REVIEW ON ANOMALY DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Shameem Akthar K[1], VR Nagarajan[2]*

Abstract

Object detection is a critical task that focuses on determining the position of a target in each frame with its corresponding coordinates. During the object detection phase, feature vectors are obtained using 2D correlation, a statistical method that directly measures the similarity between two video frames, irrespective of changes in lighting conditions or object translations. However, this technique cannot handle image scaling and rotation. CNN popularized a two-stage technique in which a classifier is used to a sparse set of potential object locations to produce highly accurate object detectors. One-stage detectors, on the other hand, are used over a regular and dense sampling of potential item locations, making them more efficient and uncomplicated but less accurate than two-stage detectors. To train the deep learning neural networks to be correlated with specific objects, their classification, attributes, and training data are required. In the case of video analytics, the video must be divided into individual frames to identify each object or item. Video analytics can be used to analyze video evidence, such as identifying individuals wearing pink pants, using algorithms. This study proposes the use of computer intelligence (CI) and artificial intelligence (AI) algorithms to detect intelligent motion and compares their respective performances. CI and AI are the two most dominant technologies in technical society.

Keywords: Malicious Activity, Deep Learning and Machine Learning anomaly, Object, Video Surveillance

I. INTRODUCTION

Detection and identification of objects are essential components of video surveillance analytics. This capability enables operators to locate and track a specific object, such as a person, vehicle, or bag, from one frame to the next. The capacity to instantly identify the object over hours of video offers crucial forensic evidence to security and police investigations [1]. Besides it delivers much-needed insights to corporate executives, and enables a wider range of people to utilize more potent applications. To be simple, object detection is the ability to distinguish audiovisual objects. But this is not a simple task when employing high-tech equipment. The "object extraction" procedure involves locating and tracking an object at any given time. Afterwards, a second technology referred as "background/foreground separation" is employed. This identifies and differentiates the scene's consistent background from its variable foreground elements [2]. The ability to extract objects from video and differentiate them from the background against which they are detected enables more advanced video analytics, such as forensic search and real-time alerting technology, which displays all extracted video objects simultaneously, allowing the entire scene's activity to be observed as opposed to in linear time [3]. A forensic inquiry may be performed. Alerts in real time could be conveyed. The deep learning methods aid to detect and identify object using computer-based techniques. Speech recognition and automatic translation services are developed using artificial intelligence and deep learning. Deep learning is an area of artificial intelligence in which computers execute tasks such as identifying objects and recognizing them in a video by being exposed to data. Massive volumes of data must be labeled and processed in order for a system to engage in deep learning. The network is then trained until it can do the original task reliably.

Department of Computer Science

Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

*Corresponding Author

II. 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, and multimedia semantic annotation and indexing, and ensuring worldwide security in public places like airports, train stations, shopping malls, crowded sports arenas, and military sites[4]. Additionally, intelligent visual surveillance is necessary in smart healthcare facilities to observe senior citizens' daily activities and identify any minor physical accidents that may occur by chance. Sometimes, the goal is to find, recognize, or learn about interesting occurrences, which may be referred to as "suspicious happenings," "irregular behavior," "uncommon conduct," "strange activity/event/behavior," "abnormal behavior," or "anomaly," depending on the situation.

III. RESEARCH GAP

Deep learning has been used to make a lot of models and intelligent systems that can deal with the different kinds of oddities and technological problems that come up in different applications. Clearly, these models and systems can cut down on the amount of human resources that are used and make people's lives easier. Despite this, video anomaly detection still faces numerous obstacles and problems.

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

IV. LITERATURE REVIEW

For the purpose of semantic segmentation of urban road sceneries, Zhang et al. (2018) present a neural network design with less parameters that is more efficient. We use a non-symmetric encoder-decoder structure based on ResNet in our model. Continuous factorised blocks are utilised in the encoder's initial iteration to extract low-level features. The second stage involves continuously applying a dilated block. As a result, the model maintains its shallow depth and compactness despite having a greater field of vision. The decoder increases the details while restoring the features that the encoder had shrunk to the size of the input image. Without having to start from scratch, our model may be trained from beginning to end and pixel by pixel. Our model's values are only 0.2M, which is 100 fewer than those of models like SegNet. Their model achieved superior performance.

In 2009[2], Chinese researchers proposed an algorithm for an unsupervised co-segmentation model, which can be applied to multiple foreground objects while being compatible with different backgrounds. To achieve quantitative semantic extraction, they extracted the color edges in the RGB space of each image, and this algorithm effectively separates the foreground and background objects in each image. By utilizing the correlation between different regions of an image and its interior, the coherence between the foreground and background images can be significantly improved. This algorithm can accurately classify and segment images based on their meaning without the need for supervision.

In 2021[3] ,Jiang, Li, Tan, Huang, Sun, and Kong created a multi-task semantic segmentation model. 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. Additionally, the loss function was modified and optimized to achieve multi-task information output. The proposed indoor scene semantic segmentation model showed strong performance and high efficiency and was able to clearly segment objects of different scales and adapt to uneven illumination conditions.

In 2021, Doshi and Yilmaz [4] proposed a new framework for video anomaly detection. The proposed algorithm, called MONAD, uses a statistical sequential approach to detect anomalies in videos. The authors evaluate the performance of MONAD and propose a practical approach to choose the detection threshold based on the desired false alarm rate. Additionally, they introduce 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

In 2020[5], the researchers in China proposed several deep learning-based approaches for semantic segmentation of photos. The proposed methods only conduct picture segmentation for specific purposes, and for effective analysis of indoor scenes, the robot must be aware of both the position and semantic information of objects. Therefore, there is a need for further research in deep learning-based semantic segmentation of indoor scenes. To address this, researchers proposed a method for multi-task semantic segmentation. This method combines image semantic segmentation with object recognition. The method involves dividing the image segmentation and target detection processes into two parallel branches for parallel computation, as well as using a shared feature extraction network. Although this approach may not be ideal for intricate backdrops or lighting variations.

Qinmin Ma, presented a new approach to anomaly detection in 2021[6]. The proposed approach involves constraining 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 accuracy and generalizability of the approach. However, when dealing with increasingly complex datasets, the complexity of the proposed procedure could increase.

Murugesan and Thilagamani[7] (2020) from the Department of Computer Science at M.Kumarsamy College of Engineering in India introduced a background subtraction method that employs the Maximally Stable External Region (MSER) feature extraction techniqu. 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 classification accuracy, specificity, and sensitivity of the output are reported to be 98.56%, 96.05%, and 98.21%, respectively.

In (2020)[8] the researchers put out 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

V. REVIEW ON ANOMALY DETECTION ON CROWED AND UNCROWDED AREA

In and of itself, anomaly detection is a burgeoning topic of research. Various methods have been proposed for both crowded and non-crowded situations. Existing approaches focus primarily on motion data, ignoring abnormality data resulting from object appearance changes. This makes them impervious to abnormalities not produced by motion outliers, such as a vehicle crossing a bridge with weight restrictions. Furthermore, in crowded circumstances with a dynamic background, a lot of clutter, and intricate occlusions, descriptors like optical flow, pixel change histograms, and other standard background subtraction approaches are difficult to be identified.

Anomaly detection and localization can be divided into two sub-problems:

1. Define crowd behaviors
2. Evaluate the "anomaly score" of a given behavior.

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 W ang 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%

Table1: Review on Anomaly Detection

VI. 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, the objective of this research is to develop an algorithm that is effective in detecting and tracking objects in complex surveillance environments

REFERENCES

[1]. 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 Informatics,

[2]. 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.

[3]. 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.

[4]. 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.

[5]. 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.

[6]. 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

[7]. 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.

[8]. 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