计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210900266

基于隐式视角转换的视频异常检测

冷佳旭[1,2]　谭明圮[1,3]　胡　波[1]　高新波[1]
1 重庆邮电大学图像认知重庆市重点实验室　重庆 400065
2 南京理工大学江苏省社会安全图像与视频理解重点实验室　南京 210094
3 重庆邮电大学光电工程学院　重庆 400065
(lengjx@cqupt.edu.cn)

摘　要　目前,基于深度学习的视频异常检测方法都是在单一视角下对视频片段中的异常行为或异常事物进行检测,忽视了视角信息在视频异常检测中的重要性。在单一视角下,当异常事物被遮挡或异常行为不明显时,现有算法的性能将难以得到保证。为此,文中首次将视角转换的概念引入到视频异常检测中,通过级联网络结构在多视角下进行异常判断来提升模型的鲁棒性。针对受限于数据集没有多视角的监督信息,难以实现真正的显式的视角转换问题,提出了一种基于隐式视角转换的视频异常检测方法.对初步检测结果为正常的目标帧,利用其与特定帧的光流信息,通过光流映射实现目标帧到特定帧视角的隐式视角转换,并对视角转换后的目标帧进行二次异常检测。通过多个视角来判定目标帧是否异常,为视频异常检测提供了一种新的思路。实验结果表明,所提方法对异常数据的反应更灵敏,具有更鲁棒的正常数据拟合能力,在 UCSD Ped2 和 CUHK Avenue 数据集上的 *AUC* 值分别达到了 97.0%和 88.9%。

关键词:视频异常检测;隐式视角转换;光流映射;多视角检测;深度学习
中图法分类号　TP391.41

Video Anomaly Detection Based on Implicit View Transformation

LENG Jia-xu[1,2],TAN Ming-pi[1,3],HU Bo[1] and GAO Xin-bo[1]
1 Key Laboratory of Image Cognition,Chongqing University of Posts and Telecommunications,Chongqing 400065,China
2 Jiangsu Key Laboratory of Image and Video Understanding for Social Safety,Nanjing University of Science and Technology,Nanjing 210094,China
3 School of Optoelectronic Engineering,Chongqing University of Posts and Telecommunications,Chongqing 400065,China

Abstract　Existing deep learning-based video anomaly detection methods all detect anomalies in video clips under a single view,ignoring the importance of view information in video anomaly detection. Under a single view,when anomalies are occluded or not obvious,the performance of existing algorithms will suffer drops. To avoid this problem,the author firstly introduces the concept of view transformation into video anomaly detection,which improves the robustness of the model by judging abnormalities from multiple views. However,due to the lack of multi-view supervision information in the dataset,it is difficult to achieve explicit view transformation. Specifically,in order to reflect the idea of view transformation,the author proposes a video anomaly detection method based on implicit view transformation,using the optical flow information between frames to warp the implicit view information of the previous frame to the target frame,so as to realize the implicit view transformation from the target frame to the previous frame. And then,the method performs secondary anomaly detection on the target frame after view transformation. Experimental results show that the proposed method responds more sensitively to abnormal data and has a more robust normal data fitting ability. The AUC values on the UCSD Ped2 and CUHK Avenue datasets reached 97.0% and 88.9%,respectively.

Keywords　Video anomaly detection,Implicit view transformation,Optical flow warp,Multi-view detection,Deep learning

---

到稿日期:2021-09-29　返修日期:2021-11-08
基金项目:国家自然科学基金(62036007,62050175,62102057);重庆市教委科学技术研究项目(KJQN-202100627)
This work was supported by the National Natural Science Foundation of China(62036007,62050175,62102057) and Science and Technology Research Program of Chongqing Municipal Education Commission(KJQN-202100627).
通信作者:高新波(gaoxb@cqupt.edu.cn)

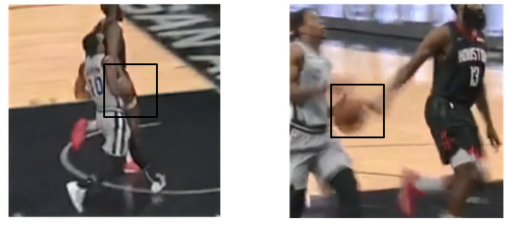
1 引言

视频监控系统是平安城市建设中重要的一环。随着视频监控系统的规模越来越大，监控视频数据越来越多，传统的监控视频理解与分析方法变得捉襟见肘，基于人工智能的视频理解与分析算法成为当前及未来的研究热点。视频异常检测作为视频理解中的关键技术，在维护社会稳定与公共安全方面意义重大。

视频异常检测指利用当前视频场景中的客观规律或正/异常行为特征表示的差异性对视频片段中的异常行为或异常事物进行检测[1]。但是，在不同的场景中，异常数据的定义是不同的（如一个人在公众场合拿刀和在厨房里拿刀这两种行为，前者是异常行为，而后者是正常行为），这给模型的设计带来了巨大挑战。从判断一个行为是否属于异常行为角度看，异常检测可以被看作一个分类任务，可利用有监督学习对异常数据进行分类检测。但对于正常数据而言，异常数据发生的概率较低，难以构造用于有监督模型训练的数据集。因此，无监督学习的方法[2-13]成为视频异常检测的主流。在训练阶段，只有正常数据进行训练，从而拟合正常数据的分布；在测试阶段，根据正/异常数据特征表示的差异性对异常数据进行检测。

目前，视频场景中的异常检测算法主要分为两大类。一是基于重构判别的视频异常检测算法，即利用自编码器在只含有正常数据的训练集上进行重构训练，实现较小的重构误差；在测试时，将重构误差较大的输入数据视为异常数据。二是基于预测模型的视频异常检测算法，该类算法从人类认知角度出发，认为正常数据通常是符合客观规律的，能够根据之前的行为对接下来的行为进行可靠预测，而异常行为因具有未知性而难以进行预测。因此，这一类方法通过预测误差来进行异常判别。由于深度自编码器[2]能够对高维信息（如视频数据）进行很好的建模，因此基于重构判别的方法大多以深度自编码器作为基础架构。然而，由于深度神经网络具有强大的表示能力，用于形成异常数据的特征在一定程度上可以由正常数据的特征进行表示，这使得深度自编码器的泛化能力“过强”，从而导致异常数据的重构误差较小，进而造成异常数据的漏检。

针对上述问题，上海科技大学高盛华课题组[4]首次提出了基于预测模型的异常检测算法，其利用异常数据的不可预测特性减弱了模型的泛化能力。随后，Ye 等[5]在此基础上结合神经科学的预测编码机制，提出了一种新的预测帧生成方法。Wang 等[12]采用多路径的卷积 GRU 预测框架来学习正常视频中的时空依赖关系，以促进生成更好的预测帧，进一步提升了模型的鲁棒性。但该类模型需要利用下一帧的信息来进行监督学习，不能对异常数据进行实时检测。为此，研究学者提出了基于记忆引导的异常检测模型[6-8]，其利用记忆模块存储正常数据的代表性特征，通过编码器的编码对记忆项进行匹配，再利用匹配好的记忆项作为解码器的输入进行解码，从而加大了正常数据和异常数据的得分差距，避免了深度自编码器的过度重构。但是，现有的异常检测模型都缺乏对视频数据中视角信息的关注，只是对目标帧的单一视角进行异常检测，这容易导致对单一视角下异常不突出的目标帧检测失败。由于在不同的视角下检测到异常数据的难易程度是不同的，因此在当前视角下看起来正常的行为或被遮挡的异常事物在转换视角后可以很容易地被发现。图 1 展示了不同视角下的异常行为，相较于图 1(a)的视角，图 1(b)的视角可以更容易看出黑队球员的手碰到了篮球，因此其更容易判断异常。



(a) (b)

图 1 不同视角下的异常行为

Fig. 1 Abnormal behavior shows from different views

为了进一步提升异常检测模型的鲁棒性，本文首次将视角转换的概念引入到视频异常检测中，通过隐式地将当前视角转换到其他视角，并联合多视角信息来进行异常判断。Zhu 等[9]为了得到当前帧不同视角的特征表示，首先利用历史帧和当前帧得到光流，然后通过光流信息把历史帧映射到当前帧，这种光流映射的方式在一定程度上体现了视角转换的思想。受此启发，我们提出了基于光流映射的隐式视角转换方法，通过光流映射将前一帧的视角信息映射到目标帧，从而实现帧与帧之间隐式的视角转换。本文在基于记忆引导的深度自编码器模型的基础上，结合基于光流映射的隐式视角转换方法，加强模型对视频数据中视角信息的关注，并利用级联结构减弱模型对异常数据的拟合能力。该方法首先利用基于记忆引导的深度自编码器对目标帧进行初步异常检测，如果检测结果为正常（可能是因为模型过度重构或者当前视角下异常数据不突出而导致的错误判断），则通过基于光流映射的隐式视角转换方法实现隐式视角转换后再次进行异常检测。最后，所提方法通过多个视角和二次检测来判定当前视频帧是否异常，从而提升模型的鲁棒性。

本文的主要贡献可总结为以下 3 点：

(1)首次在视频异常检测来中引入视角转换的概念；

(2)提出了适用于当前视频异常检测数据集的隐式视角转换方法，通过多视角下的异常检测来提升模型的鲁棒性，并通过级联结构在一定程度上解决了模型泛化能力“过强”的问题；

(3)所提隐式视角转换是一种通用模块，既可用于重构任务，又可用于预测任务。

2 相关工作

2.1 视频异常检测

传统检测方法[14-16]依赖特征的选择，对检测场景和检测的异常行为或异常事物有着特定的要求。目前，基于深度学习的视频异常检测算法快速发展，深度自编码器成为了异常

检测算法主流的特征提取和特征重构框架。假设基于重构判别的异常检测算法仅在正常数据上学习的模型不能准确地表示重构异常，从而能够以重构误差作为异常得分来检测异常。为了考虑视频数据中的时序信息，Hasan 等[3]提出了卷积自编码器(Convolutional AutoEncoder，ConvAE)，其相较于传统的深度自编码器更适用于视频数据，成为了目前大多数方法的基础框架。Zhao 等[10]提出了同时提取时间和空间特征的自编码器(Spatio-Temporal AutoEncoder，STAE)，其在自编码器中利用 3D 卷积网络进行特征提取和重构。随后，Luo 等[11]提出了基于卷积长短时记忆网络(Convolutional Long Short Term Memory，ConvLSTM)的自编码器，其利用卷积神经网络提取每个视频帧的空间信息，并通过卷积长短时记忆网络来存储与运动信息相对应的所有过去帧，从而更好地识别编码正常事件的外观和运动的变化。由于自编码器具有强大的泛化能力，当利用自编码器对输入进行重构时，异常数据会以较小的重构误差重构成功，进而被网络判定为正常数据，造成异常判别误差。为了解决自编码器过度重构的问题，Gong 等[6]提出了记忆增强的深度自编码器(Memory-augmented Deep AutoEncoder，MemAE)。对于给定的输入，该算法并不直接将其编码器得到的编码输入解码器，而是将其作为索引项来检索记忆矩阵中最相关的项，然后将这些项聚合并传递到解码器，以增加异常数据与常规数据之间的差异性。随后，Park 等[7]采用了不同的记忆模块读取和更新方案，让网络学习到由记忆引导的正常数据的正态性，从而进行异常检测(Learning Memory-guided Normality for Anomaly Detection，MNAD)。Cai 等[8]首先充分利用外观和运动信号的先验知识，以明确捕获它们在高级特征空间中的对应关系；然后结合多视图特征，以获得常规事件的更必要且更强大的特征表示，增加异常事件与常规事件之间的距离。Liu 等[4]在结构上进行创新，提出了基于预测模型的异常检测算法，这类算法的性能主要取决于生成器性能的好坏以及如何得到更有效的用于优化预测帧的视频帧信息。Ye 等[5]受神经科学的预测编码机制的启发，利用该机制生成预测帧，并且在预测误差优化阶段采用变种 U 型网络(U-Net[17])结合 ConvLSTM 的网络结构，更利于时序信息的表达。Wang 等[12]考虑到监控视频的特点，利用多路径的卷积 GRU 预测网络处理不同尺度的目标和区域中的语义信息并学习到正常视频中的时空依赖性；此外，还利用损失函数减小背景噪声的影响，从而获得质量更好的预测帧。为了将重构模型和预测模型进行结合，Nie 等[13]提出了一种混合模型，首先利用重构模型对输入视频序列对应的光流进行重构，然后将重构的光流作为先验信息来引导预测模型生成预测帧，最后将融合后的重构误差和预测误差作为异常得分进行异常检测。然而，这些方法都缺乏对视频数据中视角信息的关注，只是对目标帧的单一视角进行异常检测，容易导致单一视角下异常不突出的目标帧检测失败。

2.2 光流映射

在视频视觉跟踪、视角合成甚至光流估计领域上，已有特定视角转换的操作，这些操作在相关的领域都取得了一定的效果。Xiong 等[18]提出了第一个联合训练光流和视频语义分割任务的深度学习框架(Deep Feature Flow for Video Recognition，DFF)，实现了利用光流把一些关键帧的特征图映射到其他帧的操作。Zhu 等[9]为了得到当前帧不同视角的特征表示，首先利用历史帧和当前帧得到光流，然后利用光流信息把历史帧映射到当前帧，最后将映射过来的帧和本来的当前帧进行融合。Liu 等[19]通过预测生成的光流将前一帧的视角信息合成到后一帧。受此启发，本文将光流映射的方法引入到视频异常检测中，实现目标帧的隐式视角转换。

3 基于隐式视角转换的视频异常检测

本文在基于记忆引导的深度自编码器模型的基础上引入隐式视角转换的方法，对转换视角后的目标帧进行二次异常检测。首先利用基于记忆引导的深度自编码器模型对输入视频帧进行初步检测；然后对检测结果为正常的视频帧进行隐式的视角转换；最后将视角转换后的视频帧作为模型的输入进行二次异常检测，得到最终的检测结果。算法的整体框架如图 2 所示。

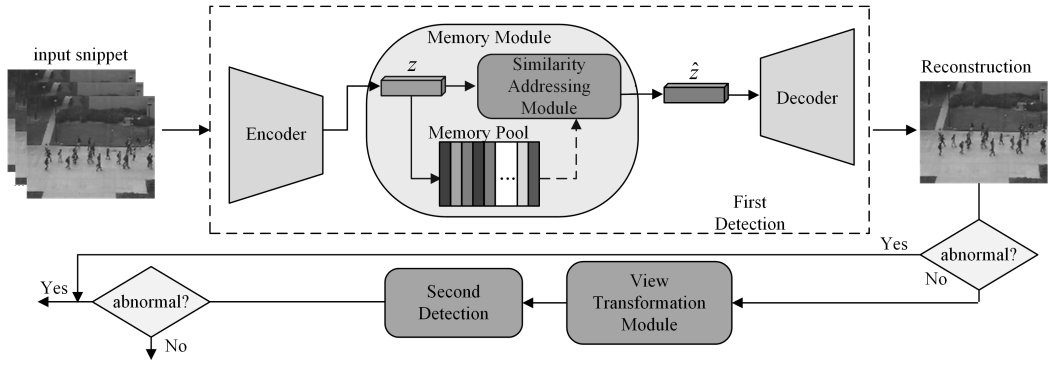


图 2 本文方法的整体框架图

Fig. 2 Overview of the proposed method

3.1 基于记忆引导的深度自编码器

基于记忆引导的自编码器在编码器与解码器中间增加一个记忆模块，通过使用记忆模块中的记忆项替代编码器提取到的特征编码，增大了正常数据与异常数据之间的差异性，从

而减弱了模型对异常数据的拟合能力。该模型主要包括编码器、记忆模块和解码器 3 个部分。其中，编码器对输入特征进行编码得到特征表达 z，并将其作为索引项用于匹配记忆模块中存储的记忆项；记忆模块包括记忆池和相似度寻址模块，特征表达 z 和记忆池中的记忆项进行相似度计算，得到对应的相似度权重，再以该权重作为对应系数，将记忆项进行线性组合得到重组特征表达 $\hat{z}$；解码器对重组特征表达 $\hat{z}$ 进行解码，重构出目标帧，若 $\hat{z}$ 能进行质量较好的重构，则将其用于记忆池的更新。通过记忆模块里的特征重组，解码器的输入并不直接等于编码器的输出，从而降低了异常数据重构成功的概率，在一定程度上解决了模型过度重构的问题。

3.2 基于光流映射的隐式视角转换

本文的初步异常检测模型是在单一视角下的视频异常检测，当异常信息被隐藏或不突出时，初步检测就会失败。通过视角转换，可以让原本异常不突出的视频帧将异常数据暴露，从而降低检测的难度。然而，由于数据集的限制，无法实现显式的视角转换。为此，本文提出了基于光流映射的隐式视角转换方法。利用视频帧之间的光流信息（光流有两个通道，分别代表 x 和 y 的偏移，该偏移量可表示隐式的视角信息），通过光流映射实现目标帧到特定帧视角的隐式视角转换。基于光流映射的隐式视角转换方法的整体逻辑框架如图 3 所示。其输入是一组相邻的视频帧，首先通过光流生成网络生成输入视频帧之间的光流，然后将光流信息映射（warp）到目标帧中，从而得到隐式视角转换后的目标帧。

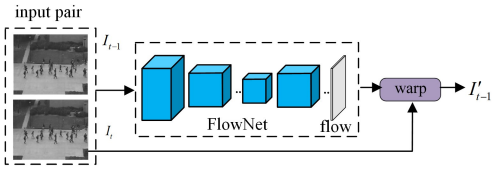


图 3 基于光流映射的隐式视角转换框架

Fig. 3 Network of the proposed implicit view transformation

3.2.1 光流生成网络

不同的光流生成网络的适用场景不同，对于同一个任务，不同的光流生成网络生成的光流的质量也不同。在基于视频帧光流的视角转换方法中，生成光流的质量是影响性能的关键因素。本文将第二代基于卷积神经网络的无监督光流生成网络（Evolution of Optical Flow Estimation with Deep Networks，FlowNet2）[20] 中针对小位移情况提出的光流生成模块（FlowNet2-SD）作为本文的光流生成网络。FlowNet2-SD 首先将第一代光流生成网络（Learning Optical Flow with Convolutional Networks，FlowNet）[21] 编码模块中大小为 7×7 和 5×5 的卷积核均换为多层 3×3 大小的卷积核，以提高对小位移的分辨率；其次，在解码模块的反卷积层之间，均增加一层卷积层，以便对小位移输出更加平滑的光流预测。

3.2.2 隐式视角转换

本文方法将前一时刻的视频帧与当前时刻的目标帧生成的光流信息映射到当前目标帧上，从而让目标帧得到前一帧的隐式视角信息。warp 操作是一种点对点的映射关系，本质上是通过双线性插值函数将光流值映射到目标帧。如图 3 所示，当输入是第 $t-1$ 帧和第 t 帧时，通过生成的光流，把第 $t-1$ 帧的隐式视角信息映射到第 t 帧，从而实现目标帧的视角转换。基于视频帧光流的视角转换的实现方法如式(1)所示：

$$I'_{t-1}=W(I_{t-1}+Flow(I_{t-1},I_t)) \tag{1}$$

其中，I'_{t-1} 表示转换视角后的目标帧，融合了第 $t-1$ 帧的视角信息；I_{t-1} 表示目标帧的前一帧；$Flow(I_{t-1},I_t)$ 表示利用光流生成网络得到的输入视频帧之间的光流；W 表示 warp 操作。

根据灰度不变假设：第 $t-1$ 帧上点 $p_{t-1}(x,y)$ 对应于输出视频帧上的点 $p_t(p_{t-1}+\delta p)$，δp 表示光流 $Flow(I_{t-1},I_t)$ 对应于点 p_{t-1} 的值。因此，可通过第 $t-1$ 帧上各点的值和对应的光流值得到视角转换后目标帧上各点的值。当 $p_{t-1}+\delta p$ 对应的值是浮点数时，采用双线性插值进行计算。利用光流映射后的目标帧的每个通道是二维的，通道上各点值的计算方式如式(2)－式(4)所示：

$$I_t^c(p_t)=\sum_q K(q,p_{t-1}+\delta p)I_{t-1}^c(q) \tag{2}$$

$$K(q,p)=g(q_x,p_x)\cdot g(q_y,p_y) \tag{3}$$

$$g(a,b)=\max(0,1-|a-b|) \tag{4}$$

其中，c 代表光流映射后的目标帧的单个通道；K 代表双线性插值核，如式(3)所示；q 代表进行双线性插值的空间点（共 4 个）。

上述双线性插值过程如图 4 所示。根据图 4 以及式(3)、式(4)，可将式(2)进行展开：

$$\begin{aligned}I_t^c(p_t)=&v(1-u)I_{t-1}(q_3)+uvI_{t-1}(q_4)+u(1-v)I_{t-1}\\&(q_2)+(1-u)(1-v)I_{t-1}(q_1)\end{aligned} \tag{5}$$

其中，u 和 v 是介于[0,1]的光流值。

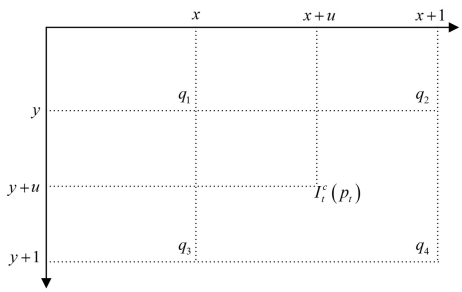


图 4 在第 $t-1$ 帧上进行双线性插值

Fig. 4 Bilinear interpolation on frame $t-1$

3.3 最终异常得分计算

视频异常检测算法通过模型重构（或预测）得到目标帧与真实目标帧之间的重构误差（或预测误差），从而得到该目标帧的异常得分，并将其作为目标帧是否异常的判别标准。为了保证实验公平，本文采用与初步检测模型相同的异常得分计算公式。对视角转换后的目标帧进行二次检测后，得到一个新的异常得分，本文引入一个权重参数 $\beta=0.8$ 用于平衡初步检测的异常得分和新的异常得分，得到目标帧最终的异常得分，如式(6)所示：

$$S_t=\beta S_t'+(1-\beta)S_t'' \tag{6}$$

其中，S_t'，S_t''，S_t 分别表示初步检测、二次检测和最终的异常得分。

4 实验

4.1 数据集、实验环境与评价标准

数据集：本文在两个公开数据集上进行实验验证。其中，UCSD Ped2 数据集[22]包括 16 个训练视频和 12 个测试视频，异常数据体现在场景中骑自行车和开卡车。CUHK Avenue 数据集[16]包括 16 个训练视频和 21 个测试视频，异常行为有 47 种，如在正常行走的环境中奔跑或者在场景中扔东西，以及错误的行进路线等。

实验环境：本文使用主流的 PyTorch 框架[23]来实现基于隐式视角转换的视频异常检测方法，并采用 Adam[24]优化器进行训练。为了实验的公平性，初步检测模型的实验参数与基础框架[7]保持相同。二次检测时不更新参数，其中隐式视角信息取目标帧与其前一帧的光流信息。本文的所有实验均在 NVIDIA GeForce RTX 2080 Ti GPU 和 Intel 24 核的 i9-10920X CPU 上进行。

评价标准：受试者工作特征曲线（Receiver Operating Characteristic Curve，ROC）下方的面积大小（Area Under Curve，AUC）是衡量异常检测模型性能好坏的标准，其值越大代表模型性能越好。本文通过最终的异常得分曲线进行 AUC 值的计算。

4.2 实验结果与可视化分析

4.2.1 实验结果比较

为了验证本文方法的有效性，以 MNAD[7]算法为基础框架进行实验，并和其他目前的主流视频异常检测算法进行对比，包括 Spatio-Temporal AutoEncoder（STAE）[10]，Convolutional AutoEncoder（ConvAE）[3]，Stacked Recurrent Neural Network（StackRNN）[25]，Temporally-coherent Sparse Coding（TSC）[25]，Memory-augmented Deep AutoEncoder（MemAE）[6]，Future Frame Prediction Network（Frame-pred）[4]，Unmasking the abnormal events（Unmasking）[26]，Appearance-Motion Memory Consistency Network（AMMC-Net）[8]。表 1 为各模型的 AUC 值对比结果。

表 1 模型的 AUC 值对比

Table 1 AUC comparison with the state of the art

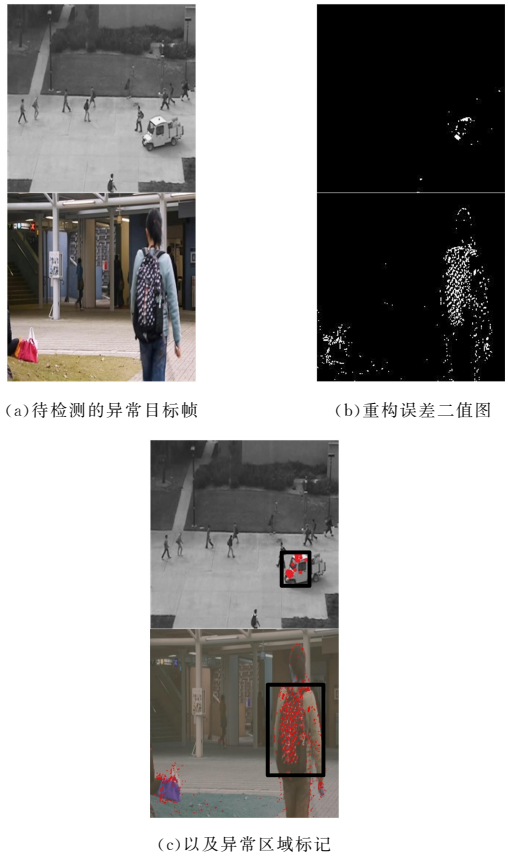
（单位：%）

Methods	UCSD Ped2	CUHK Avenue
ConvAE[3]	85.0	80.0
STAE[10]	91.2	77.1
TSC[25]	91.0	80.6
StackRNN[25]	92.2	81.7
Unmasking[26]	82.2	80.6
MemAE[6]	94.1	82.8
Frame-pred[4]	95.4	84.9
AMMC-Net[8]	96.6	86.6
MNAD(recon)[7]	90.2	82.8
Ours(recon)	**90.4**	**83.0**
MNAD(pred)[7]	96.9	88.5
Ours(pred)	**97.0**	**88.9**

表 1 中，MNAD(recon)表示 MNAD 算法的重构模型，MNAD(pred)表示该算法的预测模型。本文方法在 MNAD 重构模型上的应用，使得在数据集 UCSD Ped2 和 CUHK Avenue 上的 *AUC* 值相比 MNAD 本身分别提高了 0.2%和 0.2%；相比 ConvAE 分别提高了 5.4%和 3.0%。由于 MNAD 重构模型本身以 ConvAE 为基础框架，相较于 3D 卷积模型，ConvAE 采用 2D 卷积进行特征提取和重构，减少了记忆模块存储的记忆项数量，且不利于时间特征的提取，因此模型在重构任务上的性能稍显落后，但其计算速度更快，模型复杂度更低。本文方法在 MNAD 预测模型上的应用，使得在数据集 UCSD Ped2 和 CUHK Avenue 上的 *AUC* 值相比 MNAD 算法本身分别提高了 0.1%和 0.4%，相比 ConvAE 分别提高了 12.0%和 8.9%，均达到了目前主流异常检测算法在该数据集上最高的 AUC 值。

4.2.2 异常区域可视化

模型异常检测过程的可视化结果如图 5 所示。图中第一行、第二行分别表示 UCSD Ped2 数据集和 CUHK Avenue 数据集上的结果。其中，第一行表示的异常数据是人行道上的货车，第二行表示的异常数据是错误行进方向的行人。为了对异常区域进行可视化，本文首先把重构帧和真实帧逐像素作差，然后将得到的重构误差图进行二值化，最后将重构误差二值图在目标帧原图上进行映射，从而标记出异常区域。从重构误差二值图中可以看出，目标帧中的正常区域能够进行很好的重构，而异常区域重构效果较差。



(a)待检测的异常目标帧 (b)重构误差二值图

(c)以及异常区域标记

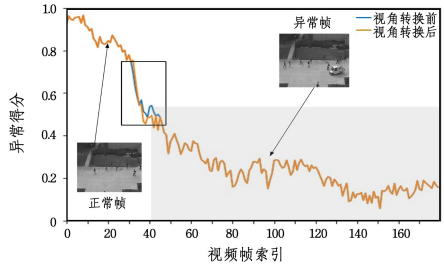
图 5 异常检测结果可视化

Fig. 5 Visualization of anomaly detection results

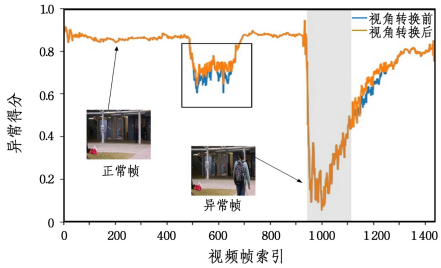
4.2.3 隐式视角转换前后异常得分曲线对比

根据式(6)可得到测试集上每个视频帧的最终异常得分，

并且异常得分越接近于零，对应的视频帧越有可能是异常的。图 6 给出了模型在 UCSD Ped2 和 CUHK Avenue 数据集上某个测试视频片段的异常得分曲线在视角转换前后的区别，其中图 6(a)为 UCSD Ped2 数据集上的结果，图 6(b)为 CUHK Avenue 数据集上的结果。图 6(a)中蓝色曲线部分为初步异常检测的异常得分曲线，橙色为进行隐式视角转换后二次检测的最终异常得分曲线图(纵坐标表示异常得分，横坐标表示视频帧的索引值，阴影区域代表异常发生时的视频片段)。对比黑色方框中的区域，相比初步检测(蓝色曲线部分)，当实现隐式视角转换后进行二次检测时，橙色曲线表示的异常得分下降更快，即模型对异常数据的反应更灵敏，说明隐式视角转换后，异常变得更突出，检测起来更容易。对比图 6(b)方框中的区域，视角转换后(橙色曲线)，模型对具有一定变化的正常视频帧的得分更稳定，说明视角转换后，具有一定变化的正常数据更符合一般的正常特征，表明算法对正常数据的定义更鲁棒，即算法在一定程度上提高了对正常数据的拟合能力。



(a)模型在 UCSD Ped2 数据集上的结果



(b)模型在 CUHK Avenue 数据集上的结果

图 6 UCSD Ped2 和 CUHK Avenue 数据集上异常得分曲线在视角转换前后的对比(电子版为彩色)

Fig. 6 Comparison of anomaly score curves on UCSD Ped2 and CUHK Avenue dataset before and after the view transformation

结束语 本文提出了基于隐式视角转换的视频异常检测算法，首次将视角转换的思想引入到视频异常检测领域，更多地关注视频数据中的视角信息，为异常检测提供了一种新的思路。通过实验验证，本文提出的方法是一个通用的模块，既可以与重构模型相结合，也可以和预测模型相结合，且均可以在一定程度上提升模型的性能。本文设计的视角转换方法受限于没有可监督的多视角信息，难以生成真实的新视角视频帧，实现显式的视角转换，因此所提方法性能受到了限制。在未来工作中，我们将构建具有不同视角的视频异常检测数据集，训练新视角的图像生成网络，实现更为稳定的真正意义上的视角转换。此外，我们将尝试设计自适应的视角选择网络，从多视角中选择异常更突出的视角，以降低检测难度，从而提升模型的鲁棒性。

参 考 文 献

[1] PENG J L, ZHAO Y L, WANG L M. An Overview of Video Anomaly Behavior Detection Based on Deep Learning[J]. Laser & Optoelectronics Progress, 2020, 58(6): 1-17.

[2] KINGMA D P, WELLING M. Auto-encoding variational bayes [C]// Proceedings of the International Conference on Learning Representations. Banff, Canada, 2014.

[3] HASAN M, CHOI J, NEUMANN J, et al. Learning Temporal Regularity in Video Sequences[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 733-742.

[4] 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. Utah, USA, 2018: 6536-6545.

[5] YE M C, PENG X J, GAN W H, et al. Anopcn: Video anomaly detection via deep predictive coding network[C]// Proceedings of the ACM International Conference on Multimedia. Nice, France, 2019: 1805-1813.

[6] 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 International Conference on Computer Vision. Seoul, Korea, 2019: 1705-1714.

[7] PARK H, NOH J, HAM B. Learning memory-guided normality for anomaly detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 14372-14381.

[8] CAI R C, ZHANG H, LIU W, et al. Appearance-Motion Memory Consistency Network for Video Anomaly Detection [J]. AAAI Conference on Artificial Intelligence, 2021, 35(2): 938-946.

[9] ZHU Z, WU W, ZOU W, et al. End-to-End Flow Correlation Tracking with Spatial-Temporal Attention[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah, USA, 2018: 548-557.

[10] ZHAO Y R, DENG B, SHEN C, et al. Spatio-Temporal AutoEncoder for Video Anomaly Detection[C]// Proceedings of the ACM International Conference on Multimedia. New York, USA, 2017: 1933-1941.

[11] LUO W X, LIU W, GAO S H. Remembering history with convolutional LSTM for anomaly detection[C]// Proceedings of the IEEE International Conference on Multimedia and Expo. Hong Kong, China, 2017: 439-444.

[12] WANG X Z, CHE Z P, YANG K, et al. Robust Unsupervised Video Anomaly Detection by Multi-Path Frame Prediction[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021.

[13] LIU Z A, NIE Y W, LONG C J, et al. A Hybrid Video Anomaly Detection Framework via Memory-Augmented Flow Recon-

struction and Flow-Guided Frame Prediction[C]// Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021: 13588-13597.

[14] SALIGRAMA V, KONRAD J, JODOIN P. Video Anomaly Identification[J]. IEEE Signal Processing Magazine, 2010, 27(5): 18-33.

[15] LEYVA R, SANCHEZ V, LI C T. Video Anomaly Detection with Compact Feature Sets for Online Performance[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3463-3478.

[16] LU C, SHI J, JIA J. Abnormal Event Detection at 150 FPS in MATLAB[C]// Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia, 2013: 2720-2727.

[17] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2015: 234-241.

[18] ZHU X Z, XIONG Y W, DAI J F, et al. Deep Feature Flow for Video Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 4141-4150.

[19] LIU L, ZHANG J N, HE R F, et al. Learning by Analogy: Reliable Supervision from Transformations for Unsupervised Optical Flow Estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 6488-6497.

[20] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 1647-1655.

[21] FISCHER P, DOSOVITSKIY A, ILG E, et al. FlowNet: Learning Optical Flow with Convolutional Networks[C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2016: 2758-2766.

[22] LI W X, MAHADEVAN V, VASCONCELOS N. Anomaly detection and localization in crowded scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1): 18-32.

[23] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An imperative style, high-performance deep learning library[C]// Proceedings of the Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 8026-8037.

[24] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]// Proceedings of the International Conference on Learning Representations. San Diego, USA, 2015: 1-15.

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

[26] IONESCU R T, SMEUREANU S, ALEXE B, et al. Unmasking the abnormal events in video[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2914-2922.

LENG Jia-xu, born in 1989, Ph.D, lecture, is a member of China Computer Federation. His main research interests include computer vision and object detection.

GAO Xin-bo, born in 1972, Ph.D, professor, is one of the talent bank of 10000 advisor for innovation and entrepreneurship, is a board member of China Computer Federation. His main research interests include artificial intelligence and pattern recognition.

(责任编辑:李亚辉)