

视频监控下利用记忆力增强自编码的行人异常行为检测

孙敬波, 季 节

(济宁学院 数学与计算机应用技术学院, 山东 曲阜 273155)

摘要: 随着视频监控数据的快速增长, 对大规模视频数据的自动异常检测的需求越来越大, 基于深度自编码器重构误差检测方法已经被广泛探讨。但是, 有时自编码器“泛化”得很好, 能够很好地重建异常并导致漏检。为了解决这个问题, 提出了采用记忆力模块来增强自动编码器, 称为记忆力增强自编码 (Memory-augmented autoencoder, Memory AE) 方法。给定输入, Memory AE 首先从编码器获取编码, 然后将其用作查询以检索最相关的记忆项来进行重建。在训练阶段, 记忆内容被更新以表示正常数据的原型元素。在测试阶段, 将学习到的记忆元素固定下来, 从正常数据的几个选定的记忆记录中获得重建, 因此重建将趋向于接近正常样本。因此, 将加强对异常的重构误差以进行异常检测。对两个公共视频异常检测数据集, 即 Avenue 数据集和 ShanghaiTech 数据集的研究证明了所提出方法的有效性。

关键词: 异常事件检测; 视频监控; 自编码网络; 记忆力增强; 深度学习

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20210680

Memory-augmented deep autoencoder model for pedestrian abnormal behavior detection in video surveillance

Sun Jingbo, Ji Jie

(School of Mathematics and Computer Application Technology, Jining University, Qufu 273155, China)

Abstract: With the rapid growth of video surveillance data, there is an increasing demand for video anomaly detection, and reconstruction error detection methods based on deep autoencoders have been widely discussed. However, the autoencoder generalizes well, can reconstruct the anomaly well and lead to missed detection. In order to solve this problem, this paper proposes to adopt a memory module to enhance the autoencoder, which is called the Memory-augmented autoencoder (Memory AE) method. Given the input, Memory AE first obtains the encoding from the encoder, and then uses it as a query to retrieve the most relevant memory items for reconstruction. In the training phase, the memory content is updated and encouraged to represent prototype elements of normal data. In the test phase, the learned memory elements are fixed, and reconstruction is obtained from several selected memory records of normal data, thus the reconstruction will tend to be close to normal samples. Therefore, the reconstruction of abnormal errors will be strengthened for abnormal detection. Experiments on two public video anomaly detection datasets, namely Avenue dataset and ShanghaiTech dataset, proves the effectiveness of the proposed method.

Key words: anomalous event detection; video surveillance; auto-encoding network; memory-augmented model; deep learning

收稿日期:2021-09-16; 修订日期:2021-10-12

作者简介:孙敬波,男,讲师,硕士,主要从事软件工程、大数据技术、人工智能方面的研究。

0 引言

作为一种高级计算机视觉任务,视频异常检测指的是自动检测给定视频序列中的异常事件,能够有效区分视频序列中的异常和正常活动以及异常类别。在过去的几年中,研究者们开展了许多异常检测相关的研究^[1-9]。与正常事件相比,很少发生或发生概率低的事件通常被认为是异常。但在实际中,由于未知的事件类型和异常的定义不明确,很难建立有效的异常检测模型。大多数现有的异常检测方法都是基于这样一个假设,即任何与学习到的正常模式不同的模式都被视为异常。在这个假设下,不同场景中的相同活动可能被表示为正常或异常事件。例如,两个人在街头打架的打斗场景可能被认为是异常的,而这两个人在进行拳击运动时则是正常的;一个小孩因为恐慌而在街上奔跑可能被认为是异常的,但如果他忘记带伞,在下雨天奔跑可能是正常的;动物接触人可能被认为是不正常的,而在海豚亲吻人的情况下可能是正常的。此外,高维视频数据中存在大量冗余的视觉信息,这增加了视频序列中事件表示的难度。

根据参考文献[6],异常检测方法通常可以分为两种类型。一种是通过重建误差设计的,专注于对视频序列中的正常模式进行建模^[3-5, 7-8, 10-11]。这类方法的目标是在训练阶段学习正常模式的特征表示模型,在测试阶段利用异常样本和正常样本之间的差异来确定测试数据的最终异常分数,例如重建成本或特定阈值^[7-14]。尽管基于重建的异常检测方法擅长重建视频序列中的正常模式,但关键问题是它们严重依赖于训练数据。另一类方法则将异常检测视为分类问题^[15-18],通过使用已训练的分类器提取诸如光流直方图(histogram of optical flow, HOF)或动态纹理(dynamic texture, DT)之类的特征来预测视频序列的异常分数。这类方法的性能高度依赖于训练样本。为了获得令人满意的性能,提取有效且有判别力的特征对于该类异常检测方法至关重要。然而,这两类方法通常以相对简单的方式对事件之间的相互关系进行建模,例如仅捕获线性关系,这对于许多现实世界设置中复杂、高度非线性的关系是不够的。

近十年前,基于深度学习的方法被应用于视频检测领域,并取得了较大进展。例如,自编码器(autoen-

coders, AE)使用重建误差来检测异常,并且一系列方法在此基础上进行了改进。此外,生成对抗网络(generative adversarial networks, GAN)和长短时记忆(long Short-term Memory, LSTM)网络也被应用于解决异常检测问题。然而,自编码器可能泛化能力较强,导致能够重构异常事件。参考文献[14]指出,由于没有异常训练样本,异常样本的重构是不可预测的,可能会导致异常样本获得更大的重构误差。如果某些异常与正常训练数据共享共同的组成模式(例如图像中的局部边缘),或者解码器“太强”而无法很好地解码某些异常编码,那么自编码器能够很好地重建异常。

为了克服自编码器的不足,文中采用记忆力模块来对深度自动编码器进行增强,引入一种记忆力增强自编码(memory-augmented autoencoder, Memory AE)方法。当输入新测试样本时,Memory AE不会直接将其编码并输入解码器,而是将其用作查询以检索记忆模块中相关的内容,然后将这些内容汇总后传递给解码器,这个过程是通过基于注意力的注意寻址来实现的。进一步,文中提出使用可微的收缩算子来诱导记忆寻址权重的稀疏性,这能够鼓励记忆内容接近特征空间中的查询。在训练阶段,编码器和解码器同时对记忆模块进行更新,以获得较低的平均重构误差。在测试阶段,学习到的记忆内容是固定的,将使用少量的正常记忆项进行重构,选择这些作为输入编码的邻域,重建误差会非常明显。在几个公共基准数据集上的实验表明,Memory AE的性能检测效果达到现有技术发展水平。

1 算法原理

图1为Memory AE的整体网络结构,分三个子结构:编码器(用于编码输入和生成查询)、解码器(用于重建)和记忆模块(具有记忆和相关的寻址运算)。如图1所示,给定待测事件,编码器首先获得其编码值。通过使用编码值作为查询,记忆模块通过基于注意力的寻址算子检索记忆模块中最相关的内容,然后将传递给解码器进行重构。在训练期间,编码器和解码器以最小化重构误差进行参数优化,同时更新记忆模块以记录编码的正常数据的原型元素。给定一个测试样本,该模型仅使用记忆模块中记录的有限的正常模式来执行重建。这样,重建趋于接近正常样本,

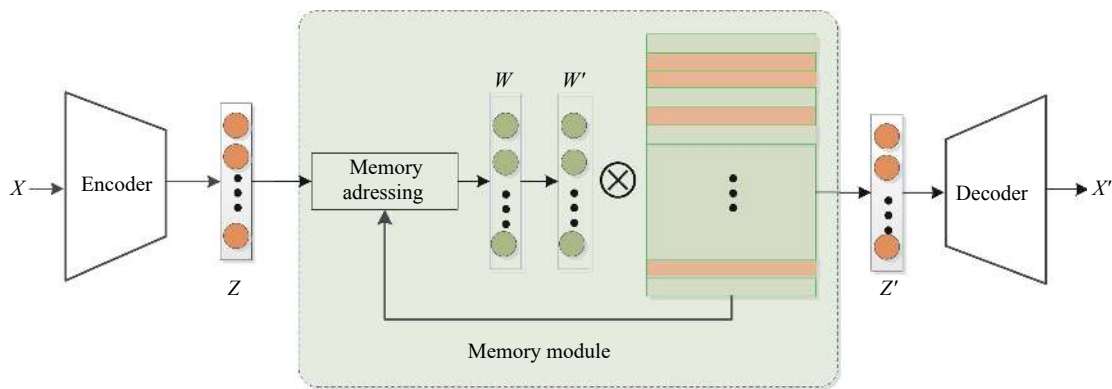


图 1 基于记忆力增强自编码的异常检测方法流程

Fig.1 The flow chart of Memory AE based anomaly detection method

正常样本的重建误差较小, 异常的误差较大。

1.1 编码器和解码器

编码器和解码器分别是自编码器的两部分, 前者将输入数据映射到特征空间获得其编码值, 后者将编码值重构成输入数据。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:

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

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

式中: x 和 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); z 为 x 的编码结果。 w_1 和 w_2 为编码器和解码器的参数, 能够通过最小化输入 x 和重构输入 x' 之间的重构误差获得:

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

通过重建正常样本的误差^[19-20] 判定是否异常, 自编码器已成功用于解决异常检测任务中。然而, 异常样本的重构应该是不可预测的, 可能会导致异常样本获得更大的重构误差。为了解决这个问题, 在下一节中给自编码器引入了记忆模块, 提出了记忆力增强自编码方法。

1.2 记忆力模块

与自编码器不同, 文中提出的方法包含一个记录原型编码模式的记忆力模块和一个用于访问记忆模块的寻址操作。

1.2.1 基于注意力的表示

注意力模块被设计成一个矩阵 $M \in \mathbb{R}^{N \times C}$, 包含了 N 个固定维度的实时向量。为了简便, 假设 C 为 z 的维度, 然后让 $z \in \mathbb{R}^C$ 。给定行向量 m_i , 对于 $\forall i \in [N]$ 表示为其第 i 行, 其中 $[N]$ 表示从 1 到 N 的整数。每个 m_i 表

示一个记忆力项, 给定一组查询 $z \in \mathbb{R}^C$, 记忆力网络获得 \hat{z} 回复一个软地址向量 $w \in \mathbb{R}^{1 \times N}$ 如下:

$$\hat{z} = wM = \sum_{i=1}^N w_i m_i \quad (4)$$

式中: w 为一个行向量, 所有项总和为 1; w_i 表示 w 的第 i 项。权值向量 w 可以通过 z 计算获得。如公式 (4) 所示, 权值 w 需要来接近记忆力模块。混合参数 N 定义为记忆力模块的最大容量。由于 Memory AE 对 N 的设置不敏感, 足够大的 N 能够很好地适用于每个数据集。

1.2.2 用于记忆寻址的注意力

在 Memory AE 中, 记忆模块 M 被设计用于详细地记录训练阶段的原始正常模式。将记忆力模块定义为内容可寻址记忆, 其寻址方案根据查询 z 和记忆力项的权值 w 。如图 1 所示, 可以通过 softmax 操作计算每个权值 w_i :

$$w_i = \frac{\exp(d(z, m_i))}{\sum_{j=1}^N \exp(d(z, m_j))} \quad (5)$$

其中, $d(\cdot, \cdot)$ 表示相似度测度。将其定义为一个余弦相似度:

$$d(z, m_i) = \frac{z m_i^T}{\|z\| \|m_i\|} \quad (6)$$

如公式 (4)~(6) 所示, 记忆模块检索与 z 最相似的记忆项以获得表示。由于记忆大小的限制和稀疏寻址技术, 每次只能寻址少量内记忆项。因此, 记忆模块的有效行为可以解释为: 在训练阶段, Memory AE 中的解码器仅限于使用非常少的寻址记忆项来进行重建, 从而需要有效利用记忆项。而重建时需要迫使记忆模块在输入的正常模式中记录最具代表性的

原型模式;在测试阶段,给定训练好的记忆,只能检索记忆中的正常模式进行重建。因此,正常样本能够得到较好的重建。反之,异常输入的编码值能够被检索到的正常模式代替,从而导致异常样本存在较大的重构错误。

1.3 训练

给定一个包含 T 个样本的数据集 $\{\mathbf{x}^i\}_{i=1}^T$, 让 \mathbf{x}' 表示训练样本 \mathbf{x}^i 的重构样本。首先执行最小化重构:

$$R(\mathbf{x}^i, \mathbf{x}') = \|\mathbf{x}^i - \mathbf{x}'\|_2^2 \quad (7)$$

式中: \mathbf{w}' 表示每个样本 \mathbf{x}^i 的记忆寻址权重; 2 范数用于衡量重构误差。为了进一步促使 \mathbf{w}' 的稀疏性, 在训练时最小化稀疏正则化 \mathbf{w}' 。考虑到所有的 \mathbf{w}' 是非负的且有 $\|\mathbf{w}'\|_1 = 1$, 通过最小化 \mathbf{w}' 的熵以代替最小化 $\|\mathbf{w}'\|_1$:

$$E(\mathbf{w}') = \sum_{j=1}^T -w_j \cdot \log(w_j) \quad (8)$$

通过结合公式 (7)~(8) 的损失函数, Memory AE 的目标函数为:

$$L(\theta_e, \theta_d, M) = \frac{1}{T} \sum_{i=1}^T R(\mathbf{x}^i, \mathbf{x}') + \alpha E(\mathbf{w}') \quad (9)$$

式中: α 为训练过程中的超参数, 用于平衡前后两项。在实际中, α 设定为 0.0002。在训练过程中, 通过反向传播和梯度下降来更新记忆 M 。在反向传播中, 只有具有非零寻址权重 w_i 的记忆项的梯度才能为非零。

1.4 测试

模型训练好以后, 可以通过如下公式计算第 t 帧 (x, y) 像素位置的重构误差:

$$e(x, y, t) = \|I(x, y, t) - h(I(x, y, t))\|_2 \quad (10)$$

式中: $h(\cdot)$ 表示整个模型。给定第 t 帧的像素级重构误差, 可以通过求和获得整帧图像的重构误差 $e(t) = \sum_{(x,y)} e(x, y, t)$ 。那么, 可以最后计算出该帧的异常得分:

$$s(t) = 1 - \frac{e(t) - \min_e e(t)}{\max_e e(t)} \quad (11)$$

然后设定门限值判断其是否为异常:

$$s(t) > \theta \quad (12)$$

2 实验

验证提出的方法有效性, 并与现有其他方法进行对比。目前, 采用了两个公开数据集进行实验, 即 Avenue 数据集和 ShanghaiTech 数据集, 并且采用帧

级 AUC 和 EER 作为量化评价指标。

2.1 实验数据及评价指标

Avenue 数据集使用分辨率为 640×360 pixel 的固定摄像头捕捉和记录香港城市大学的街头活动。该数据集有 16 个包含正常人类行为的训练视频片段以及 21 个包含异常事件和人类行为的测试视频片段, 共有 30652 帧, 所有测试视频都有目标级的标注, 即用一个矩形区域来标记空间位置的异常。正常行为是人行道上行走的人, 而异常事件是人们乱扔垃圾/丢弃物品、徘徊、走向摄像机、在草地上行走和丢弃物体。

ShanghaiTech 数据集对于异常事件检测来说是一个非常具有挑战性的集合。与其他数据集不同, 它包含 13 个具有不同光照条件和摄像机角度的场景, 共有 330 个训练视频和 107 个测试视频。测试集共包含 130 个具有像素级标注的异常事件。整个数据集共有 316154 帧, 其中训练集 274515 帧, 测试集 42883 帧, 异常帧共 17090 帧, 视频帧的分辨率为 480×856 pixel。

该模型在 NVIDIA GTX1080TI、显存 8 GB 的平台上进行实验, 软件环境为 PyTorch 和 Python 3.6。为了衡量文中提出的用于视频异常检测方法的有效性, 采用了帧级接收者操作特征曲线 (receiver operating characteristic, ROC) 的曲线下面积 (area under curve, AUC) 作为评价指标。对于帧级评价指标来说, 如果一个帧的至少一个像素被标记为异常, 则该帧被认为是异常的, 帧级 AUC 是通过将帧级检测结果与帧级真实标签进行比较计算得到的。

2.2 实验设置

所有视频帧都被调整为 227×227 , 然后转换为灰度图像。模型的输入是 $227 \times 227 \times 5$, 即采用了连续 5 帧作为模型的输入。编码器的结构采用常见的 $C128 \times (3 \times 3) - C64 \times (3 \times 3) - C64 \times (3 \times 3) - C128 \times (4 \times 4)$ 的结构, 即首先为 128 个大小 3×3 卷积核的卷积层, 接下来采用 64 个大小 3×3 卷积核的卷积层, 然后采用 64 个大小 3×3 卷积核的卷积层和 128 个大小 3×3 卷积核的卷积层, 可以获得 64 维的向量 z_c 。每个卷积层后有一个 Batch Normalization 层和一个 ReLU 激励层。解码器则采用完全相反的结构, 包含有 4 个反卷积层。注意力模块设置为 $N = 1000$, 考虑到视频数据

的复杂性,让每个记忆片段记录特征图中的一个像素上的特征,对应视频片段的一个子区域。因此,记忆模块是一个 1000×64 的矩阵。整个模型参数优化选取的是 Adam 优化器,初始学习率为 0.0001,迭代次数为 1000。动量 (momentum) 参数为 $\rho_1 = 0.9, \rho_2 = 0.999$,批尺寸为 128。

2.3 实验结果

为了证明提出的方法在视频异常检测方面的有效性,将其与现有的 12 种不同方法进行了比较。其中, MPPCA(概率主成分分析器的混合)+SF(社会力)^[17]、MDT(动态纹理的混合)^[18] 是基于手工特征的方法; Conv-AE^[8]、3D Conv^[19]、Stacked RNN^[20] 和 ConvLSTM-AE^[21]、MemNormality^[22] 和 Cluster-AE^[23] 都是基于自编码器的方法; AbnormalGAN^[24]、Pred+Recon^[25] 是基于生成对抗网络的方法。

表 1 显示了各种方法的帧级视频异常检测结果,可以观察到,在两个数据集的结果中,基于自编码器的方法通常优于基于手工制作的特征的方法,取得了更高的帧级 AUC。这是因为手工制作的特征通常是基于其他任务而提取的,可能是次优的。在基于自编码器的方法中, ConvLSTM-AE 优于 Conv-AE,这是因为前者可以更好地捕获时间信息。此外,基于深度对抗网络的方法比大多数基线方法表现更好。所提出的 Memory AE 方法在 Avenue 数据集上取得了 85.7% 的帧级 AUC,比效果最好的 Pred+Recon^[25] 方法领先 0.6%; 而提出的方法在 ShanghaiTech 数据集上取得了 75.3% 的帧级 AUC,比其他方法在帧级 AUC 上领先 2% 以上,效果非常明显。这主要是因为提出的基于 Memory AE 方法采用了记忆片段,能够很好地重建异常并引入一些随机错误。此外,相对于 Avenue 数据集,在 ShanghaiTech 数据集上取得了较高的帧级 AUC,主要是因为 ShanghaiTech 数据集包含多场景及此前其他数据集中未出现的异常事件,更为复杂。为了验证 ShanghaiTech 数据集上单场景的

检测结果,采用了单场景视频片段进行训练和测试,采用了 83 个片段 (25%) 用于训练, 34 个片段 (32%) 用于测试,取得了 86.3% 的帧级 AUC,达到了和 Avenue 数据集中近似的水平。综上可知,提出的于 Memory AE 方法可以灵活地应用于不同类型的数。仅通过使用重构误差,所提出的方法可以以最少的特定知识获得较好的结果。

表 1 同现有技术水平方法比较结果 (以帧级 AUC% 的形式)

Tab.1 Comparison with the state of the art methods in terms of AUC%

Method	Avenue	ShanghaiTech
MPPCA+ SF ^[17]	56.2%	-
MDT ^[18]	77.4%	-
Conv-AE ^[8]	80.0%	60.9%
Conv3D-AE ^[19]	80.9%	-
Stacked RNN ^[20]	81.7%	68.0%
ConvLSTM-AE ^[21]	77.0%	-
MemNormality ^[22]	88.5%	70.5%
ClusterAE ^[23]	86.0%	73.3%
AbnormalGAN ^[24]	-	72.4%
Pred+Recon ^[25]	85.1%	73.0%
Proposed method	85.7%	75.3%

为了评估预定义记忆模块在检测视频异常事件方面的性能,通过改变记忆模块大小并在 Avenue 数据集上进行实验,并在表 2 中给出了帧级 AUC 值。可以发现给定足够大的记忆模块大小, Memory AE 方法可以稳健地产生最优的结果。当记忆模块大小大于 1000 后,对检测结果影响较小,但是采用较大的记忆模块大小会导致更大的计算量,因此选用记忆模块大小为 1000。

图 2(a)、(b) 分别展示了提出方法在 Avenue 数据集和 ShanghaiTech 数据集中的一些视频片段上估计

表 2 记忆模块大小对于 Avenue 数据集实验结果 (帧级 AUC%) 的影响

Tab.2 The influence of the number of memory module size on the experimental results of the Avenue data set (frame-level AUC%)

Size of memory module	500	1000	1500	2000	2500
Result	78.2%	85.7%	85.3%	85.7%	85.8%

的帧级异常分数。图中的蓝线描绘了帧级异常得分，绿色和红色阴影段分别表示地面真实正常和异常视频段。绿色框中的帧是来自常规视频片段的正常帧，红色框中的帧是来自异常视频片段的异常帧。一些异常事件如丢纸屑、人行道上骑自行车、殴打等能够被检测出。

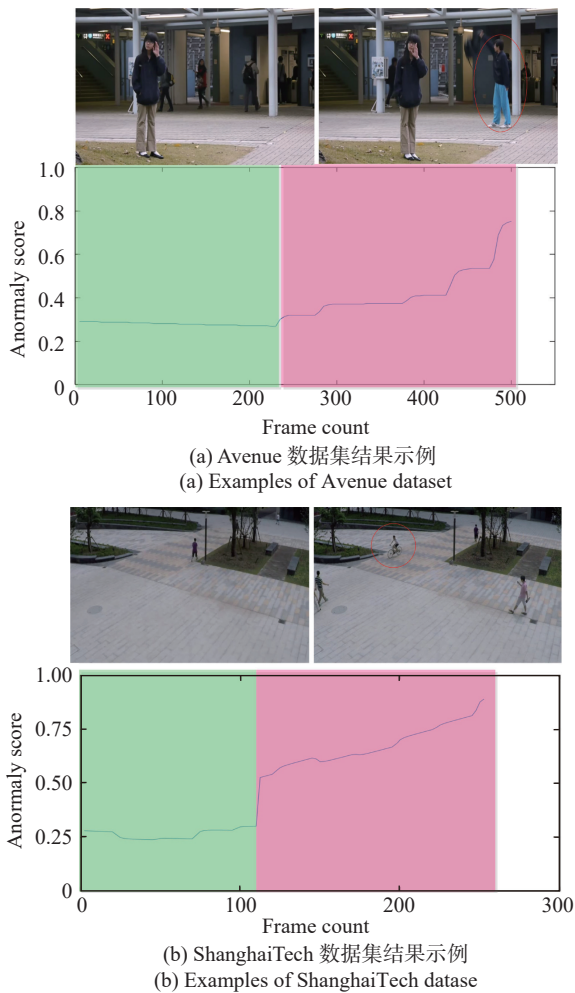


图 2 部分检测结果示例

Fig.2 Examples of the detection results

3 结 论

文中提出了一种记忆力增强自编码器 (Memory AE) 来提高基于自编码器的异常检测方法的性能。给定输入，提出的 Memory AE 方法首先使用编码器获得编码表示，然后使用编码作为查询来检索记忆模块中最相关的模式以进行重建。由于记忆模块被训练来记录典型的正常模式，所提出的 Memory AE 可以很好地重建正常样本并扩大异常的重建误差，加强

了重建误差作为异常检测标准的作用。在两个数据集上的实验证明了所提出方法的通用性和有效性。将来会研究使用寻址权重进行异常检测，考虑到所提出的内存模块是通用的，并且与编码器和解码器的结构无关，将其集成到更复杂的基础模型中，并将其在更具挑战性的数据集集中进行实验。

参考文献：

- [1] Dong F, Zhang Y, Nie X. Dual discriminator generative adversarial network for video anomaly detection[J]. *IEEE Access*, 2020, 8: 88170-88176.
- [2] Doshi K, Yilmaz Y. Any-shot sequential anomaly detection in surveillance videos [C]//Proceedings of International Conference on Computer Vision/Pattern Recognition workshop, 2020: 934-935.
- [3] Doshi K, Yilmaz Y. Continual learning for anomaly detection in surveillance videos [C]//Proceedings of International Conference on Computer Vision/Pattern Recognition Workshop, 2020: 254-255.
- [4] Dutta J K, Banerjee B. Online detection of abnormal events using incremental coding length [C]//AAAI Conference on Artificial Intelligence, 2015: 3755-3761.
- [5] Feng Y, Yuan Y, Lu X. Learning deep event models for crowd anomaly detection[J]. *Neurocomputing*, 2017, 219: 548-556.
- [6] Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection [C]//IEEE International Conference on Computer Vision, 2017: 1705-1714.
- [7] Ravanbakhsh M, Nabi M, Sangineto E, et al. Abnormal event detection in videos using generative adversarial nets [C]//IEEE International Conference on Image Processing, 2017: 1577-1581.
- [8] Hasan M, Choi J, Neumann J, et al. Learning temporal regularity in video sequences [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 733-742.
- [9] Liu W, Luo W, Lian D, et al. Future frame prediction for anomaly detection—A new baseline [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6536-6545.
- [10] Park H, Noh J, Ham B. Learning memory-guided normality for anomaly detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2020, 14: 372-381.
- [11] Zaheer M Z, Lee J-H, Astrid M, et al. Old is gold: Redefining the adversarially learned one-class classifier training paradigm [C]//IEEE Conference on Computer Vision and Pattern

- Recognition, 2020, 14: 183-193.
- [12] Zhu X, Liu J, Wang J, et al. Anomaly detection in crowded scene via appearance and dynamics joint modeling [C]//IEEE International Conference on Image Processing, 2013: 2705-2708.
- [13] Colque R V M, Júnior C A C, Schwartz W R. Histograms of optical flow orientation and magnitude to detect anomalous events in videos [C]//Sibgrapi Conference on Graphics, Patterns and Images, 2015: 126-133.
- [14] Zong B, Song Qi, Min M, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection [C]//International Conference on Learning Representations, 2018.
- [15] Sabokrou M, Fayyaz M, Fathy M, et al. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes [C]//Computer Vision and Image Understanding, 2018, 172: 88-97.
- [16] Li C, Han Z, Ye Q, et al. Visual abnormal behavior detection based on trajectory sparse reconstruction analysis [J]. *Neurocomputing*, 2013, 119(7): 94-100.
- [17] Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010: 1975-1981.
- [18] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(1): 18-32.
- [19] Zhao Y, Deng B, Shen C, et al. Spatio-temporal autoencoder for video anomaly detection [C]//ACM International Conference on Multimedia, 2017: 1933-1941.
- [20] 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: 1-8.
- [21] Luo W, Liu W, Gao S. Remembering history with convolutional LSTM for anomaly detection [C]//IEEE International Conference on Multimedia and Expo (ICME), 2017: 439-444.
- [22] Park H, Noh J, Ham B. Learning memoryguided normality for anomaly detection [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 14372-14381.
- [23] Chang Y, Tu Z, Xie W, Yuan J. Clustering driven deep autoencoder for video anomaly detection [C]//European Conference on Computer Vision (ECCV), 2020: 329-345.
- [24] Ravanbakhsh M, Nabi M, Sangineto E, et al. Abnormal event detection in videos using generative adversarial nets [C]//IEEE International Conference on Image Processing (ICIP), 2017: 1577-1581.
- [25] Liu W, Luo W, Lian D, et al. Future frame prediction for anomaly detection—A new baseline [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6536-6545.