摘要

在现代社会，视频监控设备已经在多个社会领域得到普及，深入到了人们日常生活之中。从这些监控视频中进行异常事件检测，具有重大的意义，尤其是在安防监控、生产质控、病人监护、智能家居等场景。但是，视频数据的迅猛增长显著增加了用人工方法检测异常事件的难度。因此，开发一种智能视频异常事件检测方法，实现视频异常事件检测的自动化，对以上多个场景具有重要的价值。

半监督方法是视频异常事件检测的主流任务定义方法。在训练阶段，半监督方法只使用无标注的正常视频序列数据，学习正常视频序列的特征。在测试阶段，如果检测到输入视频序列的特征和正常视频序列特征有显著差异，就会将其标记为异常。

当前视频异常视频检测的半监督深度学习方法主要存在两大问题：（1）时域平滑性不足：当前基于长短时记忆网络-自编码器框架的多数生成帧方法，其相邻的输出帧之间在时域上缺少平滑性。（2）泛化能力不足：在本文做领域调研的时候，未发现有方法能够在多个数据集上都能取得领域前沿的检测性能，即当前方法的泛化能力存在不足。

本文工作可以总结如下：

第一，针对当前方法时域平滑性不足的问题，本文提出了基于长短时记忆网络-自编码器框架的时域平滑提升方法，通过在训练目标函数中加入额外的约束项，限制长短时记忆网络输出的相邻特征隐向量之间的差异，从而使得输出帧序列在时域上更具平滑性。实验表明，先前基于长短时记忆网络-自编码器的各类方法，经过时域平滑提升方法处理后都能够得到显著的性能提升。

第二，针对当前方法泛化能力不足的问题，本文提出了基于生成对抗网络的新视频异常事件检测框架，通过在现有框架上引入两个针对不同具体问题的新模块，改善了方法的泛化能力。实验表明，该方法在多个公开数据集取得了和领域最新前沿成果相当的性能，并显著提升了方法在不同数据集上的泛化能力。另外，针对上述框架检测速度慢的问题，本文通过引入自注意力机制网络替代了长短时记忆网络，实现了部分计算的并行化，在检测性能仅略有下降的前提下加快了检测速度。

综上所述，本文主要对当前视频异常事件检测的半监督深度学习方法进行了改进。首先，提出了一种基于长短时记忆网络-自编码器框架的时域平滑提升方法。其次，提出了一种基于生成对抗网络的新检测框架，提高了算法的泛化能力。为了加快检测速度，本文还引入了自注意力机制网络替代长短时记忆网络，实现了部分计算的并行化。实验结果表明，本文提出的方法在多个公开数据集上取得了优异性能。这些成果为视

频异常事件检测领域提供了新的研究思路和技术基础，有望在安防监控、生产质控、病人监护和智能家居等应用场景中发挥重要作用。

关键词：计算机视觉；深度学习；视频异常检测

Chuan Lin (Computer Science and Technology (Intelligence Science and Technology))
Directed by: Prof. Shiliang Zhang

ABSTRACT

In modern society, video surveillance has become widespread across various social sectors, deeply integrating into people's daily lives. Detecting abnormal events from these surveillance videos holds significant importance , particularly in scenarios such as security monitoring, production quality control, patient care, and smart homes. However, the rapid growth of video data significantly increases the difficulty of detecting abnormal events using manual methods. Therefore, developing an intelligent video anomaly detection method for automating video anomaly detection is of great value in multiple aforementioned scenarios.

Semi-supervised methods are the mainstream task definition for video anomaly detection. In the training phase, semi-supervised methods use unlabeled normal video sequence data to learn the features of normal video sequences. In the testing phase, if the features of the input video sequence significantly differ from those of the normal video sequence, it is marked as abnormal.

There are two main issues in existing semi-supervised deep learning methods for video anomaly detection: 1) Insufficient temporal smoothness: most of the generative frame methods based on the Long Short-Term Memory-Autoencoder (LSTM-AE) framework lack temporal smoothness between adjacent output frames. 2) Inadequate generalization capabilities: during the domain research of this thesis, no method was found to achieve satisfactory detection results on different datasets, indicating insufficient generalization capabilities of current methods.

The contributions of this thesis can be summarized as follows:

First, to address the issue of insufficient temporal smoothness in current methods, this thesis proposes a temporal smoothness enhancement method based on the LSTM-AE framework. By adding an additional constraint term to the training objective function, the difference between adjacent feature latent vectors output by the LSTM is limited, thus making the output frame sequence smoother in the temporal domain. Experiments show that various previous LSTM-AE-based methods exhibit significant performance improvements after applying the temporal smoothness enhancement method from this thesis.

Second, to address the inadequate generalization capabilities of current methods, this

thesis proposes a new video anomaly detection framework based on Generative Adversarial Networks (GANs). By introducing two novel modules to address specific problems in the existing framework, the generalization capabilities of the method are improved. Experimental results show that the method proposed in this thesis achieves comparable performance to state-of-the-art results on multiple public datasets and significantly enhances its generalization capability across different datasets. Furthermore, to address the slow detection speed of the aforementioned framework, this thesis introduces a self-attention mechanism network to replace the LSTM, achieving parallelization of some computations while accelerating detection speed with only a slight performance decrease.

In summary, this thesis primarily improves the current semi-supervised deep learning methods for video anomaly detection. First, a temporal smoothness enhancement method based on the LSTM-AE framework is proposed. Next, a new detection framework based on GANs is proposed to improve the algorithm's generalization capability. To accelerate the detection speed, a self-attention mechanism network is introduced to replace the LSTM, achieving parallelization of some computations. Experimental results show that the proposed methods in this thesis achieve excellent performance on multiple public datasets. These achievements provide new research ideas and technical foundations for the field of video anomaly detection and are expected to play an essential role in various application scenarios such as security monitoring, production quality control, patient care, and smart homes.

KEY WORDS: Computer Vision； Deep Learning； Video Anomaly Detection