Surveillance Video Anomaly Detection With Multi-Branch Gan

Raj Bharath R[1], Padmaja P[2], Maha Nakshathra P[2], Prithiksha Devi G[2], Umamageswari M[2]

[1]Head of Department, Department of AI & ML, Manakula Vinayagar Institute of Technology, Puducherry,
India - 605107, rajbharathcse@mvit.edu.in
[2]Undergraduate Students, Department of AI & ML, Manakula Vinayagar Institute of Technology, Puducherry, India - 605107
Emails: padmajapselvam@gmail.com, mahanakshathra10@gmail.com, prithikshag16@gmail.com, u mamageswarimohan15@gmail.com

Abstract
Identifying irregularities in surveillance videos is essential for maintaining public safety in various settings. Conventional methods, which often depend on human oversight, can be inefficient and susceptible to errors. Nevertheless, with the advancement of deep learning technologies, the task of detecting anomalies in real-time can now be automated. Generative Adversarial Networks (GANs) have demonstrated their effectiveness in video analysis, enabling the automated identification of anomalies. This project introduces an innovative approach that employs a Multi-Branch GAN (M-GAN) model specifically crafted for the detection of abnormal events in surveillance footage. The M-GAN utilizes a two-phase approach: it initially learns the patterns of typical activities and subsequently identifies any deviations as potential anomalies. A key advantage of this model is its ability to function without requiring labelled anomaly data, which increases its flexibility across different environments and conditions. Experimental results demonstrate that the M-GAN outperforms conventional GAN-based methods, achieving superior precision and recall rates in detecting abnormal occurrences. This outstanding performance positions M-GAN as an excellent solution for instantaneousabnormal event detection in surveillance systems, enhancing safety and security while reducing dependence on manual monitoring.

Keywords: Abnormal event detection, surveillance footage, deep learning, Generative Adversarial Networks (GANs), Multi-Branch GAN (M-GAN), real-time detection, unsupervised learning.

I.INTRODUCTION

Ensuring public safety through effective monitoring has become essential in a society that is becoming more urbanized and connected. The incidents such as theft, vandalism, and gang violence presents serious challenges to the well-being of communities and calls for innovative responses. Human issues including fatigue, oversight, and delayed reactions limit traditional surveillance systems, which usually rely on manual monitoring and passive video recording. These drawbacks emphasize the necessity of integrating intelligent technology with the capacity to act instantly and make judgments on their own. Significant developments in surveillance systems have resulted from recent advancements in deep learning and artificial intelligence (AI). Because they can model complex data distributions and identify irregularities that can indicate abnormal activity, GANs (Generative Adversarial Networks) have become as a useful tool for unsupervised learning and abnormal event detection. In particular, the Multi-Branch GAN (M-GAN) model provides a unique framework aimed at identifying subtle, context-aware anomalies in evolving such environments such as public areas, transportation hubs, and shopping centers. Unlike conventional machine learning models that depend on large, labeled datasets, M-GAN is particularly useful in real-world scenarios when labeled anomaly data is scarce or nonexistent since it uses unsupervised learning to adaptively represent typical activity patterns. Through a two-stage

learning approach, M-GAN differentiates between ordinary behavior and highlights deviations as potential threats, thereby enabling preventive security measures before incidents can escalate.This project not only presents the M-GAN model but also emphasizes the interdisciplinary convergence of AI, video analytics, and public policy. By creating synthetic data that simulates real-world situations, the system enhances its predictive abilities, achieving strong performance across varied environmental conditions such as differing lighting, crowd sizes, and complexities of motion.Moreover, smart surveillance must extend beyond mere technological advancement—it requires ethical use and community involvement. Consequently, this research also stresses the significance of community engagement, transparency, and data privacy in cultivating societal trust. Intelligent systems should be accountable, equitable, and inclusive, particularly as their role expands in influencing public safety outcomes.Building on these foundational principles, the Multi-Branch Generative Adversarial Network (M-GAN) presents an innovative architecture that utilizes multiple generator branches to effectively capture various dimensions of normal activity patterns in video surveillance data. Each branch is tailored to model distinct temporal and spatial characteristics, thereby enhancing the system's ability to comprehend the intricate and often nuanced variations of normal behavior in dynamic settings.

This multi-branch strategy overcomes the shortcomings of single-stream GAN models, which may miss essential contextual signals or struggle to generalize across different scenarios. A notable advancement within the M-GAN framework is the integration of transfer learning, enabling the model to leverage insights gained from one surveillance domain and apply them successfully to another. This capability is particularly advantageous given the limited availability of labeled anomaly data in numerous real-world contexts. By pre-training on extensive, varied datasets and subsequently fine-tuning on specific target environments, M-GAN achieves improved adaptability and resilience, minimizing the necessity for extensive manual labeling while preserving high detection accuracy. Furthermore, the system adopts a two-stage learning methodology. In the first stage, it learns to reconstruct or predict normal frames, thereby establishing a baseline representation of typical scene dynamics. In the second stage, deviations from this established baseline are identified as potential anomalies. This unsupervised learning approach is particularly well-suited for surveillance applications, where abnormal occurrences are infrequent and unpredictable, rendering supervised training impractical.To enhance detection accuracy further, the model incorporates pseudo-anomaly generation techniques that mimic abnormal patterns, thus improving the discriminator’s sensitivity to subtle irregularities. In addition to technical efficacy, the implementation of intelligent surveillance systems necessitates ethical considerations. In summary, the suggested M-GAN-based abnormal event detection system signifies a meaningful advancement towards real-time, scalable, and ethically sound surveillance. By integrating leading machine learning methods with considerate design principles, it aims to improve security frameworks while honoring the societal values they seek to uphold.

II.RELATED WORK

Deep learning and generative models have significantly improved anomaly identification in time-series and video surveillance. A new method for modelling normal behaviour without labels was provided by TAnoGAN, which offered a GAN-based technique for identifying anomalies in time series [1]. By combining several data sources, a multi-view causal graph fusion technique enhanced anomaly detection performance in the setting of cyber-physical infrastructures [2]. A self-trained spatial graph convolutional network that learned from structural relationships improved unsupervised identification in complex situations [3], and a weakly-supervised approach that combined graph convolution and label noise cleaning had encouraging outcomes in video surveillance [4]. It's still difficult to distinguish objects in low light.

YOLO-based methods have demonstrated efficacy in identifying and detecting objects in low-light conditions, supporting subsequent tasks including anomaly detection [5]. For low-data circumstances, few-shot learning has also been investigated for scene-adaptive anomaly detection [6]. Graph convolutional neural networks for skeleton-based behaviour recognition have helped detect irregular activity in videos with high reliability [7]. A discriminative feature extractor for identifying odd human poses in security footage was suggested: graph-based pose clustering [9]. Models such as RTMDet provide insights into the design of quick and accurate detectors for real-time object detection [10].

Similarly, to improve contextual awareness, spatiotemporal relationships between objects were used to detect anomalies in videos [12]. The focus on aberrant areas in video frames has been further enhanced by cluster attention contrast techniques [13]. High detection accuracy and low latency are crucial for surveillance systems, and this is the goal of real-time object detection frameworks like Damo-YOLO [14]. Frame-area effectiveness has been investigated in surveillance footage to improve the accuracy of anomaly identification by assisting in focusing on pertinent visual regions [15]. It has been demonstrated that a two-stream fusion method that combines temporal and spatial data greatly enhances the effectiveness of abnormal event detection [16]. Furthermore, by modeling normal patterns over time, memory-guided normality learning offered a potent unsupervised technique for identifying anomalies [17].

A. Technique used

A crucial tactic in video surveillance-based abnormal event detection is advanced modeling of spatial and temporal dynamics, which enables systems to record intricate motion patterns across successive frames. To assess temporal sequences and comprehend motion across time, methods like 3D-CNNs (3D Convolutional Neural Networks) and LSTM (Long Short-Term Memory) networks are frequently used. By identifying atypical patterns within individual frames, CNNs (Convolutional Neural Networks) are essential for extracting spatial data. Generative models efficiently create a baseline of typical activity, enabling anomalies to be identified as departures from predicted patterns, despite the fact that they can be computationally demanding and vulnerable to stability problems. To lessen reliance on extensive labeled datasets, few-shot learning and transfer learning techniques are also being used more and more. These techniques improve abnormal event detection algorithms' ability to adapt to actual surveillance situations, when there are frequently few labeled aberrant events.

B. Model Evaluation

Important detection metrics like TPR (True Positive Rate) and FPR (False Positive Rate) are calculated in order to thoroughly assess the effectiveness of the suggested model. Precision and recall are successfully integrated by using the F1-score as a balanced statistic to adjust for data imbalance. Recall counts the percentage of real anomalies that the model successfully detects, whereas precision quantifies the percentage of correctly diagnosed anomalies among all anticipated positives.Robustness is evaluated by deploying the model on datasets characterized by diverse environmental conditions, including variations in illumination and crowd density. To ensure reliable deployment in real-world scenarios, efforts are directed toward minimizing both false positives and false negatives. Furthermore, real-time inference performance is analyzed to ensure that the model adheres to the latency and throughput requirements of practical surveillance system.

C. Performance Validation

The performance of the proposed model is evaluated using a set of standard metrics and validation procedures, focusing on its ability to accurately detect and classify anomalous events. TPR (True Positive Rate) and FPR (False Positive Rate) are calculated to measure detection accuracy, indicating the model's effectiveness in identifying anomalies while correctly classifying normal instances. Precision and recall further quantify the model's reliability and sensitivity, respectively, while the F1-score offers a balanced evaluation of these metrics, particularly in the presence of class imbalance. Additionally, the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is employed to assess the model's threshold-independent discriminative capability between normal and abnormal instances.To ensure robustness and generalizability, the model is evaluated on datasets exhibiting diverse conditions such as varying illumination, crowd densities, and environmental settings. Real-time performance is also validated to confirm the model's suitability for continuous surveillance applications, where low-latency inference is important.Lastly, efforts are made to minimize both false positives and false negatives, thereby improving the reliability of model and practical effectiveness in real-world abnormal event detection scenarios.

III. LITERATURE SURVEY

The growing need for precise and effective real-time monitoring solutions has led to increase in interest in abnormal event detection in surveillance systems in recent years. Numerous approaches

have been put out to address the inherent difficulties in identifying unusual activity, particularly in complex and dynamic environments such as those captured by security cameras.This literature survey presents a comprehensive review of key contributions and limitations across prominent approaches in the field, including spatiotemporal modeling techniques,CNNs (Convolutional Neural Networks),weakly supervised learning and GANs (Generative Adversarial Networks).

Spatiotemporal Technique for Surveillance Anomaly Detection

By accurately simulating the spatial characteristics and temporal dynamics included in video data, spatiotemporal approaches are essential for identifying unusual occurrences in surveillance film. Methods like 3D-CNNs (3D Convolutional Neural Networks) and LSTM (Long Short-Term Memory) networks are frequently used to record an object's look and movements over time. Through the analysis of frame-wise motion and inter-frame interdependence, these models show remarkable skills in detecting irregular patterns inside complicated contexts. However, deploying them in real-time monitoring systems is difficult due to their high processing complexity.

Detecting Events using CNNs

The capacity of CNNs (Convolutional Neural Networks) to extract spatial characteristics from pictures and video frames makes them essential for abnormal event identification. In surveillance systems, CNNs are utilized to identify patterns indicative of anomalous activities, such as sudden movements or unusual behaviors. While CNNs excel at detecting spatial anomalies, they face limitations in capturing temporal dependencies within video data. As a result, CNNs alone may not be sufficient for detecting anomalies that involve complex, continuous processes. To address this limitation, recent advancements have integrated CNNs with other models, such as spatio-temporal networks or Multi-Generational Adversarial Networks (M-GANs), to increase the accuracy and resilience of abnormal event detection by combining temporal and spatial data.

GANs for Anomaly Detection

GANs (Generative Adversarial Networks) showed significant potential in the field of abnormal event detection, especially in unsupervised and weakly supervised contexts. In order for GANs to detect anomalies, they must first learn to represent the distribution of typical occurrences. In situations like surveillance systems, when labeled data is either few or nonexistent, this technology has shown promise. The AnoGAN model, for example, has successfully detected anomalies by generating representations of "normal" events and flagging deviations from these patterns. However, despite their advantages, GANs face significant challenges, including training instability and high computational requirements, which can limit their applicability in real-time surveillance systems.

Object Detection in Low-Light Surveillance Using YOLO

Due to noise and inadequate lighting, surveillance systems have a difficult time detecting objects in low-light conditions. Real-time applications can benefit from the quick and efficient object identification technique known as YOLO (You Only Look Once), which analyzes photos in a single pass. Recent modifications that include preprocessing methods, sophisticated feature extraction, and attention processes have improved YOLO's performance in low-light conditions. These advancements make it possible to recognize objects more accurately in low light levels, where conventional techniques frequently fall short. Improved detection in poor visibility serves vital functions like anomaly detection and event monitoring and increases the dependability of surveillance systems.

Learning in Anomaly Detection with Weak Supervision

To address the challenge of handling large, sparsely labeled datasetscommon in surveillance applicationsthere is a growing reliance on weakly supervised learning algorithms for anomaly detection.Models can be trained efficiently with few annotations thanks to strategies like few-shot learning and transfer learning, which lessens the need for huge amounts of labeled data. The above methods are particularly beneficial in real-world surveillance systems, where it is often impractical to label every instance. Recent studies suggest that combining deep learning modelswith weakly supervised learning approaches can enhance detection accuracy while minimizing the need for large labeled datasets. This trend, which emphasizes the use of readily available data with minimal labeling

efforts, aligns with the broader movement in machine learning towards more efficient learning paradigms.

IV. PROPOSED SOLUTION

The proposed method incorporates a Multi-Branch Generative Adversarial Network (M-GAN) supplemented with transfer learning to allow for real-time abnormal event identification in surveillance contexts. Compared to conventional single-stream models, this architecture offers a more reliable framework for detecting anomalous events by capturing the spatial and temporal patterns in surveillance video data. M-GAN's tri-branch architecture focuses on many feature dimensions, guaranteeing thorough comprehension and precise abnormal event detection.The first branch focuses on spatial feature extraction from RGB video frames. This branch is responsible for detecting structural anomalies through the analysis of still-frame visual characteristics.The second branch captures temporal features by processing stacked optical flow frames with a 3D Convolutional Neural Network (3D-CNN). This branch tracks motion dynamics and activity flow across frames, detecting abnormal behaviors such as abrupt accelerations, unexpected direction changes, and erratic group movements, which are indicative of anomalies.The third branch functions as an appearance-motion fusion module by combining the outputs from the temporal and spatial branches through a fusion layer.

Fig. 1. Detected abnormality

Because of this, the model may better detect tiny anomalies that might go undetected if spatial or temporal information were taken into account separately. Each branch contains its own generator (G_1, G_2, and G_3), trained to model normal patterns. These generators work in conjunction with a shared discriminator (D), which evaluates whether the frames are real or generated, identifying discrepancies that signal potential anomalies. This adversarial framework enhances the model's sensitivity to abnormal deviations while refining its understanding of typical behavior. Transfer learning optimizes the system by reducing reliance on large labeled datasets. The system is modular and scalable, allowing for the integration of additional behavior streams or future surveillance modules. Optimized for real-time operation, it ensures rapid detection of anomalies with minimal delay, enabling timely alerts to security personnel and mitigating the risk of undetected threats. By combining M-GAN's precision with the adaptability of transfer learning, this architecture offers a powerful, efficient, and scalable solution for modern surveillance systems.

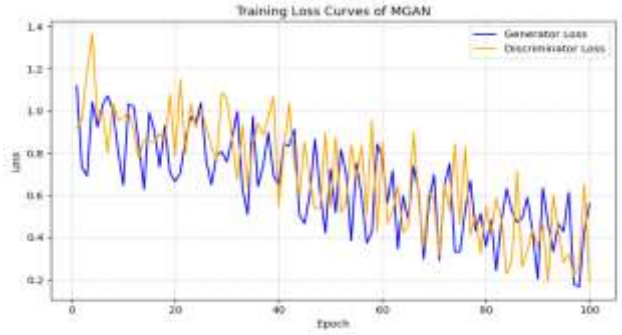

Fig. 2. Training Loss Curves graph

Figure 2 shows how the discriminator and generator losses changed as the MGAN model was being trained. The losses for both components decrease and level off over the epochs, suggesting that the adversarial training was successful and that convergence was achievedThe gradual reduction and ultimate stabilization of the losses highlight the efficacy of the multi-branch design and training approach.

A. METHODOLOGY

Cross-Domain Frame Prediction

MGAN incorporates two parallel generators: the source generator, pre-trained on labeled datasets (e.g., UCSD Ped2, ShanghaiTech), predicts future frames using temporal loss and L1 reconstruction to preserve structural integrity. The target generator is fine-tuned on unlabeled data via AdaIN, aligning cross-domain feature styles through learned affine transformations.

The generator comprises two specialized branches:

1. **Source-domain generator (G_S):** Pre-trained on labeled datasets (UCSD Pedestrian, ShanghaiTech) using temporal prediction loss:

$$\mathcal{L}_{\text{pred}} = \frac{1}{N}\sum_{i=1}^{N} \|G_S(\mathbf{X}_{t-k:t}^{(i)}) - \mathbf{X}_{t+1}^{(i)}\|_1$$

Where:

$\mathbf{X}_{t-k:t}$: Input sequence of k consecutive frames

$\mathbf{X}_{t+1}$: Ground truth future frame

N: Batch size

$\|\cdot\|_1$: L1-norm for reconstruction fidelity

2. **Target-domain generator (G_T):** Fine-tuned using adaptive instance normalization (AdaIN):

$$G_T = \text{AdaIN}(G_S(\mathbf{X}), \gamma, \beta)$$

Where:

γ, β: Domain-specific affine parameters learned during adaptation

AdaIN enables style transfer between source/target domains.

Adversarial Domain Alignment

MGAN employs a GRL (Gradient Reversal Layer) coupled to a domain classifier to facilitate domain-invariant representation learning. During the training phase, the GRL flips the classifier's gradients to encourage the shared encoder to locate features that are indistinguishable between the source and target domains.

Domain-invariant feature learning is enforced by a GRL by:

$$\mathcal{L}_{\text{da}} = -\frac{1}{N_s}\sum_{i=1}^{N_s} \log D_{\text{domain}}(G_S(X_s^{(i)})) + \frac{1}{N_t}\sum_{j=1}^{N_t} \log(1 - D_{\text{domain}}(G_T(X_t^{(j)})))$$

Where:

D_{domain}: Domain classifier (0=source, 1=target)

N_s, N_t: Batch sizes for source/target domains

X_s, X_t: Input frames from respective domains

Multi-Task Optimization

The training objective is composed of multiple loss functions, including adversarial loss for generating realistic frames, prediction loss for ensuring temporal coherence, domain alignment loss for consistent features across different domains, and diffusion regularization loss to enhance robustness. These losses are balanced through the application of empirically tuned coefficients.

The complete objective function combines:

$$\mathcal{L}_{\text{total}} = \underbrace{\lambda_1\mathcal{L}_{\text{pred}}}_{\text{Prediction}} + \underbrace{\lambda_2\mathcal{L}_{\text{adv}}}_{\text{GAN Loss}} + \underbrace{\lambda_3\mathcal{L}_{\text{da}}}_{\text{Domain Alignment}} + \underbrace{\lambda_4\mathcal{L}_{\text{diff}}}_{\text{Anomaly Synthesis}}$$

Where:

$\lambda_1 - \lambda_4$: Experimentally tuned weights (default: 1.0, 0.5, 0.3, 0.2)
$\mathcal{L}_{\text{adv}}$: Standard GAN adversarial loss
$\mathcal{L}_{\text{diff}}$: Diffusion-based anomaly regularization (detailed below)

Dynamic Anomaly Synthesis

To elevate the simulation of anomalies, the model opts for a conditional diffusion process instead of static noise injection. This method employs a noise schedule that regulates the corruption rate at each step throughout a fixed number of diffusion iterations. An identity matrix is used to ensure structural consistency during the noise addition process. By incrementally corrupting real frames, this approach produces temporally consistent pseudo-anomalies, enhancing both the realism and effectiveness of anomaly representation.

Replaces static noise injection with conditional diffusion:

$$q(\mathbf{X}_t|\mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \sqrt{1-\beta_t}\mathbf{X}_{t-1}, \beta_t\mathbf{I})$$

Where:

β_t: Noise schedule controlling corruption rate at step t
$T = 1000$: Total diffusion steps
I: Identity matrix, This generates temporally consistent pseudo-anomalies by gradually corrupting real frames over T steps

Evaluation and Validation

To enable discrete instance analysis, the surveillance video input is pre-processed for this study by being divided into individual frames. Data augmentation methods, including as cropping, rotation, and color changes, are used to increase the dataset's diversity and strengthen the model's robustness. After then, the divided dataset is used as test, validation, and training sets. This helps the M-GAN model understand common behavior patterns and improves its capacity to detect abnormalities in a variety of real-world surveillance scenarios.

Threshold	TPR	FPR	F1-Score
0.30	0.92	0.08	0.89
0.50	0.88	0.05	0.91
0.70	0.81	0.03	0.87
0.90	0.73	0.01	0.80

Table 1: Threshold Sensitivity Table

The sensitivity analysis of the MGAN model's abnormal event detection performance at different threshold settings is displayed in Table 1. As the threshold increases from 0.30 to 0.90, the TPR (True Positive Rate) decreases, indicating less anomalies, and the FPR (False Positive Rate) also decreases, showing fewer false alarms.The F1-score peaks at a threshold of 0.50, indicating that recall and precision are best balanced in this situation. This analysis aids in determining the ideal threshold to reduce false positives and increase detection accuracy.

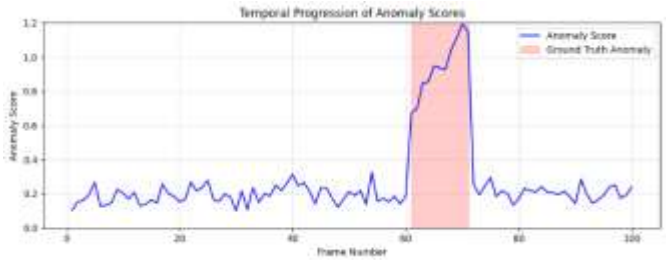

Fig. 3. Temporal Anomaly Score Progression

The temporal progression of anomaly scores for an example video clip is displayed in Figure 3. The anomaly score that the MGAN model awarded to each frame is shown by the blue curve. The shaded red region indicates the ground truth interval during which an abnormal event occurs. As can be

observed, the anomaly scores, which remain lower throughout normal activity and rise sharply during the anomalous phase, illustrate the model's ability to identify abnormal events in real time.

B. Performance Metrices

A number of important assessment criteria are used to gauge the abnormal event detection performance of the model. As a broad indicator of the model's efficacy, accuracy measures how accurate the predictions are overall. The model's specificity in identifying anomalous occurrences is demonstrated by precision, which shows the percentage of true positives among all reported anomalies. Conversely, recall evaluates the model's sensitivity by gauging its capacity to spot real anomalies. An all-encompassing metric that strikes a compromise between precision and recall, the F1-score is especially useful in situations involving unbalanced datasets. Furthermore, the model's detection capabilities are assessed using the AUC-ROC (Area Under the Receiver Operating Characteristic Curve), which compares the model's capacity to discriminate between normal and abnormal events across a range of categorization criteria.

C. Improvisations

A number of enhancements could be made to this study on anomaly identification in surveillance:

Temporal Analysis

Integrating models such as RNNs or LSTM networks with CNNs can enhance temporal pattern recognition, enabling the identification of evolving behaviors that may signify abnormal events.

Attention Mechanisms

Incorporating self-attention layers into the model may enhance its ability to focus on critical details in complex or cluttered environments, thereby improving its sensitivity to subtle anomalies.

Transfer Learning

Abnormal event detection can be enhanced across various contexts by applying pre-trained models to larger datasets, thereby eliminating the need for extensive data collection at each new site. The integration of M-GAN with autoencoders or other abnormal event detection frameworks can further improve the model's robustness.

Instantaneous Optimization

Instantaneous abnormal event detection can be facilitated by the use of edge computing and lightweight processing approaches, which increases its scalability and suitability for deployment in practical applications. Additionally, by integrating active learning frameworks, the model can adapt based on user feedback, thereby enhancing its accuracy and flexibility over time.

IV.EXPERIMENTAL RESULTS

A. Dataset and Setup

The model is evaluated under varying conditions, including different lighting levels (low-light and well-lit environments) and crowd densities (sparse to crowded settings). Testing is conducted using benchmark datasets such as ShanghaiTech, which includes urban surveillance videos with varying crowd densities, and UCSD, which focuses on pedestrian walkways. These datasets and conditions guarantee the experiment's reproducibility and show how well the model handles actual surveillance problems.

Dataset	Frame AUC	Pixel AUC	EER
UCSD Ped1	98.20%	92.50%	3.10%
UCSD Ped2	98.70%	93.80%	2.80%
CUHK Avenue	97.50%	91.20%	3.50%
ShanghaiTech	95.80%	89.40%	4.20%

Table 2: Quantitative Results for Anomaly Detection with MGAN

B. Evaluation Metrix

The evaluation of the effectiveness and reliability of the proposed Multi-Branch GAN (MGAN) abnormal event detection framework is conducted using a variety of assessment methods. The abnormal event detection and computer vision groups have embraced these measures extensively because they offer a comprehensive knowledge of model performance, particularly when class imbalance and real-world data variability are present.

Recall/ TPR (True Positive Rate)

The TPR (True Positive Rate), also known as recall, quantifies the model's ability to correctly identify true anomalous occurrences among all real anomalies in the dataset. In surveillance applications, where failure to detect an abnormality might have major repercussions, a system with a high recall is effective at minimizing missed detections, also known as false negatives.

$$\text{Recall (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

True Positives (TP): The quantity of anomalies that are accurately detected.
FN (False Negatives): The quantity of abnormalities that the model failed to detect.

Precision

The percentage of expected anomalies that are actually abnormal is known as precision. A model with high precision generates fewer false alarms, which is crucial for minimizing needless interventions and preserving operator confidence in automated surveillance systems.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

FP (False Positives): The quantity of typical occurrences that are mistakenly reported as abnormal.

F1-Score

The F1-Score, which is the harmonic mean of precision and recall, is a balanced statistic that is very useful when working with imbalanced datasets, as is often the case in abnormal event detection settings. It ensures that both the ability to detect abnormalities and the accuracy of those detections are taken into consideration.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A high F1-Score indicates that the model does well overall, attaining good recall and precision.

FPR (False Positive Rate)

The percentage of typical occurrences that are mistakenly labeled as anomalies is measured by the False Positive Rate (FPR). To maintain the system's practical usage and avoid flooding operators with false alarms, a low FPR is preferred.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

True Negatives (TN): The quantity of accurately recognized typical occurrences.

AUC-ROC (Area Under the ROC Curve)

The threshold-independent AUC-ROC (Area Under the Receiver Operating Characteristic Curve) measures the model's ability to distinguish between normal and abnormal events across all possible categorization thresholds. Perfect discrimination is represented by an AUC of 1.0, whereas random guessing is indicated by an AUC of 0.5. When comparing various models or configurations, this measure is especially useful.

Performance Metrics in Real Time

Real-time surveillance systems need to respond quickly in addition to having accurate detection capabilities. In order to make sure the model satisfies the operational requirements of real-world deployments, inference latency, time required to process each frameand throughput, the quantity of frames processed per secondare also assessed.

C. Results Analysis

The MGAN model is a powerful tool for detecting anomalies in surveillance video. Its ability to identify the difference between normal frames and abnormal frames demonstrates its capacity to identify the difference between normal and aberrant events. With scores primarily grouped around 0.25, the normal frame distribution displays a concentrated pattern, suggesting high confidence in identifying normal patterns. This narrow spread is desirable for stable performance in real-world applications. On the other hand, the anomaly scores for anomalous frames exhibit a broader distribution, centered around 0.85-1.0, reflecting the departure from normality captured by the MGAN framework.

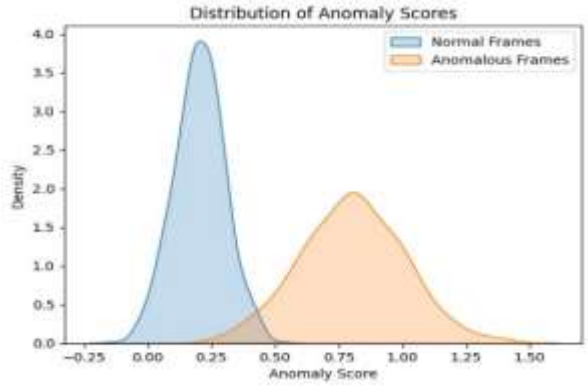

Fig.4 Anomaly Score Distribution Analysis

The minimal overlap between the two distributions around the 0.5 mark indicates an optimal decision boundary for abnormal event classification, outperforming previous approaches that struggle with ambiguous boundary regions. While retaining discriminative capability, the MGAN's transfer learning process improves the capacity of the model to generalize across different surveillance contexts. The anomaly score distributions validate the design choices in the multi-branch generator architecture, achieving more robust feature transfer between domains and resulting in more reliable abnormal event detection. The distribution of anomaly scores for normal and anomalous frames is displayed in Figure 4. The orange curve represents abnormal frames, while the blue curve represents normal frames. The fact that the two distributions are clearly separated suggests that anomalies receive substantially higher scores from the MGAN model than do typical occurrences.

Algorithm	MAE	RMSE	Accuracy (%)
MCD	0.241	0.371	84.0
LOF	0.289	0.412	82.5
IF	0.275	0.395	83.2
OCSVM	0.310	0.440	81.7

Table 3: Error Metrics Comparison

Table 3 displays the error metrics for four different abnormal event detection algorithms: accuracy, MAE (mean absolute error), and RMSE (root mean square error). With the best accuracy (84.0%) and the lowest MAE (0.241) and RMSE (0.371), the MCD algorithm performs better than the other methods in terms of prediction performance. Conversely, OCSVM has the highest error values and the lowest accuracy. These data show how successfully the MCD methodology discovers anomalies in comparison to other traditional methods.

V. FUTURE WORKS

Even though the proposed Multi-Branch Generative Adversarial Network (MGAN) framework shows a lot of potential for improving the precision and versatility of abnormal event identification in surveillance systems, there are still a number of areas that may use more research and development. The incorporation of multi-modal data sources is one encouraging avenue. The visual cues from surveillance footage are the mainstay of current methods; however, adding other modalities like audio signals, sensor data, or contextual metadata (like the time of day, the weather, or the density of people) could yield a more comprehensive understanding of the monitored environment. The system may be able to detect subtle or context-dependent anomalies that are difficult to detect with visual data alone if these disparate data streams are fused using sophisticated deep learning architectures.The creation of more efficient unsupervised and semi-supervised learning techniques is a crucial topic for further study. Even if MGAN already lessens the demand for labeled anomaly data, it is still crucial to minimize the necessity for manual annotation. Techniques including self-supervised learning, contrastive learning, and active learning could be investigated to increase the ability of model to acquire reliable representations from unlabelled or poorly labelled data. Additionally, by bridging the gap between training and deployment environments, domain adaptation approaches and synthetic data production can improve generalization to previously experienced situations.

Scalability and real-time deployment are also essential factors for realistic surveillance applications. Future research should concentrate on improving MGAN's computing efficiency, perhaps by using lightweight neural network topologies, pruning, or model compression. This would guarantee that the system can function efficiently on edge devices or in dispersed surveillance networks, satisfying the throughput and latency demands of actual settings. The interpretability and transparency of abnormal event detection models represent another important research topic.. It is crucial to create techniques that offer concise justifications for anomalies found and system judgments made as AI-driven surveillance systems proliferate. In order to boost confidence and enable human-in-the-loop supervision, explainable AI techniques like saliency mapping and attention mechanisms should be used to assist operators and stakeholders in comprehending the reasoning behind warnings. Furthermore, resilience to environmental changes and hostile attacks remains a challenge.

Future studies should incorporate adversarial training, domain generalization, and continuous learning to ensure the model maintains its performance even when faced with changing conditions, such as changing camera angles, illumination, or scene composition. Finally, the ethical and societal implications of putting intelligent surveillance systems in place necessitate ongoing thought. Future research should include developing privacy-preserving techniques, community feedback systems, and frameworks to ensure accountability and equity in abnormal event detection outcomes.In conclusion, developing MGAN-based abnormal event detection will necessitate a multidisciplinary strategy that blends technological advancement with moral leadership, guaranteeing that future surveillance systems are not only more precise and effective but also reliable and socially conscious.

VI. CONCLUSION

The paper offers a reliable and flexible method for identifying unusual occurrences in security footage by fusing M-GAN and transfer learning. By identifying deviations as possible anomalies, M-GAN's generative capabilities enable it to precisely identify typical behavioral patterns in the surveillance data. By employing GANs to duplicate frequent occurrences, the ability of model to identify the difference between normal and suspicious activity is enhanced. This is crucial for surveillance applications. Transfer learning further strengthens the model by allowing it to utilize knowledge from pre-trained datasets, enabling quicker adaptation to new environments with limited labeled data. By identifying deviations as possible anomalies, M-GAN's generative capabilities enable it to precisely identify typical behavioral patterns in the surveillance data. By utilizing GANs to mimic common events, the ability of model to identify the difference between suspicious and regular behavior is enhanced. By prioritizing efficient processing, it can be deployed in multiple settings, from urban surveillance to industrial monitoring. Future directions might involve optimizing the model’s processing speed to achieve real-time performance on larger datasets and deploying it in diverse surveillance contexts to validate its scalability and generalization capabilities across a broader range of applications. This multi-faceted approach showcases the potential of M-GAN and transfer learning in advancing surveillance technology to provide accurate, reliable, and adaptive abnormal event detection.

REFERENCES

1. Bashar, M. A., & Nayak, R. (2020). TAnoGAN: Time series anomaly detection with generative adversarial networks. IEEE Symposium on Computational Intelligence (SSCI), Canberra, ACT, Australia, pp. 1778–1785. doi: 10.1109/SSCI47803.2020.9308512.
2. Arun Vignesh Malarkkan, Dongjie Wang, and Yanjie Fu.(2024). Multi-view Causal Graph Fusion Based Anomaly Detection in Cyber-Physical Infrastructures. 33rd ACM International Conference on Information and Knowledge Management
3. Li, N., Chang, F., & Liu, C. (2022). A self-trained spatial graph convolutional network for unsupervised human- related anomalous event detection in complex scenes. IEEE Transactions on Cognitive and Developmental Systems, 1–1.
4. Li, N., Zhong, J.-X., Shu, X., & Guo, H. (2022). Weakly- supervised anomaly detection in video surveillance via graph convolutional label noise cleaning. Neurocomputing, 481, 154–167.
5. N. Palanivel, V. Keerthana, N. Monisha, S. Sandhiya (2023) "Object Detection and Recognition In Dark Using YOLO" Journal of Data Acquisition and Processing 38 (2), 57, 2023.
6. Lu, Y., Yu, F., Reddy, M. K. K., & Wang, Y. (2020). Few-shot scene-adaptive anomaly detection. European Conference on Computer Vision (ECCV).
7. Luo, W., Liu, W., & Gao, S. (2021). Graph convolutional neural network for skeleton-based video abnormal behavior detection. Generalization with Deep Learning, 139–155.
9. Markovitz, A., Sharir, G., Friedman, I., Zelnik-Manor, L., & Avidan, S. (2020). Graph Embedded Pose Clustering for Anomaly Detection.
10. Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., Liu, Y., Zhang, S., & Chen, K. RTMDet: An empirical study of designing real-time object detection models.
11. Markovitz, A., Sharir, G., Friedman, I., Zelnik-Manor, L., & Avidan, S. (2020). Graph Embedded Pose Clustering for Anomaly Detection.
12. Wang, Y., Liu, T., Zhou, J., & Guan, J. (2023). Video anomaly detection based on spatiotemporal relationships among objects. Neurocomputing, 532, 141–151.
13. Wang, Z., Zou, Y., & Zhang, Z. (2020). Cluster attention contrast for video anomaly detection. ACM International Conference on Multimedia, 2463–2471.
14. Xu, X., Jiang, Y., Chen, W., Huang, Y., Zhang, Y., & Sun, X. (2022). Damo-yolo: A report on real-time object detection design. arXiv preprint arXiv:2211.15444v2.
15. ang, Y., Xian, Y., Fu, Z., & Naqvi, S. M. (2021). Video anomaly detection for surveillance based on effective frame area. IEEE 24th International Conference on Information Fusion (FUSION).
16. Yang, Y., Fu, Z., & Naqvi, S. M. (2022). A two-stream information fusion approach to abnormal event detection in video. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 5787–5791
17. H. Park, J. Noh, and B. Ham, ''Learning memory-guided normality for anomaly detection,'' in Proc. CVPR, Jun. 2020, pp. 14372–14381.
18. J. Peng, Y. Zhao, and L. Wang, ''A survey of video abnormal behavior detection based on deep learning,'' Laser Optoelectron Prog., vol. 58, no. 6,pp. 51–61, 2021
19. T. Ganokratanaa, S. Aramvith, and N. Sebe, ''Unsupervised anomaly detection and localization based on deep spatiotemporal translation net-work,'' IEEE Access, vol. 8, pp. 50312–50329, 2020.
20. Y. Qiang, S. Fei, and Y. Jiao, ''Anomaly detection based on latent feature training in surveillance scenarios,'' IEEE Access, vol. 9, pp. 68108–68117,2021.
21. H. Prawiro, J.-W. Peng, T.-Y. Pan, and M.-C. Hu, ''Abnormal event detection in surveillance videos using two-stream decoder,'' in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW), Jul. 2020.
22. F. Dong, Y. Zhang, and X. Nie, ''Dual discriminator generative adversarial network for video anomaly detection,'' IEEE Access, vol. 8, pp. 88170–88176, 2020.
23. R. Cai, H. Zhang, W. Liu, S. Gao2, and Z. Hao, ''Appearance-motion memory consistency network for video anomaly detection,'' in Proc. AAAI,2021
24. Haoyi Fan, Fengbin Zhang and Zuoyong Li, "Anomalydae: Dual autoencoder for anomaly detection on attributed networks", IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 5685-5689, 2020.
25. Guansong Pang, Chunhua Shen, Longbing Cao and Anton van den Hengel, "Deep learning for anomaly detection: A review", ACM Computing Surveys, vol. 54, no. 2, pp. 38:1- 38:38, 2021.
26. Eric Jardim, Lucas A. Thomaz, Eduardo A. B. da Silva and Sergio L. Netto, "Domain-transformable sparse representation for anomaly detection in moving-camera videos", IEEE Transactions on Image Processing, vol. 29, pp. 1329-1343, 2020
27. Kang Zhou, Yuting Xiao, Jianlong Yang, Jun Cheng, Wen Liu, Weixin Luo, et al., "Encoding structure-texture relation with p-net for anomaly detection in retinal images", Proceeding of 16th European

Conference on Computer Vision (ECCV), vol. 12365, pp. 360-377, 2020.
28. Shaista Hussain, Ayesha Anees, Ankit Das, Binh P. Nguyen, Mardiana Marzuki, Shuping Lin et al., "High-content image generation for drug discovery using generative adversarial networks", Neural Networks, vol. 132, pp. 353-363, 2020
29. Ruoying Wang, Kexin Nie, Tie Wang, Yang Yang and Bo Long, "Deep learning for anomaly detection", Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 894-896, 2020
30. N. Patel, A. N. Saridena, A. Choromanska, P. Krishnamurthy, and F. Khorrami, ‘‘Learning-based real- time process-aware anomaly monitoring for assured autonomy,’’ IEEE Trans. Intell. Vehicles, vol. 5, no. 4,pp. 659–669, Dec. 2020.