Journal of Information Systems Engineering and Management
2025, 10(15s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimized ST-GCN Model for Suspicious Human Activity Recognition using Edge Computing

[1]Likhith SR, [2]R. Mahalakshmi,

[1] *School of CSE and School of IS,Presidency University, Bengaluru, Karnataka, India.*
[1]*likhith.sr@presidencyuniversity.in*
[2] *School of CSE and School of IS,Presidency University, Bengaluru, Karnataka, India.*
[2] *mahalakshmi@presidencyuniversity.in*

ARTICLE INFO

Received: 01 Dec 2024

Revised: 26 Jan 2025

Accepted: 05 Feb 2025

ABSTRACT

Introduction: In automated video surveillance applications, detecting abnormal human activity is incredibly difficult to classify them. In these systems, labeling of human activities relies on the visual aspects and motion patterns observed in videos. But, a significant portion of many traditional methods, as well as traditional neural models, either disregard or face challenges in leveraging temporal features for predicting Human Action Recognition (HAR).

Objectives: The main objective is identifying different suspicious human activities from videos and achieving precise and effective HAR.

Methods: The automatic detection of aberrant human activity in a surveillance system was resolved by the Deep Learning (DL) based edge computing in our proposed work. The videos are first turned into frames and secondly, the spatio temporal features are retrieved from the key frames using a DL model Spatiotemporal-Graph Convolutional Network (ST-GCN). Finally, to recognize anomalous activity from video, the collected features are loaded into an optimized gated recurrent unit (OGRU). Results: The experimentation is carried out on the two benchmark datasets and achieved better accuracies of 98.1% (Dataset 1) and 98% (Dataset 2) respectively.

Conclusions: This work effectively recognizes all kinds suspicious human activities from vast set of video files.

Keywords: Abnormal Human Activity, Video Surveillance, Spatiotemporal-Graph Convolutional Network, Optimized Gated Recurrent Unit

INTRODUCTION

During the last two decades, computer vision has emerged as a crucial technology with a wide range of applications that replace human intermission. Computer vision can process and analyze digitized images and videos to obtain a high level of understanding and extract features [1]. Additionally, these systems are made to automate a number of tasks performed by the human visual system. Computer vision is used in many interdisciplinary applications like video surveillance, navigation, automatic inspection, object modeling, and process control. One of the major applications of computer vision is video surveillance, which is employed to observe and monitor the activities of humans in public places [2] [3]. In recent times, intelligent video surveillance models have been utilized in place of human supervision for detecting, tracking and obtaining a high level of object understanding [4] [5]. The processing and analysis of video data are emphasized by the necessity for public safety [6][7].

EC (edge computing) is a growing computer model and it overcomes computational sources from the centralized cloud center to the edges. Computations are performed at the edge of EC and it minimizes latency and ensures low bandwidth load, better time response and enhanced data security [8] [9] [10].

Recognizing Human Action Recognition (HAR) involves deciphering a sequence of intricate and diverse sub-actions. The aim of abnormal detection is to collect activities using videos and exploit ML models for extracting

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

essential features [11]. The combination of Deep Learning (DL) and EC will surely provide new insights into addressing existing issues and open up more appealing applications [12]. Various ML approaches for detection have been proposed by numerous researchers because abnormal activity detection has become a ubiquitous problem. Since ML techniques struggle to manage streaming data, this work presents an optimized DL model for extracting the spatial and temporal features of videos [13][14].

The main objectives of this work are:

- To introduce an automated DL based edge computing model for detecting abnormal human activities.
- To introduce Spatiotemporal- Graph Convolutional Network (ST-GCN) for extracting the spatial and temporal features.
- To introduce an optimized Gated Recurrent Unit (GRU) for recognizing anomalous activity from video.

LITERATURE SURVEY

Related works based on the HAR using different approaches are given in this section:

Kumar et al. [15] developed CNN with LSTM model for abnormal HAR. In this existing work, fuzzy logic was used to extract the frames and the features were extracted using the pre-trained CNN model. At last, the DL model LSTM was used for finding the anomaly behaviour.

Vallathan et al. [16] determined suspicious activities using the DL model in the IoT environment. Frame rate conversion was carried out and the kernel density approach was used for identifying the past terms. At last, the malicious activity was determined by the Random Forest (RF) model.

Vrskova et al. [17] presented the ConvLSTM model for HAR using the UCF crime and AIRTLab datasets. Wazwaz et al. [18] proposed three smartphone accelerometers to detect the HAR in the IoT environment. Three approaches were created, trained, and employed for achieving the required accuracy at the cloud and at the edge of the IoT with a proper response time.

Huang et al. [19] presented HAR model on the basis of the EC based GRU model. This existing work considered EC for optimizing wearable device's processing time and energy consumption. Then, the convolutional model was utilized for pre-processing more training data efficiently.

PROPOSED METHODOLOGY

The suggested abnormal HAR model successfully balances a minimized computational load while preserving accuracy when compared to other literature methods. The framework encompasses an initial module dedicated to spatio-temporal Feature extraction and Abnormal HAR detection. Subsequently, for each identified individual, the HAR model is activated. This process comprises several stages: an input data pre-processing, the utilization of a DL model centered on the ST-GCN for spatio-temporal feature extraction, and the classification of these features through an OGRU.

Selecting key frames

In this work, the pre-processing stage is considered through an automatic typical frame selection from the input frames. Instead of extracting features from the entire dataset, our focus is on isolating key features from the selected typical frames. The videos are first turned into frames and keyframes from a batch of frames are selected for the spatio-temporal feature extraction process.

Feature extraction

For acquiring the complex spatio-temporal features and identifying the HAR, the DL model ST-GCN- OGRU is introduced. Figure 1 presents the proposed ST-GCN- OGRU model and it has two stages like: (a) ST-GCN is introduced to learn the spatio-temporal features and (b) OGRU is introduced to identify HAR which has GRU layer, FC (fully connected) layer and output layer.

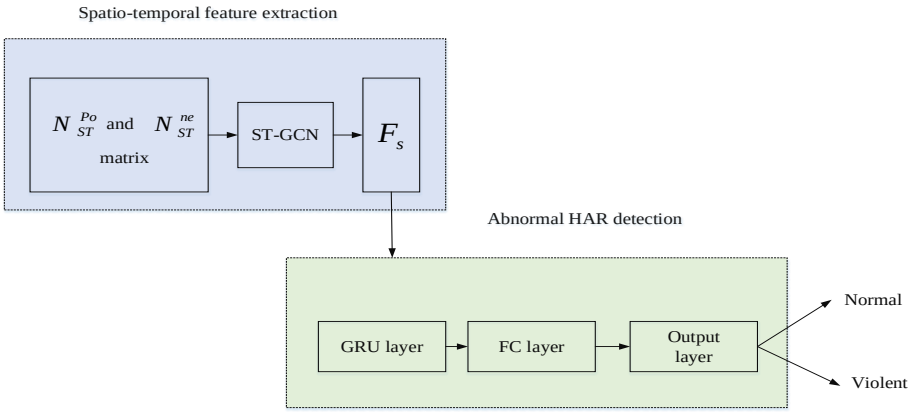


Figure 1: Proposed ST-GCN- OGRU model

ST-GCN: In this stage, ST-GCN is presented for capturing spatio-temporal features with respect to matrix relationship. The ST-GCN has dual GCN, one GCN is utilized for learning positive N_{ST}^{Po} and another one GCN is utilized for learning negative features N_{ST}^{ne}. In this work, initially positive and negative matrices are constructed and then features are obtained from N_{ST}^{Po} and N_{ST}^{ne}. If the self correlative values are all equal to 1, proceed to normalize the N_{ST}. Then, the value of N_{ST}^{Po} and N_{ST}^{ne} with respect to the self correlative values given as:

$$N_{ST}^{Po} = \max\left\{0, N_{ST}\right\} \quad (1)$$

$$N_{ST}^{ne} = \max\left\{0, N_{ST}\right\} + Id \quad (2)$$

where Id is the identity. In addition two Laplacian negative and positive matrices with respect to N_{ST}^{Po} an N_{ST}^{ne} are given as:

$$La_{+} = D_{+}^{-1} N_{ST}^{Po} \quad (3)$$

$$La_{-} = D_{-}^{-1} N_{ST}^{ne} \quad (4)$$

where D_{+}^{-1} and D_{-}^{-1} are the diagonals. For La_{+}, the GCN_{+} has blocks like $0,1.2,.....,n.$ every graph node at l^{th} layer is given as $l = 0,1.2,.....,n.$ the activation matrix in the l^{th} layer L_{+}^{l} is given as:

$$L_{+}^{l} = sig\left(La_{+} L_{+}^{l-1} \Theta_{+}^{l-1} W_{+}^{l}\right) \quad (5)$$

where Θ_{+}^{l-1} is the filter matrix variables, W^{l} is the trainable variable, sig is the activation term. The terms like Θ_{+}^{l-1} and W^{l} are considered as one parameter and the Equation (5) is written as:

$$L_{+}^{l} = sig\left(La_{+} L_{+}^{l-1} W_{+}^{l}\right) \quad (6)$$

For the n^{th} output layer, the propagating term is given as:

$$L_{+}^{n} = La_{+} L_{+}^{n-1} W_{+}^{n} \quad (7)$$

where W_{+}^{n} and L_{+}^{n-1} are the weighting variable and positive feature.

Similarly, for La_{-}, the GCN_{-} has blocks like $0,1.2,.....,n.$ every graph node at l^{th} layer is given as $l = 0,1.2,.....,n.$ the activation matrix in the l^{th} layer L_{+}^{l} is given as:

$$L_{-}^{l} = sig\left(La\_L_{-}^{l-1}\Theta_{-}^{l-1}W_{-}^{l}\right) \quad (8)$$

For the n^{th} output layer, the propagating term is given as:

$$L_{-}^{n} = La\_L_{-}^{n-1}W_{-}^{n} \quad (9)$$

where W_{-}^{n} and L_{-}^{n-1} are the weighting variable and positive feature. Finally, for obtaining the spatio-temporal features F_s , L_{+}^{n} and L_{-}^{n} is integrated using the W_{+}^{n} and W_{-}^{n} . The term F_s is given as:

$$F_s = \left[L_{+}^{n}L_{-}^{n}\right] \times W_{+}^{n}W_{-}^{n} \quad (10)$$

Abnormal HAR detection

After extracting the spatio-temporal features F_s, the term F_s is fed to the OGRU to identify the HAR. There are two gates like update gate u and the reset gate r are present in the OGRU as shown in the Figure 3. The u determines the selection of whether the hidden state C_g should be updated using a candidate state $\tilde{C}_g$ and it is given as:

$$u_g = sig\left(W_u[x_g, C_{g-1}]\right) \quad (11)$$

The r is utilized for determining whether the prior C_g is eliminated and it is given as:

$$r_g = sig\left(W_r[x_g, C_{g-1}]\right) \quad (12)$$

The term $\tilde{C}_g$ is given as:

$$\tilde{C}_g = sig(W_C[r_g \times x_g, C_{g-1}]) \quad (13)$$

The term C_g is given as:

$$C_g = (1-u_g) \times C_{g-1} + u_g \times \tilde{C}_g \quad (14)$$

At last, the final outcome of the OGRU is given as:

$$C_{OGRU,t} = sig\left(W_0 C_g\right) \quad (15)$$

where W_r, W_C and W_0 are the weighting matrices.

4. RESULTS & ANALYSIS

In this section, we outline the experimental configuration using the Python platform and the datasets utilized. Then, the suggested ST-GCN- OGRU network is compared over several models. The objective is to showcase the superiority of the suggested ST-GCN- OGRU.

Dataset details

Dataset 1: The UCF Crime dataset [22] is an extensive collection comprising 128 hours of video footage. This dataset encompasses 1900 untrimmed, real-time surveillance videos featuring a diverse range of scenarios. Within these videos, 13 authentic anomalies are documented, which included the instances of Abuse, Vandalism, Accident, Arrest, Assault, Road Burglary, Fighting, Robbery, Stealing, Shoplifting, Shooting, Explosion, and Arson.

Dataset 2: Violence detection AIRTLab dataset [23] comprises 350 video clippings categorized as violent and non-violent designed for training and testing models dedicated to detect violence in videos. Notably, the non-violent clippings are intentionally stored for incorporating characteristics such as exultation, claps, and hugs. Figure 2 defines the samples images of the dataset 1 and 2.

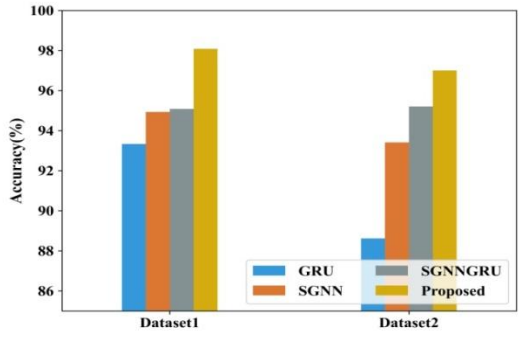
(a)

(b)

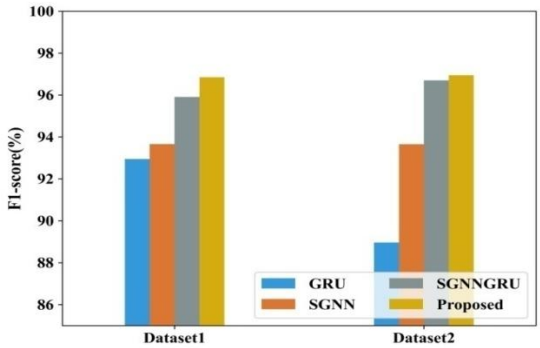
Figure 2: Samples images of the (a) dataset 1 and (b) 2

Comparative analysis

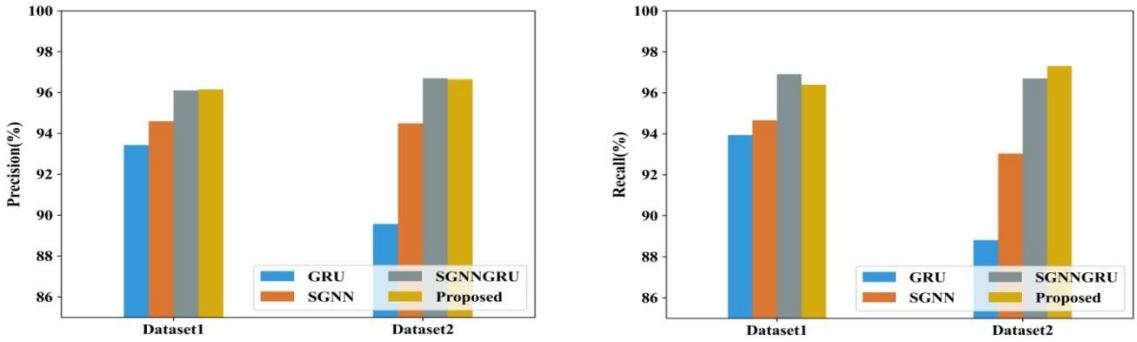
The evaluation of the suggested ST-GCN-OGRU concerning the binary classification problem involves metrics such as accuracy, F1-score, recall, and precision. It is essential to establish the definitions of True Positive U_p, True Negative U_n, False Positive V_p, and False Negative V_n before delving into these concepts. In the context of binary classification, we assume the two classes are labeled as positive and negative. Comparative analysis is made for the approaches like GRU, SGNN (spatio graph neural network), and SGNNGRU are compared with the proposed ST-GCN-OGRU.



(a)



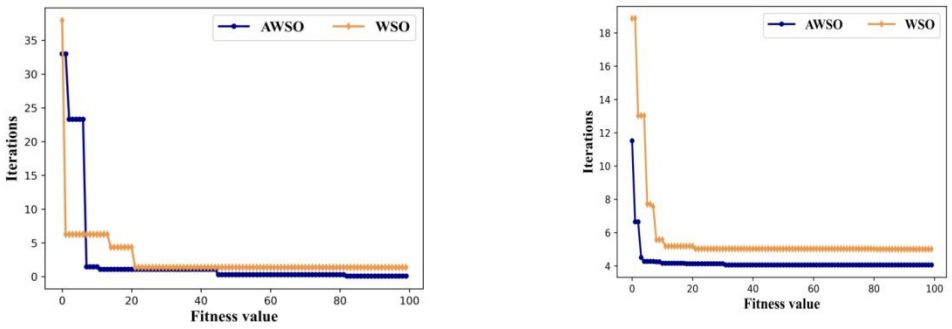
(b)



(c) (d)

Figure 3: Comparative analysis

Figure 3 depicts the comparative analysis with respect to the measures like accuracy, F1-score, recall and precision. It is observed that the proposed ST-GCN-OGRU achieved better accuracies of 98.1% (Dataset 1) and 97.6% (Dataset 2); F1-score of 97.7% (Dataset 1) and 97.8% (Dataset 2); Precision of 97.2% (Dataset 1) and 97.7% (Dataset 2); Recall of 97.1% (Dataset 1) and 98% (Dataset 2); it is observed that the proposed ST-GCN-OGRU outperformed the conventional models.



(a) (b)

Figure 4: Convergence analysis of (a) Dataset 1 and (b) Dataset 2

Figure 4 defines the convergence analysis of the standard WSO and the proposed AWSO for Dataset 1 and Dataset 2. The graph is plotted between the fitness value and iterations. It is observed that the fitness values of the proposed AWSO are better than the standard WSO.

CONCLUSION

The abnormal HAR holds substantial importance within video surveillance, serving as a crucial element for early anomaly detection. This capability improves security across diverse domains such as security surveillance, emergency, sociality and healthcare. This work presents the ST-GCN-OGRU for spatio-temporal feature extraction and HAR detection with respect to EC environment. The performance of the GRU was optimized by the AWSO and it improved the convergence speed. The analysis was carried out on the UCF Crime and AIRTLab datasets. AUC values achieved by the ST-GCN-OGRU are 0.985 (Dataset 1) and 0.981 (Dataset 2) respectively. Future work endeavors could focus on enhancing the suggested ST-GCN-OGRU for recognizing abnormal HAR in video feeds characterized by distortions, commonly observed in video sequences obtained with moving cameras. Additionally, hand-held devices must prioritize ensuring the affinity of the hardware with the processing of real data.

REFERENCES

[1] Patrikar, Devashree R., and Mayur Rajaram Parate. "Anomaly detection using edge computing in video surveillance system." International Journal of Multimedia Information Retrieval 11, no. 2 (2022): 85-110.

[2] Aishwarya, D., and R. I. Minu. "Edge computing based surveillance framework for real time activity recognition." ICT Express 7, no. 2 (2021): 182-186.

[3] Kumar, Manoj, Anoop Kumar Patel, and Mantosh Biswas. "Real-time detection of abnormal human activity using deep learning and temporal attention mechanism in video surveillance." Multimedia Tools and Applications (2023): 1-17.
[4] Kuppusamy, P., and V. C. Bharathi. "Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance–A survey." Measurement: Sensors 24 (2022): 100510.
[5] Lentzas, Athanasios, and Dimitris Vrakas. "Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review." Artificial Intelligence Review 53, no. 3 (2020): 1975-2021.
[6] Sernani, Paolo, Nicola Falcionelli, Selene Tomassini, Paolo Contardo, and Aldo Franco Dragoni. "Deep learning for automatic violence detection: Tests on the AIRTLab dataset." IEEE Access 9 (2021): 160580-160595.
[7] Kim, Siyeon, Sungjoo Hwang, and Seok Hwan Hong. "Identifying shoplifting behaviors and inferring behavior intention based on human action detection and sequence analysis." Advanced Engineering Informatics 50 (2021): 101399.
[8] Kim, Siyeon, Sungjoo Hwang, and Seok Hwan Hong. "Identifying shoplifting behaviors and inferring behavior intention based on human action detection and sequence analysis." Advanced Engineering Informatics 50 (2021): 101399.
[9] Li, Jun, Xianglong Liu, Mingyuan Zhang, and Deqing Wang. "Spatio-temporal deformable 3d convnets with attention for action recognition." Pattern Recognition 98 (2020): 107037.
[10] Shreyas, D. G., S. Raksha, and B. G. Prasad. "Implementation of an anomalous human activity recognition system." SN Computer Science 1 (2020): 1-10.
[11] Liu, Yuchao, Sunan Zhang, Ziyue Li, and Yunpu Zhang. "Abnormal behavior recognition based on key points of human skeleton." IFAC-PapersOnLine 53, no. 5 (2020): 441-445.
[12] Subramanian, R. Raja, and V. Vasudevan. "A deep genetic algorithm for human activity recognition leveraging fog computing frameworks." Journal of Visual Communication and Image Representation 77 (2021): 103132.
[13] Jaouedi, Neziha, Noureddine Boujnah, and Med Salim Bouhlel. "A new hybrid deep learning model for human action recognition." Journal of King Saud University-Computer and Information Sciences 32, no. 4 (2020): 447-453.
[14] Maqsood, Ramna, Usama Ijaz Bajwa, Gulshan Saleem, Rana Hammad Raza, and Muhammad Waqas Anwar. "Anomaly recognition from surveillance videos using 3D convolution neural network." Multimedia Tools and Applications 80, no. 12 (2021): 18693-18716.
[15] Kumar, Manoj, and Mantosh Biswas. "Abnormal human activity detection by convolutional recurrent neural network using fuzzy logic." Multimedia Tools and Applications (2023): 1-17.
[16] Vallathan, G., A. John, Chandrasegar Thirumalai, SenthilKumar Mohan, Gautam Srivastava, and Jerry Chun-Wei Lin. "Suspicious activity detection using deep learning in secure assisted living IoT environments." The Journal of Supercomputing 77 (2021): 3242-3260.
[17] Vrskova, Roberta, Robert Hudec, Patrik Kamencay, and Peter Sykora. "A new approach for abnormal human activities recognition based on ConvLSTM architecture." Sensors 22, no. 8 (2022): 2946.
[18] Wazwaz, Ayman A., Khalid M. Amin, Noura A. Semari, and Tamer F. Ghanem. "Enhancing human activity recognition using features reduction in iot edge and azure cloud." Decision Analytics Journal 8 (2023): 100282.