DOI: https://doi.org/10.15276/aait.07.2024.15
UDK 004.932.72

Anomaly detection in crowded scenes: technologies, challenges and opportunities

Ruslan Y. Dobryshev
ORCID: https://orcid.org/0009-0007-8639-3157; rdobrishev@gmail.com
Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

The paper discusses advancements in intelligent video surveillance systems, particularly focused on anomaly detection in crowded environments. These systems aim to enhance public safety by automatically detecting unusual behavior and potential threats in real-time. Traditional video surveillance, relying heavily on human monitoring, faces limitations like reduced concentration and slow response times. In contrast, intelligent surveillance uses machine learning and AI algorithms to process vast amounts of video data, identifying patterns that deviate from normal behavior. Crowd anomaly detection is essential in densely populated areas like transportation hubs, stadiums, and public squares. The diversity of anomalies, ranging from minor disruptions to serious threats such as theft or terrorist attacks, presents a challenge for these systems. Anomalies can be difficult to detect due to their unpredictable nature, and what constitutes an anomaly varies depending on the context. The paper highlights the need for robust systems that can adapt to various environmental conditions and distinguish between normal variations and genuine threats. While supervised machine learning models show promise, they often require large amounts of labeled data, which is difficult to obtain in real-world settings. Unsupervised models and deep learning techniques, such as Convolution Neural Networks, have been effective in analyzing crowd behavior and detecting anomalies. However, these methods still face challenges, including scalability, high false positive rates, and the need for real-time processing in large-scale environments. The paper concludes by addressing the limitations of current crowd anomaly detection methods, such as their computational costs, ethical concerns, and inability to detect novel anomalies. It suggests directions for future research, including the integration of advanced learning techniques to improve system performance and scalability.

Keywords: Intelligent video surveillance; anomaly detection; crowded environments; machine learning; public safety; deep learning; real-time monitoring; computer vision

***For citation*: Dobryshev R. Y. "Anomaly detection in crowded scenes: technologies, challenges and opportunities". *Applied Aspects of Information Technology*. 2024; Vol.7 No.3: 219–230. DOI: https://doi.org/10.15276/aait.07.2024.15**

INTRODUCTION, FORMULATION OF THE PROBLEM

In recent years, the field of intelligent video surveillance has experienced significant growth, driven by rapid advancements in artificial intelligence, machine learning, and big data technologies. These developments have made it possible to automate the analysis of video streams, significantly improving the efficiency and effectiveness of security monitoring systems [1].

The core objective of intelligent video surveillance systems is to automatically detect and analyze events in real-time, providing alerts for potential threats, identifying suspicious behavior, and ensuring public safety. A critical task in this domain is the detection of anomalies, particularly in crowded scenes such as transportation hubs, stadiums, public squares, and large shopping centers [2].

Anomalies in such environments can vary widely, ranging from minor disruptions to serious threats, including theft, violence, or even terrorist attacks (Fig. 1). The ability to quickly and accurately detect these anomalies is essential for enhancing public safety and enabling timely intervention in critical situations [3].

The importance of intelligent video surveillance has risen dramatically in recent years, especially in urban environments where the density of people and activities has increased exponentially. As cities become more congested and large-scale events become more frequent, the need for robust security systems becomes paramount [4].

Traditional methods of video surveillance, which rely heavily on human operators monitoring multiple screens, are proving inadequate to meet the demands of modern security requirements. Human operators face significant challenges in maintaining concentration over extended periods, leading to reduced effectiveness in detecting suspicious activities or responding to incidents in real time [5].

The rise of intelligent video surveillance technologies provides a solution to this problem by automating the analysis process, reducing the reliance on human oversight, and enhancing the precision and speed with which potential threats can be identified.

These systems utilize advanced algorithms and AI techniques to process vast amounts of video data, detecting patterns and behaviors that deviate from the norm. This capability is especially crucial in crowded environments where the sheer volume of people and activities can obscure anomalous behavior.

By leveraging machine learning and deep learning models, intelligent surveillance systems can continuously learn and improve their detection capabilities, making them more adept at identifying threats and preventing incidents before they escalate.

Anomaly detection in crowded scenes is a critical aspect of public safety and security. Crowded environments pose unique challenges, as they often serve as potential targets for criminal activities or emergencies [6].

The detection of anomalies in such settings is essential for mitigating risks and preventing disasters. For example, in a busy train station, an abandoned bag or erratic behavior could signal a security threat, such as a bomb or a potential terrorist attack. Similarly, the early detection of panic or chaotic movement in a large crowd could help security personnel respond swiftly to prevent injuries or loss of life [7].

Beyond security concerns, anomaly detection in crowded environments also plays a role in managing public order and ensuring the smooth flow of activities.

For instance, monitoring crowd behavior in stadiums or concert venues can help organizers manage crowd control, reduce the risk of stampedes, and ensure that the event proceeds safely. The ability to identify unusual behavior, such as individuals attempting to access restricted areas or groups forming unexpectedly, allows for proactive interventions that can prevent incidents before they occur [8].

The significance of anomaly detection in these environments cannot be overstated, as the consequences of failing to detect threats can be catastrophic. Real-time monitoring and detection systems that can quickly alert authorities to suspicious activities are crucial for minimizing the impact of security breaches and ensuring the safety of large groups of people [9].

While the importance of anomaly detection in crowded scenes is clear, the task itself is highly complex. One of the primary challenges lies in the diversity of anomalies that can occur. Anomalous behavior can take many forms, from subtle deviations in individual actions to large-scale disruptions involving multiple people [10].

This diversity makes it difficult to develop a one-size-fits-all approach to anomaly detection. Furthermore, what constitutes an anomaly may vary depending on the context and environment. For instance, running in a park may be considered normal behavior, but running in a crowded airport terminal could be deemed suspicious. The same action can have different implications based on situational factors, adding a layer of complexity to the task of anomaly detection.

Another significant challenge is the unpredictability of anomalies. Unlike pre-defined events, anomalies often do not follow a predictable pattern, making them difficult to detect using traditional rule-based systems [11].

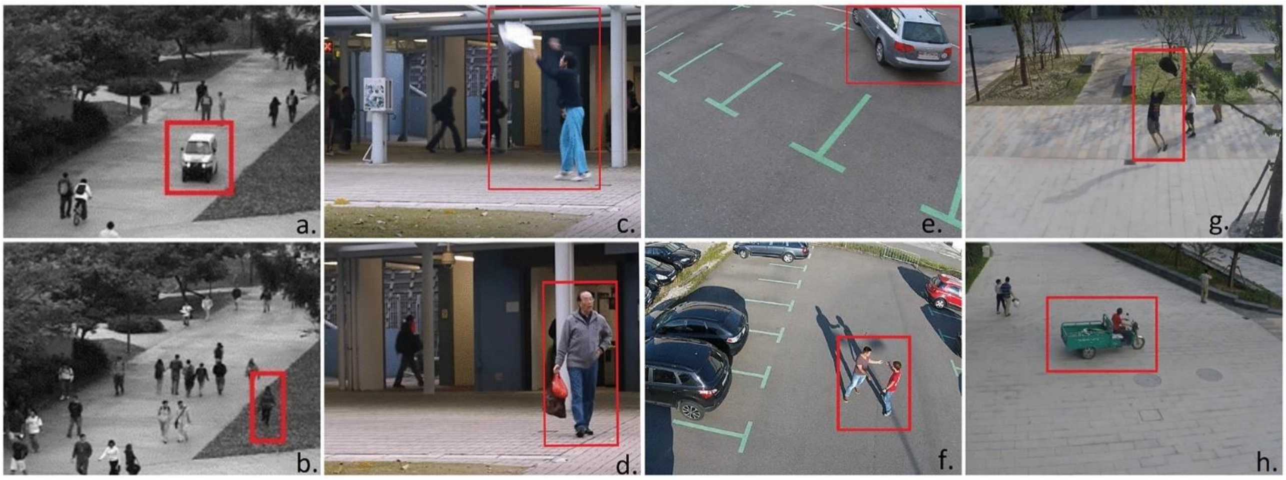

Fig. 1. **Illustration of some crowded scenes anomalies**
Source: **compiled by the [1]**

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

Machine learning and deep learning techniques have shown promise in addressing this issue by learning patterns from large datasets and identifying deviations that may indicate an anomaly. However, these models often require large amounts of labeled data to be effective, which can be difficult to obtain in real-world scenarios where anomalies are rare and diverse.

Moreover, crowded environments are often characterized by noise and interference, such as fluctuations in lighting, occlusion of objects or people, and dynamic movement patterns [12].

These factors can complicate the detection process, as algorithms must differentiate between normal variations in behavior and genuine anomalies. Environmental conditions, such as weather changes in outdoor venues or variable lighting in indoor spaces, further add to the complexity, requiring robust models that can adapt to different settings and conditions [13].

Despite significant advances in machine learning and computer vision, current methods for anomaly detection in crowded scenes still have notable limitations. Many existing approaches rely on supervised learning techniques that require large amounts of labeled data for training.

In practice, however, gathering labeled data for every possible type of anomaly is infeasible, particularly in crowded environments where anomalies are rare and unpredictable. As a result, supervised models may struggle to detect novel or previously unseen anomalies, limiting their effectiveness in real-world applications [14].

Another common limitation of existing methods is their tendency to produce a high number of false positives. Crowded environments are inherently dynamic, with constant fluctuations in movement, density, and behavior.

Algorithms designed to detect anomalies in such settings often flag normal variations as anomalies, leading to a flood of false alarms that can overwhelm security personnel and reduce the overall efficiency of the system. This issue highlights the need for more precise detection methods that can accurately distinguish between genuine threats and benign behavior [15].

Additionally, scalability remains a significant challenge for many anomaly detection systems. As the number of video cameras and the amount of data being processed increase, algorithms must be able to handle large-scale video streams in real-time without compromising accuracy or performance.

Current systems often struggle to meet these demands, particularly in large, complex environments with multiple cameras and high volumes of video data. The development of scalable, real-time anomaly detection systems capable of processing large video streams efficiently remains an open research challenge.

Thus, **the purpose of this study** is to analyze the latest achievements and challenges in the field of anomaly detection in crowded scenes.

1. CROWD ANOMALY DETECTION

To comprehend the general features of a crowd in a video, such as its density, flow, and demographic information, crowd anomaly detection is the process of analyzing the crowd. Congestion levels in a crowd may be measured by calculating the density of the crowd, which is the number of people per unit space. One way to determine the degree of mobility present in a crowd is to observe the flow of the crowd, which refers to the direction and speed of movement of individuals.

Over the last several years, a number of different methods have been developed to identify patterns in crowds. Image processing, computer vision, and machine learning are some of the aforementioned approaches. In order to extract information regarding the density and movement of a crowd from video footage, image processing methods may be used.

Some examples of these approaches are background removal and blob identification. Extraction of information on the demographic features of a crowd is accomplished via the use of computer vision methods such as object detection and tracking.

Neural networks and more recently, deep learning models are examples of machine learning methods that are used in order to examine the information that has been retrieved and to create predictions about the audience.

The identification of particular individuals or groups of people who are acting in an atypical manner, such as loitering, running, or moving against the flow of the crowd, is another definition of the phenomenon known as anomaly detection occurring in crowds.

Computer vision methods, such as object identification and object tracking, are used in this approach in order to monitor the movement of people inside a crowd. Following the tracking of a person, the behavior of that individual may be examined to determine whether or not it is normal or aberrant. The identification of anomalies and the study of such anomalies in social crowds has emerged as an important topic of research in recent years. The identification of crowd anomalies is a problem that is both practical and hard for computer vision [16].

This is because there are many different types of anomalous occurrences. When there are unusual crowds or unusually high levels of congestion, it is now able to do an automatic security analysis of the behavior of the crowd. Automated identification of anomalous behavior in the crowd is of the highest significance because of activities such as terrorist activities, fights, weird and suspicious motions, and other similar activities.

Within the context of conventional systems, it is the job of the operators to oversee the security monitoring in order to guarantee the highest level of safety. This is a considerable problem, which ultimately leads to decisions that are both expensive and wrong [17].

Consequently, the development of a system that is devoid of faults and does not experience any feeling of weariness, and that offers functionality in real time, will have adequate impacts on the management of crowd behavior.

Both the amount and quality of research in the field of crowd anomaly detection have increased as a result of the development of a number of complex algorithms and the availability of high processing horsepower. The issues that are involved in crowd behavior analysis are depicted by computer vision algorithms that make use of image processing, machine learning, and pattern recognition among other techniques [18].

A few of its most important uses are the management of crowds, the installation of video surveillance systems, and the development of intelligent public spaces. It is possible for the intelligent environment, which is vital for public safety, to assist in the redirection of the crowd and to assist the planner in designing the public area making the most of the space that is available.

2. CROWD ANOMALY DETECTION TAXONOMY

The flow and overall structure of a crowd anomaly detection system are shown in Fig. 2, which may be found here.

The raw video data that is gathered by the CCTV cameras may be considered the starting point for the anomaly detection system. The data from the sensors are then subjected to a process of feature extraction after being pre-processed using a variety of techniques to reduce the amount of signal noise. Components such as color, texture, motion, and form might be included in this category of characteristics. It is the purpose of this endeavor to recognize patterns within the video that may assist in distinguishing normal conduct from abnormal behavior [19].

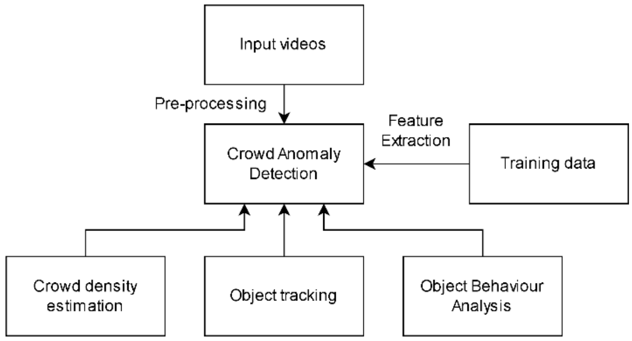

Fig. 2. **Crowd anomaly detection system structure**
Source: **compiled by the author**

The calculation of the crowd density, the tracking of objects, and the analysis of the behavior of objects are the three primary components that are commonly included in any crowd anomaly detection system. An explanation of how these components interact with one another is provided here. Estimating the density of the crowd is the first stage in the process of detecting crowd anomalies. This involves determining the number of individuals present in the scene.

Computer vision techniques such as background removal, foreground identification, and optical flow are some examples of the methods that may be used to accomplish this goal. One of the objectives is to determine the areas of the scene where people are present and then to estimate the number of persons that are present in each of those areas. Object tracking: After the crowd density has been approximated, the next step is to follow individual objects (i.e., individuals) in the scene over time.

This procedure is known as object tracking. In most cases, this is accomplished via the use of a tracking algorithm that provides a one-of-a-kind identifier to every item and automatically updates its location as it travels across the scene. Kalman filtering, particle filtering, and graph-based tracking are examples of some of the more prevalent tracking methods.

There are a great number of different tracking algorithms accessible available. Analysis of object behavior: After the tracking of individual items has been completed, the next stage is to study the behavior of the objects in order to identify peculiarities. To do this, it is often necessary to compare the behavior of each item to a regular behavior model that has been specified. For instance, the system could search for items that are traveling in a direction that is not typical, moving at a much faster or slower rate than what is anticipated, or remaining in the same location for a long length of time.

Rule-based methods, machine learning, and anomaly detection algorithms are just some of the many various approaches that may be used for the purpose of analyzing the behavior of objects. The flow of the system as a whole is often iterative, with the components associated with crowd density estimates and object tracking operating constantly and in real time [20].

Object behavior analysis is a component that is often executed at regular intervals (for example, every few seconds) in order to identify any irregularities in the behavior of the objects that are being monitored. The system is able to create a warning or initiate some other action (for example, turning on lights, sounding an alarm, or informing security staff) whenever it detects an abnormality.

The estimate of crowd density is a significant topic of study that has been shown to have practical implications in the management and monitoring of crowds in densely populated areas such as subway stations, sports stadiums, and convention centers. As a result of rising urbanization and population, it is not uncommon for a huge throng to congregate in a short amount of time [21].

The ability to accurately forecast the arrival of crowds and determine the density of those crowds is essential for the successful planning and administration of events that include crowds. As a result of the recent COVID-19 epidemic, social distancing regulations were put into place in order to prevent the virus from spreading.

This further emphasizes the need of crowd density assessment. Both counting objects and estimating the density map are the two basic strategies that are used in the process of crowd density estimate.

Because of their superior performance in both image and video sequences, CNN-based algorithms are the ones that are recommended. For the purpose of crowd density estimate and counting, techniques that are based on CNN are being used. Some examples of these techniques are Scale Pyramid Module Network and Attention Networks. Through the use of the attention mechanism, Attention Networks are able to count persons in photographs while taking into account scales.

This is accomplished by picking suitable global and local scales. In computer-aided design (CAD) systems, the process of tracking the crowd is very important since it requires the analysis of picture sequences to detect the velocity and trajectory of objects, particularly people [22].

To begin, the method involves identifying items inside a movie and then filtering them so that they may be tracked. The observation of the movement of pedestrians is a key component in the process of comprehending the behavior of crowds. The task of tracking items may be difficult since it entails keeping track of one or more objects over a period of time.

The movement of pedestrians in a crowd requires the use of automated tracking devices in order to keep up with such movement. In this method, the first and most critical step is to identify and define regions of interest, also known as ROIs. The view, position, and resolution of the camera are some of the characteristics that might make this task challenging from a variety of perspectives. After the features representative of the ROI have been retrieved, the tracker will then be able to follow the object of interest [23].

Training for these models may be either supervised or unsupervised from the beginning. It is possible to assess the level of congestion for the purpose of taking the appropriate measures, such as dispersing the crowd, when the density of the crowd reaches a certain threshold that has been established beforehand. After the objects have been monitored, the anomaly is investigated by making use of the items that are being tracked and the behavior of those objects. It is possible to make a conclusive determination in real time on whether the status of the crowd is normal or aberrant.

Crowd Anomaly detection based on the availability of labels. As can be seen in Fig. 3, the models that are used for the purpose of crowd anomaly detection are classified into two categories: supervised/semi-supervised neural networks, unsupervised neural networks, hybrid neural networks, and one-class neural networks.

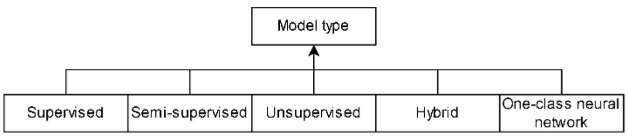

Fig. 3. **Crowd anomaly detection based on the availability of labels**
***Source:* compiled by the author**

Supervised learning is a sort of anomaly detection that is dependent on labeled data. In this type of anomaly detection, the abnormal behavior and normal behavior are explicitly established in advance. Similarly to supervised learning, semi-supervised learning is a method of anomaly detection that makes use of labelled data. However, it also makes use of unlabeled data in order to improve the performance of the model. In contrast to supervised learning, unsupervised learning does not need the use of labeled data in order to discover anomalies. In its place, it employs methods like as clustering and dimensionality reduction in order to analyze the data and find patterns. Radical Basis Function (RBF) and Support Vector Machine (SVM) classifiers are utilized as inputs in deep hybrid models rather than more standard approaches. RBF and SVM classifiers are employed for one class. An anomaly detector and a feature extractor are both need to be constructed on complex, high-dimensional domains in order to construct a viable anomaly detection model. In order to differentiate between the average data points and the outliers, one-class neural networks (OC-NN) combine the capabilities of deep neural networks with the purpose of extracting more rich representations of data. These representations may take the form of a hyperplane or a hypersphere [24].

Crowd anomaly detection based on the type of anomaly. Another way to classify crowd anomaly detection methods is according to the kind of abnormality that they are able to identify and deal with. Point anomalies, contextual or conditional anomalies, and collective or group anomalies are the three primary forms of anomalies that are dealt with, as seen in Fig. 4.

A significant portion of the published works are dedicated to highlighting abnormalities.

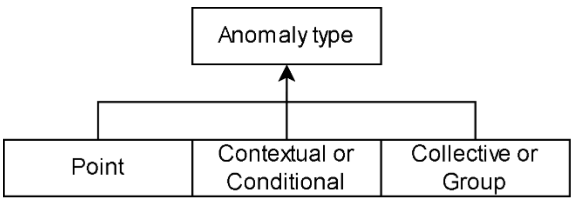

Fig. 4. **Crowd anomaly detection based on the type of anomaly**
***Source:* compiled by the author**

Point anomalies are typically indicative of an irregularity or deviation that takes place of its own accord and may or may not have a particular significance.

A contextual anomaly, also known as a conditional anomaly, is a data instance that may be considered unusual in a certain context. This kind of anomaly is also known as a conditional anomaly. The presence of a contextual anomaly is discovered by taking into account both behavioral and contextual factors. Additionally, time and location are two components of context that are used rather often. On the other hand, the behavioral features may consist of a pattern of money expenditure, the occurrence of system log events, or any other characteristic that is representative of usual behavior [25].

A collective or group anomaly is an aberrant collection of individual data points in which each individual point, when evaluated independently, seems to be a typical instance of data but, when observed collectively, exhibits an unexpected feature. Collective or group anomalies are also known as groups of anomalies.

3. RECENT PROGRESS ON CROWD ANOMALY DETECTION

Over the last several years, the area of crowd anomaly detection and analysis has seen substantial development. This surge in growth may be attributed to the growing awareness of the significance of comprehending and effectively regulating the behavior of crowds. The proliferation of surveillance cameras and social media has resulted in an increasing quantity of data that may be used for the purpose of evaluating the behavior of crowds.

Numerous researches have been carried out on a variety of areas of crowd anomaly identification as a direct consequence of this consequence. On the other hand, being able to get a full grasp of the current state of research in this topic might be challenging due to the abundance of material that is currently accessible.

For the purpose of determining the aberrant behavior of crowds, a comprehensive analysis of the many traditional methodologies, including the Spatial-Temporal Technique (STT), optical flow, Gaussian mixture model (GMM), and Hidden Markov model (HMM), has been carried out in [26]. Recent research in the field of crowd anomaly analysis has made extensive use of deep learning techniques. [27] provides a presentation of the many characteristics of the Convolution Neural Network (CNN) as well as the numerous optimization strategies that are used in the context of crowd behavior research.

In [28], an additional literature study was carried out on the topic of intelligent video surveillance applications that make use of a variety of deep learning approaches for crowd identification. Specifically, this research investigated the use of Long-Short Term Memory (LSTM) networks, as well as VGG16 and YOLO technologies. In addition, [29] carried out an in-depth investigation on the several deep learning techniques that are used for crowd counting and analysis, both of which are essential elements of crowd anomaly detection.

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

For the purpose of displaying the different developments in CNN and classical machine learning techniques, a study and comparison of the numerous approaches that are applied for counting the crowd has been conducted [30].

The approach of Aggregation of Ensembles (AOE), which is a mixture of four classifiers over sub-ensembles of three tuned Convolutional Neural Networks (CNNs) on crowd datasets, is presented by [31] as a way to identify irregularities in videos of crowded situations by the use of majority voting. In order to complete the processing of each sub-ensemble individually, a distinct classifier is used.

For the purpose of social multiple-instance learning, a dual branch network was suggested as a framework in [32]. The neural network that is used in this method is a two-stream neural network, which includes both an interactive dynamic stream and a spatiotemporal stream. The RGB video clips are processed by the spatiotemporal stream computer. After the video has been segmented into smaller segments using a video segmentation approach, a 3D ConvNets (C3D) model that has previously been trained on Kinetics and UCF-101 is used in order to extract features from the video. After the features have been input into a one-dimensional dependency attention capture module, the output is the next thing that is fed into a fully connected network. It is the force maps of social interactions that are received as input by the second channel. In order to build maps of the social interactions that take place inside a scene, the social force model [33] is used. Using the receiver operating characteristic (ROC) curve and the area under the curve (AUC) measures, they determine whether or not their technique is successful when applied to the UCF Crime dataset [34].

In [35], a generic adversarial network-based anomalous behavior analysis in the enormous crowd was developed. This analysis takes into consideration a case study of the Hajj pilgrimage. The optical flow technique was used in this study in order to extract the dynamic characteristics. For the purpose of distinguishing between normal and deviant actions in big crowd scenarios, it employs a transfer learning method in conjunction with U-Net and Flownet. With a 99.4 % and 97.1 % accuracy rate, respectively, this system has shown a very high level of accuracy in smaller video scenes, such as UMN and UCSD.

Understanding behavior becomes a challenging endeavor when one considers the fact that anomalies are not well characterized and would occur only in a very small percentage of cases. In an effort to build learning algorithms that are capable of accurately identifying anomalies in videos, researchers have been working to find solutions to these problems. Generative Cooperative Learning (GCL) is the foundation of a novel technique that has been used in [36] with the purpose of solving the low frequency of abnormalities and contributing to the avoidance of human annotations.

According to [37], an architecture for a deep convolutional neural network (DCNN) was presented for the purpose of identifying crowd abnormalities. Ten convolutional layers and three max pooling layers were used in the design, which was based on the VGG16 model. They used six convolutional layers with a dilation rate of two and a kernel size of three by three in order to count the number of people in the crowd.

The majority of the robustness of crowd anomaly detection is evaluated based on the temporal consistency among the frames. Through the use of optical flow analysis, the temporal characteristics were taken into consideration for the purpose of anomaly prediction based on motion information. On the other hand, there need to be some kind of system that can differentiate between false sequences and actual sequences in terms of temporal consistency, which might result in either abnormal or typical behavior.

Both the optical flow approaches, which are useful for studying the short-term temporal connection between consecutive frames, are able to identify anomalous activity with a high degree of accuracy. However, these strategies ultimately fail, particularly in films where events that are dependent on a long-term temporal connection take place. An innovative approach that is built on a bi-directional architecture that creates inconsistencies on three separate levels, including temporal-sequence, cross-modal, and pixel, has been presented as a means of resolving such problems. Regularization of the predictive consistency is achieved by the implementation of the bidirectional predictive network in this study. The discriminator that was constructed in the course of the study is able to identify the long-term temporal link that exists within the video sequences. On a number of datasets, including UCSD Pred2, ShanghaiTech, and CUHK Avenue, this technique delivered superior results to any other learning method that is considered to be state-of-the-art.

4. LIMITATIONS AND CHALLENGES

In situations involving crowds that include a range of variables, a crowd anomaly detection algorithm that is reliable will attempt to assess both the local and global density of the crowd and will reliably forecast the behavior of the crowd. Changes in the environment have a substantial influence on the performance of the model. As a result, it is of the utmost importance to first have an understanding of these challenges and the ways in which they could influence the performance of the model. Having a comprehensive understanding of these challenges is beneficial to the process of developing models that are more dependable.

It is challenging to observe the behavior of crowds in a variety of contexts. To provide an example, it is a lot less difficult to count and monitor the persons in photos that were taken by a single CCTV camera (like those in the Mall dataset), as opposed to counting people in images that were acquired by a large number of security cameras (like those in the WorldExpo'10 dataset).

In the course of drone surveillance, the scene is often altered, which, when paired with other differences such as size fluctuations, makes the task more challenging.

When objects that belong to the same class (in this example, people) appear at different sizes in a single shot as well as across many photographs, this phenomenon is referred to as scale variation. The size of a picture is determined by two factors: the distance (between the camera and the objects) and the perspective influence that occurs inside the same image themselves [38].

Alterations in scale may also be seen in photographs that take place at varying resolutions. Within the realm of research pertaining to the identification of crowd anomalies, scale variation is among the most prominent concerns that greatly impact the performance of models.

There are varying numbers of individuals or other topics of interest in the many photographs that are available. Visuals with a low density are often easier to comprehend than those with a high density. In a similar vein, another challenge arises when the same photograph depicts different populations of individuals in different regions.

It's possible that the distribution of items in crowd images will change depending on the circumstances. As an example, seats in a sports stadium are distributed uniformly among everyone, and there is a consistent distance between each item. On the other hand, in congested streets, things may be placed in a manner that is not consistent. Crowds that are consistently scattered may be estimated with more precision than crowds that are not uniformly dispersed, provided that there are no other factors that might potentially affect the accuracy.

When it comes to video analysis, occlusion has proven to be a significant obstacle, and it is much more challenging to identify anomalies in crowds. Occlusion is a concept that defines the way in which items overlap. On the other hand, the term "inter-class occlusion" refers to the overlapping of dissimilar things, such as vehicles, walls, and other people, as opposed to "intra-class occlusion," which describes the overlapping of items that are similar to one another, such as humans.

The process of dealing with occlusion is usually challenging. The existence of occlusion makes it difficult for both the object detectors and the annotators to appropriately annotate the items and predict them. This is because occlusion makes it difficult to see the objects.

Through the interweaving of semantic information, occlusion makes it difficult to differentiate between the boundaries of objects occurring inside frames. There is also the possibility that it will be difficult to learn when the pixel values of the item are similar to those of the backdrop.

Because occlusion makes it harder to recognize people within a crowd, it may reduce the performance of crowd anomaly detection systems. This can result in either false positives or false negatives during the detection process. It is possible to lessen the impacts of occlusion by the use of techniques such as the use of several cameras or machine learning algorithms; nonetheless, it may still be a substantial difficulty in crowds that are dense and complicated [39].

The illumination of a picture might fluctuate throughout the day and in different parts of the same image due to the diverse lighting conditions that are present. since of this, learning is made more difficult since the same item (like people) in the same picture will have different pixel values. Alterations in the weather, noise and pixilation in photographs, rotation of objects, and other factors are also factors that contribute to the difficulty of computer-aided design (CAD).

The following is a list of some of the limitations of the crowd anomaly detection technologies that are currently in use:

1. In real-world applications, where computer resources are restricted, the high computational cost of some approaches, such as those based on deep learning, might be a constraint. These methods demand a huge amount of computational resources to train and test, which can be a drawback.

2. The need for annotated data: Many approaches, such as those based on deep learning, need a significant quantity of labeled data in order to

train the models. This might be a drawback in real-world applications where the data is either restricted or difficult to annotate.

3. Ethical and privacy considerations: some approaches, such as those based on face recognition, may give rise to privacy concerns. It is essential to make certain that the method is used in a manner that is in accordance with the applicable laws and regulations.

4. Many approaches are intended to identify known anomalies, and it is possible that they are not able to discover new or unknown abnormalities. A limited capacity to detect novel anomalies is one of the limitations of these methods. The detection of anomalies is the primary emphasis of many approaches; however, these methods do not provide any insight into the underlying cause of the abnormality.

5. Limited scalability: Many techniques are not built to manage huge crowds and may not be able to scale to accommodate massive volumes of data. Another limitation is that they are not scalable. Many approaches concentrate on identifying a single form of anomaly, and it is possible that they are unable to manage numerous types of anomalies at the same time. This restricts their capacity to handle multiple anomalies concurrently.

CONCLUSIONS

This study presents a comprehensive overview of crowd anomaly detection and the significance of this technique in the context of real-world surveillance and security. This paper provides an overview and summary of the most recent research on crowd analysis, with a particular emphasis on the primary components of crowd anomaly detection.

These components include crowd density estimate, object tracking, and object behavior analysis. Additionally, a summary of research based on several taxonomies is shown in the study.

Regarding the evaluation of model performance, as well as the issues that are now being faced in the field, there was also a discussion. The paper, in light of the review, provides directions for future research. These directions include the necessity of conducting model generalization for various anomalies in various scenarios, the development of application-specific crowd anomaly detections, and the efficient selection of the most appropriate models for analysis in order to reduce the amount of resources that are used without necessity and the amount of carbon emissions that are produced.

This paper also provides directions for future research, such as the incorporation of generative models, graph-based methods, reinforcement learning, transfer learning, online learning, ensemble methods, multi-task learning, domain adaptation, active anomaly detection, and meta-learning. These are all methods that have the potential to significantly improve the performance of crowd anomaly detection systems and address the limitations that are currently present in the field.

REFERENCES

1. Zaheer, M. Z., Mahmood, A., Khan, M. H., Segu, M., Yu, F. & Lee, S.-I. “Generative cooperative learning for unsupervised video anomaly detection”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022. p. 14744–14754. DOI: http://dx.doi.org/10.48550/arXiv.2203.03962.

2. Chen, K., Loy, C. C., Gong, S. & Xiang T. “Feature mining for localised crowd counting”. *Proceedings of the British Machine Vision Conference (BMVC)*. 2012. p. 1–11. DOI: http://dx.doi.org/10.5244/C.26.21.

3. Tan, T. H., Gochoo, M., Huang, S. C., Liu, Y. H., Liu, S. H. & Huang, Y. F. “Multi-Resident activity recognition in a smart home using RGB activity image and DCNN”. *IEEE Sensors Journal*. 2018; 18 (20): 9718–9727. DOI: https://doi.org/10.1109/JSEN.2018.2866806.

4. Erfani, S. M., Rajasegarar, S., Karunasekera, S. & Leckie, C. “High-Dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning”. *Pattern Recognition*. 2016; 58: 121–134. DOI: https://doi.org/10.1016/j.patcog.2016.03.028.

5. Wu, C., Shao, S., Tunc, C., Dolz, J. M. & Erdogan, A. T. “An explainable and efficient deep learning framework for video anomaly detection”. *Cluster Computing*. 2022; 25 (5): 2715–2737. DOI: https://doi.org/10.1007/s10586-021-03439-5.

6. Zabłocki, M., Stawicki, J., Kołakowski, A. & Zieliński, W. “Intelligent Video Surveillance Systems for Public Spaces: A Survey”. *Journal of Theoretical and Applied Computer Science*. 2014; 8 (4): 3–18.

7. Xia, L. & Li, Z. “A new method of abnormal behavior detection using LSTM network with temporal attention mechanism”. *The Journal of Supercomputing*. 2021; 77 (4): 3223–3241. DOI: https://doi.org/10.1007/s11227-020-03391-y.

8. Luo, W., Liu, W., Lian, D. & Gao, S. “"Video anomaly detection with sparse coding inspired deep neural networks”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021; 43 (3): 1070–1084. DOI: https://doi.org/10.1109/TPAMI.2019.2944377.

9. Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A. & Xu, L. Q. “Crowd analysis: A survey”. *Machine Vision and Applications*. 2008; 19 (5-6): 345–357. DOI: http://dx.doi.org/10.1007/s00138-008-0132-4.

10. Kim, K., Byun, J., Choi, J. & Youn, H. Y. “Human activity recognition using Wi-Fi signals and skeleton data”. *IEEE Access*. 2018; 6: 61634–61647. DOI: http://dx.doi.org/10.1155/2018/6163475.

11. Kumar, K. S. & Bhavani, R. “Human activity recognition in egocentric video using HOG, GIST and color features”. *Multimedia Tools and Applications*. 2022: 81 (4): 6433–6450. DOI: https://link.springer.com/article/10.1007/s11042-018-6034-1.

12. Nishi, K. & Miura, J. “Generation of human depth images with body part labels for complex human pose recognition”. *Pattern Recognition*. 2017; 71: 402–413. DOI: https://doi.org/10.1016/j.patcog.2017.06.006.

13. Doshi, K. & Yilmaz, Y. “Any-Shot sequential anomaly detection in surveillance videos”. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020. p. 934–935. DOI: https://doi.org/10.1109/CVPRW50498.2020.00475.

14. Cai, R., Zhang, H., Liu, W., Gao, S. & Hao, Z. “Appearance-Motion memory consistency network for video anomaly detection”. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021; 35 (2): 938–946. DOI: https://doi.org/10.1609/aaai.v35i2.16177.

15. Roy, P. K. & Om, H.. “Suspicious and violent activity detection of humans using HOG features and SVM classifier in surveillance videos”. *In Advances in Soft Computing and Machine Learning in Image Processing*, *Springer*. 2018. p. 277–294. DOI: https://doi.org/10.1007/978-3-319-63754-9_13.

16. Zhang, Q., Lin, W. & Chan, A. B. “Cross-View cross-scene multi-view crowd counting”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. p. 557–567. DOI: https://doi.org/10.48550/arXiv.2205.01551.

17. Moustafa, N., Hu, J. & Slay, J. “A holistic review of network anomaly detection systems: A comprehensive survey”. *Journal of Network and Computer Applications*. 2019; 128: 33–55. DOI: https://doi.org/10.1016/j.jnca.2018.12.006.

18. Liu, Z., Nie, Y., Long, C. & Huang, Q. “A hybrid video anomaly detection framework via memory-augmented flow reconstruction”. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021. p. 13561–13570. DOI: https://doi.org/10.48550/arXiv.2108.06852.

19. Wang, T., Xu, X., Shen, F. & Hu, Y. “A cognitive memory-augmented network for visual anomaly detection”. *IEEE/CAA Journal of Automatica Sinica*. 2021; 8 (7): 1296–1307. DOI: https://doi.org/10.1109/JAS.2021.1004045.

20. Lalit, R., Gupta, A. & Singh, V. “Crowd abnormality detection in video sequences using a supervised convolutional neural network”. *Multimedia Tools and Applications*. 2022; 81 (4): 5259–5277. DOI: https://doi.org/10.1007/s11042-021-11781-4.

21. Ma, J., Dai, Y. & Hirota, K. “A survey of video-based crowd anomaly detection in dense scenes”. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2017; 21 (2): 235–246. DOI: http://dx.doi.org/10.20965/jaciii.2017.p0235.

22. Babaee, M., Dinh, D. T. & Rigoll, G. “A deep convolutional neural network for video sequence background subtraction”. *Pattern Recognition*. 2018; 76: 635–649. DOI: https://doi.org/10.1016/j.patcog.2017.09.040.

23. Fernandes, G., Rodrigues, J. J. P. C., Carvalho, L. F., Al-Muhtadi, J. F. & Proença, M. L. “A comprehensive survey on network anomaly detection”. *Telecommunication Systems*. 2019; 70 (3): 447–489. DOI: https://doi.org/10.1007/s11235-018-0475-8.

24. Dhiman, C. & Vishwakarma, D. K. “A review of state-of-the-art techniques for abnormal human activity recognition”. *Engineering Applications of Artificial Intelligence*. 2019; 77: 21–45. DOI: https://doi.org/10.1016/j.engappai.2018.08.014.

25. Guan, Y., Hu, W. & Hu, X. “Abnormal behavior recognition using 3D-CNN combined with LSTM”. *Multimedia Tools and Applications.* 2021; 80 (12): 18787–18801. DOI: https://doi.org/10.1007/s11042-021-10667-9.

26. Tripathi, G., Singh, K. & Vishwakarma, D. K. “Convolutional neural networks for crowd behaviour analysis: A survey”. *The Visual Computer.* 2019; 35 (5): 753–776. DOI: https://doi.org/10.1007/s00371-018-1499-5.

27. Blanchard, G., Lee, G. & Scott, C. “Semi-Supervised novelty detection”. *The Journal of Machine Learning Research.* 2010; 11: 2973–3009. DOI: https://doi.org/10.5555/1756006.1953028.

28. Amrutha, C., Jyotsna, C. & Amudha, J. “Deep learning approach for suspicious activity detection from surveillance video”. *IEEE Conference on Innovative Mechanisms for Industry Applications (ICIMIA).* 2020. p. 335–339. DOI: https://doi.org/10.1109/ICIMIA48430.2020.9074920.

29. Mehran, R., Oyama, A. & Shah, M. “Abnormal crowd behavior detection using the social force model”. *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition.* 2009. p. 935–942. DOI: https://doi.org/10.1109/CVPR.2009.5206641.

30. Bendali-Braham, M., Weber, J., Idoumghar, L. & Muller, P. “Recent trends in crowd analysis: A review”. *Machine Learning with Applications*. 2021; 4: 100023. DOI: https://doi.org/10.1016/j.mlwa.2021.100023.

31. Wang, P., Li, W., Zhou, Z. & Ogunbona, P. “RGB-D-Based human motion recognition with deep learning: A survey”. *Computer Vision and Image Understanding.* 2018; 171: 118–139. DOI: https://doi.org/10.1016/j.cviu.2018.04.007.

32. Zhang, D., Huang, C., Liu, C. & Xu, Y. “Weakly supervised video anomaly detection via transformer-enabled temporal relation learning”. *IEEE Signal Processing Letters.* 2022; 29: 1095–1099. DOI: https://doi.org/10.1109/LSP.2022.3175092.

33. Ullah, A., Muhammad, K., Del Ser, J., Baik, S. W. & de Albuquerque, V. H. C. “Activity recognition using temporal optical flow convolutional features and multilayer LSTM”. *IEEE Transactions on Industrial Electronics.* 2019; 66 (12): 9692–9702. DOI: https://doi.org/10.1109/TIE.2018.2881943.

34. Ladjailia, A., Bouchrika, I., Tabia, H. & Taleb-Ahmed, A. “Human activity recognition via optical flow: decomposing activities into basic actions”. *Neural Computing and Applications.* 2020; 32 (13): 16387–16400. DOI: https://doi.org/10.1007/s00521-018-3951-x.

35. Bansod, S. D. & Nandedkar, A. V. “Crowd anomaly detection and localization using histogram of magnitude and momentum”. *The Visual Computer*. 2020; 36 (3): 609–620. DOI: https://doi.org/10.1007/s00371-019-01647-0.

36. Jagadeesh, B. & Patil, C. M. “Video-Based human activity detection, recognition, and classification of actions using SVM”. *Transactions on Machine Learning and Artificial Intelligence.* 2018; 6 (3): 22–32. DOI: https://doi.org/10.14738/tmlai.66.5287.

37. Perera, P. & Patel, V. M. “Learning deep features for one-class classification”. *IEEE Transactions on Image Processing.* 2019; 28 (11): 5450–5463. DOI: http://dx.doi.org/10.1109/TIP.2019.2917862.

38. Marsden, M., McGuinness, K., Little, S. & O'Connor, N. E. “ResNetCrowd: A residual deep learning architecture for crowd counting, Violent Behaviour detection and crowd density level classification”. *Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).* 2017. p. 1–7. DOI: https://doi.org/10.48550/arXiv.1705.10698.

39. Xu, K., Sun, T. & Jiang, X. “Video anomaly detection and localization based on an adaptive intra-frame classification network”. *IEEE Signal Processing Letters.* 2022; 29: 631–635. DOI: https://doi.org/10.1109/TMM.2019.2929931.

Conflicts of Interest: the authors declare no conflict of interest

Received 29.07.2024
Received after revision 04.09.2024
Accepted 18.09.2024

DOI: https://doi.org/10.15276/aait.07.2024.15
УДК 004.932.72

Виявлення аномалій у місцях масового скупчення людей: технології, виклики та можливості

Добришев Руслан Євгенович
ORCID: https://orcid.org/0009-0007-8639-3157; rdobrishev@gmail.com
Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У статті обговорюються досягнення в області інтелектуальних систем відеоспостереження, зокрема, виявлення аномалій у місцях масового скупчення людей. Ці системи спрямовані на підвищення громадської безпеки шляхом автоматичного виявлення незвичайної поведінки та потенційних загроз у режимі реального часу. Традиційне відеоспостереження, яке значною мірою покладається на людський моніторинг, стикається з такими обмеженнями, як знижена концентрація уваги і повільний час реакції. На відміну від нього, інтелектуальне відеоспостереження використовує алгоритми машинного навчання і штучного інтелекту для обробки величезних обсягів відеоданих, виявляючи патерни, які відхиляються від нормальної поведінки. Виявлення аномалій у натовпі має важливе значення в густонаселених районах, таких як транспортні вузли, стадіони та громадські площі. Різноманітність аномалій, від незначних порушень до серйозних загроз, таких як крадіжки або терористичні атаки, є викликом для цих систем. Аномалії буває важко виявити через їхню непередбачувану природу, а те, що є аномалією, залежить від контексту. У статті підкреслюється потреба в надійних системах, які можуть адаптуватися до різних умов навколишнього середовища і розрізняти нормальні відхилення від реальних загроз. Хоча керовані моделі машинного навчання є багатообіцяючими, вони часто вимагають великих обсягів маркованих даних, які важко отримати в реальних умовах. Неконтрольовані моделі та методи глибокого навчання, такі як згорткові нейронні мережі, виявилися ефективними в аналізі поведінки натовпу та виявленні аномалій. Однак ці методи все ще стикаються з проблемами, включаючи масштабованість, високий рівень помилкових спрацьовувань і необхідність обробки в реальному часі у великомасштабних середовищах. У статті розглядаються обмеження сучасних методів виявлення аномалій у натовпі, такі як обчислювальні витрати, етичні проблеми та нездатність виявляти нові аномалії. Запропоновано напрямки майбутніх досліджень, зокрема інтеграцію передових методів навчання для покращення продуктивності та масштабованості системи.

Ключові слова: інтелектуальне відеоспостереження; виявлення аномалій; скупчення людей; машинне навчання; громадська безпека; глибоке навчання; моніторинг в реальному часі; комп'ютерний зір

ABOUT THE AUTHOR

Ruslan Y. Dobryshev - PhD Student of Artificial Intelligence and Data Analysis Department. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0009-0007-8639-3157, rdobrishev@gmail.com
***Research field*:** Deep learning; crowd analysis; video processing; motion tracking; intelligent surveillance; computer vision

Добришев Руслан Євгенович - аспірант кафедри Штучного інтелекту та аналізу даних. Національний університет «Одеська Політехніка», пр. Шевченка, 1. Одеса, 65044, Україна