Journal of Information Systems Engineering and Management
2025, 10(14s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Conv2D-LSTM-AE-GAN: Convolutional 2D LSTM Auto Encoder Generative Adversarial Network

Mrs. Swapna.C[1], Dr. B. Padmaja Rani [2] and Mr. Manoj Reddy Dasari[3]
[1]*Assistant Professor, AI & DS Department, SCTEW, India*
[2]*Professor, CSE Department, JNTUH, India.*
[3]*Morgan Stanley, Atlanta, Georgia, USA*
[1]*swapnac.jntuh@gmail.com*, [2]*padmaja_JNTUH@jntuh.ac.in*, [3]*dasarimanojreddy@gmail.com*

ARTICLE INFO

Received: 22 Nov 2024
Revised: 14 Jan 2025
Accepted: 28 Jan 2025

ABSTRACT

Surveillance video refers to video footage captured by cameras for the purpose of monitoring and recording activities in specific environments. These videos are commonly used for security purposes in places such as airports, shopping malls, streets, industrial facilities, hospitals, and other public or private spaces. The primary objective of surveillance video systems is to maintain safety, detect suspicious activities, and collect evidence for investigation. Anomaly detection in Surveillance video is an important and evolving field with applications across various industries. It involves analyzing video data to detect unusual or suspicious events, which could indicate threats, errors, or rare occurrences. While traditional methods have been useful, recent advancements in learning methods, particularly using 2D Convolutional Long Short Term Memory, Autoencoders, and Generative Adversarial Networks have made significant improvements in detecting complex anomalies. Our proposed system based on Autoencoder with Convolutional 2DLong Short Term Memory unit in Generative Adversarial Network. The model aims to learn the appropriate normal data distribution during training. Frames with a large variance in their regularity score are identified as anomalies based on this distribution. We have adopted depth-wise separable convolution with Conv2DLSTM unit in auto encoder to learn spatial and temporal features to reconstruct and differentiate generated frame with real frame in video sequence, and make the model lightweight and efficient. The entire system has been evaluated on many benchmark datasets using metrics like AUC and Equal Error Rate (EER) and shown to be reliable for complicated video anomaly identification.

Keywords: LSTM, GAN, Autoencoder.

1 Introduction:

Anomaly detection in video surveillance focuses on identifying abnormal events or activities that deviate from expected patterns. Anomaly detection plays a crucial role in the functioning of modern video surveillance systems, helping to monitor large volumes of video data effectively and alert security personnel in real-time. Anomaly detection is useful for a variety of applications, including intrusion detection in time series, surveillance, activity detection, and healthcare monitoring. Manually finding anomalies in surveillance films is a laborious and time-consuming task that requires human resources. With the proliferation of security cameras, the necessity for an automated system for detecting anomalies in videos has become increasingly apparent. These systems serve an important role in security control, crime detection, accidents, and traffic monitoring, because the amount of data accessible from multiple sources is simply too large to manually analyse.

An anomalous event occurs unexpectedly and rarely in practice, making it difficult to describe such different happenings. For example, running on the beach is regarded regular conduct, however running in a retail mall is considered unusual. As a result, the automated system will have no prior knowledge of the nature of past or future irregularities. Representational learning approaches, such as sparse coding, perform well in anomaly detection [20, 46]. Anomaly detection in videos, in particular, can be addressed in either a supervised or unsupervised learning scenario. In supervised anomaly detection, the system learns from example films that are classified as anomalous or non-anomalous [17, 37, 45]. In the unsupervised approach, the system considers each unusual or abnormal occurrence that differs from the learnt normal sample parameters to be anomalous [38]. The model can thus be taught to detect abnormalities with large amounts of unlabeled, 'regular' data. Variations in surveillance,

such as changes in scale and viewpoint, create extra ambiguity.

Several research have focused on the application of Convolutional Neural Networks (CNNs) for (supervised) anomaly identification in industrial products [43], such as evaluating concrete surfaces [5] and cracking [15]. However, in the supervised situation, there is typically an unequal balance of normal and anomalous data. The use of data augmentation has been recommended to reduce this obstacle; nonetheless, this setting still has significant limitations in addressing real-world situations [9, 47]. In contrast, Ravanbakhsh et al. have recently proposed using adversarial learning to localise anomalous activity in an unsupervised scenario [30]. Generative Adversarial Networks (GANs), in particular, because of their ability to represent high-dimensional picture data, have recently emerged as the state-of-the-art in anomaly detection. Several GAN-centric architectures, including AnoGAN [35] and GANomaly [2], have been proposed for this purpose. The goal of GAN, which promotes generated samples to resemble genuine data, is not directly related to the goal of conducting anomaly detection. As a result, in much recent work in anomaly detection, adversarial training has been tweaked to improve both training and inference for this particular purpose [14, 39]. Overall, CNN and GAN designs are inefficient for use on edge devices such as robotics, smart surveillance cameras, self-driving cars, and microcomputers. Furthermore, most unsupervised GAN architectures use shallow networks that are meant to learn just spatial characteristics, ignoring the critical temporal component of movies. Because of the enormous number of parameters, such networks are limited to low-dimensional data and are prone to overfitting.

Conv2D-LSTM-AE-GAN Contributions: This paper, inspired by the AnoGAN architecture (encoder decoder encoder pipeline) but significantly departing from it, we propose a light-weight, efficient anomaly detection architecture with a reduced number of parameters, aimed at addressing the convergence and overfitting problems with GAN training and at achieving real-time performance. Our Proposed architecture is capable of learning in main contributions are:

- *Our Autoencoder based GAN* making use of depth-separable and time distributed convolution 2D LSTM layers in Autoencoder in Generator and encoder with Conv2D LSTM layers in Discriminator of our own design, which leads to increased efficiency due to retaining the history of learning in LSTM Unit. Our model is both lightweight and more efficient.
- *Losses:* Losses in our proposed architecture improves performance with total 4 losses (Adversarial loss, Contextual loss, Encoder loss, SVD loss)

The losses in GANs drastically improves their performance, the new loss can be widely employed in other deep learning models for better representation learning, whenever few training samples are available. Our system can detect complex anomalous events occurring for a very short time, and outperforms on public alky available datasets like UCSD Ped1, UCSD Ped2, CHUK Avenue.

2 Related Work

Spatial Feature Extraction Methods for anomaly detection: In challenging scenarios, deep learning-based methods have outperformed the prior state of the art in the detection of anomalous events in films [7, 19, 21, 33, 40]. In typical architectures, handmade feature extraction methods are not as powerful as deep neural networks with hierarchical feature representation learning. In particular, deep generative models have drawn interest lately because to their ability to encode complex changes. GANs are suggested by Liu et al. as a way to identify anomalies by minimising the difference between ground truth frames and projected future video frames [19]. In order to minimise model reconstruction loss, training video frames (Vtrain = {Vi}) and a parametric representation (fθ: Vtrain → Rl) are used to approximate the normal data distribution during training. Every test frame Vj ∈ Vtest generates an anomaly score A(Vj) at test time based on the deviation from the learnt optimal representation fθ∗: A(Vj) = fδ∗ (Vj) Vj. Lastly, a threshold T, A(Vj)>T, is applied to the anomaly score in order to identify abnormalities.

Generative Adversarial Networks (GANs) are made up of two networks: a generator and a discriminator trained on unlabeled data [10, 29, 32]. The generator G seeks to capture the data distribution and generate realistic video frames by constructing a data distribution for the input data V via a mapping from a previous latent space noise distribution z. The discriminator D's goal is to determine the likelihood of the sample being outputted by the generator. The generator and discriminator compete against each other in a zero-sum min-max game: $\min_G \max_D V(D, G) = E_{V\ Pdata(V)} \log D(V) + EZ\ pz(Z) \log(1\ D(G(Z))$. In recent years, a variety of anomaly detection GAN designs have been presented. Mizra et al. [25] proposed an architecture called extended conditional GAN. This model conditions either the generator G or the discriminator D with additional information Y. The condition Y can be formulated using multimodal input data or class labels as auxiliary information. Vu et al. [41] introduced a robust anomaly detection system for movies that employs

conditional GANs to accurately detect video anomalies at various levels of representation using a layer-wise approach. The extraction of optical flow data is difficult in uncertain situations defined by untextured regions, light variations, occlusions, and quick motions. The generator's inverse mapping, E = G1, is learned by the encoder E (V, E(V)) in the bidirectional GAN (BiGAN) architecture proposed by Donahue et al. [8]. By learning to map the latent space to image data during training, the model improves results on the MNIST benchmark dataset (8) and lowers statistical complexity. 'fast-AnoGAN' is a unique mapping approach that was introduced by Schlegl et al. [34] in 2019. It can detect anomalies at the image level and localise them at the pixel level. Utilising the BiGAN design, Zenati et al. presented the Efficient GAN-Based Anomaly Detection (EGBAD) system in 2018 [44]. Conversely, Ackay and colleagues proposed the theory of integrated learning of latent space vectors and images.

The architecture captures the data distribution of image and latent space vectors using an adversarial autoencoder and an encoder-decoder-encoder pipeline. However, this method has limitations in managing spatial-temporal learning, which leads to unstable reconstructions for real-time films [2]. Vu et al. [41] proposed a robust anomaly detection system for movies that uses conditional GANs to learn representations from both intensity and motion information. A deep CNN for anomaly detection that learns a correlation between frequent object appearances (such as pedestrians, backdrops, and trees) and their corresponding motions was proposed by Nguyen and Meunier [27]. Using two U-Net blocks in the generator, Tang et al. proposed a combination of future frame prediction and reconstruction for anomaly identification [1]. The first block attempted to anticipate frames, while the second block recreated the frames produced by the first block. CNN (Convolutional Neural Network) Excellent at image and video analysis, local feature extraction, spatial relationships, robustness to noise and translation invariance Large number of parameters, high computational cost, limited interpretability. LSTM (Long Short-Term Memory) Effective at capturing long-term dependencies, handles variable length sequences, robustness to vanishing/exploding gradient problem Computational complexity, training time, limited interpretability[48],[49],[50],[51].

Temporal Feature Extraction Methods: In an encoder-decoder paradigm, Jefferson et al. created a Conv-LSTM network to predict upcoming frames and identify anomalies through reconstruction [24]. For visual anomaly detection, the similar architecture was demonstrated to be promising in [22]. The input video frames are processed by the convolutional LSTM to extract features, which are subsequently deconvolutionally rebuilt. Stacking Recurrent Neural Networks (RNNs) map identical neighboring frames to a reconstruction coefficient in Luo et al.'s Temporally-coherent Sparse Coding (TSC) technique [21]. LSTM autoencoders are ideal for extracting spatial and temporal information. Shi et al. [36] and Pa-traucean et al. [28] used layered convolutional LSTMs in an autoencoder architecture to extract features from video sequence data. Conv-LSTMs can extract both spatial representations and spatio-temporal information from a sequence of video frames, as well as forecast future frames with high accuracy. In Conv-LSTMs, the size of the convolutional filter in the hidden-to-hidden connection determines how much information the hidden state receives in the preceding time step. Large transitional kernels are used to record faster motions, while smaller kernels are adequate for slower motions [41]. But as the number of parameters increases, GAN performance drastically declines, leading to GANs commonly failing on high-dimensional data. Our strategy reduces model size and the chance of overfitting during GAN training by utilizing depth-wise separable convolutions within the LSTM. To handle the spatial and depth dimensions of the video frames, depth-wise separable convolution is followed by pointwise convolution. LSTM Good at capturing temporal dependencies Suitable for time-series data Longer training times Difficulty in parallelizing. CNN High accuracy and performance Effective in handling spatial features Requires large datasets for training Computationally expensive [48],[49],[50],[51].

3. Architecture & Proposed Methodology

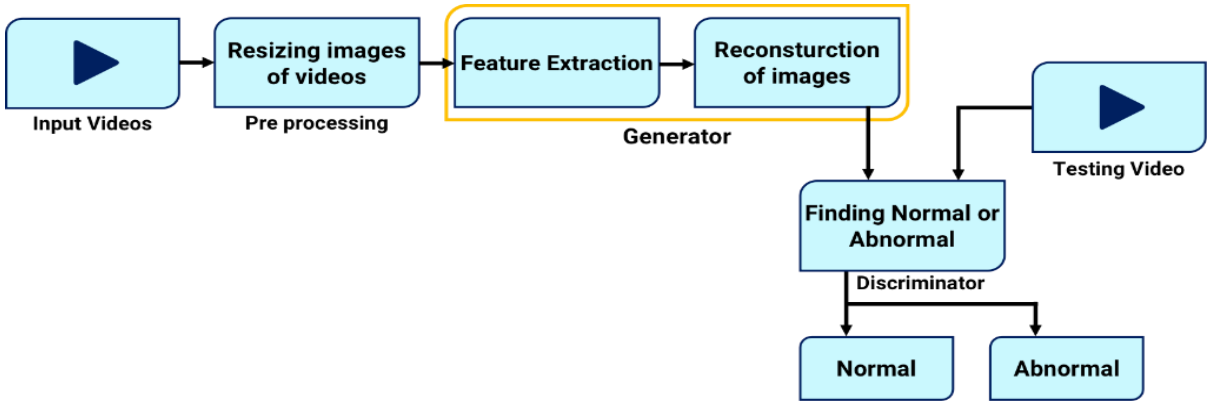

Figure 1: Block diagram for video anomaly detection.

To identify abnormalities in video data, the Anomaly Detection method in video shown in Figure 2 combines Discriminative and Generative models. In order to ensure uniformity and lower computing cost, the process starts with the preprocessing step-by-step technique illustrated in Algorithm1, where incoming videos are broken up into frames at a rate of 24 frames per second and downsized to 128x128 pixels. Sequential sets of seven scaled frames are aggregated and saved as individual clips during the preprocessing phase. These clips are then utilized as input to train the Generative Adversarial Network model. GAN contain 2 blocks Generator and Discriminator. The Generator is an Autoencoder that uses convolutional LSTM, and transposed convolutional layers to recreate the input frames during the training phase. The Encoder-based Discriminator then assesses both the real and reconstructed frames to determine which are generated and which are real. To improve the Generator and Discriminator, a variety of losses are computed and backpropagated, such as Adversarial loss, Contextual loss, Encoding loss, and SVD losses. Once training is complete, the model is saved for testing. The same preparation procedures are used to a test video during the testing phase. The trained Discriminator then assesses the test frames after they have been rebuilt by the Generator. Anomalies are indicated by low scores, which are determined by how closely a frame resembles a real frame. Effective detection of anomalous occurrences is made possible by classifying frames below a predetermined threshold as Abnormal or Anomaly and those above the threshold as Normal.

3.1 .Architecture

Proposed Architecture framework for Conv2D-LSTM-AE- GAN given in figure 3. The diagram illustrates the (Convolutional 2D LSTM Auto Encoder Generative Adversarial Network) architecture for anomaly detection in video frames. It comprises two primary components: the Generator and the Discriminator.

The Generator takes resized video frames (128x128) (step by step procedure shown in Algorithm1) as input and reconstructs the frames using a combination of Time Distributed Separable Conv2D layers, ConvLSTM2D layers, and Conv2DTranspose layers. These layers extract spatial and temporal features, progressively refine the reconstruction, and output the generated frame. The generated frame is compared with the real frame by the Discriminator, which also employs a similar architecture, including ConvLSTM2D and Time Distributed Dense layers, to classify frames as real or generated (step by step procedure shown in Algorithm 2). Losses from Generator and Discriminator (Adversarial loss, Contextual loss, Encoding loss, and SVD losses) are backpropagated to improve both components iteratively. During testing, low-scoring frames from the Discriminator are flagged as anomalous, while high-scoring ones are classified as normal (step by step procedure shown in Algorithm3). This architecture leverages spatiotemporal features to enhance anomaly detection accuracy.

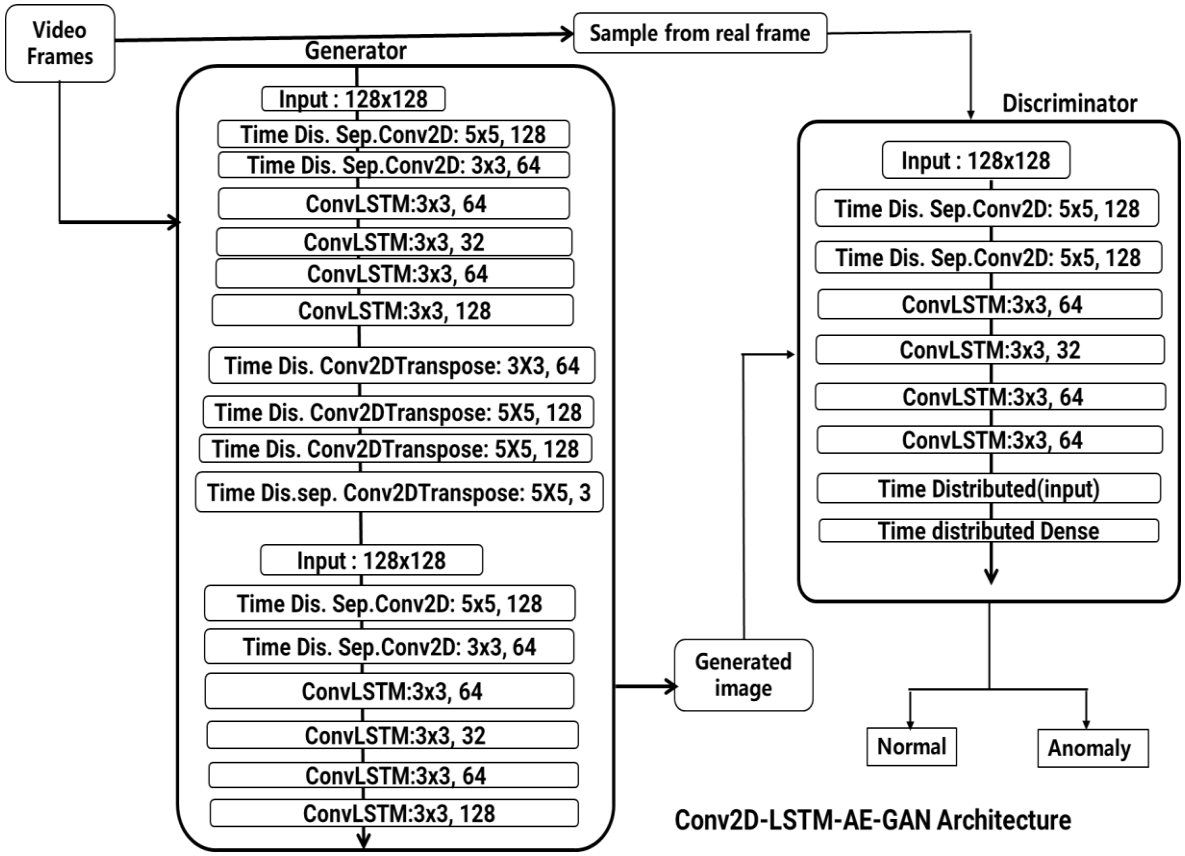

Figure 2 Proposed Conv2D-LSTM-AE-GAN Architecture

Algorithm 1 Preprocessing Algorithm for video anomaly detection

Input: video1 to videon,

Output: clips []

Parameters: N=Total no. of frames, M=No. of clips.

Method: For All video1 to videon.

1.Begin

2.For each video

2.1 Convert video into frames (24 frame per sec.)

3.For each frame in video

3.1 resized_frame: =resize(frame,128,128))

3.2 allframes[] . append[resized_frame]

4.N: =count(allframes) /*N =Total no. of frames*/

5.strides i:1 to 3, count: =0, clips: =[ ], clip:=[ ]

6.For each stride 1 to 3

6.1 For each frame of allframes

6.1.1 clip.append[frame]

6.1.2 count++

6.1.3 if count: ==7 /*No. of sequential frames= 7*/

6.1.4 clips. Append [clip]

7.Return clips [ ] M:=Count(clips) /* M=No. of Clips*/

8.End

Time Complexity of Preprocessing Algorithm O(NxM)

N=No. of Frames, M=No. of clips (Reviewer 1)

Algorithm 2 Training Algorithm for Conv2D-LSTM-AE-GAN Architecture Framework

Input: clips[ ] C1 to Cn (HXW=128X128)

Output: LSTM_GAN_Model

Parameters: Bat-Batches, Ep-Epochs, E=no. of. Epochs, n=no. of frames

Af1-Tanh activation function, Af2-Sigmoid Activation function, K = No. of Filters

B= no. of batches (Ex.4) Xn =no. frames for training.

Method: For clips Ci: C1 to Cn

1.Begin

2.For each Ep q :1 to E

3. For each Bat j:1 to B

3.1 Z: =E ϕ(ConvLSTM(Wi*Ci+ bi),K) /*C=Input frame*/

/* Z= Latent space */ /* E=Encoder with convLSTM=Convolutional LSTM layers*/

3.2 Ć=Dɵ(Z) /*D=Docoder, Ć=Reconstructed frame */

/* ϕ,ɵ denote hidden parameter */

3.3 Ladv:=Dis(Af2(C, Ć)) /*Dis= Discriminator*/

4. Ź= Eϕ(ConvLSTM(Ć)) /* Ź = Latent space of reconstructed frame */

5. Compute Losses

5.1 Ladv:=Mean(C- Ć) /* Ladv=Adversarial Loss */

5.2 Lcnt:=Mean(abs(C- Ć)) /* Lcnt=Contextual Loss*/

5.3 Lenc:= Mean(Z - Ź)^2 /* Lenc=Encoder Loss */

5.4 Lsvd:= Mean(abs(difference in U,S,VT Matrices of C, Ć) /* Lsvd =SVD Loss*/

5.5 LGAN= Lsvd + Ladv + Lenc + Lcnt

6. Back propagate and update weights:

$\Delta W := \nabla W \sum i=1 n LGAN$

7. Train LSTM_GAN_Model using LGAN

8. Compile with optimizer, loss and weights:

LSTM_GAN_Model . compile (adam , LGAN, weights)

9. Trained LSTM_GAN_Model

10. End

Time Complexity: O(B*H*W*K*E)

B=Batch size, H=Height of frame, W=Width of frame, K=No. of Filters, E= No. of Epochs

Algorithm 3 Testing Algorithm for Conv2D-LSTM-AE-GAN Architecture Framework

Input: Test Video frames t1 to tn , trained LSTMGANMODEL

Output: Normal and Anomaly frames.

Parameters: thr-Threshold, LGAN(t'i)-Model generated frame

tn=No.of frames for testing

Method:

1. Begin

2. For each ti: t1 to tn

2.1 LGAN(t'i):=LSTM_GAN_MODEL(ti)

2. 2 Calculate RS:= | ti-LGAN(t'i)| /*RS= Reconstruction Score*/

3. Sort RS

4. min_RS:=min(RS), max_RS=max(RS)
/*min_RS & max_RS- minimum & maximum value of RS */

5. calculate mid:=min_RS/max_RS

6. For each ti: t1 to tn
   6.1 Calculate NS[]:=RS[]-mid /* NS= normalized RS score */
   6.2 Calculate ReS[]:=1-NS[] /*ReS= Regularity Score */

7. Calculate Threshold:
   7.1 Median(m):=statistics.median(ReS[])
   7.2 Standard Deviation(sd):=statistics.stdev(ReS[])
   7.3 Threshold(thr):=m-sd

8. For each frame ti:t1 to tn:
   8.1 If ReS[i]<thr, classify ti as Anomaly
   8.2 If ReS[i]>=thr, classify ti as Normal

9. Mark each frame ti and Anomaly or Normal

10. End

Testing Time Complexity: O(H*W+S+T)

H=Height of frame, W=weight of frame, S=no. of steps for Score Calculation,

T=no. of steps for threshold calculation (Reviewer 1)

4. Experimental Results:

To test our Architecture, we conducted experiments on several benchmark databases, namely USCD Ped1, Ped2 and Avenue datasets.

4.1. Data Sets:

A. CUHK Avenue Dataset:

CUHK Avenue dataset has 21 test videos and 16 training videos with a frame size of 360 x 640 pixels. The videos have a total of 30652 frames (15328 for training and 15324 for testing). There are many bizarre scenes, like people tossing objects, fleeing, and jumping on roadways.

B. UCSD Ped Dataset:

A stationary camera installed at a height and looking down on pedestrian pathways was used to collect the UCSD Anomaly Detection Dataset. The walkways had varying densities of people, from very few to many. Bikers, skateboarders, tiny carts, and pedestrians crossing a walkway or in its surrounding grass are examples of often occurring anomalies.

B.1.Ped1: footage showing crowds of people moving in both directions from and toward the camera, with some perspective distortion. Contains 34 examples of training videos(6800 frames) and 36 examples of testing videos(7200 frames) with Resolution 158X238.

B.2.Ped2: Portrays pedestrians walking parallel to the camera plane, with 16 videos for training (2550 frames) and 12 for testing videos (2010 frames) with Resolution 240X360. Video sources from 13 distinct scenarios, with varying lighting and camera perspectives, are included in the corpus. It contains 130 anomalous events total—330 video examples for training and 107 for testing. Complex, unexpected human actions, the existence of strange things, and movements in the incorrect direction are examples of anomalies.

Public Datasets	Anomalies	Anomaly Events	Duration	Resolution	Training Videos	Testing Videos	Anomalous Frames
UCSD Ped1	40	Skaters, Bikers,Small carts, Walking Sideways	5 min	238X158	34	36	4005
UCSD Ped2	12	Skaters, Bikers,Small carts, Walking Sideways	5 min	360X240	16	12	1636
CHUK Avenue	47	Throwing, Running. Loitering	30 min	640X360	16	21	3820

Table 1: Characteristics of the datasets used in this work

4.2 Implementation:

Our architecture is implemented using Python, Keras, and the TensorFlow framework. Hardware 16 GB RAM and IRIS Xe Graphics used to implement. We utilize resized frames extracted from videos,(as described in Algorithm 1). These frames are divided into batches and used to train the architecture (as described in Algorithm 2) in specific number of epochs. The trained model is then saved and loaded during the testing phase. In the testing process (as described in Algorithm 3), the test video is fed into the architecture. The generator reconstructs the frames, which are evaluated by the discriminator. Based on a predefined threshold (as described in Algorithm3), the frames are classified as either Anomalous or Normal.

4.2.1 Regularity Score: During testing, Low Regularity Score frames from the Discriminator are flagged as Anomalous, while High-scoring ones are classified as normal.

Figure 3 Normal Frames of Test Video 5

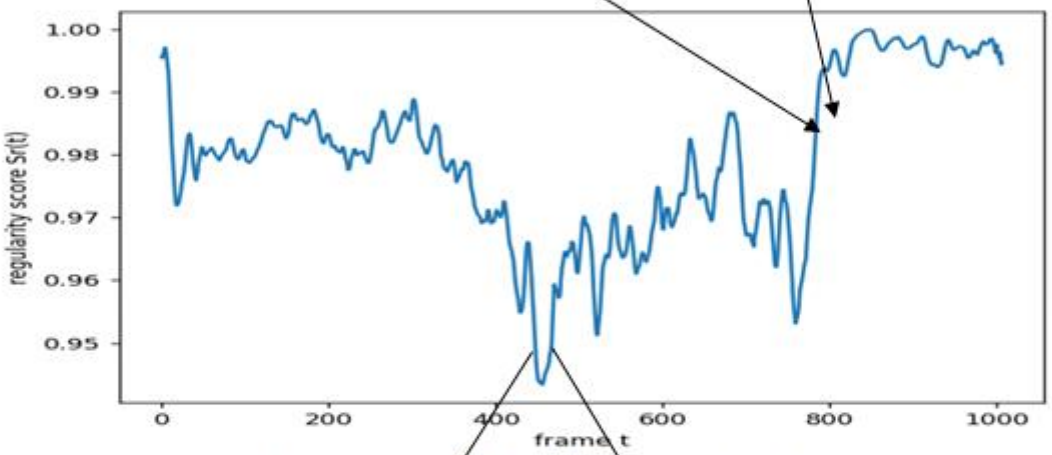

Figure 4 Regularity Score of Test Video 5

Figure 5 Anomaly Frames of Test video 5

4.2.2 Evaluation Metrics:

In the field of video anomaly detection, two commonly used anomaly detection criteria are Equal-Error Rate (EER) and Area under Curve.

i) AUC- Area Under the ROC(Receiver Operating Characteristics)Curve

ii) EER-Equal Error Rate is defined as the percentage of misclassified frames when the TPR is equal to the FNR. It is effective for the detection of video anomalies, the lower the equal error rate value is best result.

AUC is measured between 0 and 1. The better the results, the higher the AUC value execution.

4.2.3 Result Analysis:

For computing EER and AUC, first we need to compute True Positive (TP), False Positive (FP), True

Negative (TN), False Negative (FN) using the following:

TPR = TP/TP+FN ----------(1)

FPR = FP/FP+TN ----------(2)

FNR=FN/TP+FN ----------(3)

AUC=1+TPR-FPR/2 = 0.91 ----------(4)

EER = FNR+FPR/2 = 0.08995 ----------(5)

These values are depicted in Table 1 for Test Video 5

Table 2. TP, FP, TN and FN of Avenue Dataset Test Video 5 using Conv-LSTM-ED2D-GAN

Avenue Dataset Sample Test Video		Actual Test Video Frames	
		Anomaly	Normal
Predicted Test Video Frames	Anomaly	186(TP)	20(FP)
	Normal	34(FN)	766(TN)

Plot of AUC curve drawn shown in the figure. From the plotted AUC curve it is inferred that AUC 0.91 and EER 0,08995.

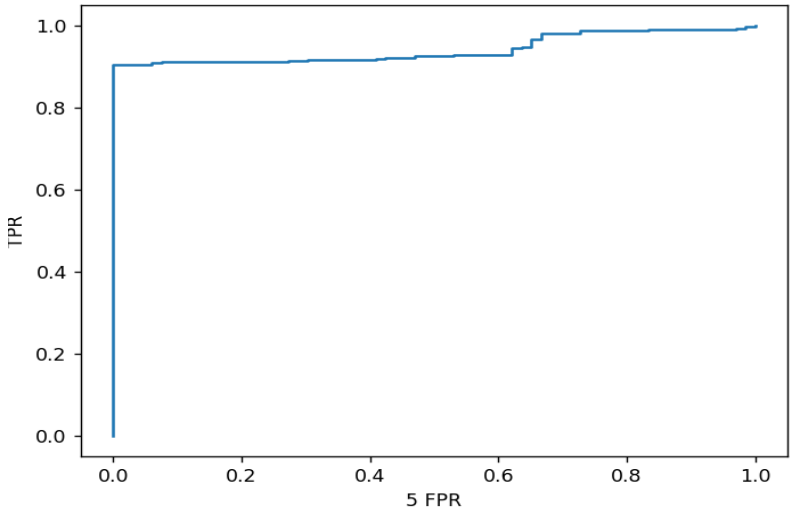

Figure 7 AUC Curve of Conv2D-LSTM-AE-GAN for Test Video 5

4.2.4 Performance Comparison with Existing Models:

The Conv2D-LSTM-AE-GAN Architecture's AUC and EER outcomes compared with existing unsupervised models are shown in Table 2. The suggested Conv2D-LSTM-AE-GAN Architecture framework achieved an AUC of 75.17 and an EER of 25.53 for the UCSD Ped1 dataset. The evaluated AUC and EER for the UCSD Ped2 dataset were 80.30 and 22.60, respectively. The suggested Conv2D-LSTM-AE-GAN Architecture framework performed best on the CHUCK Avenue dataset, with an AUC of 90.6 and an EER of 11.56.

Table 2 Comparison of Proposed Method with the Existing Methods (Unsupervised Methods)

	UCSD Ped1		**UCSD Ped2**		**Avenue**	
Unsupervised Methods	**AUC**	**EER**	**AUC**	**EER**	**AUC**	**EER**
Conv-AE [2016] [16]	81.1	27.9	90.0	21.7	70.2	25.1
MLAD [2019] [17]	82.34	23.50	99.21	2.49	52.82	38.82
SVDGAN [2021] [18]	73.26	28.75	76.98	23.46	89.82	21.55
Conv-LSTM-ED2D-GAN	75.17	25.53	80.30	22.60	90.6	11.56

5. Conclusion:

Even for anomalous frames, the suggested framework shows good reconstruction skills, but it sometimes has trouble identifying particular anomalies. For LSTM-based models like ours to activate and function properly, there must be enough sequential data. Interestingly, the suggested approach performs noticeably better on the CHUCK Avenue dataset than previous research. The proposed lighter GAN architecture outperforms state-of-the-art unsupervised anomaly detection methods while utilizing fewer parameters, thanks to the utilization of temporal blocks for improved spatiotemporal feature learning and an original SVD loss for more robust GAN learningIn the future, the system's accuracy can be increased further by incorporating a memory module or a cutting-edge 3D feature extraction algorithm.

References:

[1] Integrating prediction and reconstruction for anomaly detection. *Pattern Recognition Letters*, 129:123–130, 2020. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec. 2019.11.024.

[2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. Ganomaly: Semi- supervised anomaly detection via adversarial training. In C. V. Jawahar, Hongdong Li, Greg Mori, and Konrad Schindler, editors, Computer Vision – ACCV 2018, pages 622–637, Cham, 2019. Springer International Publishing.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017.

[4] D. V. S. Chandra. Digital image watermarking using singular value decomposition. In The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002., volume 3, pages III–III, 2002. doi: 10.1109/MWSCAS.2002.1187023.

[5] F. Chen and M. R. Jahanshahi. Nb-cnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. IEEE Transactions on In- dustrial Electronics, 65(5):4392–4400, 2018. doi: 10.1109/TIE.2017.2764844.

[6] MyeongAh Cho, Taeoh Kim, and Sangyoun Lee. Unsupervised video anomaly de- tection via flow-based generative modeling on appearance and motion latent features.CoRR,abs/2010.07524,2020.URLhttps://arxiv.org/abs/2010.07524.

[7] Yong Shean Chong and Yong Haur Tay. Abnormal event detection in videos using spatiotemporal autoencoder. In Fengyu Cong, Andrew Leung, and Qinglai Wei, editors, Advances in Neural Networks - ISNN 2017, pages 189–196, Cham, 2017. Springer International Publishing.

[8] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. Adversarial feature learning. ICLR, 2017.

[9] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for in- creased cnn performance in liver lesion classification. Neurocomputing, 321:321– 331, 2018. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.09.

[10] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[11] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. Learning temporal regularity in video sequences. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 733–742. IEEE Computer Society, jun 2016. doi: 10.1109/CVPR.2016.86.

[12] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computa- tion, 9:1735–80, 12 1997. doi: 10.1162/neco.1997.9.8.1735.

[13] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with condi- tional adversarial networks. In 2017 IEEE Conference on Computer Vision and Pat- tern Recognition (CVPR), pages 5967–5976. IEEE Computer Society, jul 2017.doi:10.1109/CVPR.2017.632.URLhttps://doi.ieeecomputersociety.org/10.1109/CVPR.2017.632.

[14] B Ravi Kiran, Dilip Mathew Thomas, and Ranjith Parakkal. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. Journal of Imaging, 4(2):36, 2018.

[15] Jin-Hwan Lee, Sung-Sik Yoon, In-Ho Kim, and Hyung-Jo Jung. Diagnosis of crack damage on structures based on image processing techniques and R-CNN using un- manned aerial vehicle (UAV). In Hoon Sohn, editor, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, volume 10598, pages 265 – 272. International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2296691. URL https://doi.org/10.1117/12.2296691.

[16] Sangmin Lee, Hak Gu Kim, and Yong Man Ro. Stan: Spatio-temporal adversarial networks for abnormal event detection. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1323–1327, 2018. doi: 10. 1109/ICASSP.2018.8462388.

[17] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36 (1):18–32, 2014. doi: 10.1109/TPAMI.2013.111.

[18] Zachary C Lipton and Subarna Tripathi. Precise recovery of latent vectors from gener ative adversarial networks. arXiv preprint arXiv:1702.04782, 2017.

[19] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection - a new baseline. pages 6536–6545, 06 2018. doi: 10.1109/CVPR. 2018.00684.

[20] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In 2013 IEEE International Conference on Computer Vision, pages 2720–2727, 2013. doi: 10.1109/ICCV.2013.338.

[21] W. Luo, W. Liu, and S. Gao. A revisit of sparse coding based anomaly detection in stacked rnn framework. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 341–349, 2017. doi: 10.1109/ICCV.2017.45.

[22] W. Luo, W. Liu, and S. Gao. Remembering history with convolutional lstm for anomaly detection. In 2017 IEEE International Conference on Multimedia and Expo (ICME), pages 439–444, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society. doi: 10.1109/ICME.2017.8019325. URL https://doi.ieeecomputersociety.org/10.1109/ICME.2017.8019325.

[23] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly de- tection in crowded scenes. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1975–1981. IEEE, 2010.

[24] J. Medel. Anomaly detection using predictive convolutional long short-term memory units. 2016.

[25] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. CoRR, abs/1411.1784, 2014. URL http://arxiv.org/abs/1411.1784.

[26] Marc Moonen, Paul Van Dooren, and Joos Vandewalle. A singular value decomposi- tion updating algorithm for subspace tracking. SIAM Journal on Matrix Analysis and Applications, 13(4):1015–1038, 1992.

[27] Trong Nguyen Nguyen and Jean Meunier. Anomaly detection in video sequence with appearance-motion correspondence. In 2019 IEEE/CVF (ICCV), pages 1273–1283, 2019. doi: 10.1109/ICCV.2019.00136.

[28] Viorica Patraucean, Ankur Handa, and Roberto Cipolla. Spatio-temporal video au- toencoder with differentiable memory. CoRR, abs/1511.06309, 2015. doi: https://doi.org/10.17863/CAM.26485.

[29] A. Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learn- ing with deep convolutional generative adversarial networks. CoRR, abs/1511.06434, 2016.

[30] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regaz- zoni, and Nicu Sebe. Abnormal event detection in videos using generative adversarial nets. In 2017 IEEE International Conference on Image Processing (ICIP), pages 1577– 1581, 2017. doi: 10.1109/ICIP.2017.8296547.

[31] Rowayda A Sadek. Svd based image processing applications: state of the art, contri- butions and research challenges. arXiv preprint arXiv:1211.7102, 2012.

[32] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved

techniques for training gans. In Proceedings of the 30th In- ternational Conference on Neural Information Processing Systems, NIPS'16, page 2234–2242, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

[33] Dinesh Jackson Samuel and Fabio Cuzzolin. Unsupervised anomaly detection for a smart autonomous robotic assistant surgeon (saras) using a deep residual autoencoder. IEEE Robotics and Automation Letters, 6(4):7256–7261, 2021. doi: 10.1109/LRA. 2021.3097244.

[34] T. Schlegl, Philipp Seeböck, S. Waldstein, G. Langs, and U. Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis, 54:30–44, 2019.

[35] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ipek Oguz, Pew-Thian Yap, and Dinggang Shen. Unsupervised anomaly detection with generative adversarial net- works to guide marker discovery. In Information Processing in Medical Imaging, pages 146–157, Cham, 2017. Springer International Publishing.

[36] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang- chun Woo. Convolutional LSTM network: A machine learning approach for precipi- tation nowcasting. In Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, pages 802–810, 2015.

[37] Dinesh Singh and C. Krishna Mohan. Graph formulation of video activities for abnormal activity recognition. Pattern Recogn., 65(C):265–272, May 2017. ISSN 0031- 3203. doi: 10.1016/j.patcog.2017.01.001. URL https://doi.org/10.1016/ j.patcog.2017.01.001.

[38] Angela A. Sodemann, Matthew P. Ross, and Brett J. Borghetti. A review of anomaly detection in automated surveillance. IEEE Transactions on Systems, Man and Cyber- netics Part C: Applications and Reviews, 42(6):1257–1272, Deceber 2012. ISSN 1094-6977. doi: 10.1109/TSMCC.2012.2215319.

[39] H. Song, C. Sun, X. Wu, M. Chen, and Y. Jia. Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. IEEE Transactions on Multimedia, 22(8):2138–2148, 2020. doi: 10.1109/TMM.2019.2950530.

[40] W. Sultani, C. Chen, and M. Shah. Real-world anomaly detection in surveillance videos. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6479–6488, 2018. doi: 10.1109/CVPR.2018.00678.

[41] Hung Vu, Tu Dinh Nguyen, Trung Le, Wei Luo, and Dinh Phung. Robust anomaly detection in videos using multilevel representations. In Pascal Van Hentenryck and Zhi-Hua Zhou, editors, Proceedings of AAAI19-Thirty-Third AAAI conference on Ar- tificial Intelligence, number 1 in Proceedings of the AAAI Conference on Artificial Intelligence, pages 5216–5223, United States of America, 2019. Association for the Advancement of Artificial Intelligence (AAAI). doi: 10.1609/aaai.v33i01.33015216. URL https://aaai.org/Conferences/AAAI-19/. AAAI Conference on Artificial Intelligence 2019, AAAI 2019 ; Conference date: 27-01-2019 Through 01- 02-2019.

[42] Xuanzhao Wang, Zhengping Che, Bo Jiang, Ning Xiao, Ke Yang, Jian Tang, Jieping Ye, Jingyu Wang, and Qi Qi. Robust unsupervised video anomaly detection by multi- path frame prediction. IEEE Transactions on Neural Networks and Learning Systems, pages 1–12, 2021. doi: 10.1109/TNNLS.2021.3083152.

[43] D. Weimer, B. Scholz-Reiter, and M. Shpitalni. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. Cirp Annals-manufacturing Technology, 65:417–420, 2016.

[44] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ra- maseshan Chandrasekhar. Efficient gan-based anomaly detection, 2018. URL https://openreview.net/forum?id=BkXADmJDM.

[45] Y. Zhang, H. Lu, L. Zhang, and X. Ruan. Combining motion and appearance cues for anomaly detection. Pattern Recognit., 51:443–452, 2016.

[46] B. Zhao, Li Fei-Fei, and E. Xing. Online detection of unusual events in videos via dynamic sparse coding. CVPR 2011, pages 3313–3320, 2011.

[47] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. Proceedings of the AAAI Conference on Artificial Intelligence, 34 (07):13001–13008, Apr. 2020. doi: 10.1609/aaai.v34i07.7000. URL https://ojs. aaai.org/index.php/AAAI/article/view/7000.

[48] Abdelhafid Berroukham, Khalid Housni, Mohammed Lahraichi, Idir Boulfrifi, Deep learning-based methods for anomaly detection in video surveillance: a review, Feb. 2023, [Online]. Available: https://www.researchgate.net/publication/365243648_Deep _learning-based_methods_for_anomaly_detection_in_video_surveillance_a_review.

[49] Amnah Aldayri, Waleed Albattah, Taxonomy of Anomaly Detection Techniques in Crowd Scenes, Aug. 2022, [Online]. Available: https://www.mdpi.com/1424- 8220/22/16/6080.

[50] Kallepalli Rohit Kumar, Nisarg Gandhewar, Anomaly Detection in Surveillance System Using Machine Learning Techniques – A Review, 2022, [Online]. Available: https://www.semanticscholar.org/paper/Anomaly-Detection-In-Surveillance-SystemUsing-A-Kumar-Gandhewar/2673e050034c07bea00c4513ab93f7a1ec3a0a87. (Reviewer 3 comments)

[51] Karan Thakkar, Kuldeep Kadiya, Mr. Jigar Chauhan, Anomaly Detection in Surveillance Video, 2021, [Online]. Available: https://www.semanticscholar.org/paper/ ANOMALY-DETECTION-IN-SURVEILLANCE-VIDEO-Thakkar-Kadiya/32248d2a b0430dd21646d00a6cd17549cd985e52.

[52] Dinesh Jackson Samuel Fabio Cuzzolin Faculty of Technology, Design and Environment Visual Artificial Intelligence Laboratory Oxford Brookes University Oxford, " SVD-GAN for Real-Time Unsupervised Video Anomaly Detection".

[53] Moses, M. B., Nithya, S. E. & Parameswari, M. (2022). Internet of Things and Geographical Information System based Monitoring and Mapping of Real Time Water Quality System. International Journal of Environmental Sciences, 8(1), 27-36. https://www.theaspd.com/resources/3.%20Water%20Quality%20Monitoring%20Paper.pdf