

DOI: 10.13382/j.jemi.B2003610

# 基于一类神经网络的视频异常事件检测方法<sup>\*</sup>

蒋卫祥<sup>1</sup>  李  功<sup>2</sup>

(1. 常州信息职业技术学院  软件与大数据学院  常州  213164; 2. 南京信息工程大学  计算机与软件学院  南京  210044)

**摘  要:**视频异常事件检测一直是一个具有挑战性的问题,现有的方法往往把视频特征提取和异常检测模型建立两个步骤独立设计,导致方法无法达到最优。针对该问题,设计了一种一类神经网络方法用于视频异常检测。该方法结合了自编码器的逐层数据表示形式能力以及一类分类能力,隐藏层的特征是针对异常检测的特定任务而构建的,从而获得了一个超平面以将所有正常样本与异常样本分开。实验结果表明,提出的方法在 PED 子集和 PED2 子集上分别达到了 94.9% 的帧级 AUC 和 94.5% 的帧级 AUC,在 Subway 数据集上实现了 80 个正确事件检测,证实了该方法在工业和城市环境中的广泛适用性。

**关键词:**视频监控;异常事件;一类神经网络;深度学习;自编码网络

**中图分类号:** TP391.4; TN942      **文献标识码:** A      **国家标准学科分类代码:** 520.2040

## One-class neural network for video anomaly detection and localization

Jiang Weixiang<sup>1</sup>  Li Gong<sup>2</sup>

(1. School of Software and Big Data, Changzhou College of Information Technology, Changzhou 213164, China;  
2. School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** Due to the vague definition of abnormal events and the scarcity of its own samples, the detection of video abnormal events has always been a challenging problem. Existing methods often separate the two steps of video feature extraction and anomaly detection model establishment, it leads to the method that cannot reach the optimum. This paper follows the idea of distance-based anomaly detection, and proposes a one-class neural network method for video anomaly detection. This method combines the layer-by-layer data representation ability of the autoencoder and the one-class classification ability. The features of the hidden layer are constructed for the specific task of anomaly detection, thereby obtaining a hyperplane to separate all normal samples from abnormal samples. The experimental results on two benchmark data sets show that the proposed method achieves 94.9% frame-level AUC and 94.5% frame-level AUC on the PED subset and PED2 subset, respectively, and achieves 80 correct event detections on the Subway dataset, confirming the wide applicability of the method in industrial and urban environments.

**Keywords:** video surveillance; anomalous event; one-class neural network; deep learning; auto-encoder

## 0  引  言

随着社会公共安全需求的不断增长,监视摄像机已经作为城市基础设施建设被广泛部署,为视频分析提供数据源。然而,监视视频分析面临的一个主要挑战是检测异常事件,这需要消耗大量的人力。近年来,这种劳动强

度大的任务被设计为异常检测问题<sup>[1-3]</sup>,该问题旨在辨识意外事件或模式。与传统计算机视觉中的分类识别问题有所不同,异常检测具有如下特点:1)列出所有可能的异常样本是非常困难的;2)由于出现次数非常少,收集足够的异常样本难度也非常大。为了解决这些问题,目前最流行的方法之一是使用正常事件的视频作为训练数据来学习模型,然后将不符合所学习模型的样本判定为异常

事件。

根据对先前方法进行总结,这些方法可以分为3类。

1) 基于距离的方法<sup>[4-5]</sup>,这类方法从根据训练样本训练模型,并通过测量与该模型的偏差以确定测试样本的异常分数以判定异常。文献[4]提出首先采用降噪自编码以提取视频块的外观和运动特征,然后训练一类支持向量机(SVM)来实现异常检测。文献[5]通过合并来自预训练的卷积神经网络(CNN)提取的特征变化来获得异常得分。2) 基于概率的方法<sup>[6-7]</sup>,这种方法类似于基于距离的方法,不同之处在于检测模型具有概率的解释,例如概率图模型或者一个高维的概率分布。文献[6]提出了一种混合动态纹理的表示事件,并对混合动态纹理生成过程建模,根据判别显着性假设检验进行异常检测。文献[7]使用多个固定位置监控提取光流场,并根据存储在该监缓冲区中的分布来计算观察的可能性。3) 基于重构的方法,这种方法将输入分解为它们的共同组成部分,然后通过共同组成部分重构输入,从而最大程度地减少“重构误差”。通常,重构误差大的样本被判定为异常。常见的重构方法包括稀疏重构<sup>[8]</sup>,自编码重构<sup>[9-10]</sup>等。

近年来,在数据驱动的框架下,深度学习技术如深度神经网络(deep neural networks, DNN)、CNN、自编码网络(auto-encoder, AE)、生成对抗网络(generative adversarial nets, GAN)<sup>[11]</sup>等被应用在计算机视觉、语音识别、自然语言处理<sup>[12-13]</sup>等领域,并获取了极好的效果。实际上,深度学习技术也被应用于解决异常检测问题,如文献[4,7,10],主要是基于两种思路:1) 提供了除手工特征外更高层的特征表示,这种思路基于CNN的思想,通过在图像分类任务上<sup>[8,14]</sup>预先训练好的网络对视频帧进行高层的特征表示;2) 学习自编码器<sup>[10]</sup>或生成对抗网

络<sup>[15]</sup>来学习重建或预测正常视频帧,然后将重建误差用于判定异常。

与以上方法不同的是,本文提出了一种新的基于距离的异常检测方法。这种方法基于一类神经网络(one-class neural network, ONN),它是一类分类器在深度学习框架下的扩展。与其他基于距离的异常方法不同的是,隐藏层中的数据表示是由ONN直接驱动的,因此可以针对异常检测任务进行设计。而其他这类方法中,往往采用手工特征或者CNN提取的特征,然后再采用其他的一类分类器如一类支持向量机(one-class SVM, OC-SVM),特征提取和异常检测两个阶段是完全独立,无法联合优化。ONN结合了自编码器的逐层数据表示形式能力以及一类分类能力,从而获得了一个超平面以将所有正常样本与异常样本分开。方法通过对视频帧和光流图的大小相同的局部区域块分别训练ONN,以检测外观异常和运动异常,并将二者融合以确定最终的检测结果。在两个公开数据集上的实验结果表明,本文提出的方法检测效果达到现有技术发展水平。

## 1 算法原理

本文所提出的方法总流程如图1所示。在训练阶段,训练样本的RGB图像和光流图被密集采样,然后分别学习了两个自编码网络,并且复制预训练自动编码器的编码器层和ONN网络组合后联合优化参数,学习异常检测模型;在测试阶段,给定测试区域的RGB图和光流图,输入到外观异常检测模型和运动异常检测模型中,将输出得分进行融合设定检测门限判定该区域是否异常。

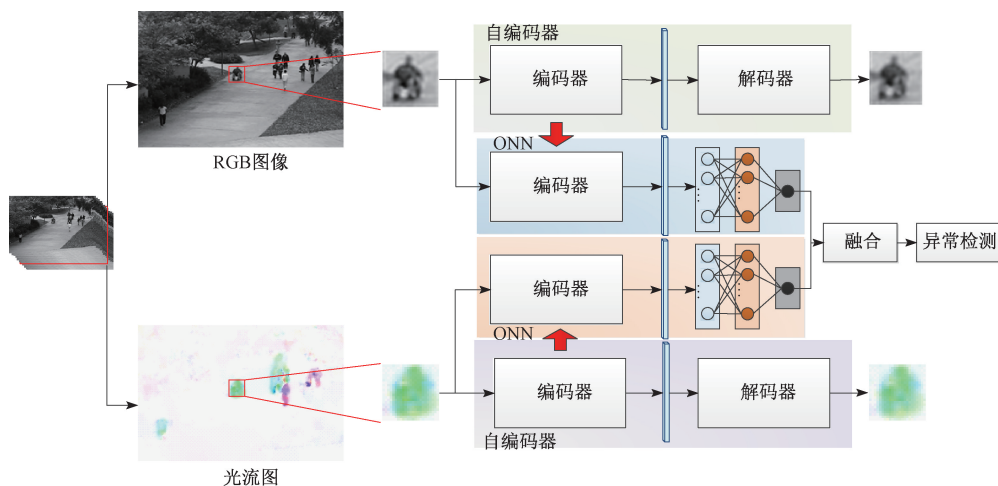


图1 基于一类神经网络的视频异常事件检测方法

Fig. 1 The flow chart of video anomaly detection based on one-class neural network

### 1.1 自编码器

自编码器<sup>[16]</sup>将输入数据映射到隐层空间得到其隐层表示(也成为特征或者编码),并且通过其隐层表示能够重构原来的输入数据。自编码器由编码器 $f_{w_1}(\cdot)$ 和解码器 $g_{w_2}(\cdot)$ 组成,可分别表示为:

$$\mathbf{z} = f_{w_1}(\mathbf{x}) \quad (1)$$

$$\mathbf{x}' = g_{w_2}(\mathbf{z}) \quad (2)$$

式中: $\mathbf{x}$ 和 $\mathbf{x}'$ 分别为自编码器的输入和重构的输入(即自编码器的输出); $\mathbf{z}$ 为 $\mathbf{x}$ 的隐层表示; $W_1$ 和 $W_2$ 为神经网络的参数,能够通过最小化输入 $\mathbf{x}$ 和重构输入 $\mathbf{x}'$ 之间的重构误差获得。

$$\min_{w_1, w_2} \|\mathbf{x} - \mathbf{x}'\|_2^2 \quad (3)$$

自编码器的隐层表示往往作为有效的特征,直接输入到后续的模式识别模型中。

### 1.2 OC-SVM

OC-SVM 是一种被广泛使用的无监督异常检测方法。其实是 SVM 的一种特殊形式,能够学习一个超平面,使得在核希尔伯特空间所有数据点和原点分开,而且使得从该超平面到原点的距离最大化。在 OC-SVM 模型中,除了原点外的所有数据都被标注为正例样本,而原点则被标注为负例样本。具体来说,给定没有任何类别信息的训练样本集  $\mathbf{X}$  和核希尔伯特空间投影函数  $\Phi(\mathbf{X})$ , 一个特征空间  $F$  中的超平面或者线性决策函数  $f(\mathbf{X}_{n_i}) = \omega^T \Phi(\mathbf{X}_{n_i}) - r$ , 用于从原点中分离尽可能多的投影向量  $\Phi(\mathbf{X}_{n_i})$ ,  $n = 1, 2, \dots, N$ 。  $\omega$  和  $r$  是超平面的法向量和截距,为了求取这两个参数,需要优化这个问题:

$$\min_{w, r} \frac{1}{2} \|\omega\|_2^2 + \frac{1}{v} \cdot \frac{1}{N} \sum_{n=1}^N \max(0, r - \langle \omega, \Phi(\mathbf{X}_{n_i}) \rangle) - r \quad (4)$$

其中  $v \in (0, 1)$  是用来权衡间隔大小和训练误差两者重要性的参数。

### 1.3 一类神经网络模型原理

在 OC-SVM 的基础上,介绍了用于无监督异常检测的 ONN 原理。该方法可以看作是使用 OC-SVM 等效损耗函数设计的神经网络结构。通过构建 ONN, 能够利用和完善从无监督迁移学习中获得,专门用于异常检测的特征。那么,这将可以鉴定那些正负决策边界是高度非线性的复杂数据集中的异常。

本文设计了一个简单的前馈卷积神经网络  $F(\cdot)$ , 它仅有一个隐藏层和一个输出节点。那么, ONN 的目标函数可以表示为:

$$\min_{w, r} \left[ \frac{1}{2} \|\omega\|_2^2 + \frac{1}{v} \cdot \frac{1}{N} \sum_{n=1}^N \max(0, r - \langle \omega, F(\mathbf{V}\mathbf{X}_{n_i}) \rangle) + \right.$$

$$\left. \frac{1}{2} \|\mathbf{V}\|_2^2 - r \right] \quad (5)$$

其中  $\omega$  表示隐藏层到输出层的标量输出。对比式 (5) 和 (4), 可以发现采用了  $\langle \omega, F(\mathbf{V}\mathbf{X}_{n_i}) \rangle$  代替了  $\langle \omega, \Phi(\mathbf{X}_{n_i}) \rangle$ 。这种变化使得可以利用自动编码器获得的转移学习功能,并创建一个额外的神经网络层以完善功能以进行异常检测。但是,更改的代价是目标函数变得不凸,因此模型参数的求取无法全局最优。

根据文献[17]可以采用交替最小化的方法对其进行优化,首先固定  $r$  优化  $\omega$  和  $\mathbf{V}$ , 需要求解

$$\arg\min_{w, V} \left[ \frac{1}{2} \|\omega\|_2^2 + \frac{1}{v} \cdot \frac{1}{N} \sum_{n=1}^N l(y_n, \hat{y}_n(\omega, \mathbf{V})) + \frac{1}{2} \|\mathbf{V}\|_2^2 \right] \quad (6)$$

其中,  $l(y, \hat{y}) = \max(0, y - \hat{y})$ ,  $y_n = r$ ,  $\hat{y}_n(\omega, \mathbf{V}) = \langle \omega, F(\mathbf{V}\mathbf{X}_{n_i}) \rangle$ , 类似地,优化  $r$  可以表示为:

$$\arg\min_{w, V} \left[ \frac{1}{Nv} \sum_{n=1}^N \max(0, r - \hat{y}_n) \right] - r \quad (7)$$

实际上,对于神经网络参数  $(\omega, \mathbf{V})$  的优化可以采用标准的反向传播算法。具体求解过程可以参考文献[17]。实际上,可以采用卷积自编码网络作为特征提取器对原始数据进行特征提取,然后输入到 ONN 中,卷积自编码和 ONN 可以联合优化。当求得  $r$  后,一旦实现收敛,可以通过决策函数判断样本是否为异常:

$$S_n = \text{sgn}(\hat{y}_n - r) \quad (8)$$

### 1.4 基于一类神经网络的视频异常事件检测方法

受双流神经网络的启发,通过对视频帧和光流图的局部区域块分别训练 ONN 模型,分别检测外观异常和运动异常,并将二者融合以确定最终的检测结果。

对于视频帧进行密集采样,训练了一个深度自动编码器以获取输入的特征,然后复制预训练自动编码器的编码器层,并将其作为输入提供给 1.3 节设计的 ONN 中,然后联合训练 ONN 的参数获得外观异常检测模型。测试时,将密集采样的视频帧输入到该模型中得到外观正常得分  $\hat{y}_{\text{appearance}}$ 。同理,对相应的光流图像训练出运动检测模型,并在测试时获得与视频帧相应区域的光流图得到运动异常得分  $\hat{y}_{\text{motion}}$ 。那么,可通过外观正常得分  $\hat{y}_{\text{appearance}}$  和运动异常得分  $\hat{y}_{\text{motion}}$  的加权和设定门限判定该区域是否为异常:

$$S = \hat{y}_{\text{appearance}} + \alpha \hat{y}_{\text{motion}} < \theta \quad (9)$$

## 2 实验

本文在两个网上公开的视频异常检测数据集上评估了该方法的性能,它们分别是 USCD 数据集<sup>[6]</sup>和 Subway 数据集<sup>[7]</sup>。USCD 数据集提供了帧级标注,因此根据所

提出的模型输出的帧级得分测量的接收器工作特性 (ROC) 曲线的曲线下面积 (area under the curve, AUC) 作为评价指标。而对于 Subway 数据集, 帧级的标注未提供, 只提供了事件级的标注, 因此在该数据集上采用事件级评价标准进行评价。

对于两个数据集, 每帧都被调整为大小  $420 \times 280$ , 运动图像由文献中提供的光流法<sup>[18]</sup> 计算得到。为了检测外观异常和运动异常, 分别采用了 RGB 图像和其对应的视频帧图, 训练了两个独立的 ONN 网络, 以检测外观和运动异常。在训练阶段, 所有的图像 (包括 RGB 图像和动态流图) 都通过大小为  $28 \times 28$  步长为 14 的滑动窗进行采样, 对于每个模型获得约 20 000 个图像块作为最终的训练集。在测试阶段, 测试样本通过相同大小的滑动窗获得, 输入到模型中计算结果, 这意味着每帧图像最终输出大小为  $15 \times 10$  得分图。自编码中编码编码器结构采用常见的 ConvTran ( $32, 3 \times 3$ )-ConvTran ( $64, 3 \times 3$ )-ConvTran ( $128, 3 \times 3$ )-FullyConnect32 的结构, 解码器则采用相反的结构。这意味着每个  $28 \times 28$  区域提取出来 64 维的特征, 并输入到 ONN 的 16 个节点的隐藏层中, 最终输出正常得分。训练阶段批量大小设置为 128, 初始学习率为 0.001, 并训练最多 5,00 次迭代。每个神经网络层采用 Xavier-Glorot<sup>[19]</sup> 进行权重初始化, 而且采用 ReLU 作为激活函数, 最后的全连接层中连接了 0.3 的 Dropout 层。整个方法在 python 和 tensorflow 环境中实现的, 硬件条件为 NVIDIA GeForce GTX 1080Ti 的 GPU, 32 GB 内存和 I7-9700 的 CPU。

2.1 USCD 数据集实验结果

USCD 异常数据集包含两个子集 Ped1 和 Ped2, 这两个子集是从可俯瞰人行道的静态摄像机获取的。异常事件为是非行人物体 (如车辆) 或者异常的行人运动。两个子集之间的主要差异是拍摄视角方向 (在 Ped1 中朝向和远离相机, 平行于 Ped2 中的相机平面), 部分样例如图 2 所示。此外, Ped1 它包含 34 个正常和 36 个异常大小为  $238 \times 158$  的视频片段, 每个视频片段包含 200 帧; 而 Ped2 它包含 12 个正常和 16 个异常大小为  $320 \times 240$  的视频片段, 每个视频片段包含 150~200 帧。



图 2 USCD 数据集部分示例

Fig. 2 Examples of normal events (first row) and abnormal events (second row) of USCD dataset

在这个数据集中, 对比方法包括 SF (social force)<sup>[20]</sup>、MDT (mixture of dynamic textures)<sup>[6]</sup>、Sparse Reconstruction<sup>[8]</sup>、Detection at 150FPS<sup>[1]</sup>、AMDN (appearance and motion DeepNet)<sup>[4]</sup>、Learning Temporal Regularity<sup>[3]</sup>、AbnormalGAN<sup>[15]</sup> 等, 实验结果如表 1 所示。其中, 前 3 种为手工特征结合异常检测模型, AMDN<sup>[4]</sup> 为深度特征结合异常检测模型, 而最后 3 种方法则为端对端的深度学习方法, 方法的结果均从各文献获得的。

表 1 UCSD 数据集帧级检测结果

Table 1 Comparison with the state of the art methods in terms of AUC% for the UCSD dataset

| 方法                                          | PED1 | PED2 |
|---------------------------------------------|------|------|
| SF <sup>[20]</sup>                          | 67.5 | 55.6 |
| MDT <sup>[6]</sup>                          | 81.8 | 82.9 |
| Sparse Reconstruction <sup>[8]</sup>        | 87.6 | 86.3 |
| Detection at 150 FPS <sup>[1]</sup>         | 91.8 | -    |
| AMDN <sup>[4]</sup>                         | 92.1 | 90.8 |
| Learning Temporal Regularity <sup>[3]</sup> | 81.0 | 90.0 |
| AbnormalGAN <sup>[15]</sup>                 | 94.3 | 93.1 |
| 本文                                          | 94.9 | 94.5 |

由表 1 可以看出, 端对端的深度学习方法检测性能最好, 深度特征+异常检测模型检测结果其次, 而手工特征+异常检测模型检测结果相对较差。主要是因为端对端的深度学习方法中特征提取和模型建立两个阶段是针对异常检测任务进行设计, 联合优化后性能达到了最优。而本文提出的方法作为一种端对端的深度学习方法, 把特征提取和一类分类的功能封装在一个网络中, 因此也取得了较好的结果。具体来说本文提出的方法在 PED1 和 PED2 上取得了 94.9% 和 94.5% 的帧级 AUC, 优于其他所有对比方法。

2.2 Subway 数据集实验结果

这个数据集也包含两个子集, Entrance 和 Exit, 分别捕获地铁站入口门和出口门的监控视频。它们的长度分别为 96 和 43 min, 每帧大小为  $512 \times 384$ 。这两个视频中的异常事件主要是错误的方向 (如乘客从入口处出来)、没有付款、游荡、异常交互 (如一个人笨拙地走路以避免另一个人) 以及其他情况 (如突然加速), 部分样例如图 3 所示。根据文献[7] 提供的标注划分训练样本和测试样本, 即 Entrance 子集的前 15 min 和 Exit 前 5 min 作为训练集, 其余部分作为测试集。值得注意的是, 两个子集上的实验是独立进行的。

由于这个数据集不提供帧级标注, 因此本文采用文献[15] 中评估方案来确定实验中的异常事件。详细来说, 在视频序列上应用持久性算法来定位局部最小值, 其中每个最小值指示异常事件。为了减少可能其他事件对于检测结果影响, 将发生时间附近的事件合并为一个异



图 3 Subway 数据集部分示例

Fig. 3 Examples of normal events (first row) and abnormal events (second row) of Subway dataset

常事件检测结果。对比方法包括 Ground truth Dynamic Sparse Coding<sup>[21]</sup>、Learning Temporal Regularity<sup>[3]</sup>、AMDN<sup>[4]</sup>。

表 2 为事件级的检测结果对比,表明模型检测到了大多数异常事件。本文提出的方法在 Entrance 子集上比现有其他方法至少多检测一个事件,而且具有更少虚警。

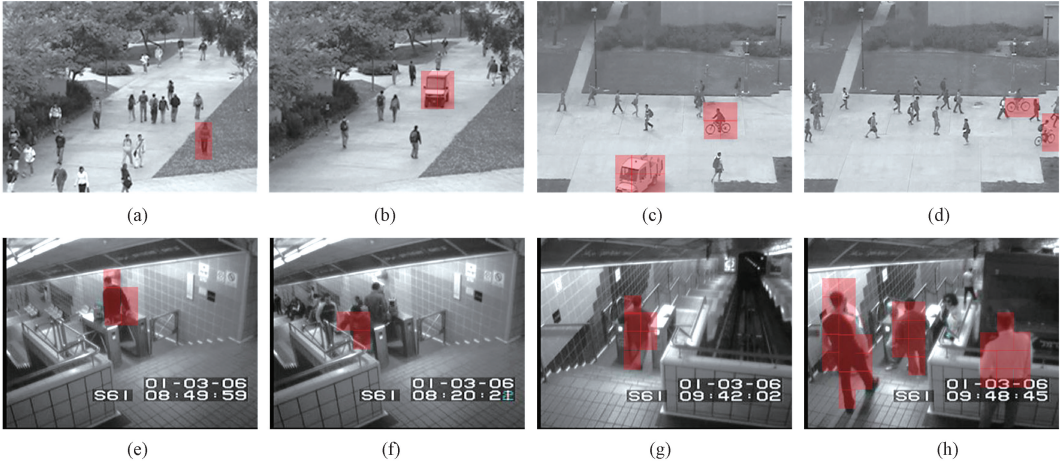


图 4 部分检测结果示例

Fig. 4 Examples of the detection results

3 结 论

本文采用基于距离的异常检测的思路,提出了一种一类神经网络方法用于视频异常检测。类似于一类 SVM,一类神经网络采用相似的损失函数训练参数,其主要优势在于隐藏层的特征是针对异常检测的特定任务而构建的。这种方法与最近提出的使用深度学习作为特征提取器再单独训练异常检测模型的混合方法大不相同,因为混合方法中的特征提取是通用的,并且不是专门面向异常检测任务的。在两个基准数据集上进行的实验结果表明,该方法同时具备准确性和鲁棒性,证实了该方法在工业和城市环境中的广泛适用性。下一步研究将考虑采用其他信息代替光流图来表示视频的运动信息。

而在 Exit 子集上,本文提出的方法和其他方法都将所有的异常事件检测出来,但是只检测出一个虚警。以上结果证明了 Perceptual GAN 方法在 Subway 数据集上的有效性。从图 4(a)~(d)可以看出,本文方法能够检测各种不同类型的异常事件,包括在人行道上骑车、人行道上出现小车和人在草坪上行走等。图 4(f)~(i)为该数据集上的一些正确检测结果。

表 2 Subway 数据集事件级检测结果

Table 2 Comparison with the state of the art methods in terms of events number for the subway dataset

| 方法                                          | Entrance (66) |    | Exit (19) |    |
|---------------------------------------------|---------------|----|-----------|----|
|                                             | TP            | FA | TP        | FA |
| Ground truth                                | 66            | —  | 19        | —  |
| Dynamic Sparse Coding <sup>[21]</sup>       | 60            | 4  | 19        | 2  |
| Learning Temporal Regularity <sup>[3]</sup> | 61            | 5  | 17        | 5  |
| AMDN <sup>[4]</sup>                         | 61            | 5  | 19        | 1  |
| 本文                                          | 61            | 4  | 19        | 1  |

参考文献

[ 1 ] LU C, SHI J, JIA J. Abnormal event detection at 150 FPS in MATLAB [ C ]. Proceedings of IEEE International Conference on Computer Vision ( ICCV ), 2013:2720-2727.

[ 2 ] SABOKROU M, FATHY M, HOSEINI M, et al. Real-time anomaly detection and localization in crowded scenes [ C ]. Proceeding of IEEE Conference Computer Vision and Pattern Recognition Workshops, 2015:56-62.

[ 3 ] HASAN M, CHOI J, NEUMANNY J, et al. Learning temporal regularity in video sequences [ C ]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

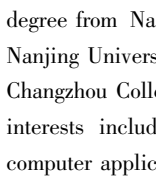
[ 4 ] XU D, YAN Y, RICCI E, et al. Detecting anomalous events in videos by learning deep representations of appearance and motion [ J ]. Computer Vision and Image

- Understanding, 2017, 156: 117-127.
- [5] SALIGRAMA V, CHEN Z. Video anomaly detection based on local statistical aggregates [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2112-2119.
- [6] MAHADEVAN V, LI W, BHALODIA V, et al. Anomaly detection in crowded scenes [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 1975-1981.
- [7] ADAM A, RIVLIN E, SHIIMSHONI I, et al. Robust real-time unusual event detection using multiple fixed location monitors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 30(3): 555-560.
- [8] LUO W, LIU W, GAO S. A revisit of sparse coding based anomaly detection in stacked RNN framework [C]. IEEE International Conference on Computer Vision, 2017: 341-349.
- [9] 戴俊, 王俊, 朱忠奎, 等. 基于生成对抗网络和自动编码器的机械系统异常检测 [J]. 仪器仪表学报, 2019, 40(9): 16-26.
- DAI J, WANG J, ZHU ZH K, et al. Anomaly detection of mechanical systems based on generative adversarial network and auto encoder [J]. Chinese Journal of Scientific Instrument, 2019, 40(9): 16-26.
- [10] YIRU Z, BING D, CHEN S, et al. Spatio-temporal autoencoder for video anomaly detection [C]. Proceedings of the 25th ACM International Conference on Multimedia, 2017: 1933-1941.
- [11] 陈亮, 吴攀, 刘韵婷, 等. 生成对抗网络 GAN 的发展与最新应用 [J]. 电子测量与仪器学报, 2020, 34(6): 70-78.
- CHEN L, WU P, LIU Y T, et al. Development and application of the latest generation against the network of GAN. Journal of Electronic Measurement and Instrumentation, 2020, 34(6): 70-78.
- [12] 蓝金辉, 王迪, 申小盼. 卷积神经网络在视觉图像检测的研究进展 [J]. 仪器仪表学报, 2020, 41(4): 167-182.
- LAN J H, WANG D, SHEN X P. Research progress on visual image detection based on convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2020, 41(4): 167-182.
- [13] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521: 436-444.
- [14] RAVANBAKSH M, NABI M, MOUSAVI H, et al. Plug-and-Play CNN for crowd motion analysis: An application in abnormal event detection [C]. IEEE Winter Conference on Applications of Computer Vision, 2018: 1689-1698.
- [15] RAVANBAKSH M, NABI M, MOUSAVI H, et al. Abnormal event detection in videos using generative adversarial nets [C]. IEEE International Conference on Image Processing, 2017.
- [16] SUN C, MA M, ZHAO Z, et al. Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing [J]. IEEE Transactions on Industrial Informatics, 2019, 15(4): 2416-2425.
- [17] CHALATHY R, MENON A K, CHAWLA S. Anomaly detection using one-class neural networks [C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2017: 36-57.
- [18] BAKER S, SCHARSTEIN D, LEWIS J, et al. A database and evaluation methodology for optical flow [J]. International Journal of Computer Vision, 2011, 92: 1-31.
- [19] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [C]. Proceedings of Machine Learning Research, 2010: 249-256.
- [20] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1-8.
- [21] ZHAO B, LI F, XING E. Online detection of unusual events in videos via dynamic sparse coding [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2011: 1-8.

## 作者简介



**蒋卫祥**, 2003 年于南京大学获得学士学位, 2004 年于南京大学获得硕士学位, 现为常州信息职业技术学院副教授, 主要研究方向为软件工程、图像处理、计算机应用技术。  
E-mail: gssds\_123@21cn.com



**Jiang Weihang** received his B. Sc. degree from Nanjing University in 2003, M. Sc. degree from Nanjing University in 2004. Now he is an associate professor at Changzhou College of Information Technology. His main research interests include software engineering, image processing and computer application technology.



**李功**, 2001 年于南京大学获得硕士学位, 2004 年于南京大学获得博士学位, 现为南京信息工程大学副教授, 主要研究方向为计算机应用技术。  
E-mail: ligong\_nj@163.com

**Li Gong** received his M. Sc. degree from Nanjing University in 2001, Ph. D. degree from Nanjing University in 2004. Now he is an associate professor at Nanjing University of Information Science & Technology. His main research interest includes computer application technology.