Sketch and Size Orient Malicious Activity Monitoring for Efficient Video Surveillance Using CNN

K. Lokesh[1], M. Baskar[2]*

Research Scholar, Department of Computer Science and Engineering-School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu 603 203, India[1]
Associate Professor, Department of Computing Technologies-School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamilnadu 603 203, India[2]

***Abstract*—Towards malicious activity monitoring in organizations, there exist several techniques and suffer with poor accuracy. To handle this issue, an efficient Sketch and Size orient malicious activity monitoring (SSMAM) is presented in this article. The model captures the video frames and performs segmentation to extract the features of frames as shapes and size. The video frames are enhanced for its quality by applying High Level Intensity Analysis algorithm. The quality improved image has been segmented with Color Quantization Segmentation. Using the segmented image, the feature are extracted and applied with scaling and rotation for different number of size and angle. Such features extracted have been trained with convolution neural network. The CNN model is designed to perform convolution on two levels and pooling as well. At the test phase, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance. The proposed approach improves the performance in activity monitoring and video surveillance.**

***Keywords*—Video surveillance; deep learning; activity monitoring; malicious activity; SSMAM; SSM**

I. Introduction

Growth of information and communication technology has been adapted to various scientific, medical and security problems. Industrial security is the most concern in recent days and they tends to enforce the human monitoring in the units of organization. Video surveillance is the process of monitoring the human activity in various locations and units of any organization. In this way, they can monitor the activity of human, worker towards enforcing security and maintaining the human resource [5, 6]. By monitoring the human activity, the malicious event happening within or outside the organization can be tracked and used towards enforcing security.

The video frames captured through set of video devices and cameras are the key input for the video surveillance system of any organization. By using the images of video, the human object can be tracked in successive frames and their activity can be classified. Image processing plays vital role in identifying the human texture and can be used to extract the features from the frame to support the classification of activity. In order to perform event or activity classification, the system has to maintain number of frames and objects of humans about different activities. By maintaining such set of frames and objects, the input frame can be classified for its activity [7].

To perform video surveillance there are numbers of machine learning algorithms described in literature. For example, support vector machine (SVM), Decision Tree, Bayesian Classifier, Neural Network are used in earlier articles [11]. The methods capture the video frame and perform background subtraction which is learned from the different previous frames. With the subtracted image, the method would identify the human object to extract the object and train the model. The issue with the machine learning algorithms is the dimensionality and missing features. In order to achieve higher performance in video surveillance, it is necessary to maintain the set of human objects under different activities. With the trace, the method can identify the class of any video frame for its activity. The machine learning algorithms are little uncomfortable to maintain and handling such huge volume of frames. This is where the deep learning algorithms come to play.

The deep learning algorithms are capable of converting the huge volume of features into small set without the loss of feature and data. For example, Convolution neural network is more popular on handling such higher volume data. It would convert the features of video frames into tiny feature set and used to perform classification. With all these consideration, this article presents a novel Sketch and Size Orient Malicious Activity Monitoring (SSMAM) model. The model concentrates on extracting the human features and sketch features. By extracting the sketch and size features, the model has been trained with number of features at each class with convolution neural network. The SSMAM-CNN model is capable of handling various class of activities and support the detection of malicious activity in the premises. By adapting SSMAM-CNN model the human activity can be greatly monitored.

A. Problem Statement

The methods discussed in literature uses variety of features from object, graph, edge, etc. towards human activity monitoring. But still they suffer to achieve higher accuracy towards malicious activity monitoring. The accuracy of video surveillance and activity monitoring is greatly depending on the kind of feature considered. The human sketch is the most effective one which can be used in activity monitoring to detect any malicious activity happening in the venue.

B. Contributions

- The problem of activity monitoring is approached with sketch and size features along with other features.
- CNN model is designed for training and classification.
- Sketch Support Measure (SSM) is estimated to perform classification.

The article is structured to provide general introduction about the problem in Section I. Section II briefs the literature review around the problem and Section III presents the detailed working of the model. Section IV is dedicated to present the experimental result and Section V discuss the conclusion.

II. RELATED WORK

Different methods are analyzed around video surveillance and activity monitoring. Set of methods around the problem is discussed here.

A Gaussian Mixture Model with Universal attribute model is presented in [1], which monitor the violent activities by computing super action vector. A Growing self-organizing map (GSOM) is presented in [2], which uses traditional deep learning and multistream learning to use unlabeled data to handle the over fitting problem to identify malicious activity.

An Efficient Marine Organism Detector (EMOD) is discussed in [3], which uses attention relation to detect malicious activity. A deep learning based video surveillance model is presented in [4], which classify the objects under different categories and identify the region in the frame to perform supervised learning. A human action recognition (HAR) model is presented to monitor the activity in aircraft surveillance. The method uses temporal features and use LSTM to perform classification.

A visual surveillance framework is presented in [6], to monitor the human fall which uses deep CNN and uses aggregated heuristic visual features in detecting the occurrence of falls. An audio video based activity recognition model is presented in [7], towards video surveillance.

A human activity recognition model is sketched in [8], which uses data enhancement techniques to collect discriminative features from various activities. The features captured are transformed with existing model to perform recognition. A hybrid visual geometry group based bidirectional short term memory model is presented in [9], towards monitoring the moments of animals and produces alarm according to the activity of the animals. A transformer network based LSTM model is presented in [10], towards recognizing the activity of the human.

A multipedestrian multicamera tracking (IMPMCT) model is presented in [11], which uses spatial and temporal features towards monitoring the pedestrian monitoring. An Interactive video surveillance as an edge service (I-ViSE) model is presented in [12], which works according to the feature queries. The method adapts features of human body, color and cloths in activity monitoring. An LSTM based baby activity recognition model is presented in [13], which uses pose features of baby towards recognition.

A Class-privacy Preserved Collaborative Representation (CPPCR) based multi-modal human action recognition model is presented in [14], which uses temporal structure in recognizing the activity. An LSTM with 3DCNN model is presented in [15], towards detecting illegal activity in the environment.

A realtime surveillance model is presented in [16], towards detecting malicious behavior using deep transfer learning. A posture recognition system is presented in [17], which uses mobilenetV2 towards estimating the human posture and LSTM is used to extract the features. Machine learning algorithm is used for classification. A secure surveillance scheme is presented in [18], to support healthcare system with enabled internet of Things. A deep learning multi layered CNN with LSTM is presented in [19], to capture the physical activities of persons and performs activity classification.

A machine learning based harmful weapon detection model is presented in [20], which uses different pistol classes and objects. The method uses sliding window classification model to perform classification. Lightweight Deep Neural Network (LDNN) with Convolution Long Short Term Memory (ConvLSTM) model is presented in [21], towards detecting abnormal activity in the environment. A Graph Convolution Network with 3DCNN is presented in [22], towards detecting abnormal behavior. 3DNN is used to extract the features and GCN has been used to perform classification. An audio and visual based activity detection model is sketched in [23], which uses both audio and video data in detecting abnormal activity in video frames. The method uses PSO and social force model towards feature extraction and classification. A Faster Region-Based Convolutional Neural Networks (Faster RCNN) is presented in [24], to detect the firearms in organization. The model uses ensemble learning to detect the human face and guns using Weighted Box Fusion techniques.

A spatio temporal attention fusion slowfast network based model is presented in [25], which uses spatial and temporal features to perform classification. A 2D and 3D feature based HAR is presented in [26], which extract the geometric features using deep and machine learning algorithms like CNN to classify the activity of human using LSTM and SoftMax classifier. A multi perspective abnormal posture recognition model is presented in [27], which works according to multi view cross information and posture features towards detecting illegal activity.

All the above discussed approaches suffer to achieve higher performance in activity monitoring and video surveillance.

Research Gap: According to the literature, the existing works does not consider the sketch and size features with the combination of other object features. This affects the performance of activity monitoring.

III. SKETCH AND SIZE ORIENT MALICIOUS ACTIVITY MONITORING WITH CNN (SSMAM-CNN)

The model captures the video frames and performs segmentation to extract the features of frames as shapes and size. The video frames are enhanced for its quality by applying High Level Intensity Analysis algorithm. The quality improved image has been segmented with Color Quantization

Segmentation. Using the segmented image, the feature are extracted and applied with scaling and rotation for different number of size and angle. Such features extracted have been trained with convolution neural network. The CNN model is designed to perform convolution on two levels and pooling as well. At the test phase, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance.

The functional architecture of SSMAM-CNN model has been presented in Fig. 1 and the functions of the model are briefed in this section.

A. High Level Intensity Analysis Preprocessing

The high level intensity analysis algorithm enhances the quality of video frame given. To perform this, the method fetch the image and extract the RGB layers. Among the three layers, the method use the red layer features to enhance the image quality. The method initializes a window size and crops the features from the red layer. Among the window feature, the method computes the max intensity value (MIV) and least intensity value (LIV). Using these two, the method traverse through each pixel and computes the Max Intensity Distance and Least Intensity Distance. Based on the value of distances, the method identifies the closest intensity sector and computes the Intensity Normalization Value. Based on that, the pixel has been adjusted for its intensity. The quality enhanced image has been used to perform segmentation and activity monitoring.

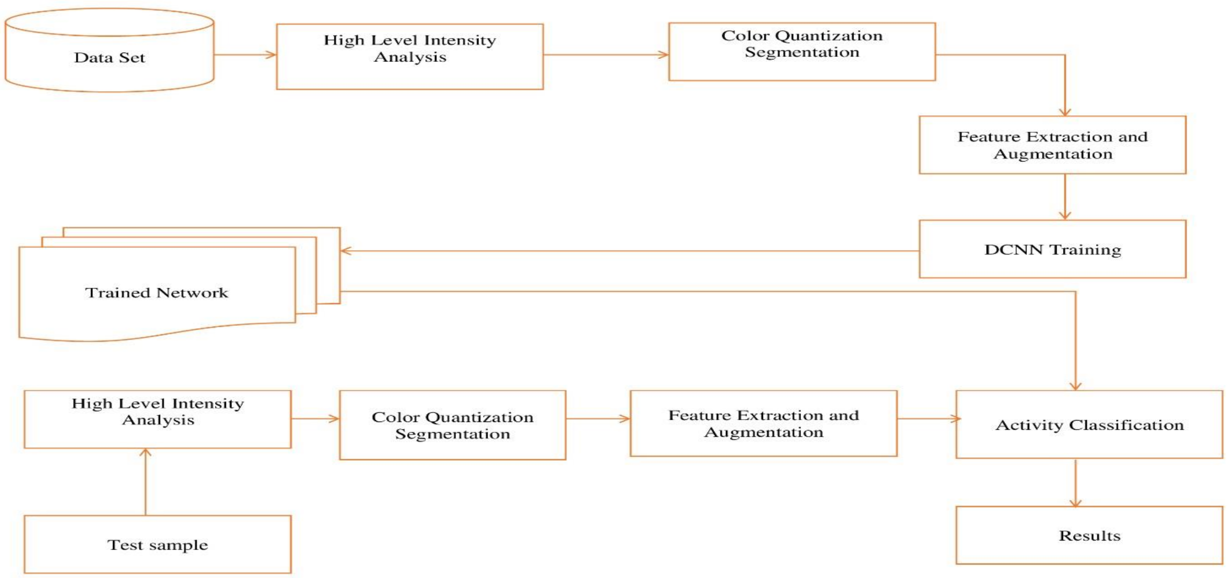

Fig. 1. Architecture SSMAM-CNN model.

Algorithm:

Given: Video Frame Vf
Obtain: Enhanced Video Frame Evf.
Start
Read Vf.
[RGB]= Extract RGB Layers (Vf)
Initialize window size Ws = 5
For each window w
Compute Max Intensity Value Miv = $\underset{i=1}{\overset{Size(w)}{Max}}(W(R(i))$
Compute least intensity value LIV = $\underset{i=1}{\overset{Size(w)}{Min}}(W(R(i))$

For each pixel p
Compute Max intensity Distance Mid = Dist(p(R), Miv)
Compute Least intensity Distance Lid = Dist (P (R), LIV)
if Mid<Lid then
Compute intensity normalization value Inv = p(R)+Mid/3
Evf(W(p(R))) = Inv
End
End
End
Stop

The high level intensity normalization algorithm computes the INV value for different pixels of red layers of the frame to adjust the quality of the frame. The enhanced image has been used towards segmentation and activity monitoring.

B. Colour Quantization Segmentation

The colour quantization algorithm segments the image according to the colour features. The quantization is performed based on the RGB values obtained from the frame given. To perform this, the method reads the red green black layer features and for each layer feature, the method computes the Quantization Measure (QM) which is measured based on the overall feature value and identifies the non-dominant value. To obtain this, the method generates the histogram for each layer and identifies the feature value with same numbers. Further, the method segments the identified pixels with same quantization measure to perform segment the image. The pixels selected with the same quantization value has been grouped and produces segmented image.

The colour quantization segmentation algorithm groups the pixels according to the QM value identified from the histogram of all layer features to produce segmented image.

Algorithm:

Given: Enhanced Image Eimg
Obtain: Segmented Image Simg
Start
    Read Eimg
    [R,G,B]= Extract RGB features.
    For each layer R, G, B
        Histogram Histo = Generate Histogram (R, G, B)
        For each histogram value h
            For each other layer histogram Hi
                Compute QM = $\sum_{i=1}^{size(Hi)} Hi(i) == h$
            End
        End
    End
    For each pixel pi
        If Pi(R,G,B)== QM(hi) then
            Add to group
            Simg(pi)= 0
        End
    End
Stop

IV. Feature Extraction and Augmentation

The proposed model extract the sketch features from the segmented image and computes the size of the sketch. According to the sketch feature, the method generates various size of sketch to produce data augmentation. Augmented feature set has been used to perform training and testing.

A. DCNN Training

The activity data set given has been fetched and the frames of the data set has been used to perform high level intensity analysis which enhance the image. Further, the method applies color quantization segmentation to segment the image. From the segmented image, the method extract the sketch and size features and produces augmented data. With the augmented data, the method generates the convolution neural network, which comes with two level of convolution. At the first level, the sketch features are convolved to one dimension feature and at the second convolution, the method generates unique size of feature set. The max pooling layers adjust the values of the feature. The method generates number of neurons according to the number samples and augmented data given. Each has been initialized with the unique sample given. At the first convolution layer, the sketch features are convolved to unique standard size and at the second convolution layer, the neurons convert the features into one dimensional array. Finally, the size features are append to the feature set. Generated network has been used to perform activity classification.

B. Activity Classification

At the test phase, the method applies high level intensity analysis and color quantization segmentation on the given video frame. Further, the method extract the same set of features and performs convolution to obtain same set of feature lengths and the neurons are designed computes the value of Sketch Support Measure (SSM) towards various class of activity. According the value of SSM, the method classifies the user activity towards efficient video surveillance.

Algorithm:

Given: Video Frame Vf
Obtain: Activity_Class Ac
Start
    Read Vf and CNN.
    Eimg = Perform high level intensity analysis on Vf.
    Simg = Apply color quantization segmentation (Eimg)
    [sketch, size] = Feature Extraction and Augmentation on Simg.
    Pass through CNN
    At convolution layer one
        Convert sketch feature to standard size.
    Perform Max pooling
    At convolution layer two
        Convert sketch feature to one dimension value.
    Perform max pooling
    At output layer
        For each class of activity
            For each neuron n
                Compute Sketch support measure SSM.

$$SSM = \frac{\sum_{i=1}^{size(n(f))} count(n(f).value == Vf(f(i)))}{size(Vf(f))}$$

            End
        Compute cumulative SSM = $\frac{\sum SSM}{size(neurons)}$
        End
    Activity_Class = Choose class with maximum SSM.
Stop

The activity classification algorithm computes the SSM value for various class of activity according to the feature considered. Based on the value of SSM, the method performs classification.

V. Results and Discussion

The sketch and size based malicious activity monitoring model is hardcoded with MATLAB. The performance of the method is evaluated under various factors. The method uses various data sets with different poses and shapes. The results obtained have been compared with the results of other approaches.

The evaluation detail used towards evaluating the performance of various methods are measured and presented in this section. The methods discussed in literature are evaluated for their performance with different data sets according to Table I. Their performances are evaluated under various parameters and the results produced by the approaches are compared with others in this section.

TABLE I. EVALUATION DETAILS

Parameter	Value
Number of Activities	20
Total Images	10000
Tool	MATLAB
No of users	700

TABLE II. ANALYSIS OF MALICIOUS ACTIVITY DETECTION ACCURACY

Malicious Activity Detection Accuracy % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DMPMAM	82	88	95
DFI_SVQA	72	75	79
MulVIS	76	81	85
SSOcT	71	77	82
SSMAM_CNN	85	91	97

The accuracy of methods in detecting malicious activity has been counted and analyzed in Table II, where the SSMAM_CNN scheme stimulates higher accuracy than other techniques.

The accuracy in detecting malicious activity is measured and compared in Fig. 2. The SSMAM_CNN model improves the performance at the increasing number of activity classes and videos.

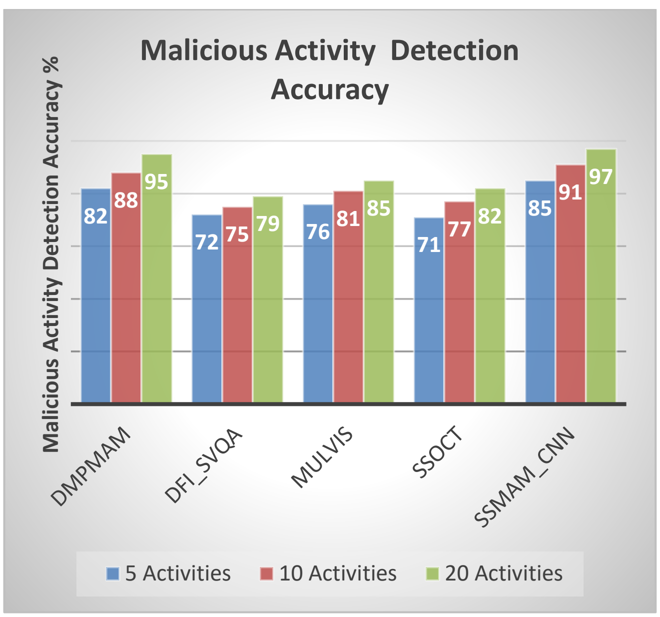

Fig. 2. Malicious activity detection accuracy.

The ratio of false detection is measured and compared in Table III, where the SSMAM_CNN approach has produced fewer ratios than other techniques.

TABLE III. FALSE RATIO IN MALICIOUS ACTIVITY DETECTION

False Ratio % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DMPMAM	18	12	5
DFI_SVQA	18	25	21
MulVIS	24	19	15
SSOcT	29	23	18
SSMAM_CNN	15	9	3

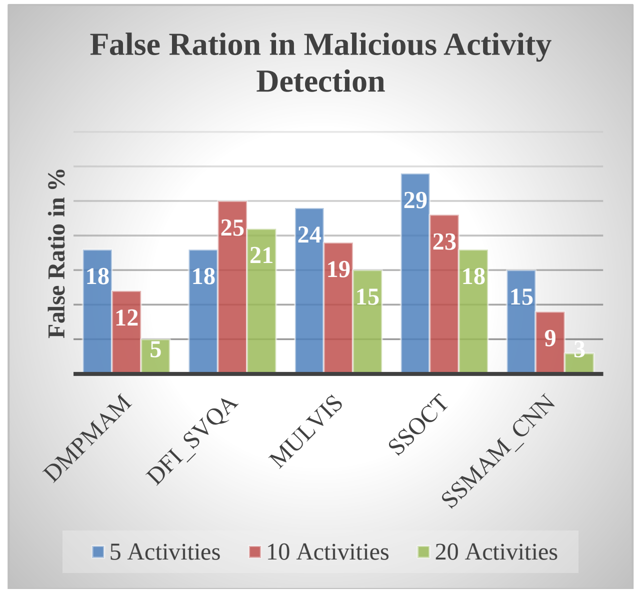

Fig. 3. False ratio in malicious activity detection.

The false classification ratio introduced by various approaches in malicious activity detection is measured & compared in Fig. 3. The SSMAM_CNN models introduces negligible false detection ratio compare to others in all the cases.

TABLE IV. TIME COMPLEXITY IN MALICIOUS ACTIVITY DETECTION

Time Complexity in Malicious Activity Detection % vs No of Activities			
	5 Activities	10 Activities	20 Activities
DFI_SVQA	67	77	88
DMPMAM	21	32	45
MulVIS	58	73	83
SSOcT	48	62	77
SSMAM_CNN	18	26	37

The time complexity in Millie seconds produced by various methods in detecting the malicious activity is measured and compared in Table IV, where the SSMAM_CNN approach produced less time complexity compare to other approaches.

The time complexity produced by various approaches in malicious activity detection is measured and compared in Fig. 4. The SSMAM_CNN model has produces less time complexity in all the test cases.

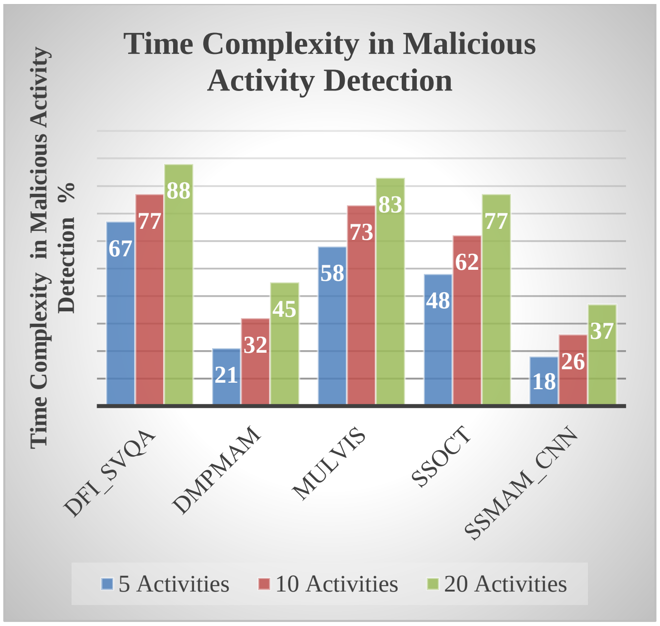

Fig. 4. Analysis time complexity in malicious activity detection.

VI. Conclusion

To maximize the accuracy of malicious activity detection, this article presented a novel sketch and size based malicious activity detection model with CNN (SSMAM_CNN). Apart from using object and shape features towards the problem, this approach uses the sketch and size features of the person to classify the activity. By adapting sketch and size features with the problem, the accuracy has been greatly improved. The model applies high level intensity analysis to enhance the quality of the image. Further, the method applies color quantization segmentation to segment the image. Next, the features are extracted and augmented data has been produced to train the convolution neural network. At the test phase, the method computes the sketch support measure (SSM) towards various class of activity features. According to the value of SSM, the class with maximum SSM value has been selected to perform the classification. The proposed model improves the performance of malicious activity detection accuracy up to 97%.

Further, the research can be extended by adapting invariant position features to stimulate the performance.

VII. Funding Statement

The authors received no specific funding for this study.

Conflicts of Interest

The authors declare they have no conflicts of interest to report regarding the present study.

References

[1] M. Mudgal, D. Punj and A. Pillai, "Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling," in Journal of Web Engineering, vol. 20, no. 1, pp. 129-146, January 2021, doi: 10.13052/jwe1540-9589.2017.

[2] R. Nawaratne, D. Alahakoon, D. De Silva, H. Kumara and X. Yu, "Hierarchical Two-Stream Growing Self-Organizing Maps With Transience for Human Activity Recognition," in IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7756-7764, Dec. 2020, doi: 10.1109/TII.2019.2957454.

[3] Z. Shi et al., "Detecting Marine Organisms Via Joint Attention-Relation Learning for Marine Video Surveillance," in IEEE Journal of Oceanic Engineering, vol. 47, no. 4, pp. 959-974, Oct. 2022, doi: 10.1109/JOE.2022.3162864.

[4] S. A. Ahmed et al., "Query-Based Video Synopsis for Intelligent Traffic Monitoring Applications," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 8, pp. 3457-3468, Aug. 2020, doi: 10.1109/TITS.2019.2929618.

[5] M. Ding, Y. Ding, L. Wei, Y. Xu and Y. Cao, "Individual Surveillance Around Parked Aircraft at Nighttime: Thermal Infrared Vision-Based Human Action Recognition," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no. 2, pp. 1084-1094, Feb. 2023, doi: 10.1109/TSMC.2022.3192017.

[6] Y. Zhang, X. Zheng, W. Liang, S. Zhang and X. Yuan, "Visual Surveillance for Human Fall Detection in Healthcare IoT," in IEEE MultiMedia, vol. 29, no. 1, pp. 36-46, 1 Jan.-March 2022, doi: 10.1109/MMUL.2022.3155768.

[7] S. Cristina, V. Despotovic, R. Pérez-Rodríguez and S. Aleksic, "Audio- and Video-Based Human Activity Recognition Systems in Healthcare," in IEEE Access, vol. 12, pp. 8230-8245, 2024, doi: 10.1109/ACCESS.2024.3353138.

[8] Amarendra Reddy Panyala, M. Baskar, "Real-time GB pattern convolution neural network-based brain image classification", AIP Conf. Proc. 3075, 020056 (2024), https://doi.org/10.1063/5.0218626.

[9] B. Natarajan, R. Elakkiya, R. Bhuvaneswari, K. Saleem, D. Chaudhary and S. H. Samsudeen, "Creating Alert Messages Based on Wild Animal Activity Detection Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 67308-67321, 2023, doi: 10.1109/ACCESS.2023.3289586.

[10] S. Juraev, A. Ghimire, J. Alikhanov, V. Kakani and H. Kim, "Exploring Human Pose Estimation and the Usage of Synthetic Data for Elderly Fall Detection in Real-World Surveillance," in IEEE Access, vol. 10, pp. 94249-94261, 2022, doi: 10.1109/ACCESS.2022.3203174.

[11] W. Liu, G. Wei, Y. Wang and R. Wu, "Indoor Multipedestrian Multicamera Tracking Based on Fine Spatiotemporal Constraints," in IEEE Internet of Things Journal, vol. 10, no. 11, pp. 10012-10023, 1 June1, 2023, doi: 10.1109/JIOT.2023.3235148.

[12] S. Y. Nikouei, Y. Chen, A. J. Aved and E. Blasch, "I-ViSE: Interactive Video Surveillance as an Edge Service Using Unsupervised Feature Queries," in IEEE Internet of Things Journal, vol. 8, no. 21, pp. 16181-16190, 1 Nov.1, 2021, doi: 10.1109/JIOT.2020.3016825.

[13] M. M. E. Yurtsever and S. Eken, "BabyPose: Real-Time Decoding of Baby's Non-Verbal Communication Using 2D Video-Based Pose Estimation," in IEEE Sensors Journal, vol. 22, no. 14, pp. 13776-13784, 15 July15, 2022, doi: 10.1109/JSEN.2022.3183502.

[14] C. Liang, D. Liu, L. Qi and L. Guan, "Multi-Modal Human Action Recognition With Sub-Action Exploiting and Class-Privacy Preserved Collaborative Representation Learning," in IEEE Access, vol. 8, pp. 39920-39933, 2020, doi: 10.1109/ACCESS.2020.2976496.

[15] C. Gupta et al., "A Real-Time 3-Dimensional Object Detection Based Human Action Recognition Model," in IEEE Open Journal of the Computer Society, vol. 5, pp. 14-26, 2024, doi: 10.1109/OJCS.2023.3334528.

[16] K. Rezaee, M. R. Khosravi and M. S. Anari, "Deep-Transfer-Learning-Based Abnormal Behavior Recognition Using Internet of Drones for Crowded Scenes," in IEEE Internet of Things Magazine, vol. 5, no. 2, pp. 41-44, June 2022, doi: 10.1109/IOTM.001.2100138.

[17] P. Nguyen Huu, N. Nguyen Thi and T. P. Ngoc, "Proposing Posture Recognition System Combining MobilenetV2 and LSTM for Medical Surveillance," in IEEE Access, vol. 10, pp. 1839-1849, 2022, doi: 10.1109/ACCESS.2021.3138778.

[18] J. Khan et al., "SMSH: Secure Surveillance Mechanism on Smart Healthcare IoT System With Probabilistic Image Encryption," in IEEE Access, vol. 8, pp. 15747-15767, 2020, doi: 10.1109/ACCESS.2020.2966656.

[19] A. Manocha and M. Bhatia, "IoT-Inspired Monitoring Framework for Real-Time Stereotypic Movement Analysis," in IEEE Systems Journal,

vol. 16, no. 2, pp. 2788-2796, June 2022, doi: 10.1109/JSYST.2021.3084370.

[20] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," in IEEE Access, vol. 9, pp. 34366-34382, 2021, doi: 10.1109/ACCESS.2021.3059170.

[21] Reddy Panyala, A., & Manickam, B. (2024). Generative adversarial network for Multimodal Contrastive Domain Sharing based on efficient invariant feature-centric growth analysis improved brain tumor classification. Electromagnetic Biology and Medicine, 1–15. https://doi.org/10.1080/15368378.2024.2375266.

[22] Y. Hao et al., "An End-to-End Human Abnormal Behavior Recognition Framework for Crowds With Mentally Disordered Individuals," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 8, pp. 3618-3625, Aug. 2022, doi: 10.1109/JBHI.2021.3122463.

[23] A. -U. Rehman, H. S. Ullah, H. Farooq, M. S. Khan, T. Mahmood and H. O. A. Khan, "Multi-Modal Anomaly Detection by Using Audio and Visual Cues," in IEEE Access, vol. 9, pp. 30587-30603, 2021, doi: 10.1109/ACCESS.2021.3059519.

[24] R. Chatterjee, A. Chatterjee, M. R. Pradhan, B. Acharya and T. Choudhury, "A Deep Learning-Based Efficient Firearms Monitoring Technique for Building Secure Smart Cities," in IEEE Access, vol. 11, pp. 37515-37524, 2023, doi: 10.1109/ACCESS.2023.3266514.

[25] H. Wang, B. Dong, Q. Zhu, Z. Chen and Y. Chen, "Spatio-Temporal Attention Fusion SlowFast for Interrogation Violation Recognition," in IEEE Access, vol. 11, pp. 103801-103813, 2023, doi: 10.1109/ACCESS.2023.3316724.

[26] M. Waheed et al., "An LSTM-Based Approach for Understanding Human Interactions Using Hybrid Feature Descriptors Over Depth Sensors," in IEEE Access, vol. 9, pp. 167434-167446, 2021, doi: 10.1109/ACCESS.2021.3130613.

[27] M. Xu, L. Guo and H. -C. Wu, "Robust Abnormal Human-Posture Recognition Using OpenPose and Multiview Cross-Information," in IEEE Sensors Journal, vol. 23, no. 11, pp. 12370-12379, 1 June1, 2023, doi: 10.1109/JSEN.2023.3267300.