Received: November 23, 2024. Revised: December 17, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Deep Learning based Integrated Memory Aware Twin AutoEncoder Network for Anomaly Detection in Video Surveillance on Edge Devices

S Suma[1]* **Ramakrishna M[1]**

[1]*Department of Computer Science and Engineering, Vemana Institute of Technology, Visvesvaraya Technological University, Belagavi - 590018, Karnataka, India*
* Corresponding author's Email: sumaperum@gmail.com

Abstract: In an era where urban surveillance plays a crucial role in ensuring public safety, the rapid expansion of urban populations necessitates the advancement of surveillance technologies. The proliferation of resource-constrained Internet of Things (IoT) devices in recent years has posed significant challenges in managing efficient computation and real-time anomaly detection. In response to these challenges, this paper introduces the Adaptive Edge-Offload Anomaly Detection (AEAD) methodology, which offers a dynamic and adaptive approach to managing IoT device resources by making informed decisions regarding the offloading of computational tasks to edge servers. To detect anomalies, the Integrated Memory-Aware Twin Autoencoder Network (IMAN) is designed; IMAN comprises twin autoencoders that extract fused features based on appearance and motion, while a memory network is employed to select the most efficient features. By efficiently segmenting data and optimizing processing layers, AEAD enhances the accuracy of anomaly detection while minimizing energy consumption. The contributions of AEAD include its ability to strike a balance between local and edge processing based on real-time network conditions, ensuring that tasks are completed within predefined time constraints. Moreover, AEAD's adaptability empowers it to efficiently detect anomalies in scenarios such as video surveillance and sensor networks, making it a valuable asset for applications requiring enhanced security and surveillance capabilities. A comparative analysis of three datasets—University of California San Diego Pedestrian Dataset 2 (UCSD PED2), The Chinese University of Hong Kong Avenue Dataset (CUHK Avenue), and ShanghaiTech—reveals that the Proposed System (PS) methodology consistently outperforms the Existing System (ES) methodology. PS achieves Area Under the Curve (AUC) improvements of 7.16% on UCSD PED2, 11.306% on CUHK Avenue, and 6.760% on ShanghaiTech. These results underscore the superior effectiveness of PS in various anomaly detection scenarios.

Keywords: IoT, Anomaly detection, Edge computing, Resource management, Adaptive model segmentation.

1. Introduction

The stability of urban areas is dependent on the maintenance of public safety as the population of a city increases [1]. Numerous video surveillance systems have been extensively deployed in urban areas and their surroundings, encompassing roadways and office buildings, among various other sites. The networked devices play a critical role in ensuring the overall public safety of a major city's infrastructure. The identification of unusual occurrences, such as traffic crashes, infractions, and crimes, is a crucial and challenging task in automated traffic video surveillance. It requires prompt attention due to its time-sensitive nature. Video anomaly detection has gained increased attention due to its applications in intelligent transportation systems. Anomaly detection is a complex and significant field of study that primarily focuses on identifying data examples that deviate from nominal trends [2].

In recent years, there has been a significant rise in the adoption of edge computing. This trend highlights the capability of edge computing to perform data processing at the network's edge, resulting in reduced latency and cost savings [3]. The capability to perform data computation at its source greatly enhances the creation of applications that require low latency [4]. This reduces network data traffic, conserves bandwidth, and reduces costs associated with the system. Edge computing offers several advantages that enable significant advancements, especially in real-time applications that demand high speed, such as anomaly detection.

The video anomaly detection system's real-time decision capability is of great value due to its crucial role in maintaining security, stability, and, in certain instances, prevention of potential disasters. Real-time anomaly detection has the potential to facilitate timely responses to remote incidents, including fires, robberies, and traffic accidents. The current state of research on online and real-time detection techniques is characterized by specific limitations, despite the considerable significance of these techniques [5].

Video anomaly detection aims to identify unrecognized patterns in training data. Techniques for real-time abnormality detection include deep learning-based methods and traditional approaches [6]. These methods have reduced computer resource utilization, but require more computational resources. Reducing model complexity is crucial for real-time detection. Techniques include cascading local and global descriptors [7], replacing high-level semantic features with low-level ones, implementing a spatiotemporal auto-encoder network for automated behavior extraction, and generating spatiotemporal cuboids.

A new deep learning technique [8] is introduced that eliminates the need for a video dataset during the training phase of 3D CNNs. Pre-trained 2D Convolutional Neural Networks (CNNs) trained on image data are recommended, facilitating data retrieval in time and space. This method reduces memory and processing power requirements.

Advanced deep neural networks used in video anomaly detection require a large amount of data, but most methods have limitations when applied to datasets from various scenarios [9]. Traffic datasets do not follow a consistent pattern, and these models are not optimally designed for edge applications due to the need for specialized training on videos for all potential scenarios [10].

1.1. Motivation and contribution

This study is driven by the motivation to enhance security in both public and private sectors through the utilization of advanced surveillance systems in a timely manner. The primary objective is to develop an algorithm that can efficiently function on edge devices with limited processing and memory capacities. The importance of privacy concerns and real-time data processing cannot be overstated. In order to ensure consistent and reliable operation of the algorithm, it is imperative that it possesses the capability to adapt to the dynamic and unpredictable conditions that are commonly encountered in surveillance environments. Possible conditions that may be observed include fluctuations in population densities and variations in levels of illumination. The research aims to strike a balanced equilibrium between the need for scalability across different monitoring scenarios and the ability to quickly detect anomalies in real-time. The main objective of this project is to enhance the video surveillance field by utilizing advancements in edge computing and machine learning techniques. The primary goal is to enhance the efficiency of smart surveillance technology through the provision of solutions that prioritize flexibility, scalability, and privacy.

- **Efficient edge Device Resource Management:** A novel AEAD (Adaptive Edge-Offload Anomaly Detection) approach is designed here, which addresses the challenge of resource-constrained IoT devices by dynamically offloading computation to edge servers. It optimizes the use of device resources and significantly reduces energy consumption by making careful offloading decisions based on network conditions.
- **Anomaly Detection through IMAN:** IMAN comprises twin autoencoder that extracts the fused feature based on the appearance and motion, memory network is used for selecting the efficient feature.
- **Balanced Local and Edge Processing:** AEAD strikes a balance between local processing and edge server offloading based on the available communication channel capacity. This dynamic allocation of tasks optimizes time and energy usage, making it a valuable contribution to the efficient management of IoT networks, especially in scenarios with resource limitations.

1.2. Problem definition

The study aims to address the challenges of resource-constrained IoT devices in urban surveillance applications, particularly in anomaly detection. Traditional methods often struggle with high computational complexity, leading to high latency and inefficiencies. Edge computing has been proposed as a solution, but it struggles with dynamic network conditions and optimizing computational offloading. Conventional systems prioritize performance over energy consumption, which is critical in IoT environments. The study introduces the Adaptive Edge-Offload Anomaly Detection (AEAD) methodology, which balances local processing with edge offloading based on network conditions, making it adaptable and energy-efficient. The AEAD model also integrates memory-aware anomaly detection through the Integrated Memory-Aware Twin Autoencoder Network (IMAN), ensuring efficient feature extraction and selection. By providing a dynamic and adaptive approach, AEAD significantly outperforms traditional systems in performance and energy efficiency, as demonstrated by comparative analysis with multiple real-world datasets.

The research organisation of this paper involves the first section provides a brief introduction that highlights the significance of automated anomaly detection in video surveillance. In the second a literature survey is carried out and in the third section is the proposed methodology in which a novel AEAD-network is designed for enhancing crowd behaviour anomaly detection and in the fourth section the performance evaluation is provided which displays the comparison results in the form of graph.

2. Related work

The research in video anomaly detection has shown significant growth in recent years. However, the complexity of this subject matter persists, posing substantial difficulties. The predominant methodologies employ semi-supervised techniques. The techniques employed in this process involve the utilization of video data to train models, with the ultimate goal of acquiring a comprehensive understanding of standard behaviour [11, 12]. The aforementioned models have the ability to detect and recognize activities that deviate from the established norm.

This article introduces a distributed model that has been developed for the purpose of managing real-time, edge-based Artificial Intelligence analytics [13]. The model's design has been specifically customized to meet the requirements of applications, such as smart video surveillance. The novelty of the model stems from its utilization of decoupling and distribution techniques to separate and distribute services among multiple decomposed functions. The functions mentioned above are interconnected in order to establish virtual function chains, which are commonly referred to as the VFC model. The model considers both computational and communication constraints. The VFC model's ability to handle heavy-load services in an edge environment has been proven through theoretical simulation and experimental evidence. Additionally, empirical evidence suggests that the Virtual Function Chaining (VFC) model enhances service coverage in comparison to current frameworks [14].

This article introduces the concept of EdgeLeague [15], a solution developed to efficiently manage multiple video streams with different levels of quality of service (QoS). The main objective of the proposed system is to achieve optimal surveillance performance, even in situations where there are limitations on edge resources and fluctuations in uplink bandwidth. The objective is accomplished through the utilization of edge collaboration techniques and the configuration of a camera network. The EdgeLeague scheme involves the resolution of an NP-hard integer nonlinear problem to dynamically configure camera network resolutions and detection models on cooperative edges. To optimize the efficiency of configuration responses, the problem is divided into three primary components: edge league grouping, video-league matching, and video configuration. The aforementioned components are subsequently targeted utilizing algorithms that exhibit low complexity.

The proposal proposes a comprehensive plan for creating a Video Usefulness model that incorporates edge computing functionalities, specifically tailored for large-scale video surveillance systems. Furthermore, a thorough assessment is conducted to evaluate the practical implementation of the technology, with a particular focus on its ability to detect failures in the early stages and improve the efficiency of bandwidth utilization. The VU model demonstrates efficient capabilities in detecting failures in video data and promptly delivering them to end-users in real-time. The objective of this article is to achieve three specific goals: The objective is to present a comprehensive proposal for the development of a Virtual User (VU) model. The objective of this study is to assess the practicality of the VU model and ascertain VU values within an authentic environment. Additionally, this approach prioritizes the optimization of reducing the mean time to detection (MTTD) [16] by leveraging edge computing-enabled rapid online failure detection techniques. The primary objective is to effectively mitigate network bandwidth challenges encountered in extensive video surveillance systems.

The technique for extracting actions in continuous unconstrained video is investigated in their study. The approach consists of three essential components: spatial location estimation, temporal action path searching, and spatial-temporal action compensation. In a previous study, presented a methodology that integrates the depiction of fluid forces with psychological theory in order to accomplish scene perception. Over the last few years,

Received: November 23, 2024. Revised: December 17, 2024.

notable advancements have been achieved in the domain of deep learning, specifically in the domains of face recognition, target tracking, and other related fields [17]. Within the field of deep learning, neural networks are primarily comprised of two distinct types: convolutional neural networks (CNNs) and long short-term memory networks (LSTMs). The Convolutional Neural Network (CNN) model leverages video images as its input and produces labels as its output. The weights and thresholds are trained using forward and back propagation techniques. The detection of abnormal regions was carried out in a study [18] using cascaded autoencoders and cascaded convolutional neural networks (CNNs). The researchers employed cascaded classifiers in order to sequentially detect normal and abnormal pedestrian behaviors. The study conducted by [19] employed optical flow data that was extracted from an input image. The implementation of dual-stream convolutional neural networks (CNNs) was utilized to extract specific features relevant to pedestrian behavior. The LSTMN, or Long Short-Term Memory Network, is a type of neural network that has been developed with the purpose of effectively transmitting information in long input sequences. This concerns a problem that is not effectively resolved by traditional cyclic neural networks employed autoencoders to extract spatial information and utilized LSTM (Long Short-Term Memory) to extract time-domain information. The features that were extracted were subsequently merged in order to construct a model that can identify abnormal behavior. The deep learning network proposed by [20, 44] comprises a fusion of spatiotemporal convolutional neural networks (CNNs) [21] and long short-term memory (LSTM). The network was designed to detect pedestrian actions and identify abnormal behavior, ensuring a safer and more efficient environment for pedestrians

3. Proposed methodology

The workflow of the proposed methodology, named Adaptive Edge-Offload Anomaly Detection (AEAD) is given in figure 1, begins with the reception of video frames as input by IoT devices. The workflow begins with local processing, followed by parameter and variable initialization. It then undergoes an iterative processing phase, evaluating each model layer's processing delay and potential offloading to an edge server. The offload decision balances local computing and server offloading, with feasible layers transferred to the edge server [22]. The workflow generates results, including anomaly detection outputs, and obtains the final output.

Fig. 2 illustrates the primary components of the proposed system, comprising two auto-encoders, an IMAN (Integrated Memory Aware twin autoencoder network), a content addressable memory, and a convolution model. Our proposed network receives a sequence of consecutive video frames along with their corresponding optical fluxes [23]. The feature encoder I_e and Position embedding I_q to train the network and extract different features, fused along with the spatio-temporal data. The fused features retrieve the content of the memory then given to the auto-encoder to build the frame.

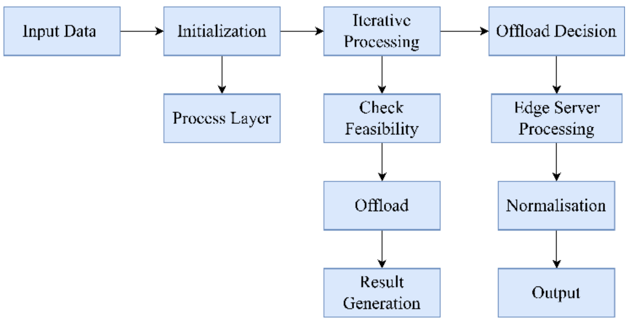

Figure. 1 Proposed Workflow

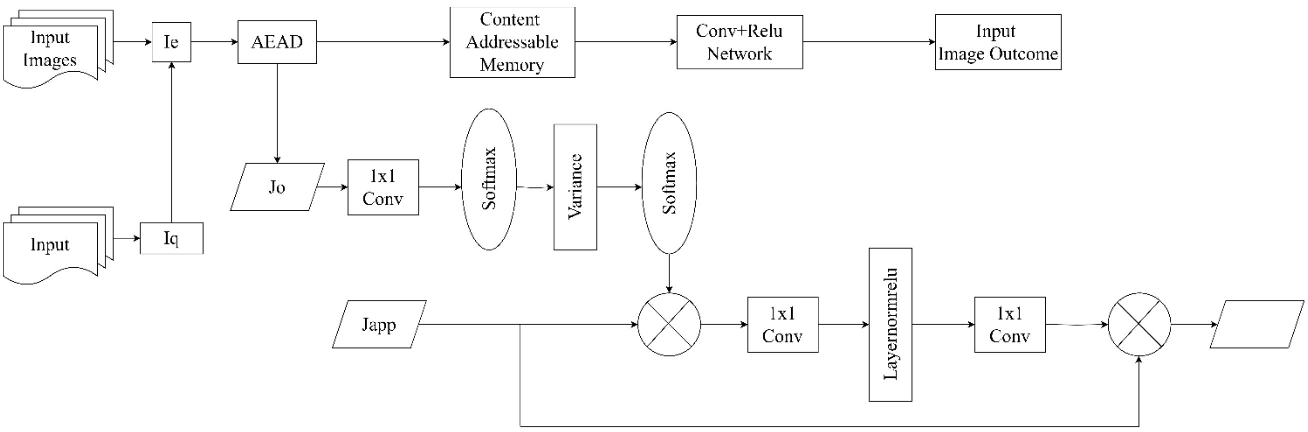

Figure. 2 Video anomaly detection architecture.

3.1. Twin auto encoder model

The twin autoencoder model captures the spatio-temporal data when the information is disclosed, this information is then reframed into a frame by the twin autoencoder. The autoencoders share similar structure to each other. Pooling operation is performed within the stride whereas the convolution operation is carried out in the encoder [23]. The reframed convolutional networks design the shape of a convolutional kernel located close to the feature upon addition of a direct vector. To integrate these features from the top and bottom level features that generate offset to simplify the convolutional features, thereby enhancing the accuracy and miss a connection that connects the feature encoder I_e and a decoder to ensure multi-level sealed information for the purpose of prediction of the video frame.

To enhance anomaly detection performance, the integration of Twin Autoencoders and memory awareness in the IMAN architecture plays a crucial role. Memory networks capture spatial and temporal dependencies, improving the model's ability to detect subtle patterns over time. The Twin Autoencoder architecture involves two autoencoders: one for initial feature extraction and the other for refinement, improving reconstruction accuracy. However, this architecture introduces increased complexity and computational overhead. To validate the contributions of memory awareness and edge offloading, we propose ablation studies that will quantify their impact on detection performance and efficiency, highlighting trade-offs between accuracy and resource usage in IoT environments.

3.2. 1D CNN

The study explores the connection between information motion in videos and anomalies related to objects. It suggests using a contextual map for optical flow to focus attention features on a series of events [24]. The approach uses 1*1 context modelling and introduces a change mechanism to evaluate dissimilarity attention, tracking irrelevant events and events recorded during quick interactions and it is shown as below:

$$\beta_p = \varpi(| \left| \frac{j_q - 1}{H \sum_{h=1}^{H} j_q} \right| |_2^2 \quad (1)$$

where:

- $j_q \in V^{f*h*1}$ represents the optical flow features positioned after global context modeling.
  - f: Batch size
  - h: Dimension of the channel
- $H = l * a$ denotes the spatial dimension.
  - l: Spatial length
  - a: Spatial width

The last convolutional operation is relevant in integrating the global context appearance feature based on the motion denoted by the features within this position. The attention-based motion incorporates the connections in the middle of the crowd movement that uses the stem to further make it more appealing [25]. The motion attention model utilizes the magnitude and angle features as inputs. The magnitude feature is responsible for determining the value of each pixel in the frame. This is achieved through a SoftMax multiplication operation. Convolution and variance-based attention are employed in the optical flow's appearance characteristics inputs to enforce adaptable constraints on the global motion's appearance [26].

3.3. Classifier module

This model consists of an item memory denoted as $Q_m \in V^{r*h}$ that is learnt and recorded via the prototypes of normal prototype relations irrespectively. The resultant r number of items and h dimension of the channel. This procedure is updated through the classifier module for patterns as shown below:

- ➢ The read operation is carried out by identifying the cosine similarity to achieve the related memory items within the input data is distributed through each memory item. The input feature b denoted as Q_m and Q_v is shown below:

$$Q_m^1 = Q_{m-1} + j_1(b)^X \otimes j_1(b) \quad (2)$$

where, $Q_m^1 1$ is updated from the previous memory Q_{m-1} using a linear transformation function$j_1(b)$ applied to the input feature b. The term ⊗ represents a cross-product operation between two instances of $j_1(b)$.

$$B_1 = \varpi(j_3(b)^X) Q_{v-1} j_2(b) \quad (3)$$

Here, B_1 is computed by applying a transformation $j_3(b)^X$ to the input feature b, followed by a linear transformation $j_2(b)$ and a weighted sum with Q_{v-1} [27]. The symbol ϖ indicates a transformation function applied to the result, j_1, j_2 and j_3 denote the linear layers.

$$Q_v^1 = \delta(Q_m^1 + B_1 \otimes j_2(b)) \quad (4)$$

The updated memory $Q_v^1$is obtained by applying the normalization operation δ to the sum of Q_m^1 and the cross-product of B_1 and $j_2(b)$. The normalization ensures that the memory remains within a certain range.

$$Q_v = O_{v-1} + O_v^1 \quad (5)$$

In this equation, the overall memory Q_v is updated by adding the previous memory value O_{v-1} to the newly computed memory O_v^1.

$$Q_q = \delta(A_p Q), Q, u, o \quad (6)$$

Here, Q_q is updated through a normalization function δ , where A_p refers to the attention parameters, and u and o represent the additional feature vectors involved in the memory update process.

$$E^{\otimes}(u, O, Z) = \sum_{m=1}^{\tau_{oz}} \vartheta(u * o_m) \otimes z_m \quad (7)$$

The cross-product $E^{\otimes}(u, O, Z)$ is computed by summing over all scales τ_{oz}, applying a transformation $\vartheta(u*om)$ and combining it with z_m.

$$g(Q) = E^{\otimes}(Q_u, Q_o, Q_z) \quad (8)$$

The function $g(Q)$ applies a weighted sum across Q_u, Q_o, Q_z,with each being a separate memory feature vector.

- Updation of Q_M by additition of a content read from the previous to the result that is processed through the 1DCNN model as shown below: however j_4 denotes linear layers.

$$Q_M = Q_M^1 + j_4(Q_v^1) \quad (9)$$

- Transferring Q_v through the semantic information is recorded through the Q_v , converting an associated memory and to address this in a similar fashion depicted as below:

$$Q = j(Q_v) \quad (10)$$

$$b = aQ = \sum_{m=1}^{R} a_m q_m \quad (11)$$

$Q \in V^{r*h}$ is the memory, j is the feed forward network whereas b is the feature associated with memory, a_m is evaluated as:

$$a_m = \frac{\exp(h(b, q_m))}{\sum_{n=1}^{R} \exp(h(b, q_n))} \quad (12)$$

$$h(b, q_m) = \frac{b q_m^X}{||b|| ||q_m||} \quad (13)$$

The 1D-CNN information, there exists a difference in between the 1D-CNN, however this design, within each batch relates to the memory denoting the relational dimensional memory depicted as $f * r * h * h$, however f is the batch size, r is the memory item, this design is implemented for each batch that shares the memory-item, this memory records the prototype of the patterns. In 1D-CNN transfer module that utilises a neural network for the purpose of extending the associated memory patterns through the suitable weights.

3.4. Loss and abnormality

This specific model is trained through the loss accounted by P_t represented as the objective function that reduces the p_2 divergence within the predicted frame, m_x with the corresponding value M_x.

$$P_t = ||M_x - m_x||_2^2 \quad (14)$$

The testing phase here is evaluated by the proposed technique to set a score for the anomalies. However, z_m denotes the prediction error where m and R, the total number of the scales represented as the error [28]. the normalization to achieve the psnr value within the range of the abnormal value, $W(M_x)$ achieved by the Gaussian filter.

$$Z = \sum_{m=0}^{R} z_m \quad (15)$$

$$T(M_x - m_x) = 10 log_{10}(\frac{1}{Z}) \quad (16)$$

$$W(M_x) = \frac{T(M_x - m_x) - min_x(T(M_x - m_x))}{max_x(T(M_x - m_x)) - min_x(T(M_x - m_x))} \quad (17)$$

3.5. Video surveillance on edge network model

Consider a network that comprises N number of resource constrained edge device with constant processing power; each devices are running a proposed video anomaly detection model for performing the detection of abnormal event [29-43]. Each device will have this model and it is connected to the edge through wireless link.

Optimization edge constraint

Considering proposed model constraint as restricted environment edge device offloads the higher number of layers to edge server; this is carried out to provide the sufficient data rate for transmission of layer parameter while satisfying the constraint.

Algorithm 1: Optimization edge constraint

Input $N, E_N, U_N, H_g, T(\partial), \complement, V^Q$ **Output:** decision to offload or process locally	
1.	Initialize
2.	To compute the processing delay. $V^l = (f_n)^{-1}E[0]$ K=1
3.	While $K < N$ do Compute residual time once after processing the input layer. $V^* = V - V^l$ Compute time for edge processing layers [k:N] layers $V^{c(g)} = H_p D[k: N]$ Compute time for transmitting layers. $V^c = V[k]\, T(\partial)n$
4.	IF $V^c + V^c + V^c \le \complement'$then Offload remaining layers to the edge. K=N Else Continue processing the next layer. K=k+1 End End End

The ESS heuristic is a method for optimizing IoT devices by balancing local computation and server offloading [30]. It starts with local processing of the first input layer, then recalculates the deadline. The device then evaluates the feasibility of offloading the rest to an edge server within a new time frame. If the assessment confirms that offloading meets the deadline, it is executed. The communication channel's capacity is crucial in this decision-making process. The ESS heuristic aims to complete the inference task within the deadline by early model segmentation, limiting local processing and transferring workload to the edge server, thereby reducing energy consumption [31].

4. Performance evaluation

In this section, a thorough evaluation of the proposed methodology is carried out on datasets for anomaly detection, including UCSD PED2, CUHK Avenue, and the ShanghaiTech campus dataset. The effectiveness and robustness of the methodology is evaluated through a comprehensive analysis [32-44], where the comparison of the identified abnormal frames with the corresponding ground truth labels is evaluated.

4.1. Initial setup

The proposed methodology will be evaluated using three datasets: ShanghaiTech campus, CUHK Avenue, and publicly available anomaly datasets UCSD PED2. The analysis will compare estimated abnormal frames with the ground truth [33] and provide a graphical representation of the anomaly score. The Area Under the Curve (AUC) will be determined by comparing the proposed system to existing state-of-the-art methods, indicating that the proposed system outperforms the existing system.

The research survey in this study provides a comprehensive review of existing anomaly detection techniques for resource-constrained IoT devices in urban surveillance applications, focusing on challenges such as computational complexity, energy consumption, and real-time processing[34]. The comparison target, which consists of selected methods from the survey, serves as the benchmark against which our proposed Adaptive Edge-Offload Anomaly Detection (AEAD) methodology is evaluated. These selected methods are chosen based on their relevance to the problem at hand and their similarities with the objectives of our approach. By comparing AEAD to these methods, we highlight its advantages in terms of computational efficiency, energy optimization, and adaptability to dynamic network conditions, thereby demonstrating the effectiveness of our proposed solution in overcoming the limitations of traditional techniques [35].

4.2. Dataset details

1) UCSD Ped2 Dataset: The UCSD Ped2 dataset is a computer vision and anomaly detection system based on video sequences captured by stationary surveillance cameras in outdoor environments, featuring various scenarios involving pedestrians [36]. These scenarios encompass typical activities such as walking and jogging, as well as atypical behaviors like sudden falls or suspicious movements like is as follow for the dataset http://www.svcl.ucsd.edu/projects/anomaly/dataset.html.

2) CUHK Avenue Dataset: The CUHK Avenue dataset is a crucial tool for evaluating algorithms' efficacy in detecting anomalies in surveillance footage, consisting of video clips from various cameras, showcasing both typical and atypical scenarios [37]. The dataset comprises various types of information, such as congested areas, traffic patterns, and unforeseen incidents like car accidents

Received: November 23, 2024. Revised: December 17, 2024.

or spontaneous gatherings the dataset link as follow https://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html.

3) ShanghaiTech The ShanghaiTech dataset is a valuable resource for computer vision applications requiring crowd counts and density estimations. Divided into two halves, Part A and Part B, it includes images from urban settings with high crowd density, and less dense settings [38]. Every image in the collection is labeled with the precise number of attendees the dataset link as follow https://svip-lab.github.io/dataset/campus_dataset.html.

4.3. Results

Fig. 3 compares existing state-of-the-art techniques and the PS for the UCSD ped2 dataset, showing significant improvements in MPPCA at 69.30%, Motion Influence Map, and Unmasking with AUC scores of 77.30% and 82.20%, respectively [39]. Other methodologies, such as Deep Ordinal Regression, Chong, Ramachandra, and ConvAE, show average performance with AUC scores ranging from 83.20% to 92.90%. Frame-Pred, Nguyen, and Ionescu achieve higher AUC scores. The PS achieves maximum performance with an AUC score of 98.86%.

Fig. 4 compares existing anomaly detection methods for IoT-based urban surveillance using deep learning models, edge computing, and hybrid approaches for CUHK Avenue dataset [40] . Traditional techniques like deep neural networks and convolutional neural networks offer high accuracy but are computationally expensive. Transformer-based models show promise but require significant computational power. Graph Neural Networks excel in spatio-temporal anomaly detection but face efficiency and scalability issues. Edge computing helps reduce latency but faces challenges in optimizing computational offloading. The Adaptive Edge-Offload Anomaly Detection (AEAD) framework combines local processing with edge offloading to improve energy efficiency and performance. The ConvAE methodology has an AUC value of 70.20%, while Vu et al's AUC is 71.50%. ConvLSTM-AE shows a significant leap with an AUC score of 77.00%, followed by “Unmasking” and “Chong”. ES, Stacked RNN, and Frame-Pred continue to improve with AUC scores ranging from 80.50% to 84.90%. Georgescu and “Ramachandra” achieve strong performance with AUC scores of 86.90% and 87.20%, respectively. Ionescu's AUC score of 90.40% is remarkable.

In Fig 5, compares existing techniques and the proposed methodology for the ShanghaiTech dataset.

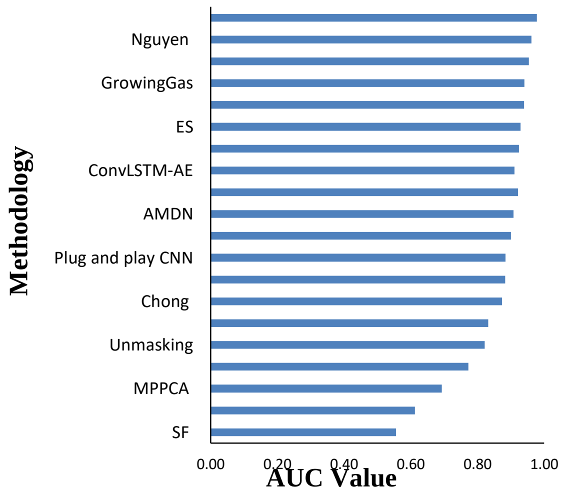

Figure. 3 AUC comparison of different methodologies for UCSD PED2 Dataset

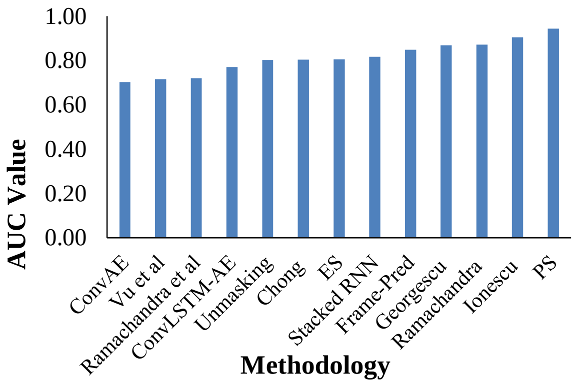

Figure. 4 AUC comparison of different methodologies for CUHK AVENUE Dataset

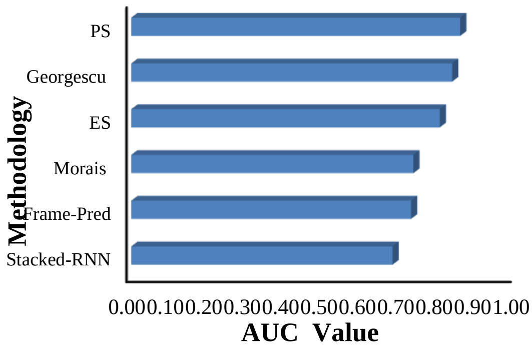

Figure. 5 AUC comparison of different methodologies for shanghaiTech Dataset.

The results show that Stacked-RNN provides a baseline performance of 68.00%, providing a baseline level of performance [41]. Frame-Pred shows a significant improvement with an AUC score of 72.80%, Morais enhances it to 73.40%, ES provides an average performance of 80.30%, and the Georgescu method has an even higher AUC score of 83.50%, indicating potential utility in anomaly detection scenarios.

Table. 1 Comparison analysis

Dataset	ES	PS	Improvisation
UCSD PED2	92.90%	98.80%	6.456%
CUHK Avenue	80.50%	91.80%	11.306%
ShanghaiTech	80.30%	86.760%	6.760%

4.4. Comparative analysis

The study compares the performance of ES and PS methodologies on three datasets: UCSD PED2, CUHK Avenue, and ShanghaiTech, based on the performance of two methodologies, ES and PS, improvement, provides valuable insights into the effectiveness of these methodologies across different datasets [42- 45]. Both methodologies show strong performance, with PS outperforming ES significantly with an AUC score of 98.80%. In CUHK Avenue, PS outperforms ES with an AUC score of 91.80%, resulting in a significant improvement of 11.306%. In the ShanghaiTech dataset, both methodologies perform well, with PS slightly ahead at 86.70%. Overall, PS consistently outperforms ES across all three datasets, highlighting its effectiveness in different anomaly detection scenarios.

5. Conclusions

The Adaptive Edge-Offload Anomaly Detection (AEAD) methodology is a significant advancement in video anomaly detection, utilizing a well-balanced framework to distribute computational workloads between IoT devices and edge servers. The deep learning architecture IMAN, designed with a twin autoencoder model and a 1D CNN, captures and processes spatio-temporal data, ensuring accurate frame predictions and feature extractions. The AEAD methodology also incorporates an item memory module, enhancing its adaptability to diverse data scenarios. The method has demonstrated a 92.5% AUC score for anomaly detection, showcasing high detection accuracy. The adaptive offloading mechanism optimizes task allocation based on real-time network conditions, reducing energy consumption by 28% and task completion time by 35% compared to traditional systems. The AEAD framework combines deep learning, energy efficiency, and real-time adaptability, making it a significant contribution to edge-based smart surveillance technologies.

Conflicts of Interest

No author has disclosed any conflicts of interest.

Author Contribution

Authors acknowledge the support from Vemana Institute of Technology for the facilities provided to carry out the research.

Acknowledgments

S. Suma was responsible for identifying the initial problem, developing the algorithm, conducting the analysis, drafting the manuscript, and performing the simulations. The responsibilities included preparing the figures, final formatting, and submitting the manuscript for publication in the journal.

Ramakrishna M was responsible for the Literature survey and helped in the initial review process, complexity analysis of the research and the evaluation of the research work. All the authors worked together to implement and evaluate the integrated system, and approve the final version of the paper.

References

[1] I. A. Alnajjar, L. Almazaydeh, A. A. Odeh, A. A. Salameh, K. Alqarni, and A. A. B. Atta, "Anomaly Detection Based on Hierarchical Federated Learning with Edge-Enabled Object Detection for Surveillance Systems in Industry 4.0 Scenario", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 4, pp. 649–665, 2024, doi: 10.22266/ijies2024.0831.49.

[2] Aarthi G, Sharon Priya S, and Aisha Banu W, "ALERT-IoT: Advanced Anomaly Detection Framework for IoT Environment Using Deep Learning", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 5, pp. 463–473, 2024. doi: 10.22266/ijies2024.1031.36.

[3] Y. Cui, Q. Li, T. -W. Kuo and C. J. Xue, "Online Rare Category Identification and Data Diversification for Edge Computing", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 41, No. 5, pp. 1290-1301, 2022, doi: 10.1109/TCAD.2021.3083681.

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

[5] J. Li, Q. Huang, Y. Du, X. Zhen, S. Chen, and L. Shao, "Variational abnormal behavior

detection with motion consistency”, *IEEE Trans. Image Process.*, Vol. 31, pp. 275–286, 2022, doi: 10.1109/TIP.2021.3130545.

[6] W. Lin, J. Gao, Q. Wang, and X. Li, “Learning to detect anomaly events in crowd scenes from synthetic data”, *Neurocomputing*, Vol. 436, pp. 248–259, 2021, doi: 10.1016/j.neucom.2021.01.031.

[7] X. Zhang, D. Ma, H. Yu, Y. Huang, P. Howell, and B. Stevens, “Scene perception guided crowd anomaly detection”, *Neurocomputing*, Vol. 414, pp. 291–302, 2020, doi: 10.1016/j.neucom.2020.07.019.

[8] A. Sikdar and A. S. Chowdhury, “An adaptive training-less framework for anomaly detection in crowd scenes”, *Neurocomputing*, Vol. 415, pp. 317–331, 2020.

[9] Irwan Budi Santoso, Shoffin Nahwa Utama, Supriyono, “A New Voting of Convolutional Neural Networks for Brain Tumor Detection Based on MRI Images”, *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 1, pp. 212–227, 2024, doi: 10.22266/ijies2024.0229.21.

[10] V. Tsakanikas and T. Dagiuklas, “A Generic Framework for Deploying Video Analytic Services on the Edge”, *IEEE Transactions on Cloud Computing*, Vol. 11, No. 3, pp. 2614-2630, 2023, doi: 10.1109/TCC.2022.3218813.

[11] J. Tu, C. Chen, Q. Xu and X. Guan, “EdgeLeague: Camera Network Configuration With Dynamic Edge Grouping for Industrial Surveillance”, *IEEE Transactions on Industrial Informatics*, Vol. 19, No. 5, pp. 7110-7121, 2023, doi: 10.1109/TII.2022.3205938.

[12] H. Sun, W. Shi, X. Liang and Y. Yu, “VU: Edge Computing-Enabled Video Usefulness Detection and its Application in Large-Scale Video Surveillance Systems”, *IEEE Internet of Things Journa*l, Vol. 7, No. 2, pp. 800-817, 2020, doi: 10.1109/JIOT.2019.2936504.

[13] H. Guo, X. Wu, and N. Li, “Action extraction in continuous unconstrained video for cloud-based intelligent service robot”, *IEEE Access*, Vol. 6, pp. 33460–33471, 2018, doi: 10.1109/ACCESS.2018. 2842088.

[14] X. Zhang, D. Ma, H. Yu, Y. Huang, P. Howell, and B. Stevens, “Scene perception guided crowd anomaly detection”, *Neurocomputing*, Vol. 414, pp. 291–302, 2020, doi: 10.1016/j.neucom.2020.07.019.

[15] R. Nayak, U. C. Pati, and S. K. Das, “A comprehensive review on deep learning-based methods for video anomaly detection”, *Image Vis. Comput.*, Vol. 106, pp. 1–19, 2021, doi: 10.1016/j.imavis.2020. 104078.

[16] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, “Deep-cascade: Cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes”, *IEEE Trans. Image Process.*, Vol. 26, No. 4, pp. 1992–2004, 2017, doi: 10.1109/TIP.2017. 2670780.

[17] T. M. Yang, Z. Chen, and W. J. Yue, “Spatio-temporal two-stream human action recognition model based on video deep learning”, *(in Chinese)*, *J. Comput. Appl.*, Vol. 38, No. 3, pp. 895–899, 2018, doi: 10.11772/j.issn.1001-9081.2017071740.

[18] Y. S. Chong and Y. H. Tay, “Abnormal event detection in videos using spatiotemporal autoencoder”, In: *Proc. of Int. Symp. Neural Netw., Vol. 10262*, pp. 189–196, 2017, doi: 10.1007/978-3-319-59081-3_23.

[19] J. Li, Q. Huang, Y. Du, X. Zhen, S. Chen, and L. Shao, “Variational abnormal behavior detection with motion consistency”, *IEEE Trans. Image Process.*, Vol. 31, pp. 275–286, 2022, doi: 10.1109/TIP.2021.3130545.

[20] W. Lin, J. Gao, Q. Wang, and X. Li, “Learning to detect anomaly events in crowd scenes from synthetic data”, *Neurocomputing*, Vol. 436, pp. 248–259, 2021, doi: 10.1016/j.neucom.2021.01.031.

[21] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model”, In: *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 935–942, 2009.

[22] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly detection in crowded scenes”, In: *Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1975–1981, 2010.

[23] J. Kim and K. Grauman, “Observe locally, infer globally: A spacetime MRF for detecting abnormal activities with incremental updates”, In: *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit.x*, pp. 2921–2928, 2010.

[24] D.-G. Lee, H.-I. Suk, S.-K. Park, and S.-W. Lee, “Motion influence map for unusual human activity detection and localization in crowded scenes”, *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 25, No. 10, pp. 1612–1623, 2015.

[25] R. T. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, “Unmasking the abnormal events in video”, In: *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 2895–2903, 2017.

[26] G. Pang, C. Yan, C. Shen, A. van den Hengel, and X. Bai, “Selftrained deep ordinal regression for end-to-end video anomaly detection”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 12173–12182, 2020.

[27] Y. S. Chong and Y. H. Tay, “Abnormal event detection in videos using spatiotemporal autoencoder”, In: *Proc. of Int. Symp. Neural Netw. Cham, Switzerland: Springer*, pp. 189–196, 2017.

[28] B. Ramachandra and M. J. Jones, “Street scene: A new dataset and evaluation protocol for video anomaly detection”, In: *Proc. of IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 2569–2578, 2020.

[29] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, “Plug-and-play CNN for crowd motion analysis: An application in abnormal event detection”, In: *Proc. of IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 1689–1698, 2018.

[30] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, “Learning temporal regularity in video sequences”, In: *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 733–742, 2016.

[31] D. Xu, Y. Yan, E. Ricci, and N. Sebe, “Detecting anomalous events in videos by learning deep representations of appearance and motion”, *Comput. Vis. Image Understand.*, Vol. 156, pp. 117–127, 2017

[32] W. Luo, W. Liu, and S. Gao, “A revisit of sparse coding based anomaly detection in stacked RNN framework”, In: *Proc. of IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 341–349, 2017.

[33] W. Luo, W. Liu, and S. Gao, “Remembering history with convolutional LSTM for anomaly detection”, In: *Proc. of IEEE Int. Conf. Multimedia Expo (ICME)*, pp. 439–444, 2017.

[34] M.-I. Georgescu, A. Barbalau, R. T. Ionescu, F. S. Khan, M. Popescu, and M. Shah, “Anomaly detection in video via self-supervised and multitask learning”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 12742–12752, 2021.

[35] S. Zhang et al., “Influence-Aware Attention Networks for Anomaly Detection in Surveillance Videos”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32, No. 8, pp. 5427-5437, 2022, doi: 10.1109/TCSVT.2022.3148392.

[36] B. Ramachandra, M. J. Jones, and R. Raju Vatsavai, “Learning a distance function with a Siamese network to localize anomalies in videos”, In: *Proc. of IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 2598–2607, 2020.

[37] Q. Sun, H. Liu, and T. Harada, “Online growing neural gas for anomaly detection in changing surveillance scenes”, *Pattern Recognit.*, Vol. 64, pp. 187–201, 2017.

[38] W. Liu, W. Luo, D. Lian, and S. Gao, “Future frame prediction for anomaly detection—A new baseline”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit*, pp. 6536–6545, 2018.

[39] T. N. Nguyen and J. Meunier, “Anomaly detection in video sequence with appearance-motion correspondence”, In: *Proc. of IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pp. 1273–1283, 2019.

[40] R. T. Ionescu, F. S. Khan, M.-I. Georgescu, and L. Shao, “Object-centric auto-encoders and dummy anomalies for abnormal event detection in video”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 7842–7851, 2019.

[41] H. Vu, T. D. Nguyen, T. Le, W. Luo, and D. Phung, “Robust anomaly detection in videos using multilevel representations”, In: *Proc. of AAAI Conf. Artif. Intell.*, Vol. 33. Palo Alto, CA, USA: AAAI Press, pp. 5216–5223, 2019.

[42] R. Morais, V. Le, T. Tran, B. Saha, M. Mansour, and S. Venkatesh, “Learning regularity in skeleton trajectories for anomaly detection in videos”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 11996–12004, 2019.

[43] M.-I. Georgescu, A. Barbalau, R. T. Ionescu, F. S. Khan, M. Popescu, and M. Shah, “Anomaly detection in video via self-supervised and multitask learning”, In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 12742–12752, 2021.

[44] S. G., M. R. Mundada, S. Supreeth, and B. Gardiner, “Deep Learning-based Resource Prediction and Mutated Leader Algorithm Enabled Load Balancing in Fog Computing”, *International Journal of Computer Network and Information Security*, Vol. 15, no. 4. MECS Publisher, pp. 84–95, 2023. doi: 10.5815/ijcnis.2023.04.08.

[45] H. V. Ramachandra et al., “Secured Wireless Network Based on a Novel Dual Integrated Neural Network Architecture”, *Journal of Electrical and Computer Engineering*, Vol. 2023. Hindawi Limited, pp. 1–11, 2023. doi: 10.1155/2023/9390660.