ENHANCING VIDEO SURVEILLANCE THROUGH REAL-TIME ANOMALY DETECTION AND CONTINUOUS LEARNING WITH STEP INCREMENTAL LEARNER (SIL)

[1]Naresh K, [2]Dr.G.Thippanna, [3]Dr. G Venkata Rami Reddy

[1]Research Scholar Department of Computer Science and Engineering Niilm University Haryana.
[1]kuna48@gmail.com
[2]Associate professor Department of Computer Science and Engineering Niilm University Haryana
[2]gt.pana2012@gmail.com
[3]Professor Department of Information Technology JNT University Hyderabad
[3]gvr_reddy@yahoo.co.in

Abstract:

Video surveillance is indispensable in contemporary business and urban landscapes, playing a pivotal role in development, safety, and security. However, the exponential growth of modern environments and the widespread deployment of surveillance systems have rendered manual monitoring impractical. To address this challenge, machine learning techniques, particularly anomaly detection, have gained prominence.Anomaly detection is crucial in video surveillance, complementing object and activity detection. Yet, it is fraught with challenges owing to the unpredictability of anomalies and the ambiguity in distinguishing them from normal behavior. Moreover, normal behavior evolves over time, compounding the complexity of the problem.This study introduces the Step Incremental Learner (SIL) system, a real-time solution that leverages spatial and temporal characteristics to detect anomalies. SIL amalgamates systematic learning and active learning techniques, continually adapting to the evolving landscape of normal behavior.The computational model employs a spatiotemporal autoencoder to represent typical behavior, featuring Convolutional Neural Network (ConvNet) and Convolutional Long Short-Term Memory (ConvLSTM) layers for spatial and temporal learning. Anomaly detection hinges on reconstruction error, with predefined thresholds for anomaly identification.Active learning, coupled with fuzzy aggregation, empowers SIL to accommodate changing environments and assimilate human observer feedback. This dynamic mechanism ensures that SIL continuously enhances its understanding of normalcy.The study showcases compelling experimental results, affirming SIL's superiority over alternative methods. Notably, SIL consistently outperforms benchmark approaches, particularly on the challenging Ped 2 dataset. The research underscores SIL's efficacy across diverse surveillance scenarios, including the detection of bicycles, cars, pedestrians, and crowds.

Keywords: Video Surveillance, Anomaly Detection, Spatiotemporal Modeling, Machine Learning, Active Learning, Fuzzy Aggregation

DOI: 10.31838/ecb/2023.12.5.509

1. Introduction

Video surveillance has become an integral component of contemporary business and urban environments, playing a crucial role in their development, functioning, and sustainability. Video surveillance systems contribute significantly to various aspects such as safety, security, organization, monitoring of individuals, processes, and activities. As industrial sectors adopt advanced technologies, delve into the cyber world, and adopt energy-efficient layouts, the landscape becomes increasingly complex. Moreover, urban areas are experiencing a surge in population, leading to the proliferation of multi-story structures, higher pedestrian and vehicular traffic volumes.

The rapid expansion of modern environments, both horizontally and vertically, has led to a substantial increase in the deployment of visual security systems like Closed-Circuit Television (CCTVs). However, manually monitoring and analyzing every frame of recorded footage is impractical. To address this challenge, machine learning methods have been categorized into object detection, activity detection, and anomaly detection.

Anomaly detection is a vital component of video surveillance systems, complementing object and activity detection. Detecting anomalies is particularly challenging because these deviations from normal behavior are often unpredictable, posing difficulties for human observers as well. Anomaly detection refers to identifying any behavior that significantly deviates from what is considered normal or expected.Because of the computational complexity and the processing cost of video data, which must take into account both the spatial and temporal aspects, anomaly identification can be a

challenging task. In addition, the line that separates behavior that is typical from behavior that is unusual is typically blurry, making it difficult to differentiate between the two. Typical conduct that shifts over the course of time and in reaction to changing aspects of the surrounding environment adds still another degree of complexity.

Despite efforts to address the limits of computing complexity and video processing overheads, the problem of normal behavior changing over time as a result of temporal factors and time remains largely unexplored in the existing body of work. This is the case despite the fact that such a problem exists. The authors of this study want to investigate the question of why and how normative behaviors shift over the course of history.

The Step Incremental Learner (SIL) approach is one that has been presented as a means of getting around these challenges. The SIL is able to detect anomalies in real time by looking at the typical behavior's spatial and temporal features in detail. It adheres to the tenets of active learning and structured learning procedures in its instructional design. In order to differentiate between normal and pathological behaviors, SIL first acquires a foundational knowledge based on the data it has access to. In addition to this, it adapts itself automatically to the changing conditions in order to provide accurate anomaly detection.

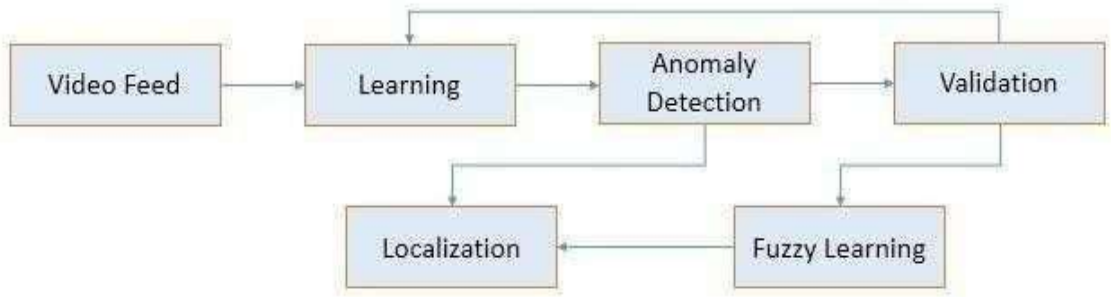

Figure 1: Overview of proposed SIL system

Figure 1 provides a high-level overview of the SIL classification system. At first, the live feed from the surveillance cameras is used to feed data into a spatiotemporal model so that it may be used to forecast normal behavior. This may be done in order to identify abnormal activity. After that, the trained model is utilized to identify and localize outliers inside the time window that was previously determined. During the phase of the anomaly identification process known as fuzzy aggregation, the input and expertise of human analysts is leveraged in order to increase the accuracy of the process. This feedback loop is essential to the effectiveness of SIL in its ability to detect anomalies since it both accelerates the learning process and guarantees that the baseline behavior model is always accurate.

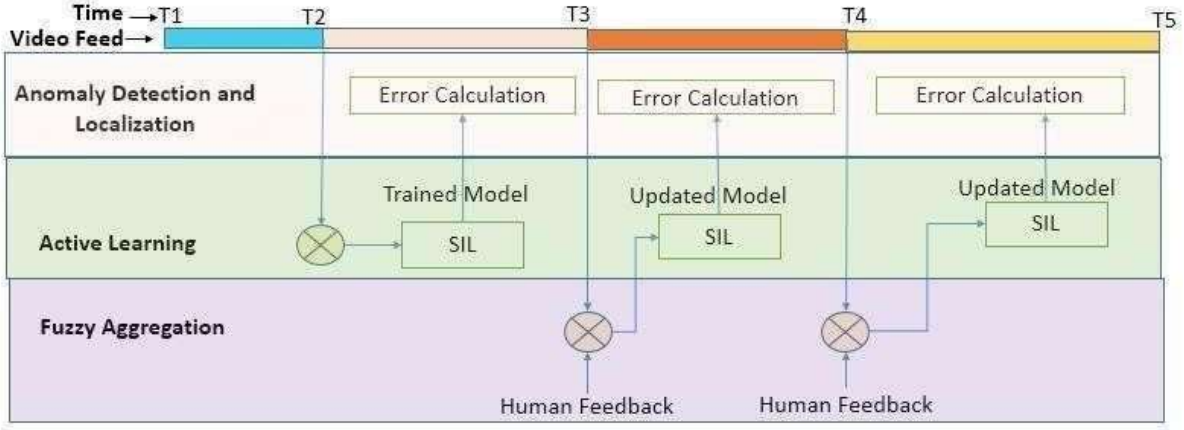

Figure 2: Functional system overview

The following presents a computational model designed for the identification of anomalies in video surveillance. The training video stream, denoted as Xtrain, comprises a series of video frames illustrating typical behavior within a specific camera perspective. In the real-world scenario, R encompasses the entire video sequence from that camera perspective.

During the testing phase, a video feed labeled as Xtest is utilized, encompassing both normal and abnormal video frames within R. The objective is to leverage the knowledge gained from Xtrain to construct a representation of typical behavior, subsequently applying this model to Xtest in order to detect and categorize abnormalities.In contrast to previous research, the SIL technique incorporates two significant improvements: (i) the ability to continually update its knowledge based on incoming streaming content, and (ii) a mechanism for incorporating feedback from human observers when anomalies are detected, facilitating the continuous learning and refinement of the SIL system.

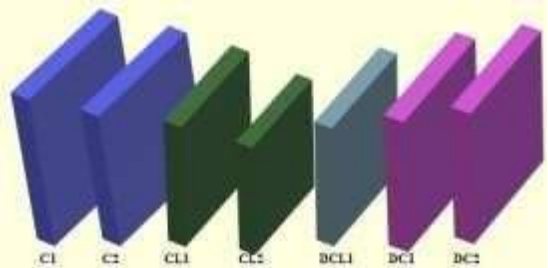

Figure 3: auto encoder architecture

2. Learning with Spatiotemporal Dimensions

The algorithm that is utilized for monitoring cameras has as its end aim the construction of a spatial-temporal model of behaviors that are acceptable and permitted based on the information provided by Xtrain. In order to accomplish this, the SIL model makes use of a spatiotemporal autoencoder in order to record the visual features of the input video stream. Autoencoders rely on unsupervised backpropagation to better match target values with input data so that they can minimize the amount of mistakes that occur during the reconstruction process. Convolutional Neural Network (ConvNet) layers are used to comprehend spatial characteristics, while Convolutional Long Short-Term Memory (ConvLSTM) layers are used to grasp temporal features, and together they make up the spatiotemporal autoencoder in the proposed design.

First Layer (the Input): In this layer, the input is given some preliminary processing. The pre-processing stage helps the model learn more effectively by increasing its potential to do so. One frame at a time, the data from an input stream is reconstructed. Following that, it will be in grayscale. Massive volumes of information are compressed as a result of the process. The number of pixels has been cut down to a pitiful 224 by 224. This is then normalized between zero and one using the range zero to one. The input consists of sequential frames, which are used to form a temporal cuboid with a length of T. Longer-duration motion can be accounted for by increasing the size of the sliding window that is used to govern the temporal cuboid; however, doing so results in an exponentially greater amount of time being required for the model to converge.

Table I is a listing of the filters and kernel sizes that are used by the SIL model.

Table 1: Spatiotemporal Auto encoder Architecture

ID	Input Tensor	Operations	Output Tensor
C1	T*224*224*1	CV-F:128;K:27*27;S:4	T*56*56*128
C2	T*56*56*128	CV;F:64;K:13*13;S:2	T*28*28*64
CL1	T*28*28*64	CL;F:64;K:3*3	T*28*28*64
CL2	T*28*28*64	CL;F:32;K:3*3	T*28*28*64
DCL1	T*28*28*32	CL;F:64;K:3*3	T*28*28*32
DC1	T*28*28*64	DCV;F:64;K:13*13;S:2	T*56*56*128
DC2	T*56*56*128	DCV;F:128;K:27*27;S:4	T*224*224*1

In a recurrent neural network (RNN), the layers known as convolution long short-term memory (ConvLSTM) are responsible for processing input sequences by storing information regarding the data in question in a memory.Long short-term memory (LSTM) units make the RNN's generic building blocks, also known as neurons, more effective. A cell, an input gate, an output gate, and a forget gate are the components that make up an LSTM unit.The input gate is

responsible for determining the acceptable range for an input value. The output gate determines how long the currently-selected input value is utilized in the computation that activates the unit, whereas the forget gate determines how long the values from earlier time steps are stored in the unit. Both gates are controlled by the same set of inputs and outputs. The information that is saved in the cell can be maintained for an endless amount of time.

Despite the fact that it was designed for modeling long-range temporal correlations and is commonly used for this purpose, the LSTM has limitations when it comes to dealing with geographical data because it does not preserve spatial information in the transitions between its states. However, in order to properly detect anomalies, it is necessary to acquire the temporal regularity of the surveillance video stream while simultaneously preserving its spatial organization. As a result, we use a variation of LSTM known as convolutional LSTM (ConvLSTM), which incorporates convolution structures into both the input-to-state transitions as well as the state-to-state transitions. The ConvLSTM is able to work around this constraint because it constructs its inputs, hidden states, gates, and cell outputs as three-dimensional tensors. The third dimension represents the spatial dimension, which enables the network to function normally despite the restriction. In place of traditional matrix operations, it makes use of the convolution operator for both its inputs and its gates. In order to facilitate the extraction of spatiotemporal characteristics from input frame sequences using the ConvLSTM, some modifications have been made. These equations serve as a representation of the ConvLSTM model.

3. Anomaly Detection and Localization

The SIL model offers a means to obtain a pixel-level assessment of the normalcy of input video data. However, because the training phase lacks samples of such images, the trained autoencoder cannot accurately reproduce anomalous or previously unseen sequences. This phenomenon is harnessed to analyze and detect irregularities within the input footage.

The computation of the reconstruction error (E) involves taking the square root of the sum of squared vector values, as illustrated in the equation below. Here, X represents the input temporal cuboid, X' is the regenerated temporal cuboid, T denotes the time window, w stands for width, and h signifies the height of the video frame.

The reconstruction error serves as a metric to assess the abnormality of each temporal cuboid. To distinguish between normal and abnormal behavior, an anomaly threshold (Θ) is established. In real-world video surveillance applications, human observers may select an appropriate threshold based on the desired sensitivity level for the surveillance task. A lower Θ value would result in heightened responsiveness, potentially leading to more frequent alerts. Conversely, a higher Θ value would reduce sensitivity, potentially causing the system to overlook subtle anomalies within the surveillance area.

Additionally, we introduce a temporal threshold (τ), defined as the minimum number of consecutive video frames exceeding τ required to identify an event as an anomaly. This threshold helps prevent false-positive anomaly alerts triggered by abrupt changes in the surveillance video stream, such as occlusion, motion blur, or intense lighting fluctuations. Figure 4 provides an overview of the anomaly detection approach based on reconstruction error.

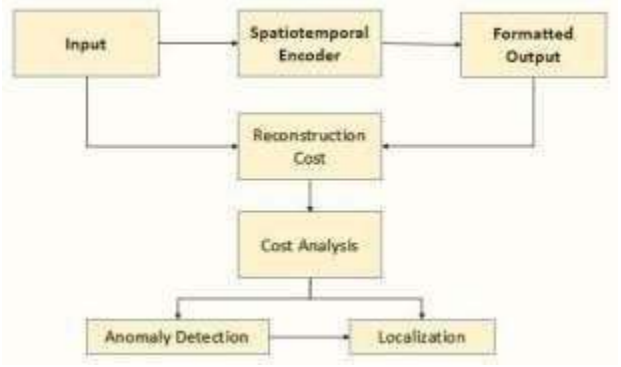

Figure 4: Anomaly Identification

4. Active Learning with Fuzzy Aggregation

In the context of real-world video surveillance scenarios, active learning plays a pivotal role in facilitating the detection of anomalies within ever-evolving situations. Our objective is to train the learning model to recognize acceptable typical behavior, as determined by the anomaly detection process using the previously outlined deep learning model. However, in dynamic environments characterized by unexpected new normal behaviors and the transformation of existing abnormal behaviors into normal ones, it is crucial for the detection system to adapt and identify these novel instances. SIL addresses this challenge by leveraging fuzzy aggregation to continually train the learning model to recognize unknown or newly emerging normal behaviors relevant to the current surveillance scenario.

This approach draws inspiration from the human brain's ability to construct a foundational knowledge base that continuously evolves as new information becomes available. Initially, SIL is trained using pre-identified typical behaviors within the monitoring environment and subsequently employed to detect anomalies. If a video frame is flagged as an anomaly, it signifies that the training error of the input cuboid exceeds the anomaly threshold. These flagged frames are then subjected to validation by a human observer.

The input from human observers serves the purpose of proactively feeding the learning model with continuously evolving normal behavior. Consequently, if a detected video frame is determined to be erroneous by the human observer, it is classified as 'normal,' and this information is integrated into the ongoing learning process. Following human observer input, the video frames labeled as 'normal' are used to continually train and enhance the SIL model's understanding of the concept of normalcy.

As depicted in the illustration, the SIL model undergoes continuous updates through two primary mechanisms: (i) the incorporation of spatiotemporal data from the continuously received surveillance video stream and (ii) active human observer input regarding observed anomalies. To ensure consistency and stability throughout the learning iterations, the SIL model's continuous learning is augmented by fuzzy aggregation of video frames.

During the detection phase, all video frames are assigned a fuzzy measure denoted as 'g,' which depends on their reconstruction error. These frames are then categorized into a limited number (n) of sets based on their 'g' values. The method subsequently selects the 'k' video frame cuboids with the highest 'g' values from each set of fuzzy measurements to train the SIL model during the continuous learning phase. The values of 'k' and 'n' are initially determined based on the duration of the video surveillance stream utilized for continuous learning, as defined by the equation specifying the scene selection for continuous training.

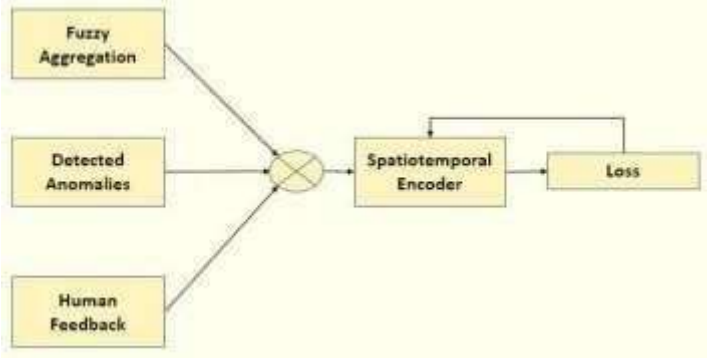

Figure 5: Active Learning using fuzzy aggregation

The dataset used for regular training iterations currently includes two key components: (i) human-verified false positive detections and (ii) temporal cuboids extracted from normal behavior using fuzzy aggregation. This combination ensures that the detection model remains updated to capture novel normal behaviors while maintaining stability in recognizing previously known normal behaviors. The effectiveness of this fuzzy aggregation technique has been demonstrated in continuous learning scenarios for IoT stream mining, text mining, and video stream mining, preserving a balance between stability and adaptability.

Following the scene selection process, the SIL model undergoes continuous training using the selected representations from the input video data. This process results in an updated model that reflects the most recent expected and acceptable behaviors within the surveillance environment. The enhanced SIL model is then employed for future anomaly detection tasks.The CUHK Avenue dataset was constructed using a stationary 640x360-pixel video camera capturing pedestrian movement at City University of Hong Kong. This compilation comprises 16 training videos depicting typical human behavior and 21 videos showcasing behaviors considered abnormal. Common sidewalk behaviors observed include

littering, walking toward the camera, strolling on the grass, and discarding items.The UCSD Pedestrian Dataset, focusing on two pedestrian paths, consists of videos with dimensions of 238x158 pixels. This package includes two datasets, Ped1 and Ped2, representing various crowd scenarios ranging from sparse to dense. The training video examples in Ped1 primarily depict people walking, while the test dataset contains samples of pedestrians crossing sidewalks, grassy areas, and vehicle-related activities. Ped1 comprises 34 training video samples and 36 test videos, while Ped2 includes 16 training videos and 12 test videos.

The experimental setup is as follows:
1) Initially, the efficiency and accuracy of the proposed spatiotemporal autoencoder for anomaly detection are tested using three standard datasets.
2) Next, we evaluate the continuous learning potential of the ISTL model using the UCSD Ped1 and Ped2 datasets, reclassifying previously identified anomalous events as normal.
3) Finally, we conduct a runtime examination to demonstrate the real-time processing capabilities of our algorithm.

Preprocessing of input samples involves resizing all samples to a dimension of 224x224 pixels and normalizing pixel values to a range between 0 and 1. The choice of a temporal cuboid depth, T=8, is based on the frame rate of the training data, corresponding to approximately one-third of a second. This selection balances the capture of motion between successive frames while mitigating model complexity due to an excessive number of input cuboids. In cases where the input surveillance data has a lower frame rate, longer motion can be recorded with shorter temporal depths.

In the course of this experiment, a learning model was trained using a learning rate of 0.01, and it required 1500 iterations to complete the process. It was determined that stochastic gradient descent was the best method for optimizing the spatiotemporal autoencoder model, and mean squared error was chosen as the metric to utilize for quantifying reconstruction loss. We stopped the training when there was no longer any improvement in the loss, which prevented the model from getting overfit and allowed us to avoid this problem by using an early stopping regularization strategy. The training procedure consisted of three successive rounds, each of which split the available data in the proportions of 60%, 20%, and 20% respectively. During the active learning stage, the reconstruction error was utilized as the fuzzy measure that was utilized. A temporal threshold and an anomaly threshold were utilized in order to detect and localize any anomalies that may have been present. In order to discover the optimal parameters for each dataset that was utilized in our testing, we investigated a range that went from 1 to 9.

5. Results - Anomaly Detection

Anomaly detection was conducted using three feature-based methods and four distinct deep learning approaches. The first method, proposed by Mehran et al., focuses on identifying abnormal crowd behavior by implementing a social force model. The second method, introduced by Kim and K. Grauman, utilizes the Markov field for anomaly detection. In order to evaluate anomaly detection, the first two approaches, which use social force and Markov random field principles, are integrated into the third approach using this method. In this combination method, the terms "anomalies" and "outliers" are used interchangeably.The following is an explanation of the deep learning models that were utilized for the purpose of evaluation. First, the feed-forward model that was proposed by Hasan et al. is a convolutional model. This type of model learns attributes and classifiers simultaneously through the use of an autoencoder component.

Anomaly localization and identification in the process incorporate both the anomaly threshold and the temporal threshold. To determine the optimal λ value, a range spanning from 0 to 9 was explored for each of the three datasets. Interestingly, it was observed that the optimal λ value varied across the three datasets. This discrepancy can be primarily attributed to differences in camera positioning during image capture for each dataset. Some datasets involve close-range image capture, while others were taken with the camera positioned farther from the scene.

Dataset	Optimal AUC/EER	Anomaly Threshold(μ)	Temporal Threshold(λ)
Pad 1	67.4/19.3	0.36	5
Pad 2	92.5/7.7	0.41	9
Avenue	79.8/19.2	0.31	2

Table 2:Selection of Anomaly Threshold and Temporal Threshold

TABLE III
COMPARISON OF AUC AND EER

Model	Ped 1 AUC/EER	Ped 2 AUC/EER	Avenue AUC/EER
SF (2009)	57.5/31.0	45.6/39.0	NA
MPCCA (2009)	63.8/38.0	59.3/28.0	NA
MPCCA + SF (2010)	72.2/33.0	63.3/39.0	NA
Conv-AE (2016)	78.0/25.9	88.0/22.5	69.2/22.1
S-RBM (2017)	73.3/38.4	78.4/19.5	75.8/29.2
ConvLSTM-AE (2017)	73.5/NA	82.1/NA	72.0/NA
Unmasking (2017)	66.4/NA	84.2/NA	79.3/NA
Proposed (SIL)	79.2/18.8	94.1/7.7	72.8/21.2

Table IV: Anomaly Detection for Cycling Scenario

Dataset	Prior Active Learning	After active Learning
UCSD Ped 1	11/14	3/14
UCSD Ped 2	5/7	2/7

TABLE V
PROCESSING TIME ANALYSIS (SECONDS PER FRAME)

Process	Ped 1	Ped 2	Avenue
Pre-processing	0.0009	0.0009	0.0009
Representation	0.0298	0.0294	0.0292
Detection	0.0017	0.0017	0.0016
Localization	0.0044	0.0044	0.0039
Total	0.0369	0.0371	0.0360
FPS	~23	~23	~23

Figure 5.6 Anomaly identification using Ped 1 dataset

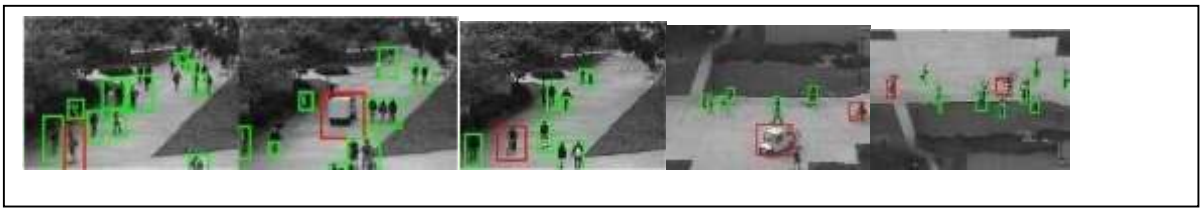

Figure 5.7: Anomaly identification using Ped 2 dataset

Figure 5.8: Anomaly identification using CUHKAvenue dataset

Conclusion

The performance of the deep learning models was evaluated using metrics such as Equal Error Rate (EER), Area Under the Curve (AUC), and frame-level Receiver Operating Characteristic (ROC) curves. In summary, our approach outperforms all manual methods and achieves comparable performance to deep learning representation-based methods when applied to the Ped 1 and Avenue datasets.Furthermore, our proposed SIL technique surpasses all similar models, including the benchmark Conv-AE (2016) approach, when tested on the Ped 2 dataset.Figure 1 illustrates a qualitative analysis of localized anomaly patches.In the UCSD Ped 1 dataset, SIL successfully identifies anomalies such as bicycles and cars on pathways, pedestrians crossing paths, gatherings of crowds, and pedestrians pushing trolleys. We specifically examined scenarios involving bicycles on pedestrian walkways from the UCSD Ped 1 and Ped 2 datasets to showcase SIL's active learning capability. As these scenarios are considered normal, all bicycle anomaly detections were categorized as such.

References

[1]B. Son, Y. Her, and J.-G. Kim, "A Design and Implementation of Forest-Fires Surveillance SystembasedonWirelessSensorNetworks for SouthKoreaMountains,"p.8,2006.

[2]M. Carvajal-Yepes et al., "A global surveillance system for crop diseases," Science, vol. 364, no.6447,pp.1237–1239,Jun.2019,doi:10.1126/science.aaw1572.

[3]K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learningtechniques for suspicious behavior recognition in intelligent surveillance system," Int. j. inf. tecnol.,vol.14,no.1,pp.397–410,Feb.2022,doi:10.1007/s41870-019-00364-0.

[4]F. T. Al-Dhief et al., "A Survey of Voice Pathology Surveillance Systems Based on Internet of Thingsand Machine Learning Algorithms," IEEE Access, vol. 8, pp. 64514–64533, 2020, doi:10.1109/ACCESS.2020.2984925.

[5]K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine LearningTechniques for Cyber Security in the Last Decade," IEEE Access, vol. 8, pp. 222310–222354, 2020,doi:10.1109/ACCESS.2020.3041951.

[6]K. A. Joshi and D. G. Thakore, "A Survey on Moving Object Detection and Tracking in VideoSurveillance System,"vol.2, no.3, p. 5, 2012.

[7]H. V. Padullaparti, P. Chirapongsananurak, M. E. Hernandez, and S. Santoso, "Analytical ApproachtoEstimateFeederAccommodationLimitsBasedonProtectionCriteria," IEEEAccess,vol.4,pp.

4066–4081,2016,doi:10.1109/ACCESS.2016.2589545.

[8]R. Sagar, R. Jhaveri, and C. Borrego, "Applications in Security and Evasions in Machine Learning: ASurvey,"Electronics,vol.9,no.1,p. 97,Jan.2020,doi:10.3390/electronics9010097.

[9]A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learningtechniques for pedestrian detection and tracking: A survey," Neurocomputing, vol. 300, pp. 17–33,Jul.2018, doi:10.1016/j.neucom.2018.01.092.

[10]W.-T. Chen, P.-Y. Chen, W.-S. Lee, and C.-F. Huang, "Design and Implementation of a Real TimeVideo Surveillance System with Wireless Sensor Networks," in VTC Spring 2008 - IEEE VehicularTechnology Conference, Marina Bay, Singapore, May 2008, pp. 218–222. doi:10.1109/VETECS.2008.57.

[11]S. N. Jyothi and K. V. Vardhan, "Design and implementation of real time security surveillancesystem using IoT," in 2016 International Conference on Communication and Electronics Systems(ICCES),Coimbatore, India,Oct.2016,pp.1–5.doi:10.1109/CESYS.2016.7890003.

[12]T. He et al., "Energy-efficient surveillance system using wireless sensor networks," inProceedings of the 2nd international conference on Mobile systems, applications, and services-MobiSYS'04, Boston,MA,USA,2004, p.270.doi:10.1145/990064.990096.

[13]T. He et al., "Energy-efficient surveillance system using wireless sensor networks," inProceedings of the 2nd international conference on Mobile systems, applications, and services-MobiSYS'04, Boston,MA,USA,2004, p.270.doi:10.1145/990064.990096.

[14]C. Clavel, T. Ehrette, and G. Richard, "Events Detection for an Audio-Based Surveillance System,"in 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 2005,pp.1306–1309.doi:10.1109/ICME.2005.1521669.

[15]G. Sreenu and M. A. Saleem Durai, "Intelligent video surveillance: a review through deeplearning techniques for crowd analysis," J Big Data, vol. 6, no. 1, p. 48, Dec. 2019, doi:10.1186/s40537-019-0212-5.

[16]S. Tiwari, H. Ramampiaro, and H. Langseth, "Machine Learning in Financial Market Surveillance:ASurvey,"IEEEAccess,vol.9,pp.159734–159754,2021,doi:10.1109/ACCESS.2021.3130843.

[17]J. Tejedor, J. Macias-Guarasa, H. Martins, J. Pastor-Graells, P. Corredera, and S. Martin-Lopez,"Machine Learning Methods for Pipeline Surveillance Systems Based on Distributed AcousticSensing: A Review," Applied Sciences, vol. 7, no. 8, p. 841, Aug. 2017, doi: 10.3390/app7080841.[18]J.Hou,C.Wu,Z.Yuan,J.Tan, Q.Wang,andY.Zhou,"ResearchofIntelligentHomeSecurity

[18].Surveillance System Based on ZigBee," in 2008 International Symposium on Intelligent InformationTechnology Application Workshops, Shanghai, China, Dec. 2008, pp. 554–557. doi:10.1109/IITA.Workshops.2008.223.

[19]Dr. E. Baraneetharan, "Role of Machine Learning Algorithms Intrusion Detection in WSNs: ASurvey," JITDW,vol.02,no.03,pp.161–173,Aug.2020, doi:10.36548/jitdw.2020.3.004.

[20]B. Lavi, M. F. Serj, and I. Ullah, "Survey on Deep Learning Techniques for Person Re-IdentificationTask." arXiv, Jul. 19, 2018. Accessed: Jul. 01, 2022. [Online]. Available:http://arxiv.org/abs/1807.05284

[21]C.N.T.SiebelandS.J.Maybank,"TheADVISORVisualSurveillanceSystem,"p.9.

[22]T.Zhang,A.Chowdhery,P.(Victor)Bahl,K.Jamieson,andS.Banerjee,"TheDesignand Implementation of a Wireless Video Surveillance System," in Proceedings of the 21st AnnualInternational Conference on Mobile Computing and Networking, Paris France, Sep. 2015, pp. 426–438.doi: 10.1145/2789168.2790123.

[23]N. H. Tandel, H. B. Prajapati, and V. K. Dabhi, "Voice Recognition and Voice Comparison usingMachine Learning Techniques: A Survey," in 2020 6th International Conference on AdvancedComputing and Communication Systems (ICACCS), Coimbatore, India, Mar. 2020, pp. 459–465. doi:10.1109/ICACCS48705.2020.9074184.

[24]Z. Guan, L. Bian, T. Shang, and J. Liu, "When Machine Learning meets Security Issues: A survey,"in 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR), Shenyang, Aug.2018,pp.158–165.doi:10.1109/IISR.2018.8535799.