Ingénierie des Systèmes d'Information
Vol. 27, No. 2, April, 2022, pp. 233-241
Journal homepage: http://iieta.org/journals/isi

Detection and Localization of Abnormal Events for Smart Surveillance

Baliram Sambhaji Gayal[1*], Sandip Raosaheb Patil[2]

[1] Dept. of Electronics & Telecommunication Engineering, JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune-33, India
[2] Dept. of Electronics & Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering for Women, Dhankawadi, Pune-43, India

Corresponding Author Email: bsgayal.sae@sinhgad.edu

https://doi.org/10.18280/isi.270207

Received: 8 February 2022
Accepted: 19 April 2022

Keywords:
vision only system, multi-model system, deep convolutional neural network (DCNN), bi-directional prediction model, GWO based deep CNN, SMO based deep CNN

ABSTRACT

In this study, the methods of anomaly detection are proposed. Background substitution (BG) is used for extracting the motion and indicating the attention region's locations, which are employed. Then the regions are fed into the "Deep Convolutional Neural Network (DCNN)". With the advantages of DCNN, for properly exploiting the spatiotemporal relationships, a network is developed for distinguishing anomalous and normal events. Besides this, the anomaly detection techniques are also described. The related databases are provided in this study. Many techniques for anomaly detection are discussed in this study with the help of the neural network. The different types of anomaly events are discussed here. All the data related to these anomaly events are discussed in the dataset. Different types of models related to the CNN model are also discussed in this study. And the anomaly techniques are also considered for discussion in this study.

1. INTRODUCTION

Foundations and Governments are paying more attention to forestalling crime and the threats of terrorism. In this context, advancements and ventures identification with anomaly identification is required, including image-change recognition, interloper discovery, and surveillance framework are continuously developing. Many surveillance frameworks are commercially incorporated into the framework of analog surveillance as sensor data light, for example, ultrasonic waves, and infrared beams. The framework of computerized surveillance is halfway overseen depending on the servers on CCTV (Closed-Circuit Television) and comparative sensors image data. These frameworks may have a mark-able result if there is a misinterpretation made. With the help of these ways, operational effectiveness and heartiness in several environmental conditions are needed. The system, which is important for learning the conceptual categories of the image schema, is categorized into 2 parts. One is the "vision only system" and the other one is the "multi-model system". The first system utilizes the visual priors for the construction of models for the object classes. And the second system is the combination of video and audio or text streams for matching the image categories with annotations or commentaries. The "vision-only system" constructs the models, which need the supervision of the inputs in the form of activity or object priors. And in the "multi-model system" the streams of co-occurrence languages are used for enhancing the space of the feature used for correlation and categorization is sometimes mediated by the attentive focus. "Video surveillance" is hugely used in different fields, like medical monitoring, security guards, traffic monitoring, etc. In these fields, the use of anomaly detection is huge, especially in discovering the different irregularities. Since some 1st hand datasets of videos are obtained without any abnormal event and labels are tough to find the definition prior. Besides this, the unsupervised models are the more existing and practical methods in the detection of the abnormality and it mainly focuses on the learning of the general patterns.

2. MOTIVATION

The method of anomaly detection can be divided into three parts, such as supervised, weakly supervised, and unsupervised. There are some approaches provided in this study for each of these methods. The unsupervised method can be completed with the approaches of "classic machine learning" and "deep learning". The datasets of anomaly detection are also discussed in this study. In the dataset, the related terms are UCSD, subway, avenue, UMN, CADP, DAD, A3D, DADA, DoTA, UCF Crime, Street scene, Iowa DOT, etc. And this dataset's average frames are also provided in the dataset table. While observing the safety & security of public violence [1]. The surveillance system plays an important role in this context. Anomalous incident identification like illegal activities, robberies, and traffic accidents, is important for video surveillance. Video surveillance has become widely accepted in recent years after knowing the great advantages of the Convolutional Neural Network (CNN). In this study, the "attention region localization" approach is discussed with the subtraction of the background. In this approach, the first step is giving attention to the regions. Once the desired region is found then it will feed into Deep Convolutional Neural Network (DCNN) action recognition. Also, a "hybrid approach" incorporates a bilateral filter and background subtraction for localizing the attention region to detect the efficient anomaly. Another approach is based on the behavior

for classifying the non-violent and violent videos. It is a suggestion from an author that tracking can be used as an anomaly for modeling normal motion. "Bilateral BG subtraction approach" is used for detecting visual attention.

The excess of the paper is arranged as follows: Section II exposes the various existing techniques for the detection of anomalies. Section III exposes the proposed anomaly detection model using the optimized classifier. Section IV exposes the proposed approaches and methods. Section V reveals the different datasets used for anomaly detections. Finally, section 6 concludes the paper.

The conventional techniques for the detection of anomaly-related benefits and drawbacks are detailed in the motivation section. In addition to that, various demands and challenges involved in anomaly detection are also taken into account.

2.1 Literature review

2.1.1 Bi-directional prediction

In this survey, this total paper different factors will be declared but, in this section, the bi-direction prediction will be discussed and the survey report on this factor will be discussed properly [2]. In this case, as shown in Figure 1, two or more picture is combined and the different candidate predictions are normally combined with that and the final prediction will be discussed according to this. This type of picture is normally known as the B picture. This is the compression technique of the picture. In this case, the codec uses different information from the decompressed type frame and this codec separates in two different directions. One is the back and another one is the Forward. For this reason, it is called a bidirectional prediction. After this prediction, the motion estimation normally increases. For the application of this prediction model, two different anchor pictures are necessary. In this research, a Bidirectional network is proposed for the application in deep learning. In this case, the total case study will be developed based on the incorporation of the optical flow of the images. In this case, any advanced autoencoder is not used for the compression technique. This result will represent that the total prediction method is focused on the current frame whereas the autoencoder normally uses the residual compression.

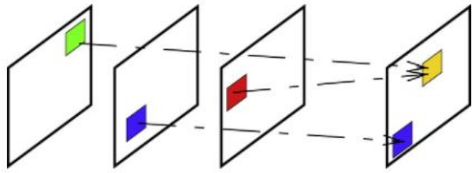

Figure 1. Bi-directional prediction model

2.1.2 GWO based deep-CNN

In this model, 3 constraints are employed, such as welded beam, compression or tension spring, and designs of pressure vessels [3]. These issues have different inequality and equality constraints that are why the GWO-model must be equipped with a method of handling the constraints for optimizing the problems about the constraints. Handling constraints is very challenging while the functions of fitness directly affect the agent's position (GSA). For the algorithm without fitness, any type of handling of constraints can be accepted without any modification in the mechanism of the algorithm (like PSO and GA) [4]. The GWO agents are updating their positions according to the delta, beta, and alpha locations. No relation is present between the function of fitness and search agents. According to this, the best function for controlling the constraints is the penalty function, where a huge number of values in the objective function is assigned by the search agents. If the positions of the agents violate the constraints, then they are replaced automatically by a new agent. Any type of penalty function is employed for penalizing the search agents according to the violation level. In this context, if the locations are found to be less efficient than a wolf, then it is replaced by a new agent automatically. And this is replaced in the next step [5]. A scaler and simple penalty functions for the remained problems, except the problem of compression or tension springs that utilizes complex-penalty functions. In following Figure 2, shown GWO based Deep CNN model for understanding concept in details.

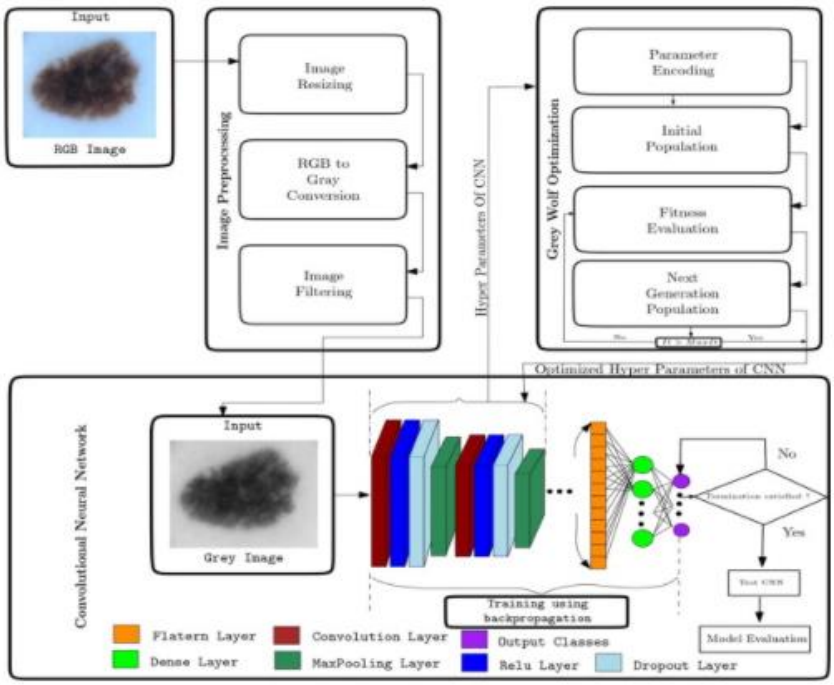

Figure 2. GWO-CNN model [6]

2.1.3 SMO based Deep-CNN

"Spider Monkey Optimization (SMO)" algorithm is a metaheuristic method and it is based on the social behavior of the spider monkeys. Spider monkeys normally live in the 40-50 members of the swarm [7]. The algorithm based on SMO can able to satisfy the following requisites, Labour division: And spider monkeys divide their work into different groups. Self-organization: Group sizes are selected for meeting the availability of food. The SMO adopts a collaborative process of trial-and-end that consists of 6 phases; "local leader phase, local leader learning phase, local leader decision phase, global leader phase, global leader, learning phase, and global leader decision phase" [8]. The SMO-CNN model is to feature engineering and select the best-attributed features, which positively affect the accuracy of the classification. The function of fitness and rate of convergence is utilized in the result of SMO. The main advantage of this model is that it helps in the quick convergence; it also helps in the efficient classification of the optimal parameters that make this technology an optimal choice in the reduction of dimensionality. The hyper-parameters of this technique are activation function, learning rate, etc [9]. These parameters are turned before the input processing. The learning rate is optimized with the help of the Adam optimizer. In Figure 3, shown SMO based Deep CNN model for understanding concept and general steps involved for implementation.

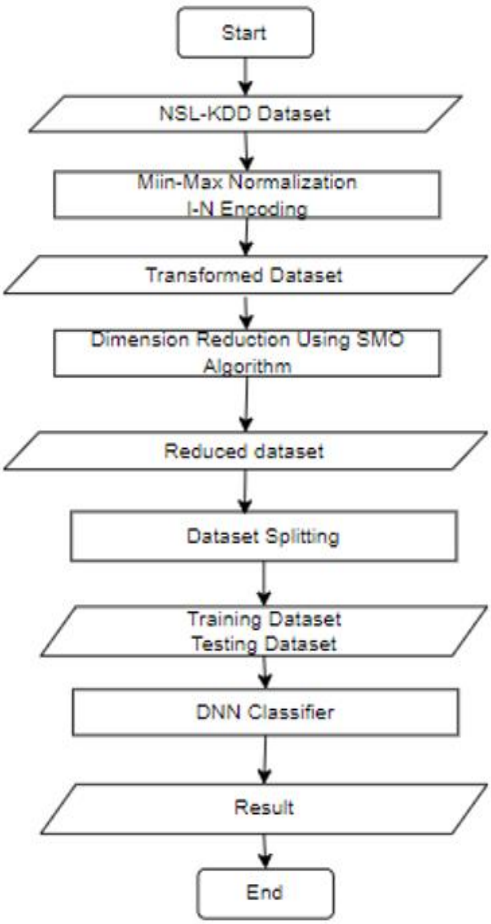

Figure 3. SMO-deep neural network

2.1.4 Conv-LSTM

The total data collection process can be characterized by the proper time series. For this reason, advanced techniques are normally applied to the LSTM or the 'Longand Short term type memory'. This is a neural network type architecture. In this case, the model normally passes the previous state (which is hidden) in the next step of the total sequences. The information is stored before and after that, this information is normally used for the other sequences. In this technique, the proper order of the data is the priority of this technique. In this case, the different images are normally passed from the different layers of convolution [10]. This will affect the creation of the 1D array. During the repetition time of this process in a given time set, the inputs normally act as the proper input layers of the system. This system can be applied to spatiotemporal type prediction. This system can declare the future state of the present cell. This can be calculated by the grid input and the different types of past states of the neighbor cell. Proper declaration of the padding can help to ensure the proper state of the different rows and the other columns of the cell. After this conventional operation can be applied. In Figure 4 shown Conv-LSTM model. Bangare et al. [11-16] and Joseph et al. [17] have sued machine learning for the research work. Shelke et al. [18] have worked in emotion analysis. Bangare et al. [19-21] have shown the object detection techniques.

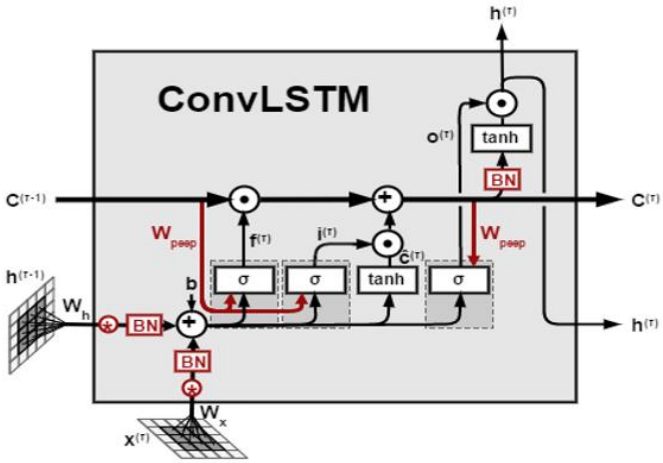

Figure 4. Conv-LSTM model [22]

2.1.5 Bi-directional LSTM

This model is different from the simple LSTM. The result or output of this model is depending on the upcoming frames from the sequence as well as the previous frame [23]. In Figure 5, shown Bidirectional LSTM model.The BD-RNN structure is simple with 2-stacked RNNs, one stack is in the direction of backward, and the other one is in the direction of forward. And both of these stacks' hidden states are combined with the output. In this study, the multi-layer BD-LSTM is used where every layer has 2 cells, one is for the pass backward and another one is for the pass which is in the forward direction. The features are extracted with the help of ResNet-50 and which is used for deciding whether the event is normal or anomaly, which is fed to Bidirectional-LSTM that has a multi-layer. And it is present in chunk forms for taking the decision of anomaly [24]. There are 1000 features available and the first one forms an input from at time t, and the next feature is chunked at the form of t+1. In the training stage of this model, the training data are passed. And the stage which is hidden is merged with the backward and forward passes [25]. The backpropagation is utilized for the adjustment of the weight and bias. The inner part of this model represents bw and fw. The result of this frame is measured according to the next and previous frames, until the layers process in both directions.

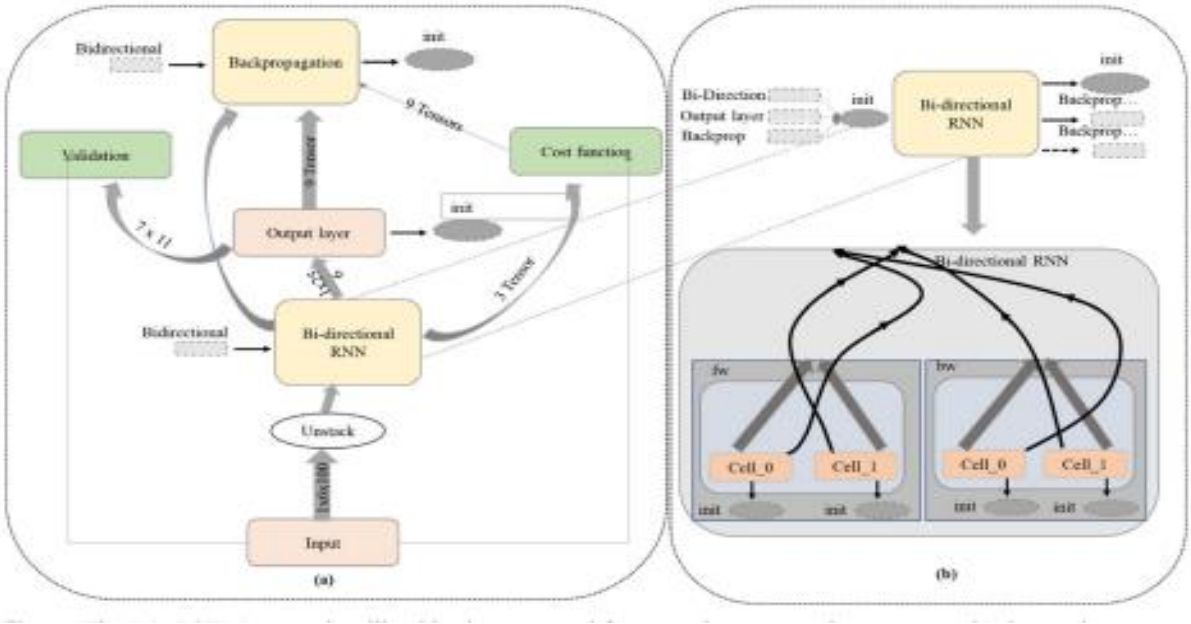

Figure 5. Bidirectional LSTM model [26]

3. PROPOSED ANOMALY DETECTION MODEL

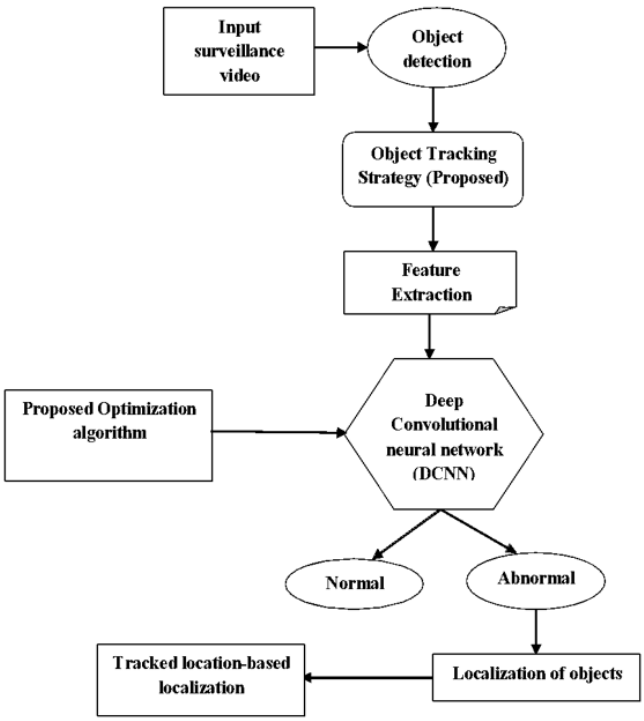

Figure 6. Proposed anomaly detection model

The main aim of the research is to develop an automatic anomaly detection strategy for the detection of anomalies in surveillance videos. In above Figure 6, shown proposed anomaly detection model using Deep CNN. Initially, the input surveillance video will be subjected to object detection using the threshold method. Then, the object tracking strategy will be developed with the integration of the features obtained from the tracking algorithms. Once the objects are tracked, the features will be extracted from the feature extraction step, where the significant features are extracted for the classification process. The extracted features will be statistical features extracted using GLCM descriptors, object speed, and shape-based features. The selected features will be fed as input to the Deep Convolutional Neural Network (DCNN), where the optimal weights are tuned using the proposed algorithm. Deep-CNN will classify the video as normal or abnormal. If any anomalies are detected in the surveillance video, the object localization will be done to track the location of the abnormal object in the video. Different databases can be used for the implementation such as the UCSD database, Avenue, ShanghaiTech Campus, Subway dataset, and the implemented model will be analyzed and compared with the existing methods in order to reveal the effectiveness of the developed method. The proposed technique will be evaluated on the different performance metrics, such as accuracy, sensitivity, and specificity, MOTP, EER, AUC and ROC.

4. APPROACHES AND METHODS FOR ANOMALY DETECTION

The methods of anomaly techniques are:

4.1 Unsupervised methods

Unless the practical anomaly events occur with low profits, capturing all types of anomalies is tough. The capturing of videos is easily accessible from public surveillance and social media. This method helps in the detection of anomaly events with "only normal videos in the training set" [27]. This method is unable to meet satisfactory performance in the complex scenarios of the real world.

4.1.1 Classical machine learning

Initially the unsupervised method has adopted the "classic machine learning" with the features of hand-craft. Firstly, the optical flow features are extracted, and find out the patterns are with the probabilistic PCA ("Principal Component Analysis") mixture. A "space-time MRF (Markov Random Field)" is constructed for modeling the relationship between

the local regions of the videos. The anomaly and normal frames are identified with "BoW (Bag of Words)", and "LDA (Latent Dirichlet Allocation)". The DT-based mixture model is introduced for temporal normalcy and a" discriminant saliency detection" is used to measure the spatial normality.

4.1.2 Deep learning

By the deep learning techniques, the recent works get the advantages of the resources of power consumption and large-scale dataset. Both the discriminative regular patterns and motion features with an "FCN (Fully Convolutional Network)" based autoencoder can be learned. The score of regularity is computed according to the AE model's error reconstruction. For making a better relationship with the video, "LSTM (Long Short-Term Memory)" and FCN are combined which can improve the AE framework's performance [28]. In all the AE-based methods the anomaly events are identified according to the error reconstruction.

From the strength of the framework of VS based on CNN, a deep-CNN framework is developed for the surveillance of video summarization. A VS framework based on the CNN model is proposed for outdoor and indoor IoT surveillance-network [29]. This framework generates high-quality videos effectively for the users so that they could use this as evidence from the "Mean of Scores (MOS)" and experimental results, which are assigned by the users. This framework recruits an extraction pipeline, which utilizes entropy, aesthetics, and memorability features to determine every frame's importance within a segmented shot. The memorability score of images is predicted from the model of CNN. And the entropy and aesthetic features are maintaining the diversity and interestingness of the final VS. A mechanism of hierarchical weightage fusion is designed for producing the aggregated scores of every frame from the segmented shot. An in-depth and comprehensive experimental evaluation is done for verifying the proposed solution [30]. As per the result of the experiments it can be seen that this solution is able to improve the effectiveness and quality rather than the state-of-the-art schemes, so this framework can be adopted by the IoT surveillance network.

4.1.3 Deep shot segmentation based on the feature

The coverage, interestingness, and diversity of the summary of a video are dependent on the shot segmentation. Because of this, its implementation of it is becoming challenging. The "shot-segmentation" scheme is important in different applications, like detection of anchors in the videos of news, VS, indexing videos, and the needed contents retrievals. As an example, an author focused on the retrieval and indexing of the videos, which are ecological, used shot segmentation. Also, in the literature, a sparse method of shot segmentation based on coding is presented, which is based on the summarized videos. The methods which are discussed belong to a particular specific domain, and this is not suitable for the streams of surveillance video for their limited performance of them. Then the challenges and issues of surveillance videos are considered, by exploring the segmentation of CNN into an appropriate shot. In this method, 2 frames are selected from 2 sets that are based on the extracted features from the deep-CNN model for deciding the shot, whether they are from the same shot or not. And its features of it are extracted from the Squeeze Net CNN model. This model actually a model of classification, but global features of discriminations is learned from this model, which is useful for the other domains also.

4.2 Weakly supervised methods

With the increased number of video data presented on the platforms of social media like YouTube, a huge amount of anomaly videos annotation and access is possible [31]. For some scenarios of certain applications, where the activities of the anomaly are defined well, their performance can be developed with the introduction of supervision information. The videos of training are labeled as anomalous or normal. The problems of weakly supervised are formulated as "multiple instance learning (MIL)". The frame of the normal video must be normal and the anomalous videos are framed anomalies [32]. A framework of C3D is utilized for the extraction of the features of Spatio-temporal and the generation of anomaly scores. For distinguishing the anomaly and normal frames with the help of this method, the score of the forces of loss function of a (-) ve video should be higher than the normal video's highest score.

4.3 Supervised methods

In the recent scenarios where the objects and backgrounds are defined well, such as the cars, and roads for detecting traffic accidents on the highways. The solution to this problem is the detection of the object and geometric prior knowledge from the public datasets.

Firstly, Faster-RCNN is applied for detection of the vehicles, and then an LSTM module is applied which is attention-based for learning the score of the accidents. Despite the object detection application, the semantic segmentation that uses space and background should be modeled and perspective detection leverage the geometric priors [33]. The dynamics of the vehicles are then represented by the matrix of spatial-temporal. According to the IOU ("Intersection Over Unit"), the events of the anomaly are identified. The starting time of the accident is estimated according to the algorithm of curve fitting.

5. DATASET

The conventional techniques for the detection of anomaly-related benefits and drawbacks are detailed in the motivation section. In addition to that, various demands and challenges involved in anomaly detection are also taken into account.

5.1 Avenue

This dataset consists of typical events like throwing an object and running are considered an anomaly in this dataset.

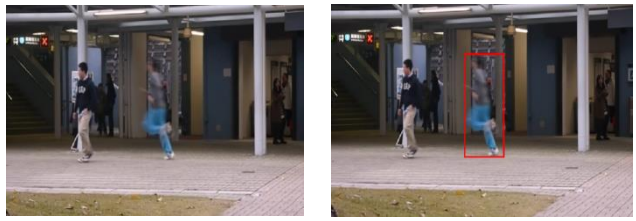

Figure 7. Avenue dataset original and anomaly detected frame

Avenue dataset contains 21 testing and 16 training video clips. These videos are captured in the avenue of the CUHK

campus with more than 30652 frames including 15324 testing and 15328 training frames [29, 34]. The training videos were captured in a normal situation. And the testing videos are included with both the normal and abnormal events. The events like new objects, running, suspicious actions, throwing, etc. all are considered anomaly events in this dataset. In above Figure 7, shown original frame and after processing anomaly detected frame from Avenue dataset.

5.2 UCSD

The dataset of UCSD captured data with the help of 2 poses of the camera on the campus of UCSD [35]. On this campus, most of the pedestrians walk. The set of training contains only the normal frame and in the test set, both the anomaly and normal frames are present. This dataset only consists of pedestrians so the other entities like carts and bikers are anomaly instances in this entity. This dataset contains 2 subsets, called ped 1 and ped 2. The common anomalies include the skaters, small carts, bikers, and the people who are walking on the roads, or in the grass around. Some examples are also given for the people who are in a wheelchair, which are also considered anomaly events. All the abnormal situations are naturally occurring, whereas they are not staged for the purposes of assembling the dataset. All these data are divided into two subsets, based on the different scenes. As well as the video footage are also divided into different clips of almost 200 frames.

Figure 8. UCSD Ped1 dataset original and anomaly detected frame

Peds 1 clips of groups of people walking towards and any away from the camera and some amount of perspective distortion [36]. Contains 34 training video samples and 36 testing video samples.

Peds 2 scenes with pedestrian movement parallel to the camera plane. Contains 16 training video samples and 12 testing video samples.

Figure 9. UCSD Ped2 dataset original and anomaly detected frame

This dataset is divided into two distinct scenes. The first one is UCSD-Peds1 and the second one is UCSD-Peds2. This dataset contains videos that capture a scene of a crowded pedestrian area. The area is restricted for vehicles, therefore, any vehicles, cyclists, or skaters found moving in the area should be considered an anomaly. Videos in the UCSD-Peds1 data set are captured from two different angles where the camera is fixed with no lighting variations. This data set contains 98 video sequences in total, where each video sequence consists of 200 video frames. UCSD-Peds1 comprises 70 video sequences. For the training phase, we have used 34 video sequences while 36 of them are used for testing purposes. Some of the normal and abnormality detected frames from the dataset UCSD-Peds1 are shown in Figure 8. UCSD-Peds2 contains 28 videos sequences. For training, we have used 16 of these sequences, while 12 of them are used for testing. The ground truth data for both scenes are provided to evaluate the model. Some of the normal and abnormality detected frames from datasets of UCSD Peds2 are shown in Figure 9.

5.3 ShanghaiTech campus

This dataset is collected from the University of Shanghai under the condition of camera viewpoint and complex light. In this dataset, the anomaly events are the cars and bikes. ShanghaiTech Campus dataset has 13 scenes with complex light conditions and camera angles. It contains 130 abnormal events and more than 270, 000 training frames, and 42883 testing frames. In total training and testing, frames are 317398 [37]. The pixel-level ground truth of abnormal events is also annotated in the dataset. The anomalies are the cars, bicycles, bikes, skaters, suspicious actions, running, etc. In Figure 10, shown original frame and after processing anomaly detected frame from ShanghaiTech campus dataset.

Figure 10. ShanghaiTech dataset original and anomaly detected frame

5.4 Subway

This dataset is divided into 2 parts, subway exit, and subway entrance. In the subway station, only one surveillance video is contained [38]. Moving in the wrong direction is an anomaly in this dataset.

Figure 11. Subway original and anomaly detected frame

In Figure 11 shown subway original and anomaly detected frames.

5.5 UCF

The UCF-Crime dataset is a large-scale, first-of-its-kind dataset containing 128 hours of video. It includes 1900 lengthy and uncut real-world surveillance videos with 13 actual anomalies such as Abuse, Arson, Assault, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, and Vandalism. These anomalies were chosen because they pose a serious threat to public safety. This data set can be utilized for two different purposes. First, all anomalies in one group are considered, whereas all typical behaviors are considered in another. Second, for identifying each of the 13 unusual activities. Iowa Department of Transportation, for example. The average frames for this dataset are also listed in the dataset (Table 1). In Figure 12, shown frames from UCF-Crime dataset which having robbery and burglary anomalous events.

Figure 12. UCF dataset robbery and burglary frames

5.6 UMN ("University of Minnesota")

5 videos are included in this dataset, which are from several sides. In this entity, running is considered an anomaly.

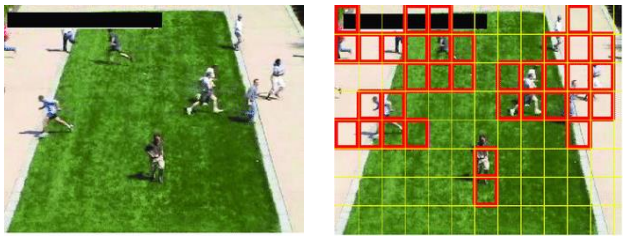

Figure 13. UMN Original and anomaly detected frame

In Figure 13 shown original and anomaly detected frames.

5.7 DAD ("Dashcam Accident Dataset")

This entity is especially for the detection of the accident. In this entity, the normal pattern is the movement of the vehicles, and the anomaly events are the traffic accidents, such as a collision between car and car, motorbike between a motorbike, etc.

In Figure 14 shown accidents as anomaly detected frames.

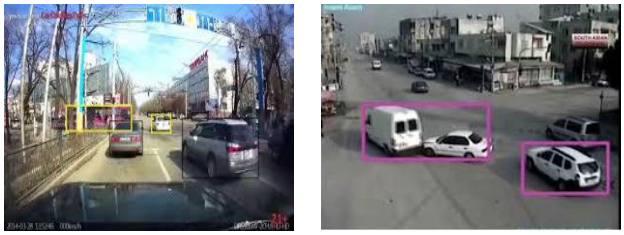

Figure 14. DAD accident detected frames

5.8 CADP ("Car Accident Detection and Prediction")

In this dataset, the main focus is on the accidents captured on the CCTVs. This entity includes the videos captured from different types of cameras, weather conditions, and qualities [39]. Here the anomalies are traffic accidents. In Figure 15 shown without accident frame means normal frame and with accident frame means anomaly detected frames.

Figure 15. CADP no accident frame and accident frame

Table 1. Various dataset summary

Dataset	Video's	Average frames	Example anomalies
Avenue	35	840	The new object, run, throw, suspicious actions, wrong directions
UCSD (Ped 1)	68	200	Small carts, bikers, Bicycle, skaters, wheelchair
UCSD (Ped 2)	28	160	Small carts, bikers, Bicycles, Skaters
Shanghai Tech	435	725	Cars, bikers, bicycles, Skaters, running, suspicious movement/actions
Subway entrance	1	121750	Wrong direction, No Payment
Subway exit	1	64900	Wrong direction, No Payment
UMN	5	1300	Run
A3D	1499	80	Traffic accidents
DoTA	4655	150	Collision, traffic anomalies
DADA	2000	321	Traffic accidents
Iowa DOT	199	27100	Traffic accidents
CADP	1415	365	Traffic accidents
DAD	1729	99	Traffic accidents
Street scenes	80	2500	Illegally car parking, jaywalking
UCF crime	1897	7250	Accident, fighting, burglary, abuse, arrest, arson, fighting, robbery

In Table 1, shown various dataset summary, which include dataset name, number of video's, average number of frames and examples of anomalies present.

This approach is based on texturing that can retain the observed area's edges and eliminate the noises [40]. This technique can build a proper BG model for clearly observing the moving object's area as a "visual attention region" and the part will be considered as the uninterested area. From the figure, the comparison can be visualized between the subtraction of bilateral BG with the 2 approaches, called "Improved model of Mixture of Gaussians (MOG_2)" and "K-Nearest Neighbours (KNN)". In these methods, the images are

processed pixel-by-pixel. The "bilateral texture-based approach" can remove the noise and can point out the desired region. The datasets are based on the anomalies of the avenue, UCSD Peds 1 and Peds 2, ShanghaiTech, subway, UCF, UMN, DAD, and CADP are collected in this study [41]. The different types of data based on different anomaly events are considered here. The anomaly event can be the new objects, running, suspicious actions, throwing, etc. And these events are captured by testing and training videos [42]. In the avenue dataset, the videos are included with 30652 frames consisting of 15324 testing and 15328 training frames. And in UCSD Peds 1 contains 34 training video samples and 36 testing video samples and in Peds 2 16 training video samples and 12 testing video samples are included. ShaighaiTech contains 130 abnormal events and more than 270, 000 training frames, and 42883 testing frames. In total training and testing, frames are 317398.

6. CONCLUSION

Surveillance systems play a vital part in ensuring the safety and security of the general population. Identification of unusual incidents such as illicit activity, robberies, and traffic accidents is critical for video monitoring. In the automatic detection and localization of anomalies, computer technology plays a critical role. In this paper, many strategies for detecting abnormalities are discussed. For sequential frameworks, the Deep-CNN framework is introduced. The bidirectional prediction model, GWO-based DCNN, SMO-based DCNN, Conv-LSTM, and bidirectional-LSTM model all follow this paradigm. This research also looks at several databases of abnormalities. UCSD, subway, avenue, UMN, CADP, DAD, A3D, DADA, DoTA, UCF Crime, Street scene, Iowa DOT, and others are among the connected datasets. The average frames for this dataset are also listed in the dataset table.

Despite the adoption of a variety of frameworks for learning, separating these learned representations for complicated anomalous behaviors remains unsatisfactory. In addition, there is a compromise between generalization and performance for greater application improvement. The current methods for detecting anomalies function as warning systems. Deep learning-based methods have proved very successful in the detection of abnormalities in computer vision, image processing, and video processing, and are now widely employed. Surveillance systems are improving every day, and their costs are decreasing.

REFERENCES

[1] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M., Baik, S.W. (2021). CNN features with bi-directional LSTM for real time anomaly detection in surveillance networks. Multimedia Tools and Applications, 80(11): 16979-16995. https://doi.org/10.1007/s11042-020-09406-3

[2] Shehzed, A., Jalal, A., Kim K. (2019). Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection. In 2019 International Conference on Applied and Engineering Mathematics (ICAEM), pp. 163-168. https://doi.org/10.1109/ICAEM.2019.8853756

[3] Mirjalili, S., Mirjalili, S.M., Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69: 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[4] Shao, Z., Cai, J., Wang, Z. (2017). Smart monitoring cameras driven intelligent processing to big surveillance video data. IEEE Transactionson Big Data, 4(1): 105-116. https://doi.org/10.1109/TBDATA.2017.2715815

[5] Biradar, K.M., Gupta, A., Mandal, M., Vipparthi, S.K. (2019). Challenges in time-stamp aware anomaly detection in traffic videos. ArXivpreprint arXiv:1906.04574.

[6] Source: https://ars.els-cdn.in/content/image/1-s2.0-S1319157821001270-gr2.jpg.

[7] Bansal, J.C., Sharma, H., Jadon, S.S., Cler, C.M. (2014). Spider monkey optimization algorithm for numerical optimization. Memeticcomputing, 6(1): 31-47. https://doi.org/10.1007/s12293-013-0128-0

[8] Khan, A., Ali Shah, J., Kadir, K., Al battah, W., Khan, F. (2020). Crowd monitoring and localization using deep convolutional neural network: A review. Applied Sciences, 10(14): 4781. https://doi.org/10.3390/app10144781

[9] Li, A., Miao, Z., Cen, Y., Zhang, X.P., Zhang, L.,Chen, S. (2020). Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. Pattern Recognition, 108: 107355. https://doi.org/10.1016/j.patcog.2020.107355

[10] Liu, G., Chen, H., Sun, X., Quan, N., Wan, L., Chen, R. (2017). Low-complexity nonlinear analysis of synchro phasor measurements for events detection and localization. IEEE Access, 6: 4982-4993. https://doi.org/10.1109/ACCESS. 2017.2772287

[11] Bangare, S.L., Patil, M., Bangare, P.S., Patil, S.T. (2015). Implementing tumor detection and area calculation in MRI image of human brain using image processing techniques. Int. Journal of Engineering Research and Applications, 5(4): 60-65.

[12] Bangare, S.L., Dubal, A., Bangare, P.S., Patil, S.T. (2015). Reviewing Otsu's method for image thresholding. International Journal of Applied Engineering Research, 10(9): 21777-21783. https://dx.doi.org/10.37622/IJAER/10.9.2015.21777-21783

[13] Bangare, S.L., Pradeepini, G., Patil, S.T. (2018). Regenerative pixel model and tumour locus algorithm development for brain tumour analysis: a new computational technique for precise medical imaging. International Journal of Biomedical Engineering and Technology, 27(1-2): 76-85. https://dx.doi.org/10.1504/IJBET.2018.093087

[14] Bangare, S.L., Pradeepini, G., Patil, S.T. (2017). Neuro endoscopy adapter module development for better brain tumor image visualization. International Journal of Electrical and Computer Engineering, 7(6): 3643. https://dx.doi.org/10.11591/ijece.v7i6.pp3643-3654

[15] Bangare S.L. (2022). Classification of optimal brain tissue using dynamic region growing and fuzzy min-max neuralnetwork in brain magnetic resonance images. Neuroscience Informatics, 2(3): 100019. https://doi.org/10.1016/j.neuri.2021.100019

[16] Bangare, S.L., Pradeepini, G., Patil, S.T. (2017). Brain tumor classification using mixed method approach. In 2017 International Conference on Information Communication and Embedded Systems (ICICES), pp. 1-4. https://doi.org/10.1109/ICICES.2017.8070748

[17] Joseph, L.L., Shrivastava, P., Kaushik, A., Bangare, S.L., Naveen, A., Raj, K.B., Gulati, K. (2021). Methods to identify facial detection in deep learning through the use of real-time training datasets management. EFFLATOUNIA-Multidisciplinary Journal, 5(2).

[18] Shelke, N., Chaudhury, S., Chakrabarti, S., Bangare, S.L., Yogapriya, G., Pandey, P. (2022). An efficient way of text-based emotion analysis from social media using LRA-DNN, Neuroscience Informatics, 100048. https://doi.org/10.1016/j.neuri.2022.100048

[19] Bangare, P.S., Uke, N.J., Bangare, S.L. (2012). Implementation of abandoned object detection in real time environment. International Journal of Computer Applications, 57(12).

[20] Bangare, P.S., Uke, N.J., Bangare, S.L. (2012). An approach for detecting abandoned object from real time video. International Journal of Engineering Research and Applications (IJERA), 2(3): 2646-2649.

[21] Bangare, P.S., Bangare, S.L., Yawle, R.U., Patil, S.T. (2017). Detection of human feature in abandoned object with modern security alert system using Android Application. In 2017 International Conference on Emerging Trends & Innovation in ICT(ICEI), pp. 139-144. https://doi.org/10.1109/ETIICT.2017.7977025

[22] https://miro.medium.in/max/471/1*u8neecA4w6b_F1NgnyPP0Q.png.

[23] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M., Baik, S.W. (2021). CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. MultimediaTools and Applications, 80(11): 16979-16995. https://doi.org/10.1007/s11042-020-09406-3

[24] Muhammad, K., Khan, S., Baik, S.W. (2020). Efficient convolutional neural networks for fire detection in surveillance applications. In Deep Learning in Computer Vision, 63-88, CRC Press.

[25] Fan, Y., Wen, G., Li, D., Qiu, S., Levine, M.D, Xiao, F. (2020). Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder. Computer Vision and Image Understanding, 195: 102920. https://doi.org/10.1016/j.cviu.2020.102920

[26] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M., Baik, S.W. (2021). CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications, 80(11): 16979-16995. https://doi.org/10.1007/s11042-020-09406-3

[27] Vu, H., Nguyen, T.D., Travers, A., Venkatesh, S., Phung, D. (2017). Energy-based localized anomaly detection in video surveillance. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 641-653. https://doi.org/10.1007/978-3-319-57454-7_50

[28] Sabokrou, M., Fathy, M., Moayed, Z., Klette, R. (2017). Fast and accurate detection and localization of abnormal behavior in crowded scenes. Machine Vision and Applications, 28(8): 965-985. https://doi.org/10.1007/s00138-017-0869-8

[29] Murugesan, M., Thilagamani, S. (2020). Efficient anomaly detection in surveillance videos based on multilayer perception recurrent neural network. Microprocessors and Microsystems, 79: 103303. https://doi.org/10.1016/j.micpro.2020.103303

[30] Deepak, K., Chandrakala, S., Mohan, C.K. (2021). Residual spatiotemporal auto encoder for unsupervised video anomaly detection. Signal, Image and Video Processing, 15(1): 215-222. https://doi.org/10.1007/s11760-020-01740-1

[31] Sabokrou, M., Pourreza, M., Fayyaz, M., Entezari, R., Fathy, M., Gall, J., Adeli, E. (2018). Avid: Adversarial visual irregularity detection. In Asian Conference on Computer Vision, 488-505.

[32] Santhosh, K.K., Dogra, D.P., Roy, P.P. (2020). Anomaly detection in road traffic using visual surveillance: A survey. ACM Computing Surveys (CSUR), 53(6): 1-26. https://doi.org/10.1145/3417989

[33] Rojek, I., Studzinski, J. (2019). Detection and localization of water leaks in water nets supported by an ICT system with artificial intelligence methods as a way forward for smart cities. Sustainability, 11(2): 518. https://doi.org/10.3390/su11020518

[34] Dhiman, C., Vishwakarma, D.K. (2019). A review of state-of-the-art techniques for abnormal human activity recognition. Engineering Applications of Artificial Intelligence, 77: 21-45. https://doi.org/10.1016/j.engappai.2018.08.014

[35] Tariq, S., Farooq, H., Jaleel, A., Wasif, S.M. (2021). Anomaly detection with particle filtering for online video surveillance. IEEE Access, 9: 19457-19468.

[36] Zhang, X., Yang, S., Zhang, J., Zhang, W. (2020). Video anomaly detection and localization using motion-field shape description and homogeneity testing. Pattern Recognition, 105: 107394. https://doi.org/10.1016/j.patcog.2020.107394

[37] Wang, Z., Yang, Z., Zhang, Y.J. (2020). A promotion method for generation error-based video anomaly detection. Pattern Recognition Letters, 140: 88-94. https://doi.org/10.1016/j.patrec.2020.09.019

[38] Ullah, H., Altamimi, A.B., Uzair, M., Ullah, M. (2018). Anomalous entities detection and localization in pedestrian flows. Neurocomputing, 290: 74-86. https://doi.org/10.1016/j.neucom.2018.02.045

[39] Verma, K.K., Singh, B.M., Dixit, A. (2019). A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system. International Journal of Information Technology, 1-14. https://doi.org/10.1007/s41870-019-00364-0

[40] Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., Baik, S.W. (2018). Efficient deep CNN-based fire detection and localization in video surveillance applications. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(7): 1419-1434. https://doi.org/10.1109/TSMC.2018.2830099

[41] Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M. (2010). Visual object tracking using adaptive correlation filters. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2544-2550. https://doi.org/10.1109/CVPR.2010.5539960

[42] Chen, D., Wang, P., Yue, L., Zhang, Y., Jia, T. (2020). Anomaly detection in surveillance video based on bidirectional prediction. Image and Vision Computing, 98: 103915. https://doi.org/10.1016/j.imavis.2020.103915