

DOI: 10.19650/j.cnki.cjsi.J2107458

基于卷积变分自编码器的异常事件检测方法*

于晓升¹, 许茗¹, 王莹¹, 王思齐¹, 胡楠²

(1. 东北大学机器人科学与工程学院 沈阳 110169; 2. 沈阳建筑大学信息与控制工程学院 沈阳 110168)

摘要:针对传统的异常事件检测方法中手工提取的浅层特征无法统一表达不同场景的表观和运动信息问题,提出了一种基于卷积变分自编码器(ConVAE)的异常事件检测方法。首先构建了一个卷积变分自编码器网络,以视频帧序列作为网络输入,提取场景的深度特征;接着采用多变量高斯模型对在所有帧序列的深度特征图中位于同一位置的特征向量进行拟合,获得对应于原始输入中不同感受野的多变量高斯模型;对于测试样本的深度特征在其对应的高斯模型中的拟合程度,作为异常检测的判断标准。使用UCSD异常事件检测数据集对本文提出的方法进行实验验证,在帧级别与像素级别两种级别的度量标准中,本文算法分别取得了95.7%和69.9%的受试者工作特征曲线下面积值。

关键词:卷积变分自编码器;异常事件检测;深度特征;多变量高斯模型

中图分类号: TP391.4 TH701 **文献标识码:** A **国家标准学科分类代码:** 510.99

Anomaly detection method based on convolutional variational auto-encoder

Yu Xiaosheng¹, Xu Ming¹, Wang Ying¹, Wang Siqi¹, Hu Nan²

(1. Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110169, China;

2. School of Information and Control Engineering, Shenyang Jianzhu University, Shenyang 110168, China)

Abstract: The hand-crafted features in the traditional anomaly detection algorithms can not represent the appearance and motion patterns in a unified way for different scenes. In this paper, we propose a novel anomaly detection algorithm based on the convolutional variational auto-encoder (ConVAE). Firstly, a ConVAE which takes the raw frames series as input is constructed to extract the deep features of the scene. These deep features can represent the appearance and motion patterns more specifically. And then multiple Gaussian models are employed to fit the deep feature vectors of the corresponding receptive fields. The fitted Gaussian models which correspond to the receptive fields are used to decide the deep feature of the corresponding receptive fields from the test sample is anomalous or not. The proposed anomaly detection algorithm is evaluated the UCSD anomaly detection datasets. Experimental results show that the area under curve (AUC) of the proposed method are 95.7% and 69.9% in frame-level and pixel-level, respectively.

Keywords: convolutional variational auto-encoder; anomaly detection; deep features; multiple Gaussian models

0 引言

智能视频监控系统无论在工业领域还是民用领域都有着广泛的应用。在智能监控系统中,海量的监控视频数据需要被实时快速准确地处理,进而完成相关的监控任务。异常事件检测,即自动检测识别监控场景中发生的异常事件,是智能视频监控系统中的一项重要的任务。

作为智能视频监控系统中的事件级目标检测任务,异常事件检测需要同时综合考虑目标事件的空间信息与时间信息,在视频中对异常事件进行快速准确地标记。异常事件检测一直以来都受到研究者的广泛关注,是图像处理 and 计算机视觉领域中的一个热点研究问题^[1-2]。

在智能视频监控系统中,异常事件检测面临许多的挑战。其中,最大的挑战源于异常事件本身,即如何明确定义事件是否异常^[3-4]。通常情况下,需要借助于场景中

收稿日期:2021-01-30 Received Date: 2021-01-30

* 基金项目:中央高校基本科研业务费专项资金项目(N2026005)、中国科学院光电信息处理重点实验室开放基金项目(OEIP-O-202005)、辽宁省教育厅科学研究经费项目(lnqn202014)资助



经常发生的事件对异常事件进行定义,例如行人场景中 出现骑行自行车和驾驶汽车的情况。在不同的场景中, 异常事件的定义往往是不同的,一个场景中出现的异常 事件在另外一个场景中很可能属于正常事件。从模式识 别的范畴来说,异常事件检测属于一类新奇点检测,即通 过对仅包含正常数据的样本进行学习,实现对未知数据 样本中是否包含异常数据进行判断和定位。

现有的异常事件检测方法主要包括如下两个类别: 基于传统计算机视觉方法和基于深度学习方法。基于计 算机视觉方法主要从分析场景中局部或者个体的时空模 式着手,提取手工设计特征,即浅层特征,再结合分类器 实现异常事件的检测^[5]。因此,研究者们设计开发了多 种不同的特征描述方法,描述场景中目标的形状和运动 信息。应用于异常事件检测常用的浅层特征包括梯度方 向直方图(histogram of oriented gradients, HOG),三维时 空梯度向量以及光流方向直方图(histogram of oriented flows, HOF)等^[6-7]。尽管如此,这些方法均属于普适性 较强的特征提取方法,与针对特定目标设计的特征描述 方法相比,局限性较为明显。此外,在其他一些研究工 作中,主要针对场景中的运动信息进行分析,例如通过局部 轨迹或者像素光流图描述运动模式^[8-9]。但是这类方法 对场景环境要求较高,当光照变化剧烈或多目标相互遮 挡严重时,检测结果不理想。

近年来,随着深度学习在计算机视觉领域的兴起,研 究者们开始尝试采用深度学习方法进行异常事件检测。 与图像分类^[10]、目标检测^[11]等传统计算机视觉任务不 同,事件是否异常在不同的场景中有着不同的定义。因 此,基于深度学习的异常事件检测方法通常采用自编码 器(auto-encoder, AE)、变分自编码器(variational auto- encoder, VAE)等无监督学习方法提取场景的深度特 征^[12]。相比于传统的手工设计的特征,深度网络能够从 场景中学习到更丰富更独特具有针对性的特征^[13]。在 完成场景深度特征获取的基础上,一些工作基于不同的 全卷积网络^[14](fully convolutional network, FCN)结构,重 构或者生成一幅完整的新的视频帧或者一段完整的视频 片段,然后通过计算新生成的数据与原始输入数据的整 体差异来估计原始输入数据的异常性。但是,这类方法 由于整合了输入与输出的全部信息,导致其无法确定异 常事件在场景中的具体位置。另外一些工作^[2,15-17]则采 用概率密度估计模型,如单类支持向量机、高斯模型等, 对已获取的特征进行异常性判断,定位异常事件。但是 在该类方法中,来自场景中不同区域的所有特征均共享 同一个模型,这极大地影响了检测结果的准确性^[18]。

针对上述问题,本文提出了一种基于卷积变分自编 码器(convolutional variational auto-encoder, ConVAE)的 异常事件检测方法。该方法采用 ConVAE 作为主体结

构,具有如下 3 个方面优点:1)对原始输入帧序列无需进 行任何目标实例级别分析作为预处理,直接提取表观和 运动特征;2)可同时提取稠密分布于输入帧序列中的所 有感受野的表观与运动特征,无需对输入帧序列进行切 块处理,算法效率较高;3)采用 ConVAE 提取场景的深度 特征,提取的特征具有高独立性和低耦合性。本文使用 UCSD^[19]异常事件检测数据集验证本文方法的有效性。 大量实验结果证明,本文算法能够快速准确检测视频监 控场景下的异常事件,取得了较好的结果。

1 基于 ConVAE 的异常事件检测

本文所提算法的流程如图 1 所示,该算法主要包含 如下 3 个步骤:

1)以若干帧视频组成的张量形式的帧序列作为网络 输入,采用所构建的 ConVAE 网络提取场景特征,实现场 景表观与运动信息的深层次表达。因此,在 ConVAE 输 出的特征图中,每个位置的特征向量均可视为其在输入 中对应感受野的表观与运动信息的特征表达。

2)统计所有帧序列的深度特征,通过多变量高斯模 型拟合深度特征^[20-21]中位于同一位置的所有特征向量。 因此,不同位置的特征向量所对应的感受野有其对应的 多变量高斯模型。

3)基于拟合的高斯模型对测试样本每一个位置的 特征向量进行异常性判断,拟合程度较低的特征向量即 被判为异常,并将该特征向量映射至原始测试样本,作为 异常事件检测的最终结果。

1.1 基于 ConVAE 的异常事件特征提取

异常事件通常指的是场景中在表观和运动模式上明 显有别于其它经常发生事件的特殊事件。由于不同实际 场景下的异常事件定义往往不同,因此无法明确定义具 体事件类别,进而导致基于有监督学习的特征提取方法 无法工作。因此需要一个无监督学习方法对场景中的背 景和正常事件的表观与运动模式进行建模,在此基础上, 构建异常事件判别准则,检测出现在未知视频数据中的 异常事件。在本文所提方法中,用做训练样本的视频数 据集中不包含异常事件,采用 ConVAE 提取场景的深度 特征表达该场景的表观与运动信息。

本文参考了文献[12]中提出的卷积自编码器网络 结构,设计了 ConVAE 网络结构,如图 2 所示。该网络采 用全卷积网络结构,分别由 1 个编码器网络、1 个解码器 网络构成。编码器网络可根据所使用的卷积核尺寸不同 划分为 4 个卷积模块。卷积模块 1 包含 1 个卷积层与 1 个最大值池化层,其中卷积核尺寸为 7×7 且步长为 1×1 ,最大值池化层尺寸为 2×2 ;卷积模块 2 包含 3 个卷 积层,其中所有卷积核尺寸为 5×5 且步长为 1×1 ;卷积

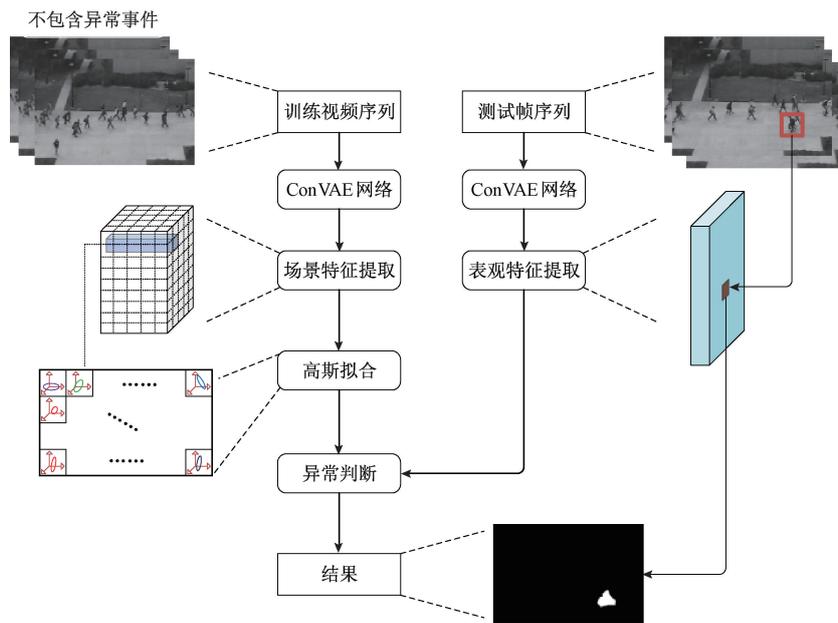


图 1 算法流程

Fig. 1 The flow of the proposed algorithm

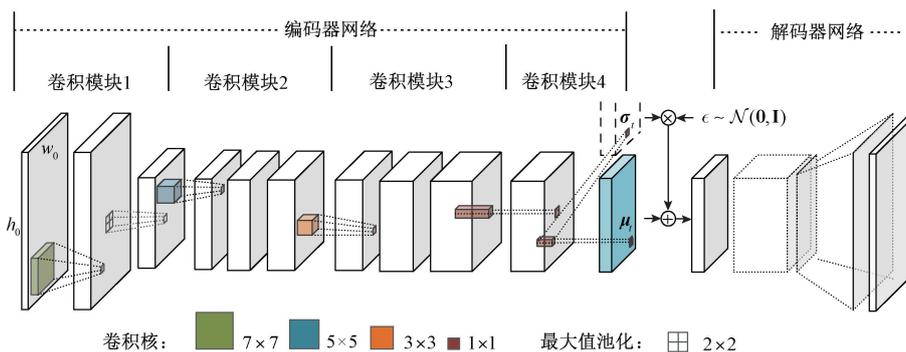


图 2 ConVAE 网络结构

Fig. 2 ConVAE network architecture

模块 3 包含 3 层卷积层,卷积核尺寸均为 3×3 且步长为 1×1;卷积模块 4 包含 2 层卷积层,卷积核尺寸均为 1×1 且步长为 1×1;除卷积模块 4 最后一层卷积层外,其他所有卷积层均采用 ReLU 非线性函数作为激活函数^[12]。解码器网络结构采用与编码器网络结构完全对称的网络结构,只是将所有卷积层和最大池化层分别由转置卷积层和最大反池化层替代。

本文的研究对象是由若干帧监控图像组成的帧序列,采用所设计的 ConVAE 网络提取该帧序列的深度特征可以同时表达场景的表观与运动模式。假设 \mathcal{I} 为一个不包含任何异常事件的训练数据集,该数据集由 T 帧画面构成, f_t 表示第 t 帧图像,第 $t-1$ 帧图像与第 $t-2$ 帧图像的对应像素平均为 $f_t^{(-1)}$,第 $t+1$ 帧图像与第 $t+2$ 帧图像的对应像素平均为 $f_t^{(+1)}$,即:

$$f_t^{(-1)} = \frac{f_{t-2} + f_{t-1}}{2} \tag{1}$$

$$f_t^{(+1)} = \frac{f_{t+1} + f_{t+2}}{2} \tag{2}$$

由此,可以构建一个帧序列 $\mathcal{I}_t = \{f_t^{(-1)}, f_t, f_t^{(+1)}\}$ 作为 ConVAE 网络的输入,用以检测视频帧 f_t 是否包含异常事件。对于整个训练数据集,帧序列集合可表示为 $\mathcal{I} = \{\mathcal{I}_t \mid t = 3, 4, \dots, T-2\}$ 。

在 ConVAE 网络中,编码器的输出为用于描述隐层变量分布的参数,即均值与标准差。在本文中,帧序列 \mathcal{I}_t 作为网络输入,假设其分辨率 $h_0 \times w_0$,经过编码器网络处理,输出张量形式均值图 μ_t 与标准差图 σ_t 作为描述隐层变量的分布参数, $\mu_t, \sigma_t \in \mathbf{R}^{h \times w_l \times c_l}$,其中 h, w 分别表示张量图的高度与宽度, c 表示张量图的通道数, l 是编码



器网络的层数。在此基础上,采用重参数技巧从隐层变量的后验分布中生成新的隐层变量样本。最后,将新生成隐层变量输入解码器网络,获得网络的输出结果 $\hat{\mathcal{I}}_t$ 。

本文构建的 ConVAE 网络的损失函数分别由重构损失和 KL(kullback-leibler, KL) 散度两个部分构成,其中重构损失用于确保网络输出的结果 $\hat{\mathcal{I}}_t$ 能够尽可能接近原始输入帧序列 \mathcal{I}_t 。因此,这部分损失函数定义如下:

$$\mathcal{L}_{con} := \|\mathcal{L}_t - \hat{\mathcal{I}}_t\|^2 \quad (3)$$

编码器网络输出均值张量 $\boldsymbol{\mu}_t$ 和标准差 $\boldsymbol{\sigma}_t$,由该均值与标准差作为分布参数所描述的后验概率通常是假设近似于具有对角协方差的高斯分布,因此采用 KL(Kullback-Leibler, KL)散度量两个概率分布 $N(\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^2)$ 和 $N(0, \mathbf{I})$ 之间的差异性,即:

$$\mathcal{L}_{KL} := D_{KL}(N(\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^2) \| N(0, \mathbf{I})) \quad (4)$$

KL 散度的无偏估计定义如下:

$$D_{KL}(N(\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^2) \| N(0, \mathbf{I})) = -\frac{1}{2} \sum_{k=1}^K (1 + \log(\sigma_{ik}^2) - (\boldsymbol{\mu}_{ik})^2 - \sigma_{ik}^2) \quad (5)$$

式中: $K = h_t \times w_t \times c_t$ 为均值张量的所有元素个数,标准差张量也具有相同元素个数。损失函数的完整形式定义如下:

$$\mathcal{L} = \mathcal{L}_{con} + \mathcal{L}_{KL} \quad (6)$$

ConVAE 的核心功能是学习输入的特征表达的后验概率分布,即以均值 $\boldsymbol{\mu}$ 和标准差 $\boldsymbol{\sigma}$ 为参数的高斯分布。从数值计算角度来看,均值 $\boldsymbol{\mu}$ 可被视为输入数据的一种统计表达而标准差 $\boldsymbol{\sigma}$ 则为一个噪声强度调节器,用以调节在网络隐层变量中添加的噪声强度。因此,在本文中,对于输入帧序列 \mathcal{I}_t ,编码器网络输出的均值张量 $\boldsymbol{\mu}_t$ 可作为场景表观与运动模式的特征表达 $\boldsymbol{\theta}_t$ 。如果编码器输出层(第 l 层)输出的均值张量 $\boldsymbol{\mu}_t$ 的尺寸为 $h_l \times w_l \times c_l$,那么特征表达 $\boldsymbol{\theta}_t$ 可认为包含 $h_l \times w_l$ 个特征向量,每个特征向量的维度为 c_l 。对于特征张量 $\boldsymbol{\theta}_t$ 上的每个位置 (i, j) ,其中 $i \in [1, h_l], j \in [1, w_l]$,特征向量定义如下:

$$\boldsymbol{\theta}_t(i, j) = [\boldsymbol{\theta}_t(i, j, 1), \boldsymbol{\theta}_t(i, j, 2), \dots, \boldsymbol{\theta}_t(i, j, c_t)]^T \quad (7)$$

由卷积神经网络的结构可知,每个特征向量 $\boldsymbol{\theta}(i, j)$ 均来自其对应的感受野,每个感受野则是原始输入帧序列 \mathcal{I}_t 的一个子区域。由此可见,本文充分利用了卷积神经网络的连接结构,稠密地将帧序列 \mathcal{I}_t 切分为相互交叠的且包含不同表观与运动信息的感受野,并且实现了一次性从所有感受野中提取特征 $\boldsymbol{\theta}_t = \{\boldsymbol{\theta}_t(i, j) \mid i = 1, 2, \dots, h; j = 1, 2, \dots, w\}$ 。与传统卷积自编码器相比,本文构建的 ConVAE 网络提取的特征能够对感受野所属于的某一类别的表观与运动模式进行深度表达,而不仅仅是存在于训练数据集中的若干特定样本的表观与运动模式进行深度表达,并且特征向量中各元素之间更趋于相互独立且解耦。

1.2 异常事件识别与定位

训练集是由不包含任何异常事件的帧序列集合构成,本文构建的 ConVAE 网络提取的这些帧序列的表观与运动特征可视为正常特征。通过对这些正常特征进行建模,可用以检验之后的测试集中的特征是否为异常。本文采用多变量高斯模型拟合源自不同感受野但处于相同位置的正常特征,构建它们各自的高斯模型。即在特征张量图中的每个位置 (i, j) ,构建一个专属的变量数为 c_t 的多变量高斯模型 G_{ij} ,用以拟合来自帧序列集合 \mathcal{I} 的特征向量 $\{\boldsymbol{\theta}_t(i, j) \mid t = 3, 4, \dots, T - 2\}$:

$$\boldsymbol{\theta}_t(i, j) \sim N(\boldsymbol{\mu}_{ij}^c, \sum_{ij}^c) \quad (8)$$

式中: $\boldsymbol{\mu}_{ij}^c$ 和 \sum_{ij}^c 分别表示 G_{ij} 的均值向量和协方差矩阵。

本文将拟合得到的所有多变量高斯模型作为表示正常事件的基准模型,进而对异常事件进行判断,提出如下的判断准则:在编码器提取的输入帧的深度特征中,如果某个位置上的特征向量在其对应位置的混合高斯模型中拟合程度较低,那么可以判定该特征向量及其对应的感受野具有异常性;反之则判定为正常的特征向量。衡量拟合程度的标准定义如下:

$$A_t(i, j) = G_{ij}(\boldsymbol{\theta}_t^{test}(i, j) \mid \boldsymbol{\mu}_{ij}^c, \sum_{ij}^c) \quad (9)$$

式中: $\boldsymbol{\theta}_t^{test}(i, j)$ 表示 t 时刻输入帧的深度特征。

本文定义一个判定阈值 τ ,当 $A_t(i, j) \geq \tau$ 时,特征向量及其对应的感受野则判定为正常;当 $A_t(i, j) < \tau$ 时,特征向量及其对应的感受野则判定为异常。此外,对于不同的场景,阈值的选择也是不同的。通过对所有特征向量进行判断,并将异常的特征向量映射回原始测试样本中即可得到异常事件检测的最终结果。

2 实验结果与分析

本文采用提出的方法检测异常事件,并与文献[5]提出的基于稀疏字典学习的异常事件检测算法,文献[8]提出的基于混合动态纹理分析的异常事件检测算法,文献[16]提出的基于稀疏重构的异常事件检测算法以及文献[17]提出的基于形状运动模型深度网络的异常事件算法的实验结果相比较,验证本文所提算法的有效性。检测算法使用的所有图像数据均来自 UCSD 异常事件检测数据集。该数据集中的所有数据均为一个固定位置摄像头拍摄的一个行人通道上的视频监控数据,由美国加利福尼亚大学圣迭戈分校的计算机视觉实验室整理发布。在不同的监控视频中,行人的密度是不同的。除行人外,行人通道上的有时会出现非行人目标,例如汽车以及不同于行走的人体运动模型,例如滑滑板、骑自行车等。该数据集将人们在行人通道中正常行走认定为正

常事件,对于通道上出现非行人目标以及出现不同于行走的人体运动模式等情况,则均认定为异常事件。此外,该数据集中的所有异常事件均是自然发生的,并无事先安排。该数据集包括 32 个训练用视频序列以及 36 个测试用视频序列,训练用视频序列中不包括任何异常事件,而测试用视频序列则正常事件和异常事件均包含,每个视频序列大约 200 帧左右,每一帧视频图像分辨率为 158×238 pixel。

为了评估本文所提异常事件检测算法性能,本文采用受试者工作特征(receiver operating characteristic, ROC)曲线、ROC 曲线下面积(area under curve, AUC)以及等错误率(Equal Error Rate, EER)作为评价指标,将本文提出的异常事件检测算法与其他异常事件检测算法比较。与此同时,本文借鉴文献[22]的提出的度量方式,将同时采用帧级别与像素级别两种级别的度量标准。这两种度量方式所关注的对象均为某一视频图像是否有异常事件发生。在帧级别的度量标准中,只要图像中存在一个及以上像素被判断为异常,那么无论判断是否正确,则该帧视频图像均被认定存在异常事件。而在像素级别的度量标准中,如果被判定为

异常的像素区域占比手工标记的标准真值异常区域不低于 40%,那么该帧视频图像则被认定存在异常事件。在这两种度量标准中,帧级别的度量更加侧重宏观上展现检测系统的性能,而像素级别的度量更加严格,只有在检测的异常事件区域正确的情况下,才能判定检测结果有效。

图 3 所示为本文提出的算法在 UCSD 数据集上进行异常事件检测的部分结果,其中,图 3 第 1 行为人工标记的标准真值异常区域,图 3 第 2 行为本文算法的异常事件检测结果。由图 3 可以看出,本文算法将出现在行人通道上的机动车,滑滑板的人和骑自行车的人准确地标记出来,认定为异常事件,证明了本文算法的有效性。在图 3(b)中,滑滑板的人出现在人群密集的地方,与行人在表现特征上特别相似,密集的人群对异常事件检测产生了强烈的干扰,但本文所提算法仍然能够准确地检测异常事件,这是因为本文算法对场景的表现特征和运动模型进行了深度特征表达,对复杂场景具有较强的鲁棒性。在图 3(a)和(c)中,场景中同时出现多个异常事件,本文算法准确地检测了全部异常事件,达到了很好的效果。

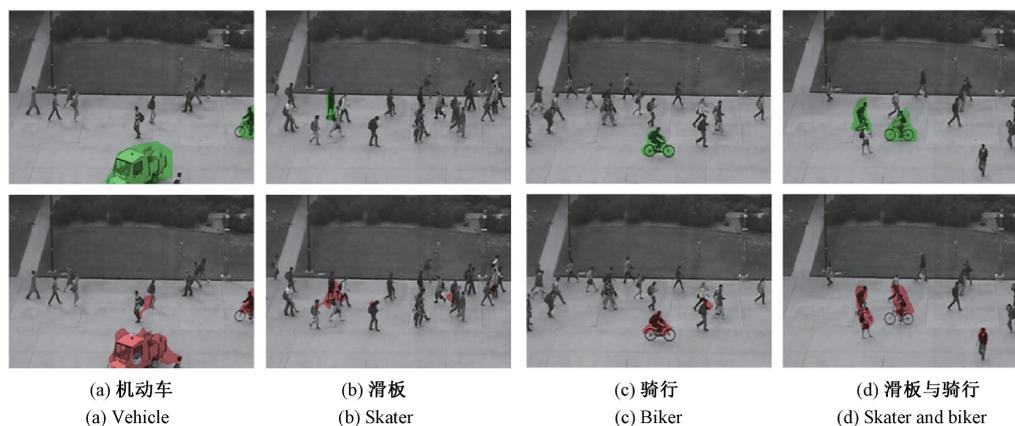


图 3 本文算法在 UCSD 数据集上的检测结果示例

Fig. 3 Detection results by the proposed method on UCSD dataset

图 4 所示为在 UCSD 数据集下 5 种算法各自的 ROC 曲线,其中图 4(a)为各种算法的帧级别 ROC 曲线,图 4(b)为各种算法的像素级别 ROC 曲线。ROC 曲线主要展现了检测算法的可靠性,其横坐标和纵坐标分别为假阳率和真阳率,ROC 曲线越靠近坐标系(0,1),表示该算法性能越好。由图 4(a)可以看出,在帧级别度量标准下,本文所提算法在检测性能上明显优于文献[5]的基于稀疏字典学习的异常事件检测算法和文献[8]的基于混合动态纹理分析的异常事件检测算法,与文献[16]的基于稀疏重构的异常事件检测算法以及文献[17]的基于形状运动模型深度网络的异常事件算法不相上下。而

在图 4(b)所示的像素级别度量标准下,本文所提算法的检测性能则明显优于其他 4 种算法,上述实验结果表明本文所提算法不仅能够准确地判断每一视频图像中是否包括异常事件,并且能够精准地定位监控场景中发生异常事件的区域位置,达到了很好的检测效果。

表 1 和图 5 分别展示了在 UCSD 数据集下 5 种算法的 AUC 与 EER 两种数值指标的对比结果。AUC 值表示 ROC 曲线下方面积,其数值越高,表示算法的性能越好。EER 为 ROC 曲线中错分正负样本概率相等的点,其数值越小,表示算法的性能越好。由表 1 和图 5 可以看出,本文算法性能优于其他 4 种算法。其中,与文献[5]的算

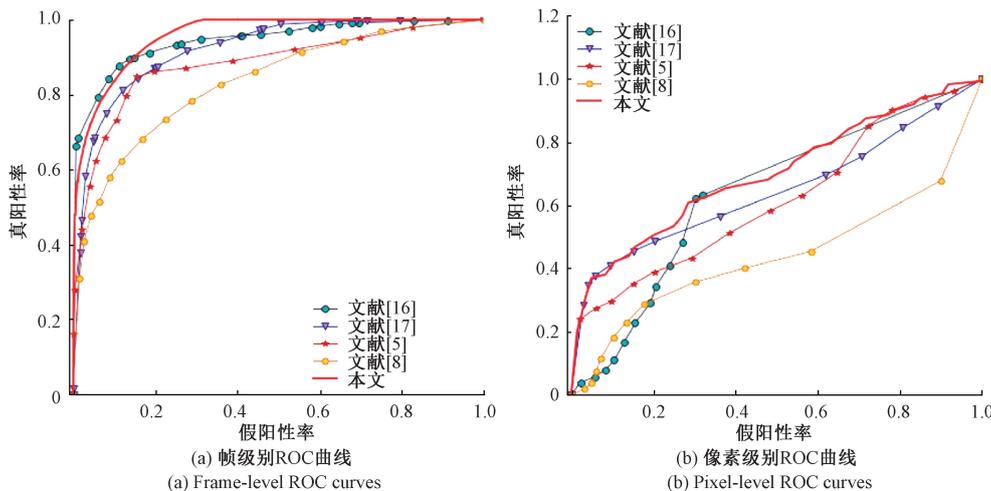


图 4 各种算法在 UCSD 数据集中的帧级别 ROC 曲线对比和像素级 ROC 曲线对比

Fig. 4 The frame-level ROC and the pixel-level ROC comparisons of different methods on UCSD dataset

表 1 5 种算法的 AUC 对比结果

Table 1 AUC comparison with five methods

方法	帧级别/%	像素级别/%
文献[5]	81.8	44.1
文献[8]	91.8	63.8
文献[16]	92.1	67.2
文献[17]	94.7	64.0
本文	95.7	69.9

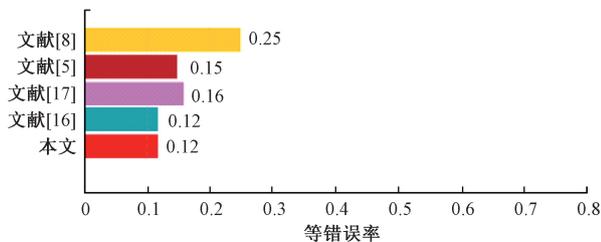


图 5 各种算法帧级别等错误率比较

Fig. 5 EER comparison of frame-level performance with different methods

法和文献[8]算法相比,优势明显。而与文献[16]的算法和文献[17]算法相比,检测性能有了一定程度的提高。具体地,与文献[16]的算法相比,在帧级别度量标准下,本文算法的 AUC 值高出 1%,而 EER 值相同;但是在像素级别度量标准下,本文算法的 AUC 值领先 5.9%,与文献[17]的算法相比,在帧级别度量标准下,本文算法的 AUC 值大约高出 3.6%,EER 值则比其少 0.04;在像素级别度量标准下,本文算法的 AUC 值高出 2.7%。

由上述的对比结果可以看出,与这 4 种算法相比,本文算法具有最好的异常事件检测性能,能够更加准确地定位场景中异常区域。

3 结 论

本文提出一种基于卷积变分自编码器的异常事件检测方法,用于实现视频监控场景下的异常事件检测。该方法采用卷积变分自编码器作为主体结构,将视频帧序列作为网络输入,以提取场景的深度特征;在此基础上,采用多变量高斯模型对在所有深度特征中位于同一位置的特征向量进行拟合,构建异常的判别准则,对测试样本所有位置的深度特征进行异常性判断,最终将被判断为异常的特征向量映射至原始测试样本,作为异常事件检测的最终结果。实验结果证明,与现有的异常事件检测方法相比,本文算法能够更加快速准确地检测视频监控场景下的异常事件,取得了较好的结果。在接下来的工作中,如何进一步提高本文方法性能,减少误检、漏检的情况,将是工作的重点。

参考文献

[1] YE R, LI X. Collective representation for abnormal event detection [J]. Journal of Computer Science and Technology, 2017, 32(3) : 470-479.

[2] 胡海洋, 张力, 李忠金. 融合自编码器和 one-class SVM 的异常事件检测 [J]. 中国图象图形学报, 2020, 25(12) : 2614-2629.

HU H Y, ZHANG L, LI ZH J. Anomaly detection with autoencoder and one-class [J]. Journal of image and



- graphics, 2020, 25(12): 2614-2629.
- [3] 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.
- [4] LIU W, LUO W X, LIAN D Z, et al. Future frame prediction for anomaly detection-a new baseline [C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018:6536-6545.
- [5] LU C, SHI J, JIA J. Abnormal event detection at 150 fps in Matlab [C]. Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 2013: 2720-2727.
- [6] SALIGRAMA V, CHEN Z. Video anomaly detection based on local statistical aggregates[C]. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 2012: 2112-2119.
- [7] HUGO MORA COLQUE R V, CAETANO C, LUSTOSA DE ANDRADE M T, et al. Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(3): 673-682.
- [8] MAHADEVAN V, LI W, BHALODIA V, et al. Anomaly detection in crowded scenes[C]. Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010: 1975-1981.
- [9] ZHANG Y, LU H, ZHANG L, et al. Video anomaly detection based on locality sensitive hashing filters[J]. Pattern Recognition, 2016, 59: 302-311.
- [10] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely connected convolutional networks[C]. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 2261-2269
- [11] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.
- [12] RIBEIRO M, LAZZARETTI A E, LOPES H S. A study of deep convolutional auto-encoders for anomaly detection in videos [J]. Pattern Recognition Letters, 2017, 105: 13-22.
- [13] SULTANI W, CHEN C, SHAH M. Real-word anomaly detection in surveillance videos[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018:6479-6488.
- [14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015: 3431-3440.
- [15] WANG S Q, ZENG Y J, LIU Q, et al. Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection[C]. Proceedings of the 26th ACM International Conference on Multimedia, New York, USA, 2018: 636-644.
- [16] YU B, LIU Y, SUN Q. A content-adaptively sparse reconstruction method for abnormal events detection with low-rank property [J]. IEEE Transactions on Systems Man Cybernetics-Systems, 2017, 47(4): 704-716.
- [17] 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, 2018, 156: 117-127.
- [18] 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.
- [19] 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: 1-8.
- [20] 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 Q, 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.
- [21] 曲建岭, 余路, 袁涛, 等. 基于一维卷积神经网络的滚动轴承自适应故障诊断算法 [J]. 仪器仪表学报,



2018, 39(7): 134-143.

QU J L, YU L, YUAN T, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural network [J]. Chinese Journal of Scientific Instrument, 2018, 39(7): 134-143.

- [22] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model [C]. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009: 20-25.

作者简介



于晓升(通信作者),2006年于英国贝特福德大学获得学士学位,2007年于英国利物浦大学获得硕士学位,2014年于东北大学获得博士学位,现为东北大学讲师,主要研究方向为图像处理与计算机视觉。

E-mail:yuxiaosheng@mail.neu.edu.cn

Yu Xiaosheng (Corresponding author) received his B. Sc.

degree in 2006 from University of Bedfordshire, received his M. Sc. degree in 2007 from University of Liverpool, received his Ph. D. degree in 2014 from Northeastern University, now he is a lecturer in Northeastern University. His main research interests include image processing and computer vision.



胡楠,2010年于沈阳建筑大学获得学士学位,2012年于东北大学获得硕士学位,2016年于东北大学获得博士学位,现为沈阳建筑大学讲师,主要研究方向为图像处理和机器视觉。

E-mail:hunan0806@163.com

Hu Nan received his B. Sc. degree in 2010 from Shenyang Jianzhu University, received his M. Sc. degree in 2012 from Northeastern University, received his Ph. D. degree in 2016 from Northeastern University, now he is a lecturer in Shenyang Jianzhu University. His main research interests include image processing and machine vision.