Form 3

Abstract of Thesis

Name (Savath SAYPADITH)

Title

A Study on Anomaly Detection in Surveillance Video using Generative Adversarial Network
(敵対的生成ネットワークを用いた監視カメラ映像に対する異常検知手法に関する研究)

Abstract of Thesis

Videos from Closed-Circuit Television (CCTV) cameras are rapidly generated every minute in accordance with an increasing number of cameras either in public places or private places in order to increase the efficiency, safety, and security due to criminal and terrorist attacks. The monitoring proficiency of anomaly events in hundred surveillance cameras using human labor is ambitious. To overcome this problem, developing intelligent computer vision algorithms to automatically detect events in a video scene is a viable solution. Anomaly detection in the video has recently gained attention due to its importance in the intelligent surveillance system. Real-world anomaly events are complicated and it is difficult to define every specific event. Although anomaly detection algorithms have reached the accuracy level under certain condition, the algorithm may still be affected by the external and internal variation such as the illumination, direction of movement object, motion velocity, occlusion and similar object motion.

Even though the performance of the state-of-art methods has been competitive in the benchmark dataset, the trade-off between the processing time and the accuracy of the anomaly detection should be considered. This dissertation proposes a framework for detecting anomalies in video, which designs a "multi-scale U-Net" network architecture based on generative adversarial network (GAN) structure for unsupervised learning to detect anomaly in video. To improve the training and testing of the neural network, Shortcut Inception Modules (SIMs) and residual skip connections are used in the generator network. Instead of using traditional convolution layers, an asymmetric convolution was used to reduce the number of training parameters without impacting detection accuracy. A multi-scale U-Net kept useful features of an image that were lost during training caused by the convolution operator. The generator network is trained by minimizing the reconstruction error on the normal data and then using the reconstruction error as an indicator of anomalies in the testing phase. This dissertation evaluates the performance with three benchmark datasets including UCSD Pedestrian, CUHK Avenue and ShanghaiTech datasets. The experimental results demonstrate that the framework surpasses the state-of-the-art learning-based methods, which achieved 95.7%, 86.9%, and 73.0% in terms of AUC. The multi-scale U-Net reduces the number of network parameters by 22.6% compared to the original U-Net architecture. In average, the proposed architecture takes 0.041 seconds per frame. As a result, the complete pipeline can run at 24 frames per second (fps), which is on par or slightly better than the baseline network architecture, which can run at roughly 22 fps.

This dissertation also proposes a joint representation learning for video anomaly detection. The proposed architecture extracts features from the object appearance and their associate motion features via different encoders based on ResNet network architecture. The network architecture is designed to combine spatial and temporal features, which share the same decoder. Using a joint representation learning approach, the proposed architecture effectively learn both appearance and motion features to detect anomalies in various scene scenarios. The experiments on three benchmark datasets demonstrate the remarkable detection accuracy with respect to existing state-of-the-art methods, which achieve 96.5%, 86.9%, and 73.4% in UCSD Pedestrian, CHUK Avenue, and ShanghaiTech datasets, respectively.

様式 7

論文審査の結果の要旨及び担当者

氏　名　（　Savath Saypadith　）		
	（職）	氏　名
論文審査担当者		
主 査	教授	尾上 孝雄
副 査	教授	櫻井 保志
副 査	准教授	谷口 一徹
副 査	准教授	Supavadee Aramvith（チュラロンコン大学）

論文審査の結果の要旨

本論文は、監視カメラ映像に対する異常検知手法のフレームワークに関する研究の成果をまとめたものであり、以下の主要な結果を得ている。

1. マルチスケールU-Netに基づく異常検知手法の提案

ビデオの異常を検出するためのフレームワークとして、教師なし学習のための敵対的生成ネットワークを用（GAN）構造に基づく「マルチスケールU-Net」ネットワークアーキテクチャを設計している。ニューラルネットワークの性能向上をめざして、生成ネットワークにShortcut Inception Modules（SIM)と残差スキップ接続を使用するアーキテクチャである。従来の畳み込み層に代えて、非対称畳み込みを使用し、検出精度を落とすことなく学習パラメータ数の削減を達成している。本マルチスケールU-Netは、畳み込み演算子に起因して学習中に失われた画像の特徴を保全する機能を持つ。また、生成ネットワーク部では、正常データに対する再構成誤差を最小化することによって学習を行い、認識時には再構成誤差を異常検知指標として使用する。提案手法を、UCSD Pedestrian、CUHK Avenue、ShanghaiTechの3つのベンチマークデータセットで性能評価した結果、Area Under the Curve(AUC)値で95.7%、86.9%、73.0%を達成し、従来手法を上回る性能を達成している。また、本マルチスケールU-Netは、従来のU-Netアーキテクチャと比較して、ネットワークパラメータの数を22.6%削減している。提案アーキテクチャでのフレーム処理時間は0.041秒であり、これは、24フレーム/秒のフレームレートを達成している。

2. 時空間特徴併合による異常検知手法の提案

異常検知性能の向上をめざして、物体の外観とそれらに関連する運動特徴から特徴を抽出する手法を提案している。提案アーキテクチャは、ResNetネットワークアーキテクチャに基づくエンコーダを介して、空間的特徴と時間的特徴をそれぞれ抽出、併合する設計となっており、デコーダは共有するものである。時空間特徴を併合することにより、さまざまなシーンにおける異常検出が可能となる。UCSD Pedestrian, CHUK Avenue, ShanghaiTechの各データセットにおいて、それぞれ96.5%、86.9%、73.4%のAUC値を達成し、従来手法に対して検出精度の向上に成功している。

以上のように、本研究による敵対的生成ネットワークを用いた監視カメラ映像に対する異常検知手法に関する一連の研究成果は、Society5.0時代の安全・安心な社会生活を実現する目的などで今後ますます重要となる観点からも非常に有用である。また、認識精度の向上のみならず処理時間の短縮に関しても議論されており、本論文はシステム実用化にも寄与するものと期待できる。従って、博士（情報科学）の学位論文として価値あるものと認める。