文章编号:1674-2974(2017)10-0130-09 DOI:10.16339/j.cnki.hdxbzkb.2017.10.018

基于深度学习特征的异常行为检测*

王军,夏利民†
(中南大学 信息科学与工程学院,湖南 长沙 410075)

摘 要:已有的异常行为检测大多采用人工特征,然而人工特征计算复杂度高且在复杂场景下很难选择和设计一种有效的行为特征.为了解决这一问题,结合堆积去噪编码器和改进的稠密轨迹,提出了一种基于深度学习特征的异常行为检测方法.为了有效地描述行为,利用堆积去噪编码器分别提取行为的外观特征和运动特征,同时为了减少计算复杂度,将特征提取约束在稠密轨迹的空时体积中;采用词包法将特征转化为行为视觉词表示,并利用加权相关性方法进行特征融合以提高特征的分类能力.最后,采用稀疏重建误差判断行为的异常.在公共数据库CAVIAR和BOSS上对该方法进行了验证,并与其它方法进行了对比试验,结果表明了该方法的有效性.

关键词:异常行为;深度学习特征;堆积去噪编码器;特征提取;稠密轨迹
中图分类号:TP391 **文献标志码**:A

Abnormal Behavior Detection Based on Deep-Learned Features

WANG Jun,XIA Limin†
(College of Information Science and Engineering,Central South University ,Changsha 410075,China)

Abstract: Most existing methods of abnormal behavior detection merely use hand-crafted features to represent behavior,which may be costly. Moreover,choice and design of hand-crafted features can be difficult in the complex scene without prior knowledge. In order to solve this problem,combining the stacked denoising autoencoders (SDAE) and improved dense trajectories,a new approach for abnormal behavior detection was proposed by using deep-learned features. To effectively represent the object behavior,two SDAE were utilized to automatically learn appearance feature and motion feature,respectively,which were constrained in the space-time volume of dense trajectories to reduce the computational complexity. The vision words were also exploited to describe the behavior using the method of bag of words. In order to enhance the discriminating power of these features,a novel method was adopted for feature fusing by using weighted correlation similarity measurement. The sparse representation was applied to detect abnormal behaviors via sparse reconstruction costs. Experiments results show the effectiveness of the proposed method in comparison with other state-of-the-art methods on the public databases CAVIAR and BOSS for abnormal behavior detection.

Key words: abnormal behavior; deep-learned features; SDAE; feature extraction; dense trajectories

* **收稿日期**:2017-01-11
基金项目:国家自然科学基金资助项目(50808025),National Natural Science Foundation of China(50808025)
作者简介:王军(1971—),男,山西应县人,中南大学博士研究生,讲师
†通讯联系人,E-mail:xlm@mail.csu.edu.cn

异常行为检测是智能监控系统的核心,近年来,已引起了学术界和工业界的广泛关注,并成为计算机视觉的重要研究课题.然而,场景的复杂性和异常行为的多样性,使得异常行为检测仍然是一项具有挑战性的工作.

根据所用特征,异常行为检测方法可分为两大类:基于视觉特征的异常检测和基于轨迹的异常检测.在基于物体轨迹的异常检测方面,Junejo 等[1]以轨迹的位置、速度、空时曲率等作为特征,提出了基于 DBN 的异常行为检测.Yang 等[2]提出了基于轨迹和多示例学习的局部异常检测方法.Li 等[3]用三次样条曲线表示目标轨迹,提出了基于稀疏表示的异常检测方法.Mo 等[4]提出了基于轨迹联合稀疏表示的异常行为检测方法.Kang 等[5]用 HMM 表示轨迹的运动模式,提出了基于 HMM 的异常检测方法.然而,在目标遮挡情况下,跟踪性能明显下降,导致异常检测率低.

为了解决上述问题,采用视觉特征进行异常行为检测,如方向梯度直方图 HOG,3D 空时梯度,光流直方图 HOF 等.Saligrama 等[6]利用局部空-时特征和最优决策规则进行异常检测.Nallaivarothayan[7]以光流加速度、光流梯度直方图作为特征,提出了基于 MRF 的异常事件检测.Wang 等[8]引入加速度信息,构建混合光流方向直方图,采用稀疏编码进行异常检测.Zhang 等[9]结合运动信息和外观信息进行异常检测.Mehran 等[10]利用光流特征构建社会力模型 SFM 来表示人群行为和异常行为识别.Wang 和 Schmid[11]提出了基于稠密轨迹行为识别,沿着稠密轨迹提取 HOG,HOF 以及运动边界描述符 MBH 来表示行为,利用 SVM 进行行为识别.然而,这些视觉特征是人为设计的,很难有效地反映行为特性,并且计算复杂.

目前,基于深度学习特征表示已成功地用于异常行为检测.Xu 等[12]采用堆积去噪编码器提取外观特征和运动特征,利用多类 SVM 进行异常行为检测.Erfani 等[13]利用深度置信网络提取行为特征,利用 SVM 进行异常行为检测.Zhou 等[14]采用空时卷积神经网络提取行为特征,进行人群异常行为检测和定位.Fang 等[15]以图像的显著性信息和多尺度光流直方图作为低层特征,然后采用深度学习网络 PCANet 从这些低层特征中提取更有效的特征用于异常事件检测.Hu 等[16]构建了一个深度增强慢特征分析网络,并用于行为特征提取和异常检测.基于深度学习的异常行为检测是借助于深度结构,通过多层非线性变换从原始图像中学习和挖掘复杂的非线性行为特征表示,相比基于人工特征的异常检测,由于深度学习得到的特征往往具有一定的语义特征和更强的区分能力,能更有效地表示行为特性,因此基于深度学习的异常行为检测准确率更高.同时基于深度学习的异常行为检测降低了计算复杂性.然而,基于深度学习的异常行为检测的性能远没有达到人们的期望,主要原因是:1)深度学习方法需要大量的样本用于训练,而一般的行为数据库样本相对比较少;2)为了平衡计算代价,基于深度学习的行为表示通常采用下采样策略,导致信息丢失[17].

为了解决该问题,受文献[11－12]的启发,我们结合堆积去噪编码器和改进的稠密轨迹,提出了一种基于深度学习特征的异常行为检测方法,利用深度学习提取行为特征,采用稀疏重建进行异常行为检测.该方法基本思想是利用堆积去噪编码器的学习能力和稠密轨迹方法的优良采样性能和特征提取策略来提取有效的行为特征.首先利用堆积去噪编码器,沿行为兴趣点轨迹提取深度外观特征和运动特征;由于不同行为视频中目标个数不同,从而兴趣点个数及其对应的轨迹数不同,导致整个特征长度不同,为了解决该问题,采用词包法将特征转化为行为视觉词表示;在此基础上,为了提高特征的分类能力,利用加权相关性法对这二种特征进行融合;最后,采用稀疏重建进行异常行为检测.在公共数据库 CAVIAR 和 BOSS 上利用文中方法和其它几种方法进行了对比实验.

1　改进的稠密轨迹提取

首先介绍稠密轨迹的提取[11].把图像划分成大小为 $W\times W$ 的网格,并在每个网格上进行采样.在采样过程中,以网格的中心作为采样点,对每个采样点进行三值插值处理得到兴趣点.为了得到足够多的兴趣点,按 $1/\sqrt{2}$ 比例增加空间尺度,由于视频分辨率的限制,可选择 8 个空间尺度.通过实验发现,当 $W=5$,在所有数据库上都可以得到很好的实验结果.

由于在没有结构的均匀区域内跟踪不到任何点,因此,将这些区域内的点去掉,这些点对应的自相关矩阵的特征值很小,为此设置一个阈值,舍弃自相关矩阵的特征值小于阈值的兴趣点.阈值设置为:

$$T=0.001\times\max_{i\in I}\min(\lambda_i^1,\lambda_i^2) \tag{1}$$

式中:$(\lambda_i^1,\lambda_i^2)$是图像 I 中点 i 自相关矩阵的特征值.

得到兴趣点后,利用光流法在每个空间尺度上对兴趣点进行跟踪,得到其运动轨迹.设第 t 帧图像 I_t 的稠密光流为 $w_t=(u_t,\ v_t)$,这里 u_t 和 v_t 分别为光流的水平分量和竖直分量,则 I_t 中某一点 $z_t=(x_t,\ y_t)$在下一帧 I_{t+1} 中的位置可通过对 w_t 进行中值滤波平滑得到:

$$z_{t+1}=(x_{t+1},y_{t+1})=(x_t,y_t)+(\boldsymbol{M}*w_t)\mid(x_t,y_t) \tag{2}$$

式中:$\boldsymbol{M}$ 为中值滤波的核,其大小为 3×3.采用中值滤波可以很好地保留跟踪过程中边界上的点.

把各帧对应的兴趣点连接起来就得到该兴趣点的轨迹:$(z_t,z_{t+1},\cdots,z_{t+L-1})$.在跟踪过程中可能会出现轨迹漂移,为了避免这种现象,我们限制采样帧的长度 $L=15$,另外,由于一条持续 5 帧图像的轨迹才是"可靠的",短于 5 帧的轨迹自动被删除.为了得到稠密的轨迹,在每一帧图像中,如果在一个 $W\times W$ 的邻域内没有任何跟踪点,那么选择新的点作为采样点,并对新采样点进行跟踪.在后处理阶段,我们删掉静态轨迹和突然漂移比较大的轨迹,因为前者不包含运动信息,后者很有可能是跟踪误差所致.

2 行为深度特征

稠密轨迹附近包含丰富的行为信息,我们利用稠密轨迹和堆积去噪编码器 SDAE 提取行为深度外观特征和运动特征.首先介绍堆积去噪编码器,然后描述基于稠密轨迹和堆积去噪编码器的行为深度学习特征提取方法.

2.1 堆积去噪编码器

去噪编码器 DAE[18] 是一种三层神经网络,用于从噪声数据 $\tilde{\boldsymbol{x}}_i$ 重建原数据 $\boldsymbol{x}_i$,DEA 包含二个部分:编码器和解码器.DEA 学习就是学习两个映射函数 $f_e(\boldsymbol{W};\boldsymbol{b})$和 $f_d(\boldsymbol{W}';\boldsymbol{b}')$,其中 $\boldsymbol{W},\boldsymbol{b}$ 表示编码器部分的权值矩阵和偏差向量,$\boldsymbol{W}',\boldsymbol{b}'$对应于解码器的参数.对于噪声数据 $\tilde{\boldsymbol{x}}_i$,编码器的隐层输出为 $\boldsymbol{y}$:

$$\boldsymbol{y}=f_e(\tilde{\boldsymbol{x}}_i)=s(\boldsymbol{W}\tilde{\boldsymbol{x}}_i+\boldsymbol{b})=\frac{1}{1+\mathrm{e}^{-(\boldsymbol{W}\tilde{\boldsymbol{x}}_i+\boldsymbol{b})}} \tag{3}$$

解码器目的是从噪声数据 $\tilde{\boldsymbol{x}}_i$ 重建原数据 $\boldsymbol{x}_i$:

$$\boldsymbol{z}_i=f_d(\boldsymbol{y})=s(\boldsymbol{W}'\boldsymbol{y}+\boldsymbol{b}') \tag{4}$$

给定一组训练样本 $\boldsymbol{X}=\{\boldsymbol{x}_i\}_{i=1}^N$,通过求解下列优化问题来学习 DEA 的参数$(\boldsymbol{W},\boldsymbol{W}',b,b')$:

$$\min_{\boldsymbol{W},\boldsymbol{W}',b,b'}\sum_{i=1}^{N}\|\boldsymbol{x}_i-\boldsymbol{z}_i\|_2^2+\lambda(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{W}'\|_F^2)+\beta\sum_{j=1}^{K}KL(\mu\|\hat{\mu}) \tag{5}$$

式中:第一项表示重建误差,第二项为权值惩罚项,第三项为稀疏性约束,λ,β 为平衡参数,μ 是稀疏性参数,它表示隐节点的系数水平,K 是隐含层的节点个数,$\hat{\mu}_j$ 是隐含层第 j 个节点对所有训练样本的平均阈值化激活值,如果平均激活值大于 0.5 时,$\hat{\mu}_j=1$,否则 $\hat{\mu}_j=0$.第三项稀疏性约束为:

$$KL(\mu\|\hat{\mu}_j)=-\mu\log\hat{\mu}_j+(1-\mu)\log(1-\hat{\mu}_j) \tag{6}$$

利用梯度下降法对式(5)进行求解,可确定 DAE 参数.

将多个 DAE 逐层堆叠形成 SDAE,其中低层 DAE 的输出作为上一层 DAE 的输入.SDAE 的训练,采用从低层向高层逐层训练的方式对各层中的 DAE 进行训练.训练好的 SDAE 可用于从输入数据中学习到有效的特征表示.

2.2 行为深度特征提取

我们分别利用两个 SDAE 在以轨迹为中心的 3D 体积中提取行为的外观特征和运动特征.3D 体积的大小为 $N\times N\times L$,$L=15$ 为轨迹的长度,N 取 32.为了嵌入结构信息,首先,将该立方体划分为 $n_\sigma\times n_\sigma\times n_\tau$ 的时-空网格,本文取 $n_\sigma=2$,$n_\tau=3$;然后在这些网格(大小为 16×16×5)中用 SDAE 提取深度特征;最后将所有网格的特征结合得到与该轨迹对应的深度特征.

图 1 表示提取行为深度特征的 SDAE 的结构,它包括两个 SDAE,其中一个用于学习外观特征,称为外观堆积去噪编码器(ASDAE),另一个用于学习运动特征,称为运动堆积去噪编码器(MSDAE).每个 SDAE 包括两个部分:编码器和解码器.编码器的输入层节点数等于输入数据的维数,然后,每层节点数逐层减少一半,直到"瓶颈"隐层,解码器的结构与编码器对称,"瓶颈"隐层的输出就是深度特征,实验结果表明,当隐层数为 5 时系统检测率最高.

2.2.1 外观深度特征

在灰度视频图像中,以轨迹为中心的 3D 体积中各个网格区域的图像(16×16×5)作为 ASDAE 的输入,其"瓶颈"隐层的输出作为外观深度特征,然后,将体积中各个网格区域的外观深度特征连接起来得到该轨迹(该兴趣点)的外观深度特征.由于

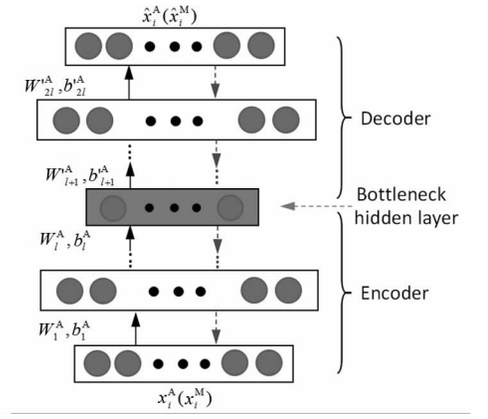


图 1　堆积去噪编码器的结构

Fig.1　The structure of SDAE

网格区域大小为 16×16×5,所以 ASDAE 的输入层节点个数为 1 280,各隐层节点个数分别为:640,320,160,80,40;一条轨迹对应的外观特征维数为 480(2×2×3×40).

2.2.2　运动深度特征

根据光流场提取运动深度特征.以轨迹为中心的 3D 体积中各个网格区域的光流场(16×16×2×5)作为 MSDAE 的输入,其"瓶颈"隐层的输出作为运动深度特征,然后,将体积中各个网格区域的运动深度特征连接起来得到该轨迹(该兴趣点)的运动深度特征.这里 ASDAE 的输入层节点个数为 2 560,各隐层节点个数分别为:1 280,640,320,160,80;一条轨迹对应的外观特征维数为 960(2×2×3×80).

将两类特征合在一起,则每一条轨迹可用1 440 维的向量表示.相比一般深度特征而言,我们的深度学习特征具有以下优点:

1)不需要大量的行为样本,适合于样本少的行为数据库识别.因为我们用深度网络不是直接提取整个行为特征,而是提取行为区域中的采样点(兴趣点)的特征,而采样点的个数足以训练深度网络.

2)尽可能保留了行为信息.这是因为我们采用的是稠密轨迹采样,这是一种浓密的采样,信息丢失较少,同时保留了结构信息.

3)计算复杂度不高.因为我们只是在稠密轨迹附近提取特征,而包含少量运动信息的区域并未计算.

3　基于加权相关性的特征融合

3.1　行为特征的词包表示

由上可知,对于一条轨迹可以得到一个 1 440 维的特征向量.由于不同视频中目标个数不同,导致兴趣点的个数以及对应的轨迹条数不同,行为特征维数不同,因此,利用词包法将行为特征用视觉词表示,以统一行为的维数.分别根据外观深度特征和运动深度特征利用词包法建立外观视觉字典和运动视觉字典,然后分别由外观视觉词和运动视觉词表示外观深度特征和运动深度特征.

具体方法是:首先,提取轨迹的外观深度特征/运动深度;然后对所有轨迹外观深度特征/运动深度进行聚类,得到 N_v 个类中心,每个类对应一个视觉词.对于测试行为样本,根据最近邻原则,把它的每条轨迹分类到每一类,于是可得到各个视觉词在样本中出现的频率,这些频率就构成样本的视觉词表示.

经过多次实验得到外观视觉词和运动视觉词的个数分别为 230 和 470.因此,外观深度特征和运动深度特征分别用 230 维和 470 维的视觉词向量表示.

3.2　基于加权相关性的特征融合

为了提高特征的分类能力,采用基于加权相关性的特征融合方法将外观深度特征和运动深度特征结合在一起形成 700 维的特征向量.

为了表示方便,将外观深度特征和运动深度特征分别用 $\boldsymbol{y}^1,\boldsymbol{y}^2$ 表示,则融合后的特征 $\boldsymbol{y}$ 为:

$$\boldsymbol{y}=(\omega_1\boldsymbol{y}^1,\omega_2\boldsymbol{y}^2) \tag{7}$$

这里 ω_i 是加权系数,且 $(\omega_1)^2+(\omega_2)^2=1$.我们根据类内一致性和类间可分性来确定该加权系数.

类内一致性:一般希望同类中的样本在特征空间尽可能接近.但是通常在同一类中样本特征会出现较大的方差,因此,没有必要要求同类中的所有样本都相互接近.一种权衡的方法是保证同类中同一近邻内的样本尽可能接近.设 $\boldsymbol{y}_i=(\omega_1\ \boldsymbol{y}_i{}^1,\omega_2\ \boldsymbol{y}_i{}^2)$,$\boldsymbol{y}_j=(\omega_1\ \boldsymbol{y}_j{}^1,\omega_2\ \boldsymbol{y}_j{}^2)$ 表示第 i,第 j 个样本,则类内一致性定义为:

$$S_c=\sum_{i=1}^{N}\sum_{j\in N_k^+(F_i)}\frac{\langle\boldsymbol{y}_i,\boldsymbol{y}_j\rangle}{\|\boldsymbol{y}_i\|\ \|\boldsymbol{y}_j\|}=\sum_{i=1}^{N}\sum_{j\in N_k^+(F_i)}\frac{\sum_{k=1}^{2}\omega_k^2\boldsymbol{y}_i^k\boldsymbol{y}_j^k}{\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_i^k)^2}\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_j^k)^2}} \tag{8}$$

式中:$N_k^+(F_i)$表示样本 F_i 的、且与 F_i 属于同一类的 k 个最近邻样本的索引集.

类间可分性:要求特征具有好的区分性,即不同类的两个样本在特征空间尽可能远离.但这样的样本对很多,为了减少计算量,只考虑特征空间分界面附近的样本对.于是定义类间可分性为:

$$S_c=\sum_{i=1}^{N}\sum_{j\in N_k^-(F_i)}\frac{\langle \boldsymbol{y}_i,\boldsymbol{y}_j\rangle}{\|\boldsymbol{y}_i\|\,\|\boldsymbol{y}_j\|}=\sum_{i=1}^{N}\sum_{j\in N_k^-(F_i)}\frac{\sum_{k=1}^{2}\omega_k^2\boldsymbol{y}_i^k\boldsymbol{y}_j^k}{\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_i^k)^2}\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_j^k)^2}} \tag{9}$$

式中:$N_k^-(F_i)$表示样本 F_i 的、且与 F_i 不同类的 k 个最近邻样本的索引集.

融合后的特征应具有良好的类内一致性和类间可分性,因此通过求解下列优化问题确定加权系数:

$$\max_{\omega}\{(S_c-S_b)+\lambda_s\|\omega\|\}\quad \text{s.t.}\quad \omega_k>0,\ \|\omega\|=1 \tag{10}$$

式中,λ_s是控制参数.

采用梯度下降法求解式(10),即:

$$\omega_k(t+1)=\omega_k(t)+\eta\frac{\partial L}{\partial\omega_k}\Big|_{\omega_k=\omega_k(t)} \tag{11}$$

式中:t 为迭代次数,η 为迭代步长,$L=(S_c-S_b)+\lambda_s\|\omega\|$ 为目标函数.

$$\frac{\partial L}{\partial\omega_k}=\sum_{i=1}^{N}\left\{\sum_{j\in N_k^+(x_i)}\frac{\partial h_{ij}(\omega)}{\partial\omega_k}-\sum_{j\in N_k^-(x_i)}\frac{\partial h_{ij}(\omega)}{\partial\omega_k}\right\}+2\lambda_s\omega_k \tag{12}$$

这里:

$$h_{ij}=\frac{\sum_{k=1}^{2}\omega_k^2\boldsymbol{y}_i^k\boldsymbol{y}_j^k}{\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_i^k)^2}\sqrt{\sum_{k=1}^{2}\omega_k^2(\boldsymbol{y}_j^k)^2}} \tag{13}$$

$$\frac{\partial h_{ij}(\omega)}{\partial\omega_k}=\frac{2f_{ij}^k b_{ii}^{1/2}b_{jj}^{1/2}+b_{ij}(b_{jj}^{1/2}f_{ii}^{1/2}/b_{ii}^{1/2}+b_{ii}^{1/2}f_{jj}^{1/2}/b_{jj}^{1/2})}{\sqrt{\sum_{k=1}^{2}\omega_k^2(f_{ik})^2}\sqrt{\sum_{k=1}^{2}\omega_k^2(f_{jk})^2}} \tag{14}$$

$$f_{ij}^k=\boldsymbol{y}_i^k\boldsymbol{y}_j^k \tag{15}$$

$$b_{ij}=\sum_{k=1}^{2}\omega_k^2\boldsymbol{y}_i^k\boldsymbol{y}_j^k \tag{16}$$

4 基于稀疏重建的异常行为检测

得到行为的特征后,采用稀疏重建来进行异常行为检测,其基本思想是任何行为可以用一组正常训练样本的稀疏线性组合表示.对于正常行为,稀疏重建误差小,而异常行为稀疏重建误差比较大.因此,我们可根据重建误差[19]来进行异常检测.

设有 C 类正常行为,每个行为用上述特征向量表示,$\boldsymbol{D}=[\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_C]$表示稀疏字典,其中 $\boldsymbol{D}_i$ 是由 K 个第 i 类行为构成的子字典,对于测试样本 $\boldsymbol{y}$ 可表示为:

$$\boldsymbol{y}=\boldsymbol{D}\alpha \tag{17}$$

这里,$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_c]^{\mathrm{T}}$为稀疏编码向量.稀疏重建的关键是字典学习和求解稀疏编码.

4.1 字典学习

给定训练样本集 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_N\}\in R^{m\times N}$,$\boldsymbol{y}_i\in R^m$ 表示第 i 个正常样本的特征向量,目的是学习字典 $\boldsymbol{D}$ 和稀疏编码向量 $\boldsymbol{\alpha}$,使 $\boldsymbol{Y}$ 可以通过字典的加权和来重建,即:$\boldsymbol{Y}=\boldsymbol{D\alpha}$,也就是求解下列优化问题:

$$\min_{D,\alpha}\|\boldsymbol{Y}-\boldsymbol{D\alpha}\|_F+\lambda\|\boldsymbol{\alpha}\|_{2,1} \tag{18}$$

这里 $\boldsymbol{\lambda}$ 控制参数,第一项是重建误差,第二项是稀疏性约束.这是一个非凸的优化问题.但是如果 $\boldsymbol{D}$ 和 $\boldsymbol{\alpha}$ 中的一个固定,则问题就变为线性的.因此,通过依次固定 $\boldsymbol{D}$ 和 $\boldsymbol{\alpha}$,可以导出 $\boldsymbol{D}$ 和 $\boldsymbol{\alpha}$,具体算法[20]如下:

1)输入训练样本集 $\boldsymbol{Y}$,初始字典 $\boldsymbol{D}^0\in\boldsymbol{R}^{m\times K}$,$i=0$;

2)令 $i=i+1$;

3)固定 $\boldsymbol{D}$ 利用式(22)求 $\boldsymbol{\alpha}^i$;

4)固定 $\boldsymbol{\alpha}$ 利用式(21)求 $\boldsymbol{D}^i$;

5)重复 2),3),4)步,直到收敛;

6)输出字典 $\boldsymbol{D}^i$.

算法中每一步确定 $\boldsymbol{D}^i$ 和 $\boldsymbol{\alpha}^i$ 就是求解:

$$\boldsymbol{D}^i=\underset{D}{\operatorname{argmin}}\|\boldsymbol{Y}-\boldsymbol{D\alpha}\|_F \tag{19}$$

$$\boldsymbol{\alpha}^i=\underset{\alpha}{\operatorname{argmin}}\|\boldsymbol{Y}-\boldsymbol{D\alpha}\|_F+\lambda\|\boldsymbol{\alpha}\|_{2,1} \tag{20}$$

式(19)采用 K-SVD 算法来求解.由于 $\|\boldsymbol{\alpha}\|_{2,1}$ 是非平滑的,所以,式(20)对应的优化问题是凸的、非平滑的优化问题,用一般优化算法求解会导致收敛速度很慢,在此,利用 Nesterov 提出的方法[21]来求解.考虑目标函数 $f_0(x)+g(x)$,其中,$f_0(x)$是凸的且平滑,而 $g(x)$是凸的、非平滑,Nesterov 采用

$$P_{Z,L}(x)=f_0(Z)+\langle\nabla f_0(Z),x-Z\rangle+L\|x-Z\|_F^2+g(x) \tag{21}$$

来近似表示 Z 处的 $f_0(x)+g(x)$.这里 L 是 Lipschitz 常数.这样每次迭代,只需求解 $\min_x P_{Z,L}(x)$.

定义 $f_0(\boldsymbol{\alpha})=\|\boldsymbol{Y}-\boldsymbol{D\alpha}\|_F, g(\boldsymbol{\alpha})=\lambda\|\boldsymbol{\alpha}\|_1$，则有：

$$P_{Z,L}(\boldsymbol{\alpha})=f_0(Z)+\langle\nabla f_0(Z),\boldsymbol{\alpha}-\boldsymbol{Z}\rangle+L\|\boldsymbol{\alpha}-\boldsymbol{Z}\|_F^2+\lambda\|\boldsymbol{\alpha}\|_{2,1} \tag{22}$$

类似文献[21]，可得到式(20)的解：

$$\underset{\alpha}{\operatorname{argmin}}P_{Z,L}(\alpha)=H_{\lambda/L}\left(Z-\frac{1}{L}\nabla f(Z)\right) \tag{23}$$

式中，$H_\tau:\boldsymbol{M}\in\boldsymbol{R}^{k\times k}\rightarrow\boldsymbol{N}\in\boldsymbol{R}^{k\times k}$

$$\boldsymbol{N}_i=\begin{cases}0, & \|\boldsymbol{M}_i\|\leqslant\tau;\\(1-\tau/\|\boldsymbol{M}_i\|)\boldsymbol{M}_i, & \text{otherwise}\end{cases} \tag{24}$$

这里，$\tau=\lambda/L$，$\boldsymbol{M}=\boldsymbol{Z}-(1/L)\nabla f(\boldsymbol{Z})$，$\boldsymbol{M}_i$是原数据的第 i 行，$\boldsymbol{N}_i$是计算得到的矩阵的第 i 行.

4.2　异常行为检测

给定一个字典 $\boldsymbol{D}$，对于测试样本 y 可用式(17)表示，其中，稀疏编码 $\boldsymbol{\alpha}$ 可通过求解式(25)获得：

$$\boldsymbol{\alpha}^*=\min_{\alpha}\|\boldsymbol{y}-\boldsymbol{D\alpha}\|_2+\lambda\|\boldsymbol{\alpha}\|_1 \tag{25}$$

一旦得到最优的稀疏编码 $\boldsymbol{\alpha}^*$，可以计算稀疏重建代价(SRC)[19]：

$$S(\boldsymbol{y},\boldsymbol{\alpha}^*,\boldsymbol{D})=\|\boldsymbol{y}-\boldsymbol{D\alpha}\|_2+\lambda\|\boldsymbol{\alpha}^*\|_1 \tag{26}$$

对于正常行为，稀疏重建代价较小，而异常行为稀疏重建代价较大.因此如果

$$S(\boldsymbol{y},\boldsymbol{\alpha}^*,\boldsymbol{D})>\varepsilon \tag{27}$$

则 $\boldsymbol{y}$ 为异常行为.式中，ε 是预先设置的阈值.

5　实验与结果

为了验证所提方法的有效性，我们在公共数据库 CAVIAR 和 BOSS 上进行了实验，并与以下 4 种方法进行对比：1) Mo 等[4]提出的基于轨迹联合稀疏表示的异常行为检测方法；2)Wang 等[11]提出的基于稠密轨迹方法；3)Xu 等[12]提出的基于外观深度网络和运动深度网络的异常检测方法；4)Zhou 等[14]提出的基于空时卷积神经网络的异常检测方法.

我们以 ROC 曲线评估异常行为检测效果，ROC 曲线以 FPR 为横坐标，TPR 为纵坐标.其中：

$$TPR=\frac{TP}{TP+FN},FPR=\frac{FP}{FP+TN} \tag{28}$$

这里，TP (True positive)是真异常行为，FN(False negative)假正常行为，FP (False positive)是假异常行为，TN(True negative)真正常行为. 我们选择不同的阈值 ε，分别计算 FPR 和 TPR，然后，作 ROC 曲线.ROC 曲线越靠近上方，曲线下面积 AUC 越大，则检测的准确率越高，否则检测准确率越低.

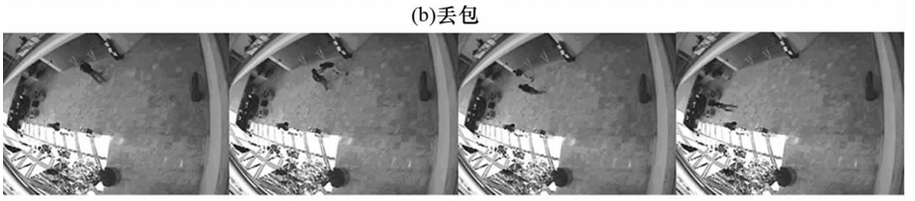
5.1　CAVIAR 数据库

在实验中，我们利用 CAVIAR 数据库的第一部分数据进行实验.这些数据是利用广角镜头拍摄的，包括行走、闲逛，休息、跌倒或晕倒等单人行为，以及会谈、两人走近和分开、两人打架等交互行为.每个场景包括 3 至 5 片段，每个片段持续 40 至 60 s，分辨率为 384×288，共采用了 26 个片段.该数据库可用于单人异常行为和两人的交互异常行为.

由于每片段包含几种行为，因此采用人工方法将每片段分成多个镜头，使每个镜头只包含一个行为.结果数据库分成 1 200 个行为镜头，其中，200 个异常行为和 1 000 正常行为.其中打架、追赶、闲逛、

(a)晕倒

(b)丢包



(c)打架

图 2　CAVIAR 数据库

Fig.2　CAVIAR dataset

丢包和晕倒等异常行为个数分别为 44,39,42,37 和 38.图 2 是该数据库的一些异常行为的视频帧.其中,图 2(a)为一个人穿过大厅时晕倒;图 2(b)是一个人丢包;图 2(c)是两个人打架.实验中,我们首先提取稠密轨迹;其次随机选择 500 万条轨迹训练 SDA,并采用 K-均值聚类法得到 230 个外观视觉词和 470 个运动视觉词(视觉词的个数分别从 100 到 1 000 变化时,发现当外观视觉词为 230 个、运动视觉词为 470 个可取得最好性能);然后随机选择 800 个正常行为学习特征融合参数,得到:$\omega_1=0.3$,$\omega_2=0.7$;最后,用 800 个正常行为学习稀疏字典.以全部异常行为样本和余下的 200 个正常样本作为测试样本,测试结果为 AUC=0.996,表明文中方法的检测率很高.图 3 给出了几种检测方法的 ROC 曲线.表 1

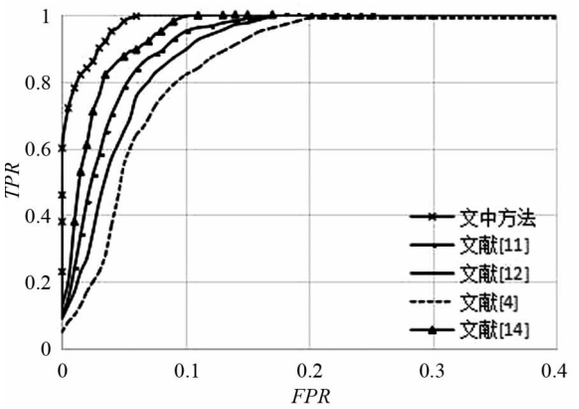


图 3 在 CAVIAR 数据库上异常行为检测的 ROC 曲线

Fig.3 The ROCs for abnormal detection on CAVIAR dataset

是几种检测方法 ROC 曲线下的面积 AUC.结果表明,对于单人异常行为和两人的交互异常行为,文中方法比其它 3 种方法具有更好的检测准确率.

表 1 5 种不同方法在 CAVIAR 数据库上的 AUC

Tab.1 AUC of five methods on CAVIAR dataset

方法	文献[4]	文献[11]	文献[12]	文献[14]	文中方法
AUC	0.910	0.960	0.953	0.962	0.996

5.2 BOSS 数据库

BOSS 数据库是在列车上通过 9 个摄像头拍摄的,包括 14 个视频序列,每段视频持续 1 到 5 min.分辨率为 720×576,帧率 25 fps.14 个视频包括 3 个正常行为视频,11 个异常行为视频.异常行为包括:抢手机、打架、抢报纸、骚扰、晕倒、恐慌.该数据库可用于单人、多人的异常行为检测.

同样采用人工方法将每视频分成多个镜头,每个镜头只包含一个行为.结果这个数据库分成1 600 个行为镜头,其中,400 个异常行为和 1 200 正常行为,抢手机、打架、抢报纸、骚扰、晕倒、恐慌等异常行为数分别为 52,70,65,58,65,90. 图 4 是该数据库的一些异常行为的视频帧.其中图 4(a)为打架;图 4(b) 是晕倒;图 4(c) 为抢手机.实验过程和 CAVIAR 数据库上的实验一样,随机选择 500 万条轨迹训练 SDA、确定外观视觉词和运动视觉词;选择 1 000个正常行为学习特征融合参数($\omega_1=0.35$,$\omega_2=0.65$)和稀疏字典.用全部异常行为样本和余下的 200 个正常样本作为测试样本,测试结果为 AUC=0.982,即我们的方法有很高的检测率.图 5 是几种检测方法的 ROC 曲线.表 2 是几种检测方法 ROC 曲线下的面积 AUC.结果表明,对于多人的交互异常行为,文中方法比其它 3 种方法具有更好的检测准确率.

图 4 BOSS 数据库

Fig.4 BOSS dataset

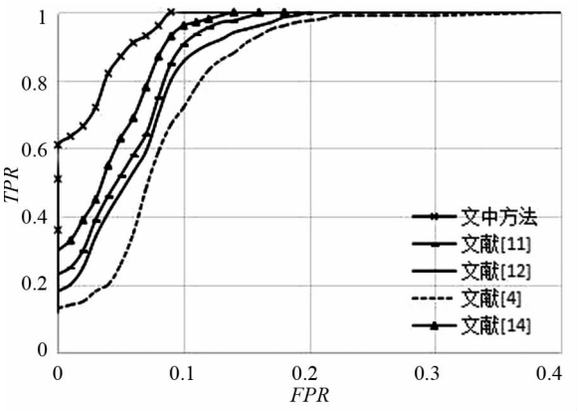


图 5　在 BOSS 数据库上异常行为检测的 ROC 曲线

Fig.5　The ROCs for abnormal behavior detection on BOSS

表 2　5 种不同方法在 BOSS 数据库上的 AUC

Tab.2　AUC of five methodson BOSS dataset

方法	文献[4]	文献[11]	文献[12]	文献[14]	文中方法
AUC	0.897	0.945	0.936	0.950	0.982

以上结果表明,无论是单人的异常行为检测还是多人异常行为检测,与其它方法相比,我们的方法都具有更高的检测准确率.文献[4]采用基于目标轨迹的异常行为检测,由于视频中有多人交互情况,不可避免存在遮挡问题,使得跟踪性能明显下降,导致异常检测率比其它方法低.文献[11]采用稠密轨迹特征(HOG,HOF 和 MBH),这些特征是目前最好的人工特征[17],因此检测准确率明显提高.文献[12]和[14]分别采用堆积自动编码器和空时卷积神经网络提取行为特征,这些特征能较好地描述人体行为,因此其检测率超过 93%.但这两个方法检测率并不比基于稠密轨迹方法高,这是因为:1)这些方法需要大量的样本用于训练深度学习网络,而行为数据库样本相对比较少;2)为了平衡计算代价,提取行为表示通常下采样策略,采用导致信息丢失.而我们的方法是沿行为兴趣点的稠密轨迹提取行为的外观特征和运动特征.一方面,由于在稠密轨迹附近有丰富的运动信息,利用堆积去噪编码器强大的学习能力可提取有效的行为特征;另一方面,用深度网络不是直接提取整个行为特征,只是提取行为区域中的采样点的特征,而这些采样点的个数足以训练深度网络,因此不需要大量的样本训练深度网络,解决了训练样本不足对深度学习的影响,所以异常行为检测率比其他几种方法更高.

5.3　计算复杂性

我们对所提方法的计算复杂度进行了测试,所有实验都在工作站 (2.8 GHz CPU,32GB RAM)上进行.表 3 为几种方法分别在 CAVIAR,BOSS 中每帧平均计算时间.从表中可看出,基于目标轨迹的异常行为检测计算时间最少,而基于稠密轨迹的异常行为检测计算时间最长,我们的方法与其它二个基于深度学习的异常行为检测计算时间适中,表明我们的方法是非常有效的.这是因为这几种方法在异常识别阶段的时间都差不多,并占总时间的很小部分,主要计算时间花费在特征提取上.基于目标轨迹的异常行为检测用 B 样条函数近似目标轨迹,取 B 样条函数 50 个控制点的坐标作为行为特征,显然这些特征的提取需要时间最少.而基于稠密轨迹的异常行为检测所用时间主要在光流场计算、稠密轨迹提取和 HOG,HOF 和 MBH 的计算,其中特征 HOG,HOF 和 MBH 的计算需要很长时间;文献[12]和[14]分别采用深度网络提取特征,因此计算速度快;相比文献[12]和[14],我们的方法多了一步稠密轨迹提取,该过程约占总计算的 10%,但由于我们只是提取稠密轨迹附近的特征,那些包含少量运动信息的图像区域并未计算,所以总的计算时间并未显著提高.

表 3　5 种不同方法的计算复杂度

Tab.3　Computation complexity of five methods

方法	文献[4]	文献[11]	文献[12]	文献[14]	文中方法
CAVIAR	0.21	0.84	0.41	0.39	0.42
BOSS	0.23	0.91	0.44	0.43	0.45

6　结　论

结合堆积去噪编码器和改进的稠密轨迹,提出了一种基于深度学习特征的异常行为检测方法.首先,利用堆积去噪编码器沿着稠密轨迹提取深度外观特征和深度运动特征;然后,为了提高特征的分类能力,利用加权相关性方法对这二种特征进行融合;最后,采用稀疏重建进行异常行为检测.为了验证文中方法的有效性,在公共数据库 CAVIAR 和 BOSS 上对文中方法进行了测试,并与其它几种方法进行了对比,结果表明,文中方法具有更高的检测率和较低的计算复杂度.

参考文献

[1]　JUNEJO I N.Using dynamic Bayesian network for scene modeling and anomaly detection[J].Signal Image and Video Processing,2010,4(1):1－10.

[2]　YANG W Q,GAO Y ,CAO L B. TRASMIL:A local anomaly

detection framework based on trajectory segmentation and multi-instance learning[J]. Computer Vision and Image Understanding,2013,117(10):1273－1286.

[3] LI C,HAN Z,YE Q M,*et al*. Visual abnormal behavior detection based on trajectory sparse reconstruction analysis[J]. Neurocomputing,2013,119(6):94－100.

[4] MO X,MONGAV,BALA R,*et al*. Adaptive sparse representations for video anomaly detection[J]. IEEE Transactions on Circuits and Systems for Video Technology,2014,24(4):631－645.

[5] KANG K,LIU W B ,XONG W W. Motion pattern study and analysis from video monitoring trajectory [J]. IEICE Transactions on Information and Systems,2014,97(6):1574－1578.

[6] SALIGRAMA V,CHEN Z. Video anomaly detection based on local statistical aggregates[C]// IEEE Computer Vision and Pattern Recognition. Providence:IEEE,2012:2112－2119.

[7] NALLAIVAROTHAYAN H,FOOKES C,DENMAN S. An MRF based abnormal event detection approach using motion and appearance features[C]// IEEE International Conference on Advanced Video and Signal based Surveillance. Seoul:IEEE,2014:343－348.

[8] WANG Q, MA Q ,LUO C H,*et al*. Hybrid histogram of oriented optical flow for abnormal behavior detection in crowd scenes[J]. International Journal of Pattern Recognition and Artificial Intelligence,2016,30(2):1－14

[9] ZHANG Y ,LU H C,ZHANG L H,*et al*. Combining motion and appearance cues for anomaly detection[J].Pattern Recognition,2016,51(C):443－452.

[10] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]//IEEE Conference on Computer Vision & Pattern Recognition. Miami Beach:IEEE Comp Soc,2009:935－942.

[11] WANG H,SCHMID C. Action recognition with improved trajectories[C]//IEEE International Conference on Computer Vision.Sydney,Australia:IEEE Comp Soc,2013:3551－3558.

[12] XU D,RICCI E. Learning deep representations of appearance and motion for anomalous event detection [C]//British Machine Vision Conference. Swansea:Spring,2015:1－12.

[13] ERFANI S M,RAJASEGARAR S,KARUNASEKERA S,*et al* . High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning[J]. Pattern Recognition,2016,58(C):121－134.

[14] ZHOU S F,SHEN W,ZENG D,*et al*. Spatial- temporal convolutional neural networks for anomaly detection and localization in crowded scenes[J]. Signal Processing -Image Communication,2016,47:358－ 368.

[15] FANG Z, FEI F, FANG Y. Abnormal event detection in crowded scenes based on deep learning [J]. Multimedia Tools and Applications,2016,75(22):14617－14639.

[16] HU X,HU S Q,HUANG Y P,*et al*. Video anomaly detection using deep incremental slow feature analysis network[J]. IET Computer Vision,2016,10(4):258－267.

[17] ZHU F,SHAO L,XIE J,*et al*. From handcrafted to learned representtations for human action recognition: A survey[J]. Image and Vision Computing,2016,55(1):42－52.

[18] VINCENT P,LAROCHELLE H,LAJPIE I,*et al*. Stacked denoising autoencoders:Learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research,2010,11(12):3371－3408.

[19] CONG Y,YUAN,LIU J.Sparse reconstruction cost for abnormal event detection[C]// IEEE Computer Vision and Pattern Recognition. Colorado Springs:IEEE,2011,3449－3456.

[20] 陈炳权,刘宏立. 基于稀疏分解的分块图像压缩编码算法[J]. 湖南大学学报:自然科学版,2014,41(2):95－101.
CHEN Bingquan,LIU Hongli. Block compressed image coding based on sparse decomposition[J]. Journal of Hunan University:Natural Sciences,2014,41(2):95－101.(In Chinese)

[21] NESTEROV Y. Gradient methods for minimizing composite objective function[J]. Mathematical Programming, 2013, 140(1):125－161.