基于深度自编码-高斯混合模型的视频异常检测方法

钟友坤[1]，莫海宁[2*]

(1. 河池学院 物理与机电工程学院，广西 宜州 546300；
2. 广西科技大学 宏达威爱科技学院，广西 柳州 545006)

摘　要：由于异常定义的模糊性和真实数据的复杂性,视频异常检测是智能视频监控中最具挑战性的问题之一。基于自动编码器 (AE) 的帧重建 (当前或未来帧) 是一种流行的视频异常检测方法。使用在正常数据上训练的模型,异常场景的重建误差通常比正常场景的重建误差大得多。但是,这类方法忽略了正常数据本身的内部结构,效率较低。基于此,提出了一种深度自动编码高斯混合模型 (DAGMM)。首先利用深度自动编码器获得输入视频片段的生成低维表示和重构误差,并将其进一步输入高斯混合模型 (GMM)。而估计网络则通过高斯混合模型预测能量概率,然后通过能量密度概率判断异常。DAGMM 以端到端的方式同时联合优化深度自动编码器和 GMM 的参数,能够平衡自动编码重建、低维表示的密度估计和正则化,泛化能力强。在两个公共基准数据集上的实验结果表明,DAGMM 达到了现有最高技术发展水平,在 UCSD Ped2 和 ShanghaiTech 两个数据集上分别取得了 95.7% 和 72.9% 的帧级 AUC。

关键词：视频监控；　异常事件；　自编码网络；　高斯混合模型；　深度学习

中图分类号：TP391.4　　**文献标志码：**A　　**DOI：**10.3788/IRLA20210547

A video anomaly detection method based on deep autoencoding Gaussian mixture model

Zhong Youkun[1]，Mo Haining[2*]

(1. Physics and Mechanical & Electrical Engineering School, Hechi University, Yizhou 546300, China;
2. HTC VIVEDU School of Technology, Guangxi University of Science and Technology, Liuzhou 545006, China)

Abstract: Due to the vagueness of anomaly definition and the complexity of real data, video anomaly detection is one of the most challenging problems in intelligent video surveillance. Frame reconstruction (current or future frame) based on autoencoder (AE) is a popular video anomaly detection method. Using a model trained on normal data, the reconstruction error of abnormal scenes is usually much larger than that of normal scenes. However, these methods ignore the internal structure of the normal data and are memory-consuming. Based on this, a deep auto-encoding Gaussian mixture model (DAGMM) was proposed. Firstly, the deep autoencoder was used to obtain the low-dimensional representation of the input video segment and the reconstruction error, and then further input into a Gaussian mixture model (GMM). The energy probability was predicted through the Gaussian

收稿日期:2021−08−07；　**修订日期:**2021−08−21

基金项目:国家自然科学基金 (61662007)

作者简介:钟友坤,男,高级实验师,硕士,主要从事计算机及其机电设计应用方面的研究。

通讯作者:莫海宁,男,副教授,硕士,主要从事数据挖掘与信息管理方面的研究。

mixture model, and then the anomaly was judged through the energy density probability. The proposed DAGMM can simultaneously optimizes the parameters of the deep autoencoder and GMM in an end-to-end manner, and balance auto-encoding reconstruction, density estimation and regularization of low-dimensional representation, and has strong generalization ability. Experimental results on two public benchmark datasets show that DAGMM has reached the highest level of technological development, achieving 95.7% and 72.9% frame-level AUC on the UCSD Ped2 and ShanghaiTech dataset, respectively.

Key words: video surveillance; anomalous event; auto-encoding network; Gaussian mixture model; deep learning

0 引 言

视频监控系统越来越多地出现在各种公共场景和私人场所中,以监控人类活动并防止犯罪发生。毫无疑问,这需要有人观看监控视频,并在发生与正常情况不同的事情时进行判断并发出警报。然而,这些异常事件并不经常发生,因此大多数时候监控这些视频的人不会看到任何异常行为。这些不寻常的事件可以被认为是异常,可以将其定义为不符合正常情况的模式,发现这些不符合正常模式的任务称为异常检测。基于此,研究人员一直在尝试设计一种强大的异常检测算法,以自动监控和检测监控视频中的异常事件。

异常检测是一项具有挑战性的任务[1]:首先,异常事件的定义往往取决于当时的环境,很难准确地区分正常事件和异常事件。其次,构成异常的不同可能性是无限的。第三,异常数据点,尤其是真实世界的数据,往往与可能被定义为正常的数据点非常接近。这些原因导致异常检测任务十分困难,是过去几年研究人员在提出新解决方案时一直在考虑的问题。

近十年前,大多数研究人员都专注于基于轨迹的异常检测[2−3]。主要思想是:如果感兴趣的对象没有符合学习到的正常轨迹模式,视频将被标记为异常。然而,这种方法的一个主要缺点是遮挡,因为该方法严重依赖于持续检测跟踪感兴趣的对象。由于这些缺点,研究者们开始采用底层特征进行特征提取。这些基于低级特征的方法依赖于外观、运动和纹理特征的使用[4−6]。大量的方法已经使用了各种底层运动特征表示来表示视频,如社会力模型、光流直方图等,但是这些仅基于运动的特征是不够充分的。动态纹理、描述空间和运动的光流特征、光流空间局部直方图和基于均匀局部梯度模式的光流等特征被提出[6−7]。

尽管这些传统方法在基准数据集上取得了成功,但泛化能力较差,在其他场景中使用时它们仍然无效。此外,它们无法适应以前从未见过的异常。由于这些原因,研究人员探索使用深度神经网络来完成异常检测任务。这些神经网络能够自动学习有用的判别特征,从而消除了创建手工特征的麻烦,这也使其在用于不同场景时更具适应性。深度学习被证明对各种计算机视觉任务有效[8−9],例如图像中的特征提取、图像分类、对象检测、视频分析和许多其他任务。深度学习技术主要侧重于创建新网络结构或设计适合特定问题的组件。现有的基于深度学习的视频异常检测方法可以分为四类:(1) 基于重构的方法[10−11]:这类方法假设是正常样本的重构误差会更低,因为它们更接近训练数据,而对于不正常的样本,假设或预期重建误差会更高。这类方法往往基于自编码,它能够将输入编码为更紧凑的表示的同时保留重要的判别特征,并且还能够将该特定编码解码回其原始形式。(2) 基于预测未来帧的方法[12−13]:这类方法主要是通过对基于现有帧对未来帧进行预测,看其是否符合现有帧的模式进行异常判断。这类方法基于生成对抗网络,它包含生成器和鉴别器两个网络,前者能够模拟原始数据分布,后者则给出输入是否来自生成器的概率。(3) 基于分类的方法[14−15]:这类问题可以看成对一段视频片段进行直接分类,给出其正常或是异常类别。由于正负例训练样本不在一个数量级,这类方法集中于利用卷积神经网络创建紧凑、高效且鲁棒的特征。(4) 基于异常得分的方法[16−23]:将问题定义为回归问题,其中目标是提供异常分数,然后将其用作确定视频片段或帧是否异常的手段。

与这些方法不同的是，文中基于深度自编码高斯混合模型 (deep autoencoding Gaussian mixture model, DAGMM)，提出一种新的异常检测方法。DAGMM 包含一个压缩网络和一个估计网络：压缩网络通过深度自动编码器对输入视频片段进行降维，根据低维特征和重构误差特征输入到估计网络中；而估计网络则通过高斯混合模型预测能量概率，然后通过能量密度概率判断异常。通过同时最小化来自压缩网络的重建误差和来自估计网络的样本能量，可以联合训练一个降维结构，直接帮助目标密度估计任务。与之前方法不同的是，文中方法能够同时对事件表示 (压缩网络) 和异常检测模型 (估计网络) 进行联合优化，泛化能力强。在几个公共基准数据集上的实验表明，DAGMM 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 DAGMM 的整体网络结构，分两个子结构，左边部分为压缩网络，是一个深度全卷积自编码网络，通过这个自编码可以得到输入视频块$\boldsymbol{x}$的低维表示z_r，同时得到输入$\boldsymbol{x}$与重构的$\boldsymbol{x}'$之间的重构误差z_c，然后进行拼接操作形成$\boldsymbol{z}$；右边为估计网络，也是一个多层的全连接前馈神经网络，输入为$\boldsymbol{z}$，经过多层全连接得到一个概率分布，这个概率分布的长度即为混合高斯分布中高斯成分的个数。那么可以通过这个输出概率判断输入是否为异常。

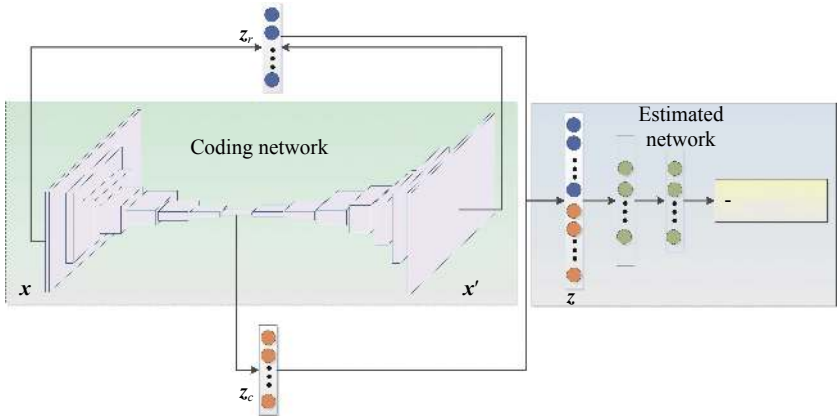


图 1 基于深度自编码-高斯混合模型的异常检测方法流程

Fig.1 Flow chart of abnormal event detection method based on DAGMM

1.1 压缩网络原理

如图 1 所示，压缩网络提供的低维表示包含两个特征来源：(1) 深度自编码器学习到的低维表示；(2) 由重构误差导出的特征。给定一个样本$\boldsymbol{x}$，压缩网络计算其低维表示$\boldsymbol{z}$如下：

$$z_c = f_{w_1}(\boldsymbol{x}) \tag{1}$$

$$\boldsymbol{x}' = g_{w_2}(z_c) \tag{2}$$

$$z_r = h(\boldsymbol{x}, \boldsymbol{x}') \tag{3}$$

$$\boldsymbol{z} = [z_c, z_r] \tag{4}$$

式中：$\boldsymbol{x}$和$\boldsymbol{x}'$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_r为$\boldsymbol{x}$的隐层表示；w_1和w_2为自编码网络中编码器和解码器的参数；$h(\cdot,\cdot)$表示计算重构误差的函数。然后，获得的特征$\boldsymbol{z}$被传入估计网络中。

1.2 估计网络原理

给定输入样本的低维表示$\boldsymbol{z}$，估计网络在高斯混合模型 (Gaussian mixture model, GMM) 的框架下进行密度估计。与传统 GMM 不同的是，估计网络在训练阶段无需采用期望最大化 (expectation-maximization, EM) 算法中的交替迭代的方法，通过未知混合分量分布ϕ、混合均值μ和混合协方差Σ即可直接估计 GMM 的参数并评估样本的似然。为了实现这一点，估计网络通过利用多层神经网络来预测每个样本的混合成分。给定特征$\boldsymbol{z}$和混合成分的个数特征K(K的大小将在 2.3 节中进行讨论)，则样本属于高斯混合分布中各个分布的概率为：

$$\boldsymbol{p} = MLN(\boldsymbol{z}; \theta_m) \tag{5}$$

$$\hat{\gamma} = \mathrm{softmax}(\boldsymbol{p}) \tag{6}$$

式中：$\hat{\gamma}$表示预测结果，是个K维的向量；$\boldsymbol{p}$为多层感知器的输出。给定N个样本和预测结果，那么可以估计出 GMM 的参数如下：

$$\hat{\phi}_k = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N} \tag{7}$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} z_i}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \tag{8}$$

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} \left(z_i - \hat{\mu}_k\right)\left(z_i - \hat{\mu}_k\right)^{\mathrm{T}}}{\sum_{i=1}^{N} \hat{\gamma}_{ik}} \tag{9}$$

式中：$\hat{\gamma}_i$为隐层表示z_i属于第i个分量的概率；$\hat{\phi}_k, \hat{\mu}_k, \hat{\Sigma}_k$分别表示 GMM 中第$k$个成分的混合概率、均值和方差。那么，样本的概率分布可以表示为：

$$E(z) = -\log\left(\sum_{k=1}^{k} \hat{\phi}_k \frac{\exp\left(-\frac{1}{2}\left(z_i - \hat{\mu}_k\right)^{\mathrm{T}} \Sigma_k^{-1} \left(z_i - \hat{\mu}_k\right)\right)}{\sqrt{\left|2\pi\hat{\Sigma}_k\right|}}\right) \tag{10}$$

其中，$|\cdot|$表示矩阵的行列式。

1.3 目标函数

基于以上压缩网络和估计网络的基本原理，DAGMM 的的目标函数可以表示为：

$$J(W_1, W_2, \theta_m) = \frac{1}{N}\sum_{i=1}^{N} L(\boldsymbol{x}_i, \boldsymbol{x}'_i) + \frac{\lambda_1}{N}\sum_{i=1}^{N} E(z_i) + \lambda_2 P\left(\hat{\Sigma}\right) \tag{11}$$

式中：N为训练样本个数；λ_1和λ_2为用于平衡三项的参数。目标函数共包含三项，第一项中$L(\boldsymbol{x}_i, \boldsymbol{x}'_i)$为压缩网络中深度自编码器造成的重构误差的损失函数。如果压缩网络能够使重构误差较低，那么低维表示可以更好地保留输入样本的关键信息；第二项$E(z_i)$是可以观察到输入样本的概率。通过最小化样本概率，可以通过寻找压缩和估计网络的最佳组合，以最大化观察输入样本的可能性；第三项是惩罚项，防止矩阵不可逆。

1.4 预测

对于给定的测试样本y，如果其为正常样本，那么一定能够和某个 (某几个) 高斯分量相关联，因此其隐层表示属于某个高斯分量的条件概率一定会相对比较大。反之，异常样本无法与任何一个高斯分量相关联，那么其属于所有高斯分量的条件概率都很小。

在预测阶段，首先根据公式 (1)、(3)、(4) 获得其隐层表示z_y，然后通过公式 (10) 计算隐层表示z_y的能量，并设定门限值ξ判断其是否为异常：

$$E(z_y) > \xi \tag{12}$$

2 实 验

2.1 实验数据及评价指标

为了验证文中提出的方法，在两个公开可用的视频异常数据集上评估提出的方法，即 UCSD PED1 行人数据集[4] 和 ShanghaiTech[18] 校园数据集。这两个数据集都有自己的挑战和独特的特征，例如异常类型、视频质量、背景位置等，两个数据集的简要介绍如表 1 所示。因此，需要对两个数据集分别训练模型并进行测试。

UCSD 行人异常检测数据集由 Mahadevan 等人创建的，目的是评估他们的异常检测方法。该数据集包含由安装在高处的固定相机以 10 帧/s 的速度拍摄的俯瞰人行道的视频。在这个数据集中，异常事件包括人行道上的非行人和异常的行人运动，具体来说，一些异常示例包括骑自行车的人、溜冰者、猫等。该数据集有两个子集，其中每个子集对应于特定场景。文中仅采用第一个场景即 UCSD PED2 进行实验，包含 16 个训练片段和 14 个测试片段，共 4 560 帧，分辨率为 320×240。

ShanghaiTech 数据集的提出是由于现有基准数据集缺乏场景多样性。与之前的数据集相比，ShanghaiTech 数据集的视频数量更多，总共有 330 个训练视频和 107 个测试视频，包括 13 个不同的场景和大量不同的异常类型。该数据集中视频的分辨率为 856×480。此外，异常事件包括人行道上自行车、追逐和争吵等突然运动引起的异常。

帧级评价指标被用于评估检测方法的性能。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。为了使用帧级标准进行评估，时间标签用于确定度量标准的真阳性和假阳性，并通过公式 (12) 和公式 (13) 来计算方法的

检测率 (true positive rate, TPR) 和虚警率 (false positive rate, FPR):

$$TPR = \frac{TP}{TP+FN} \tag{13}$$

$$FPR = \frac{FP}{FP+TN} \tag{14}$$

对于这两个标准，计算接收者操作特征曲线 (ROC) 的曲线下面积 (AUC) 以衡量模型的最终性能。给定具有不同阈值的分类模型，接收者操作特征曲线 (ROC) 说明了模型的性能。

表 1 基准数据集概述

Tab.1 Overview of benchmark datasets

Attributes	UCSD Ped2	ShanghaiTech
Frames	4 560	317 398
Scene	Single	Multi
Labels	Spatial & Temporal	Spatial & Temporal
Resolution	360×240	856×480
Anomalies	Biker, cart, etc	Chasing, brawlingsudden motion, etc

2.2 实验设置

对于两个数据集，每帧都被调整为大小 420×280，且转换为灰度图。网络的输入为 28×28×3 长方体，意味着 420×280×3 的连续 3 帧视频可以划分为 15×10 的长方体。压缩网络中编码器的结构类似于参考文献 [19] 中的 Model C。CNN 的结构采用常见的 C64×(3×3)−C128×(3×3)−C256×(3×3)−C64×(4×4) 的结构，即首先为 64 个大小 3×3 卷积核的卷积层，接下来采用 128 个大小 3×3 卷积核的卷积层，然后采用 256 个大小 3×3 卷积核的卷积层和 256 个大小 4×4 卷积核的卷积层，可以获得 256 维的向量z_c。解码器的网络结构与编码器相反，而重构向量z_r为 2 352 维，因此预测网络的输入为 2 608 维，其中全连接层网络节点数目分别设置为 500 和 50，输出层为 softmax 层，所有的激励函数均设置为 tanh。公式 (11) 中λ_1和λ_2分别设置为 0.1 和 0.01，公式 (5) 中K设置为 16。整个模型参数优化选取的是 Adam 优化器，初始学习率为 0.000 1，迭代次数为 1 000。动量 (momentum) 参数为 $\rho_1 = 0.9, \rho_2 = 0.999$，批尺寸为 128。实验硬件平台为 NVIDIA GTX1070TI，显存 8 GB，软件环境为 Tensorflow 1.15 和 Python 3.6。

2.3 实验结果

为了验证文中提出方法的优势，将所提出的方法同十余种方法进行了对比。这些方法包含采用手工特征的方法，如 MPPCA[3]、动态纹理 (MDT)[4]、MT-FRCN[5] 等；也包括采用深度学习特征的方法，如包括 2D 卷积自动编码器方法 (Conv2D-AE)[10]、3D 卷积自动编码器方法 (Conv3D-AE)[10]、基于卷积长短时记忆网络的自动编码器方法 (ConvLSTM-AE)[20]、堆叠循环神经网络 (StackRNN)[21] 和基于生成对抗网络的方法 (GAN)[18]。

表 2 以帧级 AUC 的形式给出了两个数据集的检测结果。通过表 1 可以看出，文中提出的方法优于其他对比方法。与基于手工特征的方法相比，所提出方法的结果在 UCSD Ped2 数据集上的准确率至少提高了 3.5% 帧级 AUC(95.7% vs 92.2%)。值得注意的是，由于 ShanghaiTech 是近几年提出的新数据集，帧数较多，对于计算需求也比较大，目前为止没有基于手工特征的方法在该数据集上进行验证。与同深度学习的方法相比，文中提出的方法在两个数据集上取得了最佳的检测结果。具体来说，提出的方法在 UCSD Ped2 数据集和 ShanghaiTech 数据集上分别比 Baseline[18] 方法好 0.3% 和 0.1% 帧级 AUC。但是，提出的方法在 ShanghaiTech 数据集上取得 72.9% 帧级 AUC，相对

表 2 与现有技术发展水平检测方法结果对比 (以 AUC% 的形式)

Tab.2 Comparison with the state of the art methods in terms of AUC%

Method	UCSD Ped2	ShanghaiTech
MPPCA[3]	69.3%	-
MDT[4]	82.9%	-
MT-FRCN[5]	92.2%	-
Conv2D-AE[10]	85.0%	60.9%
Conv3D-AE[10]	91.2%	-
ConvLSTM-AE[20]	88.1%	-
StackRNN[21]	92.2%	68.0%
Baseline[18]	95.4%	72.8%
Proposed method	95.7%	72.9%

于在 UCSD Ped2 数据集取得的 95.7% 帧级 AUC 来说要低很多，这主要是因为 ShanghaiTech数据集相对于 UCSD Ped2 数据集更为复杂，包含多场景、多帧数、以及此前其他数据集中未出现的异常事件。

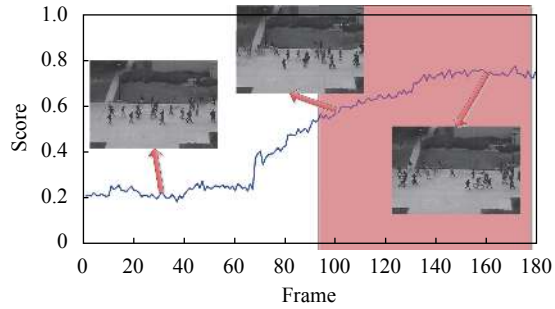
为了评估公式 (5) 中预先定义了高斯混合分量个数K在检测视频异常事件方面的性能，通过改变高斯混合分量个数K并在 UCSD Ped2 数据集上进行实验，实验结果以帧级 AUC 的形式给出。表 3 为在 UCSD Ped2 数据集上的检测结果。可以看出，当 K<16 时，检测结果随着K值增大而提升，这主要是因为较小的K值导致原本区别很大的样本被聚集在同一个高斯混合分量下，损失了较多的局部细节信息；而当K > 16 时，检测性能不再随着K值增大而变化。但是，根据公式 (7)~(9)，采用较大的K值会导致更大的计算量。

表 3 高斯混合分量个数 K 对于 UCSD Ped2 数据集实验结果 (帧级 AUC%) 的影响

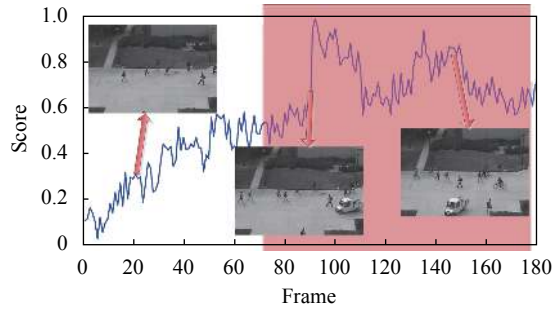
Tab.3 Influence of the number of Gaussian mixture components number K on the experimental results of the UCSD Ped2 data set (frame-level AUC%)

K	AUC%
2	92.3%
4	94.5%
8	951%
16	95.7%
32	95.6%
64	95.7%

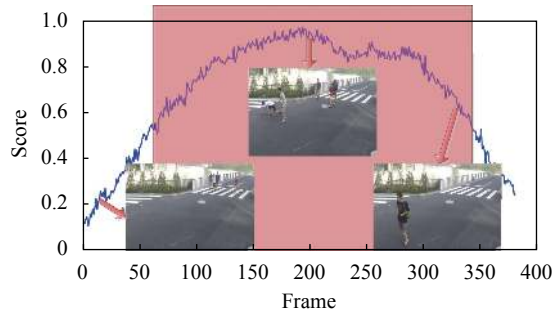
图 2(a)~(d) 分别展示了文中提出的方法在 UCSD Ped2 数据集和 ShanghaiTech 数据集中的一些视频片段上估计的帧级异常分数。其中，横坐标为视频帧数，纵坐标异常分数已经归一化到 1，红色区域表示真实注释的异常帧。通过图 2 可以看出，异常得分较大的区域和真实标注的异常帧能够基本吻合上，一些异常事件如人行道上出现自行车、小车，打架推搡基本能够检测出。此外，图 2 还提供了一些具有正常或异常事件的关键帧。当异常事件突发，如图 2(b) 场景中出现小车时，异常得分突然增加；相反，如果导致异常的对象在摄像头视野中消失，如图 2(c) 所示，则异常得分会降低。



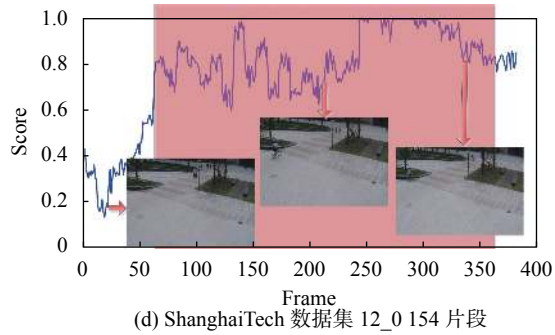
(a) UCSD Ped2 数据集片段 1
(a) UCSD Ped2 dataset clip 1



(b) UCSD Ped2 数据集片段 4
(b) UCSD Ped2 dataset clip 4



(c) ShanghaiTech 数据集 07_0 006 片段
(c) ShanghaiTech dataset clip 07_0 006



(d) ShanghaiTech 数据集 12_0 154 片段
(d) ShanghaiTech dataset clip 12_0 154

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 DAGMM 网络，结合了深度自编码和高斯混合模型，用于监控视频中的异常检测。

DAGMM 由两个主要部分组成：压缩网络和估计网络，其中压缩网络将样本投影到低维空间，保留异常检测的关键信息，估计网络在框架下评估低维空间中的样本能量高斯混合建模。DAGMM 能够实现端对端训练，估计网络预测样本混合隶属度，从而无需交替程序即可估计 GMM 中的参数，估计网络引入的正则化有助于压缩网络摆脱吸引力较小的局部最优，并通过端到端训练实现低重构误差。在两个数据集上的实验证明了文中提出的方法与最先进的方法相比具有竞争力。

参考文献：

[1] 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.

[2] 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.

[3] Jiang F, Yuan J, Tsaftaris S A, et al. Anomalous video event detection using spatiotemporal context [J]. ***Computer Vision and Image Understanding***, 2011, 115(3): 323-333.

[4] Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scene [J]. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, 2014, 36(1): 18-32.

[5] Reddy V, Sanderson C, Lovell B. Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2011: 55-61.

[6] Wang S, Zhu E, Yin J, et al. Video anomaly detection and localization by local motion based joint video representation and OCELM [J]. ***Neurocomputing***, 2018, 277: 161-175.

[7] Kaur P, Gangadharappa M, Gautam S. An overview of anomaly detection in video surveillance [C]//International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018.

[8] Schmidhuber J. Deep learning in neural networks: An overview [J]. ***Neural Networks***, 2015, 61: 326-366.

[9] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. ***Nature***, 2015, 521: 436-444.

[10] 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.

[11] Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1-8.

[12] Ravanbakhsh M, Sangineto E, Nabi M, et al. Abnormal event detection in videos using generative adversarial nets [C]// Proceedings of the IEEE International Conference on Image Processing (ICIP) 2017: 1-5.

[13] Ravanbakhsh M, Sangineto E, Nabi M, et al. Training adversarial discriminators for cross-channel abnormal event detection in crowds [C]//Winter Conference on Applications of Computer Vision, 2019: 1896-1904.

[14] Narasimhan MG, S SK. Dynamic video anomaly detection and localization using sparse denoising autoencoders [J]. ***Multimedia Tools Appl***, 2018, 77(11): 1317313195.

[15] Sabzalian B, Marvi H, Ahmadyfard A. Deep and sparse features for anomaly detection and localization in video [C]//4th International Conference on Pattern Recognition and Image Analysis (IPRIA), 2019: 173-178

[16] Lin S, Yang H, Tang X, et al. Social MIL: Interaction-aware for crowd anomaly detection [C]//16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2019: 1-8.

[17] Fan Y, Wen G, Li D, et al. Video anomaly detection and localization via gaussianmixture fully convolutional variational autoencoder [J]. ***Computer Vision and Image Understanding***, 2020, 195: 102920.

[18] Liu W, Luo W, Lian D, et al. Future frame prediction for anomaly detection-a new baseline [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6536-6545.

[19] Springenberg J, Dosovitskiy A, Brox T, et al. Striving for simplicity: The all convolutional net [C]//International Conference on Learning Representations, 2015.

[20] 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.

[21] 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.

[22] Wang Dong, Zhang Xiaojun, Dai Lihua. Video anomaly detection and localization via deep Gaussian process regression [J]. ***Chinsese Journal of Scientific Instrument***, 2021, 35(3): 158-164. (in Chinese)

[23] Yu Bo, Tian Fuqing, Liang Weige. Fault diagnosis based on a deep convolution variational autoencoder network [J]. ***Journal of Electronic Measurenment and Instrument***, 2018, 39(10): 27-35. (in Chinese)