Received: February 27, 2024. Revised: April 7, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Residual Attention based Long-Short Term Memory with Self Gated Rectified Linear Unit for Anomalous Behavior Detection

Kiran Kalla[1]* **Jaya Suma Gogulamanda[2]**

[1] *Department of Artificial Intelligence and Data Science, Ramachandra College of Engineering, Affiliated to JNTUK Kakinada, Eluru, India*
[2] *Department of Information Technology, JNTUGV College of Engineering, Vizianagaram, India*
* Corresponding author's Email: kallakiran1974@gmail.com

Abstract: The detection of anomalous behavior in video surveillance is a focus of present research which has huge value and extensive application probabilities. Due to the difficulty of human movement and the feasibility of environments, anomalous behavior detection has some challenges. This paper proposed a deep learning-based approach for detecting the weapon and anomalous behavior. The Residual Attention-based Long-Short Term Memory (LSTM) with Self Gated Rectified Linear Unit (SGReLU) is proposed to enhance the detection accuracy. The median filter is used for preprocessing which removes noise from the UCF-Crime dataset and feeds into Histogram of Oriented Gradients (HOG) for feature extraction. Then, the Reverse Learning Chimp Optimization Algorithm (RL-COA) is utilized for hyperparameter optimization which attains the individual's reverse solution and then preserves the individual with high fitness values. At last, the Residual Attention-based LSTM with SGReLU is utilized for the classification process. This model overcomes neuron dead issues by allowing negative values for some neurons and minimizing the probability of inactive neurons. The proposed model attained better results on UCF-Crime dataset through the metrics like accuracy, precision, recall, f1-score and AUC values of about 98.84%, 98.62%, 98.47%, 98.35% and 98.21% correspondingly that ensures accurate detection when compared to existing techniques such as ResNet18 with Simple Recurrent Unit (SRU), Residual attention-based LSTM and Convolutional LSTM.

Keywords: Chimp optimization algorithm, Histogram of oriented gradients, Long-short term memory, Residual attention module, Self gated rectified linear unit.

1. Introduction

The detection of anomalous behavior enables the computers to change humans to monitor and observe the video [1]. The surveillance video plays a substantial role in numerous applications like municipal management, safety, and traffic control [2]. Through, the huge number of video data produced through high-growing video equipment each instant has taken huge challenges to abnormal event detection [3]. Consequently, it is enormously significant for progress in abnormal behavior detection which does not rely on manual clarifications and machine learning techniques to find anomalous automatically from video [4]. This technique is efficient in minimizing manpower costs and material sources that enhance the reliability and safety of the monitoring process [5, 6]. In the progressively intricate security system of society, video anomalous behavior detection is crucial in several scenarios [7]. The anomalous detection requires high resources and manpower and it has applications like commercial value and received enhancing attention from industry [8].

The video surveillance system recognizes unusual sights and identifies abnormal activities that create security issues [9]. The identification of anomalous events is referred to as hesitant behavior that requires retail malls, trains, airports and bus stations [10]. The purpose of a circuit camera is to produce a security, prevention, crime examination and reduction of insurance costs [11]. Through, the prevention effects of circuit cameras differ from various periods and crime categories [12]. The numerous amount of areas are measured through a video camera and some factors are essential to the person's conditions like fatigue and attention loss which makes the system ineffective [13]. The Hidden Markov Model (HMM) and its variants are mostly utilized in the detection of anomalous behavior [14]. This method presents the human behavior as a vector and its possibility is utilized as identical among test sequences and behavior patterns [15]. The research contribution is given as follows:

- The median filter is used for pre-processing which removes the noise from UCF-Crime dataset and feeds into HOG for feature extraction.
- The RL-COA is utilized for hyperparameter optimization which attains the individual's reverse solution and then preserves the individual with high fitness values.
- The proposed Residual Attention -based LSTM with SGReLU is utilized for classification which overcomes the neurons' dead issues by allowing negative values for some neurons and minimizing the probability of inactive neurons.

This research is organized as follows: Section 2 describes the literature review. Section 3 explains the proposed methodology. Section 4 elaborates results and discussion and Section 5 provides the conclusion.

2. Literature review

This section describes some of the recent literature work based on weapon and anomalous behavior detection using machine learning and deep learning techniques.

Maryam Qasim and Elena Verdu [16] developed a deep convolutional and recurrent model for video anomaly detection. The Deep Convolutional Neural Network (DCNN) and Simple Recurrent Unit (SRU) are utilized to create an automated approach that finds video anomalies. The ResNet was utilized to take large-scale features from video while the SRU gathered the temporal features. This model outperforms every individual transfer learning and numerous ensemble approaches. However, this model required highly labeled data which consumes more time and labor-intensive tasks.

Waseem Ullah [17] suggested a Residual Attention-based LSTM for effective anomaly recognition. The developed model was a functional in surveillance environment through minimized time complexity. It extracts the CNN features from the video and fed into a developed model that recognizes anomalous in video. The residual attention-based LSTM allows to focus on a relevant region of an input sequence which enhances the model accuracy. However, this model was sensitive to noisy data and tried to identify the optimum hyperplane.

Zheng Xu and Yuanyao Lu [18] introduced a Multi-branch CNN and Gated Recurrent Unit (GRU) for abnormal behavior detection. The multi-branch CNN was utilized to extract the spatial and an encoder was passed to GRU to extract temporal features in multiple videos which provided the output as a decoder. The multi-branch CNN is capable of capturing and influencing the spatial dependences and relationships within medical images. However, it required a large number of labeled training data.

Roberta Vrskova [19] implemented a Convolutional Long Short-Term Memory (ConvLSTM) for abnormal human activity recognition. The video comprised various classes of abnormal activities through ConvLSTM structure. The ConvLSTM contained dual Conv2D and one ConvLSTM layer which have the same kernel size and filters. The ConvLSTM captured long-term dependences in consecutive data and it learns hierarchical features through its convolutional operations. However, it was unable to manage long-term dependences and consecutive information in input data.

Yi Hao [20] developed a Spatio-Temporal Consistency-Enhanced Network (STCEN) for detecting video anomalies. In the developed model, the 2D CNN based decoder and 3D CNN based encoder are major parts. Resampling approach was utilized to latent space vector if the model was trained by actual data which cause poor performance when it considered abnormal data. Furthermore, input clip with produced frame was integrated into reformed video clips that was feed into discriminator part. At the testing phase, the abnormal data produced a variance appearance that impact the model capability. However, STCEN struggled to manage corrupted or noisy spatio-temporal data.

Lin Wang [21] implemented a Double-Flow convolutional LSTM with Variational Autoencoder (DF-ConvLSTM-VAE) for detecting video anomalies. The implemented model was a probabilistic distribution of actual video and it was used to construct videos without anomalous objects. The DF-ConvLSTM-VAE was adoptable for asymmetric structure and enhance the network structure width for attain huge training accuracy. However, the size of foreground target was less that affect the target feature extraction.

G. Rajasekaran and J. Raja Sekar [22] introduced an Optimized Pyramidal Lucas-Kanade Technique with CNN (OPLKT-CNN) for abnormal behavior detection. Initially, the optical flow angle changes of present and past frames are distinguished which

Received: February 27, 2024. Revised: April 7, 2024.

result in noisy and angle variations. Then, the MIIs are created through multiplying angle variations with present frames of optical flow magnitudes then, it was applied into input as 3layer CNN for detecting abnormal behavior. However, OPLKT-CNN prone overfitting issues which reduces the model accuracy. The existing techniques has a limitation such as required highly labeled data which consumes more time and labor-intensive tasks. It was sensitive to noisy data and tried to identify the optimum hyperplane. It required a large number of labeled training data and struggled to manage corrupted or noisy spatio-temporal data. it was not adopted to handle the long-term dependencies and sequential information in input data. The size of foreground target was less that affect the target feature extraction and prone overfitting issues which reduces the model accuracy.

3. Proposed methodology

In this section, the Residual Attention -based LSTM with SGReLU is proposed for early and accurate detection of weapon and anomalous behavior. The UCF-Crime dataset is used in this research that comprises 14 classes and 1900 videos. The preprocessing is done by using a median filter which removes the noise from the UCF-Crime dataset. The Histogram of Oriented Gradients (HOG) is utilized for feature extraction and Reverse Learning based Chimp Optimization Algorithm (RL-COA) is utilized for hyperparameter optimization. At last, the Residual Attention-based LSTM with SGReLU is utilized for classification. It overcomes the neurons' dead issues by allowing negative values for some neurons and minimizing the probability of inactive neurons. Fig. 1 presents a work flow of the proposed methodology.

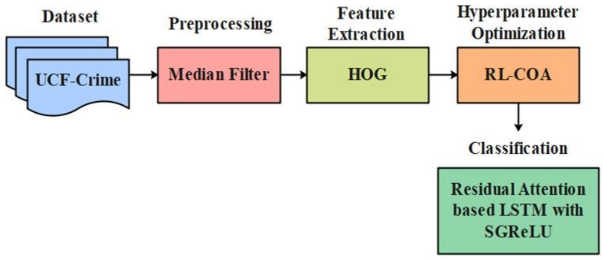

Figure. 1 Work flow the proposed methodology

3.1 Dataset

The dataset used in this analysis is a UCF-Crime dataset [23] which comprises 14 classes and 1900 videos, a total of 128 hours, 7274 average frame rate, MP4 format, and a frame rate is 30fpt. The size of the video is 240×320, which is near to 250p. The dataset includes 800 normal and 810 anomalous video events in training. The rest videos contain 150 normal and 140 anomalous video events in a testing set temporally. Fig. 2 presents the sample images of the UCF-Crime dataset.

3.2 Preprocessing

The UCF-Crime dataset is preprocessed by using a median filter which removes noise from the UCF-Crime dataset. It processed by shifting pixels in the image, adjusting each score by median value of neighbor pixel. It eliminates noise without fading image sharpness which is signified in Eq. (1).

$$\hat{f}(x,y) = median\{g(s,t)\},\ where\ (s,t) \in S_{xy} \quad (1)$$

Here, S_{xy} is a coordinates group in rectangular image window which comprises center at (x, y). The $\hat{f}(x,y)$ is a restored image, $g(s,t)$ is a corrupted and measured area in S_{xy} , s and t are coordinate variables in S_{xy} at point (x, y).

3.3 Feature extraction

The preprocessed features are extracted by utilizing a Histogram of Oriented Gradients (HOG) Which totals the gradient orientation incidences in local spatial region of image that is known as cell. Through extracting HOF features, the image gradients are estimated by forming histogram gradients in each cell. The attained histograms are normalized and provides HOG descriptor of certain blocks. Primarily, frames are transformed into grayscale which is followed through gradient estimation by converting image into vertical and horizontal masks [24]. Each pixel orientation θ is calculated through ratio of gradients in vertical and horizontal directions which is given in Eq. (2).

$$\theta(x,y) = arctan\ \frac{G_x(x,y,z)}{G_y(x,y,z)} \quad (2)$$

Where, $G_x(x,y,z)$ and $G_y(x,y,z)$ represent the gradients of vertical and horizontal masks, respectively. Each block is separated into $M \times N$ cells, where $M \leq N$. For each cell, pixel is estimated by weight that are magnitude gradients at each pixel and votes are gathered in orientation. It is a gradient magnitude at pixels (x, y, z). The HOG features are extracted from each cell as produced features which is equivalent to bins and features are normalized and provided in Eq. (3).

Received: February 27, 2024. Revised: April 7, 2024.

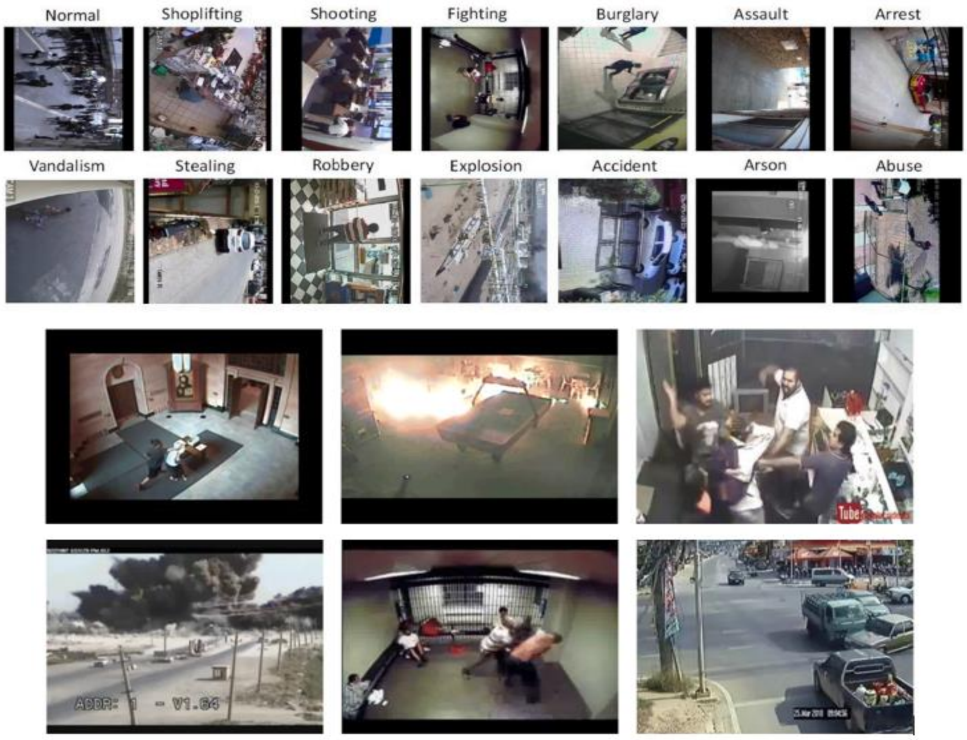

Figure. 2 Dataset sample images

$$N_f = \frac{(\Sigma_{(x,y,z)\epsilon Block} G(x,y,z)+\epsilon)}{(\Sigma_{(x,y,z)\epsilon Block} G(x,y,z)+\epsilon)} \quad (3)$$

The $G(x, y, z)$ is a magnitude of gradients at the pixels (x, y, z),N_f is a normalized histogram features, in which each cell from dimension vector are equivalent to oriented bins number and entire cell in block which is a last descriptor in a block. The extracted features are given to next process.

3.4 Hyperparameter optimization

Hyperparameter optimization is the process of finding the right combination of hyperparameter values to achieve maximum performance on the data. This paper utilized the Reverse Learning based Chimp Optimization Algorithm (RL-COA) for optimizing the hyperparameters. To enhance the population diversity and individual quality of COA, RL is utilized to attain the individual's reverse solution and then preserve the individual with high fitness values. In this chimp population, the chimps are categorized into Attacker, Barrier, Chaser, and Driver based on the variety of capabilities and intelligence which individuals demonstrate during the hunting. Every chimp species has an independent capability and utilizes its search approach to explore and predict the prey location. Consider there are N chimps and the position of the ith chimp is X_i. The chimp behavior and the prey surrounding and its position update equations are presented in Eqs. (4), (5), (6) and (7).

$$D = \left|C.X_{prey}(t) - m.X_{chimp}(t)\right| \quad (4)$$

$$X_{chimp}(t+1) = X_{chimp}(t) - A.D \quad (5)$$

$$A = f \cdot (2 \cdot r_1 - 1), C = 2 \cdot r_2 \quad (6)$$

$$m = chaotic\_value \quad (7)$$

Where, r_1 and r_2 are random vectors within the range of [0,1], X_{prey} and X_{chimp} are a position vector of prey and chimp, f is the non-linear decay factor. The t is the present iterations, D is a difference among two quantities, A is the random vector within the range of $[-f, f]$, m is the chaotic factor indicating the behavior incentives on the individual chimp position, C is the random variable. The other chimp positions are determined through the Attacker, Barrier, Chaser, and Driver and its position update equation is given in Eqs. (8) and (9).

$$\begin{cases} X_1 = |X_{Attacker} - A_1 D_{Attacker}| \\ X_2 = |X_{Barrier} - A_2 D_{Barrier}| \\ X_3 = |X_{Chaser} - A_3 D_{Chaser}| \\ X_4 = |X_{Driver} - A_4 D_{Driver}| \end{cases} \quad (8)$$

$$X(t+1) = \frac{X_1+X_2+X_3+X_4}{4} \quad (9)$$

Where, $D_{Attacker}$, $D_{Barrier}$, D_{Chaser} and D_{Driver} are distance among four types of chimps and prey in present population, $X_{Attacker}$, $X_{Barrier}$, X_{Chaser} and X_{Driver} are chimps position vector to the prey. V_1, V_2, V_3 and V_4 are preys position update vector, $X(t+1)$ is a position of $t+1$ chimps. A_1, A_2, A_3 and A_4 are similar to A. In reverse learning, the upper and lower sides of the horizontal axis are circulated with mediums A and B that have various refractive indexes and the vertical axis is the normal line which is presented in Eqs. (10) and (11).

$$\begin{cases} sin\theta_1 = ((u+l)/2 - x)/|PO| \\ sin\,\theta_2 = (x' - (u+l)/2)/|OQ| \end{cases} \quad (10)$$

$$\eta = \frac{sin\theta_1}{sin\theta_2} \quad (11)$$

Where u and l is the upper and lower bounds of the search space, $x \in [u, l]$ and O is the midpoint of the interval $[u, l]$. η is the refraction index, $k = |PO|/|OQ|$, then it is determined in Eq. (12).

$$x'_i = \frac{u_i+l_i}{2} + \frac{u_i+l_i}{2k\eta} - \frac{x_i}{k\eta} \quad (12)$$

Where, u_i and l_i is the upper and lower bounds i th dimensional vector correspondingly. x_i is an individual population, x'_i is an updated position of solution vector x_i in search space. At the final iteration, because of the particle distribution near large-quality solution which is difficult to find the global optimal solutions. So, the hyperparametric ω is presented that adaptively adjusts based on various iterations to enhance the randomness of the solution and enhance the capability to escape from the local optima which is given in Eq. (13), the individuals with lower fitness values are eliminated which is shown in Eq. (14).

$$\begin{cases} x'_i = \frac{u_i+l_i}{2} + \frac{u_i+l_i}{2\omega} - \frac{x_i}{\omega} \\ \omega = \frac{\sigma}{2} - \left(\frac{e^{t/T}-1}{e-1}\right)^{\sigma} \end{cases} \quad (13)$$

$$x_{update} = \max\_fitness(x_i, x'_i) \quad (14)$$

Where the t and T are the present and maximum iterations, σ is a sigmoid function which manages the ω attenuation rate. Table 1 presents the hyperparameter with its range.

Table 1. Hyperparameter with its range

Hyperparameter	Range
Learning rate	0.001 – 0.2
Max epochs	5 – 20
Momentum	0.0 – 1
Dropout	0.1 – 0.9

This paper optimized the learning rate of 0.001, maximum epochs of 15, momentum of 0.9 and dropout of 0.5. After attaining every chimp's reverse position, individuals with high fitness values are reserved through the greedy approach.

3.5 Classification

The optimized features are classified through Residual Attention-based LSTM with SGReLU which helps to capture complex features in input data which leads to quick convergence at training. The LSTM is a kind of Recurrent Neural Network (RNN) that can learn long-term dependency, particularly in sequence prediction problems. It has huge memory power for remembering the outcomes of every node for a much-extended time to generate the output for the next node effectively. Let us consider that x_t, h_t and C_t are input, control and cell state at time t. The $(x_1, x_2, \ldots, x_m)$ are the sequence of inputs, LSTM evaluates h-sequence ($h_1, h_2, \ldots, h_m$) and C-sequence ($C_1, C_2, \ldots, C_m$). The mathematical formulae are represented in Eqs. (15)-(20).

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (15)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (16)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (17)$$

$$\bar{C_t} = PReLU(W_c . [h_{t-1}, x_t] + b_c) \quad (18)$$

$$C_t = f_t \times C_{t-1} + i_t \times \bar{C_t} \quad (19)$$

$$h_t = o_t \times PReLU(C_t) \quad (20)$$

Where, σ represents the sigmoid function, $\bar{C_t}$ is a candidate cell state, $\times$ represents element-wise multiplication, W_f, b_f, W_i, b_i and W_o, b_o represents weight and bias of forget, input and output gate respectively, W_C, b_C represents weight and bias of cell state, i_t, o_t and f_t represents input, output and forget gates respectively. Every LSTM unit

Received: February 27, 2024. Revised: April 7, 2024.

comprises memory cell state C_t at time t, which is managed by these three gates.

3.5.1. Residual attention based LSTM

Attention based neural networks have attained numerous techniques in the prediction and classification of images/videos. The different attention mechanisms were developed previously, the residual based attention model is utilized in the research for obtaining an active weighted sum of features extracted from the frames of video. Rather than using a frame as LSTM input, utilized characteristics of the weighted image which compensates for the attention in the residual attention model. Residual-based attention is employed to specify high-level layers, sequence data, and layer formulation by searching residual functions that are given to the input layer. The mathematical formula of residual learning function is represented as Eq. (21).

$$Y = f\left(\ddot{X}, \ddot{W}\right) + \ddot{X} \tag{21}$$

Where, $\ddot{X}\ and\ Y$ represents the input and output sequential data vectors of layers, $f\left(\ddot{X}, \ddot{W}\right)$ represents the residuals learned from relative layers. In residual, the outcome of these layers creates a sequence of certain input and nonlinear residuals. The major benefit of the method is it develops shortcut functions between certain layers for highly efficient model training as well as beneficial for protecting the vanishing gradients problems. This research normalizes the data by evaluating normalization in residual LSTM to an easy effective hidden state, normalizes data of neurons in LSTM and minimizes the training time. The mathematical formula is represented from Eqs. (22)-(24).

$$\tilde{n}_t = \frac{1}{h}\sum_{i=0}^{h}(ht)i \tag{22}$$

$$\delta_t = \sqrt{\frac{1}{h}\sum_{i=0}^{h}((ht)i - \tilde{n}_t)^2} \tag{23}$$

$$\acute{y}_t = f\left(\frac{\dot{g}}{\delta_t} \odot (ht - \tilde{n}_t) + b\right) \tag{24}$$

Where, $(ht)i$ represents the hidden state in every LSTM layer of ith neuron, $\dot{g}$ and b represents trainable weights which are utilized for reducing input activation function f, $\odot$ is an element-wise multiplication, t is represented as a time step. The threshold value is applied to 0.5 in every residual LSTM layer to minimize overfitting. The research utilized compression and decompression with attention mechanism-based LSTM to improve the performance of video encoding. It evaluated the decompression for block generation which inputs video features to the next block, which depends on frames generated through the model. This technique efficiently produces video frames utilizing two kinds of inputs like $1D$ feature vector and $2D$ video frame information. The research utilized the SGReLU activation layer with residual LSTM which evaluates the short and long-term dependences using latent correlation between features at different locations. In the video encoding model, the video frame is an input given for residual attention-based LSTM which needs a single feature block for sequence. The mathematical formulae are represented from Eq. (25), (26) and (27).

$$k_t = \frac{1}{h}\sum_{i=1}^{h} \widehat{W}_t^i f_i \tag{25}$$

$$S_t = \widehat{W}^T SGReLU(\widehat{W}_h ht + M_h R_f + b_h) \tag{26}$$

$$A_t = SGReLU\ (S_t) \tag{27}$$

Where, $\widehat{W}^T$, $\widehat{W}_h$, M_h and b_h represents parameters learned for features of frame f_i respect to attention weight $\widehat{W}_t^i$ for returning to the score S_t. A_t represents the output. The extracted 35-frame sequences are passed through residual attention-based LSTM and the final prediction is done by utilizing SGReLU layer.

3.5.2. Self gated rectified linear unit

The limitations of ReLU and other activation functions are causing neuron death, vanishing gradient and output offset. To overcome these issues, this paper proposed an SGReLU that is a non-monotonic constant function that preserves the equal binary format. The nature of activation functions is stated as bivariate function $b(x, g(x))$, where, x is an unfiltered pre-activation that is given as input to the final bivariate function which is presented in Eq. (28).

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha.x.\sigma(x), & x < 0 \end{cases} \tag{28}$$

Where, $\sigma(x)$ is a sigmoid function that is equivalent to $\frac{1}{1+e^{-x}}$ and α is a hyperparameter within the range of [0.1, 1]. As an output, the SGReLU resolves the problems of non-zero gradients and saturation that speed up the learning procedure. The absence of a sigmoid function minimizes the power

Received: February 27, 2024. Revised: April 7, 2024.

operations that lead to faster processing. The self-gated approach is utilized to make the negative function area adaptable through a single input. The SGReLU delivers a negative bump for small inputs that removes neuron death issues. The first SGReLU derivative function is shown in Eq. (29).

$$f'(x) = \begin{cases} 1, & x \geq 0 \\ \alpha \left[\frac{1}{1+e^{-x}} + \frac{xe^{-x}}{(1+e^{-x})^2}\right], & x < 0 \end{cases} \quad (29)$$

Where, α manages the depth of the negative dump, for non-zero inputs, the derivative is one in a positive region. Through, the negative region is an integration of two functions so a partial derivative is accomplished to attain the gradient at backpropagation time.

4. Experimental result

The residual attention-based LSTM with SGReLU is stimulated with Python with system configuration: 16GB RAM, windows 10 OS and i7 processor. The AUC, accuracy, precision, recall and f1-score are exploited for measuring residual attention-based LSTM with SGReLU performance. The formulation of the metrics is shown in Eqs. (30)-(34).

$$AUC = \frac{\sum R_i(I_I) - I_I(I_I+I)/2}{I_I + I_f} \quad (30)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (31)$$

$$Precision = \frac{TP}{TP+FP} \quad (32)$$

$$Recall = \frac{TP}{TP+FN} \quad (33)$$

$$F1-score = 2 \times \frac{Precision \times recall}{Precision+recall} \quad (34)$$

Where, R_i is the rate of ith image, I_I and I_f is the number of positive and negative images, TP, TN, FP and FN denote the True Positive, True Negative, False Positive and False Negatives correspondingly.

4.1 Quantitative and qualitative analysis

This Residual Attention-based LSTM with SGReLU performance is measured by accuracy, precision, recall, f1-score and AUC. Fig. 3 shows the exponential curve for accuracy using UCF-Crime dataset. Fig. 4 shows the exponential curve for loss using UCF-Crime dataset. Fig. 5 shows the exponential curve for confusion matrix for UCF-Crime dataset.

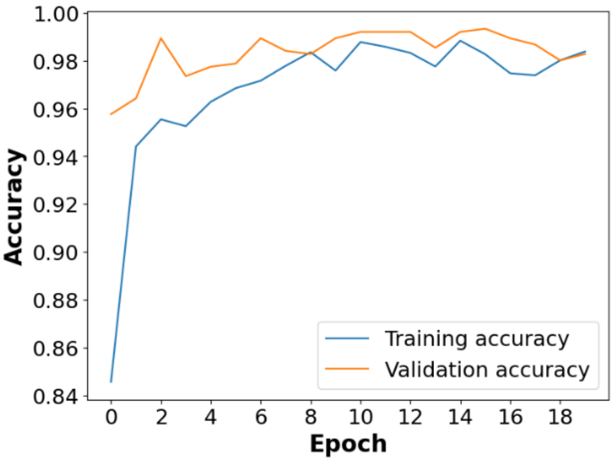

Figure. 3 Epoch v/s Accuracy

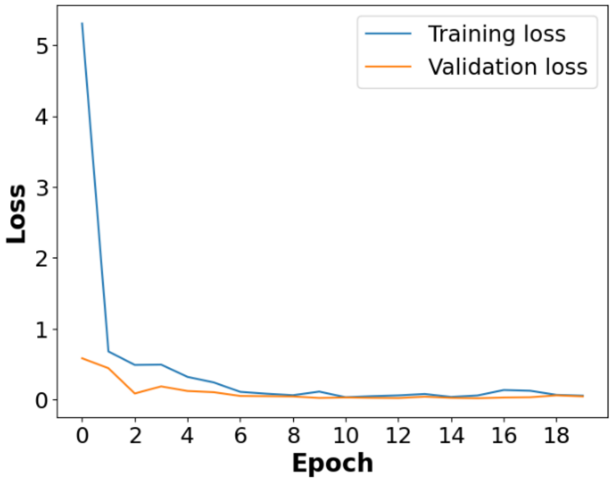

Figure. 4 Epoch v/s Loss

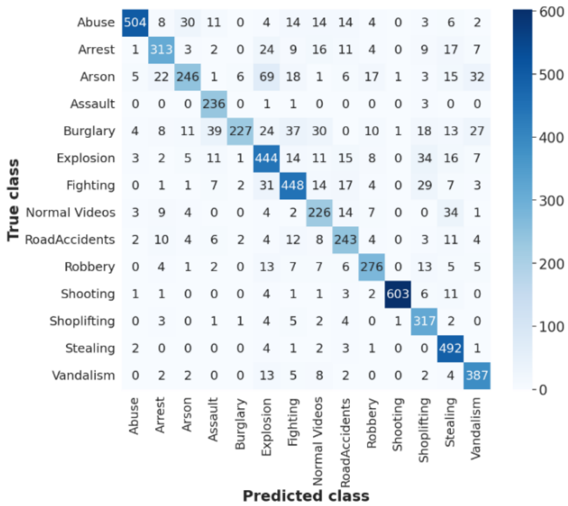

Figure. 5 Confusion Matrix

 DOI: 10.22266/ijies2024.0630.45

Table 2. Performance of hyperparameter optimization

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
WOA	90.84	90.68	90.53	90.47	90.28
GOA	92.59	92.51	92.47	92.38	91.77
COA	93.73	93.65	93.58	93.31	93.12
RL-COA	95.68	95.59	95.51	95.46	95.35

Table 3. Performance of activation function

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
ReLU	92.44	92.39	92.32	92.25	92.21
PReLU	94.73	94.64	94.59	94.43	94.17
LReLU	96.38	96.27	96.19	95.75	95.49
SGReLU	97.51	97.48	97.43	97.38	97.23

Table 4. Performance of classification

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
RNN	93.71	93.67	93.54	93.48	92.83
CNN	94.58	94.49	94.41	94.37	94.18
LSTM	96.66	96.41	96.39	96.12	95.74
Residual Attention-based LSTM with SGReLU	98.84	98.62	98.47	98.35	98.21

Table 5. Comparative Analysis

Dataset	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
UFC-Crime	ResNet18+SRU [16]	88.65	89.51	N/A	88.65	88.71
	Residual attention-based LSTM [17]	78.43	87	78	81	96
	Multi-branch CNN-GRU [18]	86.78	N/A	N/A	N/A	N/A
	Proposed Residual Attention-based LSTM with SGReLU	**98.84**	**98.62**	**98.47**	**98.35**	**98.21**
Avenue	STCEN [20]	N/A	N/A	N/A	N/A	86.6
	DF-ConvLSTM-VAE [21]	N/A	N/A	N/A	N/A	87.2
	Proposed Residual Attention-based LSTM with SGReLU	**96.43**	**95.51**	**95.38**	**92.19**	**89.76**
UMN	OPLKT+CNN [22]	91.67	95.78	90.67	93.09	N/A
	Proposed Residual Attention-based LSTM with SGReLU	**94.59**	**96.46**	**93.24**	**94.61**	**93.35**

Table 2 show the performance of RL-COA on the UCF-Crime dataset. The Whale Optimization Algorithm (WOA), Grasshopper Optimization Algorithm (GOA) and COA performance are restrained and compared to RL-COA. The RL-COA achieves best results through accuracy, precision, recall, f1-score and AUC values about 95.68%, 95.59%, 95.51%, 95.46% and 95.35% correspondingly when compared to existing techniques.

Table 3 show the performance of SGReLU on the UCF-Crime dataset. The performance of ReLU, Parametric ReLU (PReLU), and Leaky ReLU (LReLU) are restrained and compared to SG-ReLU. The SG-ReLU achieves better results through accuracy, precision, recall, f1-score and AUC values of about 97.51%, 97.48%, 97.43%, 97.38% and 97.23% correspondingly when compared to existing techniques.

Table 4 show the performance of residual attention-based LSTM with SGReLU on the UCF-Crime dataset. The performance of RNN, CNN, and LSTM are restrained and compared to residual attention-based LSTM with SGReLU. The residual attention-based LSTM with SGReLU achieves best results through accuracy, precision, recall, f1-score, and AUC values of about 98.84%, 98.62%, 98.47%, 98.35% and 98.21% correspondingly when compared to other existing techniques.

4.2 Comparative analysis

The comparative analysis of residual attention-based LSTM with SGReLU by metrics like accuracy, precision, recall, f1-score, and AUC as shown in Table 5. The existing result such as [16-22] are employed to estimate classifier ability. The Residual Attention-based LSTM with SGReLU is trained, tested and validated by UFC-Crime, Avenue and UMN dataset. The attained result shows that the Residual Attention-based LSTM with SGReLU achieves best results through accuracy, precision, recall, f1-score and AUC values of about 98.84%, 98.62%, 98.47%, 98.35% and 98.21% for UCF-Crime dataset while comparing other classifiers.

4.2.1. Discussion

The proposed model advantages and existing model limitations are examined. The ResNet18+SRU [16] was unable to manage long-term dependences and consecutive data. The Residual attention-based LSTM [17] required a large number of labeled training data. The Multi-branch CNN-GRU [18] required highly labeled data which consumes more time and labor-intensive tasks. The ConvLSTM [19] was sensitive to noisy data and tried to identify the optimum hyperplane. The CNN-based autoencoder-GAN [20] is computationally intensive and struggles with noisy data. The proposed Residual Attention-based LSTM with SGReLU tackle these existing model limitations. The residual attention-based LSTM with SGReLU overcomes the neuron dead issues by allowing negative values for some neurons and minimizing the probability of inactive neurons.

5. Conclusion

This paper proposed a deep learning-based approach for detecting the weapon and anomalous behavior. The Residual Attention based LSTM with SGReLU is proposed to enhance the detection accuracy. The median filter is utilized for preprocessing which removes the noise from UCF-Crime dataset and feeds into HOG for feature extraction. Then, the RL-COA is utilized for hyperparameter optimization which attain the individual's reverse solution and then preserved the individual with high fitness values. At last, the Residual Attention-based LSTM with SGReLU is utilized for classification process. This model overcomes the neurons dead issues through allows negative values for some neurons and minimizing the probability of inactive neurons. The proposed model attained better result on UCF-Crime dataset through accuracy, precision, recall, f1-score and AUC values of about 98.84%, 98.62%, 98.47%, 98.35% and 98.21% correspondingly which ensures accurate detection when compared to existing techniques. The future work is to develop a deep learning and generative technique for identifying more classes of abnormal behaviors.

Notation

Notation	Description
S_{xy}	Group of coordinates
$\hat{f}(x,y)$	Restored image
$g(s,t)$	Calculated and corrupted area
s and t	Coordinate variables in S_{xy} at point (x,y)
$G_x(x,y,z)$ and $G_y(x,y,z)$	Vertical and horizontal masks
θ	Pixel orientation
$G(x,y,z)$	Magnitude of gradients
N_f	Normalized histogram features
r_1 and r_2	Random vectors within the range of [0,1]
f	Non-linear decay factor
t	Present iterations
D	Difference among two quantities
A	Random vector within the range of $[-f,f]$
m	Chaotic factor
C	Random variable
u and l	Upper and lower bounds of the search space
O	Midpoint of the interval $[u,l]$
η	Refraction index
T	Maximum iterations
$X_{Attacker}, X_{Barrier}$ and X_{Driver}	Chimps position vector to the prey
$D_{Attacker}, D_{Barrier}$ and D_{Driver}	Distance among four types of chimps and prey in present population
X_{prey} and X_{chimp}	Position vector of prey and chimp
V_1, V_2, V_3 and V_4	Preys position update vector
$X(t+1)$	Position of $t+1$ chimps
x_i	Individual population
x'_i	Updated position of solution vector
x_t, h_t and C_t	Input, control and cell state at time t
$(x_1, x_2, \dots, x_m)$	Sequence of inputs
σ	Sigmoid function
W_f, b_f	Weight and bias of forget gate
W_i, b_i	Weight and bias of input gate
W_o, b_o	Weight and bias of output gate
W_C, b_C	Weight and bias of cell state
i_t, o_t *and* f_t	Input, output and forget gates
$\bar{C}_t$	Candidate cell state
$\odot$	Element-wise multiplication

$\ddot{X}$ *and* Y	Input and output sequential data vectors of layers
$f(\ddot{X}, \ddot{W})$	Residuals learned from relative layers
$(ht)i$	Hidden state in every LSTM layer of ith neuron
$b(x, g(x))$	Bivariate function
x	Unfiltered pre-activation
R_i	Rate of ith image
I_I and I_f	Number of positive and negative images
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

References

[1] M. Kumar, and M. Biswas, "Abnormal human activity detection by convolutional recurrent neural network using fuzzy logic", *Multimedia Tools and Applications*, 2023.

[2] F.J. Rendón-Segador, J.A. Álvarez-García, J.L. Salazar-González, and T. Tommasi, "Crimenet: Neural structured learning using vision transformer for violence detection", *Neural Networks*, Vol. 161, pp. 318-329, 2023.

[3] S.A. Jebur, K.A. Hussein, and H.K. Hoomod, "Abnormal Behavior Detection in Video Surveillance Using Inception-v3 Transfer Learning Approaches", *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, Vol. 23, No. 2, pp. 210-221, 2023.

[4] A. Mehmood, "LightAnomalyNet: A Lightweight Framework for Efficient Abnormal Behavior Detection", *Sensors*, Vol. 21, No. 24, p. 8501, 2021.

[5] P.V. Kolaskar, A.R. Maitre, P.R. Khopkar, S.S. Gaikwad, and D. Abin, "Anomaly Motion Detection and Tracking for Real-Time Security System", In: *Proc. of Computer Networks and Inventive Communication Technologies: Proceedings of Third ICCNCT 2020*, Springer Singapore, pp. 707-717, 2021.

[6] S. Pouyan, M. Charmi, A. Azarpeyvand, and H. Hassanpoor, "Propounding first artificial intelligence approach for predicting robbery behavior potential in an indoor security camera", *IEEE Access*, Vol. 11, pp. 60471-60489, 2023.

[7] C. Liu, R. Fu, Y. Li, Y. Gao, L. Shi, and W. Li, "A self-attention augmented graph convolutional clustering networks for skeleton-based video anomaly behavior detection", *Applied Sciences*, Vol. 12, No. 1, p. 4, 2022.

[8] E. Arif, S.K. Shahzad, M.W. Iqbal, M.A. Jaffar, A.S. Alshahrani, and A. Alghamdi, "Automatic Detection of Weapons in Surveillance Cameras Using Efficient-Net", *Computers, Materials & Continua Tech Science Press DOI*, Vol. 72, No. 3, p. 10, 2022.

[9] Z.K. Abbas, and A.A. Al-Ani, "An adaptive algorithm based on principal component analysis-deep learning for anomalous events detection", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 29, No. 1, pp. 421-430, 2023.

[10] A. Nazir, R. Mitra, H. Sulieman, and F. Kamalov, "Suspicious Behavior Detection with Temporal Feature Extraction and Time-Series Classification for Shoplifting Crime Prevention", *Sensors*, Vol. 23, No. 13, p. 5811, 2023.

[11] G. Giorgi, W. Abbasi, and A. Saracino, "Privacy-Preserving Analysis for Remote Video Anomaly Detection in Real Life Environments", *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, Vol. 13, No. 1, pp. 112-136, 2022.

[12] L. Xia, and Z. Li, "A new method of abnormal behavior detection using LSTM network with temporal attention mechanism", *The Journal of Supercomputing*, Vol. 77, pp. 3223-3241, 2021.

[13] J. Pan, L. Liu, M. Lin, S. Luo, C. Zhou, H. Liao, and F. Wang, "An Improved Two-stream Inflated 3D ConvNet for Abnormal Behavior Detection", *Intelligent Automation & Soft Computing*, Vol. 30, No. 2, 2021.

[14] S. Habib, A. Hussain, W. Albattah, M. Islam, S. Khan, R.U. Khan, and K. Khan, "Abnormal activity recognition from surveillance videos using convolutional neural network", *Sensors*, Vol. 21, No. 24, p. 8291, 2021.

[15] T. Saba, A. Rehman, R. Latif, S.M. Fati, M. Raza, and M. Sharif, "Suspicious activity recognition using proposed deep L4-branched-ActionNet with entropy coded ant colony

system optimization”, *IEEE Access*, Vol. 9, pp. 89181-89197, 2021.

[16] M. Qasim, and E. Verdu, “Video anomaly detection system using deep convolutional and recurrent models”, *Results in Engineering*, Vol. 18, p. 101026, 2023.

[17] W. Ullah, A. Ullah, T. Hussain, Z.A. Khan, and S.W. Baik, “An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos”, *Sensors*, Vol. 21, No. 8, p. 2811, 2021.

[18] Z. Xu, and Y. Lu, “Abnormal behavior detection algorithm based on multi-branch convolutional fusion neural network”, *Multimedia Tools and Applications*, Vol. 82, pp. 22723-22740, 2023.

[19] R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, “A new approach for abnormal human activities recognition based on ConvLSTM architecture”, *Sensors*, Vol. 22, No. 8, p. 2946, 2022.

[20] Y. Hao, J. Li, N. Wang, X. Wang, and X. Gao, “Spatiotemporal consistency-enhanced network for video anomaly detection”, *Pattern Recognition*, Vol. 121, p.108232, 2022.

[21] L. Wang, H. Tan, F. Zhou, W. Zuo, and P. Sun, “Unsupervised anomaly video detection via a double-flow convlstm variational autoencoder”, *IEEE Access*, Vol. 10, pp. 44278-44289, 2022.

[22] G. Rajasekaran, and J. Raja Sekar, “Abnormal Crowd Behavior Detection Using Optimized Pyramidal Lucas-Kanade Technique”, *Intelligent Automation & Soft Computing*, Vol. 35, No. 2, pp. 2399-2412, 2023.

[23] UCF-Crime dataset link: https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset (accessed on 21/12/2023).

[24] E. Akanksha, and P.R. Koteswara Rao, “A Feature Extraction Approach for Multi-Object Detection Using HoG and LTP”, *International Journal of Intelligent Engineering & Systems*, Vol. 14, No. 5, 2021, doi: 10.22266/ijies2021.1031.24.