IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Anomaly Detect Net: A Deep Learning Framework for Anomaly Detection in Video Data

[1] Ganga B, [2] N Navya Shree, [3] Dr. Lata B T, [4] Dr. Venugopal K R

[1] Research Scholar, [2]M-tech Student, [3] Associate Professor, [4] Former Vice-Chancellor of BU

[1 2 3 4]Department of CSE,

[1 2 3 4] University of Visvesvaraya College of Engineering, Bengaluru University, Bengaluru, India

Abstract: Anomaly detection is a critical task with applicationsspanning various domains, including manufacturing, healthcare, and security and surveillance. This study introduces a comprehensive deep learning-based anomaly detection frameworkfor video data using PyTorch. The architecture, built upon the "FNN" (Feedforward Neural Network) is "AnomalyDetectNet", employs multiple layers for feature extraction and classification, incorporating innovative techniques like Multiple Instance Learning (MIL) and the hyperbolic tangent (tanh) activation function to distinguish between regular and anomalous data effectively. Evaluation of benchmark datasets, particularly the UCF-Crime dataset, through metrics like the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), demonstrates the model's proficiency in anomaly detection. Comparative analysis against a baseline approach reveals notable improvements, with our "AnomalyDetectNet" model achieving an accuracy of 84.84%, surpassing the baseline's 84.30%. This work contributesa potent tool for real-world applications in surveillance, security, and anomaly detection across diverse scenarios, advancing the field of anomaly detection research.

Index Terms **- Feedforward neural network (FNN), Multiple instance learning (MIL), AnomalyDetectNet, hyperbolic tangent (tanh).**

I. INTRODUCTION

The identification of anomalies is a crucial and pervasive task in numerous real-world industries, such as banking, security, manufacturing, and healthcare. Finding unusual or unexpected patterns in a dataset that deviate significantly fromthe standard deviation is its primary objective. This work hasa big impact since it finds threats and hidden information in different data streams. In the realm of security and surveillance, anomaly detection is crucial because it acts as a firstline of defence, drawing attention to questionable activitiesand potential threats that could otherwise go unnoticed. Itis becoming increasingly obvious that accurate and scalable anomaly detection algorithms are needed as data becomesmore complex every day, ranging from live video streams to high-dimensional sensor data.

There have been many different procedures and techniques established, each with its own set of advantages and disadvantages. This study conducts a comparison analysis, looking at the outcomes of several benchmark approaches applied to a wide variety of datasets.

The study of anomaly detection methods entails an understanding of the various difficulties posed by diverse datastreams, ranging from video surveillance to sensor data inmanufacturing environments. An interdisciplinary approachtakes into account the nuances of anomaly detection acrosssectors and the need for tailored solutions that may be madeto match specific environments. The project's objective isto address the inherent difficulties in accurately identifyinganomalies in dynamic and complicated datasets, which necessitates the development of algorithms that can swiftly andaccurately identify anomalies across a range of data formats.The study incorporates ideas from [1,2,6] that have addressed the anomaly detection problem. The research is framedand the state of the art is assessed using the findings andtechniques used in these studies. In doing so, an attempt is made to draw attention to the intrinsic difficulties of thework at hand as well as the adaptability and limits of current methodologies across many areas.

The widely recognized UCF-Crime dataset is used, serving as a standard for anomaly detection assessment. Performance metrics, including the Area Under the Receiver OperatingCharacteristic Curve (AUC-ROC), are utilized to gauge the models' ability to discriminate between anomalies and routine data.

Apart from evaluating the efficacy of existing anomaly detection methods, the evaluation of this paper aims to providethe groundwork for future advancements in the domain. The investigation attempts to throw light on the minute detailsand flaws of the methods now in use to guide the development of more dependable and adaptable anomaly detection systems. This research seeks to contribute to the ongoing development of anomaly detection techniques by utilizing a strict evaluation process to offer solutions that deal with the complex and constantly evolving nature of modern data streams.

1.1 Motivation

The motivation behind this research project is highlighted by the pressing need to improve security, discourage criminal activity, and allocate resources inside surveillance systemsas efficiently as possible. Our research aims to improve the accuracy of identifying suspicious activity in surveillance video by utilizing cutting-edge anomaly detection techniquesdriven by deep learning and computer vision.

1.2 Contribution

The "AnomalyDetectNet" model's main contribution to this study is that it serves as an effective and flexible anomaly detection method. Because of its flexible engineering, it can betailored to a variety of surveillance circumstances, making it an effective tool for identifying regular and irregular patterns in data streams.

1.3 Organization

The following sections of the article are as follows: Section II provides an overview of similar work; Section III provides a background study; Section IV provides a thorough explanationof our methodology, implementation, and results; and Section V concludes with a summary of our findings.

II. Related Work

In [1] The proposed method offers a lightweight approach tovideo anomaly detection, departing from traditional multiple instance learning (MIL) and emphasizing temporal relationships among video segments. It employs a self-attention mechanism to automatically extract features, significantly reducing parameter complexity (only 1.3% compared to existing methods) while maintaining or surpassing frame-level detectionaccuracy on benchmark datasets (UCF-Crime, ShanghaiTech, XD-Violence). Future enhancements could include scalability to larger datasets, improved interpretability, bias mitigation, and real-world deployment considerations. However, potentialdisadvantages may involve limitations in handling specificanomaly types and video conditions, necessitating further evaluation in diverse scenarios.

In [2] The proposed framework introduces a lightweight CNN-based model with a residual attention-based LSTM for video anomaly recognition in smart cities. It effectively reduces time complexity while achieving state-of-the-art accuracy in complex surveillance environments. By leveraging deep CNN spatial features and LSTM sequence learning, it adapts well to smart surveillance systems. Experimentalresults show significant accuracy improvements, with 1.77%, 0.76%, and 8.62% increases on UCF-Crime, UMN, and Avenue datasets, respectively. Future enhancements may explorealternative deep learning models and generative techniquesfor recognizing additional anomaly classes, but challenges could include computational resource requirements for real-time applications and adaptability to dynamic surveillance scenarios.

In [3] The proposed method introduces an unsupervisedapproach for video anomaly detection, using deep 3D convolutional networks (C3D) to extract spatiotemporal features, surpassing hand-crafted features. It optimizes sparse coding and unsupervised feature learning jointly, enhancing feature representations. Experimental results demonstrate its effectiveness in outperforming existing methods on video surveillance datasets. However, it currently lags behind supervised deep neural networks in performance. Future enhancements may explore self-supervised signals and extensions to other video analysis tasks with expensive labeling requirements, but potential drawbacks include computational complexity for real-timeapplications.

In [4] proposed Visual Cloze Completion (VCC) method addresses key challenges in video abnormal event detection (VAD) by achieving precise event localization using appearance and motion cues. It leverages cloze tests to create Visual Cloze Tests (VCTs), emphasizing semantic understanding through DNN-based patch completion. VCC effectively exploits temporal context and offers enhanced versions for performance gains. It demonstrates state-of-the-art VAD results. Future enhancements could explore non-fixed camera scenarios and weakly supervised VAD applications. However, potential disadvantages include increased computational complexity due to ensembling strategies and data requirements forDNN training.

In [5] This paper introduces a novel autoencoder architecture for video anomaly detection, separating spatiotemporal information into two sub-modules to capture spatial and temporal regularities while generating concise representations of normal events. The spatial autoencoder reconstructs individualframes to learn spatial regularities, while the temporal autoencoder efficiently models temporal regularities using RGB differences between consecutive frames. To enhance the detectionof fast-moving anomalies, a variance-based attention moduleis incorporated. Additionally, a deep K-means cluster strategy improves feature learning and data representation. Combining spatial and motion autoencoders and cluster-based evaluation achieves state-of-the-art performance on benchmark datasets. Future enhancements could explore more intricate temporal modeling, scalability, and real-time implementation. Potential drawbacks may include increased computational demands anddata requirements for training.

In [12] This paper introduces a novel abnormality detection method for crowded scenes using Generative Adversarial Nets (GANs). Trained exclusively on normal data, these GANscannot generate abnormal events. During testing, the real data are compared to GAN-reconstructed appearance and motion representations to detect abnormalities via local differences. Experimental results demonstrate the method's superiority in frame-level and pixel-level abnormality detection tasks. Futurework may explore alternative motion representation techniqueslike Dynamic Images. Potential drawbacks include the needfor substantial normal data for GAN training and limitationsin detecting complex abnormalities.

In [13] The paper presents Deep-cascade, a method forefficient anomaly detection and localization in crowded video scenes. It utilizes a cascade of two deep networks to quickly identify normal patches and detect anomalies, reducing computation time while maintaining accuracy. The approach combines advanced feature learning with a cascade structure, yielding results comparable to existing methods on standard benchmarks. Future improvements may involve fine-tuninghyperparameters and addressing data requirements for deep network training.

In [14] This paper presents an innovative framework for one-class classification and novelty detection in images and videos using an adversarial approach. The architecture involves two modules, Reconstructor and Discriminator, collaborating to learn the target class concept. Reconstructor effectively reconstructs target class samples and distorts non-conforming samples, enhancing the Discriminator's ability to distinguish testing samples. The method proves versatile across various anomaly and outlier detection tasks,

demonstratingsuperior performance on MNIST, Caltech-256 image datasets,and UCSD Ped2 video dataset compared to baseline and state-of-the-art methods. Future enhancements might explore more complex data representations and extension to other domains, while potential drawbacks could include sensitivity to hyperparameters and computational demands for adversarial training.

In [15] This paper presents a novel approach to abnormal event detection in videos, reframing it as a one-versus-rest binary classification problem rather than traditional outlier detection. It introduces an unsupervised feature learning framework using object-centric convolutional auto-encoders and a supervised classification method based on clustering training samples into normality clusters. The approach outperformsstate-of-the-art methods across four benchmark datasets, showing significant gains in frame-level AUC. Future enhancementsmay explore object segmentation and tracking. However, potential drawbacks could include sensitivity to clustering parameters and increased computational complexity due tomulti-class classification.

In [16] introduces UBnormal, a groundbreaking benchmark for video anomaly detection that addresses open-set and closed-set scenarios. Unlike existing datasets, UBnormal provides pixel-level annotations for abnormal events during training, enabling the use of supervised learning methods.It allows a fair comparison between open-set and closed-set models and showcases improved performance on Avenue and ShanghaiTech datasets. However, UBnormal relies on virtual characters and simulated actions, limiting its real-world applicability. Future work may explore augmenting other anomaly detection datasets, although the impact on closed-set scenarioslike UCF-Crime may be less significant.

In [17] This paper introduces the Multiple Instance Self-Training (MIST) framework for weakly supervised video anomaly detection. MIST efficiently refines task-specific discriminative representations with video-level annotations. It employs a multiple instance pseudo label generator with a sparse continuous sampling strategy to produce reliable clip-level pseudo labels and a self-guided attention-boosted featureencoder that automatically focuses on anomalous regions during feature extraction. Through a two-stage self-training process, MIST achieves significant improvements on public datasets, performing comparably to or even outperforming existing supervised and weakly supervised methods, with a notable frame-level AUC of 94.83% on the ShanghaiTechdataset. Future work could explore advanced pseudo-labeling strategies, while potential limitations may include sensitivityto hyperparameters and computational requirements.

In [18] this paper presents the Memory-Augmented Autoencoder (MemAE) as an improvement for autoencoder-based unsupervised anomaly detection. MemAE incorporates amemory module that stores prototypical normal data patterns, enhancing the reconstruction process. By using this memory, MemAE can effectively reconstruct normal samples while magnifying the reconstruction error for anomalies, making ita robust anomaly detection criterion. The method is adaptable to various data types and applications, demonstratingits generality and effectiveness through experiments on diverse datasets. Future enhancements may involve exploring addressing weights for anomaly detection and integrating the memory module into more complex models for challenging applications. However, MemAE may come with increasedcomputational demands and require careful hyperparameter tuning.

In [19] This paper introduces a robust Gaussian Processes (GP) mixture framework for addressing challenges in multipleinstance learning (MIL) involving noise and multimodality. Bysimultaneously considering multiple instances using a latent mixture model, the framework enhances the model's resilienceto outliers and diverse scenarios. It incorporates a Distributionally Robust Optimization (DRO) constraint, enabling adaptiveinstance selection without requiring a fixed parameter value. To handle high-dimensional data common in MIL, the GP kernel is enriched with adaptive basis functions learned by a deep neural network. Experiments on video anomaly detectiontasks validate the model's effectiveness, with potential future enhancements focusing on handling complex multimodal scenarios and scalability for large datasets.

In [20] This paper introduces an innovative approach to anomaly detection in surveillance videos, focusing on the correspondence between object appearances and their associated motions. The model combines a reconstruction network and an image translation model, sharing the same encoder,to establish this correspondence. Training is achieved using normal event videos, enabling the model to estimate frame-level scores for unknown inputs. Experimental results on six benchmark datasets showcase the competitive performanceof this approach compared to state-of-the-art methods. The method's advantages include its ability to capture complex event patterns and motions, while potential future enhancements might involve refining anomaly scoring strategies and extending its applicability to more diverse surveillance scenarios. One limitation could be the need for a substantial amount of normal data for training, which might not always be readilyavailable in real-world applications.

In [21] The research presents a unique weakly supervised anomaly detection strategy for noisy video-level annotationsin real-world video data. This method significantly reduces label noise because it incorporates binary clustering, enabling the core network and clustering to work together and perform better. On the difficult UCF-crime and ShanghaiTech datasets,the findings are outstanding, displaying significant frame-level AUC values of 78.27% and 84.16%, respectively. This study offers a viable method for improving anomaly identificationin video data under suboptimal labelling circumstances.

III. BACKGROUND STUDY

The paper [6] proposes a novel weakly-supervised video anomaly detection method based on robust temporal feature magnitude learning. The method is designed to address the problem of false positive alarms caused by dominant negative instances in weakly supervised video anomaly detection. The main approach is that a feature extractor is used to extract features from each video snippet Then, a temporal featuremagnitude learning function is trained to learn the temporal dynamics of the feature magnitudes and a multiple instance learning (MIL) framework is used to identify abnormal video snippets. The feature extractor used in the method is a 3D convolutional neural network and the temporal feature magnitude learning function is implemented as a convolutional neural network with dilated convolutions and self-attention.

The MIL framework works by first clustering the feature magnitudes from all video snippets into two clusters: normal and abnormal. Then, each video is assigned a label of normal or abnormal based on the majority label of its constituentsnippets. The temporal feature magnitude learning function is designed to be robust to the dominant negative instances. Thisis done using a dilated convolution operation to capture long-range temporal dependencies and a self-attention mechanism to learn the relative importance of different features.

The proposed model called Robust Temporal Feature Magnitude (RTFM), addresses the challenge of differentiating between abnormal and normal video snippets with weaklylabeled training data. It operates on a dataset of weakly labeled training videos where each video snippet is representedas a feature vector with corresponding binary video-level labels. The RTFM model consists of two main components: a temporal feature extractor and a snippet classifier.

Theoretically, RTFM is motivated by the notion of classifying snippets as normal or abnormal by employing feature magnitude (L2 norm), which makes a softer assumption than conventional MIL methods. The top k temporal feature snip- pets are the main focus of the model, where k is the numberof top instances to take into account. Effective training of the snippet classifier depends on the top-k features.

A Multi-Scale Temporal Network (MTN) in the model records the local and global temporal dependencies between video clips at multiple resolutions. MTN effectively discovers these relationships by means of dilated convolutions over the temporal dimension. The modeling of the global temporalenvironment and the capture of long-range dependencies are done using a self-attention module.

Finally, feature magnitude learning is achieved through a loss function that maximizes the separability between normal and abnormal videos based on the top-k largest snippet feature magnitudes.

IV. IMPLEMENTATION

Fig. 1. provides a visual representation of the anomaly detection system workflow, including important phases includingfeature extraction, data preprocessing, anomaly classification,and performance assessment. The system is based on AnomalyDetectNet, which is a modified neural network architecture that is based on the FNN (Feedforward Neural Network) and isintended to distinguish between normal and abnormal data by use of a series of integrated layers. The system begins with pre-processing the data, which involves cleaning and preparing theinformation before extracting relevant characteristics that are necessary for identifying anomalies. The AnomalyDetectNet anomaly classification algorithm uses these criteria to distinguish between normal and abnormal occurrences. The model'sfundamental layers are feature extraction and classification. The fully connected layer (fc1), which is the first step in the feature extraction process, turns input data of dimension input dim into a new representation with hidden dim neurons, the eq.(1) results in this transformation. The output of fc1 is then applied element-wise by a Rectified Linear Unit (ReLU) activation function to add non-linearity eq.(2), where eq.(3)is the outcome of applying a dropout layer with a dropout probability of drop p to avoid overfitting.

$$fc1\_output = fc1(input\_data) \tag{1}$$

$$relu\_output = relu(fc1\_output) \tag{2}$$

$$Dropout\_output = dropout(relu\_output) \tag{3}$$

The data is processed further before being transferred tothe classification layer. A second fully connected layer (fc2) isapplied to the feature-extracted representation, further bringingdown the dimensionality to 1 by eq.(4). The extracted featuresare mapped by this layer to a single output value. The final output is fed via the hyperbolic tangent (tanh) activation function for binary classification in the context of anomaly detection eq.(5) The output is condensed by the tanh activationfunction into the range [-1, 1]. For the purpose of detecting anomalies, numbers closer to 1 denote anomalies, whereas values closer to -1 denote regular data. With the use of hyperparameters like input dim, hidden dim, and drop p, the FNN uses these processes to categorize input data as either normal or anomalous.

$$fc2\_output = fc2(dropout\_output) \tag{4}$$

$$anomaly\_score = torch.tanh(fc2\_output). \tag{5}$$

This anomaly detection model uses a Multiple Instance Learning (MIL) loss function, which is an essential part of training the network. MIL is designed for situations where bags of instances, rather than single instances, are given labels.The goal is to let the model recognize particular patterns that point to anomalies inside a bag, which is essential for video-based anomaly detection.

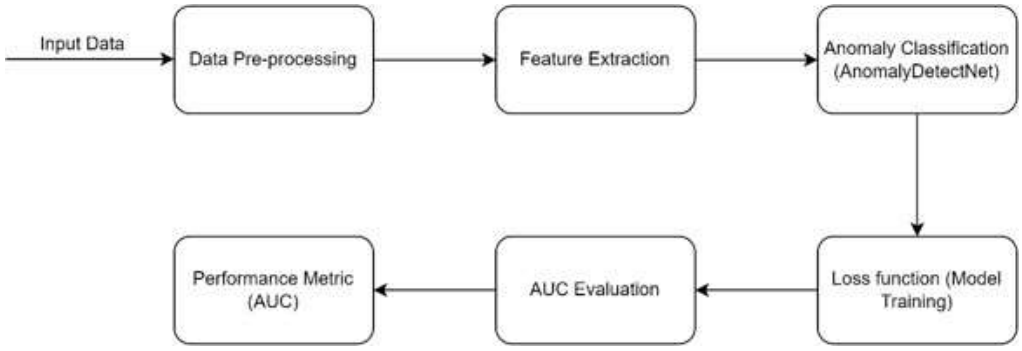

Fig. 1. Block Diagram of the Anomaly Detection System

Based on the projected scores and labels, the MIL lossfunction is calculated. It uses a variety of phrases to impose certain characteristics in the anticipated ratings. The main lossterm seeks to increase the difference between the maximum predicted score for anomalies and the maximum predictedscore for normal instances inside a bag. The hinge loss function,

$$Loss+ = F.relu(1 - (Y_{am} - Y_{nm}) \tag{6}$$

eq.(6) is used to accomplish this. Yam (y_ anomaly_ max) and Ynm (y_normal_max) stand for the highest predicted scores for anomalies and normal occurrences, respectively,within a bag.

The regularisation of smoothness and sparsity are also termsin the loss function. By penalizing large amounts of anomaly scores, the sparsity term pushes the model to deliver sparse predictions. Calculated by eq(7),

$$sparsity+ = torch.sumY_a * 0.00008 \quad (7)$$

Where Ya (y_anomaly) stands for the expected anomaly scores, by penalizing the squared disparities between nearby scores, calculated as torch, the smoothness component encourages smooth transitions between consecutive anticipated scoresand calculated by eq.(8),

$$\text{smooth+ = torch.sum}((y\_pred[i,:31]\text{-}y\ \ pred[i,1:32]) * *2) * 0.00008 \quad (8)$$

The model optimizes these goals across numerous instances byaggregating these terms and normalizing them by batch sizeto produce the final loss. Robust anomaly identification duringtraining is made possible by the MIL loss function, which successfully directs the network to discriminate anomalies from typical examples within bags of data.

The MIL loss function is used by the Adagrad (Adaptive Gradient Algorithm), which also acts as the model's (AnomalyDetectNe) optimizer and learning rate scheduler during training. The code tests the model's capacity to identify abnormalities using Area Under the Curve (AUC) metrics; if the model achieves a higher AUC score, it is saved.

Two unique classes are employed to normalise and tensorize images, which are essential for preparing frames before they are sent to the models. Pretrained models for feature extraction(resnext model) and anomaly classification (FNN) are loaded in the following steps. These models have the ability to extractsignificant characteristics from video data and identify anomalies since they have been pre-trained on relevant datasets. in order to format the videos.

The image is loaded and pre-processed for each frame,then it is added to a buffer of 16 frames and used to compute feature vectors using the feature extraction model. The anomaly prediction obtained from these characteristicsis then added to the y_ pred list after being passed throughthe classifier. The frame is designated as a "danger zone" if the prediction exceeds a predetermined threshold (0.02). The processed frames are saved in an output directory, and an MP4video is created from these frames utilizing FFmpeg. This robust methodology demonstrates the potential for real-world applications in surveillance, security, and anomaly detection scenarios.

4.1 Problem Statements

Developing an effective anomaly detection model for surveillance videos that harnesses deep learning techniques, particularly the "FNN" model, to accurately identify and categorize anomalous events within these videos. This challenge arises due to the dynamic nature and complexity of videodata, emphasizing the need for robust and efficient modelsthat enhance public safety and security while maintaining adaptability to diverse surveillance scenarios.

4.2 Dataset

The dataset used is UCF-Crime dataset, which is a comprehensive video dataset containing a diverse range of anomalousevents. It consists of 13 distinct categories of anomalies, including abuse, arrest, arson, assault, burglary, explosion, fighting, road accidents, robbery, shooting, stealing, shoplifting, and vandalism. These anomalies are captured in the form of video clips, making the dataset an invaluable resource for training and evaluating anomaly detection and recognition models. Each anomaly category is represented by multiple video clips, providing a substantial collection of real-world scenarios that can be used for various computer vision and surveillance applications. dataset has been formatted/pre-processed to be compatible with the I3D model, which includes resizing videos, extracting frames, and other modifications to fit the model's input requirements.

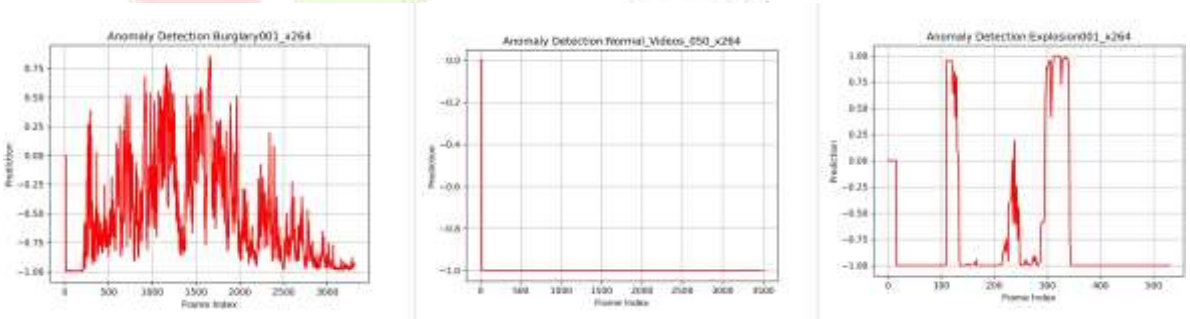

Fig. 2. Graph of the different anomaly detection i.e Burglary, normal video, and Explosion

V. RESULTS

The performance evaluation of our proposed anomaly detection model, the AnomalyDetectNet, was conducted rigorously against a baseline method, denoted as "RTFM" on the UCF- Crime dataset a widely recognized benchmark for surveillance video anomaly detection. Our "AnomalyDetectNet" model demonstrated a notable advancement in accuracy when compared to the "RTFM" baseline. Specifically, our method achieved an accuracy of 84.84%, while the "RTFM" method achieved 84.30% as shown in TABLE I and Fig. 2. which shows the anomaly detection with prediction and frames of the burglary, normal video, and explosion. This marginal yet consistentimprovement in accuracy emphasizes the effectiveness of the "AnomalyDetectNet" model in accurately identifying and categorizing anomalous events within surveillance videos.

VI. CONCLUSION

This paper presents an anomaly detection model, the "AnomalyDetectNet", which harnesses deep learning techniques to accurately identify and categorize anomalous events in surveillance videos. Anomaly detection is a critical task with applications in diverse domains, including security and surveillance. The "AnomalyDetectNet" model, with its featureextraction and classification layers, effectively distinguishes between regular and anomalous data, offering a precise mechanism for enhancing public safety. By employing Multiple Instance Learning (MIL) loss function and hyperbolic tangent (tanh) activation functions, it delivers an impressive accuracy of 84.84% on the UCF-Crime dataset, outperformingthe baseline "RTFM" method, which achieved 84.30%. This work contributes to the field of anomaly detection, offeringa robust and adaptable solution that holds promise for real-world applications, bolstering security and safety across various surveillance scenarios. Adapting the model for real-time processing of video feeds is one way to make enhancements that could increase the efficacy and timeliness of securityand surveillance systems by enabling prompt responses toabnormalities as they arise.

TABLE I
Auc Performance On Ucf-Crime

Supervision	Method	Feature	AUC (%)
Weekly Supervised	Sultani et al. [7]	C3D RGB	75.41
		13D RGB	77.92
	Zhang et al. [8]	C3D RGB	78.66
	Motion-Aware [9]	PWC Flow	79.00
		C3D RGB	81.08
	GCN-Anomaly [10]	TSN Flow	78.08
		TSN RGB	82.12
	Wu et al. [11]	13D RGB	82.44
	RTFM [6]	C3D RGB	83.28
		13D RGB	84.30
	Proposed	13D RGB	84.84

REFERENCES

[1] Yudai Watanabe, Makoto Okabe, Yasunori Harada, and Naoji Kashima "Real-World Video Anomaly Detection by Extracting Salient Featuresin Videos", IEEE, vol. 10, pp. 125052 – 125060, Nov 2022.

[2] Waseem Ullah, Amin Ullah, Tanveer Hussain, Zulfiqar Ahmad Khan and Sung Wook Baik "An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos", Journals Sensors, vol. 22, no.8, pp. 2811, Feb 2021.

[3] Wenqing Chu, Hongyang Xue, Chengwei Yao and Deng Cai, Member, IEEE "Sparse Coding Guided Spatiotemporal Feature Learning for Abnormal Event Detection in Large Videos" vol. 21, no.1, pp. 246 – 255, 2018.

[4] Siqi Wang, Guang Yu, Zhiping Cai1, Xinwang Liu, En Zhu, and Jianping Yin "Video Abnormal Event Detection by Learning to Complete Visual Cloze Tests" ACM Multimedia Conference, vol. 33, pp. 2188-2204, 2020.

[5] Yunpeng Changa, Zhigang Tua, , Wei Xie b, Bin Luoa, Shifu Zhangc, Haigang Sui a, Junsong Yuan "Video anomaly detection with spatiotem- poral dissociation", vol. 122, no. 4, Aug 2021.

[6] Yu Tian, Guansong Pang, Yuanhong Chen, Rajvinder Singh, Johan W. Verjans, Gustavo Carneiro "Weakly-supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning" IEEE International Conference on Computer Vision (ICCV), vol. 117, pp. 379- 392, Aug 2021.

[7] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 6479–6488, 2018.

[8] J. Zhang, L. Qing, and J. Miao. "Temporal convolutional network with complementary inner bag loss for weakly supervised anomaly detection." In 2019 IEEE International Conference on Image Processing (ICIP), pp 4030–4034, 2019.

[9] Yi Zhu and Shawn Newsam. "Motion-aware feature for improved video anomaly detection." arXiv preprint arXiv:1907.10211, 2019.

[10] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. Graph convolutional label noise cleaner: Train a plug- and-play action classifier for anomaly detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1237–1246, 2019.

[11] Peng Wu, jing Liu, Yujia Shi, Yujia Sun, Fangtao Shao, Zhaoyang Wu, and Zhiwei Yang. "Not only look, but also listen: Learning multimodal violence detection under weak supervision." In European Conference on Computer Vision (ECCV), 2020.

[12] Mahdyar Ravanbakhsh , Moin Nabi , Enver Sangineto , Lucio Marcenaro, Carlo Regazzoni , Nicu Sebe "abnormal event detection in videos using generative adversarial nets" Computer Vision and Pattern Recognition (cs.CV); Multimedia (cs.MM), Aug 2017.

[13] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette. Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. IEEE Transactions on Image Processing, vol. 26 no. 4 pp. 1992–2004, 2017.

[14] Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. Adversarially learned one-class classifier for novelty de- tection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[15] Radu Tudor Ionescu, Fahad Shahbaz Khan, Mariana-Iuliana Georgescu, and Ling Shao. "Object-centric auto-encoders and dummy anomalies for abnormal event detection in video." In Proceedings of the IEEE Con- ference on Computer Vision and Pattern Recognition, pp. 7842–7851, 2019.

[16] A. Acsintoae, A. Florescu, M.-I. Georgescu, T. Mare, P. Sumedrea, R. T. Ionescu, F. S. Khan, and M. Shah, "UBnormal: New benchmark for supervised open-set video anomaly detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 20143–20153, Jun. 2022.

[17] J.-C. Feng, F.-T. Hong, and W.-S. Zheng, "MIST: Multiple instance self- training framework for video anomaly detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 14004–14013, Jun. 2021.

[18] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. Van Den Hengel, "Memorizing normality to detect anomaly: Memory- augmented deep autoencoder for unsupervised anomaly detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1705–1714.

[19] H. Sapkota, Y. Ying, F. Chen, and Q. Yu, "Distributionally robust optimization for deep kernel multiple instance learning," in Proc. 24th Int. Conf. Artif. Intell. Statist., vol. 130, A. Banerjee and K. Fukumizu, Eds., Apr. 2021, pp. 2188–2196.

[20] T. N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1273–1283.

[21] Muhammad Zaigham Zaheer, Jin-ha Lee, Marcella Astrid, Arif Mah- mood, and Seung-Ik Lee. "Cleaning label noise with clusters for minimally supervised anomaly detection." arXiv preprint arXiv:2104.14770, 2021.