Research  
Artificial Intelligence—Article

## AED-Net——异常事件检测网络

王田<sup>a</sup>, 苗子琛<sup>a</sup>, 陈禹昕<sup>a</sup>, 周毅<sup>b</sup>, 单光存<sup>a,c,\*</sup>, Hichem Snoussi<sup>d</sup><sup>a</sup> School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China<sup>b</sup> Department of Electronic Engineering, Dalian Maritime University, Dalian 116026, China<sup>c</sup> School of Instrumentation Science and Opto-electronic Engineering & International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China<sup>d</sup> Institute Charles Delaunay-LM2S-UMR STMR 6281 CNRS, University of Technology of Troyes, Troyes 10010, France

## ARTICLE INFO

## Article history:

Received 22 May 2018

Revised 1 February 2019

Accepted 25 February 2019

Available online 25 May 2019

## 关键词

异常事件检测  
异常事件检测网络  
主成分分析网络  
核主成分分析

## 摘要

长期以来,在拥挤场景中检测异常事件都是一项具有挑战性的任务。为解决这一问题,本文提出了一种名叫异常事件检测网络(AED-Net)的自监督框架,它由主成分分析网络(PCANet)和核主成分分析(kPCA)组成。该框架以不同场景的监控视频序列为原始数据,通过训练PCANet以提取人群情况的高级语义。kPCA可作为一种单类分类器,被用于识别场景中的异常事件。与目前流行的一些深度学习方法相比,该框架完全是自监督的,因为它只使用正常情况下的视频序列。通过对明尼苏达大学公共监测人类活动数据集(UMN数据集)和加州大学圣地亚哥分校监测异常数据集(UCSD数据集)进行全局和局部异常事件进行检测发现,与其他最先进的方法相比,该方法具有更好的等误差率(EER)和曲线下面积(AUC)。此外,通过增加局部响应归一化(LRN)层,我们对原有的AED-Net进行了改进。结果表明,该改进版在提高框架的泛化能力方面表现出更好的性能。

© 2019 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. 引言

近年来,视频研究引起了计算机视觉界研究者们越来越多的关注。最近,有关目标跟踪[1–3]、步态识别[4,5]和活动识别[6–8]等主题的研究取得了具有竞争力的成果,并在未来具有一定的发展前景。

异常事件检测是视频研究领域最热门的问题之一,所涉及的检测包含异常视频中的特定帧。与上述任务相比,异常事件检测对国家安全和人民生活有着更大的意义,因为随着现代社会的不断发展,越来越多的监控摄像头被安装在不同的地方,每一秒就会产生大量的视

频。对于这么多的数据,我们人类是很难进行处理的,也难以确定包含在其中的任何异常事件。然而,错过监控视频中任何一个异常点,就可能会给社会造成无法承担的损失。因此,构建一个视频异常的自动检测器是十分必要的,它可以处理数百万的视频帧,并提醒人们在异常情况发生时做出及时有效的反应。

参考文献[9]描述了异常事件检测中存在的困难。列举出某一特定场景中的几种异常情况是很简单的,如行人群中有人开汽车或骑自行车的人,但列举出该场景中可能发生的所有异常事件是不切实际的;因此,在这个分类任务中有无数的正类。此外,由于缺乏异常样

\* Corresponding author.

E-mail address: [gcshan@buaa.edu.cn](mailto:gcshan@buaa.edu.cn) (G. Shan).

本,即包括异常事件在内的视频帧,训练集严重不平衡,因此多类分类模型的训练是不可行的。这些困难说明,异常事件检测任务是个难以处理的单类分类问题。

一些处理异常事件检测的方法也曾被提出。例如,参考文献[10]提出了一种基于光流方向直方图描述符的方法。由于在该研究中手工构造的特征描述符是基于人类经验构建的,它并不能完全反映训练过程中的特征。因此,它的表现比现有的深度学习方法要差一些。如参考文献[11,12]所述,深度学习方法是最近发展起来的,主要得益于大数据的可用性和高效的硬件。这些方法在计算机视觉领域得到了广泛的应用,并取得了良好的效果。在参考文献[13]中,Wang等将卷积神经网络(CNN)应用于产品质量控制中的缺陷检测。然而,最初的CNN,如人脸识别,并不适用于这一任务,因为它的训练需要不同类别的样本。考虑到主成分分析网络(PCANet)[14]在图像分类中的成功应用,参考文献[15]提出了一种基于PCANet从原始图像中提取信息进行异常事件检测的方法,该方法用一种基于聚类算法构造的单类分类器进行检测。然而,由于K-medoids聚类算法难以处理PCANet提取的高维特征,所以该方法具有很大的局限性。在本文中,我们提出了一种被称为异常事件检测网络(AED-Net)的自监督网络,用于处理视频异常检测任务,且仅提供正常样本作为训练数据。由于PCANet已经被证实能够作为一个无监督模型来提取特征,所以我们选择它作为自监督的AED-Net。此外,我们还使用了单类分类器对提取的高维特征进行处理,以确定视频帧的异常。

更具体地说,新的自监督网络使用光流图作为输入,因为这些映射非常适合表示该运动。接下来,人群状况的高级语义可以从PCANet中提取出来。随后,采用简单而有效的单分类器核主成分分析(kPCA)[16]对高维特征进行分类。AED-Net具有这两种网络的优点,经过训练可以读懂每一个视频帧并对其进行检测。更重要的是,为了改进AED-Net,我们引入了局部响应归一化(LRN)层(CNN中用于辅助泛化的技术)。值得注意的是,这种新的网络可以用未标记的数据进行训练,并且在异常检测任务中,该方法与最先进的办法相比能表现出更好的性能。新的自监督网络即使在拥挤的情况下也能有效地检测异常事件,根据对明尼苏达大学公共监测人类活动数据集(UMN数据集)和加州大学圣地亚哥分校异常检测数据集(UCSD数据集)

进行检测发现,自监督网络可以改进检测结果。

本文其余部分的结构如下。第2节简要回顾了相关工作,第3节回顾了PCANet和kPCA框架的基本算法,第4节详细介绍了AED-Net的完整体系结构以及对它的改进,第5节阐述和讨论了UMN[17]数据集和UCSD[18]数据集的实验结果,第6节给出结论。

## 2. 相关工作

一般地,用于异常事件检测的传统方法主要有两类。第一类是基于轨迹的方法,该方法在异常事件检测中已经得到了广泛的应用[19–22]。在参考文献[23–25]中,作者提取了正常事件的轨迹用于表示正常模式;与正常模式不同的轨迹则被认为是异常模式。然而,当把该方法应用于拥挤场景时,运动目标间的遮挡影响了该方法的有效性。为了解决这个问题,参考文献[26]提出了一种新的模型,用于处理人类行为的相互关联性以及改进对象间相互作用的表示方法。参考文献[27]利用离散变换原理,开发了一种可靠的多目标跟踪算法,该算法可将不同视频帧中的目标相互关联起来。然而,遮挡问题对检测结果的影响仍然很大,所以上面列出的方法不能有效地解决这个问题。因此,在我们的工作中没有采用跟踪策略。

时空方法属于另一类方法。研究人员对该方法的应用前景做过一些研究。在参考文献[28]中,Wang等提议以协方差矩阵作为特征描述符,对相邻帧的光流和偏导数进行编码。在参考文献[29–32]中,作者用像素变化直方图对运动模式建模。在参考文献[33–35]中,作者将光流分布用作基本特征,然后基于光流特征构建了异常事件检测模型。参考文献[36]提出了一种测量移动物体之间相互作用的方法。另一项研究[9]则采用了一种时空一体化数据异常检测器。也有研究人员利用图像处理中的小波变换来分析运动[37,38]。在这些研究中,精细的特征描述符是手动设计的,并且倾向于只在特定条件下才能很好地工作。然而,在我们的研究中,特征是由一个自监督网络所提取的。

随着自监督网络的迅速发展,深度学习在异常事件检测领域取得了突出的成果。与手动设计的特征不同,深度学习网络提取的特征是通过学习过程获得的。我们在AED-Net中提出了一种基于自监督学习的异常事件检测方法,该方法只需要学习正常样本即可。

### 3. 自监督特征提取与异常检测

自监督学习是一种学习范式，在这种范式中，没有外部监督信息，即标签，作为数据本身之外的真值。在此范式下，自监督学习方法仅以原始数据作为训练材料，这意味着该模型将学习从数据中提取潜在的信息。该模型在训练过程中不使用数据的类别信息。

自监督学习模型适用于异常事件检测的任务。由于没有给出外部监督信息，我们只能使用正常数据来训练模型。因此，模型必须充分理解输入视频片段中的正常数据，然后利用正常数据作为监督信息来调整其参数。表1对本文所使用的符号进行了介绍。

#### 3.1. 用于特征提取的 PCANet

传统方法和深度学习方法都已被应用于视频帧的特征提取。参考文献[10]使用全局光流描述符作为特征。然而，光流只包含了视频帧中的低级运动信息；而高级信息特征，如人的运行模式或者在视频帧中有多少人，则是无法描述的。因此，本文采用了深度学

习方法来解决这一高级特征提取问题。最受欢迎的模型是CNN，它通过层层叠加，一步一步地提取到更深的特征。然而，这个特定的模型需要强大的外部监督信息，而异常检测问题并没有提供这种信息。因此，我们选择了PCANet [14]，这是一种在不需要外部监督信息的情况下，利用深度学习能力进行特征提取的等价模型。

PCANet [14]是在当前深度学习的趋势下所提出的一种深度学习网络。尽管与其他流行的深度学习网络（如深层CNN）相比，PCANet虽简单，但它足以处理人脸识别等具有挑战性的任务。因此，该模型因其在特征提取方面具有较高的效率和较强的竞争力而被选取。

PCANet是一个级联式线性网络。图1展示了一个典型的两阶段PCANet结构。它是由CNN启发而来的，PCANet的每个阶段都由一个独立的主成分分析（PCA）滤波器组组成，而该滤波器是需要我们为了进行特征提取工作学习的。将第一阶段的特征映射线性级联到下一阶段，可以提取更高层次的特征。正如Chan等[14]所讨论的，与阶段数相对应的性能表明，虽然二级网

表1 本文所用符号的描述

Variable	Description
$S$	Raw surveillance video frame
$I$	Input of PCANet (optical flow map)
$k_1, k_2$	Size of patches in PCANet
$X$	Matrix consisting of all patches from an optical flow map at Stage 1 of PCANet
$S_i$	Matrix containing all matrices $\bar{X}$ at the $i$ th stage of PCANet
$K_j^i$	$j$ th filter of $i$ th stage of PCANet
$C$	Outputs of Stage 1 of PCANet
$Y$	Matrix consisting of all patches from an optical flow map at Stage 1 of PCANet
$O$	Outputs of Stage 2 of PCANet
$T$	Integer-valued image after binarizing and encoding outputs of Stage 2
$\mathcal{F}$	Final feature of PCANet
$F$	Inputs feature of kPCA classifier
$\mathcal{M}(F_i)$	Feature $F$ mapped into higher-dimensional space
$\kappa(F_i, F_j)$	Scalar product of $\mathcal{M}(F_i)$ and $\mathcal{M}(F_j)$
$W_j$	Eigenvectors of covariance matrix of mapped feature in higher dimensional space
$V$	Kernel matrix, $\bar{V}_{i,j} = \kappa(F_i, F_j)$
$\bar{V}$	Kernel matrix, $\bar{V}_{i,j}$ , which is the scalar product of $\bar{M}(I_i)$ and $\bar{M}(I_j)$
$R$	Reconstruction error, i.e., abnormality score
$a$	Eigenvectors of kernel matrix $\bar{V}$
$p_i$	Output value on the $i$ th feature map
$q_i$	Normalized output of $p_i$
$\delta, n, \beta, \gamma$	Hyperparameters of LRN

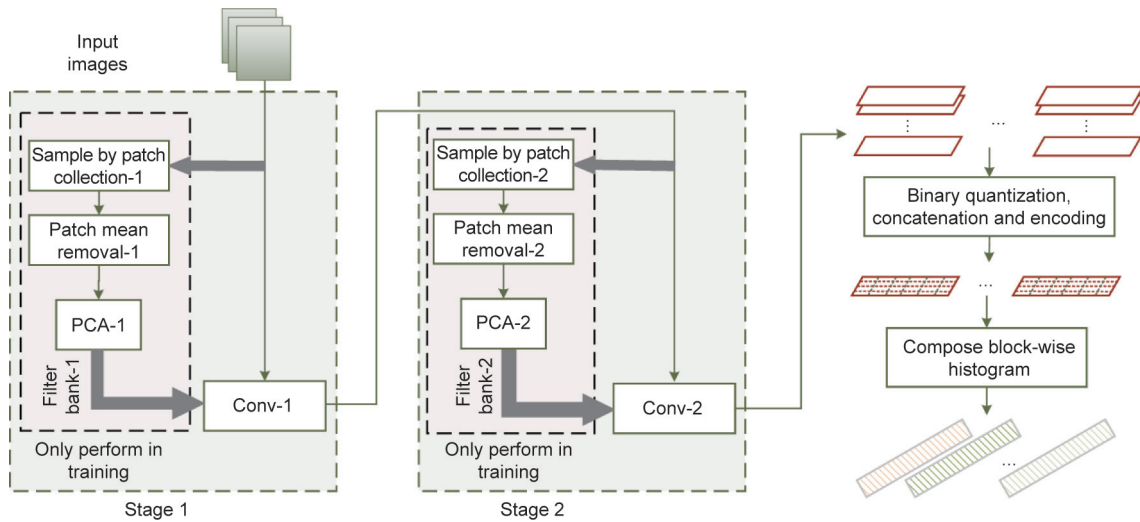


图1. 我们的方法中使用的典型的两阶段PCANet结构。Conv——卷积。

络的性能优于一级网络，但两阶段以上的网络与二级网络相比没有什么优势，因此，为了提高计算效率，两阶段PCANet对于我们的任务来说已经足够。

因此，我们使用两阶段PCANet来提取特征。在训练阶段，在第1阶段开始时，在形状为 $h \times w$ 的光流图 $I_i$ 上，围绕每个像素进行采样，采样大小为 $k_1 \times k_2$ 的小块，如图1上部灰色箭头所示。接下来，样本 $\text{patch}(x_1)$ ,  $\text{patch}(x_2)$ , ...,  $\text{patch}[x_{(h-k_1+1) \times (w-k_2+1)}]$ 被矢量化并组成样本矩阵 $X = [x_1, x_2, \dots, x_{(h-k_1+1) \times (w-k_2+1)}]$ 。然后对 $X$ 进行平均减法，得到 $\bar{X}$ 。（本文使用的所有符号见表1。）

对于输入的 $N$ 个光流图， $I = \{I_1, I_2, \dots, I_N\}$ ，PCANet首先对它们进行采样，得到以下结果：

$$S_1 = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N] \in \mathbb{R}^{k_1 k_2 \times N(h-k_1+1) \times (w-k_2+1)} \quad (1)$$

接下来，如图1下部灰色箭头所示，PCANet通过实施PCA计算基于 $I$ 的 $L_1$ 卷积核，以获得以下结果：

$$K_l^1 = \text{vec2mat}_{k_1, k_2} [s_l (S_1 S_1^T)] \in \mathbb{R}^{k_1 \times k_2}, l = 1, 2, \dots, L_1 \quad (2)$$

式中， $s_l (S_1 S_1^T)$ 表示 $S_1 S_1^T$ 的第 $l$ 个主特征向量； $\text{vec2mat}(\cdot)$ 表示将一个向量从 $\mathbb{R}^{k_1 \times k_2}$ 映射到矩阵 $M \in \mathbb{R}^{k_1 \times k_2}$ 。在第1阶段结束时，进行卷积运算以提取特征：

$$C_i^1 = I_i * K_l^1, i = 1, 2, \dots, N \quad (3)$$

式中， $*$ 表示二维（2D）卷积； $C_i^1$ 指第 $i$ 个输入 $I_i$ 的第 $l$ 个特征映射；第1阶段的输出数是 $L_1 N$ 。注意， $I_i$ 的边界为零填充，以确保输出具有与输入相同的尺寸，即 $h \times w$ 。如前所述，在测试阶段，PCANet将使用从训练阶段获得的卷积核直接对输入 $I$ 执行卷积运算。

第2阶段采用与第1阶段几乎相同的方式进行运算。

在训练阶段， $C$ 的每个输入 $C_i^l$ 都被采样成小块。在执行平均减法后，将这些小块矢量化并组成矩阵 $S_2$ ：

$$S_2 = [\bar{Y}_1^1, \dots, \bar{Y}_1^{L_1}, \bar{Y}_2^1, \dots, \bar{Y}_2^{L_1}, \dots, \bar{Y}_N^1, \dots, \bar{Y}_N^{L_1}] \in \mathbb{R}^{k_1 k_2 \times L_1 N(h-k_1+1) \times (w-k_2+1)} \quad (4)$$

式中， $\bar{Y}_j^l = [\bar{y}_{i,l,1}, \bar{y}_{i,l,2}, \dots, \bar{y}_{i,l,(h-k_1+1) \times (w-k_2+1)}] \in \mathbb{R}^{k_1 k_2 \times (h-k_1+1) \times (w-k_2+1)}$ 是 $C_i^l$ 的样本矩阵。然后，我们计算第2阶段的卷积核：

$$K_m^2 = \text{vec2mat}_{k_1, k_2} [s_m (S_2 S_2^T)] \in \mathbb{R}^{k_1 \times k_2}, m = 1, 2, \dots, L_2 \quad (5)$$

最后，通过卷积得到第2阶段的输出：

$$O_i^{l,m} = C_i^l * K_m^2, l = 1, 2, \dots, L_1, m = 1, 2, \dots, L_2, i = 1, 2, \dots, N \quad (6)$$

第2阶段的输出数为 $L_2 L_1 N$ 。

在第2阶段之后，我们将Heaviside阶跃函数 $H(\cdot)$ 的输出进行二值化，为正项分配1，为零或负项分配0。这使得网络具有非线性特征。因此，该网络能够在光流图中捕获高级语义。第2阶段 $C_i^l$ 的每一个 $L_1$ 输入在第2阶段 $O_i^{l,m}$  ( $m = 1, 2, \dots, L_2$ )中都具有 $L_2$ 实值输出。在每个像素周围，都有 $L_2$ 二进制位；我们可以将它们当作十进制数，把 $L_2$ 输出 $C_i^{l,m}$ 转换成一个整数值图像：

$$T_i^l = \sum_{m=1}^{L_2} 2^{m-1} H(O_i^{l,m}) \quad (7)$$

最后，PCANet的输出特征是根据所有 $T_i^l$ 计算的分块直方图（含有 $2^{L_2}$ 二进制）。请注意，一个直方图并不

代表整个 $T_i^l$ ，而只是它的一个区域。为了做到这一点， $T_i^l$ 被划分为 $B$ 块，然后用于计算直方图。每个分块计算一个直方图。接下来，所有的直方图被串级成一个向量 $\text{Bhist}(T_i^l)$ 。对于单输入光学图像 $I_i$ ，其特征如下：

$$\mathcal{F}_i = [\text{Bhist}(T_i^1), \dots, \text{Bhist}(T_i^{L_1})]^T \in \mathbb{R}^{(2^{L_2})L_1 B} \quad (8)$$

局部分块可以是重叠的，也可以是不重叠的。后者除了人脸检测[14]外，对其他检测都有利，因此本文将其设置为不重叠。除了选择重叠设置外，PCAnet的超参数还包括滤波器大小 $k_1$ 、 $k_2$ ，每个阶段 $L_1 L_2$ 中的滤波器数目，以及局部直方图的分块大小。

### 3.2. 一种自监督学习异常检测方法——kPCA

由于只能利用正常场景的视频序列，并且有必要区分正常帧和先前未知异常的异常帧，所以我们将此任务划分为单类分类任务是合适的。

单类分类任务的共同思想是训练一个包含训练数据（即正常数据）的分类器，从而将异常数据与正常数据分开。支撑向量域描述（SVDD）分类器就是一个很好的例子。但是，这种分类器往往会产生过大的决策边界，从而阻碍了分类器良好性能的发挥。Kemmler等[39]通过使用高斯过程先验建立了一个单类分类模型，该模型采用了从高斯过程回归和近似高斯分类得出的不同度量方法。然而，该模型在很大程度上依赖于重新参数化的核函数的超参数调整。

相反，通过学习数据的分布（通常是非线性的），kPCA分类器[16]可以根据数据的分布平滑地生成决策边界，从而使分类更加准确。

kPCA分类器的结构如图2所示。这种单类分类器的基本思想是正常帧的特征具有相似分布，而异常帧的特征分布有很大的不同。因此，为了对正常特征和异常

特征执行PCA，我们使用了基于训练特征（即正常特征）计算的PCA滤波器，并在使用后观察到了正常特征与异常特征间重构误差的显著差异。然后，我们根据这一差异进行分类。

正如Hoffmann [16]所讨论的那样，PCA不能捕获非线性结构。因此，kPCA的引入就是为了克服这个缺点，因为它可以将输入 $F_i \in \mathbb{R}^d$ 映射成高维空间中的特征： $\mathcal{M}(F_i) \in \mathbb{R}^n$  ( $n > d$ )。然后在特征空间中执行PCA。这里只需要计算 $\mathcal{M}(F_i)$ 的标量积，也就是 $[\mathcal{M}(F_i) \cdot \mathcal{M}(F_b)]$ 。标量积被核函数 $\kappa(F_i, F_j)$ 进一步替换，以执行相同的任务。在这里，核函数使用高斯核 $\kappa(F_i, F_j) = \exp\left(-\frac{\|F_i - F_j\|^2}{2\sigma^2}\right)$ 。此外，我们通过执行平均减法从 $\mathcal{M}(F_i)$ 中求得 $\bar{\mathcal{M}}(F_i)$ ，它可以进一步表示协方差矩阵在高维空间中的特征向量 $W_j$ 。因此， $W_j$ 可以按以下公式用 $\mathcal{M}(F_i)$ 表达出来：

$$W_j = \sum_{i=1}^N \alpha_i^j \left[ \mathcal{M}(F_i) - \frac{1}{N} \sum_{k=1}^N \mathcal{M}(F_k) \right] \quad (9)$$

结果显示，来自 $\alpha^j = [\alpha_1^j, \alpha_2^j, \dots, \alpha_N^j]$ ， $j = 1, 2, \dots, q$ 的 $\alpha^j$ 是核矩阵 $\bar{V}$ 的一个特征值。 $\bar{V}$ 的每一个组分 $\bar{V}_{ij}$ 都是 $\bar{\mathcal{M}}(F_i)$ 和 $\bar{\mathcal{M}}(F_j)$ 的标量积。同样地，核矩阵 $V$ 的每个组分 $V_{ij}$ 是 $\mathcal{M}(F_i)$ 和 $\mathcal{M}(F_j)$ 的标量积。因此，

$$\bar{V}_{ij} = V_{ij} - \frac{1}{N} \sum_{a=1}^N V_{ia} - \frac{1}{N} \sum_{a=1}^N V_{aj} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab} \quad (10)$$

根据Hoffmann [16]，给定特征 $F_z$ ，特征空间中的重构误差计算如下：

$$R(I_z) = [\bar{\mathcal{M}}(F_z) \cdot \bar{\mathcal{M}}(F_z)] - \left\{ [W \bar{\mathcal{M}}(F_z)] \cdot [W \bar{\mathcal{M}}(F_z)] \right\} \quad (11)$$

式中， $W = (W_1, W_2, \dots, W_q)$ 。上述公式可以更清楚地表述为：

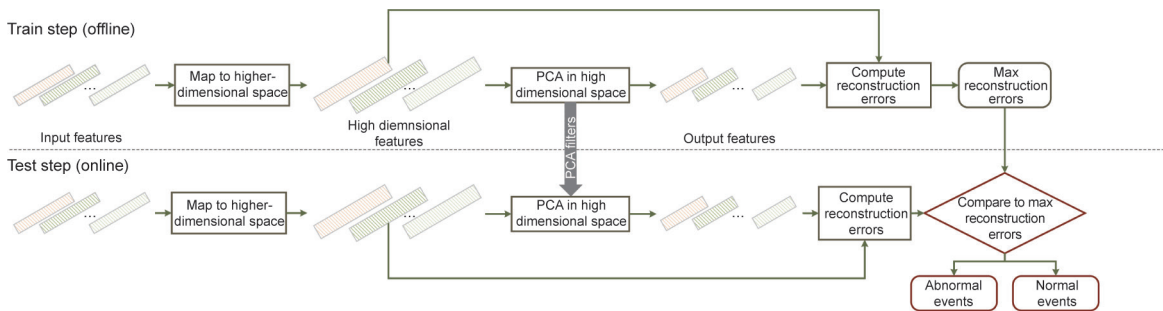


图2. 单类分类器的结构——kPCA。

$$\begin{aligned}
R(F_z) &= \|\overline{\mathcal{M}}(F_z)\|^2 - \sum_{j=1}^q [\overline{\mathcal{M}}(F_z) \cdot W_j]^2 \\
&= V_{zz} - \frac{2}{N} \sum_{a=1}^N V_{za} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab} - \sum_{j=1}^q [P_j(F_z)]^2 \quad (12) \\
&= 1 - \frac{2}{N} \sum_{a=1}^N V_{za} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab} - \sum_{j=1}^q [P_j(F_z)]^2
\end{aligned}$$

在上述方程中,  $P_j(F_z)$ 表示如下:

$$\begin{aligned}
P_j(F_z) &= \overline{\mathcal{M}}(F_z) \cdot W_j \\
&= \left[ \overline{\mathcal{M}}(F_z) - \frac{1}{N} \sum_{a=1}^N \mathcal{M}(F_a) \right] \cdot \\
&\quad \left[ \sum_{i=1}^N \alpha_i^j \mathcal{M}(F_i) - \frac{1}{N} \sum_{i,b=1}^N \alpha_i^j \mathcal{M}(F_b) \right] \quad (13) \\
&= \sum_{i=1}^N \alpha_i^j \left[ V_{zi} - \frac{1}{N} \sum_{a=1}^N V_{ia} - \frac{1}{N} \sum_{b=1}^N V_{zb} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab} \right]
\end{aligned}$$

因此, 我们得到了期望的测量形式 $R(I_z)$ 用于检测异常。

该分类器的超参数是特征向量 $q$ 的个数和核宽度 $\sigma$ 。它们的值取决于特定的实验环境。

最后, 在给定输入 $X$ 和提取特征 $F_x$ 的情况下, 我们将分类器定义如下:

$$\text{status}(X) = \begin{cases} \text{anomaly } R(F_x) > \text{threshold} \\ \text{normality } R(F_x) \leq \text{threshold} \end{cases} \quad (14)$$

上述阈值是在训练阶段计算的最大重构误差, 如图2所示。

## 4. 提出的 AED-Net

针对视频帧异常事件检测任务, 我们提出了AED-Net, 它是一种完整的自监督检测框架, 是基于对正常数据训练的自监督学习方法。为了完成基于输入视频帧的特征提取任务, 我们采用了有效的PCANet网络。对于单类分类, 我们使用了kPCA这一特殊的单类分类器来确定视频帧的异常。

### 4.1. 光流计算

最初我们获得了原始视频帧 $S$ 。为了检测这些视频帧中的异常事件, 我们首先从 $S$ 中的静态背景中分离出运动区域, 以简化检测任务。光流代表了视频帧间[40]的运动场, 适用于这一运动提取要求。

可以用Horn-Schunck(H-S)方法[41]计算光流。该方法在考虑了像素值一致性和整个图像流量变化的条件下, 构造了能量函数, 并对其进行了优化, 获得了 $u$ 和 $v$ [41]形式的光流, 它们分别是光流的水平分量和垂直分量。我们在能量函数中加入平滑约束, 以缓解孔径问题。构造的能量函数如下:

$$E = \iint [(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2)] dx dy \quad (15)$$

式中,  $E$ 是全局能量;  $I_x$ ,  $I_y$ 和 $I_t$ 是沿整个宽度方向、高度方向和时间方向上的像素值;  $\alpha$ 是控制平滑项的超参数。

然后, 为了将光流特征作为图像处理, 我们利用孟塞尔颜色系统(Munsell Color System)对光流 $u$ ,  $v$ 进行

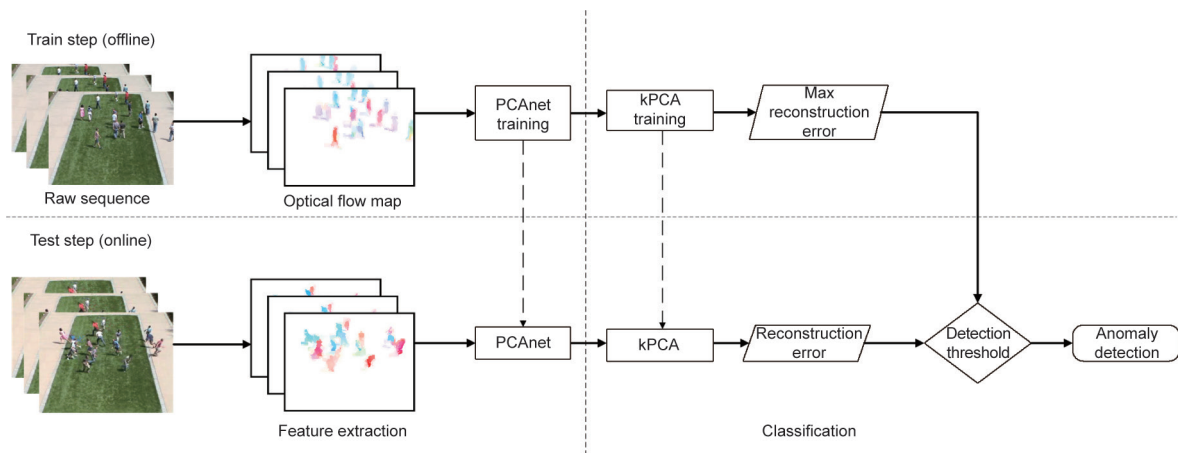


图3. AED-Net的框架结构。

了可视化处理，并得到光流图 $I$ 。

#### 4.2. AED-Net

直观地说，我们提出的AED-Net异常检测任务是为每个视频帧分配一个表示异常的分值。在训练阶段，最大重构误差应被设置为异常检测阈值。因此，在测试阶段，我们通过比较测试视频帧的分值和阈值确定了测试视频帧的异常。为了完成这项任务，我们将PCANet和kPCA结合起来构建了AED-Net。

我们提出的AED-Net框架如图3所示，以及AED-Net算法如算法1所示。首先，将光流图 $I$ 作为整个训练和测试框架的输入。接下来，对PCANet模型进行训练，以学习从时空特征中提取能更好地表示场景情况的高级信息。最后，利用分块直方图作为PCANet提取的分类特征，对kPCA进行训练，以学习正常场景的非线性数据分布以及通过重构误差确定最大正态得分作为阈值计算。

---

#### Algorithm 1. AED-Net

---

##### Input:

Optical flow maps  $I = \{I_1, I_2, \dots, I_N\}$

##### Output:

Threshold of max normality score threshold,  $K^1, K^2, \alpha$

```

1: for  $i = 1, 2, \dots, N$  do
2: Sample  $I_i$  by patches and get  $X_i$  by vectorizing and
   concatenating all patches
3: end for
4:  $\bar{X}_i = X_i - \frac{1}{N}(\sum X_i)$ 
5:  $S_1 = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N]$ 
6:  $K^1 = \underset{K^1}{\operatorname{argmin}} \|S_1 - K^1(K^1)^T S_1\| \text{ s.t. } K^1(K^1)^T = I_{L_1}$ 
7:  $C = I * K^1, C = \{C_i^l\}, i = 1, 2, \dots, N, l = 1, 2, \dots, L_1$ 
8: for  $i = 1, 2, \dots, N, j = 1, 2, \dots, L_1$  do
9: Sample  $C_i^j$  by patches and get  $Y_i^j$  by vectorizing and
   concatenating all patches
10: end for
11:  $\bar{Y}_i^j = Y_i^j - \operatorname{mean}(Y_i^j)$ 
12:  $S_2 = [\bar{Y}_1^1, \dots, \bar{Y}_1^{L_1}, \bar{Y}_2^1, \dots, \bar{Y}_2^{L_1}, \dots, \bar{Y}_N^1, \dots, \bar{Y}_N^{L_1}]$ 
13:  $K^2 = \underset{K^2}{\operatorname{argmin}} \|S_2 - K^2(K^2)^T S_2\| \text{ s.t. } K^2(K^2)^T = I_{L_2}$ 
14:  $O_i^{l,m} = C_i^l * K_m^2,$ 
    $l = 1, 2, \dots, L_1, m = 1, 2, \dots, L_2, i = 1, 2, \dots, N$ 
15:  $T_i^l = \sum_{m=1}^{L_2} 2^{m-1} H(O_i^{l,m})$ 
16:  $\mathcal{F}_i = [\operatorname{Bhist}(T_i^1), \dots, \operatorname{Bhist}(T_i^{L_1})]^T$ 

```

---

```

17: for  $i = 1, 2, \dots, N$  do
18:   for  $j = 1, 2, \dots, N$  do
19:      $V_{ij} = \kappa(\mathcal{F}_i, \mathcal{F}_j)$ 
20:   end for
21: end for
22:  $\bar{V}_{ij} = V_{ij} - \frac{1}{N} \sum_{a=1}^N V_{ia} - \frac{1}{N} \sum_{a=1}^N V_{aj} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab}$ 
23:  $\alpha = \underset{\alpha}{\operatorname{argmin}} \|\bar{V} - \alpha \alpha^T \bar{V}\| \text{ s.t. } \alpha \alpha^T = I_q$ 
24: Initialize threshold = 0
25: for  $i = 1, 2, \dots, N$  do
26:    $R(\mathcal{F}_i) = 1 - \frac{2}{N} \sum_{a=1}^N V_{ia} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab} - \sum_{j=1}^q [P_j(\mathcal{F}_i)]^2$ 
   and
27:    $P_j(\mathcal{F}_i) = \sum_{k=1}^N \alpha_k^j [V_{ik} - \frac{1}{N} \sum_{a=1}^N V_{ka} - \frac{1}{N} \sum_{b=1}^N V_{ib} + \frac{1}{N^2} \sum_{a,b=1}^N V_{ab}]$ 
28:   if  $R(\mathcal{F}_i) > \text{threshold}$  do
29:     threshold =  $R(\mathcal{F}_i)$ 
30:   end if
31: end for

```

---

在测试期间，为了使携带少量相关信息的视频帧的影响最小化，我们首先在测试视频片段中进行了前景检测，并删除了一些包含少量人员的视频帧。然后，利用事先训练的PCANet提取了 $k$ 分块特征，并通过kPCA计算出每个视频帧的测试得分。最后，将测试得分与最大正态得分进行比较，以确定框架是否异常。

#### 4.3. 通过归一化技术改进的 PCANet

在机器学习领域，算法的泛化是衡量算法处理新数据性能的一项重要而艰巨的任务。目前，深度学习领域最流行和最有效的规范化技术是批归一化 (BN) [42]。BN提高了网络的泛化能力，当给定一个样本作为输入时，输出由整个小批量决定；因此，在训练中，它不会为一个样本产生确定性的输出。BN在提高模型泛化能力方面的作用已被实验证明[42]。然而，BN并不适用于我们的自监督模型，因为它的实现过程有两个可训练的参数： $\gamma$ 和 $\beta$ 。在AED-Net中，我们无法找到训练这些参数的方法。此外，在我们的方法中，我们不使用小批量数据。然而，LRN是一种无训练参数的轻量级归一化技术，该技术适用于我们的研究，并在实验中取得了良好的效果。

Krizhevsky等[43]提出，LRN方案有助于提高模型的泛化能力。该方案引入了具有相同空间位置的连续输出之间的响应竞争。对于第 $i$ 个特征映射上的输出值 $p_i$ ，可以按以下方式计算归一化输出 $q_i$ ：

$$q_i = \frac{p_i}{\left[ \alpha + \delta \sum_{j \in \operatorname{nb}(i,n)} (p_j) \right]^\theta} \quad (16)$$

$$\operatorname{nb}(i, n) = \{j | j = \max(0, \frac{i-n}{2}) \dots \min(N-1, \frac{i+n}{2})\}$$

式中,  $\delta$ ,  $n$ ,  $\alpha$ ,  $\theta$ 是可配置的参数;  $\delta$ 表示相邻视频帧的输出权重;  $\alpha$ 是用于计算安全性的偏置项;  $\theta$ 控制归一化项的总幅度;  $n$ 表示在标准化中包含多少相邻视频帧。一旦网络初始化, 网络的特征映射就被设置好了。

为了提高该模型的泛化能力, 我们将该方案从CNN引入PCANet。在每个阶段通过卷积运算计算出特征映射后加入LRN参数。此外, LRN参数在训练前是凭经验设置的, 而且是不可学习的, 这使得LRN适合于我们的非监督框架。

## 5. 实验

我们在UMN数据集[17]与UCSD Ped1和Ped2数据集[18]上进行了局部异常事件检测的实验。这些公开的数据集对整个研究群体开放, 并被用来评估不同标准的AED-Net: 帧级标准和像素级标准。用UMN数据集对模型的容量进行帧级标准评价, 用UCSD Ped1和Ped2数据集对模型同时进行像素级标准和帧级标准评价。这两种评价标准都是基于真阳性率(TPR)和假阳性率(FPR)的, 其中“异常事件”表示为“阳性”, 而“正常状态”表示为“阴性”。将实验结果与用其他先进方法所取得的结果进行比较, 该方法的优越性得到了进一步验证。

### 5.1. UMN 数据集的检测性能

UMN数据集[17]由3个场景组成, 即草坪、室内和

广场, 分辨率为 $240 \times 320$ 。所有的场景都与人群的逃跑行为有关。在此数据集[17]中, 人群的疏散行为被指定为异常行为。我们检测了每个视频帧的异常, 并通过帧级标准来评价。图4展示了来自3个UMN场景的一对帧。为了提高计算效率, 从原始视频帧中提取的所有光流图都被调整为小尺寸, 调整后的光流图已经包含了足够多的检测信息。

为了避免无意义视频帧的干扰, 本实验采用前景检测。如图5所示, 在我们的研究中, 通过测量移动的前景斑点, 可以直接检测到包含不到3个完整人体运动形状的视频帧。

为了提高AED-Net的泛化能力, 本实验采用了一种数据增强技术。首先将光流图尺寸调整为 $120 \times 160$ , 再从调整后的光流图中裁剪出9幅大小为 $96 \times 128$ 的子图。接下来, 所有10幅光流图(一幅为 $120 \times 160$ 和9幅为 $96 \times 128$ )的尺寸都被统一调整为 $24 \times 32$ , 用于训练和测试。

在去除干扰帧后, 我们为每个场景构造了一个训练集和测试集。在草坪场景中, 我们使用了760个普通帧进行训练, 形成了7600个训练集, 而其他正常帧和异常帧则被用于测试。对于室内场景和广场场景, 用于训练的帧数分别为1100帧和1000帧。

对于这3个场景, AED-Net中的超参数设置如下: 每个阶段的滤波器大小为 $3 \times 3$ 。每个阶段都配备8个滤波器以预留足够的方差。分块大小最终为 $8 \times 8$ 。分类

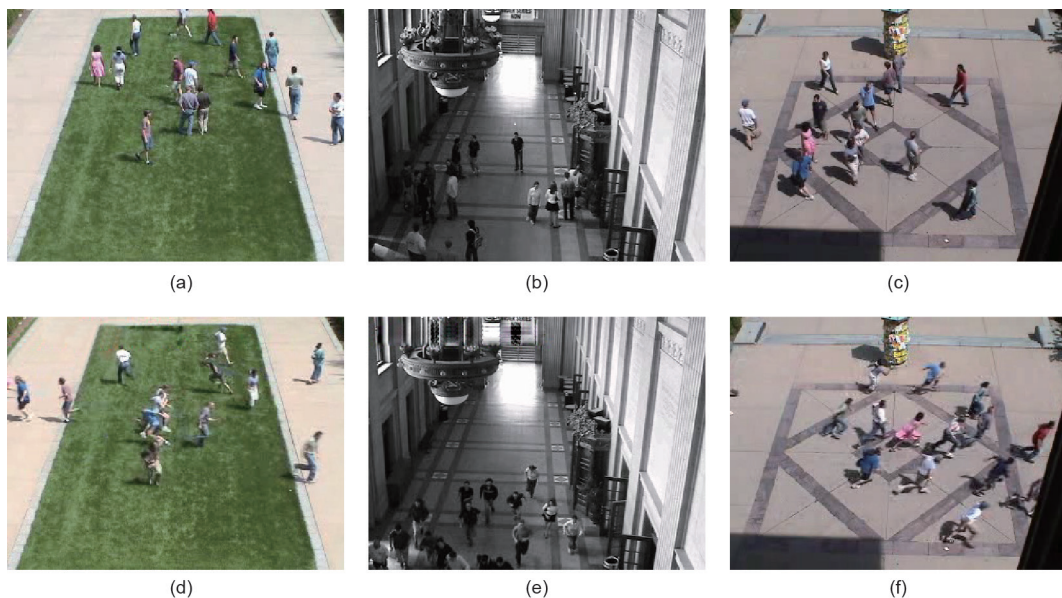


图4. 3个场景的视频帧示例。(a, d)显示了草坪场景;(b, e)显示了室内场景;(c, f)显示了广场场景。人群的疏散行为(d-f)被指定为异常行为。



器中的超参数、内核大小 $\sigma$ 和滤波器数量 $q$ 因场景不同而不同。经过交叉验证,用于草坪、室内和广场场景的 $(\sigma, q)$ 分别设置为 $(1, 2800)$ ,  $(1, 3800)$ 和 $(0.25, 4200)$ 。采用帧级标准对受试者操作特性(ROC)曲线、曲线下面积(AUC)和等误差率(EER)进行分析。在绘制ROC曲线时,用于确定异常帧的阈值被改变。实验结果以及与其他方法的比较见表2[9,15,23,34,36]。如表2所示,我们利用AUC和EER测量的方法在帧级异常检测方面取得了令人满意的结果。鉴于整个框架的简化性,这一成果是显著的,且优于最先进的方法。

## 5.2. UCSD 数据集的检测性能

UCSD数据集[18]包含从人行道上安装的摄像机中获得的分辨率为 $158 \times 238$ 的视频片段。Ped1场景中

有34个训练样本和36个测试样本, Ped2场景中有16个训练样本和12个测试样本,其中包括向不同方向行走的人。被标记为异常的视频片段存在单一异常,如汽车、骑自行车者等。图6显示了一个包含一辆汽车的异常帧。每个视频帧被分割成大小为 $12 \times 16$ 的小块,其中包含步行者或异常部分。然后将这些小块作为原始数据使用。对这些小块的异常性被认为是“基于像素级标准的异常检测”,因为它涉及对视频帧中不同区域对应的像素异常进行分类。

与前面的实验相似,本次实验也进行了前景检测以避免干扰。在此基础上,以包含骑自行车者的异常视频帧中的正常块作为训练集,以两个视频片段中的两个帧的异常块作为测试集。AED-Net中的超参数设置为: $k_1 = k_2 = 5$ ,  $L_1 = L_2 = 7$ , 分块大小为 $7 \times 7$ 。kPCA分类器

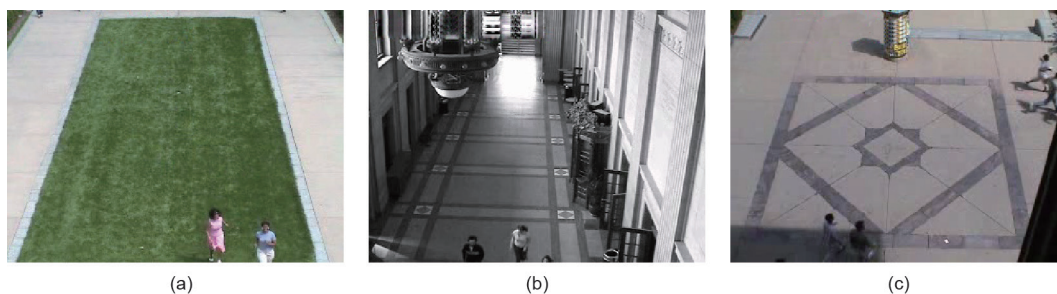


图5. 考虑视频帧中前景区域对检测的干扰,检测异常视频帧的示例。与图4一样,(a)显示了草坪场景;(b)显示了室内场景;(c)显示了广场场景。

表2 利用UMN数据集进行测试的结果比较

Method	AUC (%)	EER (%)	Ref.
Li et al.	99.5	3.7	[9]
Chaotic invariants	99.4	5.3	[23]
Social force	94.9	12.6	[36]
Sparse	99.6	2.8	[34]
Bao et al.	—	2.6	[15]
Ours	<b>99.7</b>	<b>2.4</b>	—

Bold values indicate the present work study of this paper. SF: social force.



图6. 包含异常视频片段的帧的示例。(a)包含骑自行车者的异常视频片段的帧;(b)包含汽车的异常视频片段的帧。

中的超参数设置为:  $(\sigma, q)$  为  $(0.8, 1350)$ 。

Ped1 像素级和帧级评价结果以及与其他方法的比较, 如图7和表3 [9,18,28,34,36]所示。Ped2 像素级和帧级评价结果如表4 [9,18,36]所示。在所有实验中, AED-Net框架的性能均优于现有技术方法, 特别是在AUC方面。

### 5.3. 改进型 AED-Net 的实验研究

在PCANet中添加了LRN层之后, 我们对整个框架在UCSD数据集上进行了测试, 所采用的实验设置与

在UCSD数据集上的相同。LRN的超参数设置为 $\gamma=2$ ,  $\delta=1 \times 10^{-4}$ , 其中 $n=5$ ,  $\beta=0.75$ 。

结果(如图7、表3和表4所示)表明, 在添加LRN层之后, 整个框架在检测由AUC和EER测量的异常方面表现出更好的性能。这些结果表明, 该策略通过提高泛化能力改进了我们提出的方法。

## 6. 结论

在本研究中, 我们基于一种自监督的学习方法, 提

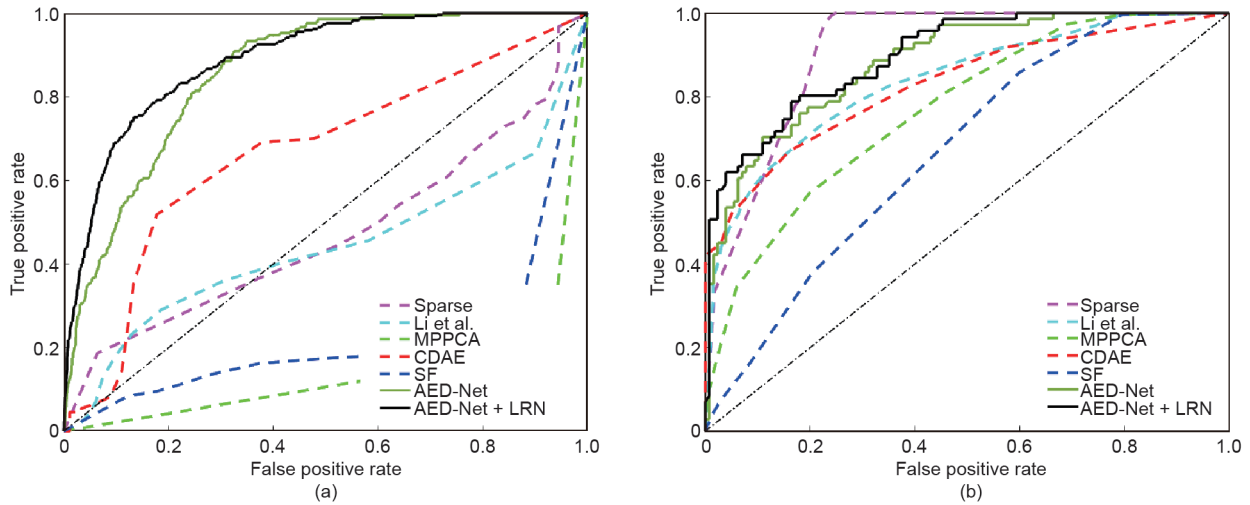


图7. Ped1场景的结果。(a) 用于Ped1的像素级ROC; (b) 用于Ped1的帧级ROC。

表3 UCSD Ped1场景的结果比较

Method	AUC (%)		EER (%)		Ref.
	Pixel level	Frame level	Pixel level	Frame level	
Li et al.	44.1	83.8	55.0	24.4	[9]
MPPCA	20.5	79.6	71.8	32.9	[18]
CDAE	65.8	82.9	36.9	26.8	[28]
sparse	46.1	90.1	53.7	18.6	[34]
SF	17.9	67.0	79.0	39.2	[36]
AED-Net	86.1	88.2	22.9	22.6	—
AED-Net + LRN	88.9	89.7	19.4	19.1	—

MPPCA: mixture of probabilistic principal component analyzers; CDAE: covariance matrix of optical flow features for detection of abnormal events.

表4 UCSD Ped2场景的结果比较

Method	AUC (%)		EER (%)		Ref.
	Pixel level	Frame level	Pixel level	Frame level	
Li et al.	70.0	85.2	29.3	18.2	[9]
MPPCA	23.5	77.6	71.2	30.4	[18]
SF	29.1	71.6	80.2	42.3	[36]
AED-Net	88.9	90.2	22.4	20.3	—
AED-Net + LRN	91.3	89.6	16.8	15.9	—

出了一个简单而有效的AED-Net框架。我们从监控视频片段中提取原始数据计算了光流图；然后用PCANet提取了这些光流图的高层特征，从而进一步确定局部异常事件和全局异常事件的异常情况。实验结果表明，该框架对全局异常事件和局部异常事件都有较好的检测效果。此外，在添加LRN层用于解决过拟合问题后，该框架的性能得到了改善。该框架获取的效果优于现有技术方法所取得的效果，这表明该框架可以有效地从原始视频中提取运动模式并用于异常检测。

## 致谢

这项工作得到了国家重点研究开发计划(2016YFE0204200)、国家自然科学基金(61503017)、中央大学基础研究基金(YWF-18-BJ-J-221)、中国航空科学基金会(2016ZC51022)和由Région Champagne-Ardenne以及FEDER(fonds européen de développement régional)资助的CAPSEC(capteurs pour la sécurité)平台的支持。

## Compliance with ethics guidelines

Tian Wang, Zichen Miao, Yuxin Chen, Yi Zhou, Guangcun Shan, and Hichem Snoussi declare that they have no conflict of interest or financial conflicts to disclose.

## References

- [1] Kalal Z, Matas J, Mikolajczyk K. P-N learning: bootstrapping binary classifiers by structural constraints. In: Proceedings of the Computer Vision and Pattern Recognition; 2010 Jul 13–18; San Francisco, CA, USA. New York: IEEE; 2010. p.49–56.
- [2] Rui C, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the European Conference on Computer Vision; 2012 Oct 7–13; Florence, Italy. Berlin: Springer; 2012. p.702–15.
- [3] Zhang K, Zhang L, Yang MH. Fast compressive tracking. IEEE Trans Pattern Anal Mach Intell 2014;36(10):2002–15.
- [4] Thapar D, Nigam A, Aggarwal D, Agarwal P. VGR-net: a view invariant gait recognition network. In: Proceedings of the IEEE International Conference on Identity, Security, and Behavior Analysis. 2018 Jan 11–12; Singapore. New York: IEEE; 2018.
- [5] Wu Z, Huang Y, Wang L, Wang X, Tan T. A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell 2017;39(2):209–26.
- [6] Bojanowski P, Bach F, Laptev I, Ponce J, Schmid C, Sivic J. Finding actors and actions in movies. In: Proceedings of the IEEE International Conference on Computer Vision. 2013 Dec 1–8; Sydney, Australia. New York: IEEE; 2013.
- [7] Cricri F, Roininen MJ, Leppanen J, Mate S, Curcio IDD, Uhlmann S, et al. Sport type classification of mobile videos. IEEE Trans Multimed 2014;16(4):917–32.
- [8] Wang T, Chen Y, Zhang M, Chen J, Snoussi H. Internal transfer learning for improving performance in human action recognition for small datasets. IEEE Access 2017;5:17627–33.
- [9] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes. IEEE Trans Pattern Anal Mach Intell 2014;36(1):18–32.

- [10] Wang T, Snoussi H. Detection of abnormal visual events via global optical flow orientation histogram. IEEE Trans Inf Forensics Security 2014;9(6):988–98.
- [11] Pan Y. Heading toward artificial intelligence 2.0. Engineering 2016;2(4):409–13.
- [12] Xing EP, Ho Q, Xie P, Wei D. Strategies and principles of distributed machine learning on big data. Engineering 2016;2(2):179–95.
- [13] Wang T, Chen Y, Qiao M, Snoussi H. A fast and robust convolutional neural network-based defect detection model in product quality control. Int J Adv Manuf Technol 2018;94(9–12):3456–71.
- [14] Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y. PCANet: a simple deep learning baseline for image classification? IEEE Trans Image Process 2015;24(12):5017–32.
- [15] Bao T, Karmoshi S, Ding C, Zhu M. Abnormal event detection and localization in crowded scenes based on PCANet. Multimedia Tools Appl 2017;76(22):23213–24.
- [16] Hoffmann H. Kernel PCA for novelty detection. Patt Recog 2007;40(3):863–74.
- [17] University of Minnesota [Internet]. Detection of unusual crowd activity; 2006 [cited 2019 February 24]. Available from: [http://mha.cs.umn.edu/proj\\_events.shtml](http://mha.cs.umn.edu/proj_events.shtml).
- [18] Mahadevan V, Li W, Bhalodia V, Vasconcelos N. Anomaly detection in crowded scenes. In: Proceedings of the Computer Vision and Pattern Recognition; 2010 Jun 13–18; San Francisco, CA, USA. New York: IEEE; 2010.
- [19] Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. Trans Pami 2000;22(8):747–57.
- [20] Zhang T, Lu H, Li SZ. Learning semantic scene models by object classification and trajectory clustering. In: Proceedings of the 2009 IEEE Conference on the Computer Vision and Pattern Recognition; 2009 Jun 20–25; Miami, FL, USA. New York: IEEE; 2009.
- [21] Siebel NT, Maybank SJ. Fusion of multiple tracking algorithms for robust people tracking. In: Proceedings of the European Conference on Computer Vision; 2002 May 28–31; Copenhagen, Denmark; 2002.
- [22] Cui X, Liu Q, Gao M, Metaxas DN. Abnormal detection using interaction energy potentials. In: Proceedings of the Computer Vision and Pattern Recognition; 2011 Jun 20–25; Colorado Springs, CO, USA. New York: IEEE; 2011.
- [23] Wu S, Moore BE, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. Proceedings of the Computer Vision and Pattern Recognition; 2010 Jun 13–18; San Francisco, CA, USA. New York: IEEE; 2010.
- [24] Khalid S. Activity classification and anomaly detection using m-medoids based modelling of motion patterns. Pattern Recognit 2010;43(10):3636–47.
- [25] Rasheed N, Khan SA, Khalid A. Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks. In: Proceedings of the 28th International Conference on Advanced Information Networking and Applications Workshops; 2014 May 13–16; Victoria, BC, Canada. New York: IEEE; 2014.
- [26] Zhang Y, Qin L, Ji R, Yao H, Huang Q. Social attribute-aware force model: exploiting richness of interaction for abnormal crowd detection. IEEE Trans Circ Syst Video Tech 2015;25(7):1231–45.
- [27] Yuan Y, Fang J, Wang Q. Online anomaly detection in crowd scenes via structure analysis. IEEE Trans Cybern 2015;45(3):548–61.
- [28] Wang T, Qiao M, Zhu A, Niu Y, Li C, Snoussi H. Abnormal event detection via covariance matrix for optical flow based feature. Multimedia Tools Appl 2018;77(13):17375–95.
- [29] Benezeth Y, Jodoin PM, Saligrama V, Rosenberger C. Abnormal events detection based on spatio-temporal co-occurrences. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009 Jun 20–25; Miami, FL, USA. New York: IEEE; 2009.
- [30] Benezeth Y, Jodoin PM, Saligrama V. Abnormality detection using low-level co-occurring events. Pattern Recognit Lett 2011;32(3):423–31.
- [31] Jiménez-Hernández H, González-Barbosa JJ, García-Ramírez T. Detecting abnormal vehicular dynamics at intersections based on an unsupervised learning approach and a stochastic model. Sensors 2010;10(8):7576–601.
- [32] Kosmopoulos D, Chatzis SP. Robust visual behavior recognition. IEEE Signal Process Mag 2010;27(5):34–45.
- [33] Adam A, Rivlin E, Shimshoni I, Reinitz D. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell 2008;30(3):555–60.
- [34] Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection. In: Proceedings of the Computer Vision and Pattern Recognition; 2011 Jun 20–25; Colorado Springs, CO, USA. New York: IEEE; 2011. p. 3449–56.
- [35] Zhao B, Li F, Xing EP. Online detection of unusual events in videos via dynamics sparse coding. In: Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition. 2011 Jun 20–25; Colorado Springs, CO, USA. New York: IEEE; 2011. p. 3313–20.
- [36] Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009 Jun 20–25; Miami, FL, USA. New York: IEEE; 2009.
- [37] Bhatnagar G, Wu QMJ, Raman B. Discrete fractional wavelet transform and

- its application to multiple encryption. *Inform Sci* 2013;223:297–316.
- [38] You X, Du L, Cheung Y, Chen Q. A blind watermarking scheme using new non-tensor product wavelet filter banks. *IEEE Trans Image Process* 2010;19(12):3271–84.
- [39] Kemmler M, Rodner E, Wacker ES, Denzler J. One-class classification with Gaussian processes. *Patt Recog* 2013;46(12):3507–18.
- [40] Burton A, Radford J. *Thinking in perspective: critical essays in the study of thought processes*. North Yorkshire: Methuen; 1978.
- [41] Horn BK, Schunck BG. Determining optical flow. *Artif Intell* 1981;17(1–3):185–203.
- [42] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv:1502.03167.
- [43] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*; 2012 Dec 3–6; Lake Tahoe, NV, USA. New York: ACM; 2012.