基于双鉴别器和伪视频生成的视频异常检测方法

郭方圆，吉根林

引用本文

郭方圆，吉根林. 基于双鉴别器和伪视频生成的视频异常检测方法[J]. 计算机科学，2024，51(8)：217-223.

GUO Fangyuan, JI Genlin. Video Anomaly Detection Method Based on Dual Discriminators and Pseudo Video Generation [J]. Computer Science, 2024, 51(8): 217-223.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600148

基于双鉴别器和伪视频生成的视频异常检测方法

郭方圆　吉根林
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
(212202032@njnu.edu.cn)

摘　要　在无监督的视频异常检测任务中,通常使用深度自编码器在仅包含正常事件的数据集上进行训练,并根据重构(预测)误差来识别异常帧。然而,这种假设在实践中并不总是成立,有时自编码器对异常事件也可以进行很好的重构(预测),从而导致异常的误检测。为了解决这一问题,提出了一种基于双鉴别器和伪视频生成的视频异常检测方法,通过鉴别器和生成器之间的对抗训练来提高生成模型对正常帧的预测能力,并抑制生成模型对伪视频帧的预测能力。此外,在生成模型中引入协调注意力,以进一步提升模型的生成能力。同时,将以往方法中的预测未来帧改为预测中间帧,有利于模型学习前向和后向的运动信息,从而提升模型的检测性能。在公开数据集 UCSD Ped2 和 CUHK Avenue 上进行实验,结果表明,AUC 值在两个公开数据集上分别达到了 98.6%和 85.9%,相比其他视频异常检测方法,所提方法可显著提高视频异常检测的性能。

关键词:视频异常检测;深度学习;生成对抗网络;伪视频;预测
中图分类号　TP183

Video Anomaly Detection Method Based on Dual Discriminators and Pseudo Video Generation

GUO Fangyuan and JI Genlin
School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China

Abstract　In unsupervised video anomaly detection tasks, deep autoencoders are typically trained on datasets containing only normal events and use reconstruction(prediction) error to identify anomalous frames. However, this assumption does not always true in practice because sometimes autoencoders can reconstruct(predict) anomalous events well, leading to false alarms. To address this issue, this paper proposes a video anomaly detection method based on dual discriminators and pseudo video generation, which enhances the generation model's prediction capability of normal frames and suppresses its prediction capability of pseudo video frames through adversarial training between the discriminator and the generator. Moreover, the introduction of coordinated attention in the generation model further improves its detection performance. Additionally, by predicting intermediate frames instead of future frames in previous methods, the model can learn forward and backward motion information, which further enhances its detection performance. Experimental results on the publicly available datasets UCSD Ped2 and CUHK Avenue demonstrate that the proposed method achieves AUC values of 98.6% and 85.9%, respectively, outperforming other video anomaly detection methods significantly.

Keywords　Video anomaly detection, Deep learning, Generative adversarial network, Pseudo-video, Prediction

1　引言

视频异常检测是一项重要的技术,在监控、安防等领域得到了广泛的应用。视频异常检测旨在通过对视频中的运动、形状、颜色等特征进行分析,检测出视频中存在的异常事件,例如盗窃、交通事故、火灾等。这些异常事件可能会导致重大的人员伤亡和财产损失,因此,视频异常检测技术的研究具有重要的实际意义。近年来,随着计算机视觉和深度学习技术的不断发展,视频异常检测的研究也得到了进一步的推进。

基于深度学习的无监督方法,是自动检测监控视频中异常事件的常用方法之一。其中,重构和预测是两种常见的方法。重构指用正常帧训练一个生成器对当前帧进行重构,预测指用正常序列的前几帧训练生成器对下一帧进行预测。这两种方法训练得到的生成器能够很好地重构(预测)正常帧,而无法重构(预测)异常帧。在测试阶段,通过生成器的重构(预测)误差来区分正常和异常。然而,文献[1-3]指出,这些方法易受到生成模型泛化性的影响,导致生成器对异常帧也能进行很好的重构(预测),从而无法区分正常帧和异常帧。

为了解决这一问题,一些研究者提出了采用记忆力模块的方法。例如,Gong 等[4]提出了 MemAE 模型,该模型在训练

到稿日期:2023-06-19　返修日期:2023-11-24
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)

过程中通过记录典型的正常模式，来获得较低的平均重构误差，但这种方法需要计算大量的参数来学习正常模式。最近，研究者提出了一种更简单的方法，通过构造不同于正常数据的伪视频数据（异常），让模型同时训练伪视频数据和正常数据，限制模型对异常帧的重构（预测）能力。例如，Zaheer 等[3]从正常数据中融合两个随机图像来生成外观异常数据，限制生成器对异常的重构能力。但此方法需要两个不同训练阶段产生的生成模型来合成伪视频数据。与此方法不同，本文提出通过伪视频合成器和具有两个鉴别器的生成对抗网络来限制生成器对异常的预测能力，以端到端的方式训练生成器。

与以往预测未来帧的方法不同，本文采用的是预测中间帧的方法，生成器能够更好地学习前向和后向的运动信息。此外，通过训练由正常序列进行跳帧处理而构造的伪视频序列，以限制生成器对异常的预测能力。为进一步限制生成器对异常的预测能力，提出了一种具有双鉴别器的生成对抗网络，其中鉴别器 D_1 用于提高生成器对正常帧的预测能力，鉴别器 D_2 用于限制生成器对异常帧的预测能力。本文的主要贡献如下：

1）提出了基于双鉴别器和伪视频生成的视频异常检测方法，使用具有两个鉴别器的生成对抗网络对视频的中间帧进行预测，两个鉴别器分别用于提升生成器预测正常序列中间帧的能力，以及限制生成器预测伪视频序列中间帧的能力，显著提高了视频异常检测的准确率。

2）提出了 CA-Unet 生成器，通过添加协调注意力帮助生成器更加精准地定位和识别感兴趣的目标，从而提升模型的检测性能。

2 相关工作

基于预测的方法是解决视频异常检测任务的主要方法之一。该方法利用正常的视频序列进行训练，使模型能够根据前几帧预测未来帧，在测试阶段，模型预测的帧和真实帧之间误差较大的帧会被认为是异常帧。Liu 等[5]提出了一种利用预测误差作为异常指标的方法来预测未来帧。Song 等[6]提出了一种利用长短时记忆网络进行异常检测的方法。然而，由于自编码器具有强大的生成能力，对异常帧的预测也可能非常准确，从而导致对异常的误判。为解决这个问题，本文提出了一种基于双鉴别器和伪视频生成的视频异常检测方法，限制生成器对异常帧的预测能力，从而提升模型的检测性能。

近年来，基于生成对抗网络（GAN）的视频异常检测方法备受关注，GAN 由生成器（Generator，G）和鉴别器（Discriminator，D）组成，通过对抗训练使生成器生成的图像更接近真实图像。Isola 等[7]提出了一种使用 U-Net 生成器和马尔可夫鉴别器的 GAN 结构用于图像翻译。在此框架的基础上，Liu 等[5]提出了一种基于生成对抗网络的视频异常检测方法，用于预测未来帧。Dong 等[8]提出了一种基于双重鉴别器的生成对抗网络结构，该方法中的双鉴别器分别作用于视频帧的运动和外观信息。尽管这些方法在一定程度上提高了视频异常检测的准确性和鲁棒性，但它们没有考虑到生成器对异常帧也能进行很好的重构（预测）的问题。而本文提出的双鉴别器分别作用于正常数据和伪视频数据，提升了生成器对正常帧的预测能力，同时限制了生成器对异常帧的预测能力。

此外，一些研究者提出采用伪视频生成的方法来解决视频异常检测任务。通过构造伪视频数据，将不同于正常数据的伪视频数据和正常数据一起输入模型进行训练，能够在一定程度上缓解视频异常检测训练过程中异常数据缺失的问题。Georgescu 等[9]通过随机选取不属于正常事件的对象作为伪视频数据，利用伪视频数据迫使自编码器只学习重构正常模式，但此方法只适用于与外观相关的异常。G2D[10]提出使用一个未经训练的生成器生成伪视频数据来训练鉴别器，该方法需要两阶段的训练。而本文以端到端的方式训练，不需要任何预先训练的网络，取得了更优的效果。

3 本文方法

3.1 网络结构

本文提出的异常检测模型的整体框架如图 1 所示。在训练阶段，网络结构由一个伪视频合成器、一个生成器 G 和两个鉴别器（D_1 和 D_2）组成。对于正常序列，采用最小化预测损失的方法进行训练，并通过鉴别器 D_1 进一步提升生成器对正常帧的预测能力。对于伪视频序列，采用最大化预测损失的方法进行训练，并通过鉴别器 D_2 进一步限制生成器对伪视频帧的预测能力。在测试阶段，使用生成器的帧级预测误差来计算异常评分。

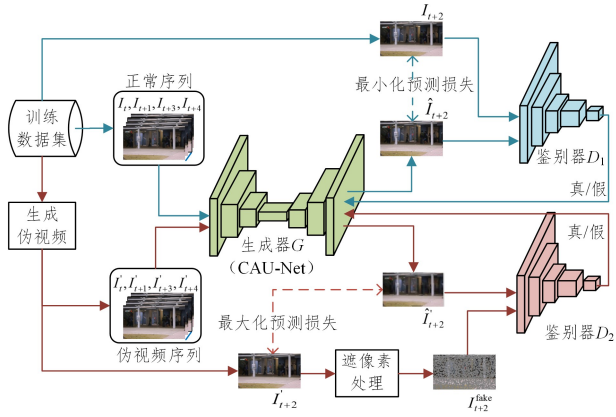


图 1 异常检测模型的整体框架

Fig. 1 Overall framework of anomaly detection model

图 1 中，$(I_t, I_{t+1}, I_{t+3}, I_{t+4})$ 为正常序列（缺中间帧），I_{t+2} 为正常序列中真实的中间帧，$\hat{I}_{t+2}$ 为生成器 G 生成的正常序列的中间帧；$(I'_t, I'_{t+1}, I'_{t+3}, I'_{t+4})$ 为伪视频序列（缺中间帧），I'_{t+2} 为伪视频序列中真实的中间帧，$\hat{I}'_{t+2}$ 为生成器 G 生成的伪视频序列的中间帧，I^{fake}_{t+2} 为 I'_{t+2} 经过遮像素处理得到的假帧。

1）伪视频合成器

用于训练的数据集只包含正常数据。从训练集 $V_j = \{I_1, I_2, \cdots, I_{k_j}\}$ 中选取长度为 n 的帧序列，V_j 是训练集中的第 j 个视频，k_j 是 V_j 的帧总数，I_t 是当前视频中的第 t 帧。得到的正常序列如式（1）所示：

$$X_N = (I_t, I_{t+1}, \cdots, I_{t+n-1}), 0 \leqslant t < k_j, t+n-1 \leqslant k_j \quad (1)$$

通过跳帧的方式生成伪视频数据。从训练集 $V_j = \{I_1,$

$I_2,\cdots,I_{k_j}\}$中每隔 $s-1(s>1)$帧选取一帧，依次选取长度为 n 的帧序列，得到伪视频序列如式(2)所示：

$$X_P=(I_t,I_{t+s},\cdots,I_{t+(n-1)*s}),0\leqslant t<k_j,t+(n-1)*s\leqslant k_j \tag{2}$$

为了表达方便，将 X_P 改写为式(3)的形式：

$$X_P=(I'_t,I'_{t+1},\cdots,I'_{t+n-1}) \tag{3}$$

本文中，参数 n 设置为 5，参数 s 设置为 2。

2)生成器 G

本文使用改进后的 U-Net[11] 网络模型作为生成器来预测中间帧，在 U-Net 模型的跳层连接中加入协调注意力[12](Coordinate Attention)来提高模型的检测性能。使用注意力机制可以帮助生成器更好地理解输入数据的局部特征和全局信息，从而生成更准确、细节丰富、自然的图像，这在以往的研究工作中得到了验证[13-14]。本文在生成器中加入的协调注意力不仅能够捕获跨通道的信息，还能捕获方向感知和位置感知的信息，从而帮助生成器更加精准地定位和识别感兴趣的目标，进而提升模型的检测性能。协调注意力 U-Net 模型(Coordinate Attention U-Net，CAU-Net)结构如图 2 所示。

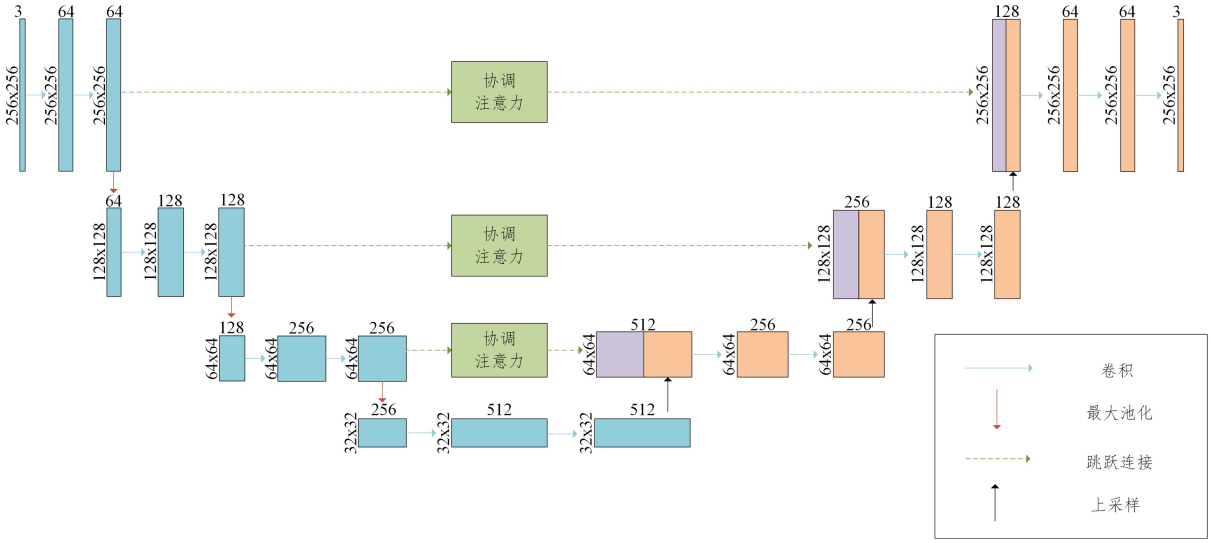


图 2 CAU-Net 结构的示意图

Fig. 2 Schematic diagram of CAU-Net structure

协调注意力通过将通道注意力分解为两个并行的一维特征编码，有效地将空间信息整合到生成的注意力图中。具体来说，协调注意力利用两个一维全局池化操作分别将垂直和水平方向的输入特征聚合成两个独立的方向感知特征图。这两个嵌入了方向特定信息的特征图被分别编码为两个注意力图，每个注意力图都捕获了输入特征图沿一个空间方向的长程依赖关系，因此位置信息可以被保留在生成的注意力图中，最后通过乘法将两个注意力图应用于输入特征图，以强调感兴趣的目标。协调注意力的处理流程如图 3 所示。

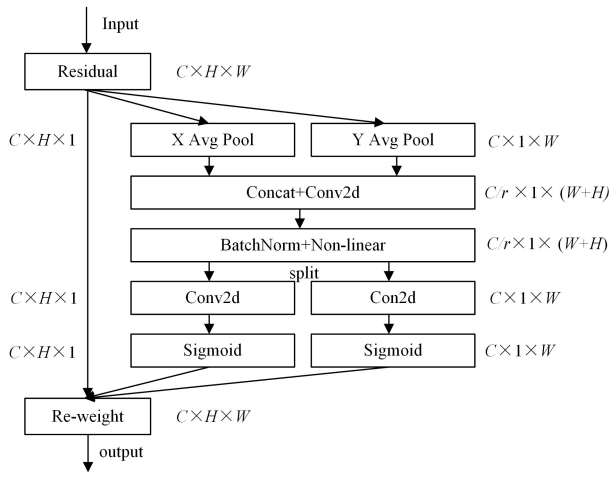


图 3 协调注意力的示意图

Fig. 3 Schematic diagram of coordinate attention

设定向生成器 G 中输入伪视频数据的概率为 p。当概率为 p 时，向生成器输入去除中间帧的伪视频序列(I_t'，I'_{t+1}，I'_{t+3}，I'_{t+4})进行训练，希望生成器预测的伪视频中间帧 $\hat{I}'_{t+2}$ 和真实的伪视频中间帧 I'_{t+2} 尽可能不相似。当概率为 $1-p$ 时，向生成器输入去除中间帧的正常帧序列(I_t，I_{t+1}，I_{t+3}，I_{t+4})进行训练，希望生成器预测的中间帧 $\hat{I}_{t+2}$ 和真实的中间帧 I_{t+2} 尽可能相似。

3)鉴别器 D_1

生成器 G 和鉴别器 D_1 组成一个生成对抗网络，以最小化生成的正常帧和真实的正常帧之间的差别。鉴别器 D_1 的任务是区分生成器预测的正常中间帧 $\hat{I}_{t+2}$ 和真实的正常中间帧 I_{t+2}。而生成器 G 的任务则是生成越来越真实的图像，以欺骗鉴别器 D_1，使其无法区分出 $\hat{I}_{t+2}$ 和 I_{t+2}。通过不断地对抗训练，生成器 G 会逐渐提高其对正常帧的预测能力，生成越来越接近真实帧的预测结果，从而提高模型的准确性和鲁棒性。这种方法在视频异常检测领域中有着广泛的应用[15-16]。

4)鉴别器 D_2

尽管可以通过最大化生成的伪视频中间帧 $\hat{I}'_{t+2}$ 和真实伪视频中间帧 I'_{t+2} 之间的预测损失，在一定程度上增大它们之间的差别，但由于伪视频帧与正常帧具有相同的背景，生成器在训练时学习了正常帧的背景，也就学习到了伪视频帧的

背景。然而，在测试阶段，通过逐像素计算真实帧与预测帧之间的差别来判断当前帧是否为异常帧时，就可能会出现因为背景误差太小而导致异常帧被误判为正常的情况。因此，需要确保在训练阶段，即使是针对伪视频帧的背景部分，也无法很好地预测，从而提升视频异常检测的性能。

生成器 G 和鉴别器 D_2 组成另一个生成对抗网络，以进一步增大 $\hat{I}'_{t+2}$ 和 I'_{t+2} 之间的差别，确保生成器对伪视频帧的背景部分也无法很好地预测。对于伪视频序列，希望生成器 G 生成的中间帧 $\hat{I}'_{t+2}$ 与真实的中间帧 I'_{t+2} 尽可能不同。因此，本文通过遮像素处理，将 I'_{t+2} 中大部分区域的像素点变成灰色，得到与真实图像极不相似的图像 I^{fake}_{t+2}，如图 4 所示。

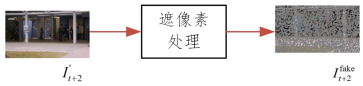


图 4 遮像素处理过程

Fig. 4 Mask pixel processing

鉴别器 D_2 的任务是区分出 I^{fake}_{t+2} 和 $\hat{I}'_{t+2}$，而生成器 G 的任务是生成越来越接近于 I^{fake}_{t+2} 的图像，以欺骗鉴别器 D_2，使其无法区分出 I^{fake}_{t+2} 和 $\hat{I}'_{t+2}$。通过不断地对抗训练，生成器 G 预测的伪视频中间帧 $\hat{I}'_{t+2}$ 会越来越接近于 I^{fake}_{t+2}。由于 I^{fake}_{t+2} 与 $\hat{I}'_{t+2}$ 极不相似，因此生成器生成的中间帧与真实中间帧也极不相似，以此来限制生成器对异常帧的预测能力。鉴别器 D_2 的有效性已在消融实验中得到验证。

3.2 损失函数

1)生成器 G 的损失函数

生成器 G 的损失函数由两部分组成，分别是预测损失和对抗损失。预测损失指生成模型中用于衡量生成器生成帧与真实帧之间相似性的度量。这里使用 L_2 一范数来测量预测误差，如式(4)、式(5)所示：

$$L_N(\hat{I}_{t+2}, I_{t+2}) = \| \hat{I}_{t+2} - I_{t+2} \|_2^2 \tag{4}$$

$$L_P(\hat{I}'_{t+2}, I'_{t+2}) = - \| \hat{I}'_{t+2} - I'_{t+2} \|_2^2 \tag{5}$$

其中，$L_N(\hat{I}_{t+2}, I_{t+2})$ 和 $L_P(\hat{I}'_{t+2}, I'_{t+2})$ 分别是输入正常序列和输入伪视频序列时的预测损失函数。对于正常序列，要求最小化预测损失；对于伪视频序列，则要求最大化预测损失。生成器 G 总的预测损失函数如式(6)所示：

$$L_{yc} = \begin{cases} L_N(\hat{I}_{t+2}, I_{t+2}), & \text{概率为 } 1-p \\ L_P(\hat{I}'_{t+2}, I'_{t+2}), & \text{概率为 } p \end{cases} \tag{6}$$

其中，p 为输入伪视频序列的概率。

对抗损失是生成器在与鉴别器对抗训练过程中用到的损失。对于鉴别器 D_1 来说，I_{t+2} 是类别 1，$\hat{I}_{t+2}$ 是类别 0；对于鉴别器 D_2 来说，I^{fake}_{t+2} 是类别 1，$\hat{I}'_{t+2}$ 是类别 0。在训练生成器时，需要将鉴别器的鉴别能力固定，训练生成器让鉴别器 D_1 错误地认为生成器生成的真实中间帧 $\hat{I}_{t+2}$ 是类别 1，让鉴别器 D_2 错误地认为生成器生成的伪视频中间帧 $\hat{I}'_{t+2}$ 是类别 1。生成器 G 的两个对抗损失函数分别如式(7)、式(8)所示：

$$L^G_{adv1}(\hat{I}_{t+2}) = \sum_{x,y} \frac{1}{2} L_{MSE}(D_1(\hat{I}_{t+2}), 1) \tag{7}$$

$$L^G_{adv2}(\hat{I}'_{t+2}) = \sum_{x,y} \frac{1}{2} L_{MSE}(D_2(\hat{I}'_{t+2}), 1) \tag{8}$$

其中，$L^G_{adv1}(\hat{I}_{t+2})$ 和 $L^G_{adv2}(\hat{I}'_{t+2})$ 分别是输入正常序列和伪视频序列时生成器 G 的对抗损失函数。其中，x 和 y 是鉴别器输出的二维向量的横纵索引；L_{MSE} 为 MSE 函数，具体的含义如式(9)所示。生成器 G 总的对抗损失函数如式(10)所示。

$$L_{MSE}(\hat{Y}, Y) = (\hat{Y} - Y)^2, Y = 1 \text{ 或 } 0, \hat{Y} \in [0,1] \tag{9}$$

$$L^G_{adv} = \begin{cases} L^G_{adv1}(\hat{I}_{t+2}), & \text{概率为 } 1-p \\ L^G_{adv2}(\hat{I}'_{t+2}), & \text{概率为 } p \end{cases} \tag{10}$$

将生成器 G 的预测损失和对抗损失加权求和得到生成器 G 总的目标损失函数，如式(11)所示：

$$L_G = \lambda_{yc} L_{yc} + \lambda_{adv} L^G_{adv} \tag{11}$$

其中，λ_{yc} 和 λ_{adv} 分别是预测损失和对抗损失所占的权重。通过不断迭代训练最小化损失函数，更新生成器 G 的参数 ω_G。

2)鉴别器 D_1 的损失函数

训练鉴别器 D_1 的目标是希望 D_1 能够将 I_{t+2} 划分为 1 类，$\hat{I}_{t+2}$ 划分为 0 类。当训练 D_1 时，需要固定生成器 G 的参数。鉴别器 D_1 的损失函数如式(12)所示。通过不断迭代训练最小化损失函数，更新鉴别器 D_1 的参数 ω_{D_1}。

$$L^{D_1}_{adv}(I_{t+2}, \hat{I}_{t+2}) = \sum_{x,y} \frac{1}{2} L_{MSE}(D_1(I_{t+2}), 1) + \sum_{x,y} \frac{1}{2} L_{MSE}(D_1(\hat{I}_{t+2}), 0) \tag{12}$$

3)鉴别器 D_2 的损失函数

训练鉴别器 D_2 的目标是希望 D_2 能够将 I^{fake}_{t+2} 划分为 1 类，将 $\hat{I}'_{t+2}$ 划分为 0 类。当训练鉴别器 D_2 时，需要固定生成器 G 的参数。鉴别器 D_2 的损失函数如式(13)所示。通过不断迭代训练最小化损失函数，更新鉴别器 D_2 的参数 ω_{D_2}。

$$L^{D_2}_{adv}(I^{fake}_{t+2}, \hat{I}'_{t+2}) = \sum_{x,y} \frac{1}{2} L_{MSE}(D_2(I^{fake}_{t+2}), 1) + \sum_{x,y} \frac{1}{2} L_{MSE}(D_2(\hat{I}'_{t+2}), 0) \tag{13}$$

3.3 异常分数

经过训练的生成器能够很好地预测正常序列的中间帧，但对于异常序列的中间帧预测效果较差，因此可以通过计算生成器预测帧和真实帧之间的差别来判断当前帧是否为异常帧。根据以往的研究工作[5]，峰值信噪比(PSNR)是一种很好的图像质量评估方法，PSNR 的值通常以分贝(dB)为单位，数值越高表示生成图像的质量越好，越接近于真实帧，说明当前被预测的帧更有可能是正常帧，PSNR 的计算式如式(14)所示：

$$PSNR(I, \hat{I}) = 10 \log_{10} \frac{[\max_{\hat{I}}^{\hat{}}]^2}{MSE} \tag{14}$$

其中，I 是真实帧，$\hat{I}$ 是生成的帧，$\max_{\hat{I}}^{\hat{}}$ 是 $\hat{I}$ 像素点的最大值。MSE(Mean Squared Error)是生成的帧和真实帧之间

像素值差异的均方误差，如式(15)所示：

$$MSE=\frac{1}{N}\sum_{i=0}^{N}(I_i-\hat{I}_i)^2 \tag{15}$$

其中，N 是当前帧的像素总数，i 是当前帧的像素索引。通过对每个测试视频中所有帧的 PSNR 值进行最大最小值归一化，将每一帧的 PSNR 值归一化到范围[0,1]，作为每一帧的异常分数。第 t 帧的异常分数 $S(t)$ 的计算式如式(16)所示：

$$S(t)=\frac{PSNR(I_t,\hat{I}_t)-\min_t PSNR(I_t,\hat{I}_t)}{\max_t PSNR(I_t,\hat{I}_t)-\min_t PSNR(I_t,\hat{I}_t)} \tag{16}$$

其中，$\max_t PSNR$ 与 $\min_t PSNR$ 分别表示当前视频前 t 帧最大的 $PSNR$ 值和最小的 PSNR 值。

测试时，根据其分数 $S(t)$ 来判断当前帧是否包含异常事件。通过设定阈值 T 来区分正常帧和异常帧，当异常分数大于阈值时判定当前帧为正常帧，当小于阈值时判定当前帧为异常帧，阈值的设定由 AUC 结果决定。

3.4 异常检测模型算法的描述

双鉴别器和伪视频生成视频异常检测算法的训练过程如算法 1 所示。

算法 1 训练双鉴别器和伪视频生成的视频异常检测算法

输入：训练轮次 E，训练数据集包含的帧数 fn，输入伪视频序列的概率 p，生成器 G 的学习率 p_G，鉴别器 D_1 和 D_2 的学习率 p_{D_1} 和 p_{D_2}，权重 λ_{yc} 和 λ_{adv}

输出：训练得到的模型参数 ω_G，ω_{D_1}，ω_{D_2}

1. initialize the model parameters ω_G，ω_{D_1}，ω_{D_2}；
2. for epoch=1 to E do
3. {for i=0 to fn do
4. {r=random(0,1)；/* 生成一个[0,1]之间的随机数，用于判断是否输入伪视频序列 */
5. if r<p{/* 输入伪视频序列训练 */
6. $\omega_G=\omega_G-p_G(\lambda_{yc}\frac{\partial L_P(\hat{I}_i',I_i')}{\partial\omega_G}+\lambda_{adv}\frac{\partial L_{adv2}^{G}(\hat{I}_i')}{\partial\omega_G})$；/* 固定鉴别器 D_2 的参数，更新生成器 G 的参数 */
7. $\omega_{D_2}=\omega_{D_2}-p_{D_2}\frac{\partial L_{adv}^{D_2}(\hat{I}_i',I_i^{fake})}{\partial\omega_{D_2}}$；/* 固定生成器 G 的参数，更新鉴别器 D_2 的参数 */}
8. else{/* 输入正常序列训练 */
9. $\omega_G=\omega_G-p_G(\lambda_{yc}\frac{\partial L_N(\hat{I}_i,I_i)}{\partial\omega_G}+\lambda_{adv}\frac{\partial L_{adv1}^{G}(\hat{I}_i)}{\partial\omega_G})$/* 固定鉴别器 D_1 的参数，更新生成器 G 的参数 */
10. $\omega_{D_1}=\omega_{D_1}-p_{D_1}\frac{\partial L_{adv}^{D_1}(\hat{I}_i,I_i)}{\partial\omega_{D_1}}$；/* 固定生成器 G 的参数，更新鉴别器 D_1 的参数 */}
11. }}
12. return ω_G，ω_{D_1}，ω_{D_2}.

4 实验结果与分析

4.1 实验数据集和评价指标

本文在两个公开的视频异常检测数据集上进行实验验证，即 UCSD Ped2[17] 和 CUHK Avenue[18]。每个数据集分为训练集和测试集，训练集中只有正常事件，而测试集中既包含正常事件又包含异常事件。UCSD Ped2 数据集是在人行道上拍摄的行人走动视频，包含 28 个视频片段，其中 16 个视频片段全部为正常行为，作为训练集；另外 12 个视频包含异常行为，例如骑自行车、开小轿车、滑滑板等，作为测试集。每个视频分辨率为 240×360 像素。CUHK Avenue 数据集中的视频是在地铁口从平视角度拍摄的，包含 37 个视频片段，其中 16 个视频序列全部为正常行为，作为训练集；另外 21 个视频片段包含异常行为，例如行人奔跑、扔掷杂物等，作为测试集。每个视频分辨率为 360×640 像素。

受试者工作的特征曲线(Receiver Operating Characteristic curve，ROC)是二分类器模型的分类性能指标。ROC 曲线的横坐标和纵坐标是不同阈值条件下，计算得出真阳性率(True Positive Rate，TPR)和假阳性率(False Positive Rate，FPR)的数值。ROC 曲线不便观察变化趋势，因此通过 ROC 曲线下的面积(Area Under Curve，AUC)对模型的分类准确度性能进行评价。在样本不平衡的情况下，AUC 依然能够对分类器作出合理的评价。AUC 的值越高，表明模型的分类性能越好。

实验中的相关参数设置如下：生成器 G 的预测损失权重系数 λ_{yc} 和对抗损失权重系数 λ_{adv} 分别设为 1.0 和 0.05，输入伪视频序列的概率 p 设为 0.01。

4.2 消融实验

为了验证本文模型各个模块的有效性，进行了消融实验，结果如表 1 所列。

表 1 消融实验

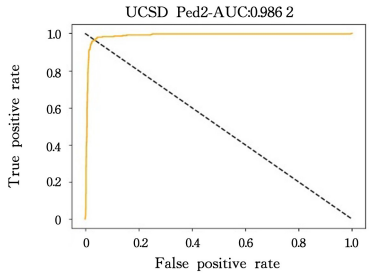
Table 1 Ablation experiment

方法	Ped2 AUC /%	Avenue AUC /%
无协调注意力	97.9	85.4
预测中间帧改为预测未来帧	97.6	85.4
无鉴别器 D_2	97.1	85.5
本文方法	**98.6**	**85.9**

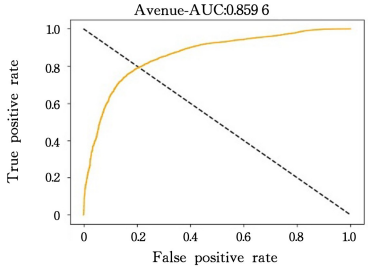
实验结果表明，没有鉴别器 D_2 的模型在两个数据集上的 AUC 分别下降了 1.5%和 0.4%，这表明鉴别器 D_2 在提升模型的检测性能方面发挥了有效作用。此外，没有添加协调注意力的模型在两个数据集上的实验结果表明，协调注意力通过对不同层的特征图进行加权，使得模型能够更好地处理不同的运动信息，提高了模型的生成能力。采用预测中间帧的模型比采用预测未来帧的模型在两个数据集上的 AUC 分别提高了 1%和 0.5%，这说明预测中间帧更有利于模型学习运动信息，从而提升模型的检测性能。

4.3 实验结果和性能评价

本文方法在 UCSD Ped2 和 CUHK Avenue 两个数据集上的 ROC 曲线如图 5 所示，横坐标表示正常数据被误判为异常的概率(False Positive Rate，FPR)，纵坐标表示异常数据被正确地判断为异常的概率(Ture Positive Rate，TPR)。异常检测的最理想情况是 TPR 等于 1，对应的 FRP 等于 0，因此曲线越接近(0，1)点，说明该方法的检测性能越好。ROC 曲线中横坐标和纵坐标差异最大的点所对应的阈值作为区分正常帧和异常帧的最佳阈值 T。



(a)在 Ped2 数据集上的 ROC 曲线



(b)在 Avenue 数据集上的 ROC 曲线

图 5 在 Ped2 和 Avenue 数据集上的 ROC 曲线

Fig. 5 ROC curves on Ped2 and Avenue datasets

本文方法和目前主流视频异常检测方法在 UCSD Ped2 和 CUHK Avenue 两个数据集上的检测性能比较如表 2 所列。在 Ped2 数据集上的 AUC 达到了 98.6%，在 Avenue 数据集上的 AUC 达到了 85.9%，均取得了较好的效果。

表 2 本文方法与其他视频异常检测方法的比较

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

Year	Method	Ped2 AUC/%	Avenue AUC/%
2018	Liu 等[5]	95.4	85.1
2019	Gong 等[4]	94.1	83.3
2019	Nawaratne 等[19]	96.5	84.5
2020	Dong 等[8]	95.6	84.9
2020	Ji 等[20]	98.1	78.3
2020	Yang 等[21]	95.9	85.9
2021	Astrid 等[22]	96.5	84.9
2022	Park 等[23]	96.3	85.3
2023	Le 等[13]	97.4	**86.7**
2023	Zhang 等[24]	97.9	85.9
2023	本文方法	**98.6**	85.9

与 Liu 等的方法相比，本文将预测未来帧改为预测中间帧，有效提高了异常检测的性能；与 Dong 等的方法相比，本文使用双鉴别器来提高生成器对正常帧的生成能力和约束模型对异常帧的生成能力，从而极大地提高了模型在异常事件检测方面的性能；而 Dong 等提出的方法使用双鉴别器仅考虑到正常帧。

由此可知，本文方法利用鉴别器 D_2 解决了生成模型也能够很好地预测异常的问题，加入协调注意力提升了模型的生成能力，同时采用预测中间帧的方式让模型更好地学习到了运动信息，从而有效提升了模型检测异常的能力。

结束语 本文提出了一种基于双鉴别器和伪视频生成的视频异常检测方法，通过生成器 G 和鉴别器 D_2 的对抗训练限制生成器对异常帧的生成能力，解决了生成器生成能力过强以至于对异常帧也能够很好预测的问题，从而提升了模型的检测性能。同时，在生成器中加入协调注意力，进一步提升了模型的检测性能。此外，本文方法将以往预测方法中的预测未来帧改为预测中间帧，有利于模型更好地学习前向和后向的运动信息。在两个公开数据集上的实验，进一步验证了本文方法在视频异常检测任务中的有效性。未来我们将考虑在伪异常生成中增加表观伪异常以及不同的运动伪异常，另外将考虑如何提升视频异常检测的实时性。

参 考 文 献

[1] MUNAWAR A, VINAYAVEKHIN P, DE MAGISTRIS G. Limiting the reconstruction capability of generative neural network using negative learning[C]// 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2017: 1-6.

[2] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]// International Conference on Learning Representations. 2018.

[3] ZAHEER M Z, LEE J, ASTRID M, et al. Old is gold: Redefining the adversarially learned one-class classifier training paradigm [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14183-14193.

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

[5] LIU W, LUO W, LIAN D, et al. Future frame prediction for anomaly detection-a new baseline[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6536-6545.

[6] SONG H, SUN C, WU X, et al. Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos[J]. IEEE Transactions on Multimedia, 2019, 22(8): 2138-2148.

[7] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.

[8] DONG F, ZHANG Y, NIE X. Dual discriminator generative adversarial network for video anomaly detection[J]. IEEE Access, 2020, 8: 88170-88176.

[9] GEORGESCU M I, IONESCU R T, KHAN F S, et al. A background-agnostic framework with adversarial training for abnormal event detection in video[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 4505-4523.

[10] POURREZA M, MOHAMMADI B, KHAKI M, et al. G2d: Generate to detect anomaly[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 2003-2012.

[11] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5 — 9, 2015, Proceedings, Part III 18. Springer Inter-

national Publishing,2015:234-241.
[12] HOU Q,ZHOU D,FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:13713-13722.
[13] LE V T,KIM Y G. Attention-based residual autoencoder for video anomaly detection[J]. Applied Intelligence,2023,53(3):3240-3254.
[14] LI H,SUN X,LI C,et al. MPAT:multi-path attention temporal method for video anomaly detection[J]. Multimedia Tools and Applications,2023,82(8):12557-12575.
[15] SHIN W,BU S J,CHO S B. 3D-convolutional neural network with generative adversarial network and autoencoder for robust anomaly detection in video surveillance[J]. International Journal of Neural Systems,2020,30(6):2050034.
[16] PARK H,NOH J,HAM B. Learning memory-guided normality for anomaly detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:14372-14381.
[17] MAHADEVAN V,LI W,BHALODIA V,et al. Anomaly detection in crowded scenes[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010:1975-1981.
[18] LU C,SHI J,AND JIA J. Abnormal event detection at 150 fps in matlab[C]// Proceedings of the IEEE International Conference on Computer Vision. 2013:2720-2727.
[19] NAWARATNE R,ALAHAKOON D,DE SILVA D,et al. Spatiotemporal anomaly detection using deep learning for real-time video surveillance[J]. IEEE Transactions on Industrial Informatics,2019,16(1):393-402.
[20] JI X,LI B,ZHU Y. Tam-net:Temporal enhanced appearance-to-motion generative network for video anomaly detection[C]// 2020 International Joint Conference on Neural Networks(IJCNN). IEEE,2020:1-8.
[21] YANG Y,ZHAN D,YANG F,et al. Improving video anomaly detection performance with patch-level loss and segmentation map[C]// 2020 IEEE 6th International Conference on Computer and Communications(ICCC). IEEE,2020:1832-1839.
[22] LU Y,YU F,REDDY M K K,et al. Few-shot scene-adaptive anomaly detection[C]// Computer Vision — ECCV 2020:16th European Conference,Glasgow,UK,Part V 16. Springer International Publishing,2020:125-141.
[23] PARK C,CHO M A,LEE M,et al. FastAno:Fast anomaly detection via spatio-temporal patch transformation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022:2249-2259.
[24] ZHANG Q,WEI H,CHEN J,et al. Video Anomaly Detection Based on Attention Mechanism[J]. Symmetry,2023,15(2):528.

GUO Fangyuan, born in 1999, master. Her main research interests include big data analysis and mining technology.

JI Genlin, born in 1964, Ph.D, professor, is a member of CCF(No. 09027S). His main research interests include big data analysis and mining technology.

(责任编辑:喻藜)