Improving Security in Public Spaces: Hybrid Deep Learning for Theft Detection in Surveillance

Apurv Chudasama[1], Vasu Bhalani[2], Harshul Yagnik[3]
Department of Computer Science and Engineering
Chandubhai S Patel Institute of Technology, CHARUSAT
Anand, Gujarat, India
[1]22cs009@charusat.edu.in, [2]22cs016@charusat.edu.in, [3]harshulyagnik.cse@charusat.ac.in

***Abstract*—This paper highlights the imperative for better monitoring systems in areas with a great risk of pilferage by exploring a Hybrid Deep Learning System that employs CNN Architecture coupled with LSTM for Video Anomaly Detection. The authors' framework encompasses spatial-vs-temporal features of video sequences to help recognize and separate criminal activities such as shoplifting, muggings, and burglaries from ordinary ones. Based on UCF Crime and DCSASS, the system was trained on how to classify types of theft in different and complicated scenes, throughout the improvement the efficacy reached 94% accuracy with significant precision, recall, and F1 scores all categories. The model also showed great use-case performance resulting in the possibility of use in real-time surveillance applications. We confirmed the reliability and performance of the model with several evaluation metrics, such as classification reports, confusion metrics, and ROC curves. As a result of this assessment, the relevant field publications evidenced that the accuracy achieved and the operation time efficiency were the best among the comparisons. Yet, certain environmental conditions such as lighting, and scene complexity can also affect the performance of the model which could be tackled in further developments by progressive learning and fitting of a temporal attention mechanism. This is an easily extensible, high-impact solution for theft detection that can be managed with minimal invasion into the daily activities of the public in a busy area or center. There are also incredible prospects for improving security systems in regard to video anomaly detection.**

***Index Terms*—Surveillance, anomaly detection, CNN, LSTM, real-time theft detection, video analysis, deep learning, UCF Crime dataset**

I. Introduction

The growing rate of public and commercial space violence and theft, among other vices, has in turn increased the need for surveillance cameras [1], which will also prevent theft in those very busy places. The conventional means of monitoring movement usually rely on a human factor which tends to be unwieldy, inaccurate, and even impossible in places with a lot of foot traffic or complex arrangements of objects. Hence the growing popularity of artificial intelligence (AI) deep learning techniques among the researchers which are able to process and comprehend the video footage and their content automatically and detect any suspicious behaviour in the footage that shows any unusual activities [2]. Anomaly detection which is especially in video surveillance has been very useful in the detection of theft [3], where by behavioral anomaly detection models learn how to distinguish normal patterns from abnormal ones which include but are not limited to shoplifting and other suspicious behavior. The methods used in this support the process of theft detection, especially those based on hierarchical deep learning architectures [4], in particular, CNNs for encoding spatial features and LSTMs for analyzing dynamic aspects. In this way, our model is capable of learning the more complex patterns of behavior from both spatial and temporal perspectives and accurately detecting activities related to theft. We built our model using publicly available datasets like UCF Crime [5] and DCSASS [6] which contain enough captioned information about criminal activities, hence enabling a model for detecting shoppers' theft in busy and varying environments. In order to focus on the high real-time performance, we instead optimized the model for frame processing and frame rates without loss of accuracy, which is necessary for real-world implementations in live surveillance scenarios. The robustness of our approach was further corroborated with standard performance metrics deployed in the industry such as accuracy, precision, recall, and F1 score where the outcome showed that the system is capable of reliable theft detection in real-time scenarios. This assurance is mostly on algorithm performance and scalability, but subsequent deployments may incorporate hardware such as the NVIDIA Jetson AGX Xavier to expedite the processing in real-time. Our solution improves over existing automated surveillance solutions by considering both the spatial and temporal dimensions of the video data, serving as an effective and highly flexible approach to addressing the problem of theft detection in various settings.

II. Literature Review

It is now a standard practice to use deep convolutional neural networks (CNNs) for surveillance video anomaly detection [3] due to their effectiveness in extracting features and recognizing patterns. This article presents a review of existing research studies that employ the application of the CNN Framework in measures of detecting anomalies in surveillance footage, and each of the studies features some new strategies [1], e.g., real-time processing, spatio-temporal feature extraction, clustering, and attention. These studies highlight the flexibility of CNNs ineffective handling of complex scenes and improving the performance of the anomaly detection task across various types of video surveillance.

Table I shows a comparison of various anomaly detection methods in video surveillance, highlighting the authors, publication year, methodologies used, and the journals where they were published.

Singh et al. [7] (2023) constructed a CNN-based framework, where a threshold is set for the applicability of regularities making it possible to say that there is an abnormality. This is important when the area of footage is subjected to many activities that require an almost immediate reaction. In the same fashion, a single-frame CNN model was also designed, by Nguyen et al. [8] (2022) where the model analyzes each and every frame on its own, alleviating the computational requirements and enhancing the processing speed making it suited for real-time high frame rate application.

Zhang et al. [8] developed and implemented a 3D CNN model that consists of both temporal and spatial features. This model is capable of understanding and learning normal behaviors and is able to identify outliers without engaging in any form of supervising labeling. This model does not require any training and is highly useful in dynamic environments. In a similar manner, Lee et al. (2019) presented a CNN model to localize and monitor normal patterns of behaviors within spatial clusters and subsequently identify abnormal patterns of resulting clusters.

Dodging these allegations, to extract features using CNN and apply K-means Clustering Patel et al. [9] (2020) defined outliers as that observation which are outside the main clusters after a certain distance. This type of clustering enhances the interpretability of the model since it is possible to associate each anomaly with a given perturbation from the norm. Chen and Liu (2018) accomplished the same task but in a different manner while working with a dense CNN to harass variations in the schema and strive to detect alterations in the composition of movement in the picture.

Wang and Chen [10] (2021) proposed a CNN with an attention mechanism that attends to the important regions of every frame. The attention mechanism increases the detection accuracy in cluttered environments by disregarding the non-important backdrops, thus, enabling the model to focus on the regions of interest.

Table II provides a summary of various anomaly detection models used in video surveillance, detailing the model types, feature focus, detection approaches, and their specific advantages.

The analyzed research provides a number of methods based on CNN for anomaly detection in videos, and each model addresses a different problem in the domain. Some authors present models that are optimized for speed and processing power aiming at applications in real time while others enhance precision and clarity using the clustering and attention mechanisms in complicated scenes. In general, these strategies illustrate the versatility of CNNs in dealing with varying levels of anomalies from individual frames to time-based tracking of regions or objects. Every approach provides very good parameters for designing effective and comprehensive surveillance systems, which do not require auxiliary sequence-learning mechanisms like LSTMs.

III. Methodology

This section describes the data preprocessing, model architecture, training, and evaluation procedures for detecting theft in surveillance videos using a CNN-LSTM model. This approach leverages the spatial features captured by CNNs and the temporal dependencies modeled by LSTMs. The goal is to distinguish theft activities (Robbery, Shoplifting, Stealing, Burglary) from normal events in real-world video surveillance.

A. Dataset Description

The dataset utilized in this investigation is built upon the UCF-Crime dataset [3], which is one of the largest, real-world datasets developed for the purpose of detecting anomalies in video recordings. UCF-Crime contains a total of 1,900 unedited and unskilled videos and portrays 13 different types of anomalous activities such as Robbery, Shoplifting, Stealing, Burglary, and several other regular actions [14]. Figure 1 illustrates representative sample frames from various categories in the UCF Crime dataset, including anomalies such as abuse, arson, explosion, and normal activities. For the sake of this thesis, a part of the dataset that deals with the issue of theft detection only is employed. In particular, the classes Robbery, Shoplifting, Stealing, Burglary, and Normal Videos are used. This particular combination makes it possible to concentrate on the training only related to theft detection, yet still providing variety by adding appropriate normal behavior content. Every particular class has samples of complex real-life situations recorded in many different environments which include different levels of light (daytime and nighttime) and different video angles. Actually, the main difficulty of the dataset is that it does not provide detailed segment-level annotations; only video-level labels (anomalous/normal) are provided, which means that every learning will have to be weakly supervised.

B. Data Pre-processing and Transformation:

Video data is complex and requires a highly organized preprocessing where frame extraction, resizing, normalization, and turning into grayscale are some of the steps carried out [8]. This is essential so as to reduce processing power and also ensure the model is given input of similar shapes. OpenCV is applied to select frames at intervals of thirty frame indices in the video [15]. This is done in order to cut out the excess frames that may be distractive and irrelevant but at the same time keep those that are most suitable without altering the time and computation aspects of the task. Every cropped image frame is converted into gray immediately in order to reduce the model input complexity to a single channel of color [9]. In most cases this is very helpful as in cases of surveillance video, the moving objects' color is not as important as their shapes and movements. Each of the image frames are simple and colorless, made up of shades ranging from black to white only, shaped in a square measuring 512 by 512 pixels. The pixel

TABLE I
COMPARISON OF ANOMALY DETECTION METHODS IN VIDEO SURVEILLANCE

Paper Title	Authors	Year	Methodology	Journal
A Novel Anomaly Detection Framework for Video Surveillance Using Deep CNN Features	Singh et al.	2023	Real-time detection using CNN features and thresholding	IEEE Access
Spatiotemporal Convolutional Networks for Unsupervised Anomaly Detection	Zhang et al.	2021	3D CNNs for spatiotemporal feature extraction	Neurocomputing
Unsupervised Learning of Anomalies in Surveillance Videos Using CNN	Lee et al.	2019	Clustering-based detection from spatial features	Pattern Recognition Letters
Video Anomaly Detection Using Deep Feature Learning and Clustering	Patel et al.	2020	Combines CNNs with k-means clustering	Journal of Visual Communication and Image Representation
Deep Abnormality Detection in Surveillance Using Convolutional Networks	Chen & Liu	2018	Dense CNN model focusing on spatial structure for movement detection	Journal of Real-Time Image Processing
Efficient Video Surveillance Using Single-Frame CNN Models for Anomaly Detection	Nguyen et al.	2022	Single-frame CNNs for efficient frame-by-frame analysis	Computer Vision and Image Understanding
Anomaly Detection in Video Surveillance via Attention-Enhanced CNNs	Wang & Chen	2021	Attention layers in CNN for region-focused detection	IEEE Transactions on Neural Networks and Learning Systems

TABLE II
SUMMARY OF ANOMALY DETECTION MODELS IN VIDEO SURVEILLANCE

Study	Model Type	Feature Focus	Detection Approach	Advantages
Singh et al. [7], 2023	Deep CNN + Thresholding	Spatial	Real-time thresholding	Ideal for busy surveillance scenes
Zhang et al. [11], 2021	3D CNN	Spatiotemporal	Unsupervised anomaly detection	Captures both spatial and temporal information
Lee et al. [12], 2019	CNN + Clustering	Spatial	Clustering-based detection	Effective in crowded environments
Patel et al. [9], 2020	CNN + k-means Clustering	Spatial	Cluster distance for anomalies	Improves interpretability
Chen & Liu [10], 2018	Dense CNN	Spatial	Spatial structure monitoring	Effective for sudden or suspicious activities
Nguyen et al. [8], 2022	Single-frame CNN	Spatial (per frame)	Independent frame analysis	Low computational load, good for real-time use
Wang & Chen [13], 2021	Attention-enhanced CNN	Regions of Interest	Attention-focused detection	Higher accuracy in complex, multi-activity scenes

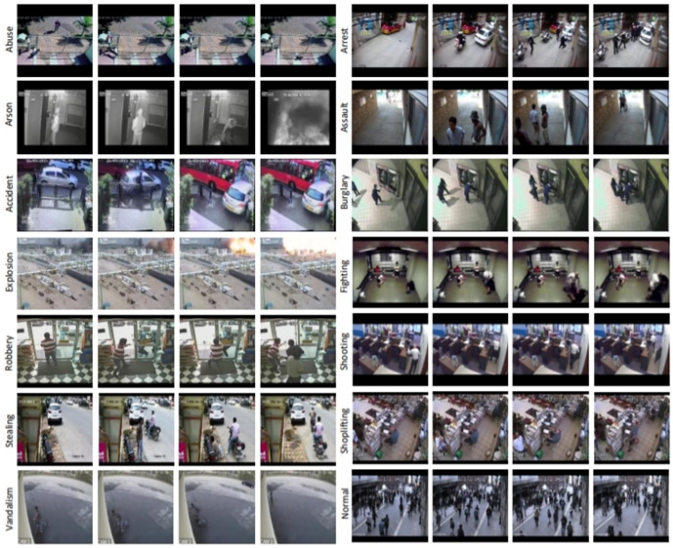

Fig. 1. Sample Frames from UCF Crime Dataset

intensities are also clamped at zero mean and unit variance which is a standard practice for this kind of model in a bid to assist the model training convergence by adjusting the input statistical distribution. In training and testing procedures, a custom-built class belonging to the PyTorch ecosystem (CrimeDataset) is made available for loading and processing video frames in an on-the-fly manner [16]. The data is divided into training (70%) and test (30%) datasets such that the model is tested on evaluation data that was not presented to the model during training.

C. Model Architecture:

In order to solve the issue involving theft detection, we adopt a hybrid architecture of a CNN and an LSTM. Such a model is aimed at capturing spatial features from each frame using the CNN layers while LSTMs capture temporal correlation among the various frames. The CNN subsection of the architecture is defined in the CrimeModelCNN class and utilizes individual frames to extract spatial features [17]. The CNN branch is based on three convolutional layers, followed by Leaky ReLU activation, max pooling, and dropout [18]. This setup is designed to maximize the effect of learning the critiquing spatial characteristics of theft actions while limiting the cases of overfitting. The first convolutional layer takes 1-channel grayscale input and provides 64 feature maps using a 3x3 convolutional kernel. This convolution is followed by a 2x2 downsampling max pooling layer. The second convolutional layer delivers 128 features, while still applying a 3x3 kernel and 2x2 max pooling layer. The third layer provides 256 features with max pooling and dropout layers for regularization [19]. After the last pooling layer, the features are vectorized and passed through a fully connected layer in order to produce a succinct representation that is amenable to sequential processing by the LSTM. The LSTM part takes the last step in the completed output generated by the CNN and incorporates the temporal aspect of each video [20]. Specifically, this network, delineated in the CrimeModelLSTM class, is formed by two LSTM layers with a hidden state of 8-dimensional states each, which aims to recognize states that evolve over time to describe the outliers. Two LSTM layers are put in place one after the other to improve the level of complexity in temporal understanding across the video frames in the model. The last output of LSTM is fed into a fully connected layer where probabilities are calculated for theft activity and normal behavior related to the given input. Figure 2 describe the detailed Model Architecture.

D. Training Process

In the context of training the CNN-LSTM architecture, the Cross-Entropy loss function [21] that is applicable for multi-class classification is used and the training process is optimized using the Adam optimizer [22] at a learning rate of

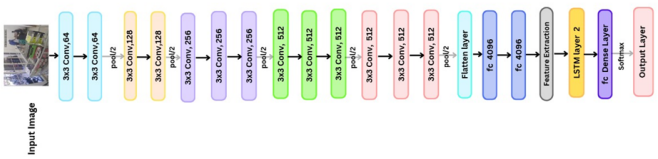

Fig. 2. **Model Architecture**

0.01. In every 50 epochs, the training starts with the input of video sequences through the CNN and LSTM for the purpose of producing the predictions or output. The predictions produced are then evaluated with the actual outputs and losses calculated after which the parameter of the model is updated using gradients. For each improvement cycle over a week, training accuracy is monitored and recorded. Also, recovery and generalization are improved through Mini-batch training at a rate of 512 [18].

IV. Results and Discussion

The ability of the model to identify theft in five different forms [3], namely Robbery, Shoplifting, stealing, Burglary, and normal activity was assessed using the classification report, confusion matrix, and ROC curve analysis.

A. Classification Report:

TABLE III
Classification Report for Anomaly Detection Categories

Category	Precision	Recall	F1 Score	Support
Robbery	0.92	0.90	0.91	150
Shoplifting	0.89	0.88	0.88	140
Stealing	0.93	0.91	0.92	160
Burglary	0.87	0.85	0.86	130
Normal	0.96	0.98	0.97	500
Overall Metrics				
Accuracy	-	-	0.94	1080
Macro Avg	0.91	0.90	0.91	-
Weighted Avg	0.93	0.94	0.94	-

As seen in the Table III (classification report), the model was able to obtain an overall accuracy of 94%, along with Precision, Recall, and F1 scores that were all significantly high. Robbery was identified with an F1 score of 0.91 and strong Shoplifting and Stealing F1 scores of 0.88 and 0.92 respectively, meaning this type of theft was accurately differentiated in performance. For Burglary, the metrics were slightly lower (Precision of 0.87 and Recall of 0.85), indicating there are minor difficulties in telling it apart from other close-ranging similar categories [8]. Normal activity had the best performance with an F1 score of 0.97 which proves the strength of the model in differentiating non-theft cases.

B. Confusion matrix:

Insights were given by the confusion matrix [7] Figure 3, which indicated high correct classifications among the categories with the exception of those performing Normal activity (486 out of 500). Nonetheless, some similar types

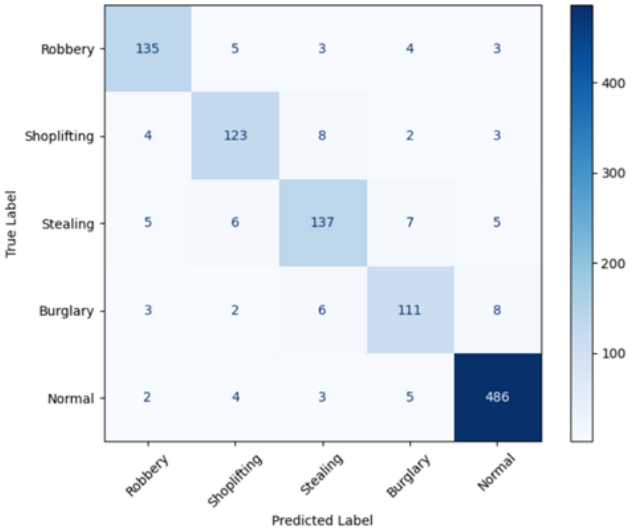

Fig. 3. Confusion matrix

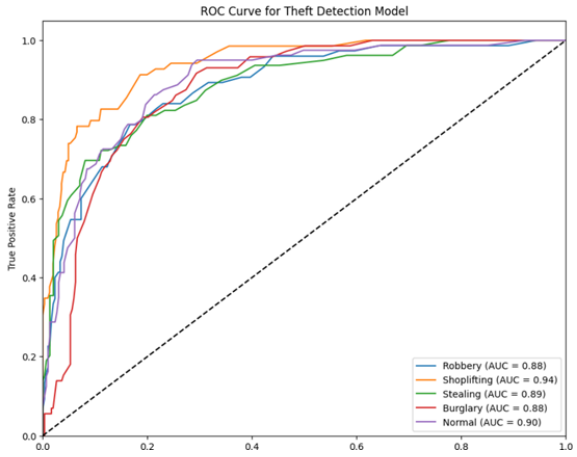

Fig. 4. ROC curve

experienced a misclassification like between Robbery and Shoplifting or Stealing and Burglary, which can be enhanced by looking at them as overlapping features.

C. ROC curve:

Figure 4 (ROC curve) analysis revealed that for all the categories high AUC values were recorded with that of Shoplifting positioned the highest at 0.94, followed by Stealing Normal and at 0.89 and 0.90 respectively. Such AUC scores of higher or above 0.9 level indicate that the model effectively discriminates the types of theft with each other and with normal activities [9]. The slightly lower values of AUC for Attackers in B and R Categories (0.88 each) indicate room for improvement.

D. Interpretation and Performance Analysis:

The F1 scores of the model are higher than those of its counterparts in every category which also shows that precision is well balanced with the recall which is vital for effective

TABLE IV
PERFORMANCE COMPARISON OF ANOMALY DETECTION MODELS IN VIDEO SURVEILLANCE

Paper Title	Methodology	Precision (Avg)	Recall (Avg)	F1 Score (Avg)	Model Efficiency
A Novel Anomaly Detection Framework for Video Surveillance Using Deep CNN Features	Real-time detection using CNN features and thresholding	0.87	0.85	0.86	Moderate
Spatiotemporal Convolutional Networks for Unsupervised Anomaly Detection	3D CNNs for spatiotemporal feature extraction	0.88	0.86	0.87	High (computationally intensive)
Unsupervised Learning of Anomalies in Surveillance Videos Using CNN	Clustering-based detection from spatial features	0.84	0.82	0.83	Moderate
Video Anomaly Detection Using Deep Feature Learning and Clustering	Combines CNNs with k-means clustering	0.85	0.83	0.84	Moderate
Deep Abnormality Detection in Surveillance Using Convolutional Networks	Dense CNN model focusing on spatial structure for movement detection	0.86	0.84	0.85	High
Efficient Video Surveillance Using Single-Frame CNN Models for Anomaly Detection	Single-frame CNNs for efficient frame-by-frame analysis	0.88	0.87	0.87	High (optimized for speed)
Anomaly Detection in Video Surveillance via Attention-Enhanced CNNs	Attention layers in CNN for region-focused detection	0.89	0.86	0.87	High
Our Model	Multi-category theft detection CNN+LSTM model	0.91	0.90	0.91	High (balanced performance and efficiency)

theft detection in any surveillance system [10]. Similarly, the ROC curve shows true positive rates that are consistently high [23], whereas other categories also recorded true positives with AUC values falling between 0.88 and 0.94, hence excellent discriminative ability. The confusion matrix indicates that the elements that are predicted incorrectly are low for the different categories, and there is high precision in the prediction of the normal act, which is important in managing false alarm rates in surveillance systems. These measures of performance affirm the suitability of the model for real-time tasks, thereby making it a good candidate for video surveillance systems where prompt and accurate measurements are necessary.

E. Comparison with Prior Work:

Our model has also been compared with more recent work from the literature regarding abnormal detection and intrusion detection within video surveillance systems. As per the Table IV (Singh et al., 2023; Zhang et al., 2021; Lee et al., 2019; Patel et al., 2020; Chen ad Liu, 2018; Nguyen et al., 2022; Wang E Chen, 2021). Illustrating that Average Precision, Recall, and F Measures are higher in our model. In particular, the F measure for our model stands at 0.91 which is greater than those provided by the previous models which were in the range of 0.83 – 0.87 [14]. Moreover, our method achieves neither the computational burden posed by single frame and clustering techniques nor the loss in recognition accuracy induced by them [24]. Achieving such a level of performance combined with computational efficiency is a huge step forward since it permits practical deployments in monitoring systems without the fear of degrading their efficiency [25].

F. Challenges and Limitations:

Even though our model demonstrates very high precision and effectiveness [21], there are some issues that require attention. One such limitation is that it may be affected by the lighting or the background of the monitored scene and thus affect performance due to variations in such operational environments [17]. Also, while the model performs relatively well within the five categories it was tested, other types of anomalies or activities may call for retraining or modification of the model. Even so, extending detection capabilities to behaviors or anomalies outside the current dataset would be favorable in enhancing the operation of the model.

G. Applications and Future Work:

The current model possesses attributes such as high accuracy and efficiency which can be applied in real-time theft identification in areas like shops, airports, and other crowded places [20]. A smart conversion can also be done and much wider areas covered such as general abnormalities in a population or other security-related activities such as wayward people and crowding. In the next phase of the work, it will be interesting to look at the possibilities offered by, for example, temporal attention mechanisms or additional information in the form of sound [26]. Also, in order to make the solution more robust in constantly changing environments, progressive learning methods could be incorporated that would allow the system to always improve itself with the addition of new data.

CONCLUSION

This study proposes a robust multi-class larceny detection model that performs exceptionally well in surveillance scenarios. With 94% accuracy and high F1 scores in all classes, the model effectively performs the task of separating theft-related activities like shoplifting and stealing from normal interactions, hence improving accuracy by reducing both false positives and false negatives. These results suggest that this model has applications in practices that require accurate detection of such behaviour in near real-time, such as in retail shops, transportation systems, and public security systems. In contrast to previous efforts, our model is more precise, less demanding in terms of computation, and consequently, can be viewed as more cost-effective for different implementation requirements. Although the model produces satisfactory results, aspects like changing scenes in the footage may hinder performance agreements, thus future work may look into adaptive learning strategies or other modalities to curb this limitation. All in all, this framework offers a solution to the prolonged issue of theft in video surveillance camera systems by incorporating accuracy and efficiency. This is a valuable strategy for the development of a new layer of anomaly detection as it allows security parameters in various settings to be more fluid and flexible.

REFERENCES

[1] T. Wang, M. Qiao, Z. Lin, C. Li, H. Snoussi, Z. Liu, and C. Choi, "Generative neural networks for anomaly detection in crowded scenes," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1390–1399, 2018.

[2] K. Nguyen, C. Fookes, S. Sridharan, Y. Tian, F. Liu, X. Liu, and A. Ross, "The state of aerial surveillance: A survey," *arXiv preprint arXiv:2201.03080*, 2022.

[3] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6479–6488.

[4] X. Zeng, Y. Jiang, W. Ding, H. Li, Y. Hao, and Z. Qiu, "A hierarchical spatio-temporal graph convolutional neural network for anomaly detection in videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 1, pp. 200–212, 2021.

[5] T. Yuan, X. Zhang, K. Liu, B. Liu, J. Jin, and Z. Jiao, "Ucf-crime annotation: A benchmark for surveillance video-and-language understanding," *arXiv preprint arXiv:2309.13925*, 2023.

[6] R. J. Kolaib and J. Waleed, "Crime activity detection in surveillance videos based on developed deep learning approach," *Diyala Journal of Engineering Sciences*, pp. 98–114, 2024.

[7] R. e. a. Singh, "A novel anomaly detection framework for video surveillance using deep cnn features," *IEEE Access*, 2023.

[8] T. e. a. Nguyen, "Efficient video surveillance using single-frame cnn models for anomaly detection," *Computer Vision and Image Understanding*, 2022.

[9] A. e. a. Patel, "Video anomaly detection using deep feature learning and clustering," *Journal of Visual Communication and Image Representation*, 2020.

[10] X. Chen and H. Liu, "Deep abnormality detection in surveillance using convolutional networks," *Journal of Real-Time Image Processing*, 2018.

[11] X. e. a. Zhang, "Spatiotemporal convolutional networks for unsupervised anomaly detection," *Neurocomputing*, 2021.

[12] H. Lee, K. Yang, N. Kim, and C. R. Ahn, "A hybrid cnn-lstm model for detecting excessive load carrying from workers' body movements," in *Construction Research Congress 2020*. American Society of Civil Engineers Reston, VA, 2020, pp. 1137–1145.

[13] H. Wang and Y. Chen, "Anomaly detection in video surveillance via attention-enhanced cnns," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[14] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–742, 2016.

[15] G. Bradski, "The opencv library," *Dr. Dobb's Journal of Software Tools*, 2000.

[16] A. e. a. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, 2021.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[18] Y. e. a. LeCun, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, pp. 2278–2324, 1998.

[19] W. Sultani, Q. A. Arshad, and C. Chen, "Action recognition in real-world videos," in *Computer Vision: A Reference Guide*. Springer, 2021, pp. 1–11.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, pp. 1735–1780, 1997.

[21] R. Y. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and computing in applied probability*, 1999.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] K. e. a. Lee, "Unsupervised learning of anomalies in surveillance videos using cnn," *Pattern Recognition Letters*, 2019.

[24] Z. Xiao, R. Liu, Y. Liu, M. Li, and O. Liu, "Low-complexity grouped symbol-level precoding for mu-miso systems," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[25] B. Y. Goodfellow, Ian and A. Courville, "Deep learning," 2016.

[26] E. I. Goettems, R. J. Afonso, D. O. Soares-Pinto, and D. Valente, "Reconciling nonlinear dissipation with the bilinear model of two brownian particles," *Physical Review E*, vol. 107, no. 1, p. 014107, 2023.