基于贝叶斯融合的时空流异常行为检测模型

陈 莹* 何丹丹

(江南大学轻工过程先进控制教育重点实验室 无锡 214122)

摘 要：针对直接利用卷积自编码网络未考虑视频时间信息的问题，该文提出基于贝叶斯融合的时空流异常行为检测模型。空间流模型采用卷积自编码网络对视频单帧进行重构，时间流模型采用卷积长短期记忆(LSTM)编码-解码网络对短期光流序列进行重构。接着，分别计算空间流模型和时间流模型下每帧的重构误差，设计自适应阈值对重构误差图进行二值化，并基于贝叶斯准则对空间流和时间流下的重构误差进行融合，得到融合重构误差图，并在此基础上进行异常行为判断。实验结果表明，该算法在UCSD和Avenue视频库上的检测效果优于现有异常检测算法。

关键词：异常行为检测；贝叶斯融合；时空流

中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2019)05-1137-08

DOI: 10.11999/JEIT180429

Spatial-temporal Stream Anomaly Detection Based on Bayesian Fusion

CHEN Ying HE Dandan

(*Key Laboratory of Advanced Control Education in Light Industry Process*, *Jiangnan University*, *Wuxi* 214122, *China*)

Abstract: Focusing on the problem that convolutional auto-encoder network based anomaly detection ignores time information, a novel anomaly detection model based on Bayesian fusion of spatial-temporal stream is proposed. A convolution auto-encoder network is used in spatial stream model to reconstructs video frames, and a convolutional Long Short-Term Memory (LSTM) encoder-decoder network is used to reconstruct short-term optical sequence in the temporal stream model. Then, the reconstruction errors under spatial and temporal stream are calculated separately. Meanwhile, an adaptive thresholds is designed to obtain the reconstruction binary error maps. Finally, the Bayesian fusion strategy is developed to combine the reconstruction error of spatial and temporal stream to obtain the final fusion reconstruction error map based on which the abnormal behavior can be determined. Experimental results show that the proposed algorithm is superior to the existing anomaly detection algorithms in UCSD and Avenue datasets.

Key words: Anomaly detection; Bayesian fusion; Spatial-temporal stream

1 引言

公众对公共安全问题的逐渐重视，并且在智能监控和安防领域的快速发展下，基于监控视频的人群异常行为检测已成为图像处理、机器视觉等相关领域的研究热点。

在基于监控视频的人群异常行为检测领域中，基于重构[1–3]的异常行为检测模型较为流行。该类模型假设正常行为的类似度较高，正常行为样本的相关性较大，一个正常行为可利用其他的正常行为来联合表示。一般情况下，先在正常行为的特征数据的基础上训练得到表示正常行为的某种基，以训练得到的基为基准，重构正常行为会得到相比于异常行为重构较小的重构误差。

在基于重构的异常行为检测模型中，深度学习相关的重构模型也逐渐发展起来。2015年，文献[4]使用自编码(Auto-Encoder, AE)网络来重构原始数据，并在2016年提出使用全卷积神经网络(Fully Convolutional Neural Network, FCNN)进行快速异常行为检测[5]。Xu等人[6]提出外观及运动深度网络(Appearance and Motion DeepNet, AMDN)深度网络。2017年，Dimokranitou等人[7]分析了几种深度网络在异常行为检测方面的应用，并提出使用生成模型中的对抗自编码(Adversarial Auto-Encoders,

收稿日期：2018-05-07；改回日期：2019-01-29；网络出版：2019-02-20
*通信作者： 陈莹 chenying@jiangnan.edu.cn
基金项目：国家自然科学基金(61573168)
Foundation Item: The National Natural Science Foundation of China (61573168)

AAE)网络学习图片中的正常模式。袁静等人[8]在稀疏去噪自编码网络的基础上，通过增加梯度差约束条件，来提升自编码网络中的解码效果。

2016年，Hasan等人[9]利用卷积自编码(Convolutional Auto-Encoder, Conv-AE)构建异常行为检测模型，虽然网络输入为连续多帧，但是采用的是2D卷积，在第一层卷积后，时间信息会完全消失。2017年，Chong等人[10]在Hasan的基础上进行改进，考虑到Hasan的CNN全卷积网络没有考虑时间信息，将卷积LSTM层添加到Conv-AE网络的编码层和解码层之间，但是由于在卷积自编码网络的编码层后，输入图像的时间信息已大量损失，所以卷积LSTM层对获得时间信息的作用还是很小。因此，Chong的模型还是没有很好地建立时间模型。

经动作识别论文中双流(two-stream)模型[11]的启发，针对直接利用卷积自编码网络未能直接考虑时间信息的问题，本文提出基于贝叶斯融合的空-时流(Bayesian Fusion of Spatial-Temporal Stream, BFSTS)异常行为检测模型。双流分别称为空间流卷积网络(Spatial Convolutional Network, SCN)和时间流卷积网络(Temporal Convolutional Network, TCN)。空间流卷积网络直接采用卷积自编码网络对视频帧进行重构。在时间流卷积网络中，采用卷积LSTM编码-解码网络[12]对短期光流序列进行重构。最后，分别计算双流对应的重构误差，利用自适应阈值对重构误差图进行二值化，并采用贝叶斯融合方法对重构误差值进行融合得到最终的融合重构误差值，并进行异常行为判断。

2 整体框架

2.1 异常行为检测整体框架

基于贝叶斯融合的空-时流异常行为检测整体框架如图1所示。

视频一般由空间和时间成分组成。该异常行为检测模型的模型架构主要分为空间流卷积网络和时间流卷积网络。空间流卷积网络部分，单帧梯度作为输入，携带有目标外观信息，经过卷积自编码网络(Convolutional Auto-Encoder, Conv-AE)后，得到单帧梯度的重构梯度图；时间流卷积网络部分，以连续L帧光流图像作为输入，经过Conv-LSTM编码-解码，得到重构的L帧光流图像。具体的在3.1和3.2节中介绍。

训练阶段，以重构图像与输入图像的损失值为目标函数，分别对空间和时间网络部分进行训练，得到检测模型的参数。测试阶段，以单帧梯度和连续L帧光流图像分别作为空间和时间流卷积网络的输入，经过网络后，得到重构梯度图和重构的L帧光流图像；然后，计算空间流重构误差和时间流重构误差；最后，通过贝叶斯融合方式，将两者重构误差融合，最后，利用最终融合重构误差来判断异常行为。

2.2 网络输入

在空间流卷积网络中以单帧梯度图作为输入，需进行图像梯度计算。为减少数据维度，将图像转化为灰度图像。为将原始数据转化为适合模型输入的尺寸，将每帧尺寸调整为224×224。为保证空间流卷积网络的输入图像在同一尺度上，先减去全局均值图像，该均值图像通过平均训练集中每帧每个位置上的像素点值的方式得到，再将灰度图像归一化至0～1之间。然后根据式(1)计算单帧梯度图中每个像素点的梯度。

$$\left.\begin{aligned} G_x(x,y,t) &= [P(x+1,y,t)-P(x-1,y,t)]/2 \\ G_y(x,y,t) &= [P(x,y+1,t)-P(x,y-1,t)]/2 \end{aligned}\right\} \tag{1}$$

其中$G_x(x,y,t)$, $G_y(x,y,t)$分别表示像素点(x,y,t)的x,y方向梯度。$P(x+1,y,t)$, $P(x-1,y,t)$, $P(x,y+1,t)$, $P(x,y-1,t)$分别表示像素点$(x+1,y,t)$, $(x-1,y,t)$, $(x,y+1,t)$, $(x,y-1,t)$

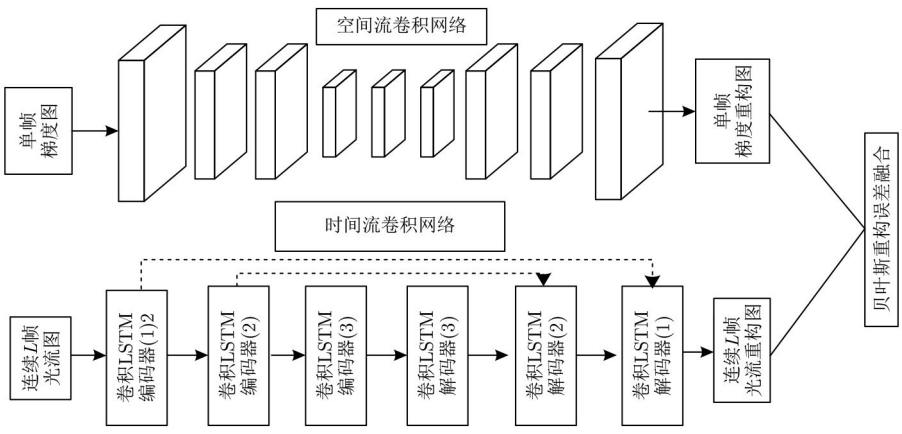


图 1 基于贝叶斯融合的时空流异常行为检测整体框架

的像素值。

在时间流卷积网络模型中，每帧图像的光流图根据文献[13]的计算方法进行计算。

3　基于贝叶斯融合的时空流异常行为检测模型

3.1　空间流卷积网络

如图2中所示为空间流卷积网络模型SCN的具体各层参数结构。与Conv-AE相同，SCN模型使用全卷积编码，包含编码层(左侧)和解码层(右侧)。由于经过全连接层后会失去空间信息，因此不包含全连接层。

解码层中包含了3层卷积层和2层池化层。输入图片尺寸为224×224，通道数为1。第一层卷积层(Conv_1)有256个滤波器，即输出尺寸为11×11，步长为4。其中，图2中每个卷积层的大小可表示卷积核尺寸×输入通道数×输出通道数(步长)，如Conv_1就可以表示为11×11×1×256(4)。经过卷积层_1层后，Conv_1的输出可表示为Conv_1层滤波器个数×卷积后图片尺寸，即256×55×55。

结合卷积层，池化层抽象了各种目的的激活，比如卷积层之后的平移不变性。池化操作中包含最大池化(max pooling)和平均池化(average pooling)，本文选择使用max pooling来保证平移不变性。第一层池化层(Poo1_1)采用max pooling，步长为2。经过Pool_1层后，图像尺寸为27×27，特征图数量为256。

Conv_2和Conv_3层分别有128和64个滤波器。最终，编码器生成64个尺寸为13×13的特征图。激活函数使用Relu函数，因为效果最好，并且使用也比较广泛。

解码层和编码层是对称的。解码层中反卷积层和反池化层中的滤波器数目和滤波器尺寸和编码层中卷积层和池化层的滤波器数目和滤波器尺寸是相互对应的。反卷积类似于卷积，是通过反向训练，达到使输出重构输入的效果。

由于使用最大池化会使反映特定位置的空间信息丢失，在解码层中使用反池化层。反池化层是池化层的反操作，重构激励的原尺度。比如，Unpool_2层利用Pool_2层得到的最大激励的位置信息和Deconv_3层对应位置上的激活信息来重构Pool_2的输入。

由最后一层的反卷积函数可得到输入G_i对应的重构图像$f_w(G_i)$，尺寸为1×224×224。为使$f_w(G_i)$尽可能地重构G_i，根据文献[9]，使用欧式距离损失函数作为空间流卷积网络的目标函数如式(2)

$$\hat{f}_{W_f}=\arg\min_{W_f}\frac{1}{2N}\sum_i\left\|G_i-f_{W_f}(G_i)\right\|_2^2 \tag{2}$$

其中，$\boldsymbol{W}_f$为各层参数，N为样本数量。

3.2　时间流卷积网络

由于空间流卷积网络作用于单独视频帧，采用网络对其重构可得到图像的空间信息表示。时间流卷积网络TCN采用Conv-LSTM编码器-解码器作为时间重构模型，在编码-解码过程中，Conv-LSTM编码器学习到能表示视频片段的固定长度特征表示。基于当前隐藏状态和前一刻的重构值，Conv-LSTM解码器使用该特征表示去重构视频片段。图3显示了Conv-LSTM编码-解码过程。

时间流卷积网络的输入是连续L帧视频帧$\boldsymbol{X}=\{\boldsymbol{x}^{(1)}\boldsymbol{x}^{(2)}\cdots\boldsymbol{x}^{(i)}\cdots\boldsymbol{x}^{(L)}\}$，其中$\boldsymbol{x}^{(i)}$表示每帧图像对应的光流，尺寸为224×224。$h_{\mathrm{E}}^{(i)}$为时刻$i(i\in\{1,2,\cdots,L\})$的编码器的隐藏状态，其中$h_{\mathrm{E}}^{(i)}\in\mathbb{R}^{224\times224\times c}$，$c$为编码器Conv-LSTM核中Conv-LSTM单元的数目。图3中每一行表示1层，共3层。左边列表示按时间序列输入，右边列表示按时间序列输出。$h_{\mathrm{E}}^{(i)}$和$\boldsymbol{x}^{(i)}$经过一个Conv-LSTM核后得到$h_{\mathrm{E}}^{(i+1)}$。当所有时刻$\boldsymbol{x}^{(i)}$训练完毕后，会得到最终状态$h_{\mathrm{E}}^{(L)}$。

解码时，编码器的最终状态$h_{\mathrm{E}}^{(L)}$会作为解码器

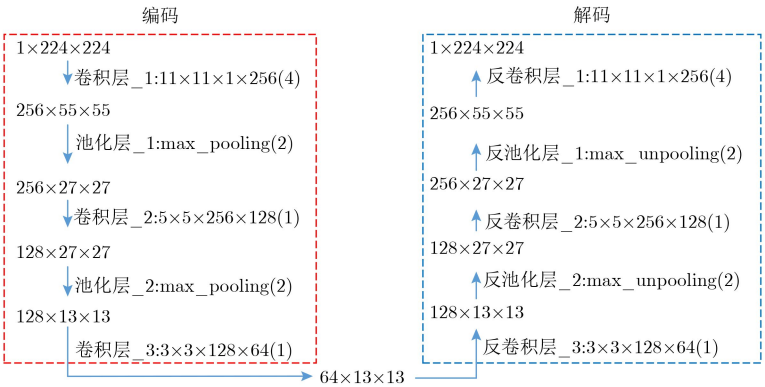


图 2 空间流卷积网络模型各层参数尺寸

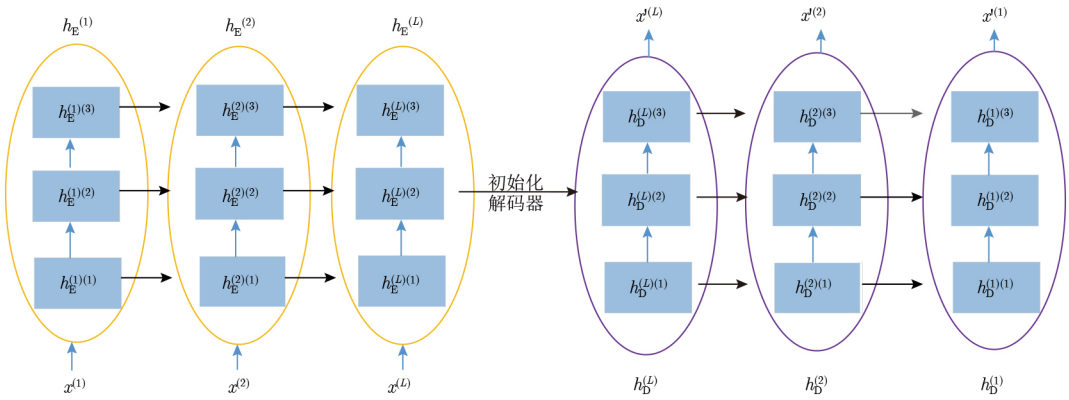


图 3 Conv-LSTM编码-解码过程

中$h_{\mathrm{D}}^{(L)}$的初始状态，即$h_{\mathrm{D}}^{(L)}=h_{\mathrm{E}}^{(L)}$。为了保证输出$\boldsymbol{x}'^{(i)}$维度和输入$\boldsymbol{x}^{(i)}$维度一致，使用1个卷积操作来计算$\boldsymbol{x}'^{(i)}$，即$\boldsymbol{x}'^{(i)}=\boldsymbol{W}_h*h_{\mathrm{D}}^{(i)}+\boldsymbol{b}_h$，其中权重矩阵$\boldsymbol{W}_h$的维度为$3\times3\times c\times L$，偏置$\boldsymbol{b}_h$的维度为$L$。解码时，时刻$i$对应的预测值$\boldsymbol{x}'^{(i)}$与其得到的隐藏状态$h_{\mathrm{D}}^{(i)}$被用来获得下一时刻的隐藏状态$h_{\mathrm{D}}^{(i-1)}$和预测值$\boldsymbol{x}'^{(i-1)}$。

经过编码-解码层后，将重构视频序列按时间反转得到有序的重构视频序列$\boldsymbol{X}'=\left\{\boldsymbol{x}'^{(1)}\boldsymbol{x}'^{(2)}\cdots\boldsymbol{x}'^{(i)}\cdots\boldsymbol{x}'^{(L)}\right\}$。为使$f_{\boldsymbol{W}_h}(\boldsymbol{X})$能够尽可能地重构$\boldsymbol{X}$，使$f_{\boldsymbol{W}_h}(\boldsymbol{X})$与$\boldsymbol{X}$之间的误差最小，目标函数可写为

$$\hat{g}_w=\arg\min_{W_h}\frac{1}{2M}\sum_{M}\sum_{i=1}^{L}\left\|x^{(i)}-x'^{(i)}\right\|_2^2 \tag{3}$$

其中，M为样本数量。

3.3 重构误差融合

在训练阶段，分别对空间流网络模型和时间流网络模型训练可得到模型参数$\boldsymbol{W}_f$和$\boldsymbol{W}_h$。在这些模型参数下，对测试视频进行异常判断。对于空间流模型而言，每像素的重构误差为$e_s(x,y,t)=\left\|P(x,y,t)-f_{W_f}(x,y,t)\right\|_2$，时间流网络模型的每像素的重构误差为$e_t(x,y,t)=\|F(x,y,t)-f_{W_h}(x,y,t)\|_2$，其中$F(\cdot)$为光流图。

传统基于加权平均的重构误差融合方法认为各原始重构误差图的重要性或贡献值没有差异，这种假设实际上有失合理性，对于不同图像上的检测结果不尽相同，性能也并不稳定。本文在显著性检测[14,15]融合思想的启发下，利用贝叶斯公式对重构误差进行融合。

假设已知空间流和时间流网络模型下的每帧图像对应的重构误差为e_s和e_t。首先，将通过空间模型得到的e_s作为贝叶斯先验概率，并用时间模型下的e_t计算观测似然概率。接着调换空间和时间下e_s和e_t的位置，将时间模型下的e_t作为贝叶斯先验概率，并根据空间模型下得到的e_s计算观测似然概率。最后将两种先验情况下得到的后验概率对应像素相加作为融合重构误差。基于贝叶斯的重构误差融合方法流程如图4所示。

具体地，当e_s为贝叶斯先验时，首先设置阈值，对e_s进行二值化操作。阈值选取的步骤：首先，可大致地获得图像对应的异常区域，计算该异常区域对应的均值(Means)；然后设置一个经验阈值(Threshold)，经验阈值设置为0.1；最后，将这个经验阈值与均值做比较，取较大的作为二值化操作的阈值。这样做的目的是为了防止当均值很小时，会将很多的非异常点划分进异常区域。

对重构误差图进行二值化后，可大致划分e_s为正常与异常区域，用A_s和NA_s表示，像素点总数记为N_{A_s}和N_{NA_s}。然后，计算e_t的重构误差在A_s和NA_s中的归一化分布直方图。像素点$z=(x,y)$的观测似然概率可用$e_t(z)$落在计算得到的直方图中的位值来表示，即

$$\left.\begin{aligned}p\left(e_t(z)\,|A_s\right)&=\frac{B_{A_s}\left(e_t(z)\right)}{N_{A_s}}\\ p\left(e_t(z)\,|\mathrm{NA}_s\right)&=\frac{B_{\mathrm{NA}_s}\left(e_t(z)\right)}{N_{\mathrm{NA}_s}}\end{aligned}\right\} \tag{4}$$

其中，$B_{A_s}\left(e_t(z)\right)$和$B_{\mathrm{NA}_s}\left(e_t(z)\right)$分别表示$e_t(z)$落在$A_s$和$\mathrm{NA}_s$区域的统计直方图中包含$e_t(z)$的位中的像素点个数。

那么，当e_s作为先验概率时，与之对应的后验概率可根据式(5)得到

$$p(A_s\,|e_t(z))=\frac{e_s(z)p\left(|e_t(z)\,|A_s\right)}{e_s(z)p\left(|e_t(z)\,|A_s\right)+\left(1-e_s(z)\right)p\left(e_t(z)\,|\mathrm{NA}_s\right)} \tag{5}$$

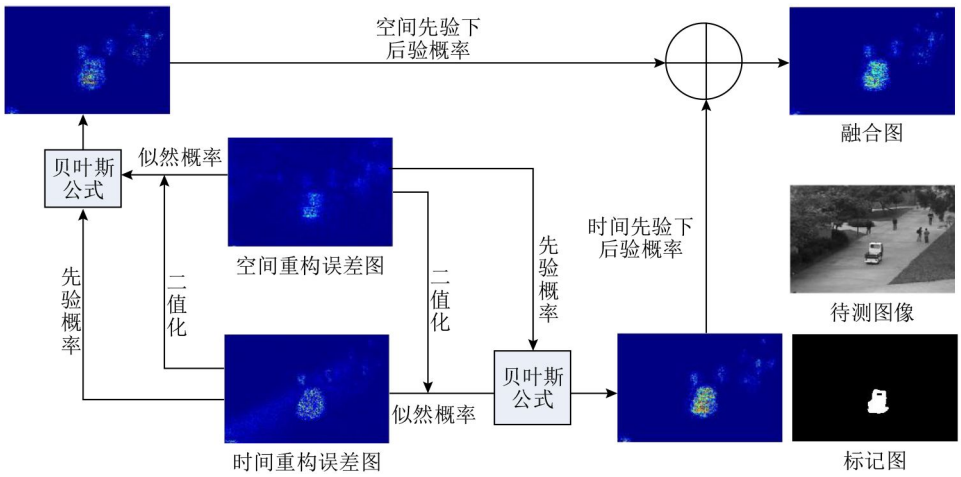


图 4 基于贝叶斯的时空流融合图

和计算e_s为贝叶斯先验时的后验概率一样，计算e_t为贝叶斯先验时的后验概率，即

$$p(A_t|e_s(z)) = \frac{e_t(z)p(e_s(z)|A_t)}{e_t(z)p(e_s(z)|A_t)+(1-e_t(z))p(e_s(z)|\mathrm{NA}_t)} \tag{6}$$

最后，基于贝叶斯公式将两个后验概率相加得到各像素的融合重构误差，即

$$\begin{aligned} e_f(z) &= e\left(e_t(z), e_s(z)\right) \\ &= p\left(A_s|e_t(z)\right)+p\left(A_t|e_s(z)\right) \end{aligned} \tag{7}$$

则每帧对应的重构误差和为$E=\sum_z e_f(z)$。

4　实验结果及分析

为了验证算法的有效性，分别在UCSD[16]和Avenue[1]数据库上进行实验。其中，UCSD数据库包含Ped1和Ped2数据库。Ped1数据库中行人走向或者远离摄像机的视频片断。其包含34个训练视频和36个测试视频，其中34个训练视频均为正常视频，36个测试视频中有10个视频有像素级别的标定。UCSD Ped2数据库是从侧面拍摄的，分辨率为320×240。其中，16个训练视频，12个测试视频，均有像素级别标定。Avenue数据库包含16个训练视频和21个测试视频，分辨率为640×360。包含的异常行为有随地扔纸、奔跑等。

实验仿真均在Matlab, python3.5, Tensorflow1.4, Window10, Intel(R) Xeon(R) i5处理器，1080Ti 11 GB GPU的计算机配置下实现。采用自适应随机梯度下降算法[17](Adaptive subGradient method, AdaGrad)对网络参数进行优化；初始化权重使用Xavier算法[18]。

4.1　规则分数

基于深度学习的方法中新增了关于异常事件统计的评价标准[13]。首先在每一帧需要计算规则分数

$$\left.\begin{aligned} S_a(t) &= \frac{E(t)-E(t)_{\max}}{E(t)_{\max}} \\ S_r(t) &= 1-S_a(t) \end{aligned}\right\} \tag{8}$$

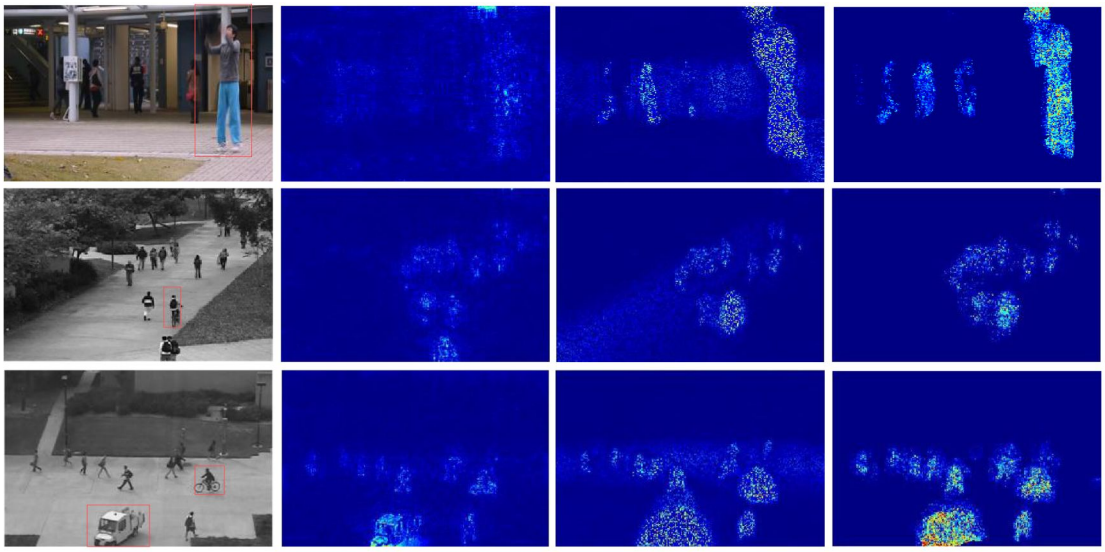
其中，$S_a(t)$为第t帧归一化的异常度，$S_r(t)$为第t帧规则分数。$S_r(t)$值越大，其对应的帧为正常的概率越大；$S_r(t)$值越小，其对应的帧为异常的概率越大。因此，需要找$S_r(t)$的局部最小来判断异常帧。基于深度学习的方法在实际应用中可通过对规则分数设置阈值判断视频中的行为是否异常。然而，与所有的二分类问题一样，不同阈值将带来不同的检测真正率和假正率，因此本文根据基于规则分数的ROC曲线和AUC对算法的性能进行评价。

4.2　重构图像显示

测试时，空间和时间流模型下分别得到重构误差图，为了显示融合后的重构误差更能突出并检测更多的异常行为，分别将空间流、时间流、空-时流贝叶斯融合后的重构图显示出来，如图5所示。

从图5中可以看出，空间流模型中异常行为和正常行为所对应的重构误差值之间不存在明显的差异。由于空间流模型的原始输入是目标的梯度信息，能够描述目标的轮廓，因此空间流模型下的重构图能较好地将目标的轮廓体现出来。相比于空间流模型，时间流模型能更好地突出异常行为与正常行为之间得差异，但是不能很好地描述目标轮廓。而在时空流贝叶斯融合重构图中，异常行为和正常行为的差异就很明显，即异常行为对应的区域颜色更亮，更突出。

从图中可以看出，图5(d)中的异常行为颜色更明显，且目标轮廓也较清晰。因此，基于贝叶斯融合的时空流异常行为检测模型能更好地检测异常行为，贝叶斯融合能使空间流模型和时间流模型下的检测结果相铺相成，突出并优化最终重构图。



(a) 原图 (b) 空间流模型重构图 (c) 时间流模型重构图 (d) BFSTS重构图

图 5 重构图显示

4.3 可视化规则分数

在重构误差的基础上计算规则分数后，分别绘制Avenue, Ped1, Ped2数据库视频的规则分数图，可直观地看出异常帧的位置，如图6所示，横坐标表示视频帧数，纵坐标表示每帧图像对应的规则分数。视频帧对应的规则分数越小，则该帧异常的可能性越大。为可视化规则分数与帧异常的关系，将真实标定(groud-truth)下的异常帧显示于规则分数图中，即图6中的红色区域。

从规则分数图中可以看出，几乎所有异常均产生强烈的向下的尖峰，对应着较低的规则分数。利用本文方法产生的规则分数较低的时间帧区域与真实异常发生的时间区域基本吻合，而真实正常行为时间区域对应的规则分数相对较高，即文中方法产生的规则分数的局部极小值与真实异常标定存在时间上的同步性。

4.4 基于规则分数的ROC曲线

将BFSTS融合后得到的基于规则分数的ROC曲线与3.1节中介绍的空间流卷积网络(SCN)及3.2节中介绍的时间流卷积网络(TCN)进行对比。Ped1和Ped2基于规则分数的ROC曲线的比较结果如图7所示。可以看出，相比于单纯采用SCN或TCN，本文所提出的BFSTS采用双流检测及融合策略能得到更高的检测率，即基于贝叶斯融合的时空流检测模型相比于单独使用空间模型或者时间模型有一定程度的提升。

4.5 方法性能比较

为了显示本文算法的先进性，将时空融合网络模型BFSTS与传统非深度方法MPPCA[16], HOFME[19], 以及基于深度学习的方法Conv-AE[9], ConvLSTM-AE[20], Unmasking[21]和Stack RNN[22]在基于规则分数的帧级别下的EER和AUC进行异常检测精度的比较，其中，BFSTS的时间流模型分别采用间断采样和连续采样两种方式提取3帧，即L=3。检测精度比较结果如表1所示。

从表1中可以看出，相比于传统非深度方法，基于深度学习的方法能够得到更好的检测精度。相比于其它深度检测方法，本文所提出的BFSTS由于融合了时间和空间的深度信息，因此获得了相对较好的检测结果，如相比于文献[9]提出的Conv-AE, BFSTS算法的AUC在Ped1及Ped2数据库上提升了约1%及4.7%。此外，可以看出无论BFSTS时间流采用间隔采样还是连续采样，其检测性能在整体上都优于其它方法，但由于相邻帧之间的关联性较大，通过相邻帧更能反应当前帧的特性，因此间隔采样效果要略低于连续采样。

5 结束语

为同时利用视频空间和时间信息来提升异常行为检测率，本文提出基于贝叶斯融合的时空流异常行为检测模型(BFSTS)。与其他异常行为检测算法相比，本文提出的贝叶斯融合的时空流异常行为检测算法可以更加准确地检测出异常帧，能进一步提升异常行为检测的EER和AUC。但还存在一定的实时性问题，本文是在决策层进行的重构误差图的融合，因此需要分别训练空间和时间流模型。异常行为检测在智能视频监控领域有着很高的实用价值，因此以后会在保证检测精度的同时，提高检测的实时性。

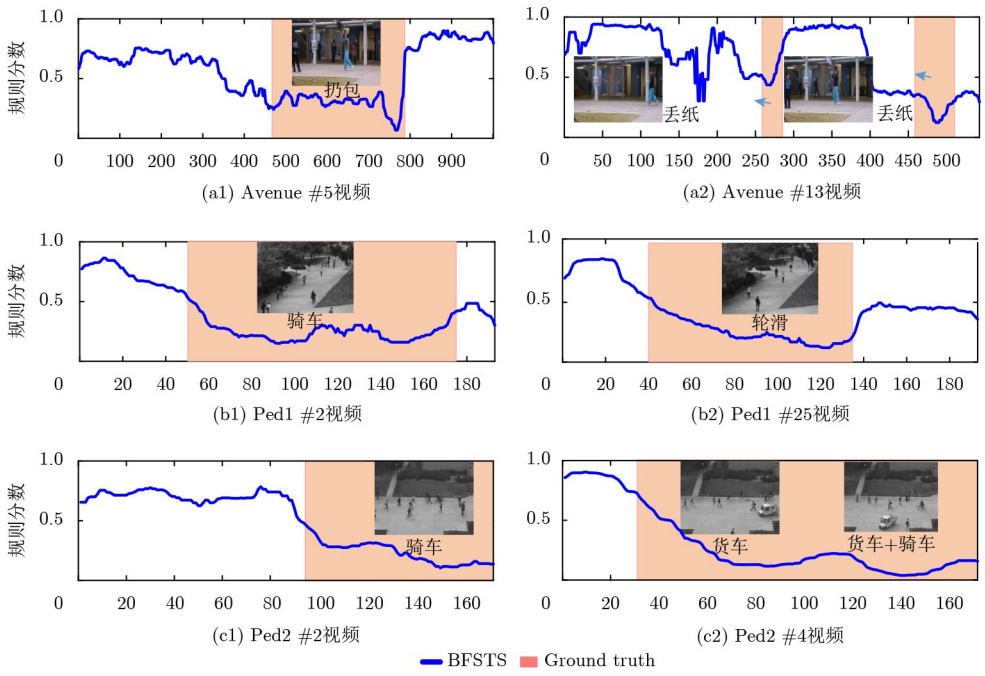


图 6 可视化规则分数图

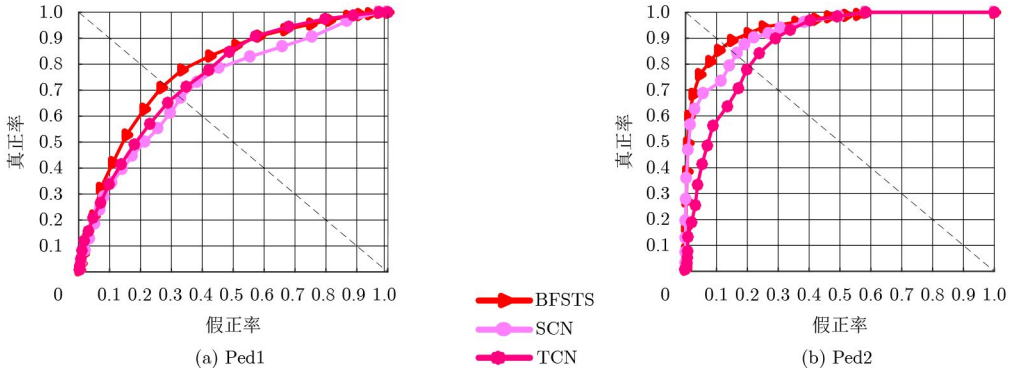


图 7 UCSD Ped1和Ped2数据库基于规则分数的ROC曲线

表 1　基于规则分数的帧级别下的EER和AUC比较(%)

方法	UCSD Ped1		UCSD Ped2	
	EER	AUC	EER	AUC
MPPCA+SF[16]	32.0	74.2	36.0	61.3
HOFME[19]	33.1	72.7	20.0	87.5
Conv-AE[9]	**27.9**	76.8	21.7	90.0
ConvLSTM-AE[20]	N/A	75.5	N/A	88.1
Unmasking[21]	N/A	68.4	N/A	82.2
Stack RNN[22]	N/A	N/A	N/A	92.2
BFSTS(间隔/连续采样)	28/**27.9**	76.5/**77.8**	16.0/**13.0**	92.7/**94.7**

参 考 文 献

[1] LU Cewu, SHI Jianping, and JIA Jiaya. Abnormal event detection at 150 FPS in MATLAB[C]. Proceedings of 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 2013: 2720–2727. doi: 10.1109/ICCV.2013.338.

[2] WEN Hui, GE Shiming, CHEN Shuixian, *et al*. Abnormal event detection via adaptive cascade dictionary learning[C]. Proceedings of 2015 IEEE International Conference on Image Processing, Quebec, Canada, 2015: 847–851. doi: 10.1109/ICIP.2015.7350919.

[3] GUO Huiwen, WU Xinyu, CAI Shibo, *et al*. Quaternion

discrete cosine transformation signature analysis in crowd scenes for abnormal event detection[J]. *Neurocomputing*, 2016, 204: 106–115. doi: 10.1016/j.neucom.2015.07.153.

[4] SABOKROU M, FATHY M, HOSEINI M, *et al*. Real-time anomaly detection and localization in crowded scenes[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA, 2015: 56–62. doi: 10.1109/CVPRW.2015.7301284.

[5] 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. doi: 10.1016/j.cviu.2018.02.006.

[6] XU Dan, YAN Yan, RICCI E, *et al*. Detecting anomalous events in videos by learning deep representations of appearance and motion[J]. *Computer Vision and Image Understanding*, 2016, 156: 117–127. doi: 10.1016/j.cviu.2016.10.010.

[7] DIMOKRANITOU A. Adversarial autoencoders for anomalous event detection in images[D]. [Ph.D. dissertation], Purdue University, 2017.

[8] 袁静, 章毓晋. 融合梯度差信息的稀疏去噪自编码网络在异常行为检测中的应用[J]. 自动化学报, 2017, 43(4): 604–610. doi: 10.16383/j.aas.2017.c150667.

YUAN Jing and ZHANG Yujin. Application of sparse denoising auto encoder network with gradient difference information for abnormal action detection[J]. *Acta Automatica Sinica*, 2017, 43(4): 604–610. doi: 10.16383/j.aas.2017.c150667.

[9] HASAN M, CHOI J, NEUMANN J, *et al*. Learning temporal regularity in video sequences[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 733–742. doi: 10.1109/CVPR.2016.86.

[10] CHONG Y S and YONG H T. Abnormal event detection in videos using spatiotemporal autoencoder[C]. Proceedings of the 14th International Symposium on Neural Networks, Hokkaido, Japan, 2017: 189–196. doi: 10.1007/978-3-319-59081-3_23.

[11] FEICHTENHOFER C, PINZ A, and ZISSERMAN A. Convolutional two-stream network fusion for video action recognition[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1933–1941. doi: 10.1109/CVPR.2016.213.

[12] SHI Xingjian, CHEN Zhourong, WANG Hao, *et al*. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]. Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 802–810.

[13] LIU Ce, FREEMAN W T, ADELSON E H, *et al*. Human-assisted motion annotation[C]. Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA, 2008: 1–8. doi: 10.1109/CVPR.2008.4587845.

[14] XIE Yulin, LU Huchuan, and YANG M H. Bayesian saliency via low and mid level cues[J]. *IEEE Transactions on Image Processin*, 2013, 22(5): 1689–1698. doi: 10.1109/TIP.2012.2216276.

[15] LI Xiaohui, LU Huchuan, ZHANG Lihe, *et al*. Saliency detection via dense and sparse reconstruction[C]. Proceedings of 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 2013: 2976–2983. doi: 10.1109/ICCV.2013.370.

[16] MAHADEVAN V, LI Weixin, BHALODIA V, *et al*. Anomaly detection in crowded scenes[C]. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 1975–1981. doi: 10.1109/CVPR.2010.5539872.

[17] DUCHI J, HAZAN E, and SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. *The Journal of Machine Learning Research*, 2011, 12: 2121–2159.

[18] GLOROT X and BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]. Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 2010: 249–256.

[19] WANG Tian and SNOUSSI H. Histograms of optical flow orientation for abnormal events detection[C]. Proceedings of 2013 IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Clearwater, USA, 2013: 45–52. doi: 10.1109/PETS.2013.6523794.

[20] LUO Weixin, LIU Wen, and GAO Shenghua. Remembering history with convolutional LSTM for anomaly detection[C]. Proceedings of 2017 IEEE International Conference on Multimedia and Expo, Hong Kong, China, 2017: 439–444. doi: 10.1109/ICME.2017.8019325.

[21] IONESCU R T, SMEUREANU S, ALEXE B, *et al*. Unmasking the abnormal events in video[C]. Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2895–2903.

[22] LUO Weixin, LIU Wen, and GAO Shenghua. A revisit of sparse coding based anomaly detection in stacked RNN framework[C]. Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 341–349. doi: 10.1109/ICCV.2017.45.

陈 莹：女，1976年生，教授，博士生导师，主要研究方向为信息融合、模式识别等.

何丹丹：女，1993年生，硕士生，研究方向为异常行为检测.