计算机科学
COMPUTER SCIENCE

基于非局部注意力生成对抗网络的视频异常事件检测方法

孙奇，吉根林，张杰

引用本文

孙奇，吉根林，张杰. 基于非局部注意力生成对抗网络的视频异常事件检测方法[J]. 计算机科学, 2022, 49(8): 172-177.

SUN Qi, JI Gen-lin, ZHANG Jie. Non-local Attention Based Generative Adversarial Network for Video Abnormal Event Detection[J]. Computer Science, 2022, 49(8): 172-177.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210600061

基于非局部注意力生成对抗网络的视频异常事件检测方法

孙 奇 吉根林 张 杰
南京师范大学计算机与电子信息学院/人工智能学院 南京 210023
(192202001@njnu.edu.cn)

摘 要 针对异常事件的不确定性,文中选择使用未来帧预测的方式对视频进行异常事件检测。通过正常样本对预测模型进行训练,使模型能够准确预测不包含异常事件的未来帧,但对于包含未知事件的视频帧,模型无法进行预测,利用生成对抗网络以及表观约束和运动约束对用于预测的生成器模型进行训练。为了减少相关目标特征丢失,提出了非局部注意力U型网络生成器(Nonlocal Attention Unet Generator,NA-UnetG)模型,提升了生成器的预测精度,同时提升了视频异常事件检测的准确度。通过公开数据集CUHK Avenue和UCSD Ped2对所提方法进行实验验证,实验结果表明,所提方法的AUC指标优于其他方法,*AUC*分别达到了83.4%和96.3%。

关键词:视频异常事件检测;生成对抗网络;视频预测;非局部注意力机制;深度学习
中图法分类号 TP391

Non-local Attention Based Generative Adversarial Network for Video Abnormal Event Detection

SUN Qi,JI Gen-lin and ZHANG Jie
School of Computer and Electronic Information/School of Artificial Intelligence,Nanjing Normal University,Nanjing,210023,China

Abstract As the uncertainty of abnormal events,the method of future frame prediction is chosen to detect abnormal events in video. The prediction model is trained with normal samples,so that the model can accurately predict the future frames without abnormal events. However,it cannot predict video frames with unknown events. Combining with apparent constraints and motion constraints,generative adversarial network is used to train the generator model for prediction. In order to reduce the loss of relative target features,a nonlocal attention Unet generator (NA-UnetG) model is proposed to improve the prediction accuracy of generator and the accuracy of abnormal video event detection. Experiments on datasets CUHK Avenue and UCSD Ped2 validate the effectiveness of the proposed method. The results show that the AUC of the proposed method is better than that of other methods,reaches 83.4% and 96.3%,respectively.

Keywords Video anomaly event detection,Generative adversarial network,Video prediction,Non-local attention mechanism,Deep learning

1 引言

视频异常事件检测是计算机视觉领域中一个有价值的研究方向。真实场景中,异常事件很少发生,并且在不同的场景下异常的定义也不同,因此异常具有稀疏性和不确定性[1]。异常的稀疏性体现在,正常运动模式的样本较多而存在异常的样本较少,正负样本的不均衡使得基于分类的异常检测方法存在一定的局限性。并且异常事件往往是不可预知的,基于分类思想去实现异常事件检测难以达到优秀的性能[2]。同时,异常的不确定性给视频异常检测任务带来了一定挑战。对于指定的视频异常往往需要特定的方法进行检测,但是该类方法不适用于通用型的异常检测任务。

在现实应用中,视频的正常事件和异常事件很难有明确的边界[3],视频中存在的异常事件可以定义为一种从未发生过的未知事件,即根据当前事件无法预测到的事件。因此,本文采用基于预测的异常检测方法,其主要思想是学习在正常运动模式下当前连续的视频帧到下一个视频帧的变化,从而在测试时通过生成当前视频帧的预测图像并将其与对应的视频真实图像作比较来进行异常事件检测。如果生成的预测图像与真实图像差异较大,则表明存在异常事件。生成对抗网络具有良好的序列预测能力,连续视频帧作为视频帧序列,可以通过生成对抗网络进行预测。

在目前的特征方法中,通用方法是使用单纯卷积神经网络对视频帧的高维特征信息进行提取[4],卷积神经网络使用

到稿日期:2021-06-04 返修日期:2021-09-11
基金项目:国家自然科学基金(41971343)
This work was supported by the National Natural Science Foundation of China(41971343).
通信作者:吉根林(glji@njnu.edu.cn)

卷积核对视频帧进行处理，主要关注图像局部的关联性，由低层神经元对局部的认知逐渐扩展到高层神经元对总体的认知。局部的认知对于整个图片来说，临近的像素之间的关联性强，而距离较远的像素之间的关联性较弱。并且，在采样和信息传递的过程中存在大量的信息丢失，因此单纯的卷积神经网络对于图像全局信息的提取存在局限性。非局部注意力机制关注较远距离的像素之间的关系[5]，能够提取全局范围内的局部目标特征的关联性。在进行特征重构时添加非局部注意力机制能够增加局部目标特征之间的关联性，保留局部特征中的有用信息和全局范围中特征之间的关联，从而提高对未来帧预测的准确性。本文的创新点如下：

(1)提出了一种基于注意力机制生成对抗网络的视频异常事件检测方法，采用生成对抗网络对视频的未来帧进行预测，利用正常样本对模型进行训练，使模型学习预测正常样本的未来帧，在测试阶段，通过对未来帧与真实帧之间的对比来判断视频中是否存在异常事件。

(2)提出了非局部注意力 U 型网络生成器，优化了视频预测生成器。添加注意力机制能使网络更加关注特征的全局性，为高层网络提供了更丰富的信息，使生成器生成更真实的未来帧。

2　相关研究

视频异常检测可以分为两个步骤，提取视频特征和训练异常检测模型。特征提取从最初的手工设计特征提取发展到基于深度学习的特征提取。手工设计特征包含方向梯度直方图[6](Histogram of Oriented Gradient，HOG)、光流直方图[7](Histogram of Oriented Optical Flow，HOF)等。Zhang 等[8]考虑到异常事件通常发生在前景区域，利用自适应混合高斯模型进行背景分割，提取视频前景进行人群异常事件检测。Adam 等[9]通过指数分布来表征光流的规则局部直方图。Mahadevan 等[4]提出混合概率主元分析(Mixture of Probabilistic PCA，MPPCA)的方法，学习正常模式光流特征的分布情况。Kim 等[10]利用混合动态纹理(Mixtures of Dynamic Textures，MDT)对视频中的局部异常进行检测。Mehran 等[11]使用社会力模型(Social Force，SF)检测拥挤场景下的异常。

深度学习的应用大大提升了特征提取的准确度，多种基于深度学习的视频异常检测方法的检测精度远远高于基于手工设计特征的方法。Xu 等[12]第一次将深度学习提取的特征用于视频异常检测，设计了多层卷积自编码器(Convolutional Auto-Encoder，Conv-AE)，用于提取特征。Hasan 等[13]利用卷积神经网络(Convolutional Neural Network，CNN)对空间特征的学习能力和长短期记忆网络(Long Short Term Memory，LSTM)对序列数据的敏感性，同时构建空间自编码器和时间自编码器，利用卷积 LSTM 自编码器(Convolutional LSTMs Auto-Encoder，ConvLSTM-AE)的方式提取时空信息。Luo 等[14]分别使用 CNN 和堆叠卷积神经网络(Stack Recurrent Neural Network，StackRNN)提取空间和事件上的特征，使用稀疏编码的方式实现视频异常检测。Morais 等[15]通过提取人体骨架特征分别获取人体的局部特征和全局特征，最终通过消息传递卷积神经网络(Message-Passing Encoder-Decoder Recurrent Neural Network，MPED-RNN)实现异常行为检测。

生成对抗网络[16](Generative Adversarial Network，GAN)是一种用于生成目标数据分布的框架，因其能够学习高度结构化的概率分布被广泛应用于序列预测。GAN 由生成器和判别器构成，训练过程即是轮流对抗训练生成器和判别器的过程，最终目标是使生成器生成的示例的分布与最优判别器判别的目标分布之间的差异最小化。GAN 被应用于图像生成和视频未来帧预测等方向。Reed 等[17]使用 GAN 实现将输入的文字描述生成图片。Villegas 等[18]使用表观判别器进行视频未来帧预测。Kwon 等[19]使用回顾性循环 GAN(Retrospective Cycle GAN)进行视频的未来帧预测。

3　视频异常检测方法

3.1　处理流程

本文的异常检测处理流程如图 1 所示，首先将视频样本拆分成视频帧的形式，然后开始训练非局部注意力 U 型网络生成器。对抗训练学习具有高度适应性的生成器，即在训练过程中令 NA-UnetG 和判别器相互对抗博弈。具体来说，在每一次训练中，将连续 4 帧(F_{t-4}，F_{t-3}，F_{t-2}，F_{t-1})输入生成器，并根据输入的连续帧预测接下来的一帧$\hat{F}_t$。在此过程中，判别器判断该视频帧$\hat{F}_t$是来自真实的视频还是由生成器生成。为了使 NA-UnetG 生成的未来帧更加逼真，同时使得判别器的鉴别能力更强，不但需要借助对抗目标函数来优化 NA-UnetG，考虑到视频数据的特殊性，还引入了额外的模块和约束条件。

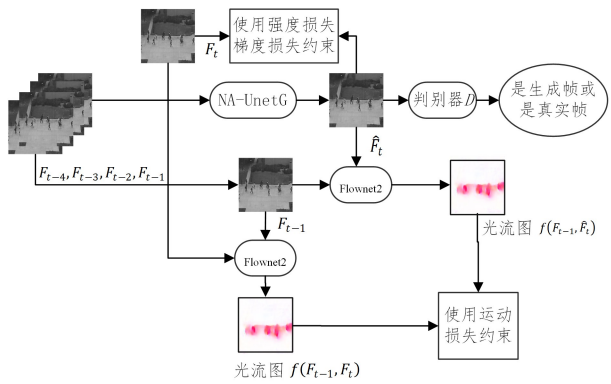


图 1　训练阶段的处理流程

Fig. 1　Processing flow in training stage

在 NA-UnetG 训练时加入表观约束条件和运动约束条件。表观约束条件是对图像空间上的表观信息进行约束，如两帧之间的欧氏距离和两帧之间的梯度损失；运动约束是为了令生成器在预测时能够考虑视频中的时间动态性特征；而光流图针对性地表达了前后帧之间的目标在空间中的运动形态随着时间的变化，因此使用 Flowenet2 模型[20]来产生光流

图，以获取帧与帧之间的时空动态特征。在此基础上，分别计算生成的未来帧与其前一帧之间的光流图和真实帧与其前一帧之间的光流图，并约束生成器能够生成符合真实运动特征的视频帧。为此，执行了 NA-UnetG 和判别器之间的光流图对抗训练，以提升生成器拟合真实未来帧的分布能力。

在测试阶段，如图 2 所示，将测试数据集中的连续 4 帧（F_{t-4}，F_{t-3}，F_{t-2}，F_{t-1}）输入到训练完成的 NA-UnetG 中，NA-UnetG 学习到的正常视频帧的分布方式可以对正常连续视频帧进行预测，生成未来帧$\hat{F}_t$，当出现意料之外的事件时，NA-UnetG 无法对包含未知事件的视频帧进行预测。因此，将生成的未来帧$\hat{F}_t$与真实帧F_t进行对比，如果差别大，则判定视频帧为异常帧。

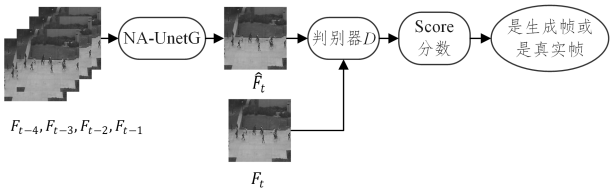


图 2 测试阶段的处理流程

Fig. 2 Processing flow in test stage

3.2 NA-UnetG 网络的结构

NA-UnetG 结构如图 3 所示。Unet 网络[21]类似于自编码器结构，通过卷积、下采样操作获得低维特征，再通过上采样与卷积操作得到高维的图像。与自编码器不同的是，为了获得更多的上下文语义信息，Unet 网络在网络结构中添加了一个跳跃连接（Skip Connection）来获取相同层次的特征，这时的跳跃连接选择加入非局部注意力机制[22]，以突出特征图中的显著特征，增加远距离像素之间的关联性，更加准确地获取注意特征，从而消除跳跃连接中的噪声和不相关的信息。

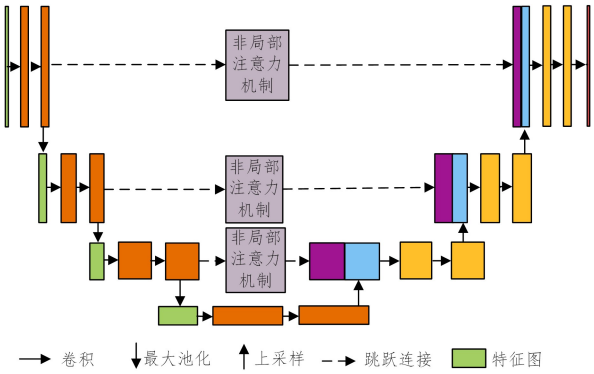


图 3 NA-UnetG 的结构

Fig. 3 Structure of NA-UnetG

生成器的结构类似于自编码器的结构，编码器部分通过卷积和最大池化得到特征的深层次表示。在这一过程中，卷积核的大小为 3×3，每一层经过两次卷积，再进行最大池化操作。池化操作后，特征的长和宽均减小至原尺寸的 1/2，且通道数不变，迭代进行 4 层操作。编码器部分使用反卷积、上采样以及连接的方式生成高维度特征。上采样操作后，特征图的长和宽均增大到原尺寸的 2 倍，通道数不变。连接是将本层通过非局部注意力机制模型处理后的特征图和下一层经过上采样得到的特征图进行融合，对得到的特征图继续做反卷积、上采样和连接操作，最终生成未来帧。

非局部注意力模型通过计算特征中每个位置的加权和，来增加全局特征之间的关联性，在较远距离相关联的特征像素点之间建立依赖联系，最终得到较远距离的显著性的相关目标的特征图。非局部注意力机制的结构如图 4 所示。首先输入一个特征图，特征图 X 经过不同的 1×1 的卷积分别生成 3 个特征图 $\boldsymbol{M}_1$，$\boldsymbol{M}_2$，$\boldsymbol{M}_3$，生成特征图的大小同原特征图的大小相同。对特征矩阵 $\boldsymbol{M}_1$ 和 $\boldsymbol{M}_2$ 的转置进行乘法运算，得到的矩阵结果经过 Softmax 操作得到空间注意力特征矩阵 $\boldsymbol{A}$，$\boldsymbol{A}$ 矩阵中每一个元素 $\boldsymbol{A}_{ji}$ 代表第 j 个位置和第 i 个位置之间的关联性，如果这两个位置之间的关联性越大，则相应的该元素值越大。然后，将特征图 $\boldsymbol{M}_3$ 与经过转置的注意力特征矩阵 $\boldsymbol{A}$ 进行相乘运算，得到全局的注意力特征图 $\boldsymbol{M}$，与初始的特征图 $\boldsymbol{X}$ 相加，最终得到空间注意力特征图 $\boldsymbol{Z}$。非局部注意力机制的公式完整定义为：

$$y_i=\frac{1}{C(\boldsymbol{X})}\sum_{\forall j}f(x_i,\hat{y}_j)g(\hat{y}_j) \tag{1}$$

$$Z_i=\boldsymbol{A}_i y_i+x_i \tag{2}$$

其中，$f(x,y)$函数用于计算 x 位置和所有可能与之相关的位置 y 之间的关联性；$g(x)$函数用于计算输入特征图在 x 位置的表示；$C(x)$用于对累加后的结果做归一化操作；y_i 是非局部特征图。最终完整的非局部化注意力特征如式(2)所示，每个特征图与经过非局部化的特征图相乘，最后增加一个残差结构 x_i，残差网络使得其可以任意连接到成型的网络中，得到空间注意力特征图 $\boldsymbol{Z}$。

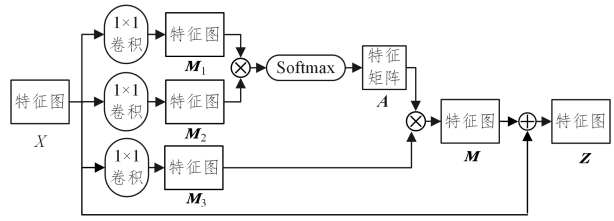


图 4 非局部注意力模型

Fig. 4 Model of nonlocal attention

3.3 约束条件

为了使 NA-UnetG 生成的未来帧与真实帧尽可能相似，同时使用表观约束和运动约束。表观约束是将生成的未来帧同对应时刻的真实帧在表观特征上进行对比。其中表观约束又分为强度约束和梯度约束，强度约束是控制两个视频帧中的像素帧间的欧氏距离，强度损失的损失函数如式(3)所示：

$$L_{\text{int}}(\hat{I},I)=\|\hat{I}-I\|_2^2 \tag{3}$$

梯度约束[23]的目的是保证生成帧与真实帧之间的梯度相近，其中 i,j 表示视频帧像素的横纵坐标，通过缩小生成帧中像素点同相邻像素点之间的距离和真实帧中像素点同相邻像素点之间的距离，使得生成的未来帧更加清晰。梯度损失的损失函数如式(4)所示：

$$L_{\text{gd}}=\sum_{i,j}(\|\,|\hat{I}_{i,j}-\hat{I}_{i-1,j}|-|I_{i,j}-I_{i-1,j}|\,\|_1+\|\,|\hat{I}_{i,j}-$$

$$\hat{I}_{i,j}|-|I_{i,j}-I_{i,j-1}|\|_1) \tag{4}$$

表观约束只是针对图像的表观特征进行约束，但对于视频来说，目标是有动作的，只依靠表观特征无法对视频的运动进行预测，因此需要同时考虑时间和空间维度上的特征。光流法可以提取到前后两帧之间像素点的运动情况，可以作为提取视频运动特征的工具。利用 Flownet2 网络提取前后两帧的运动特征，即前后两帧之间的光流图，比较生成的未来帧同上一时刻的真实帧之间的光流图和本时刻的真实帧同上一时刻的真实帧之间的光流图，运动损失公式如式(5)所示，此处直接使用训练完成的 Flownet2 模型得到前后帧的光流图。

$$L_{op}=\|f(\hat{I}_{t+1},I_t)-f(I_{t+1},I_t)\|_1 \tag{5}$$

3.4　目标函数

本文的总体框架基于生成对抗网络，生成对抗网络包含生成器和判别器，训练判别器的目标是能够准确判断视频帧是生成帧还是真实帧，固定生成器的生成能力，将生成帧判断为 0，真实帧判断为 1。判别器的损失函数如式(6)所示：

$$L_{adv}^{D}(\hat{I},I)=\sum_{i,j}\frac{1}{2}L(D(I)_{i,j},1)+\sum_{i,j}\frac{1}{2}L(D(\hat{I})_{i,j},0) \tag{6}$$

训练生成器的目标是使生成判别器无法判断生成的帧是真实帧还是生城帧，这时需要固定判别器的判别能力，使生成器生成的未来帧被判别器判断为 1。生成器的损失函数如式(7)所示：

$$L_{adv}^{G}(\hat{I})=\sum_{i,j}\frac{1}{2}L(D(\hat{I})_{i,j},1) \tag{7}$$

对于整个模型来说，训练生成器的目标函数是生成器的损失函数结合表观约束和运动约束条件。目标函数如式(8)所示：

$$\hat{G}=\min_{G}\lambda_{int}L_{int}+\lambda_{gd}L_{gd}+\lambda_{op}L_{op}+\lambda_{adv}L_{adv}^{G} \tag{8}$$

训练判别器时的损失函数如式(9)所示：

$$\hat{D}=\min_{D}L_{adv}^{D}(\hat{I}_{t+1},I_{t+1}) \tag{9}$$

3.5　异常事件检测方法

生成器能够预测出包含正常行为视频的未来帧，但无法预测包含异常事件视频的未来帧。NA-UnetG 完成未来帧预测后，将生成的未来帧与真实的未来帧进行对比，如果差异大，则将对应的真实的未来帧判定为包含异常事件的视频帧。通过判断两帧之间的峰值信噪比(Peak Signal to Noise Ratio，PSNR)来确定两帧之间的差异，其计算原理如式(10)所示。最终对于视频帧的判定使用 $Score(t)$ 作为规则分数，如式(11)所示，规则分数是将两帧的 PSNR 归一化，如果对应时刻的未来帧的 $Score(t)$ 大于阈值，则判定这一时刻的视频帧中不包含异常事件，否则判定这一时刻的视频帧中包含异常事件。非局部注意力生成对抗网络视频异常检测算法的训练过程如算法 1 所示。

$$PSNR(I,\hat{I})=10\log_{10}\frac{(\max_I)^2}{\frac{1}{N}\sum_{i=0}^{N}(I_i-\hat{I}_i)^2} \tag{10}$$

$$Score(t)=\frac{PSNR-\min_t PSNR}{\max_t PSNR-\min_t PSNR} \tag{11}$$

算法 1　训练非局部注意力生成对抗网络视频异常检测算法

输入：训练数据集包含的视频帧数 framenumber，NA-UnetG 的学习率 p_G，判别器的学习率 p_D，权重 λ_{adv}，λ_{int}，λ_{gd}，λ_{op}

输出：模型 NA-UnetG(Ped2)，NA-UnetG(Avenue)，参数 ω_D，ω_G

1. While 生成器和判别器的参数没有收敛 do

2.　{For i=0 to framenumber do

3.　$\omega_D=\omega_D-p_D\frac{\partial L_{adv}^{D}(\hat{I}_i,I_i)}{\partial\omega_D}$;} /* 固定生成器参数，更新判别器参数 */

4.　{For j=0 to framenumber do

5.　$\omega_G=\omega_G-p_G(\lambda_{adv}\frac{\partial L_{adv}^{G}(\hat{I}_j,I_j)}{\partial\omega_G}+\lambda_{int}\frac{\partial L_{int}(\hat{I}_j,I_j)}{\partial\omega_G}+\lambda_{gd}\frac{\partial L_{gd}(\hat{I}_j,I_j)}{\partial\omega_G}+\lambda_{op}\frac{\partial L_{op}(\hat{I}_j,I_j)}{\partial\omega_G})$;/* 固定判别器参数，更新生成器参数 */} }

6. return 参数 ω_D，ω_G

4　实验与结果分析

4.1　实验设置

本文分别在两个公开数据集 UCSD Ped2[24] 和 CUHK Avenue[25] 上进行实验验证，证明了算法的正确性和有效性。UCSD Ped2 数据集是在人行道俯视角度拍摄的行人走动视频，其中异常事件包含非行人实体(如机动车和自行车)进入人行道、行人快速通过人行道等。CUHK Avenue 数据集是在地铁口人行道平视角度拍摄的视频，其中异常事件包括行人快速通过人行道、行人抛掷物体、行人行进方向错误等。用于模型训练的数据集视频中不包含异常事件，用于测试的数据集视频帧中包含正常事件与异常事件。

实验使用 NVIDIA GeForce RTX3080 进行训练和测试，使用 PyTorch 框架构建相应的模型。本文实验使用基于 Adam 随机梯度下降法来优化参数，交替更新生成器和鉴别器的参数。生成器和判别器的学习率分别设置为 0.0001 和 0.00002，测试通过 ROC 曲线的曲线下面积，即 AUC(Area Under the Curve)对模型检测性能进行评判。

4.2　消融实验

为了证明添加非局部注意力机制能够提高模型的视频异常事件检测能力，本文进行了消融实验。如表 1 所列，添加非局部注意力机制的检测模型在 Avenue 和 Ped2 数据集上的 AUC 结果均高于没有添加非局部注意力机制的检测模型。实验结果证明，增加了非局部注意力机制的模型的预测能力更强，检测帧级别异常事件的准确率更高。

表 1　检测模型添加/未添加非局部注意力机制在 Ped2 和 Avenue 数据集上的 AUC 值

Table 1　AUC for detection model with/without attention mechanisms in Ped2 and Avenue datasets

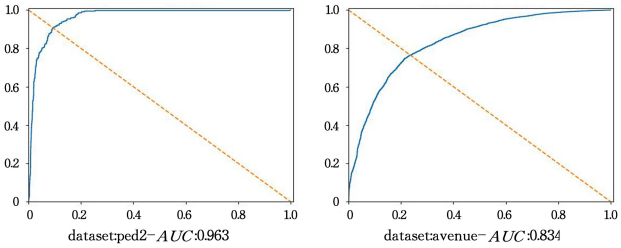
(单位：%)

	UCSD Ped2 *AUC*	CUHK Avenue *AUC*
Without non-local attention mechanisms	95.0	83.0
With non-local attention mechanisms	**96.3**	**83.4**

4.3　实验结果与分析

本节比较了基于手工特征提取的视频异常检测方法[9-11]和基于深度学习的视频异常检测方法[12-14,22,25]的实验结果。

将表 2 中的方法在相同的实验环境下对公开数据集 UCSD Ped2 和 CUHK Avenue 分别进行检测，以 AUC 作为评价指标，实验结果如表 2 所列。基于非局部注意力生成对抗网络的视频异常检测方法在 Ped2 和 Avenue 数据集上的 ROC 曲线如图 5 所示，横坐标表示异常数据被判断为正常的概率(False Positive Rate，FPR)，纵坐标表示正常数据被判断为异常的概率(True Positive Rate，TPR)。异常检测的最理想情况是 *TPR* 等于 1，对应的 *FPR* 等于 0，因此曲线越接近(0，1)点，对应方法的检测性能越好。本文方法在 Avenue 数据集 *AUC* 达到了 96.3%，在 Ped2 数据集 *AUC* 达到了 83.4%。由表 2 可知，基于深度学习的视频异常检测方法的性能优于基于手工特征的方法，对于深度学习方法，文献[22]中的方法未使用非局部注意力机制对未来帧进行预测，缺少对目标相关特征的提取，预测未来帧的能力低于本文方法。由此可见，基于非局部注意力生成对抗网络的异常检测方法能更准确地预测视频未来帧，从而有效提升视频异常事件检测的准确度。



(a) dataset：ped2-AUC：0.963　　(b) dataset：avenue-AUC：0.834

图 5　Ped2 和 Avenue 数据集上的 ROC 曲线图

Fig. 5　ROC curves in Ped2 andAvenue datasets

表 2　本文方法与其他异常检测方法的比较

Table 2　Comparison of the proposed method with other anomaly detection methods

(单位：%)

Methods	UCSD Ped2 *AUC*	CUHK Avenue *AUC*
MPPCA[7]	69.3	—
SF[11]	55.6	—
SF+MPPCA[10]	61.3	—
MDT[10]	82.9	—
ConvAE[13]	85.0	80.0
CoveLSTM-AE[13]	88.1	77.0
StackRNN[14]	92.2	81.7
MemAE[25]	94.1	83.3
FFP[23]	95.0	83.0
Our Method	**96.3**	**83.4**

结束语　本文提出了基于非局部注意力生成对抗网络的方法，以进行视频异常事件检测。训练模型的过程是对抗训练 NA-UnetG 和判别器的过程。NA-UnetG 关注全局中的相关特征，学习正常模式下视频帧中目标特征的分布情况，训练过程结合表观约束和运动约束，使预测的未来帧更加准确。测试时将连续视频帧输入到训练完备的 NA-UnetG 中，由于生成模型没有学习过异常模式，无法正确预测包含异常事件的视频帧，因此可以通过生成的未来帧与真实帧的差异在帧级别进行视频异常事件检测。

参 考 文 献

[1] WANG Z G, ZHANG Y J. Anomaly detection in surveillance videos: A survey [J]. Journal of Tsinghua University (Science and Technology), 2020, 60(6): 518-529.

[2] COLQUE R V H M, CAETANO C, DE ANDRADE M T L. Histograms of Optical Flow Orientation and Magnitude to Detect Anomalous Events in Videos [C] // The 28th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE Computer Society. 2015: 673-682.

[3] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: A survey [J]. ACM Computer, 2009, 41(3): 1-15.

[4] MAHADEVAN V, LI W X. Anomaly detection in crowded scenes [C] // The twenty-third IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2010: 1975-1981.

[5] WANG X L, GIRSHICK R B, GUPTA A, et al. Non-Local neural networks [C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2018: 7794-7803.

[6] DALA N, TRIGGS B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2005: 886-893.

[7] DALA N, TRIGGS B. Human detection using oriented histograms of flow and appearance [C] // The 9th European Conference on Computer Vision. Lecture Notes In Computer Science, 2006: 428-441.

[8] ZHANG J Y, XIE W X, ZHI K L. Abnormal crowd behavior detection based on motion effect map features of moving foregrounds [J]. Journal of Signal Processing, 2018, 34(3): 296-304.

[9] ADAM A, RIVLIN E, SHIMSHONI I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. IEEE Trans. Pattern Anal, 2008, 30(3): 555-560.

[10] KIM J, GRAUMAN K. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates [C] // 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2009: 2921-2928.

[11] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model [C] // Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2009: 935-942.

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

[13] HASAN M, CHOI J, NEUMANN J, et al. Learning temporal regularity in video sequences [C] // Proceedings of the Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 733-742.

[14] LUO W X，LIU W，GAO S H. A revisit of sparse coding based anomaly detection in stacked RNN framework[C]// IEEE International Conference on Computer Vision. IEEE Computer Society，2017：341-349.

[15] MORAIS R，LE V，TRAN T，et al. Learning regularity in skeleton trajectories for anomaly detection in videos[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2019：11996-12004.

[16] GOODFELLOW I，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial nets[C]// Neural Information Processing Systems 27，NIPS. 2014：2672-2680.

[17] REED S，AKATA Z，YAN X C，et al. Generative adversarial text to image synthesis[C]// The 33nd International Conference on Machine Learning. JMLR Workshop and Conference Proceedings，2016：1060-1069.

[18] VILLEGAS R，YANG J M，ZOU Y L，et al. Learning to generate long-term future via hierarchical prediction[C]// The 34th International Conference on Machine Learning. PMLR，2017：3560-3569.

[19] KWON Y H，PARK M. Predicting future frames using retrospective cycle GAN[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2019：1810-1820.

[20] ILG E，MAYER N. FlowNet 2.0：evolution of optical flow estimation with deep networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2017：1647-1655.

[21] RONNEBERGER O，FISC HER P，BROX T. U-Net：convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science，2015：234-241.

[22] WANG X B，GIRSHICK R，GUPTA A，et al. Non-Local Neural Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2018：1768-1777.

[23] LIU W，LUO W X，LIAN D Z，et al. Future frame prediction for anomaly detection－A new baseline[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society，2018：6536-6545.

[24] LU C W，SHI J P，JIA J Y. Abnormalevent detection at 150 FPS in MATLAB[C]// IEEE International Conference on Computer Vision. IEEE Computer Society，2013：2720-2727.

[25] GONG D，LIU L Q，LE V，et al. Memorizing normality to detect anomaly：memory-augmented deep autoencoder for unsupervised anomaly detection[C]// IEEE/CVF International Conference on Computer Vision. IEEE Computer Society，2019：1705-1714.

SUN Qi，born in 1995，master. Her main research interests include big data analysis and mining technology.

JI Gen-lin，born in 1964，Ph.D，professor. His main research interests include big data analysis and mining technology.

（责任编辑：喻藜）