IJoC

KOCON.a
The Korea Contents Association

ISSN:1738-6764, eISSN:2093-7504

Deep Learning Applied Abnormal Human Behavior Detection in Video Surveillance Systems – A Survey

Naejoung Kwak [1], Byoungyup Lee [2,*]

[1] Dept. Information Security of AI·SW Creative Convergence University, Pai Chai University; knj0125@pcu.ac.kr
[2] Dept. Information Security of AI·SW Creative Convergence University, Pai Chai University; bylee@pcu.ac.kr
* Correspondence

https://doi.org/10.5392/IJoC.2024.20.4.084

Manuscript Received 9 November 2024; Received 27 November 2024; Accepted 2 December 2024

Abstract: *As society develops, the demand for security is rapidly increasing. Accordingly, there is growing interest in research on methods to detect and prevent abnormal behavior using surveillance cameras in public places and private spaces such as shopping malls and airports for human safety. Many detection techniques based on deep learning models have been researched in the field of abnormal behavior detection. However, due to the lack of labeled abnormal behavior data, there are significant difficulties in developing an effective detection system. This paper surveys methods for deep learning methods to detect abnormal human behavior in surveillance video and presents recent techniques. First, I will introduce popular datasets that have often been used in previous research. After that, we categorized the existing methods for detecting abnormal behavior using deep learning into three types: supervised learning, unsupervised learning, and partially-supervised learning. We then explained the basic concepts and advantages of each method and summarized their shortcomings. We also briefly describe future research directions based on the advantages and disadvantages of each method. Based on this, it is expected that the technology of video surveillance systems that apply abnormal behavior detection will further develop.*

Keywords: Abnormal Human Behavior Detection; Abnormal Human Activity Recognition; Anomaly Detection; Video Surveillance; Deep learning

1. Introduction

As the need for security increases today, surveillance cameras are proliferating in public and private places such as shopping malls, airports, streets, and train stations. Surveillance cameras are used to monitor daily activities and detect abnormal behaviors in progress, such as road accidents, assaults, pedestrian fights, robberies and traffic congestion. These tasks play an important role in crime prevention and counter-terrorism. However, this process requires constant human attention, and it is a monotonous and tedious task that requires continuous observation. In addition, surveillance systems generate a lot of redundant video data, and it also takes a lot of time to manually analyze and process the large amount of stored images. Therefore, an automated system that can recognize abnormal and suspicious behaviors in stored images and alert the user is needed. A system that automatically analyzes, processes, and detects abnormal phenomena in images acquired from surveillance cameras is called an abnormal behavior detection system.

Anomalies can be defined as events that deviate from normal behavior [1], and anomaly detection is one of the most popular and widely studied topics in the fields of computer vision and machine learning. Human anomalies, or abnormal human behavior detection includes identifying abnormal behaviors or situation conversion of a subject, and behaviors that diverge from the norm are considered abnormal [2]. In surveillance cameras, these behaviors are analyzed by analyzing images from fixed cameras [3, 4]. Therefore, the recognition of abnormal human behavior is based on behavior recognition method. Traditional behavior recognition methods extract features only at important characteristic points where behaviors frequently occur. Behavior recognition is achieved by extracting specific vectors and applying them to various pattern classifiers for action recognition. However, these existing methods only recognize behavior based on predetermined features and struggle to cope with various changes in objects or actions, resulting in decreased accuracy. Recently, there has been active research aimed at solving action recognition methods using deep learning algorithms based on artificial neural networks.

Deep learning techniques for detecting abnormal human behavior can be classified into three types according to learning methods: supervised learning, unsupervised learning, and partially supervised learning [5]. Supervised learning is an approach where the model learns normal behavior patterns from the inputs during the training step [6]. Then, this model detects abnormal behavior by identifying deviations from the learned patterns or by comparing new data with the identified normal behavior clusters. Unsupervised learning is a method of training models using unlabeled data, and because there are no correct labels for the input data, the model learns the patterns in the dataset by itself. This method is reconstruction-based unsupervised learning [7] and generative-based [8]. Partially supervised learning is a algothm of learning using a partially labeled dataset, and can be classified into weakly supervised learning and semi-supervised learning [9]. Deep learning methods for detecting abnormal behavior require a lot of data, and their performance is poor when there is little data. However, in the dataset for training deep learning models, there is an imbalance in the data because there is less abnormal behavior data compared to normal behavior data. Semi-supervised learning can use labeled data to create a model, and then classify unlabeled data with that model to generate labels. Unsupervised learning can detect anomalies even without labeled data. Therefore, semi-supervised learning and unsupervised learning are used as alternatives to solve the problem of insufficient abnormal behavior data in the field of anomaly detection [10].

In this paper, we survey the recent deep learning techniques applied to the automation of abnormal behavior detection in surveillance cameras. We review the research results on supervised, unsupervised, and partially supervised learning, and also survey popular open datasets used to train models. We compare the performances of the three methods and discuss their advantage and disadvantage. We also examine open research problems in the field of abnormal human behavior detection in surveillance cameras.

The structure of this paper is as follows. Chapter 2 introduces research related to abnormal behavior detection, and Chapter 3 examines the datasets used for abnormal behavior detection. Chapter 4 examines research on applying deep learning techniques to abnormal behavior detection by classifying them into three training types: supervised learning, unsupervised learning, and partially supervised learning, and compares and analyzes each method. Chapter 5 examines future research directions, and Chapter 6 summarizes the contents of this paper and concludes.

2. Datasets

Several datasets were used to benchmark various research methods for abnormal human behavior recognition. Figure 1 shows samples of each dataset.

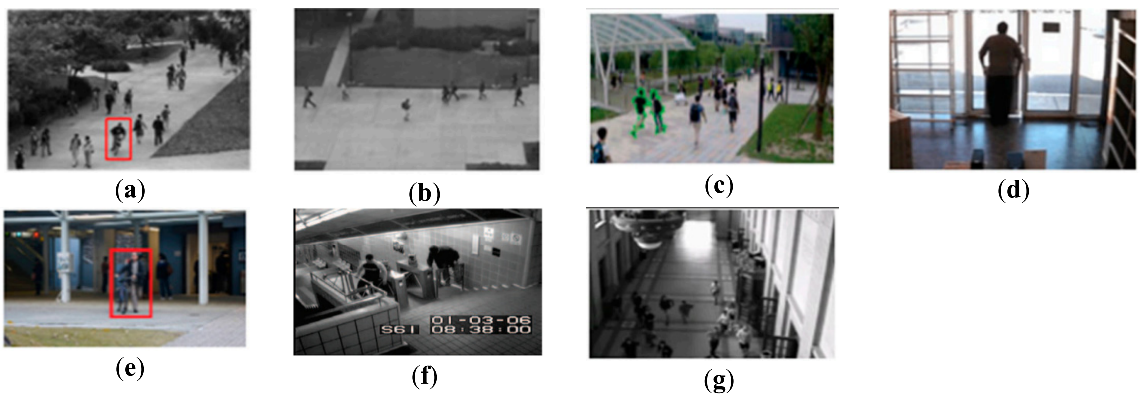

Figure 1. Samples of abnormal behaviors in each data set in Table 1: (a) Ped1; (b) Ped2; (c) ShanghaiTech; (d) UCF-Crime; (e) UMN; (f) Subway; (g) Avenue

The UCSD dataset [11] consists of 70 video acquired from above to monitor pedestrian walkways. This dataset consists of two video sets, Ped1 and Ped2, and was created in 2013. The ShanghaiTech (ST) Campus dataset [12] was created by ShanghaiTech University in 2017 and consists of 13 scenes. This dataset contains a total of 437 videos and 130 abnormal events. The UCF-Crime dataset [13] was consists of 1900 videos spanning 128 hours and covers 13 real-world anomalies, including fights, property damage, and robberies.

The UMN dataset [14] contains three indoor and outdoor scenes and consists of 11 abnormal events. This dataset includes 22 videos. The Subway data set [15] contains 209 frames at the exit and 150 frames at the entrance. This dataset contains 19 events. The Avenue dataset [16] was created in 2013. It consists of 16 videos for training and 21 videos for testing, with 14 events.

3. Abnormal human behavior detection

Human action recognition (HAR) is a fundamental research topic in computer vision and is being used usefully in various fields. Especially in environments where public safety is important, it is very important to quickly detect abnormal behaviors such as assault or abandonment. For this reason, abnormal behavior detection is used in video surveillance to monitor people's behavior and activities for the purpose of ensuring security and providing guidance. The problem of abnormal activity recognition in video is very diverse and there is no single approach that can solve all problem cases.

In the video, human action recognition is performed by analyzing spatial and visual information and extracting features. Therefore, specific general background knowledge about the video processing domain is required. Before applying deep learning techniques to recognize human behavior, features were extracted from the image sequence input in the video, and machine learning algorithms were trained based on the extracted features to recognize human behavior. In particular, methods such as random forests (RF) [17], Bayesian networks [18], Markov models [19], and support vector machines (SVMs) [20] have been used to understand and recognize human behavior. This method heavily depends on the extracted features and requires a lot of time and resources to process. Additionally, it does not fit well to diverse datasets and performs poorly in real-world cases [21].

Compared to traditional machine learning methods, deep learning is a learning process that automatically acquires representative features that suit a specific purpose by utilizing multiple hidden layers [22]. With the recent increase in computing power and available data, deep learning has been recently applied to human activity recognition and abnormal human activity detection and has shown good performance in video surveillance systems [23].

Among deep learning techniques, supervised learning is used by LSTM (long short-term memory), GRU (gated recurrent unit) and CNNs (convolutional neural networks) to spatially and temporally analyze abnormal human behaviors [24-28]. However, abnormal human behavior detection has the problem that it is difficult to detect abnormal human behaviors that were not previously defined in training data due to the lack of data [29] and the variety of human behaviors [30]. In addition, abnormal human behaviors may be considered abnormal in one situation but normal in another situation [31]. Therefore, models need to be trained effectively to detect abnormal human behaviors by considering the diversity behaviors and context dependencies of behaviors with limited labeled data. To this end, unsupervised learning and partially supervised learning methods are used to learn from unlabeled data by recognizing relationships and patterns. [32, 33]

Unsupervised learning trains models using unlabeled data that includes both normal and abnormal behaviors. Unsupervised learning includes reconstruction-based methods and generative-based methods. The reconstruction-based method analyzes only normal data and detects anomalies by checking for low reconstruction errors [34]. The generative-based method artificially generates images with learned distribution patterns, and the model effectively solves the problem of data insufficiency by distinguishing whether the images are real or fake [35, 36]. However, both reconstruction-based and generative-based approaches have difficulty identifying specific anomalous behaviors and are sensitive to changes in environment [37].

Partially supervised learning utilize both labeled and unlabeled data [38]. Partially supervised learning can be divided into semi-supervised learning and weakly supervised learning methods. Semi-supervised learning is a method that maximizes the use of non-labeled data by using labeled data as anchors [39, 40]. Weakly supervised learning is a method that prioritizes max model output with minimum reference labeled data [41, 42].

There have been several previous studies investigating abnormal human activity detection systems. Reference [43] focused on manual and deep approaches for various types of 2D and 3D data. Reference [44] detailed various techniques for single-scene video anomaly detection. Reference [45] reviews frame-based detection methods for edge computing-based abnormal behavior in video surveillance. Reference [32] investigates video surveillance systems for smart city applications, but the description of the methods used in unsupervised learning is insufficient. Reference [31] investigates vision-based human behavior recognition and describe widely used datasets. Reference [6] describes the problem of solving machine learning techniques for video surveillance systems and classify them into three categories: supervised learning, partially supervised learning, and unsupervised learning. However, their study did not focus on image-based detection. Additionally, previous research barely covers the studies published in 2023 and 2024. There is also a need for a discussion of the challenges and future applications of abnormal human behavior methods.

4. Deep learning-based abnormal behavior detection method

Deep learning can automatically learn and extract representative features of spatial and temporal information existing in image sequences. These deep learning techniques can be applied to human behavior recognition and abnormal human behavior detection to automate video surveillance systems. Deep learning techniques can be classified into supervised learning methods, unsupervised learning methods, and semi-supervised learning methods depending on the learning type and application field of the model that detects abnormal signs.

4.1 Supervised Learning

This method provides labeled data as input to the model and consists of a feature extraction part and a classification network. The deep learning network learns to obtain the desired output by repeatedly updating each parameter through the backpropagation method. The supervised learning trains the model to generate accurate detection output based on the input data, and therefore, most studies using the supervised learning use accuracy as the basic metric for evaluating model performance. Typically, the model uses CNN layers to acquire local features and integrates them with LSTM, GRU, etc. to learn the temporal relationship between features, by analyzing abnormal human behavior in space and time [46-48].

Some frameworks also use Kalman Filter, DA-CNN(dual attentional CNN), dual-stream CNN, and Bi-GRU. These studies utilize various datasets such as WED, UP-Fall, CMU, UR-Fall, MCF, Penn-Fudan, UT-Interaction, PEL, OpenImages, Hockey Fight, AVA, Ped1, Ped2, YouTube Action, UCF-50, UCF-101, VOC2007, Kinetics-600, and HMDB51. Recently, they have been used for specific works such as fall detection [49, 50] and detection of questionable activities in ATMs [51]. However, the performance of these methods depends on factors such as camera position, image quality, and presence of objects. In fall detection, detection performance deteriorates when actions such as crouching or sitting are included. It is also difficult to optimize hyper-parameters for optimal results. And while higher computational requirements increase the cost, edge devices make it difficult to detect actions, and the model can produce non-zero probabilities for specific task classes. Table 1 shows some of the different supervised learning model.

Table 1. Various Supervised learning models

Ref.	Year	Model	Datasets	Performance (AUC)
Ullah et al.[48]	2021	CNN LSTM	UCF-Crime, UMN, Avenue	0.982(UMN)
Usman et al. [52]	2022	CNN,RNN,KNN,Optical Flow	CHUCK, Avenue,UCSD, ShanghaiTech, UR fall	0.968(UCSD)
Chandrakala and Vignesh [24]	2022	Auto-encoder, C-LSTM	UCF-Crime, Crowd violence, Hockey Fight, BEHAVE	0.98(BEHAVE)
Michael Onyema et al. [27]	2023	CNN	Hockey Fight,UT-Interaction, PEL, WED, UCSD,, CMU,	0.996(CMU)
Hussain et al. [47]	2023	Dual-stream CNN	HMDB51, UCF-50, YouTube Action	0.982(UCF-50)
Ullah and Munir [28]	2023	DA-CNN Bi-GRU	HMDB51, Kinetics-600, YouTube Action, UCF	0.985(UCF-50)

4.2 Unsupervised learning

Unsupervised learning learn the inherent properties, patterns, and structures of data, such as distance and density, without any labels assigned to the data, to distinguish normal data from abnormal data. In the detection of abnormal human behavior, this method is particularly meaningful because obtaining labels for various types of abnormal behavior is sometimes difficult and inefficient [27], [53-55]. Two popular methods in this method are reconstruction-based and generative-based methods. Table 2 shows some of the unsupervised learning models.

Table 2. Various Unsupervised learning models

Ref.	Year	Model	Datasets	Performance (AUC)
Liu et al. [56]	2022	auto-endcoder	UCSD-Ped2, ShanghaiTech	0.968
Yan et al. [57]	2023	auto-endcoder	ShanghaiTech, UCF-Crime, NTU-RGB+D	0.765
Slavic et al. [58]	2022	variable auto-endcoder	Subway , Avenue	0.872
Wang et al. [59]	2022	variable auto-endcoder	USCD, Avenue	0.888
Taghinezhad and Yazdi [60]	2023	convolutional auto-encoder	UCSD, Avenue	0.976
Yu et al. [36]	2022	Generative-based	UCSD, Avenue, Subway, UCF-Crime	0.979
Huang et al. [61]	2023	Generative-based	UCSD, ShanghaiTech Avenue	0.921
Li et al. [62]	2023	Generative-based	UCSD-Ped2, ShanghaiTech, Avenue	0.968

Reconstruction-based methods train the model using only normal data and model the normal data distribution. Abnormal data is assigned a high reconstruction error by the model. If the test image is abnormal at the inference stage, the model has difficulty in reconstructing the image. Reconstruction-based detection methods include auto-encoder (AE), variational auto-encoder (VAE), and convolutional auto-encoder (CAE).

An auto-encoder is a neural network that learns input data and attempts to reconstruct a new image based on previously learned patterns, while AE consists of two structures: an encoder and a decoder. The goal of this model is to minimize the reconstruction error so that the model can reconstruct the image more accurately based on the learned data. Figure 2 shows the structure of AE

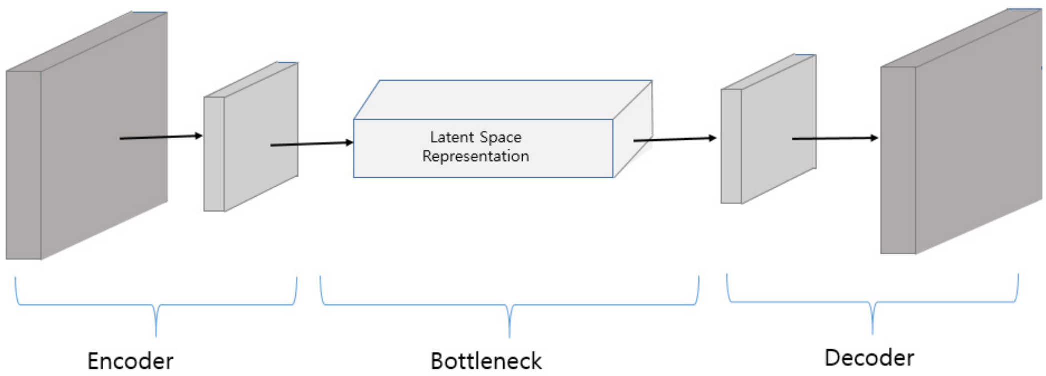

Figure 2. Auto-Encoder Structure

Liu et al. [56] proposed a memory-based connection network for video abnormality detection. Autoencoder learn specific features from data and reconstruct them. To learn which images are abnormal, memory module scoring is fed into a number of loss functions. In this model, the AUC value for detecting abnormal behaviors with a test dataset is 0.968. However, the scoring threshold should be well tuned for each environment to distinguish the behavior categories well.

On the other hand, Yan et al. [57] performed a memory clustering auto-encoder method to detect human behavior abnormalities. The auto-encoder is used to reconstruct the input sequence.

The clustering and scoring system aims to distinguish abnormal human behaviors in videos. The AUC metric is well achieved with a value of 0.765. However, there is a problem that the auto-encoder training conflicts with the clustering training.

Slavic et al. [58] proposed a self-aware module for detecting abnormal behavior in videos. VAE is introduced to reduce the dimension of video frames in the latent space. The AUC is 0.872, which is suitable for detecting abnormal behaviors.

Wang et al. [59] used a dual-stream VAE to generate a probability score to distinguish abnormal behaviors. However, the AUC score was not as high as when two separate VAEs were used.

Taghinezhad and Yazdi [60] proposed a novel multi-path network and multi-scale method for abnormal human behavior detection based on frame prediction. This study achieved the value of more than 0.976 at AUC with the Ped2 dataset. However, further studies on visual similarity, occlusion, and noise are needed to achieve improvement in the sophisticated abnormal score.

Generative-based detection is a method that attempts to generate artificial images that the model has learned. Artificial images are generated from the learned distribution patterns, and the similarity between the artificial images and the original images is evaluated [63]. The difference between the original image and the fake image is used to detect whether there is an abnormal human behavior in the captured frame. This method falls under the category of unsupervised learning, as no labels are generated. To apply the generative-based detection method, a generative adversarial network [64] is used, the representation of which is shown in Figure 3.

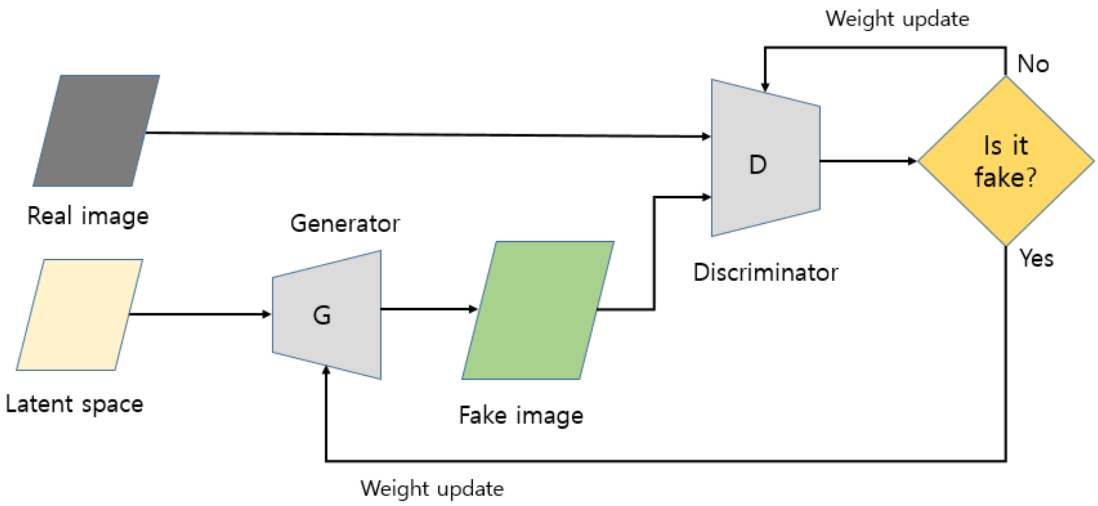

Figure 3. Generative Adversarial Network Structure

Yu et al. [36] proposed adversarial event prediction (AEP) to detect event pattern in AHB inputs. The AEP method combines reconstruction-based and generative-based detection methods. Since reconstruction-based methods have difficulty in handling diverse training data, adversarial networks provide predictions for various environmental settings in the dataset. The method uses GANs for predicting past and future frames of the video to perform detection completely according to the event sequence. The AUC value of AHB detection reaches 0.979, which is sufficient to classify unseen AHB frames during the training step. However, AEP does not utilize a preprocessing step for background detection, so its performance indicators are slightly lower in various scenarios. In addition, the dataset without events and noise also affects the performance score. Recently, Huang et al. [61] and Li et al. [62] conducted studies on AHB detection using GAN, and achieved AUC scores higher than 0.968 using the Ped2 dataset. However, these studies require a large number of parameters to be calculated. This method has a decreasing detection speed as the number of input frames increases and is difficult to set a skip interval for motion in the foreground in video anomaly detection.

4.3 Partially-Supervised learning

Partially supervised learning can use both labeled and unlabeled data. Partially supervised learning use what is learned from labeled data as a reference for the learning and prediction steps with unlabeled data [38]. There are two main types for partially supervised learning: semi-supervised and weakly supervised. Table 3 shows some of the partially supervised learning.

Table 3. Partially supervised learning types

Ref.	Year	Algorithm	Datasets	Performance (AUC)
Avital et al. [65]	2019	semi-supervised learning	CIFAR-10, SVHN	N/A
Sikdar and Chowdhury [66]	2020	semi-supervised learning	Ped1, Ped2, CUHK, UMN, ShanghaiTech	0.921
Wu et al. [67]	2021	semi-supervised learning	Ped1, CUHK	0.989
Wang et al. [68]	2023	wearkly supervised learning	CF-Crime, ShanghaiTech,	0.94
Zhang and Xue [69]	2023	wearkly supervised learning	UCF-Crime, Ped2	0.941
Shin et al. [70]	2024	wearkly supervised learning	UCF-Crime, ShanghaiTech, XD-violence	0.951

Semi-supervised learning learns the basic features of data using both labeled and unlabeled data [71]. The model is first trained with labeled data to learn the basic patterns. Then, unlabeled data is used as input to the trained model to obtain prediction values, and virtual labels are created using the prediction values. The model is trained again with a combination of labeled data and the created virtual label data to improve generalization performance. In this way, semi-supervised learning methods are used to augment label data when label data is limited [65]. Sikdar and Chowdhury [66] used a method without adaptive learning for anomaly detection. This model detects anomalous behaviors without pre-training and dynamically adjusts some parameters of the model during execution. This method achieved a performance of 0.992 in AUC using the UMN dataset. However, there is a slight performance lag due to the nature of the sparse dataset and the insufficient construction of local descriptors.Wu et al. [67] further improved the performance by using a semi-supervised relearning method. This method selectively extracts data from the original test set to create a new training data set. As a result, the AUC score on the UCSD Ped1 data set increased from 0.858 to 0.885. However, the performance of the model is affected by the performance of the baseline deep model, and sometimes the performance is often poor. Unlike the semi-supervised method, the weakly supervised learning method aims to improve the prediction results with a small amount of labeled data [65]. The weakly supervised learning model uses labels that are rougher or noisier than the label data used in the fully supervised learning model. Weakly supervised learning is used for multi-object learning [72], where groups of objects are labeled rather than as individual objects. It is also used in techniques such as expectation maximization algorithms to estimate the most likely label and optimize model parameters [73]. Regularization and loss function adjustments are used to help the model deal with noise in the labels. Recently, methods for detecting AHB using weakly supervised learning have utilized both video data [68] and image data [69]. Recent studies have achieved AUC scores of over 0.940 using the ShanghaiTech dataset. These studies used temporal features to better distinguish abnormal human behaviors. However, weakly supervised learning methods in videos have difficulties such as increased computational resources, limited interpretation of results, poor representation of local differences in long videos, and limited detection of abnormal behaviors in low-resolution videos. In addition, weakly supervised learning methods in image data take a long time to learn and have inaccurate pseudo-label generation. To improve these problems, Ullah et al. [74] proposed a dual-stream CNN architecture to detect abnormal behaviors in both surveillance and non-surveillance environments. Their performance was tested on the Violent Flow and Hockey Fight datasets, and they showed performance close to 0.990 in AUC. However, in the case of very complex video sequences, the model performance deteriorates due to misprediction.

4.4 Summary of deep learning methods for abnormal human behavior recognition

This section briefly summarizes each deep learning technique and explains its advantages and disadvantages.

Using fully supervised learning to train a model to detect abnormal human behavior is suitable when the human behavior category to be detected is determined. Detection method of abnormal human behavior using CNN is useful for detecting them at the frame level, but it is less effective at detecting long-term abnormal behavior that requires a period of time to determine whether it is abnormal. To improve this problem, some studies integrate LSTM and GRU to evaluate human behavior temporally. Supervised learning lacks data on abnormal human behavior and consumes consideravble computing resources and time due to the diverse characteristics of abnormal human behavior.

In unsupervised learning, auto-encoder methods are generally used for image dimension reduction in reconstruction-based methods. This training process is then used to compute a loss function within the model, which is used to detect abnormal human behavior within the input data. This distribution of data makes it difficult to implement the model in a heterogeneous environment, and the performance of models using auto-encoder methods varies depending on the data quality.

VAE is a method designed to solve the generalization problem that occurs in the data distribution of AE methods. In VAE, small object sizes and slightly interrupted pixel locations may occur due to the probabilistic calculations across the entire image. VAE methods can be combined with methods such as regularization and probabilistic formulation to achieve better performance in detecting abnormal behaviors in heterogeneous environments. In a reconstruction-based approach, a convolutional autoencoder (CAE) can be used to maximize the detection of each pixel in the image. CAE can distribute weights across all regions of the input image and provide generalization functions. With these functions, the model can identify abnormal human behaviors in a certain local region.

Generative-based methods show good performance in recognizing environments that are not seen in the training step. These advantages can be utilized to improve the generation module to reduce image noise, improve the quality of training images, and reduce the number of detection targets by distinguishing foreground and background objects. However, it is necessary to consider whether to focus on the adaptability of the model to objects or on reducing computation time and resources. This priority affects the balance between training and the generator and discriminator. By understanding the factors that the model considers important, training can be performed efficiently, reducing various shortcomings in identifying abnormal human behavior.

In the partially supervised learning, the semi-supervised learning uses a small amount of labeled data to use the unlabeled data as a new reference for subsequent model training. This method can be applied to various applications, especially in applying the model to the latest data that reflects the real environment. This algorithm can greatly reduce the labeling time, but if the training data used for labeling is noisy, the subsequent processing may become a meaningless task.

Weakly supervised learning focus on training models to generate higher quality information. The key is to utilize a small amount of data while generating accurate models. The model learns patterns from sparsely labeled data sets, and this leads to generalization of layers and weights. This algorithm has various advantages, such as faster detection of video streams and detection of abnormal human behaviors. This method does not require a large amount of label data and improves the generalization ability of the model in various scenarios of abnormal behaviors.

5. Future research directions

The method of recognizing abnormal behavior in camera-based surveillance systems has improved its performance with the introduction of deep learning techniques. However, each method has its own set of problems to solve.

Supervised learning require high computational resources. Therefore, it is necessary to make the model lightweight. In addition, CNN alone has limitations in detecting long-term abnormal human behavior using supervised learning, so further research is needed to improve the detection performance of abnormal behavior in the supervised learning domain by combining LSTM and GRU.

Among unsupervised learning, dealing with the variability of data distribution is an important task for reconstruction-based methods because it changes depending on the detection environment of the target object. In addition, accurately detecting AHB within a specific time frame is problematic because it is difficult to detect within a specific time frame. Therefore, it is necessary to effectively prioritize temporal AHB detection. VAE methods face the problem of pixel localization accuracy, so an alternative mechanism is needed to improve the accuracy of AHB detection.

It is very important for generative-based methods to solve problems such as gradient explosion. This problem causes rapid transitions in weights during calculation, and especially interferes with the learning process because the performance of model is highly dependent on training data. In addition, the generative model has a complex structure because the generator and discriminator exist separately. Therefore, research is needed on methods to reduce complexity or on the integration of the generator and discriminator.

The partial learning may have problems with the accuracy of labels generated automatically or manually. This is because it can be used to improve the detection performance of abnormal behavior or as a reference for defining labels for new data using limited label data. Therefore, research on how to preprocess anchor data with noise is important.

6. Conclusion

With the development of society, the need for security in public and private places is also increasing. Accordingly, video surveillance systems are being introduced everywhere to detect and predict abnormal human behavior to ensure safety. Detecting abnormal human behavior in video surveillance systems is an important task, but data sets showing such behavior are insufficient and expensive to collect. In addition, automated systems without human intervention are needed. To solve these problems, we investigated and analyzed deep learning techniques for detecting abnormal human behavior in video surveillance systems. In this paper, we first introduced popular benchmarking data sets used for training of deep learning model. Then, we classified training types of deep learning models for detecting abnormal human behaviors into supervised learning, unsupervised learning, and partially supervised learning, and briefly explained the concept of each method and presented the research results. We also described the advantages and disadvantages of each method. Based on this, we discussed several open research issues for AHB detection. It is expected that the presentation of these research issues will further develop video surveillance systems in the future.

Acknowledgments: This work was supported by Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT(MSIT, Korea)&Gwangju Metropolitan City and Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-2024-RS-2022-00156334)

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., vol. 41, pp. 1–58, 2009, doi: https://doi.org/10.1145/1541880.1541882.

[2] R. Ito, M. Tsukada, M. Kondo, and H. Matsutani, "An Adaptive Abnormal Behavior Detection using Online Sequential Learning," In Proc. IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA, pp. 436-440, 2019, doi: https://doi.org/10.1109/CSE/EUC.2019.00087.

[3] D. Kim, H. Kim, Y. Mok. And J. Paik, "Real-Time Surveillance System for Analyzing Abnormal Behavior of Pedestrians," Appl. Sci., vol. 11, no. 13, 2021, doi: http://dx.doi.org/10.3390/app11136153.

[4] A. Patwal, M. Diwakar, V. Tripathi, and P. Singh, "An investigation of videos for abnormal behavior detection," Procedia Comput. Sci., vol. 218, pp. 2264–2272, 2023, doi: https://doi.org/10.1016/j.procs.2023.01.202.

[5] P. Pandiaraja, R. Saarumathi, M. Parashakthi, and R. Logapriya, "An Analysis of Abnormal Event Detection and Person Identification from Surveillance Cameras using Motion Vectors with Deep Learning," In Proc. 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, pp. 1225-1232, 2023, doi: https://doi.org/10.1109/ICEARS56392.2023.10085466.

[6] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, "A Comprehensive Survey of Machine Learning Methods for Surveillance Videos Anomaly Detection," IEEE Access, vol. 11, pp. 114680–114713, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3321800.

[7] Y. Wang, C. Qin, Y. Bai, Y. Xu, X. Ma, and Y. Fu, "Making Reconstruction-based Method Great Again for Video Anomaly Detection," In Proc. IEEE International Conference on Data Mining (ICDM), Orlando, FL, USA, pp. 1215–1220, 2022, doi: https://doi.org/10.1109/ICDM54844.2022.00157.

[8] T. Ganokratanaa, S. Aramvith, and N. Sebe, "Anomaly Event Detection Using Generative Adversarial Network for Surveillance Videos," In Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, pp. 1395–1399, 2019, doi: https://doi.org/10.1109/APSIPAASC47483.2019.9023261.

[9] C. Zhang, G. Li, Q. Xu, X. Zhang, L. Su, and Q. Huang, "Weakly Supervised Anomaly Detection in Videos Considering the Openness of Events," IEEE Trans. Intell. Transp. Syst., vol. 23, pp. 21687–21699, 2022, doi: https://doi.org/10.1109/TITS.2022.3174088.

[10] J. Ren, F. Xia, Y. Liu, and I. Lee, "Deep Video Anomaly Detection: Opportunities and Challenges," In Proc. International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, pp. 959–966, USA,

2021, doi: https://doi.org/10.1109/ICDMW53433.2021.00125.
[11] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly detection in crowded scenes,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 1975-1981, 2010, doi: https://doi.org/10.1109/CVPR.2010.5539872.
[12] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked rnn framework,” in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 341-349, 2017, doi: https://doi.org/10.1109/ICCV.2017.45.
[13] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 6479-6488, 2018, doi: https://doi.org/10.1109/CVPR.2018.00678.
[14] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942, 2009, doi: https://doi.org/10.1109/CVPR.2009.5206641.
[15] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, “Robust real-time unusual event detection using multiple fixed-location monitors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 555–560, 2008, doi: https://doi.org/10.1109/TPAMI.2007.70825.
[16] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in matlab,” in 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, pp. 2720-2727, 2013, doi: https://doi.org/10.1109/ICCV.2013.338.
[17] C. Hu, Y. Chen, L. Hu, and X. Peng, “A novel random forests based class incremental learning method for activity recognition,” Pattern Recognit, vol. 78, pp. 277-290, 2018, doi: https://doi.org/10.1016/j.patcog.2018.01.025.
[18] Q. Xiao, and R. Song, R. “Action recognition based on hierarchical dynamic Bayesian network,” Multimed. Tools Appl., vol. 77, pp. 6955–6968, 2018, doi: https://doi.org/10.1007/s11042-017-4614-0.
[19] P. Sok, T. Xiao, Y. Azeze, A. Jayaraman, and M. V. Albert, “Activity recognition for incomplete spinal cord injury subjects using hidden Markov models,” IEEE Sens. J., vol. 18, pp. 6369–6374, 2018, doi: https://doi.org/10.1109/JSEN.2018.2845749.
[20] B. M. Abidine, L. Fergani, B. Fergani, and M. Oussalah, “The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition,” Pattern Anal. Appl., vol. 21, pp. 119–138, 2018, doi: https://doi.org/10.1007/s10044-016-0570-y.
[21] X. Hu, S. Hu, Y. Huang, H. Zhang, and H. Wu, “Video anomaly detection using deep incremental slow feature analysis network,” IET Comput. Vis., vol. 10, pp. 258–267, 2016, doi: https://doi.org/10.1049/iet-cvi.2015.0271.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, pp. 84-90, 2012, doi: https://doi.org/10.1145/3065386.
[23] O. K. Oyedotun and K. Adnan “Deep learning in vision-based static hand gesture recognition,” Neural Comput. And App., vol. 28, pp. 3941–3951, 2017, doi: https://doi.org/10.1007/s00521-016-2294-8.
[24] S. Chandrakala and L. K. P. Vignesh, “V2AnomalyVec: Deep Discriminative Embeddings for Detecting Anomalous Activities in Surveillance Videos,” IEEE Transactions on Computational Social Systems, vol. 9, no. 5, pp. 1307-1316, Oct. 2022, doi: https://doi.org/10.1109/TCSS.2021.3119957.
[25] M. Q. Gandapur, “E2E-VSDL: End-to-end video surveillance-based deep learning model to detect and prevent criminal activities,” Image Vis. Comput., vol. 123, 2022, doi: https://doi.org/10.1016/j.imavis.2022.104467.
[26] J. Ahn, J. Park, S. S. Lee, S.S.; K. H. Lee, H. Do, and J. Ko, “SafeFac: Video-based smart safety monitoring for preventing industrial work accidents,” Expert. Syst. Appl., vol. 215, 2023, doi: https://doi.org/10.1016/j.eswa.2022.119397.
[27] E. M. Onyema, S. Balasubaramanian, S. K. Suguna, C. Iwendi, B.V.V.S. Prasad, and C. D. Edeh, “Remote monitoring system using slow-fast deep convolution neural network model for identifying anti-social activities in surveillance applications,” Meas. Sens., vol. 27, 2023, doi: https://doi.org/10.1016/j.measen.2023.100718.
[28] H. Ullah and A. Munir, “Human Activity Recognition Using Cascaded Dual Attention CNN and Bi-Directional GRU Framework,” J. Imaging, vol. 9, 2023, doi: https://doi.org/10.3390/jimaging9070130.
[29] H. B. Rolf, N. M. Robertson, and D. M. Lane, “Human behaviour recognition in data-scarce domains,” Pattern Recognit. vol. 48, pp. 2377–2393, doi: https://doi.org/10.1016/j.patcog.2015.02.019.
[30] H. Tu, J. Allanach, S. Singh, K. R. Pattipati and P. Willett, “Information integration via hierarchical and hybrid bayesian networks,” IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 36, no. 1, pp. 19-33, Jan. 2006, doi: https://doi.org/10.1109/TSMCA.2005.859180.
[31] H. T. Duong, V. T. Le, and V. T. Hoang, “Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey,” Sensors, vol. 11, 2023, doi: https://doi.org/10.3390/s23115024.
[32] Y. Myagmar-Ochir and W. S. Kim, “A Survey of Video Surveillance Systems in Smart City,” Electronics, vol. 12, doi: https://doi.org/10.3390/electronics12173567.
[33] E. Şengönül, R. Samet, Q. Abu Al-Haija, A. Alqahtani, B. Alturki, and A. A. Alsulami, “An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey,” Appl. Sci., vol. 13, no. 8, 2023, doi: https://doi.org/10.3390/app13084956.

[34] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection," Image Vis. Comput., vol. 106, 2021, doi; https://doi.org/10.1016/j.imavis.2020.104078.

[35] D. Chen, P. Wang, L. Yue, Y. Zhang, Y., T. Jia, "Anomaly detection in surveillance video based on bidirectional prediction," Image Vis. Comput., vol. 98, 2020, doi: https://doi.org/10.1016/j.imavis.2020.104078.

[36] J. Yu, Y. Lee, K. C. Yow, M. Jeon, and W. Pedrycz, "Abnormal Event Detection and Localization via Adversarial Event Prediction," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 8, pp. 3572-3586, Aug. 2022, doi: https://doi.org/10.1109/TNNLS.2021.3053563.

[37] E. Sengönül, R. Samet, Q. A. Al-Haija, A. Alqahtani, B. Alturki, and A. A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey," Appl. Sci., vol. 13, 2023, doi: https://doi.org/10.3390/app13084956.

[38] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," IEEE Trans. Syst. Man Cybern. Part B (Cybern.), vol. 27, no. 5, pp. 787–795, 1997, doi: https://doi.org/10.1109/3477.623232.

[39] P. Khaire and P. Kumar, "A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments," Forensic Sci. Int. Digit. Investig, vol. 40, 2022, doi: https://doi.org/10.1016/j.fsidi.2022.301346.

[40] A. Pramanik, S. Sarkar, and S. K. Pal, "Video surveillance-based fall detection system using object-level feature thresholding," Knowl. Based Syst., vol. 280, 2023, doi: https://doi.org/10.1016/j.knosys.2023.110992.

[41] Y. Liu, J. Liu, M. Zhao, S. Li, and L. Song, "Collaborative Normality Learning Framework for Weakly Supervised Video Anomaly Detection," IEEE Trans. Circuits Syst. II Express Briefs, vol. 69, pp. 2508–2512, 2022, doi: https://doi.org/10.1109/TCSII.2022.3161061.

[42] A. M. Kamoona, A. K. Gostar, and A. Bab-Hadiashar, "Hoseinnezhad, R. Multiple instance-based video anomaly detection using deep temporal encoding–decoding," Expert. Syst. Appl., vol. 214, 2023, doi: https://doi.org/10.1016/j.eswa.2022.119079.

[43] C. Dhiman, and D. K. Vishwakarma, "A review of state-of-the-art techniques for abnormal human activity recognition," Eng. Appl. Artif. Intell., vol. 77, pp. 21–45, 2019, doi: https://doi.org/10.1016/j.engappai.2018.08.014.

[44] B. Ramachandra, M. J. Jones, and R. R. Vatsavai, "A Survey of Single-Scene Video Anomaly Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 5, pp. 2293-2312, 2022, doi: https://doi.org/10.1109/TPAMI.2020.3040591.

[45] D. R. Patrikar and M. R. Parate, "Anomaly detection using edge computing in video surveillance system: Review," Int. J. Multimed. Inf. Retr., vol. 11, pp. 85–110, 2022, doi: https://doi.org/10.1007/s13735-022-00227-8.

[46] J. Mao, P. Zhou, X. Wang, H. Yao, L. Liang, Y. Zhao, J. Zhang, D. Ban, and H. Zheng, "A health monitoring system based on flexible triboelectric sensors for intelligence medical internet of things and its applications in virtual reality," Nano Energy, vol. 118, 2023, doi: https://doi.org/10.1016/j.nanoen.2023.108984.

[47] A. Hussain, S. U. Khan, N. Khan, I Rida, M Alharbi, and S. W. Baik, "Low-light aware framework for human activity recognition via optimized dual stream parallel network," Alex. Eng. J., vol. 74, pp. 569–583, 2023, doi: https://doi.org/10.1016/j.aej.2023.05.050.

[48] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An Efficient Anomaly Recognition Framework Using an Attention Residual Lstm in Surveillance Videos," Sensors, vol. 21, no. 8, 2021, doi: https://doi.org/10.3390/s21082811.

[49] R. Espinosa, H. Ponce, S. Gutiérrez, L. Martínez-Villaseñor, J. Brieva, and E. Moya-Albor, "A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset," Comput. Biol. Med., vol. 115, 2019, doi: https://doi.org/10.1016/j.compbiomed.2019.103520.

[50] M. E. N. Gomes, D. Macêdo, C. Zanchettin, P. S. G. de-Mattos-Neto, and A. Oliveira, "Multi-human Fall Detection and Localization in Videos," Comput. Vis. Image Underst., vol. 220, 2022, doi: https://doi.org/10.1016/j.cviu.2022.103442.

[51] A. P. Kshirsagar and H. Azath, "YOLOv3-based human detection and heuristically modified-LSTM for abnormal human activities detection in ATM machine," J. Vis. Commun. Image Represent., vol. 95, 2023, doi: https://doi.org/10.1016/j.jvcir.2023.103901.

[52] I. Usman and A. A. Albesher, "Abnormal Crowd Behavior Detection Using Heuristic Search and Motion Awareness," Int. J. Comput. Sci. Netw. Secur., vol. 21 pp. 131–139, 2021, doi: https://doi.org/10.1007/s00138-011-0341-0.

[53] L. P. Cinelli, M. A. Marins, E. A. B. da Silva, and S. L. Netto, Variational Methods for Machine Learning with Applications to Deep Networks, Springer International Publishing: Cham, Switzerland, 2021.

[54] S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, "Review on Deep Learning Approaches for Anomaly Event Detection in Video Surveillance," Electronics, vol. 12, no. 1, 2022, doi: https://doi.org/10.3390/electronics12010029.

[55] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions," J. Big Data, vol. 8, no, 1, 2021, doi: https://doi.org/10.1186/s40537-021-00444-8.

[56] Q. Liu and X. Zhou, "A fully connected network based on memory for video anomaly detection," in 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS), pp. 221–226, 2022, doi:

https://doi.org/10.1109/CCIS57298.2022.10016377.
[57] M. Yan, Y. Xiong, and J. She, “Memory clustering auto-encoder method for human action anomaly detection on surveillance camera video,” IEEE Sensors Journal, vol. 23, no. 18, pp. 20715–20728, 2023, doi: https://doi.org/10.1109/JSEN.2023.3239219.
[58] G. Slavic, M. Baydoun, D. Campo, L. Marcenaro, and C. Regazzoni, “Multilevel anomaly detection through variational auto-encoders and bayesian models for self-aware embodied agents,” IEEE Transactions on Multimedia, vol. 24, pp. 1399–1414, 2022, doi: https://doi.org/10.1109/TMM.2021.3065232.
[59] L. Wang, H. Tan, F. Zhou, W. Zuo, and P. Sun, “Unsupervised anomaly video detection via a double-flow convlstm variational auto-encoder,” IEEE Access, vol. 10, pp. 44278–44289, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3165977.
[60] N. Taghinezhad and M. Yazdi, “A New Unsupervised Video Anomaly Detection Using Multi-Scale Feature Memorization and Multipath Temporal Information Prediction,” IEEE Access, vol. 11, pp. 9295-9310, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3237028.
[61] H. Huang, B. Zhao, F. Gao, P. Chen, J. Wang, and A. Hussain, “A Novel Unsupervised Video Anomaly Detection Framework Based on Optical Flow Reconstruction and Erased Frame Prediction,” Sensors, vol. 23, 2023, doi: https://doi.org/10.3390/s23104828.
[62] G. Li, P. He, H. Li, and F. Zhang, “Adversarial composite prediction of normal video dynamics for anomaly detection,” Comput. Vis. Image Underst., vol. 232, 2023, doi: https://doi.org/10.1016/j.cviu.2023.103686.
[63] M. Jovanović and M. Campbell, “Generative Artificial Intelligence: Trends and Prospects,” Computer, vol. 55, no. 10, pp. 107-112, Oct. 2022, doi: https://doi.org/10.1109/MC.2022.3192720.
[64] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in Neural Information Processing Systems, vol. 27, Curran Associates, Inc., 2014, doi: https://doi.org/10.1145/3422622.
[65] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, “Realistic evaluation of deep semi-supervised learning algorithms,“ in Proc. of the 32nd International Conference on Neural Information Processing Systems, Montreal, Curran Associates Inc.: Red Hook, NY, USA, pp. 3239–3250, 2018, doi: https://dl.acm.org/doi/10.5555/3327144.3327244.
[66] A. Sikdar, and A. S. Chowdhury, “An adaptive training-less framework for anomaly detection in crowd scenes,” Neurocomputing, vol. 415, pp. 317–331, 2020, doi: https://doi.org/10.1016/j.neucom.2020.07.058.
[67] R. Wu, S. Li, C. Chen, and A. Hao, “A. Improving video anomaly detection performance by mining useful data from unseen video frames,” Neurocomputing, vol. 462, pp. 523–533, 2021, doi: https://doi.org/10.1016/j.neucom.2021.05.112.
[68] L. Wang, X. Wang, F. Liu, M. Li, X. Hao, and N. Zhao, N. “Attention-guided MIL weakly supervised visual anomaly detection,” Measurement, vol. 209, 2023, doi: https://doi.org/10.1016/j.measurement.2023.112500.
[69] B. Zhang and J. Xue, “Weakly-supervised anomaly detection with a Sub-Max strategy,” Neurocomputing, vol. 560, 2023, doi: https://doi.org/10.1016/j.neucom.2023.126770.
[70] J. Shin, Y. Kaneko, A. S. M. Miah, N. Hassan and S. Nishimura, “Anomaly Detection in Weakly Supervised Videos Using Multistage Graphs and General Deep Learning Based Spatial-Temporal Feature Enhancement,” in IEEE Access, vol. 12, pp. 65213-65227, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3395329.
[71] N. Simmler, P. Sager, P. Andermatt, R. Chavarriaga, F. P. Schilling, and M. Rosenthal, “A Survey of Un-, Weakly-, and Semi-Supervised Learning Methods for Noisy, Missing and Partial Labels in Industrial Vision Applications,” in 2021 8th Swiss Conference on Data Science (SDS), Lucerne, Switzerland, pp. 26-31, 2021, doi: https://doi.org/10.1109/SDS51136.2021.00012.
[72] S. Park, H. Kim, M. Kim, D. Kim, and K. Sohn, “Normality Guided Multiple Instance Learning for Weakly Supervised Video Anomaly Detection,” in Proc. the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, pp. 2664-2673, 2023, doi: https://doi.org/10.1109/WACV56688.2023.00269.
[73] Z. Xu, X. Zeng, G. Ji, and B. Sheng, “Improved Anomaly Detection in Surveillance Videos with Multiple Probabilistic Models Inference,” Intell. Autom. Soft Comput., vol. 31, no. 3, pp. 1703–1717, 2022, doi: https://doi.org/10.32604/iasc.2022.016919.
[74] W. Ullah, T. Hussain, Z. A. Khan, U. Haroon, and S. W. Baik, “Intelligent dual stream CNN and echo state network for anomaly detection,” Knowl. Based Syst., vol. 253, 2022, doi: https://doi.org/10.1016/j.knosys.2022.109456.