激光与光电子学进展
先进成像

基于深度学习的视频异常行为检测研究

彭嘉丽，赵英亮*，王黎明
中北大学信息探测与处理山西省重点实验室，山西 太原 030051

摘要 视频异常行为的检测对保障公共安全至关重要，对基于深度学习的异常行为检测算法进行了分类与总结。首先，介绍了异常行为检测的整体流程。然后，根据神经网络训练的方式，从有监督学习、弱监督学习和无监督学习三个方面论述了深度学习在异常行为检测领域的发展与应用，同时分析了不同训练方式的优缺点。最后，介绍了常用数据集以及性能评估准则，分析了不同算法的性能，并展望了未来的发展方向。

关键词 图像处理；深度学习；视频异常行为检测；有监督学习；弱监督学习；无监督学习
中图分类号 TP183 **文献标志码** A **doi**：10.3788/LOP202158.0600004

Research on Video Abnormal Behavior Detection Based on Deep Learning

Peng Jiali, Zhao Yingliang*, Wang Liming
Shanxi Key Laboratory of Signal Capturing and Processing, North University of China, Taiyuan, Shanxi 030051, *China*

Abstract The detection of video abnormal behavior is paramount to ensure public safety. In this paper, the abnormal behavior detection algorithm based on deep learning is classified and summarized. First, the overall process of abnormal behavior detection is presented. Then, based on the neural network training method, the development and application of deep learning in the field of abnormal behavior detection are discussed from three aspects: supervised learning, weakly supervised learning, and unsupervised learning, and the advantages and disadvantages of different training methods are analyzed. Finally, commonly used datasets and performance evaluation criteria are presented, the performance of the different algorithms is analyzed, and future directions are discussed.

Key words image processing; deep learning; video abnormal behavior detection; supervised learning; weakly supervised learning; unsupervised learning
OCIS codes 100.3008; 100.4996; 110.4155

1 引 言

随着监控摄像头在日常生活中的普及，监控视频数据呈爆炸式增长趋势。传统的人工异常事件检测算法会耗费大量人力资源，且疲劳工作或侥幸心理导致人工检测容易出现漏检问题。国内外研究人员研究了大量的监控视频异常行为检测算法，如何实时、精准地检测和定位异常行为也成为图像处理、机器视觉等领域的研究热点。异常行为检测的难点主要包括：1）没有明确定义正常、异常行为；2）不同场景下对异常行为的定义不同，导致异常行为检测系统难以泛化；3）行为种类繁多，无法穷举；4）异常

收稿日期：2020-06-19；**修回日期**：2020-07-16；**录用日期**：2020-08-31
基金项目：电子测试技术国防重点实验室基金（6142001180410）、山西省青年科技研究基金（201901D211250）、山西省高等学校优秀成果奖（科学技术）培育项目

* **E-mail**：zhaoyl18@nuc.edu.cn

行为发生的概率远低于正常行为，正负样本不均衡，难以学习足够多的异常行为特征。

现有异常行为检测算法的综述中，Afiq 等[1]将异常行为检测分为基于高斯混合模型（GMM）、隐马尔可夫模型、光流法和时空技术等传统算法，分类不恰当，且未囊括所有传统检测算法。Kiran 等[2]总结了半监督及无监督的端到端异常行为检测算法，忽略了定义异常行为的有监督算法。胡正平等[3]综述了 2019 年以前基于深度学习的异常行为检测算法，但文章中总结的算法较少，不够详细。王志国等[4]根据算法的发展阶段、模型类型以及异常行为判别标准对异常行为检测算法进行了比较全面的 3 级分类，但没有对算法进行深入分析。

针对上述文献中存在的分类不恰当、综述不全面、不详细、不深入问题，本文首先从有监督、弱监督和无监督三个方面对基于深度学习的视频异常行为检测算法进行了全面、深入的分析。然后对异常行为检测的整体流程进行了概述，并从神经网络训练的方式对不同算法进行了归纳和对比。最后介绍了常用数据集以及性能评估准则，并对未来的研究趋势进行了思考与展望。

2　异常行为检测概述

视频异常行为的检测与定位是指利用正常和异常行为特征表示之间的差异自动检测及定位异常行为，在安防领域具有重要意义，通常由前景提取与运动目标检测、特征提取、分类及异常行为检测三部分组成，如图 1 所示。首先预处理输入的视频序列，分离冗余的背景，提取运动前景，然后根据行为特征进行分类并检测异常行为。

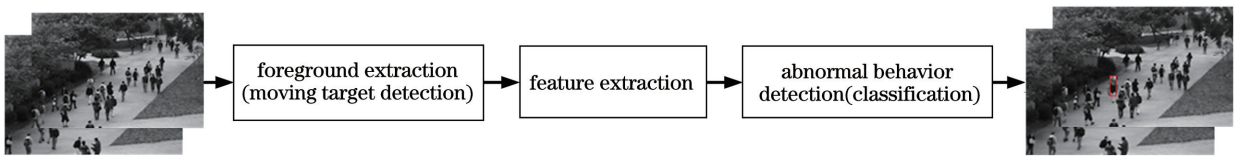

图 1　视频异常行为检测流程图

Fig. 1　Flow chart of the video abnormal behavior detection

传统的运动目标检测算法主要有帧差法、背景减除法以及光流法，随着深度学习在目标检测领域的发展[5]，目标检测网络被广泛用于前景提取及异常行为检测，如罗凡波等[6]利用 YOLOv3（You only look once）目标检测网络检测行人持棍、持枪、持刀以及面部遮挡等异常行为。

特征提取对于异常行为检测至关重要，正常、异常行为的特征区分度越高，检测精度就越高。传统基于手工构建的特征通过人为定义的低级视觉特征表征行为，如用方向梯度直方图（HOG）表征静态图像中的人体形状和轮廓信息，用光流描述相邻帧之间像素点灰度值的变化，从而表征运动信息[7]，用轨迹（trajectory）规划描述运动目标的轨迹。但手工制作的特征无法表征复杂的行为，且提取的特征比较单一，导致基于手工构建的特征泛化能力较弱。基于深度学习提取的特征能自动从海量数据集中学习数据的分布规律，提取更鲁棒的高级语义特征，对场景拥挤的情况不敏感，逐渐取代了传统特征提取算法。如彭月平等[8]利用三维卷积神经网络（C3D）[9]提取 HOG 时空特征，提高了算法对人群行为的表征能力。

传统的异常行为检测算法在提取手工构建的特征后还需训练分类器以检测异常行为。常用的分类器包括：1）基于有监督训练的分类器，如支持向量机（SVM）、随机森林[10]、朴素贝叶斯分类器；2）半监督分类器，如多实例支持向量机（MISVM）、稀疏字典；3）基于聚类的分类器，如无监督算法训练的一类支持向量机（OCSVM）、高斯分类器。基于深度学习提取特征后既可使用分类器分类正常、异常行为，也可直接使用端到端的神经网络实现异常检测。

综上所述，传统的异常行为检测算法存在人工参与多、不够客观且严重依赖场景等问题。基于深度学习的异常行为检测算法泛化能力强，易于场景迁移，能识别更多的行为类型，已成为近几年的研究热点，未来更趋向于基于深度学习的端到端异常行为检测算法。

3　深度学习在异常行为检测领域的发展及应用

深度学习具有出色的特征提取效果及强大的数据拟合能力，检测精度较高，是异常行为检测领域中的主流研究算法。按照训练神经网络的数据类型及其标签类型可将基于深度学习的异常行为检测分为有监督、弱监督以及无监督三类，如图 2 所示。

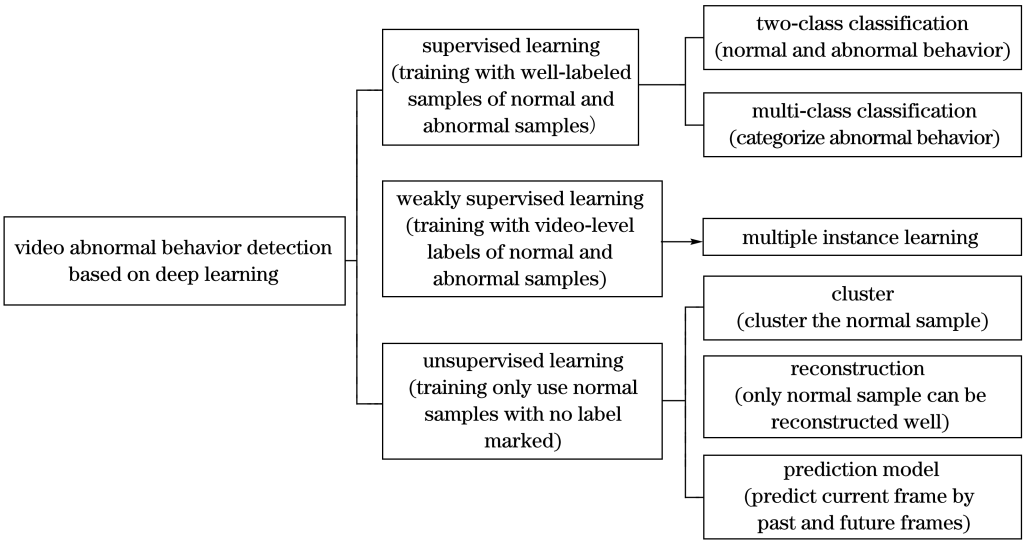

图 2　基于深度学习的异常行为检测分类

Fig. 2　Abnormal behavior detection classification based on deep learning

3.1　有监督异常行为检测

有监督算法定义异常行为时将异常检测视为二分类或多分类问题，即用详细标记的正常、异常行为样本训练神经网络，提取正常、异常行为之间更具区分性的特征。二分类即对正常、异常行为进行分类，多分类在检测异常行为的基础上进一步识别具体的异常行为。

3.1.1　二分类异常行为检测

二分类异常行为检测将正常和异常行为视为不同的类别，利用有监督学习将行为归为正常或异常行为。由于视频集成了时间和空间信息，因此异常行为检测需要使用能提取时空特征的神经网络，如三维卷积神经网络[11]、循环神经网络和双流网络模型。

与二维(2D)卷积相比，三维(3D)卷积多了一个时间维度，在行为识别任务中表现优异。胡学敏等[12]设计的深度时空卷积神经网络(DSTCNN)将视频帧划分为大小相同且互不重叠的子区域，以实现异常人群定位，然后将子区域输入改进的 C3D 模型提取行为特征并输出正常、异常行为的分类概率。Gong 等[13]提出的局部分辨率增强网络(LDA-Net)将 YOLO 网络提取的前景人体作为 3D CNN 的输入，以提取行为的时空特征，实现正常、异常行为的分类。维数的上升使超参数的数量暴增，计算代价过高，难以满足实时性要求。为了减小计算量，伪三维残差网络[14]和 2+1 维残差网络(R(2+1)D)[15]将 3×3×3 的 3 维卷积核拆分成 1×3×3 的空间卷积核和 3×1×1 的时间卷积核，为异常行为的检测提供了新思路，与 3D 卷积网络相比，(2+1)D 卷积网络在提高检测精度的同时大大减少了运算量。

循环神经网络具有处理时间序列的能力，长短期记忆(LSTM)网络作为循环神经网络的变体在长时序建模的同时解决了梯度弥散和梯度爆炸问题。胡薰尹等[16]将对光照和背景变化不敏感的三通道矫正光流运动历史图像输入 3D 卷积核，获取短时序特征，结合 LSTM 网络进一步提取长时序信息。3D-LSTM 网络结构克服了 3D CNN 只能对短时序运动信息建模的缺点，提高了算法的准确度，在 UMN 数据集上的识别准确率达到 99%，但计算复杂，实时性差。全卷积神经网络(FCN)可以实现像素级分类与定位，武光利等[17]提出利用 FCN 提取图像特征，再以时间为轴线输入到 LSTM 网络中提取行为的语义特征，最后通过上采样直接输出异常区域的标记，实现异常行为的精准定位。

针对不同动作持续时间不同的问题，张乐[18]采用双流三维卷积神经网络(TW-C3D)模型分别处理两种帧长的时空兴趣块，融合不同时间尺度的判别结果能更充分地学习行为的时空特征，同时提出结合非均匀视频帧分割和光流法提取前景的算法，以精确定位异常行为。

3.1.2　多分类异常行为识别

多分类异常行为检测将异常行为进行具体分类，以检测并识别异常行为。由于异常行为发生的概率较低，可用于训练的样本较少，李辰政等[19]利用迁移学习训练 C3D 模型提取行为特征，并对持械、打架斗殴、挟持和抢劫这 4 类危险行为进行分

类，平均识别率达到 83.2%。LSTM 网络能够记忆长时间的时序信息，对某些持续时间较长的行为具有更好的识别效果。邹云飞[20]将 YOLO 网络提取的特征向量输入 LSTM 网络，这种 CNN＋LSTM 的结构充分利用了视频的时空融合特征，可实现监狱场景下的实时暴力、非法越线和异常奔跑行为的识别，每秒可处理 63 帧图像，识别精度达到 87%。

双流网络可将外观信息和运动信息分流提取[21]，夏清[22]将原始图像及其帧差图像输入双流网络，再利用 SVM 实现分类，对拥挤和不拥挤场景下暴力行为的检测精度分别达到 93.25% 和 95.90%，延迟约为 1 s。在此基础上，苑鑫鑫[23]在空间流采用 C3D 模型提取连续视频帧的时空特征，时间流采用堆叠的光流帧作为输入，由双流分类概率线性加权融合得到最终分类结果。3D 卷积以及光流图像包含了更多的运动信息，进一步提升了模型的检测精度，在拥挤和非拥挤场景下的暴力行为识别准确率分别达到 96% 和 99%，但计算复杂，延时较长。

除神经网络外，时空流融合策略对双流模型的性能也有很大影响。在双流模型中，时间流识别精度普遍高于空间流[24]。高阳[25]研究了不同层融合时空特征对检测精度的影响，与异常得分融合相比，在全连接层融合时空流特征的识别准确率更高，对打架、抢劫、晕倒以及砸东西 4 类异常行为的识别准确率分别达到 87.9%、88.7%、88.9%、87.5%。Zhou 等[26]创造性地将 2D 和 3D 卷积所有可能的融合策略嵌入到一个概率空间中，将融合策略作为一个优化问题，通过网络对各种融合策略进行评估。

由于利用了充足的先验信息进行训练，有监督算法的识别和定位精度普遍较高，在现实生活中被广泛使用。但其只能检测预先定义好的异常行为，仅适合在已知所有可能出现异常行为的特定场景下使用，同时繁琐的人工标注限制了有监督算法的发展。

3.2　弱监督异常行为检测

弱监督算法仅给出训练样本视频级的正常或异常标签，即在训练时只知道一段视频中有没有异常行为，而不知道异常行为的具体种类及发生时间。

弱标签下的异常行为检测是典型的多实例学习问题，Sultani 等[27]使用 C3D 提取视频片段特征，结合 3 层全连接神经网络预测异常得分，在 UCF-Crime 数据集上的检测精度为 75.4%。运动信息对异常行为的检测十分关键，针对上述多实例学习排序损失忽略了潜在时间结构的问题，Zhu 等[28]用注意力模块加强网络对运动特征的学习能力，并通过实验验证了引入注意力模块可提高异常检测的精度。该算法对 UCF-Crime 数据集中逮捕、攻击和打架这 3 类异常行为的识别效果较好，但平均 AUC（Area under curve）只有 72.1%，低于文献[27]中的 75.4%，但该模型的参数更少，检测速度更高，每秒传输帧数（FPS）为 400 帧（文献[27]中算法的 FPS 仅为 300 帧）。Zhong 等[29]提出利用图卷积网络校正异常视频中正常片段的噪声标签，用校正后的标签训练动作分类器检测异常行为，在 UCF-Crime 数据集上的帧级 AUC 为 82.12%。视听结合能有效提高模型的检测性能，Wu 等[30]使用多模态信息作为输入，同时提取视频和音频特征实现在线异常行为检测，在 UCF-Crime 数据集上的帧级 AUC 达到了 82.44%。

将目标检测网络应用于异常行为检测任务中，使模型从对象的角度分析目标行为，可显著提高检测速度及模型的泛化能力。Hu 等[31]首先利用目标检测网络提取视频前景目标，然后用大尺度光流直方图描述符描述对象的行为，最后利用多实例支持向量机（MISVM）分类正常、异常行为。该系统在不需要完全标记的条件下获得了较高的识别精度，在人群遮挡严重的情况下具有较强的鲁棒性，在 UMN 数据集上的 AUC 达到了 98.9%。开集（openset）指用训练集中未出现过的异常类型测试模型性能，以增强模型的泛化能力。周培培等[32]设计了一种基于开集的边缘学习嵌入式预测（MLEP）算法，结合二维卷积编码器和 Conv-LSTM 网络，能有效区分正常、异常行为。以视频级标签训练网络时，在 Avenue 数据集上的平均 AUC 为 91.3%，以帧级标签训练时平均 AUC 达到 92.8%。

使用正常和异常行为数据训练能提高网络的学习能力，最大化正常、异常行为的特征差异。相比有监督算法使用的数据集，弱标签数据集的制作更简便，因此，弱监督异常行为检测更易操作和泛化，但误检和漏检概率较大。随着监控视频数据的暴增，即使是视频级的标签处理也会变得繁琐，因此，人们转而研究更智能化的无监督算法。

3.3　无监督异常行为检测

无监督算法无需任何标签信息，基于非正即异的思想，假设异常行为是罕见且无规律的。通过学习大量正常行为的特征表示，将不符合正常特征分

布的样本视为异常行为，具体可分为基于聚类判别、基于重构判别和基于预测模型三种算法。

3.3.1　基于聚类判别的异常行为检测

基于聚类判别的异常行为检测通过拟合正常样本空间并对正常样本进行聚类，将远离正常样本聚类中心的样本识别为异常行为。常用的聚类算法有一类分类器[33]和 GMM。

一类分类器用于寻找一个囊括正常行为特征的超平面，将不在圈内的样本判断为异常行为。Xu 等[34]提出的 AMDN(Appearance and motion deep network)将 RGB(Red，Green，Blue)图像、光流图像及其融合图像分别输入三个相同的堆叠降噪自编码器，以提取特征，融合三个一类分类器的异常行为判断结果检测异常行为。但这不是端到端的网络，丢失了实时性。雷丽莹[35]利用稀疏自编码器对 AlexNet 提取的特征进行降维处理，再输入一类支持向量机检测异常行为，但 AlexNet 容易压缩视频中的时间信息，可能会丢失运动相关性。

GMM 是多个高斯分布函数的线性组合，理论上可以拟合任意类型的分布。Fan 等[36-37]设计的高斯混合全卷积变分自编码器(GMFC-VAE)和多元高斯全卷积对抗自编码器(MGFC-AAE)利用双流自编码器提取测试样本的特征并与拟合正常样本特征空间分布的多元混合 GMM 进行对比，将不属于任何高斯分量的行为判断为异常行为。此外，针对监控视频帧中同一目标由相对位置变化引起的面积和速度变化问题，设计了一种多尺度分块结构，有效解决了由相机视角引起的透视问题。

视频帧分块卷积容易分割目标，但耗时较长，因此，Sabokrou 等[38]直接利用全卷积神经网络实现异常行为定位，每秒可处理约 370 帧任意尺寸的视频。首先利用预训练的前两层 AlexNet 提取测试样本特征，然后用第一个高斯分类器 G_1 对特征进行分类，分类规则可表示为

$$G_1\ \text{is}\begin{cases}\text{normal behavior}, & \text{if } d\leqslant\beta\\ \text{suspicious behavior}, & \text{if } \beta<d<\alpha\\ \text{abnormal behavior}, & \text{if } d\geqslant\alpha\end{cases},\tag{1}$$

若正常样本特征空间与测试样本特征的欧氏距离(d)不大于阈值 β 则为正常行为，不小于阈值 α 则为异常行为，介于 α 与 β 之间则为可疑样本。将可疑样本再输入稀疏自编码器中提取更具区分性的正常、异常行为特征并利用第二个高斯分类器分类可疑样本，检测异常行为，级联检测的方式增强了算法性能。基于聚类的算法缺乏异常行为的先验信息，无法提取正常、异常行为间有区分性的关键特征。同时，由于正常行为样本的类间差异大，需要大量正常行为数据进行训练，否则容易造成正常行为的误检。

3.3.2　基于重构判别的异常行为检测

重构图像实际上是对输入帧进行编码提取特征，再将特征解码为重构图像的操作。基于重构的算法假设仅在正常数据上学习的模型不能准确表示和重构异常，以重构误差作为异常得分以检测异常行为。

卷积自编码器(CAE)常用于重构输入图像，Ribeiro 等[39]将 CAE 对原始图像、边缘图像以及光流图像的重构误差作为异常评分，但二维卷积无法捕获时域信息。Zhao 等[40]提出的时空自编码器(STAE)利用 3D 卷积网络对视频片段进行编码，获取特征的低维表示，再用 3D 反卷积网络进行解码，在解码过程中使用了双路模型，一路重构过去的行为，一路预测下一时刻的行为，增强了运动特征。何丹丹等[41-42]基于双流结构，用 Conv-LSTM 网络自编码器重构的短期光流序列作为时间流，用卷积自编码网络重构的梯度图像作为空间流，采用贝叶斯融合算法融合双流重构误差值，将融合后重构误差超过阈值的行为视为异常行为。

同时提取视频外观和运动特征能使网络对遮挡的鲁棒性更强，Chong 等[43]利用端到端的时空自编码器重构输入帧，用 2D 卷积核提取空间特征，用 LSTM 网络获取空间特征的时间演化信息。该算法每秒可处理 140 帧图像，但不能很好地定位异常行为的终止时间，可能会产生假警报。岳海纯[44]采用 3D 编码＋LSTM 网络＋3D 反卷积解码的结构重构图像，同时在编码器和解码器之间增加跳层连接，使重构图像更完整，提高了检测精度。

Wang 等[45]设计了由两个变分自编码网络(VAE)级联而成的 S^2-VAE，首先用堆栈全连接变分自编码器(SF-VAE)生成正常样本的类 GMM 分布，从而在测试阶段滤除明显的正常样本；然后根据跳跃卷积变分自编码器(SC-VAE)的重构误差检测异常行为，结合聚类判别和重构判别得到了更精确的检测结果。生成对抗网络(GAN)由生成器和鉴别器组成，通过二者的对抗训练，生成器产生的重构误差越来越小，鉴别器的判断能力也得到了提高[46]。武慧敏[47]以 VAE 作为 GAN 的生成器，采用双流结构分别重构原始图像以及密集光流场，利用重构误差以及对抗损失优化网络，但该网络的异

常行为检测和定位不够精准且检测结果有延时，原因是密集光流的计算量大。

U-Net 的跳层连接结构将输入帧和输出帧的共同特征直接跨层传输，在减少参数量的同时提高了网络对运动信息的学习能力。Ravanbakhsh 等[48-49]利用跨通道的 U-Net 作为 GAN 的生成器，一路由 RGB 图像生成光流图，另一路由相应的光流图像生成 RGB 图像，通过生成图像与真实图像的局部差异定位异常区域。Akcay 等[50]采用编码-解码-再编码的结构，在常规 GAN 的基础上增加了一个再编码器，利用重构损失、再编码损失以及对抗损失共同优化网络，从而检测异常行为。神经网络强大的学习能力可以很好地重构异常行为，但基于重构判别的算法容易漏检异常行为，同时由于正常行为是无法穷尽的，新出现的正常行为也容易被误检。

3.3.3　基于预测模型的异常行为检测

预测模型假设正常行为是有规律且可预测的，而异常行为是不可预测的，通过预测误差即可检测异常行为，具体可分为单向预测和双向预测。单向预测将当前输入帧建模为过去 t 帧的函数，以达到预测正常视频帧的目的。Liu 等[51]使用 U-Net 作为 GAN 的生成器预测下一帧图像，并通过像素强度损失、梯度损失、光流损失和对抗损失优化网络，测试阶段将预测帧与真实帧之间的峰值信噪比(PSNR)作为异常得分检测和定位异常。Li 等[52]提出了一种具有多尺度信息的时空 U-Net 用于预测下一帧图像，在提取空间信息的 U-Net 中加入运动信息的 Conv-LSTM 网络，利用 RGB 差分代替光流损失优化生成器，减少了损失计算的时间。针对异常目标出现在视频帧边缘时容易发生漏检的问题，设计了一种新的异常分数计算函数，能更加精准地定位异常行为的起止时间。

针对单向预测模型未能充分利用时间信息的缺点，Chen 等[53]提出一种双向预测框架(Bi-prediction)，即根据目标帧的前 t 帧和后 t 帧进行双重预测，结合两种预测帧的交叉均方误差和预测帧与真实帧的均方误差作为损失函数，以 PSNR 作为异常得分。同时，提出了抑制预测图像噪声的滑动窗口方案，将注意力集中于前景目标，提高了模型的鲁棒性。该模型结构简单、参数少、泛化能力强，但只能在异常行为发生后检测异常，无法实时预警。

人类行为固有的不可预测性注定了基于预测模型的异常行为检测只能在理想状态下使用，无法投入实际应用。基于无监督的异常行为检测只利用正常样本训练的方式使网络无法针对异常行为的误检进行优化，且无监督算法中基于理想情况的假设限制了其实用性。表 1 为有监督、弱监督和无监督异常行为检测算法的优缺点，可以发现，3 类算法在识别精度、实用性和人工参与程度方面依次递减，但在智能化和泛化能力方面依次递增。

表 1　3 种异常行为检测算法的优缺点

Table 1　Advantages and disadvantages of the 3 abnormal behavior detection algorithms

Algorithm	Advantage	Disadvantage
Supervised learning	relatively less training samples is needed; the most accurate; easy to understand and apply	accurate label is time-consuming; only predefined anomaly can be detected; hard to generalize
Weakly supervised learning	relatively accurate with weak label; lower false alarm; relatively easy to use	split the difference; relatively lower accuracy
Unsupervised learning	training without any label; only normal data needed; robust and easy to generalize	higher false positive rate; poor positioning accuracy; unable to classify anomaly behavior

4　常用数据集与评估准则

4.1　常用数据集

目前视频异常行为检测领域常用的公共数据集有 UCSD、Avenue、UMN、UCF-Crime 数据集，表 2 为不同数据集的视频数量、总时长、标签情况以及异常种类。其中，UCSD 数据集共 2 个子集(Ped1 和 Ped2)，分别记录了校园内垂直和水平方向人行道上固定视角的监控视频。训练视频中只有正常行为，即人行道上的正常行走；测试视频包含正常行为和骑自行车、滑滑板、轮椅、开汽车和草地上行走等异常行为，并对异常区域进行了像素级掩码标注。

表 2　常用异常行为检测数据集的对比

Table 2　Comparison of commonly used anomaly behavior detection datasets

Dataset	Video	Length	Annotation	Anomaly category
UCSD	98	10 min	pixel-level	biker, skater, wheelchair, car, walking on the grass
Avenue	37	30 min	frame-level	run, throw, abnormal object
UMN	5	5 min	frame-level	group escape
UCF-Crime	1900	128 h	video-level	abuse, arrest, arson, assault, accident, burglary, fighting, robbery

Avenue 数据集场景固定，异常行为定义为一些奇怪的动作，如奔跑、投掷物体、游荡、异常物体和错误方向，以帧级标签标记，其训练视频中少量无标记的异常行为和测试视频中的相机抖动给检测带来了挑战。此外，训练数据中的正常模式比较简单，出现新的正常行为时易发生误检。UMN 数据集分别在草坪、室内和广场三种场景下以固定摄像机拍摄，分辨率为 320 pixel×240 pixel。正常行为包括正常的闲逛、人群聚集谈话，异常行为为人群单方向跑动、四散逃逸。该数据集主要针对群体异常行为识别，只提供帧级标注。UCF-Crime 数据集由只有视频级标签的 1900 个未剪辑的长时监控视频组成，涵盖真实场景中 13 种危害公共安全的异常事件，包括虐待、逮捕、纵火、攻击、事故、入室盗窃、爆炸、打架、抢劫、枪击、偷盗、入店行窃、破坏公物。该数据集对异常行为进行了具体分类，可用于行为识别任务。

4.2　性能评估准则

在异常行为检测领域，通过对异常分数或异常概率取不同阈值绘制的接收者操作特征曲线(ROC)定性评估算法的性能，通常用识别精确度(ACC)或 ROC 与坐标轴围成的面积(AUC)和等错误率(EER)定量评价算法的性能。ROC 以伪阳性率(FPR)为横坐标，真阳性率(TPR)为纵坐标，伪阳性率指在所有实际负例的样本中预测为正例的概率，真阳性率为所有实际正例的样本中预测为正例的概率。因此 ROC 越接近左上角，EER 越小，AUC 越大，算法性能越好。

异常行为检测需要检测异常行为发生的时间和空间位置，通常从帧级和像素级两个层次评价检测效果。在帧级准则中，只要某帧中有一个像素被检测为异常，则将该帧视为异常帧，不考虑对异常区域的定位是否准确。而像素级准则还需考虑空间定位精度，只有检测到的异常像素覆盖了至少 40%的真实异常标记时，才认为出现异常行为。

4.3　不同算法的性能

表 3 为不同异常行为检测算法在 UCSD、Avenue、Subway 和 UMN 数据集上不同评估准则下的 AUC。可以发现：1)有监督算法的异常行为检测精度较高，其次为弱监督算法；2)基于无监督算法的异常行为检测研究较多；3)用 3D 卷积或 LSTM 建模时序信息的网络性能普遍比只使用 2D 卷积的网络好，这表明运动信息的提取可以提高网络对异常行为识别的精度；4)除 S^2-VAE 外，无监督异常行为检测算法在像素级准则下的识别精度普遍偏低，这表明无监督算法对异常行为的定位能力较差；5)目前没有出现在任何场景下检测精度都最优的算法，基于深度学习的异常行为检测算法仍有很大的提升空间。

表 3　不同算法的性能对比

Table 3　Performance comparison of different algorithms　　unit：%

Algorithm	UCSD Ped1				UCSD Ped2				Avenue	
	Frame-level		Pixel-level		Frame-level		Pixel-level		Frame-level	
	EER	AUC	EER	AUC	EER	AUC	EER	AUC	EER	AUC
DSTCNN[12]	–	**99.74**	–	–	–	**99.94**	–	–	–	–
LDA-Net[13]	–	–	–	–	5.63	97.87	12.91	92.96	–	–
FCN+LSTM[17]	–	–	–	–	–	–	**6.6**	**98.2**	–	–
TW-C3D[18]	**6.29**	96.73	**9.22**	**95.27**	**5.59**	96.37	11.80	93.51	–	–
MISVM[31]	22	–	–	–	16	–	–	–	**21**	84.5

续表 3

Algorithm	UCSD Ped1				UCSD Ped2				Avenue	
	Frame-level		Pixel-level		Frame-level		Pixel-level		Frame-level	
	EER	AUC	EER	AUC	EER	AUC	EER	AUC	EER	AUC
MLEP[32]	–	–	–	–	–	–	–	–	24.8	**92.8**
AMDN[34]	16.0	92.1	40.1	67.2	17.0	90.8	–	–	–	–
OC SVM[35]	–	–	–	–	10.6	93	17.3	88	–	–
GMFC-VAE[36]	11.3	94.9	36.3	71.4	12.6	92.2	19.2	78.2	22.7	83.4
MGFC-AAE[37]	20	85	–	72.6	16	91.6	–	88	22.3	84.2
CAE[39]	–	89.5	–	–	–	54.7	–	–	–	75.4
STAE[40]	15.3	92.3	–	–	16.7	91.2	–	–	24.4	80.9
LSTM AE[43]	12.5	89.9	–	–	12.0	87.4	–	–	–	80.3
3D-LSTM AE[44]	15.9	90.9	–	–	15.8	93.6	–	–	20.7	81.8
S^2-VAE[45]	14.3	–	–	94.25	11.54	95.77	14.28	90.83	–	87.6
GAN[48]	8	97.4	35	70.3	14	93.5	–	–	–	–
Predict-GAN[51]	–	83.1	–	–	–	95.4	–	–	–	84.9
ST U-Net[52]	22.3	83.82	–	–	8.7	96.56	–	–	–	84.59
Bi-Prediction[53]	–	89.0	–	–	–	96.6	–	–	21.5	87.8

5　结　　论

概述了现有异常行为检测算法及其常用数据集和评估准则，总结了深度学习在异常行为检测领域的发展及应用，从原理上对不同算法进行了分类，最后总结了不同算法的优缺点。在异常行为检测领域，无监督算法由于无需人工标记、泛化能力强成为近几年的研究热点，但实际应用中考虑到检测及定位精度，有监督算法仍占据更多市场、落地性更强。

除了改进神经网络以外，针对无监督算法需要训练大量正常样本，有监督和弱监督中正常、异常样本不均衡的问题，基于迁移学习训练的算法可有效改善网络性能，防止过拟合。目前大部分的异常行为检测算法基于闭集测试，即测试集中出现的所有行为类别均被训练过，测试集的检测结果只作为调参标准，而基于开集训练的模型泛化能力更强。因此，在测试集中加入训练集中没有过的类别是新的研究方向。此外，利用基于深度学习的目标检测网络提取前景、滤除冗余背景信息，引入注意力机制加大有区分性的特征权重，也可以进一步提高检测精度。

参　考　文　献

[1] Afiq A A, Zakariya M A, Saad M N, et al. A review on classifying abnormal behavior in crowd scene[J]. Journal of Visual Communication and Image Representation, 2019, 58: 285-303.

[2] Kiran B, Thomas D, Parakkal R. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos [J]. Journal of Imaging, 2018, 4(2): 36.

[3] Hu Z P, Zhang L, Li S F, et al. Review of abnormal behavior detection and location for intelligent video surveillance systems[J]. Journal of Yanshan University, 2019, 43(1): 1-12.
胡正平, 张乐, 李淑芳, 等. 视频监控系统异常目标检测与定位综述[J]. 燕山大学学报, 2019, 43(1): 1-12.

[4] Wang Z G, Zhang Y J. Anomaly detection in surveillance videos: a survey[J]. Journal of Tsinghua University (Science and Technology), 2020, 60(6): 518-529.
王志国, 章毓晋. 监控视频异常检测: 综述[J]. 清华大学学报(自然科学版), 2020, 60(6): 518-529.

[5] Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.
段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12): 120005.

[6] Luo F B, Wang P, Liang S Y, et al. Crowd abnormal behavior recognition based on deep learning

and sparse optical flow[J]. Computer Engineering, 2020, 46(4): 287-293, 300.
罗凡波, 王平, 梁思源, 等. 基于深度学习与稀疏光流的人群异常行为识别[J]. 计算机工程, 2020, 46(4): 287-293, 300.

[7] Hu X M, Yu J, Deng C Y, et al. Abnormal crowd behavior detection and location based on spatial-temporal cube [J]. Geomatics and Information Science of Wuhan University, 2019, 44(10): 1530-1537.
胡学敏, 余进, 邓重阳, 等. 基于时空立方体的人群异常行为检测与定位[J]. 武汉大学学报·信息科学版, 2019, 44(10): 1530-1537.

[8] Peng Y P, Jiang R Q, Xu L. An algorithm for identifying crowd abnormal behavior based on C3D-GRNN model [J]. Measurement & Control Technology, 2020, 39(7): 44-50.
彭月平, 蒋镕圻, 徐蕾. 基于 C3D-GRNN 模型的人群异常行为识别算法[J]. 测控技术, 2020, 39(7): 44-50.

[9] Tran D, Bourdev L, Fergus R, et al. Learning spatiotemporal features with 3D convolutional networks[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4489-4497.

[10] Yang X X, Li H B, Hu G. An abnormal behavior detection algorithm based on imbalanced deep forest [J]. Journal of China Academy of Electronics and Information Technology, 2019, 14(9): 935-942.
杨欣欣, 李慧波, 胡罡. 一种基于不平衡类深度森林的异常行为检测算法[J]. 中国电子科学研究院学报, 2019, 14(9): 935-942.

[11] Ji S W, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.

[12] Hu X M, Chen Q, Yang L, et al. Abnormal crowd behavior detection and localization based on deep spatial-temporal convolutional neural network [J]. Application Research of Computers, 2020, 37(3): 891-895.
胡学敏, 陈钦, 杨丽, 等. 基于深度时空卷积神经网络的人群异常行为检测和定位[J]. 计算机应用研究, 2020, 37(3): 891-895.

[13] Gong M G, Zeng H M, Xie Y, et al. Local distinguishability aggrandizing network for human anomaly detection[J]. Neural Networks, 2020, 122: 364-373.

[14] Qiu Z F, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3D residual networks [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5534-5542.

[15] Tran D, Wang H, Torresani L, et al. A closer look at spatiotemporal convolutions for action recognition [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6450-6459.

[16] Hu X Y, Guan Y P. 3D-LCRN based video abnormal behavior recognition[J]. Journal of Harbin Institute of Technology, 2019, 51(11): 183-193.
胡薰尹, 管业鹏. 基于 3D-LCRN 视频异常行为识别方法[J]. 哈尔滨工业大学学报, 2019, 51(11): 183-193.

[17] Wu G L, Guo Z Z, Li L T, et al. Video abnormal detection combine FCN with LSTM[J]. Journal of Shanghai Jiaotong University, 2020, 120: 1-8.
武光利,郭振洲,李雷霆,等. 融合 FCN 和 LSTM 的视频异常事件检测[J]. 上海交通大学学报, 2020, 120: 1-8.

[18] Zhang L. Anomaly detection and localization in video surveillance by deep neural network[D]. Qinhuangdao: Yanshan University, 2019.
张乐. 深度神经网络视频异常目标检测与定位算法研究[D]. 秦皇岛: 燕山大学, 2019.

[19] Li C Z, Zhang X J, Zhu H T, et al. Research on dangerous behavior identification method based on transfer learning [J]. Science Technology and Engineering, 2019, 19(16): 187-192.
李辰政, 张小俊, 朱海涛, 等. 基于迁移学习的危险行为识别方法研究[J]. 科学技术与工程, 2019, 19(16): 187-192.

[20] Zou Y F. Recognition and research about abnormal behavior of human based on video[D]. Kunming: Yunnan University, 2019.
邹云飞. 基于视频的人体异常行为识别与研究[D]. 昆明: 云南大学, 2019.

[21] Wang Z W, Gao B P. Spatio-temporal fusion convolutional neural network for abnormal behavior recognition[J]. Computer Engineering and Design, 2020, 41(7): 2052-2056.
王泽伟, 高丙朋. 基于时空融合卷积神经网络的异常行为识别[J]. 计算机工程与设计, 2020, 41(7): 2052-2056.

[22] Xia Q. Research on crowd abnormal behavior detection in video surveillance [D]. Chengdu: University of Electronic Science and Technology of China, 2019.
夏清. 视频监控中的人群异常行为检测研究[D]. 成

都：电子科技大学，2019.

[23] Yuan X X. Research on video violence behavior detection algorithm of deep convolutional network [D]. Qinhuangdao: Yanshan University, 2019.
苑鑫鑫. 深度卷积网络视频暴力行为检测算法研究[D]. 秦皇岛：燕山大学，2019.

[24] Zhang M. Research on human abnormal behavior detection based on deep learning [D]. Xi′an: Xi′an University of Science and Technology, 2019.
张梦. 基于深度学习的人体异常行为识别研究[D]. 西安：西安科技大学，2019.

[25] Gao Y. Fighting behavior detection in surveillance video by two-stream convolutional networks [D]. Xi′an: Xi′an University of Technology, 2018.
高阳. 基于双流卷积神经网络的监控视频中打斗行为识别研究[D]. 西安：西安理工大学，2018.

[26] Zhou Y Z, Sun X Y, Luo C, et al. Spatiotemporal fusion in 3D CNNs: a probabilistic view [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 9826-9835.

[27] Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6479-6488.

[28] Zhu Y, Newsam S. Motion-aware feature for improved video anomaly detection [EB/OL]. [2020-06-01]. https://arxiv.org/abs/1907.10211.

[29] Zhong J X, Li N N, Kong W J, et al. Graph convolutional label noise cleaner: train a plug-and-play action classifier for anomaly detection [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1237-1246.

[30] Wu P, Liu J, Shi Y J, et al. Not only look, but also listen: learning multimodal violence detection under weak supervision [M]//Vedaldi A, Bischof H, Brox T, et al. Computer Vision-ECCV 2020. Lecture Notes in Computer Science. Cham: Springer, 2020, 12375: 322-339.

[31] Hu X, Dai J, Huang Y P, et al. A weakly supervised framework for abnormal behavior detection and localization in crowded scenes [J]. Neurocomputing, 2020, 383: 270-281.

[32] Zhou P P, Ding Q H, Luo H B, et al. Anomaly detection and location in crowded surveillance videos [J]. Acta Optica Sinica, 2018, 38(8): 0815007.
周培培，丁庆海，罗海波，等. 视频监控中的人群异常行为检测与定位[J]. 光学学报，2018，38(8)：0815007.

[33] Liu W, Luo W X, Li Z X, et al. Margin learning embedded prediction for video anomaly detection with a few anomalies [C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, August 10-16, 2019. Macao, China. California: IJCAI, 2019: 3023-3030.

[34] Xu D, Yan Y, Ricci E, et al. Detecting anomalous events in videos by learning deep representations of appearance and motion [J]. Computer Vision and Image Understanding, 2017, 156: 117-127.

[35] Lei L Y. Video anomaly detection based on deep learning [D]. Hangzhou: Hangzhou Dianzi University, 2018.
雷丽莹. 基于深度学习的视频异常检测[D]. 杭州：杭州电子科技大学，2018.

[36] Fan Y X, Wen G J, Li D R, et al. Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder [J]. Computer Vision and Image Understanding, 2020, 195: 102920.

[37] Li N J, Chang F L. Video anomaly detection and localization via multivariate Gaussian fully convolution adversarial autoencoder [J]. Neurocomputing, 2019, 369: 92-105.

[38] Sabokrou M, Fayyaz M, Fathy M, et al. Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes [J]. Computer Vision and Image Understanding, 2018, 172: 88-97.

[39] Ribeiro M, Lazzaretti A E, Lopes H S. A study of deep convolutional auto-encoders for anomaly detection in videos [J]. Pattern Recognition Letters, 2018, 105: 13-22.

[40] Zhao Y, Deng B, Shen C, et al. Spatio-temporal autoencoder for video anomaly detection [C]// Proceedings of the 25th ACM International Conference on Multimedia, October 23-27, 2017, Mountain View, CA, USA. New York: ACM, 2017: 1933-1941.

[41] He D D. Abnormal behavior detections under surveillance video scenes [D]. Wuxi: Jiangnan University, 2018.
何丹丹. 监控视频场景下的异常行为检测研究[D]. 无锡：江南大学，2018.

[42] Chen Y, He D D. Spatial-temporal stream anomaly detection based on Bayesian fusion [J]. Journal of Electronics & Information Technology, 2019, 41(5): 1137-1144.
陈莹，何丹丹. 基于贝叶斯融合的时空流异常行为检测模型[J]. 电子与信息学报，2019，41(5)：1137-

1144.

[43] Chong Y S, Tay Y H. Abnormal event detection in videos using spatiotemporal autoencoder[EB/OL]. [2020-06-03]. https://arxiv.org/abs/1701.01546.

[44] Yue H C. Abnormal events detection method based on antoencoder[D]. Changchun: Jilin University, 2020.
岳海纯. 基于自动编码器的异常行为检测[D]. 长春: 吉林大学, 2020.

[45] Wang T, Qiao M N, Lin Z W, et al. Generative neural networks for anomaly detection in crowded scenes[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1390-1399.

[46] Ouyang J, Shi Q W, Wang X X, et al. Pedestrian trajectory prediction based on GAN and attention mechanism[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141016.
欧阳俊, 史庆伟, 王馨心, 等. 基于 GAN 和注意力机制的行人轨迹预测[J]. 激光与光电子学进展, 2020, 57(14): 141016.

[47] Wu H M. A research of unspecified anomaly detection and localization in surveillance videos[D]. Chengdu: University of Electronic Science and Technology of China, 2018.
武慧敏. 视频中的非特定异常事件时空位置检测[D]. 成都: 电子科技大学, 2018.

[48] Ravanbakhsh M, Nabi M, Sangineto E, et al. Abnormal event detection in videos using generative adversarial nets[C]// 2017 IEEE International Conference on Image Processing (ICIP), September 17-20, 2017, Beijing, China. New York: IEEE Press, 2017: 1577-1581.

[49] Ravanbakhsh M, Sangineto E, Nabi M, et al. Training adversarial discriminators for cross-channel abnormal event detection in crowds[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV), January 7-11, 2019, Waikoloa Village, HI, USA. New York: IEEE Press, 2019: 1896-1904.

[50] Akcay S, Atapour-Abarghouei A, Breckon T P. GANomaly: semi-supervised anomaly detection via adversarial training[EB/OL]. [2020-06-03]. https://arxiv.org/abs/1805.06725.

[51] Liu W, Luo W X, Lian D Z, et al. Future frame prediction for anomaly detection-a new baseline[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6536-6545.

[52] Li Y Y, Cai Y H, Liu J Q, et al. Spatio-temporal unity networking for video anomaly detection[J]. IEEE Access, 2019, 7: 172425-172432.

[53] Chen D Y, Wang P T, Yue L Y, et al. Anomaly detection in surveillance video based on bidirectional prediction[J]. Image and Vision Computing, 2020, 98: 103915.