Iraqi Journal of Information and Communication Technology(**IJICT**)
Vol. 5, Issue 3, December 2022

www.ijict.edu.iq

ISSN:2222-758X
e-ISSN: 2789-7362

DETECTION OF ANOMALOUS EVENTS BASED ON DEEP LEARNING-BILSTM

Zainab K. Abbas [1], Ayad A. Al-Ani [2]
[1,2] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
zainabkudair@gmail.com [1] , ayad.a@nahrainuniv.edu.iq [2]
Received:16/2/2022, Accepted:10/5/2022
DOI:**10.31987/ijict.5.3.207**

Abstract- **Video anomaly detection in smart cities is a critical errand in computer vision that plays an imperative role in intelligent surveillance and public security but is challenging due to its different, complex, and rare events in real-time surveillance situations. Different deep learning models utilize a critical amount of training data without generalization capabilities and with high time complexity. To overcome these problems, an algorithm for reducing the size of the map of the extracted features has been suggested, and this was done by combining the features of every 15 video frames to generate the new feature vectors that will be fed into our classifier model. The values of the new feature vectors represent the summation of the values of the original feature vectors obtained from Resnet50. Finally, the new feature vectors are fed into our classifier model to detect the abnormality. A comprehensive test on a variety of anomaly detection benchmark datasets has been conducted to verify the proposed framework's functionality in complex surveillance scenarios. The numerical results were carried out on the UCF-Crime dataset, with the proposed approach achieving Area Under Curve (AUC) scores of 93.61% on the database's test set.**

keywords: **Video surveillance, Anomalous detection, Deep learning, Resnet50, BiLSTM.**

I. Introduction

Video Surveillance Systems (VSS) are widely utilized in public and private areas to increase public safety, such as shopping malls, markets, banks, hospitals, educational institutions, streets, and smart cities [1]. The accuracy and fast identification of video anomalies is usually the major goal of security applications [2]. Real-time video analysis and anomalous case detection necessitate many human resources and are subject to mistakes due to a loss of human attention over time. Automatic anomaly detection technologies based on Artificial Intelligence (AI) mechanisms are required in surveillance systems because human observation is ineffective [3]. The wide applications of abnormal event recognition in surveillance recordings, such as crime prevention, automated intelligent visual monitoring, and traffic security, take a lot of attention [4]. Most existing approaches suffer from a high percentage of false alarms. Additionally, while these strategies perform well on basic datasets, their effectiveness is restricted when applied to real-world circumstances. To address these challenges, an algorithm based on combining features to boost them and reduce the dimensions of the video feature map has been suggested to improve the performance of the model and reduce its complexity, and that was done by combining the features of 15 consecutive frames after features were extracted using pre-trained convolution neural networks Resnet50, the combination of the features of 15 consecutive frames, was done by taking the summation of their values and generating new feature vectors. These vectors will be fed into our classifier model, which is a bidirectional long-short-term memory (BiLSTM). A weakly-supervised technique based on spatiotemporal features and BiLSTM has been used to train the classifier model. The BiLSTMs have proven to be very useful when the context of the input is required. Information moves from backward to forward in a unidirectional LSTM. On the other hand, BiLSTM uses two

IJICT Iraqi Journal of Information and Communications Technology

Iraqi Journal of Information and Communication Technology(**IJICT**)
Vol. 5, Issue 3, December 2022

www.ijict.edu.iq

ISSN:2222-758X
e-ISSN: 2789-7362

hidden states to flow information not only backward to forward but also forward to backward. As a result, Bi LSTMs have a greater understanding of the context [5]. The main contributions of our research are:

- Using pre-trained Convolutional Neural networks (CNN) Resnet50 for extracting the Spatial-Temporal Features.
- Combining every 15 consecutive frames feature to decrease the size of the feature map.
- The suggested framework employs a BiLSTM architecture. Due to the forward and backward passes applied in every layer of the LSTM model to detect the abnormality.
- The work was done on the UCF-Crime dataset [6] which is a challenging benchmark dataset.

The rest of this paper is structured as follows: The second section looks at a literature assessment of existing approaches. The general recommended framework is explained in Section 3. Section 4 assesses our research's experimental outcomes and compares them to existing methodologies, with Section 5 providing a conclusion and recommendations for future research directions.

II. Related Work

A variety of scenarios have been explored to detect various types of abnormalities. All the scenarios adopted, however, were tailored to a specific situation. A lot of models previously utilized for anomaly detection have been discussed in the following sections. Waqas et al. [7] propose to learn abnormalities by utilizing both anomalous and normal videos to present a system that can recognize anomalous attitudes and alert the user to the kind of abnormal behavior. The authors suggest using the deep multiple instance ranking framework to learn anomalies from weakly labeled training videos, where the training labels (normal or anomalous) are applied at the video level rather than the clip level. The UCF Crime dataset was used in this work, and the AUC value achieved was 75.41%. Shreyas et al. [8] offered a new implementation concept in which the videos are adaptively compressed before being passed via the activity recognition system. The research was done on the UCF101 crime dataset, and the AUC value achieved was 79.8%. Anala et al. [9] describe a system that can recognize abnormal actions and alert the user depending on the type of anomaly. This work treats anomaly detection as a regression problem. The evaluation of the performance of this approach was done on normal videos only. Using solely normal data may not be the best method for detecting anomalies. The experimentation was done on the UCF-Crime dataset, and the AUC value achieved was 85%. Hao et al. [10] suggest a unique two-stream convolutional network model. The suggested model consists of RGB and Flow two-stream networks, with the combined score representing the final anomaly event detection score. The anomaly detection problem has been treated as a regression problem. The experimentation was done on the UCF-Crime dataset, and the AUC value achieved was 81.22%. Dubey et al. [11] proposed a framework that is a deep network with Multiple Ranking Measures (DMRMs). The anomaly detection problem has been treated as a regression problem. They used 3D ResNet-34 for anomaly detection purposes, and the implementation was done on the UCF-Crime dataset, and the AUC value achieved was 81.91%. Ullah et al. [2] present an efficient light-weight convolutional neural network (CNN)-based anomaly recognition framework that is functional in a surveillance environment with reduced time complexity. They used CNN (pre-trained light-weight CNN)-multilayer bidirectional LSTM for anomaly detection purposes, and the implementation was done on the UCF-Crime (4 class anomaly in addition to normal), and the AUC value achieved

ISSN:2222-758X
e-ISSN: 2789-7362

was 78.43%. Ullah et al. [4] proposed an efficient deep features-based intelligent anomalous detection framework that can operate in surveillance networks with less time complexity. The CNN-ResNet-50 with Multilayer BiLSTM was used for anomaly detection purposes. The implementation was done on the UCF-Crime database, and the AUC value achieved was 85.53%. Zaheer et al. [12] came up with a weakly supervised anomaly detection method that simply uses video-level labels to train. A pre-trained feature extractor model such as Convolution 3D (C3D), Fully Connected Network, and k-mean was used in the framework, and the implementation was done on the UCF-Crime database, and the AUC value achieved was 78.27%. By adopting a weakly supervised learning paradigm, Majhi et al. [13] proposed a method that jointly handles anomaly detection and classification in a single framework. A 3D convolutional network (I3D)-many-to-many LSTM was used for anomaly detection purposes, and the implementation was done on the UCF-Crim database, and the AUC value achieved was 82.12%. Wu et al. [14] suggested a dual branch network that takes as input multi-granularity ideas in both the spatial and temporal dimensions. The features were extracted by C3D, and the implementation was done on a new dataset (denoted as ST-UCF-Crime) that annotates Spatio-temporal bounding boxes for abnormal events in UCF-Crime. The AUC value achieved for the ST-UCF-Crime dataset was 87.65%. Zainab K. Abbas and A. A. Al-Ani [15] provide an efficient model for real-world outlier detection in a surveillance system, they suggest compressing each video using High-Efficiency Video Coding (H265) before feeding them into anomaly detection systems, then each video was fed into pre-trained Resnet50 for features extraction purposes; finally, the features vector was fed into BiLSTM for normal and abnormal class identification. The implementation of this work was done on the UCF-Crim dataset, and the AUC value achieved was 90.16 %.

III. Proposed Methodology

In this section, the overall suggested framework and its essential process are described as shown in Fig. 1. The proposed methodology is divided into three stages:

- Feature extraction.
- Reducing features map dimension.
- Using BiLSTM for training and classification.

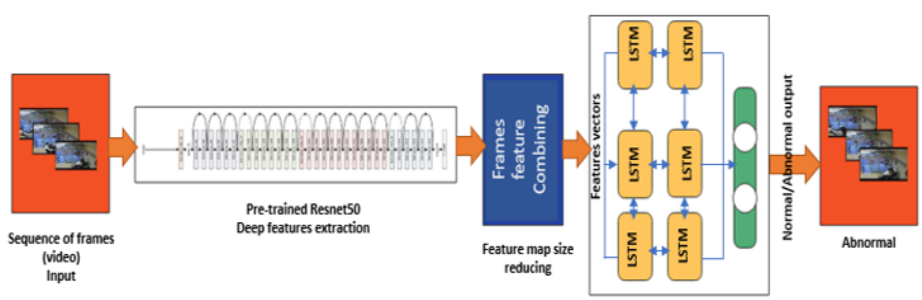

Figure 1: The suggested framework

ISSN:2222-758X
e-ISSN: 2789-7362

Fig. 2 shows the sequence of stages proposed in this work, and the flowchart of the suggested work can be seen in Fig. 3. The proposed methodology is given as follows:

Figure 2: The methodology stages

A. Feature Extraction

For the extraction of the features, the pre-trained CNN ResNet50 has been used. ResNet50 is a convolutional neural network that is 50 layers deep. This neural network takes each frame from the video as input and returns 1000 features for each frame. The Resnet model, which has a total of 50 layers with 23.5 million trainable parameters, has been used. The entire dataset was the UCF-Crime dataset. The UCF-Crime dataset includes both normal and abnormal events videos, as well as 13 various types of anomalies, such as fights, explosions, abuse, and accidents. There are 1900 surveillance movies in the collection, with an approximately equal amount of normal and abnormal videos. Whereas the training set had 800 normal and 810 anomalous samples, the testing set included the remaining 150 normal and 140 anomalous movies. The experimentation for this work was done on the UCF-Crime dataset on the video at a length equal to or less than 2 minutes. 1324 videos based on this condition have been used: 1116 videos for the training stage (90% for training and 10% for validation), and 208 videos for testing. In the preprocessing step, a center crop has been used to resize each video into 224×224, and then Resnet50 to extract the features has been used. The extracted features are taken out of the fc1000 layer and fed to the next stage to reduce the number of features.

B. Reducing The Dimensions of Features Maps

This stage involves combining the features of each of the 15 consecutive frames by taking the summation of their values and generating a new vector that represents the new anomalous features that will be fed into the classifier model, which is BiLSTM, to detect the abnormality. Different numbers of combined frames were tested, including 10, 15, 20, 25, and 30, with the best accuracy obtained when combining every 15 frames. Accordingly, this value was adopted in the work. Fig. 4 illustrates the proposed method for reducing the size of the feature map. Simultaneously, the accuracy for detecting anomaly events has improved, which is the goal of this paper.

C. BiLSTM Training

Bidirectional Long Short-Term Memory (BiLSTM) is a recurrent neural network consisting of two LSTMs, one taking the input in a forward direction and the other in a backward direction. The training of this work was done by using the Adam optimizer with a minibatch size equal to 16 and the maximum epoch number was 100 with an initial learning rate equal to 1e-5 and 450 nodes in the hidden layer with dropout equal to 0.7 and L2Regularization equal to 0.5 to reduce

Iraqi Journal of Information and Communication Technology(**IJICT**)
Vol. 5, Issue 3, December 2022
www.ijict.edu.iq
ISSN:2222-758X
e-ISSN: 2789-7362

the overfitting of the model. The model parameters have been chosen using trial and error. The loss function evaluation during the training stage is demonstrated in Fig. 5.

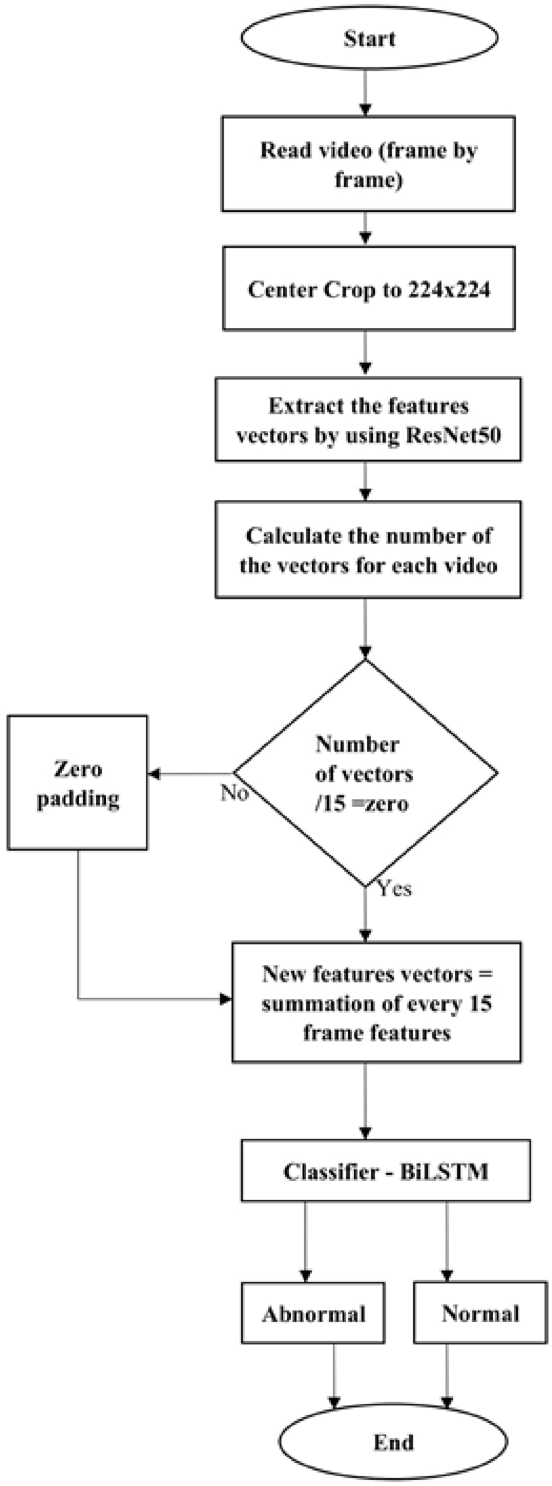

Figure 3: The flowchart of the suggested work

Iraqi Journal of Information and Communication Technology(**IJICT**)
Vol. 5, Issue 3, December 2022
www.ijict.edu.iq
ISSN:2222-758X
e-ISSN: 2789-7362

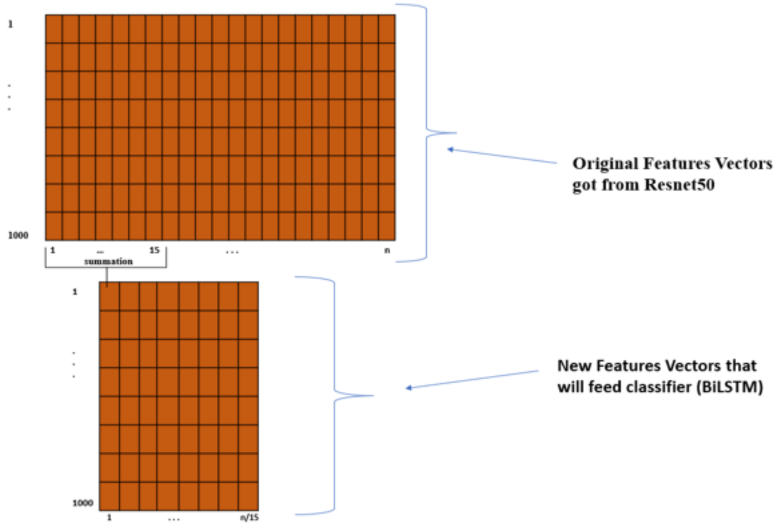

Figure 4: The proposed method for reducing the size of the features

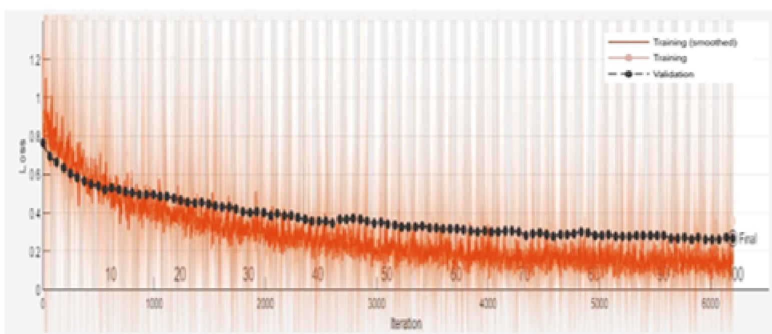

Figure 5: The loss function evaluation during the training stage

IV. Experimentation and Results

The experimentation for this work was done on the UCF-Crime dataset on the video at a length equal to or less than 2 minutes. 1324 videos based on this condition have been used: 1116 videos for the training stage (90% for training and 10% for validation), and 208 videos for testing. The receiver operating characteristics (ROC) and the AUC were utilized as performance indicators, which are commonly used in the state-of-the-art. The computer codes for the proposed work were implemented in the MATLAB software environment (version 2021a), Windows 10, Intel processor core i7, RAM 16GB, 1TB SSD hard drive, operating system 64bit, and NVIDIA GeForce MX450 graphics processing unit. The experimental results show the efficiency of our proposed framework, as it detects anomalous events with greater precision than existing methods. The ROC curve and the confusion matrix of our proposed work are shown in Fig. 6 and Fig. 7, respectively. The experimental result using ROC is presented in Fig. 6. This demonstrates that our framework accomplishes good

www.ijict.edu.iq
ISSN:2222-758X
e-ISSN: 2789-7362

performance. In addition, our proposed framework's success in reducing as possible the false-negative (anomaly classified as normal) and false-positive (normally classified as an anomaly) alarms is apparent in Fig. 7, where the value of False positive (FP) was equal to 19 and False negative (FN) was equal to 10, as shown in the confusion matrix of the model. Furthermore, we calculated the detection accuracy of our classifier, and it was 86.06%. And this illustrates the success of our proposal to reduce the size of the features map for each video at the same time, improving the accuracy of anomaly event detection. The AUC scores are compared with state-of-the-art techniques in Table I, and it can be noticed that our proposed framework achieved the highest AUC of 93.61% an increase of 3.45% compared to other existing methods.

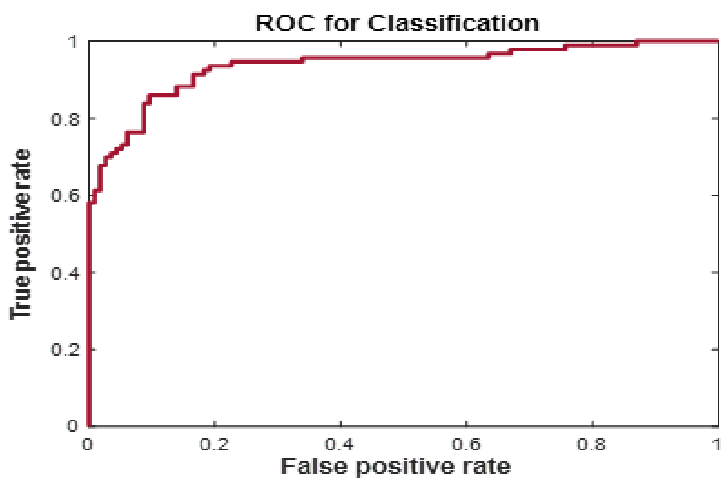

Figure 6: The ROC curve of the model

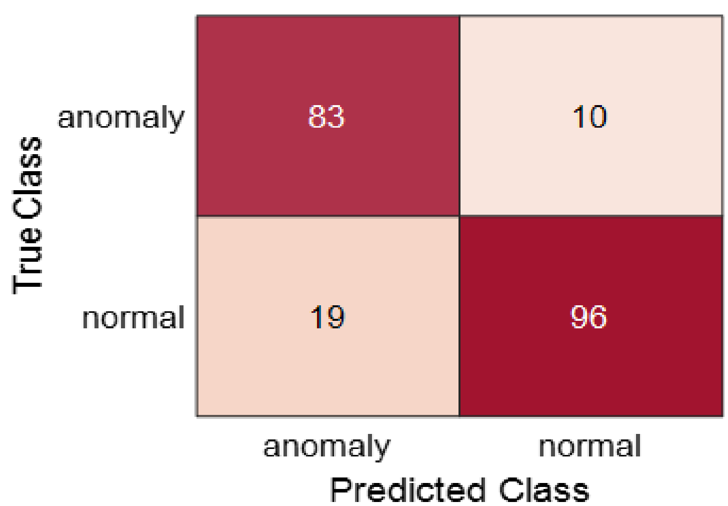

Figure 7: The Confusion matrix, Detection accuracy = 86.06%

TABLE I
Comparison of AUC score of proposed work with state-of-the-art techniques

Method of	AUC % on the UCF Crime dataset
Waqas and colleagues, 2019 [7]	75.41
Anala and colleagues, 2019 [9]	85
Shreyas and colleagues, 2020 [8]	79.8
Hao and colleagues, 2020 [10]	81.22
Dubey and colleagues, 2021 [11]	81.91
Ullah and colleagues, 2021 [2]	78.43
Ullah and colleagues, 2021 [4]	85.53
Zaheer and colleagues, 2021[12]	78.27
Majhi and colleagues, 2021 [13]	82.12
Wu and colleagues, 2021 [14]	87.65
Z. K. Abbas and A. A. Al-Ani [15]	90.16
Ours	93.61

V. Conclusion

In this paper, an efficient model for real-world outlier detection in surveillance systems has been proposed. The proposed model is based on current anomaly datasets with state-of-the-art accuracy. Firstly, each video has been resized into 224x224, and then, by using pre-trained Resnet50, the feature vectors were extracted for each video. To improve the performance of the surveillance system model, we suggest reducing the size of the feature map extracted by Resnet50 before feeding it into the classifier model. That was done by combining the video frame features, where the values of every fifteen frames are combined to generate the new feature vectors. The values of the new feature vectors represent the summation of the values of the original feature vectors obtained from Resnet50. Finally, the new feature vectors were fed into Bi-LSTM for normal and abnormal class identification. In comparison to contemporary literature on anomalous detection approaches, the proposed framework has been shown to have superior accuracy. For the UCF-Crime dataset, the experimental results show an increase in the AUC value of up to 93.61%, an increase of 3.45% compared to other methods. Furthermore, we calculated the detection accuracy of our classifier, and it was 86.06%. And this illustrates the success of our proposal to reduce the size of the feature map for each video at the same time, improving the accuracy of anomaly event detection, which is our main goal, with as little as a possible false alarm for both positive and negative. In the future, we will attempt to increase the accuracy indicator by studying feature selection algorithms and dimensionality reduction algorithms and combining them with our suggested framework.

Iraqi Journal of Information and Communication Technology(**IJICT**)
Vol. 5, Issue 3, December 2022
www.ijict.edu.iq
ISSN:2222-758X
e-ISSN: 2789-7362

REFERENCES

[1] Z. K. Abbas and A. A. Al-Ani, "A Comprehensive Review for Video Anomaly Detection on Videos" , IEEE 2022 International Conference on Computer Science and Software Engineering (CSASE), pp. 1-1, 2022.
[2] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN Features with Bi-Directional LSTM for Real-Time Anomaly Detection in Surveillance Networks" , Multimed. Tools Appl. , Vol. 80, No. 11, pp. 16979-16995, 2021.
[3] H. I. Ozturk and A. B. Can, "ADNet: Temporal Anomaly Detection in Surveillance Videos" , Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) , Vol. 12664 LNCS, pp. 88-101, 2021.
[4] 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.
[5] A. Aziz Sharfuddin, M. Nafis Tihami, and M. Saiful Islam, "A Deep Recurrent Neural Network with BiLSTM Model for Sentiment Classification" , 2018 Int. Conf. Bangla Speech Lang. Process. ICBSLP 2018, No. September, 2018.
[6] "UCF-Crime Dataset" , [Online]. Available: %09Dropbox- Anomaly-Detection-Dataset - Simplify your life.
[7] M. S. Waqas Sultani, Chen Chen, "Real-World Anomaly Detection in Surveillance Videos" , 2019.
[8] D. G. Shreyas, S. Raksha, and B. G. Prasad, "Implementation of An Anomalous Human Activity Recognition System" , SN Comput. Sci. , Vol. 1, No. 3, 2020.
[9] M. R. Anala, M. Makker, and A. Ashok, "Anomaly Detection in Surveillance Videos" , Proc. - 26th IEEE Int. Conf. High Perform. Comput. Work. HiPCW 2019, pp. 93-98, 2019.
[10] W. Hao et al. , "Anomaly Event Detection in Security Surveillance Using Two-Stream Based Model" , Hindawi Secur. Commun. Networks, Vol. 2020, 2020.
[11] S. Dubey, A. Boragule, J. Gwak, and M. Jeon, "Anomalous Event Recognition in Videos Based on Joint Learning of Motion and Appearance with Multiple Ranking Measures" , Appl. Sci. , Vol. 11, No. 3, pp. 1-21, 2021.
[12] M. Z. Zaheer, J. Lee, M. Astrid, A. Mahmood, and S. I. Lee, "Cleaning Label Noise with Clusters for Minimally Supervised Anomaly Detection" , pp. 3-6, 2021, [Online]. Available: http://arxiv.org/abs/2104.14770.
[13] S. Majhi, S. Das, F. Bremond, R. Dash, and P. K. Sa, "Weakly-Supervised Joint Anomaly Detection and Classification" , 2021, [Online]. Available: http://arxiv.org/abs/2108.08996.
[14] J. Wu et al. , "Weakly-Supervised Spatio-Temporal Anomaly Detection in Surveillance Video" , pp. 1172-1178, 2021.
[15] Zainab K. Abbas and Ayad A. Al-Ani, "Anomaly Detection in Surveillance Videos Based on H265 and Deep Learning" , In press.