块、Globle-Patch判别器以及改进后的CAF对于实验结果的影响。训练模型的系数均取自§2.1系数的最优性能,消融实验的AUC指标结果如表2所示。

表1　不同长度输入视频帧序列AUC与训练时间对比

Table 1　Comparison of AUC and Training Time of Input Video Frame Sequences of Different Lengths

AUC/%		时间/h	
$T=9$	$T=17$	$T=9$	$T=17$
85.7	86.1	3.8	6.6
86.4	86.6	3.9	6.7
86.2	86.2	4.1	6.8
86.5	86.3	3.8	6.7
86.4	86.2	3.7	6.9
86.1	86.2	3.8	6.7
86.5	86.4	3.8	6.8
85.9	86.5	3.9	7.0
86.6	86.3	4.0	6.7
86.5	86.5	3.8	6.9

表2　消融实验结果

Table 2　Results of Ablation Experiment

正向预测	双向预测	光流模块	CCA	Globle-Patch判别器	学习率衰减	AUC/%
√		√				83.1
√		√	√			83.4
√		√	√	√		83.8
√		√	√	√	√	84.2
	√	√				83.8
	√	√	√			84.3
	√	√	√	√		86.1
	√	√	√		√	85.4
	√	√	√	√	√	87.3

由表2可知,BiP-GAN模型在做相应的改进以后对于异常事件检测结果均有提升,证明了各个模块对于BiP-GAN模型检测精度提升的有效性。当PNSR数值升高时,生成器生成的预测帧图像质量与真实帧差异变小。如图7所示,当迭代次数不断增加时,特征融合后双向预测帧的PNSR数值普遍高于正向预测生成的图像帧,且明显高于反向预测生成的图像帧。由此说明本文所提的双向预测特征融合方法能够显著改善预测帧与真实图像之间的差异。

以Avenue数据集为例,图8中分别展示了采用双向预测、正向预测及反向预测模式生成器自动生成的预测帧与真实帧之间的差异。根据图8结果所示,采用双向预测生成器生成的预测帧与真实图像帧之间的差异最小,验证了BiP-GAN模型在预测帧与真实帧差异改善方面的性能。

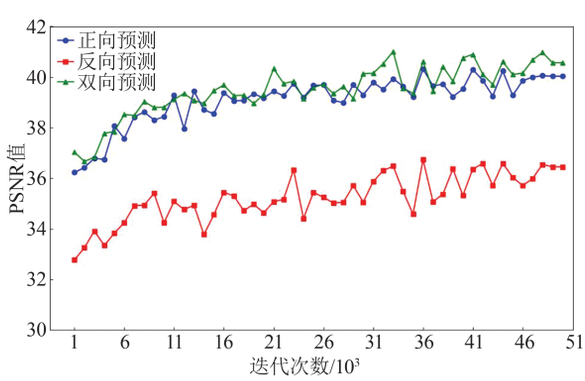

图7　双向预测与特征融合PSNR比较

Fig. 7　PSNR Comparison Between Bidirectional Prediction and Feature Fusion

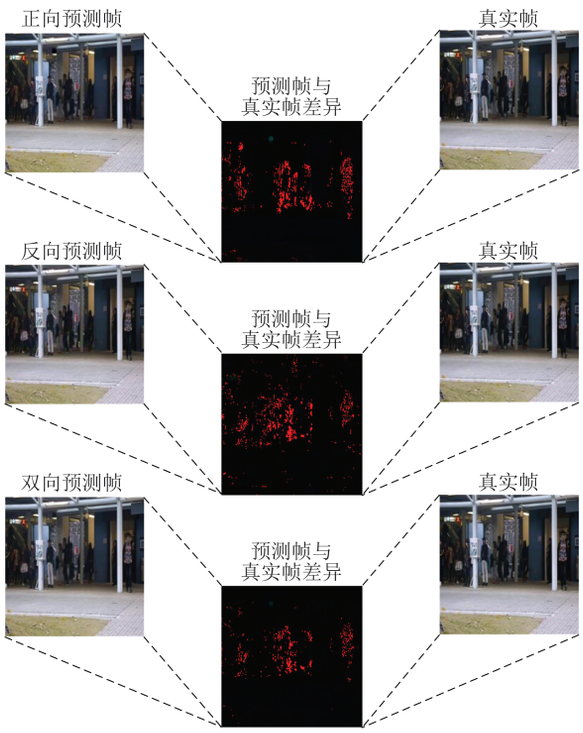

图8 双向预测生成器与正向、反向生成器在图像质量改善方面的对比

Fig. 8 Comparison of Bidirectional Prediction Generator with the Forward and Reverse Generators in Terms of Image Quality Improvement

表3展示了利用Avenue数据集添加Warm-up CAF学习率衰减策略对模型收敛速度的影响。加入Warm-up CAF策略后,BiP-GAN模型达到收敛所需的时间和迭代次数明显减少,同时模型精度没有受明显影响。这验证了Warm-up CAF策略在加速收敛方面的有效性,也为其他深度学习任务提供了实用的学习率衰减方法,有助于模型训练时使模型快速收敛,节省训练时间。

2.5 与经典算法对比

表4展示了本文模型BiP-GAN与已有经典模型在公开数据集CUHK Avenue、UCSD ped2、ShanghaiTech上的性能比较。需要强调的是,为了验证BiP-GAN性能的有效性,本文选择针对每一类数据集目前较为优越的模型作为比较对象。

以CUHK Avenue数据集为实验对象,选择FFP-GAN[9]、Dual-GAN[16]等模型作为对比模型,其中FFP-GAN模型属于较为经典的正向预测模型,并使用光流模型来添加运动特征;而Dual-GAN使用双判别器进行训练。根据表4结果可以发现,BiP-GAN模型的AUC高于目前无监督视频异常检测的经典模型方法,验证了双向预测策略及全局和局部特征结合的判别器在视频异常检测方面的有效性。UCSD pde2数据集在目前无监督视频异常检测领域获得了较高的AUC,BiP-GAN的视频异常检测AUC高于现有采用UCSD ped2数据集的典型方法。在ShanghaiTech数据集上,BiP-GAN模型的AUC也高于现有经典模型方法。

表3 Warm-up CAF学习率策略对BiP-GAN模型收敛时间和迭代次数的影响

Table 3 Influences of Warm-up CAF Learning Rate Strategy on Model Convergence Time and Iterations

学习率策略	时间/h	BiP-GAN模型收敛时迭代次数	AUC/%
CAF	3.8	36 000	85.7
	3.9	37 000	86.4
	4.1	39 000	86.2
Warm-up CAF	2.5	21 000	86.5
	2.1	18 000	86.4
	2.4	20 000	86.5

3 结 语

本研究针对GAN的视频异常事件检测,提出了一种基于BiP-GAN的行人视频异常检测模型。该模型能够在行人异常检测公开数据集上有效地检测出现异常事件的视频序列。本文提出的Warm-up与CAF相结合的学习率衰减方法,使模型能够更快地找到全局最优解,从而节省计算资源。这种学习率衰减方法可根据模型的具体情况调整公式中的对抗训练迭代次数或者更改分段函数中的第一段部分。使用该训练方法可以缩减模型的收敛时间,提高训练效率。本文的实验在公开数据集CUHK Avenue、UCSD ped2以及ShanghaiTech上证明了BiP-GAN模型的有效性。根据与现有基于预测方法进行视频异常事件检测的模型比较,BiP-GAN模型均达到较高的精度水平。但还需要进一步改进BiP-GAN模型的性能和鲁棒性,使其能够更好地

表 4　BiP-GAN模型与现有经典模型性能比较

Table 4　Performance Comparison Between BiP-GAN Model and Existing Classical Models

CUHK Avenue数据集		UCSD ped2数据集		ShanghaiTech数据集	
模型	AUC/%	模型	AUC/%	模型名称	AUC/%
ST-CaAE[15]	83.5	ST-CaAE[15]	92.9	Ada-GAN[10]	70.0
CCAE[14]	84.0	CCAE[14]	95.3	FFP-GAN[9]	72.8
FFP-GAN[9]	84.9	FFP-GAN[9]	95.4	3D U-Net[26]	73.6
ST-AE[16]	80.9	ST-AE[16]	91.2	Con-GAN[18]	73.6
3D U-Net[26]	86.0	Ada-GAN[10]	90.3	Mul-GAN[20]	73.6
Dual-GAN[16]	85.8	Con-GAN[18]	95.3	SIM-GAN[18]	73.1
SGCN[17]	82.0	TSSTGM[27]	95.2	SGCN[17]	71.7
SIGnet[28]	86.8	Mul-GAN[20]	95.4	AMAE[18]	73.6
BiP-GAN	87.3	BiP-GAN	96.2	BiP-GAN	73.9

处理复杂场景下的异常事件检测任务。

参　考　文　献

[1] GEORGESCU M I, BARBALAU A, IONESCU R T, et al. Anomaly Detection in Video via Self-Supervised and Multi-task Learning[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021.

[2] ZHANG M H, WANG J Y, QI Q, et al. Multi-scale Video Anomaly Detection by Multi-grained Spatiotemporal Representation Learning [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2024.

[3] TIAN Y, PANG G S, CHEN Y H, et al. Weakly-Supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning[C]//IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 2021.

[4] 徐晓. 基于卷积神经网络的视频监控异常事件检测研究[J]. 电子技术与软件工程, 2023(6): 193-197.
XU Xiao. Research on Abnormal Event Detection in Video Surveillance Based on Convolutional Neural Network[J]. *Electronic Technology and Software Engineering*, 2023(6): 193-197.

[5] 赵松, 傅豪, 王洪星. 伪异常选择驱动学习的视频异常检测[J]. 计算机科学, 2023, 50(5): 146-154.
ZHAO Song, FU Hao, WANG Hongxing. Video Anomaly Detection Driven by Pseudo-Anomaly Selection Learning[J]. *Computer Science*, 2023, 50(5): 146-154.

[6] 吴德刚, 赵利平, 陈乾辉, 等. 基于双尺度串行网络的视频异常行为检测[J]. 广西科学, 2023, 30(3): 575-586.
WU Degang, ZHAO Liping, CHEN Qianhui, et al. Video Abnormal Behavior Detection Based on Dual-Scale Serial Network[J]. *Guangxi Science*, 2023, 30(3): 575-586.

[7] 潘文康, 邵振峰, 廖明, 等. 利用深度时空自编码网络与多示例学习进行船只异常事件检测[J]. 武汉大学学报(信息科学版), 2024, 49(7): 1109-1119.
PAN Wenkang, SHAO Zhenfeng, LIAO Ming, et al. Ship Abnormal Event Detection Using Deep Spatiotemporal Autoencoder Network and Multi-instance Learning[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(7): 1109-1119.

[8] SHI H, DONG S, WU Y, et al. Generative Adversarial Network for Car Following Trajectory Generation and Anomaly Detection[J]. *Journal of Intelligent Transportation Systems*, 2025, 29(1): 53-66.

[9] LIU W, LUO W, LIAN D, et al. Future Frame Prediction for Anomaly Detection: A New Baseline [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.

[10] SONG H, SUN C, WU X, et al. Learning Normal Patterns via Adversarial Attention-Based Autoencoder for Abnormal Event Detection in Videos[J]. *IEEE Transactions on Multimedia*, 2019, 22(8): 2138-2148.

[11] LUO L, LI Y, YIN H, et al. Crowd-Level Abnormal Behavior Detection via Multi-scale Motion Consistency Learning[C]//AAAI Conference on Artificial Intelligence, Washington D C, USA, 2023.

[12] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning Optical Flow with Convolutional Networks[C]//IEEE International Conference on Computer Vision, Santiago, Chile, 2015.

[13] CHEN D, WANG P, YUE L, et al. Anomaly Detection in Surveillance Video Based on Bidirectional

Prediction[J]. *Image and Vision Computing*, 2020, 98: 103915.

[14] WANG J, ZHANG J, JI G, et al. Criss-Cross Attention Based Auto Encoder for Video Anomaly Event Detection [J]. *Intelligent Automation and Soft Computing*, 2022, 34(3): 1629-1642.

[15] LI N, CHANG F, LIU C. Spatial-Temporal Cascade Autoencoder for Video Anomaly Detection in Crowded Scenes[J]. *IEEE Transactions on Multimedia*, 2020, 23: 203-215.

[16] ZHAO Y, DENG B, SHEN C, et al. Spatio-temporal Autoencoder for Video Anomaly Detection [C]//ACM International Conference on Multimedia, Mountain View, CA, USA, 2017.

[17] XU J, MIAO Z, XU W, et al. Video Anomaly Detection Using Dual Discriminator Based Generative Adversarial Network[C]//IEEE International Conference on Machine Learning and Applications, Pasadena, CA, USA, 2021.

[18] LI D, NIE X, LI X, et al. Context-Related Video Anomaly Detection via Generative Adversarial Network[J]. *Pattern Recognition Letters*, 2022, 156: 183-189.

[19] SAYPADITH S, DETVONGSA S, ONOYE T. Joint Generative Network for Abnormal Event Detection in Surveillance Videos [C]//International SoC Design Conference, Jeju, Korea, 2022.

[20] LIU Y, LIU J, LIN J, et al. Appearance-Motion United Auto-Encoder Framework for Video Anomaly Detection[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022, 69(5): 2498-2502.

[21] AHIR A, PATERIYA P K, KAUR D, et al. A Review on Abnormal Activity Detection Methods [C]//International Conference on Computational Techniques, Electronics and Mechanical Systems, Pune, India, 2018.

[22] WANG B, NGUYEN T, SUN T, et al. Scheduled Restart Momentum for Accelerated Stochastic Gradient Descent[J]. *SIAM Journal on Imaging Sciences*, 2022, 15(2): 738-761.

[23] SHAMAEE M S, HAFSHEJANI S F, SAEIDIAN Z. New Logarithmic Step Size for Stochastic Gradient Descent[J]. *Frontiers of Computer Science*, 2025, 19(1): 1-10.

[24] SIMONYAN K, ZISSERMAN A. Two-Stream Convolutional Networks for Action Recognition in Videos [C]//Advances in Neural Information Processing Systems, Montreal, Canada, 2014.

[25] 肖进胜, 郭浩文, 谢红刚, 等. 监控视频异常行为检测的概率记忆自编码网络[J]. 软件学报, 2023, 34(9): 4362-4377.
XIAO Jinshen, GUO Haowen, XIE Honggang, et al. Probability Memory Autoencoder Network for Surveillance Video Abnormal Behavior Detection [J]. *Journal of Software*, 2023, 34(9): 4362-4377.

[26] YANG J X, CAI Y H, LIU D, et al. 3D U-Net for Video Anomaly Detection[C]//International Conference on Electronic Information Technology and Computer Engineering, Xiamen, China, 2021.

[27] LIU W, CAO J, ZHU Y, et al. Real-Time Anomaly Detection on Surveillance Video with Two-Stream Spatio-temporal Generative Model[J]. *Multimedia Systems*, 2023, 29(1): 59-71.

[28] ZHANG Y, NIE X, HE R, et al. Normality Learning in Multispace for Video Anomaly Detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 31(9): 3694-3706.