Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Anomaly Detection: Spotting The Unusual With Machine Learning And Deep Learning

Shameem Akthar K[1,*] , Dr. K. Lakshmi Priya[2]

[1]*Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education. Coimbatore, Email ID: s.aktharu@gmail.com*
[2]*Associate Professor, Department of Computer Science, Karpagam Academy of Higher Education. Coimbatore, Email ID: lakshmipriya.krishnan@kahedu.edu.in*

Object detection, a critical task, involves pinpointing a target's position in each video frame with corresponding coordinates. Traditional methods like 2D correlation struggle with scaling, rotation, and other transformations. Convolutional neural networks (CNNs) offer a two-stage approach with a classifier and regression, but they can be computationally heavy. Alternatively, one-stage detectors are simpler and more efficient but less accurate. This paper advocates using computer intelligence (CI) and artificial intelligence (AI) algorithms for intelligent motion detection, noting their prevalence in modern society. While CI relies on rule-based systems, AI leans on machine learning and deep learning. The paper compares their performance on various datasets, finding that AI generally excels in accuracy and robustness despite being computationally intensive. It concludes by discussing the future implications, suggesting AI will dominate due to its advantages but emphasizing the need for more efficient real-time algorithms.

Keywords: object detection, intelligent motion detection, convolutional neural networks, machine learning, deep learning.

1. Introduction

Video surveillance is a controversial topic, especially with the way privacy issues are looming over our heads. At the same time, the importance of surveillance in business cannot be understated in the current social climate, as it is fundamentally important to ensure the safety of your possessions and the individuals working for you. As a business, worrying about your employees' physical health and office property is justifiable, especially if you have a small or medium-scale outfit. CCTV monitoring is one of the most common forms of surveillance available today. There are more than 25 million operational CCTV cameras globally, and experts believe that around 67% of burglaries in businesses could be avoided by installing real-time security measures such as CCTV monitoring [1].

Many argue the United States is one of the most surveilled countries, with 15.3 closed-circuit television (CCTV) cameras for every 100 individuals. The public opinion around CCTV is varied, but there are many positives to using CCTV in public places — if they work. Many current cameras are too old or have outdated software, meaning they're no longer fit for purpose. The industry is constantly changing, with new technologies and innovative solutions

driven by increased security needs brought on by the pandemic. One area where CCTV technology has developed is artificial intelligence (AI) CCTV cameras[1].
AI CCTV cameras are network IP cameras that deliver advanced analytical functions like vehicle detection, face detection, person detection, people counting, traffic counting and license plate recognition (LPR). Advanced video analytics software is built into the camera and recorder, enabling artificial intelligence functions. For AI CCTV cameras to work, data is constantly sent to a recorder and processed via an AI layer to make sense of the raw video.
Rule-based AI cameras are manually set up with rules and reference images, such as images of humans in different postures, angles, or movements. The AI will then ask itself if anything it observes looks and moves like this. Depending on the rules set, such as ‘no one is allowed in this area at a certain time,’ if the camera observes this, it will send an alert [2]. AI CCTV cameras have shown how the security industry is constantly developing with the help of new technologies. Other features found in new CCTV camera systems include thermal cameras, solar-powered cameras and those with features like time-lapse and heat/fire detection [3].
In something reminiscent of Tom Cruise’s Minority Report, some police forces in the UK have been trialing a system that will predict how likely individuals are to commit a crime so they can stop it before it happens. The system, the first of its kind, uses over a terabyte of data from local and national police databases, including records of previously stopped and searched people and their criminal records. The police found nearly 1400 indicators that could help predict crime. What happens after individuals have been detected is debated, but support from social services has been offered as a potential solution. Governments and police departments worldwide are constantly looking for new features that will help prevent crime. CCTV is one way they hope to improve safety, especially in public spaces. Along with their self-learning systems, features like object tracking, two-way audio, and facial recognition make AI CCTV cameras an advanced and effective video security solution[3].

2. Problem Statement

Over the past decade, machine learning and human activity comprehension have garnered significant attention due to their wide-ranging and complex nature. Computer vision and machine learning can detect and track human actions, model scenes, and understand behavior, including recognizing human actions and identifying patterns. These applications have several uses, including video surveillance, human-computer interfaces, multimedia semantic annotation and indexing, and ensuring worldwide security in public places like airports, train stations, shopping malls, crowded sports arenas, and military sites. Additionally, intelligent visual surveillance is necessary in brilliant healthcare facilities to observe senior citizens' daily activities and identify any minor physical accidents that may occur by chance.

3. Deep Learning

Deep learning is an area of artificial intelligence in which computers execute tasks such as identifying objects and recognising them in a video by being exposed to data. Massive volumes of data must be labelled and processed for a system to engage in deep learning. The network is then trained until it can reliably perform the original task.

3.1 How Deep Learning Can Help with Object Detection and Identification in Video Surveillance:

Deep learning can help with object detection and identification in video surveillance in several ways. First, deep learning algorithms can be trained on large datasets of images and videos containing various objects. This allows the algorithms to learn to identify objects even in challenging conditions, such as low light or crowded scenes. Second, deep learning algorithms can be used to develop real-time object detection and identification systems. This means that the systems can detect and identify objects in video streams as they are happening. This can be useful for applications such as security monitoring and traffic management. Finally, deep learning algorithms can be used to develop intelligent video surveillance systems to detect and identify objects and understand and analyze their behavior. This can be useful for applications such as crime prevention and customer service[4,5,6].

Here are some examples of how deep learning is being used for object detection and identification in video surveillance today:

- Security cameras: Deep learning algorithms are being used to develop security cameras that automatically detect and identify people, vehicles, and other objects of interest. This can help security personnel quickly identify and respond to potential threats.
- Traffic cameras: Deep learning algorithms are being used to develop traffic cameras that automatically detect and track vehicles. This information can be used to improve traffic flow and reduce congestion.
- Retail stores: Deep learning algorithms are used to develop video surveillance systems that track customer movement and identify popular products. This information can be used to improve store layout and product placement.

Deep learning is a rapidly developing field, and new applications for object detection and identification in video surveillance are constantly being developed.

4. Research Gap

Deep learning has been used to create many models and intelligent systems that can deal with the oddities and technological problems that arise in different applications. These models and systems can clearly reduce the number of human resources used and make people's lives easier. Despite this, video anomaly detection still faces numerous obstacles and problems.

Challenges of Video Anomaly Detection

[1]. Higher false alarm
[2]. Being invalid when the model generalizes well; Inexplicability
[3]. Higher computational complexity
[4]. Expensive training; Instability; Difficulties in reproduction; Mode collapse

5. Literature Review

In this section, the authors review several recent works on video anomaly detection, focusing on the abovementioned challenges.

5.1 MONAD:

Doshi and Yilmaz (2021) proposed a new framework for video anomaly detection called MONAD. MONAD uses a statistical sequential approach to detect anomalies in videos. The authors evaluate MONAD's performance and propose a practical approach to choosing the detection threshold based on the desired false alarm rate. Additionally, they introduced a new metric based on average delay to measure timely detection in videos. However, the precision of the proposed method is not discussed in detail[8].

5.2 Multi-task Semantic Segmentation:

Jiang et al. (2021) created a multi-task semantic segmentation model for indoor environments. This model can be used for complex indoor environments using RGB-D image data and an improved Faster-RCNN algorithm for joint target detection. The authors enhanced the fusion of RGB and depth images by considering the effects of uneven lighting in the environment, which improved model training efficiency and boosted the fusion image feature information. The loss function was also modified and optimized to achieve multi-task information output. The proposed indoor scene semantic segmentation model showed strong performance and high efficiency and could segment objects of different scales and adapt to uneven illumination conditions[9].

5.3 Gaussian Distribution Constraint:

Qinmin Ma (2021) presented a new anomaly detection approach that constrains the representation of the hidden layer to a Gaussian distribution. In this study, the two main phases of anomaly detection, namely event representation and anomaly detection model setup, are transformed into hidden layer representation and Gaussian distribution constraint using a variational autoencoder (VAE). Jointly tuning the two processes enhances the approach's accuracy and generalizability. However, when dealing with increasingly complex datasets, the proposed procedure could become more complex [10].

5.4 Background Subtraction with MSER:

Murugesan and Thilagamani (2020) introduced a background subtraction method that employs the Maximally Stable External Region (MSER) feature extraction technique. This approach is suitable for pixel-wise foreground analysis and system-based anomaly detection for different objects of various sizes. The proposed method outperforms existing methods by producing better image categorization outcomes with higher accuracy and lower calculation errors. The output's classification accuracy, specificity, and sensitivity are reported to be 98.56%, 96.05%, and 98.21%, respectively[11].

5.5 DenseASPP:

In 2020, researchers released the densely connected Atrous Spatial Pyramid Pooling (DenseASPP) method, which links a number of atrous convolutional layers. As a result, multiscale features are produced that not only span a wider scale range but also do so densely and without considerably growing the size of the model. Testing DenseASPP on the Cityscapes street scene benchmark yields the best results possible[12].

5.6 The comprehensive literature reviews are as follows:

- The authors could discuss the different types of video anomaly detection algorithms, such as reconstruction-based, predictive-based, and statistical-based algorithms.
- The authors could also discuss the different applications of video anomaly detection, such as security and surveillance, industrial inspection, and medical imaging.
- Finally, the authors could suggest future research directions in video anomaly detection.

5.6 Review of Anomaly Detection in Crowded and Uncrowded Areas

Anomaly detection is a rapidly growing field of research with a wide range of applications in domains such as security, surveillance, and healthcare. In video surveillance, anomaly detection can identify unusual or suspicious behavior in crowded and uncrowded areas.

Existing anomaly detection algorithms typically focus on motion data, ignoring abnormalities resulting from object appearance changes. This makes them vulnerable to anomalies not caused by motion outliers, such as a vehicle crossing a bridge with weight restrictions. Additionally, in crowded scenes with dynamic backgrounds, noise, and complex occlusions, descriptors such as optical flow and pixel change histograms can be challenging to extract reliably.

This review summarizes recent advances in anomaly detection for crowded and uncrowded areas. We focus on deep learning-based methods, which have shown promising results in recent years.

Methods

Various deep learning-based methods have been proposed for anomaly detection in crowded and uncrowded areas. These methods can be broadly classified into two categories:

- Reconstruction-based methods train a deep learning model to reconstruct standard video frames. Anomalies are then detected as frames with high reconstruction errors.
- Predictive-based methods train a deep learning model to predict the next frame in a video sequence. Anomalies are then detected as frames with high prediction errors.

5.7 Datasets:

- A variety of datasets have been developed for evaluating anomaly detection algorithms in crowded and uncrowded areas. Some of the most popular datasets include:
- Cuhk Avenue: This dataset contains video footage of crowded and uncrowded street scenes.
- Shanghai Tech: This dataset contains video footage of crowded and uncrowded street scenes with various anomalies, such as fighting, running, and jaywalking.
- Ucsd Avenue: This dataset contains footage of crowded and uncrowded street scenes.
- Ucsd Ped2: This dataset contains video footage of crowded and uncrowded pedestrian scenes.

5.8 Output:

Anomaly detection algorithms typically output an anomaly score for each frame in a video sequence. The score measures how likely the frame is to be anomalous. Frames with high anomaly scores are then flagged for further inspection.

5.9 Results:

The following table summarizes the results of recent deep learning-based anomaly detection algorithms on popular datasets:

References	Methods	Dataset	Output
Anugrah Srivastava Et Al (2022)	CNN, Transfer Learning, Resnet-28	Hockey Dataset	99.20%
Pushpajit Khaire Praveen Kumar (2022)	Bi-Lstm, CNN	Human Action Recog-Nition Dataset In Atm.	89.1%
Fabio Et Al (2022)	CNN, Spatial Feature Selection	Cuhk Avenue	92.3%, 14.1%, 83.1%
Muhammad Ramzan (2022)	CNN	Violent-Flow Dataset And Movie Dataset	97.83%
Weichao Zhang (2021)	Gan	Cuhk Avenue And Shanghai Tech	89.2%, 75.7%
Qinmin Ma (2021)	VAE, Gaussian Distribution	Ucsd Avenue	92.3% 82.1%
Nasaruddin Et Al (2020)	CNN	Ucf Crime	98%
Juan Wang Et Al (2020)	Alexnet, SVM	Own	27.67%
Ramchandran, Anitha, And Sangaiah, Arun Kumar (2019)	Convolutional Autoencoder And Convolutional Lstm Model	Avenue , Ped1, Ped2	90.7 %, 98.4 %, 98.5 %
Waqas Sultani (2019)	Deep Multiple Instance Ranking Framework, Sparsity	Own	75.41%
Balasundaram And C. Chellappan (2018)	Split And Segment	Avenue, Own Dataset	99.77%, 98.19%
Ryota Hinami And His Associates (2017)	CNN	Avenue And Ucsd Ped2	89.2% 90.8%
Feng, Yachuang; Yuan, Yuan; And Lu, Xiaoqiang (2016)	Deep Gmm	Ped1(Frame Level), Ped1(Pixel Level)	92.5%, 64.9% 69.9%

6. Conclusion

This study aims to develop an efficient algorithm for object detection and tracking in complex video surveillance environments using computer intelligence (CI) and artificial intelligence (AI) algorithms. Human beings acquire the ability to recognize and comprehend visual information through years of learning, and vision is the human sense that enables the perception of the 3D external environment. The insights gained from this human ability serve as the foundation for emerging technologies, such as Convolutional Neural Networks (CNNs). With abundant resources and advanced methods in computer vision and deep learning, researchers can now extract more information from photos. Therefore, this research aims to develop an algorithm that effectively detects and tracks objects in complex surveillance environments.

References

[1]. Ying-Tung Hsiao, Jia-Shing Sheu, Hsu Ma(2023) "Efficient Object Detection and Intelligent Information Display Using YOLOv4-Tiny" Advances in Technology Innovation, vol. 9, no. 1, 2024, pp. 42-49.

[2]. Fabian Barreto, Jignesh Sarvaiya, Suprava Patnaik(2022) "Learning Representations for Face Recognition: A Review from Holistic to Deep Learning" Advances in Technology Innovation, vol. 7, no. 4, 2022, pp. 279-294.

[3]. Rahul Sharma, Amar Singh (2002) An Integrated Approach towards Efficient Image Classification Using Deep CNN with Transfer Learning and PCA Advances in Technology Innovation, vol. 7, no. 2, 2022, pp. 105-117.

[4]. Zhu, Jiasong, Ke Sun, Sen Jia, Qingquan Li, Xianxu Hou, Weidong Lin, Bozhi Liu, and GuopingQiu. "Urban traffic density estimation based on ultrahigh-resolution UAV video and deep neural network." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, no. 12 (2018): 4968-4981.

[5]. Zhang, X., Chen, Z., Wu, Q. M. J., Cai, L., Lu, D., & Li, X. (2018). Fast Semantic Segmentation for Scene Perception. IEEE Transactions on Industrial

[6]. Zhang, L., Sheng, Z., Li, Y., Sun, Q., Zhao, Y., & Feng, D. (2019). Image object detection and semantic segmentation based on convolutional neural network. Neural Computing and Applications.

[7]. Huang, L., He, M., Tan, C., Jiang, D., Li, G., & Yu, H. (2020). Jointly Network Image Processing: Multi-task Image Semantic Segmentation of Indoor Scene Based on CNN. IET Image Processing.

[8]. Doshi, Keval, and Yasin Yilmaz. "Any-shot sequential anomaly detection in surveillance videos." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 934-935. 2020

[9]. Jiang, D., Li, G., Tan, C., Huang, L., Sun, Y., & Kong, J. (2021). Semantic segmentation for multi scale target based on object recognition using the improved Faster-RCNN model. Future Generation Computer Systems, 123, 94–104.

[10]. Qinmin Ma School of Artificial Intelligence, Shenzhen Polytechnic, China, “Abnormal Event Detection in Videos Based on Deep Neural Networks”. Hindawi Scientific Programming , Volume 2021, Articlr ID 6412608

[11]. Murugesan, M., and S. Thilagamani. "Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network." Microprocessors and Microsystems 79 (2020): 103303.

[12]. Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, Kuiyuan Yang; "DenseASPP for Semantic Segmentation in Street Scenes".Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 3684-3692