

Research Article

Exploring the Magnitude-Shape Plot Framework for Anomaly Detection in Crowded Video Scenes

Zuzheng Wang¹, Fouzi Harrou¹, Ying Sun¹, Marc G Genton¹

1. Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology, Saudi Arabia

Detecting anomalies in crowded video scenes is critical for public safety, enabling timely identification of potential threats. This study explores video anomaly detection within a Functional Data Analysis framework, focusing on the application of the Magnitude-Shape (MS) Plot. Autoencoders are used to learn and reconstruct normal behavioral patterns from anomaly-free training data, resulting in low reconstruction errors for normal frames and higher errors for frames with potential anomalies. The reconstruction error matrix for each frame is treated as multivariate functional data, with the MS-Plot applied to analyze both magnitude and shape deviations, enhancing the accuracy of anomaly detection. Using its capacity to evaluate the magnitude and shape of deviations, the MS-Plot offers a statistically principled and interpretable framework for anomaly detection. The proposed methodology is evaluated on two widely used benchmark datasets, UCSD Ped2 and CUHK Avenue, demonstrating promising performance. It performs better than traditional univariate functional detectors (e.g., FBPlot, TVDMSS, Extremal Depth, and Outliergram) and several state-of-the-art methods. These results highlight the potential of the MS-Plot-based framework for effective anomaly detection in crowded video scenes.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

Anomaly detection in video sequences is a critical task in computer vision, particularly in scenarios involving crowded environments such as public spaces, events, transportation hubs, and marketplaces^[1]^[2]^[3]. The goal is to identify unusual or potentially dangerous events that deviate from expected crowd

behavior, such as sudden dispersals, confrontations, or individuals engaging in atypical activities^{[4][5]}. Early detection of such anomalies is crucial for ensuring public safety, managing emergencies, and preventing incidents. In crowded areas, the task of anomaly detection becomes especially challenging due to the complexity and high dimensionality of spatiotemporal data. Video sequences in these environments consist of continuous frames with numerous individuals and objects interacting, creating intricate patterns of motion and behavior^[6]. This complexity requires the use of advanced methods capable of capturing not only spatial features within individual frames, but also temporal and contextual dependencies that characterize typical crowd dynamics^{[7][8]}.

Over the past two decades, researchers have developed numerous methods to effectively identify abnormal events in video data^[9]. Trajectory-based approaches track moving objects to detect deviations from typical patterns^{[10][11][12]}. For instance, the study in^[13] explored anomaly detection through trajectory analysis using a single-class support vector machine (SVM) clustering approach, effectively identifying unusual trajectories without prior knowledge of outlier distributions. In^[14], snapped trajectories are introduced as a high-level representation that reduces computational load and identifies key scene regions. This combined approach detects anomalies in speed, direction, and finer motion details with fewer false alarms. Trajectory-based methods rely on accurate detection and tracking but are limited by crowd density, resolution, motion, and occlusion, making them more suitable for sparse crowds. However, in crowded scenes, frequent occlusions, intersecting paths, and the challenge of distinguishing similar individuals^[15] reduce their effectiveness, and they lack contextual awareness for identifying anomalies.

Alternatively, other techniques use spatiotemporal features to represent events in videos without requiring trajectory analysis. These methods capture motion dynamics by analyzing changes in pixel intensities and patterns over time^[16]. Techniques such as spatiotemporal gradients^[17], Histograms of Oriented Gradients (HOG)^[18], 3D spatiotemporal gradient^[19], and Histograms of Optical Flow (HOF)^[20] help detect collective crowd behaviors and abnormalities at the pixel or region level, enhancing anomaly detection in dense and dynamic scenes. For example, in^[21], an SVM classifier trained on HOG features is proposed to automatically detect violent activities in surveillance videos, distinguishing actions like kicking and punching. This system triggers alerts for detected violence and monitors loitering duration to flag suspicious behavior in real-time, thus enhancing traditional surveillance effectiveness. In^[22], a method is introduced for recognizing overlapping and multi-dimensional actions

using a spatiotemporal representation and enhanced Motion History Image (MHI). It employs Speeded-Up Robust Features (SURF) to capture key motion features, gradient-based optical flow for motion representation, and RANSAC (Random Sample Consensus) for outlier removal. The nearest neighbor classifier with leave-one-out cross-validation achieves improved recognition rates for complex actions in outdoor scenes compared to traditional MHI approaches. The study in^[23] introduces a hierarchical framework for video anomaly detection that integrates multi-level features, including 3D-SIFT (scale-invariant feature transform), HOF, and HOG, with Gaussian process regression. However, these low-level visual features (e.g., motion or texture) may still struggle to fully capture higher-level semantic information or complex contextual interactions in crowded scenes, limiting their effectiveness in more challenging anomaly detection scenarios^[6]. While effective in some cases, these methods often face limitations in generalizing across diverse environments and in capturing complex and subtle deviations from normal patterns.

To address these limitations, deep learning-based methods have emerged as advanced alternatives for video anomaly detection, leveraging their capacity to capture high-level semantic information and complex interactions in crowded and dynamic environments^[6]. Unlike traditional techniques, deep learning models such as Convolutional Neural Networks (CNNs)^{[24][25][26]}, Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTMs) can automatically learn and extract relevant spatial and temporal features directly from the data^{[27][28]}. These models are particularly effective in modeling nuanced deviations from normal patterns, enabling them to generalize better across diverse scenarios. Moreover, autoencoders and Generative Adversarial Networks (GANs) are commonly used to reconstruct normal patterns in video frames, with deviations in reconstruction used as indicators of potential anomalies^{[29][30][31]}. The ability of deep learning models to handle complex feature representations has significantly improved the performance and adaptability of video anomaly detection systems across various application settings^{[4][6][8][32]}. The study in^[24] presents a method for anomaly detection in crowded scenes using fully convolutional networks (FCNs) with temporal data. The model enhances feature extraction by combining a pre-trained CNN (a modified AlexNet) with a custom convolutional layer tailored to specific video data. In^[33], a spatial-temporal CNN model is proposed for anomaly detection and localization using spatial-temporal volumes with motion information from static camera scenes. By capturing both appearance and motion features through spatial-temporal convolutions, the model enhances robustness. Evaluated on four benchmark datasets, it outperforms state-of-the-art methods, especially on challenging pixel-level criteria. In^[34], a video anomaly detection

approach leverages a frame-to-label and motion (F2LM) generator to intentionally reduce the quality of abnormal regions, followed by a Destroyer that transforms these areas into zero vectors, making anomalies more prominent. This technique surpasses state-of-the-art performance on the UCSD Ped2, CUHK Avenue, and Shanghai Tech datasets. The study^[35] proposes an efficient frame-level video anomaly detection (VAD) method that utilizes transfer learning and fine-tuning on 20 CNN-based deep learning models, including variants of VGG, Xception, MobileNet, and ResNet.

Over the past decade, numerous reconstruction-based methods for video anomaly detection have emerged, leveraging models that learn to reconstruct normal video frames or sequences and using reconstruction errors to identify anomalies^{[36][37][38]}. The core idea of these approaches is that the model trained exclusively on anomaly-free data will struggle to accurately reconstruct anomalous events, resulting in higher reconstruction errors for frames with unusual or abnormal content. This discrepancy between normal and abnormal reconstructions provides a basis for detecting anomalies effectively. In^[39], Zhao et al. tackle the challenge of detecting anomalies in complex video scenes by introducing a spatiotemporal AutoEncoder (STAE) that uses deep neural networks with 3D convolutions to learn spatial and temporal features. It incorporates a weight-decreasing prediction loss for future frame generation, improving motion feature learning beyond standard reconstruction loss. In^[40], a residual spatiotemporal autoencoder (STAE) is proposed for video anomaly detection, where anomalies are identified as deviations from normal patterns using reconstruction loss. Residual connections enhance model performance, effectively reconstructing normal frames with low cost while detecting irregularities as abnormal frames. Another study in^[41] proposed a sparse coding-inspired Deep Neural Network (DNN) for video anomaly detection, known as Temporally-coherent Sparse Coding (TSC). TSC maintains frame similarity using a temporal coherence term optimized via the Sequential Iterative Soft-Thresholding Algorithm (SIATA). The enhanced stacked Recurrent Neural Network Autoencoder (sRNN-AE) model introduces data-dependent similarity, reduced model depth for real-time detection, and temporal pooling for efficiency. The study in^[42] considered a Residual Variational Autoencoder (RVAE) for unsupervised video anomaly detection, which captures complex patterns and minimizes reconstruction error through low-dimensional latent encoding and decoding. The model incorporates a ConvLSTM layer for improved spatiotemporal learning and uses residual connections to address the vanishing gradient problem. Recently, in^[43], a deep multiplicative attention-based autoencoder (DeMAAE) was introduced for video anomaly detection. DeMAAE applies a global attention mechanism at the decoder to enhance feature learning, leveraging an attention map created from encoder-decoder hidden states to guide decoding via

a context vector. In^[38], a spatial-temporal network achieved 97.4% AUC on UCSD Ped2. A Multivariate Gaussian Fully Convolutional Adversarial Autoencoder in^[44] scored 91.6% AUC on UCSD Ped2. The spatiotemporal 3D Convolutional Auto-Encoder (ST-3DCAE) in^[45] achieved 85.3% on UCSD Ped2 and 75.8% on UCF-Crime. OF-ConvAE-LSTM in^[46] reached 92.9% on UCSD Ped2 and 89.5% on Avenue. Finally,^[47] achieved 86.4% on UCSD Peds1 and 88.9% on Avenue using a deep autoencoder with regularity-based thresholding.

While deep learning models have significantly advanced video anomaly detection, statistical methods, particularly functional data analysis (FDA), remain underexplored in this area. FDA offers unique advantages by treating each frame's reconstruction error as a continuous, multivariate function over time, enabling the capture of both temporal and spatial dependencies and allowing for a more effective distinction between typical fluctuations and true anomalies by analyzing the shape and magnitude of deviations^[48]. Despite these benefits, only a few studies have investigated statistical approaches in video anomaly detection. For example, Raymaekers et al. introduced a measure of directional outlyingness specifically applied to image and video data^[49], using a statistical method for anomaly detection based on directional outlyingness. However, this approach has only been validated on images and non-crowded video sequences, emphasizing the need for further testing in more complex, crowded environments. Another statistical approach for anomaly detection in high-dimensional data is the Depthgram method by Aleman et al.^[50], designed for functional data visualization in fMRI. Depthgram uses depth-based 2D representations to identify outliers, variability, and sample composition, supporting the exploration of neuroscientific patterns across individuals and brain regions. However, it is limited to static images, such as fMRI data, and has not been explored for video applications.

This study explores video anomaly detection in crowded scenes within a Functional Data Analysis framework, offering a statistical perspective to enhance the detection process. Many statistical video anomaly detection methods^{[51][52][53][38][54][55][56][34][57]} often rely on thresholds determined in a non-automatic and non-systematic manner, focusing primarily on anomalies in the mean or variance of reconstructed residuals. To mitigate these limitations, this work explores the application of the Magnitude-Shape (MS) Plot, a Multivariate Functional Data Visualization and Outlier Detection approach, for video anomaly detection^[48]. The MS-Plot provides a statistically principled and interpretable framework by treating reconstruction errors as multivariate functional data. It effectively captures both magnitude (amplitude deviations) and shape (pattern deviations), allowing for the detection of subtle

anomalies and complex patterns indicative of abnormal behavior. By simultaneously monitoring amplitude and shape deviations, this approach provides a comprehensive statistical methodology to analyze and understand reconstruction errors in video anomaly detection. In addition, the approach trains solely on normal data, making it effective in scenarios with scarcely labeled anomalies by learning typical patterns and identifying deviations during evaluation. To illustrate, this study investigates two reconstruction-based models: a simple autoencoder and the advanced MAMA autoencoder^[34]. These models learn and reconstruct normal behavior from anomaly-free datasets, producing low reconstruction errors for normal frames and significantly higher errors for anomalous frames during testing. The integration of the MS-Plot enhances this approach by jointly analyzing magnitude and shape deviations, offering a comprehensive framework for anomaly detection. This approach was evaluated on two publicly available datasets, UCSD Ped2 and CUHK Avenue, demonstrating superior performance compared to traditional univariate functional detectors, including Functional Boxplot (FBplot)^[58], Total Variation Depth with Modified Shape Similarity (TVDMSS), Extremal Depth (ED)^[59], and Outliergram (OG)^[60]. Unlike these univariate approaches, which focus on either shape or mean deviations, MS-Plot captures complex, multivariate outliers, proving more sensitive to both subtle and significant anomalies in video data. Also, the proposed approach surpassing state-of-the-art techniques.

The remaining sections of this paper are organized as follows. Section 2 provides a brief overview of the MS-Plot approach and its application in anomaly detection, along with a recap of the simple autoencoder and other variants utilized in this study. Section 3 details the proposed approach based on MS-Plot. Section 4 discusses the datasets employed and presents the evaluation results that validate the effectiveness of our method. Finally, Section 5 concludes the study.

2. Materials and methods

2.1. Problem statement

Video anomaly detection aims to identify frames with abnormal behavior by learning normal patterns during training and detecting deviations during testing. Frame-level detection evaluates each frame for anomalies, such as unexpected objects, irregular behaviors, or environmental changes, without relying on object- or scene-based approaches. Let X denote a video consisting of multiple frames $X = \{X_1, X_2, \dots, X_T\}$, where T represents the total number of frames. Each frame X_t at time t is represented as:

$$X_t = \{p_{ij}(t) \mid 1 \leq i \leq H, 1 \leq j \leq W\}, \quad (1)$$

where H and W are the height and width of the frame, and $p_{ij}(t)$ denotes the intensity of the pixel located at (i, j) in frame t .

The goal is to develop a model f that assigns a binary label y_t to each frame X_t such that:

$$y_t = f(X_t) \quad \text{where} \quad y_t \in \{0, 1\} \quad (2)$$

Here, $y_t = 0$ indicates a normal frame, while $y_t = 1$ indicates an anomalous frame. This detection is performed at the frame level.

Frame-level anomaly detection in videos presents challenges such as diverse anomaly types and limited labeled data. Anomalies range from sudden lighting changes to unexpected behaviors, complicating model generalization. The scarcity of labeled data necessitates unsupervised or semi-supervised methods focused on normal patterns. This study employs autoencoder models combined with the MS-Plot to enhance video anomaly detection. The following sections elaborate on the MS-Plot and its role in this process.

2.2. MS-Plot for Functional Data Outlier Detection

The MS-Plot, introduced by Dai and Genton (2018) ^[48], detects outliers in functional data by evaluating magnitude and shape deviations from a central region. Extending the concept of functional directional outlyingness, it measures both the extent and direction of deviations. This measure allows the MS-Plot to effectively capture two components of outlyingness: magnitude (distance from the central region) and shape (direction of deviation), making it suitable for both univariate and multivariate functional data.

- **Mean Directional Outlyingness (MO):** This quantifies the average magnitude of a function's deviation from the central region across its domain, effectively identifying anomalies in overall amplitude or intensity. For a function $X(t)$ with distribution F_X , it is defined as ^[48]:

$$MO(X, F_X) = \int_I O(X(t), F_{X(t)})w(t)dt, \quad (3)$$

where $O(X(t), F_{X(t)})$ represents the directional outlyingness at each point t in the domain I , and $w(t)$ is a weight function.

- **Variation of Directional Outlyingness (VO):** This measures the variability in directional outlyingness across the domain, capturing deviations in the shape or structure of the function. By assessing this variability, VO enables the detection of shape anomalies or irregular patterns within the data ^[48].

$$VO(X, F_X) = \int_I \|O(X(t), F_{X(t)}) - MO(X, F_X)\|^2 w(t)dt. \quad (4)$$

Functional Directional Outlyingness (FO) quantifies the overall outlyingness of a function, defined as ^[48]:

$$FO(X, F_X) = \|MO(X, F_X)\|^2 + VO(X, F_X). \quad (5)$$

In Equation (5), FO combines magnitude outlyingness $\|MO\|$, representing deviation extent, and shape outlyingness VO , reflecting structural variability. This decomposition enhances the ability to quantify centrality and identify abnormal patterns in functional data.

The MS-Plot technique classifies observations as normal or anomalous by applying a threshold to the FO measure, which combines magnitude and shape deviations. Dai and Genton^[48] proposed calculating FO using MO and VO. Assuming normal data distribution and using random projection point-wise depth, they compute functional directional outlyingness. The method utilizes the squared robust Mahalanobis distance (SRMD) for $(MO, VO)^T$, where the covariance matrix is determined using the Minimum Covariance Determinant (MCD) algorithm^[61]. The SRMD's tail distribution is approximated with a Fisher's F distribution^[62], and curves with SRMD values exceeding the threshold are flagged as outliers.

2.3. Autoencoder-Based Video Anomaly Detection

An autoencoder is a neural network framework that learns to compress input data into a latent representation and then reconstruct it back to its original form^[63]. This process is facilitated by two key components: an encoder, which reduces the data's dimensionality, and a decoder, which restores it^[64].

- **Encoder:** transforms the input data into a compact latent space, preserving critical features and filtering out noise and unnecessary details.
- **Decoder:** Reconstructs the input data from the latent representation, aiming to reduce the reconstruction error by aligning the output closely with the original input.

Let $\mathbf{X}$ represent an input frame. The encoder function E transforms $\mathbf{X}$ into a latent representation $\mathbf{Z}$, and the decoder function D reconstructs the original frame as $\hat{\mathbf{X}}$:

$$\mathbf{Z} = E(\mathbf{X}), \quad \hat{\mathbf{X}} = D(\mathbf{Z}). \quad (6)$$

During training, the autoencoder is optimized to minimize a reconstruction loss, commonly the Mean Squared Error (MSE) between the input $\mathbf{X}$ and its reconstruction $\hat{\mathbf{X}}$:

$$L(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \hat{\mathbf{X}}_i)^2, \quad (7)$$

where N is the total number of pixels in the frame.

Figure 1 presents a flowchart of an autoencoder for video anomaly detection. Trained solely on normal frames, the autoencoder minimizes reconstruction errors for normal patterns. Anomalous frames deviating from these patterns result in higher reconstruction errors, forming the basis for anomaly detection.

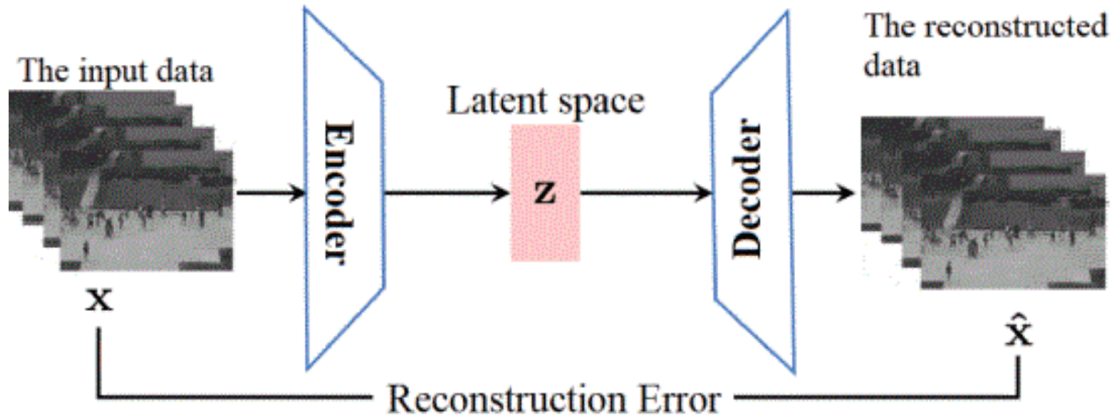


Figure 1. Basic Autoencoder for Video Anomaly Detection.

The process of using an autoencoder for video anomaly detection is summarized as follows.

1. **Training Phase:** The autoencoder is trained on a dataset of normal video frames, optimizing it to reconstruct typical, expected frames.
2. **Testing Phase:** For each test frame $\mathbf{X}_t$, the autoencoder reconstructs the frame as $\hat{\mathbf{X}}_t$ and computes the reconstruction error:

$$e_t = \|\mathbf{X}_t - \hat{\mathbf{X}}_t\|, \quad (8)$$

where e_t denotes the overall reconstruction error for frame $\mathbf{X}_t$.

3. **Anomaly Detection:** A frame is identified as an anomaly when the reconstruction error e_t surpasses a specified threshold τ .

$$y_t = \begin{cases} 1 & \text{if } e_t > \tau \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Here, $y_t = 1$ indicates an anomalous frame, while $y_t = 0$ indicates a normal frame.

Autoencoders effectively detect video anomalies by learning compact representations of normal patterns, highlighting subtle deviations through reconstruction errors^[65]. Trained exclusively on normal data, they excel in unsupervised settings, making them suitable for applications like surveillance and traffic

monitoring^[6]. Their ability to flag deviations without labeled anomalies is especially valuable in scenarios with limited anomaly data^[37]. Various enhancements have been proposed to improve their detection capabilities^{[66][39][34][57]}.

3. MS-Plot for Video Anomaly Detection with Reconstruction Models

This study introduces a framework that integrates the MS-Plot with reconstruction-based models for anomaly detection. Residuals (reconstruction errors) generated by these models are treated as multivariate functional data and analyzed using the MS-Plot to identify anomalies. Frame-level anomaly detection involves learning normal behavior patterns, calculating reconstruction errors, and applying the MS-Plot for multivariate functional outlier analysis. This enhances detection accuracy and versatility, enabling the identification of diverse anomalies in video sequences. For simplicity, the MS-Plot framework is demonstrated with an autoencoder model, but it can be extended to other reconstruction-based models.

This approach consists of three steps: training a reconstruction-based model on anomaly-free data to learn normal patterns, generating residuals by comparing input and reconstructed frames during testing, and analyzing these residuals with the MS-Plot to detect anomalies based on magnitude and shape outlyingness. This integration of reconstruction learning and multivariate analysis ensures effective anomaly detection.

3.1. Training Phase - Learning Normal Behavior

During training, only normal frames are provided to the model. The model is typically an autoencoder comprising an encoder function E and a decoder function D . The model aims to reconstruct the input frame $\mathbf{X}_t$ as $\hat{\mathbf{X}}_t$:

$$\hat{\mathbf{X}}_t = D(E(\mathbf{X}_t)) \quad (10)$$

The autoencoder is trained to minimize a reconstruction loss function $L(\mathbf{X}_t, \hat{\mathbf{X}}_t)$, which measures the difference between the original frame $\mathbf{X}_t$ and the reconstructed frame $\hat{\mathbf{X}}_t$. A common choice for L is the mean squared error (MSE):

$$L(\mathbf{X}_t, \hat{\mathbf{X}}_t) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (p_{ij}(t) - \hat{p}_{ij}(t))^2 \quad (11)$$

where $\hat{p}_{ij}(t)$ represents the intensity of the reconstructed pixel at location (i, j) .

3.2. Testing Phase - Residual Generation and Anomaly Detection

In the testing phase, the model receives frames from unseen video sequences. For each frame $\mathbf{X}_t$, it computes the reconstruction $\hat{\mathbf{X}}_t$ and the corresponding reconstruction error matrix:

$$\mathbf{E}_t = \{e_{ij}(t) \mid 1 \leq i \leq H, 1 \leq j \leq W\} \quad (12)$$

where $e_{ij}(t) = |p_{ij}(t) - \hat{p}_{ij}(t)|$ denotes the absolute difference between the original and reconstructed pixel intensities.

To enhance anomaly detection, the reconstruction error matrices $\mathbf{E}_t$ are treated as multivariate functional data. For each frame $\mathbf{X}_t$, the MS-Plot calculates MO and VO to detect deviations from normal patterns. The directional outlyingness $O(\mathbf{E}_t, F)$ quantifies the extent to which the error matrix $\mathbf{E}_t$ deviates from a reference distribution F :

$$MO(\mathbf{E}_t, F) = \int_I O(\mathbf{E}_t, F_{\mathbf{E}_t}) w(t) dt, \quad VO(\mathbf{E}_t, F) = \int_I \|O(\mathbf{E}_t, F_{\mathbf{E}_t}) - MO(\mathbf{E}_t, F)\|^2 w(t) dt. \quad (13)$$

The MS-Plot visualizes $|MO|$ against VO to distinguish normal from anomalous frames, with the relationship defined as:

$$FO = \|MO\|^2 + VO. \quad (14)$$

The MS-Plot applies a threshold on functional outlyingness FO to classify each frame $\mathbf{X}_t$ as normal or anomalous. The decision rule is:

$$y_t = \begin{cases} 1 & \text{if } FO(\mathbf{E}_t) > \tau_{FO} \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Algorithm 1 summarizes the steps for using the MS-Plot to analyze residuals and detect frame-level anomalies.

Algorithm 1 MS-Plot framework for anomaly detection based on reconstruction residuals.

Require: Functional data (reconstruction error matrices) $\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n\}$, Reference distribution $F_{\mathbf{E}}$, Threshold τ_{FO}

Ensure: Anomaly detection result for each $\mathbf{E}_i$ (0: normal, 1: anomaly)

- 1: Compute the reference distribution $F_{\mathbf{E}(t)}$ (e.g., median function) at each time point t
- 2: **for** each error matrix $\mathbf{E}_i$ **do**
- 3: **Compute directional outlyingness** $O(\mathbf{E}_i(t), F_{\mathbf{E}(t)})$ at each time point t :

$$O(\mathbf{E}_i(t), F_{\mathbf{E}(t)}) = (\mathbf{E}_i(t) - F_{\mathbf{E}(t)}) \times \text{Sign}(\mathbf{E}_i(t) - F_{\mathbf{E}(t)})$$

- 4: **Calculate Mean Directional Outlyingness (MO) for $\mathbf{E}_i$:**

$$MO(\mathbf{E}_i) = \frac{1}{T} \sum_{t=1}^T O(\mathbf{E}_i(t), F_{\mathbf{E}(t)})$$

- 5: **Calculate Variation of Directional Outlyingness (VO) for $\mathbf{E}_i$:**

$$VO(\mathbf{E}_i) = \frac{1}{T} \sum_{t=1}^T (O(\mathbf{E}_i(t), F_{\mathbf{E}(t)}) - MO(\mathbf{E}_i))^2$$

- 6: **Calculate Functional Outlyingness (FO) for $\mathbf{E}_i$:**

$$FO(\mathbf{E}_i) = \|MO(\mathbf{E}_i)\|^2 + VO(\mathbf{E}_i)$$

- 7: **end for**
 - 8: **Construct MS-Plot:** Plot $|MO(\mathbf{E}_i)|$ on the x-axis and $VO(\mathbf{E}_i)$ on the y-axis for each error matrix $\mathbf{E}_i$
 - 9: **Set Threshold and Detect Outliers:**
 - 10: **for** each error matrix $\mathbf{E}_i$ **do**
 - 11: **if** $FO(\mathbf{E}_i) > \tau_{FO}$ **then**
 - 12: Classify $\mathbf{E}_i$ as anomaly ($y_i = 1$)
 - 13: **else**
 - 14: Classify $\mathbf{E}_i$ as normal ($y_i = 0$)
 - 15: **end if**
 - 16: **end for**
 - 17: **return** Anomaly labels $\{y_1, y_2, \dots, y_n\}$ for all error matrices
-

3.3. Evaluation Metrics for Anomaly Detection in Videos

In video anomaly detection, evaluation metrics like True Positive Rate (TPR), False Positive Rate (FPR), Precision, F1 Score, and Accuracy are essential for assessing the effectiveness of identifying anomalies while minimizing false detections. These metrics quantify detection effectiveness by comparing model predictions against ground truth labels. The mathematical definitions of these metrics are summarized in Table 1, where TP, FP, TN, and FN represent True Positives, False Positives, True Negatives, and False Negatives, respectively.

Metric	Equation
True Positive Rate (TPR)	$TPR = \frac{TP}{TP+FN}$
False Positive Rate (FPR)	$FPR = \frac{FP}{FP+TN}$
Precision	$Precision = \frac{TP}{TP+FP}$
F1 Score	$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$
Accuracy	$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

Table 1. Summary of Evaluation Metrics for Video Anomaly Detection.

TPR measures the proportion of actual anomalies correctly identified, while FPR indicates the rate of normal frames misclassified as anomalies, a critical factor in reducing false alarms in video data. Precision evaluates the accuracy of predicted anomalies, reflecting the proportion of correctly identified anomalies among all flagged instances. Accuracy captures the overall proportion of correctly classified frames, encompassing both normal and anomalous ones. Finally, the F1-score provides a balanced assessment by combining Precision and Recall, particularly useful in scenarios with class imbalance.

4. Results and discussion

4.1. Data discription

In this section, the investigated anomaly detection methods are tested on two widely-used benchmark datasets: the UCSD Pedestrian (Ped2) dataset^[67] and the CUHK Avenue dataset^[17], summarized in Table 2.

Dataset	Total Frames	Training Frames	Testing Frames	Resolution (pixels)
UCSD Ped2	4,560	2,550	2,010	360×240
CUHK Avenue	30,652	15,328	15,324	640×360

Table 2. Details of the UCSD Ped2 and CUHK Avenue Datasets.

UCSD Ped2 dataset: The UCSD Ped2 dataset^[67] is a widely-used benchmark for video anomaly detection, particularly in surveillance contexts. It captures video sequences of a pedestrian-only walkway, where anomalies such as bicycles, skateboards, and vehicles disrupt typical pedestrian activity. The dataset is divided into a training set with 16 clips (2,550 frames) containing only normal pedestrian behavior and a testing set with 12 clips (2,010 frames) featuring both normal events and anomalies. This structure allows models to learn typical patterns during training and evaluate their ability to detect anomalies. Figure 2 illustrates examples from the dataset, highlighting various anomalies, including cyclists, vehicles, and skaters, which challenge models to accurately differentiate normal and abnormal events.

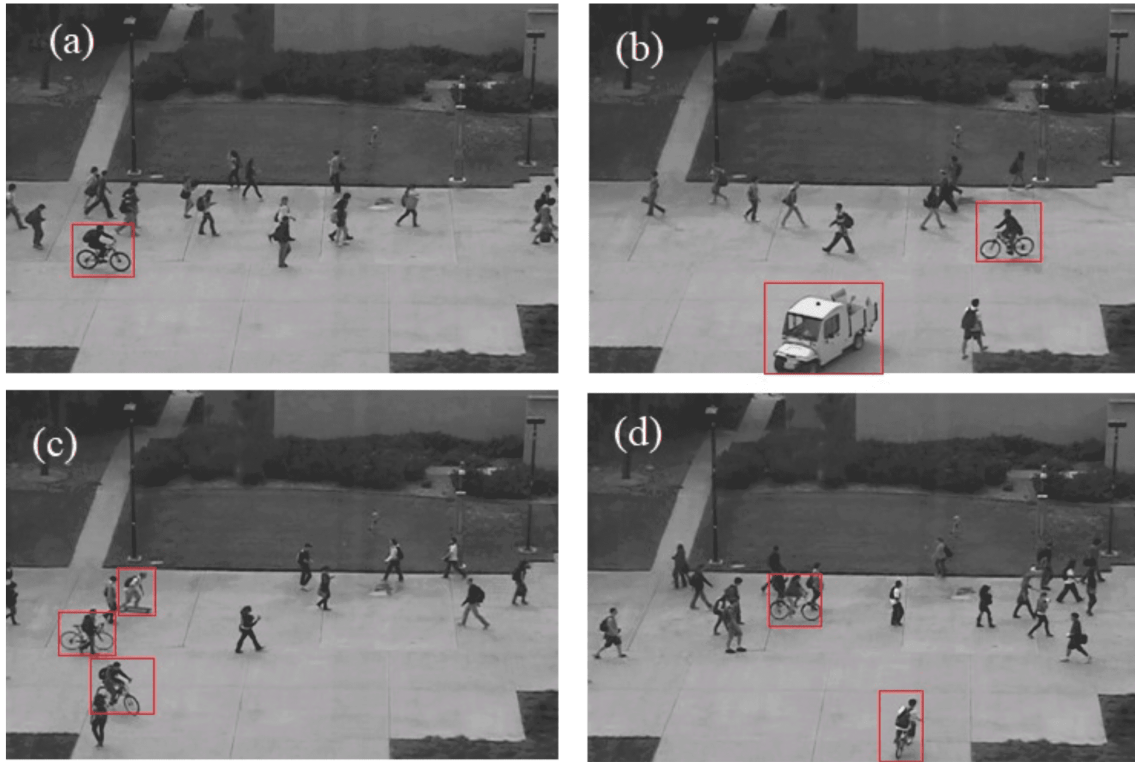


Figure 2. Sample frames from the UCSD Ped2 dataset showing anomalies: (a) Cyclist, (b) Van and cyclist, (c) Cyclist and skater, and (d) two cyclists. Red boxes highlight anomalous regions.

The UCSD Ped2 dataset is recorded at a resolution of 240×360 pixels with a frame rate of 10 frames per second (fps). It includes frame- and pixel-level annotations in the test set, marking anomalies, anomalies timing and location, and facilitating detailed performance evaluation. Frame-level annotations assess a model's ability to detect anomalies within a frame, while pixel-level annotations enable precise localization of anomalous events. Its controlled environment and clear distinction between normal and anomalous events make it a widely accepted standard for evaluating video anomaly detection models. This dataset is essential for benchmarking new methods, assessing model accuracy in anomaly detection, and comparing performance across established metrics. Table 3 outlines the anomalous frames and total video lengths in the UCSD Ped2 dataset, focusing on videos with both normal and anomalous events for MS-Plot analysis.

Test Video	Anomalous Frame Range	Total Frames
Video 1	61–180	180
Video 2	95–180	180
Video 3	1–146	150
Video 4	31–180	180
Video 5	1–129	150
Video 6	1–159	180
Video 7	46–180	180
Video 8	1–180	180
Video 9	1–120	120
Video 10	1–150	150
Video 11	1–180	180
Video 12	88–180	180

Table 3. Anomalous Frame Ranges and Total Frame Count for Each Test Video in the UCSD Ped2 Dataset.

CUHK Avenue dataset: The CUHK Avenue dataset^[17] is a widely used benchmark for video anomaly detection. Captured by a fixed camera overlooking an avenue at the Chinese University of Hong Kong, it offers consistent scene composition, varied lighting conditions, and interactions among multiple pedestrians. The dataset contains 37 video sequences, recorded at 25 frames per second (fps) with a resolution of 640×360 pixels, comprising approximately 30,652 frames. It includes 16 training videos featuring normal pedestrian activities, such as walking and entering or exiting the scene, and 21 testing videos containing both normal and abnormal events. Abnormal behaviors include running, throwing objects, loitering, walking in unusual directions, and abandoning items. Figure 3 presents examples from the CUHK Avenue dataset, showcasing a range of anomalies, such as running, walking in unconventional directions, and loitering.



Figure 3. Sample frames from the CUHK Avenue dataset showing anomalies: (a) running, (b) walking in unconventional directions, and (c) loitering.

4.2. Detection results using UCSD2 Ped2 dataset

In the first experiment, the potential of the MS-Plot technique as an effective detection framework is demonstrated by employing three distinct reconstruction-based models: a simple autoencoder and the MAMA-based model^[34]. Each model is trained exclusively on anomaly-free data to establish a baseline representation of normal behavior. The autoencoder architecture used in this study features an encoder with three convolutional blocks, each comprising a 3×3 Conv2D layer with ReLU activation followed by a 2×2 MaxPooling layer. The encoder's convolutional layers include 32, 16, and 8 filters. The decoder mirrors this structure, using 3×3 Conv2D layers with ReLU activation and 2×2 upsampling, concluding with a 3×3 Conv2D layer (three filters) and a Sigmoid activation function for output generation. Input images are resized to 64×64 pixels, normalized to [0, 1], and processed in batches of 32. The model is trained for 50 epochs using the Adam optimizer and a Mean Squared Error (MSE) loss function to measure reconstruction error.

The second approach combines the MAMA reconstruction-based model^[34] with the MS-Plot for anomaly detection. The MAMA model operates in two stages: the F2LM generator processes video frames through three parallel streams—raw frames, semantic labels (DeepLabv3), and motion data (FlowNet2)—using a Feature Transform Convolutional (FTC) block to generate high-quality reconstructions for normal events and degraded outputs for anomalies. In the second stage, the Destroyer network enhances these distinctions by suppressing low-quality regions in the F2LM output. To enhance temporal context and motion evolution, the model processes five consecutive frames at each time point. The residuals generated by the two considered reconstruction-based models are analyzed using the MS-Plot, which takes the residual matrix as input and evaluates anomalies through a multidimensional assessment of magnitude and shape outlyingness. By treating the residual matrix as

functional data, the MS-Plot detects subtle and complex anomalies, leveraging its nuanced analysis of deviations and the models' ability to handle diverse normal patterns, thereby improving sensitivity and accuracy in video anomaly detection.

Tables 4 and 5 summarize the detection performance of the investigated models on the UCSD Ped2 testing set, using the MS-Plot to analyze reconstruction errors. Table 4 highlights the results for the AE-MS-Plot method, demonstrating varying performance across the testing videos. The method achieves perfect TPR and AUC scores for certain videos (e.g., Videos 9–11), reflecting accurate detection of anomalies. However, its performance falters in scenarios like Video 8, where cyclists and skaters move in different directions, leading to notably low TPR and Accuracy scores. Similarly, Videos 1 and 5, where a bike appears in crowded areas, show higher false positive rates, indicating difficulty in distinguishing anomalies in complex scenes. These results suggest that while the AE-MS-Plot method can effectively detect clear deviations, it faces challenges in scenarios involving subtle anomalies or overlapping movements, emphasizing the need for models capable of handling diverse anomaly patterns.

Video	TPR	FPR	Accuracy	Precision	F1-score	AUC
1	100.00	46.67	84.44	81.08	89.55	78.09
2	18.60	0.00	61.11	100	31.37	77.53
3	95.89	0.00	96.00	100	97.90	98.18
4	95.33	0.00	96.11	100	97.61	97.90
5	100.00	33.33	95.33	94.85	97.36	84.66
6	98.74	9.52	97.78	98.74	98.74	95.06
7	63.70	0.00	72.78	100	77.83	83.13
8	23.33	0.00	23.33	100	37.84	63.41
9	100	0.00	100	100	100	100
10	100	0.00	100	100	100	100
11	100	0.00	100	100	100	100
12	98.92	0.00	99.44	100	99.46	99.52

Table 4. Detection performance of AE-MS-Plot on UCSD Ped2 test videos.

Dataset	TPR	FPR	Accuracy	Precision	F1-score	AUC
1	97.50	1.79	97.73	99.15	98.32	98.08
2	97.67	1.11	98.30	98.82	98.25	98.43
3	98.59	0.00	98.63	100.00	99.29	99.38
4	98.00	0.00	98.30	100.00	98.99	99.24
5	97.60	4.76	97.26	99.19	98.39	96.60
6	97.42	4.76	97.16	99.34	98.37	96.53
7	97.71	0.00	98.30	100.00	98.84	98.97
8	98.86	0.00	98.86	100.00	99.43	99.66
9	97.41	0.00	97.41	100.00	98.69	98.87
10	98.63	0.00	98.63	100.00	99.31	99.47
11	98.30	0.00	98.30	100.00	99.14	99.32
12	95.70	1.20	97.16	98.89	97.27	97.40

Table 5. Detection performance of MAMA-MS-Plot on UCSD Ped2 test videos.

The results in Table 5 demonstrate the superior performance of the MAMA-MS-Plot method in detecting different types of anomaly in crowded videos within the UCSD Ped2 dataset. The method consistently achieves high TPR, Precision, F1-scores, and AUC values across most videos, demonstrating its capability to identify both clear and subtle anomalies effectively. Compared to the AE-MS-Plot method, the MAMA-MS-Plot provides enhanced accuracy and maintains low false positive rates, emphasizing its effectiveness in handling diverse testing conditions and crowded environments. This indicates that the sophisticated reconstruction mechanisms in these models generate residuals sensitive enough for the MS-Plot to detect subtle deviations that a simpler autoencoder might miss. Overall, the MS-Plot framework enhances anomaly detection by analyzing magnitude and shape deviations in residuals over time, capturing subtle anomalies, and reducing false positives in complex scenarios.

4.3. Comparison of MS-Plot and univariate functional anomaly detection methods

The second experiment compares the MS-Plot-based approach with established univariate functional anomaly detection methods, including Functional Boxplot (FBPlot)^[58], Total Variation Depth with Modified Shape Similarity (TVDMSS)^[68], Extremal Depth (ED)^[59], and Outliergram (OG)^[60]. To apply these methods, residual matrices from the autoencoder models were reshaped into one-dimensional vectors. A brief overview of each method is provided for context before comparing their performance with the MS-Plot.

- The FBPlot, introduced by Sun and Genton (2011)^[58], extends the traditional boxplot for functional data by using depth measures, such as modified band depth (MBD), to detect outliers. It centers on the median curve and constructs an envelope representing the central 50% of data, identifying outliers as curves that deviate beyond 1.5 times the interquartile range.
- The TVDMSS method^[68], proposed by Huang and Sun (2019), identifies functional outliers using total variation depth (TVD) for magnitude deviations and modified shape similarity (MSS) for shape anomalies. Thresholds on TVD and MSS scores allow detection of outliers in magnitude, shape, or both.
- The ED^[59], proposed by Narisetty and Nair (2016), ranks functional data based on "extremeness" to detect both boundary and central outliers. Curves with ED values below a threshold are flagged as outliers.
- The OG^[60], introduced by Arribas-Gil and Romo (2014), identifies shape outliers in functional data using Modified Band Depth (MBD) for centrality and Half-Region Depth (HRD) for spread. Outliers are visualized as points deviating from the main cluster in an MBD-HRD plot, emphasizing unusual shapes.

Table 6 presents the comparative results of the proposed MS-Plot-based approaches (AE-MS-Plot and MAMA-MS-Plot) against traditional univariate functional anomaly detection methods (FBPlot, TVDMSS, ED, and OG) on the UCSD Ped2 dataset. The analysis highlights notable differences in performance, with MS-Plot-based methods, particularly MAMA-MS-Plot, demonstrating superior results across all evaluated metrics. These findings underscore the effectiveness of MS-Plot in capturing both magnitude and shape deviations, offering a more comprehensive detection framework compared to univariate techniques. For example, MAMA-MS-Plot achieves an AUC of 98.74%, along with high TPR and precision values (97.90% and 99.68%, respectively), demonstrating strong detection accuracy with a minimal false

positive rate (FPR of 1.45%). In comparison, AE-MS-Plot achieves a moderate AUC of 88.3%, with a TPR of 82.46%, precision of 94.11%, and an FPR of 23.48%. These results highlight the MS-Plot's effectiveness, with MAMA-MS-Plot excelling in accuracy and sensitivity. In contrast, univariate methods applied to autoencoder residuals perform poorly compared to MS-Plot-based approaches (Table 6). For instance, AE-FBPlot and AE-TVDMSS achieve an AUC of 52.41% with a TPR of 27.19%, reflecting low sensitivity to anomalies. Even the best univariate method, AE-ED, attains an AUC of 60.16%, far below the performance of MS-Plot-based techniques. Additionally, ED and OG methods applied to vectorized residuals show inconsistent TPR and FPR values, further highlighting their limitations.

Approach	TPR	FPR	Accuracy	Precision	F1-score	AUC
AE-MS-Plot	82.46	23.48	81.39	94.11	87.90	88.3
MAMA-MS-Plot	97.90	1.45	98.01	99.68	98.78	98.74
AE-FBPlot	27.19	22.38	36.39	84.48	41.14	52.41
AE-TVDMSS	27.19	22.38	36.39	84.48	41.14	52.41
AE-ED	48.40	28.07	52.50	89.09	62.72	60.16
AE-OG	0.00	0.00	17.60	0.00	0.00	50.00
MAMA-FBPlot	3.09	0.00	19.98	100.00	12.76	51.54
MAMA-TVDMSS	3.77	0.00	20.54	100.00	15.14	51.88
MAMA-ED	4.69	1.75	21.00	92.68	8.93	51.47
MAMA-OG	0.00	0.00	17.60	0.00	0.00	50.00
DMAD-FBPlot	0.19	0.00	17.58	100.00	0.37	50.09
DMAD-TVDMSS	0.19	0.00	17.58	100.00	0.37	50.09
DMAD-ED	11.05	17.84	23.45	74.58	19.25	46.61
DMAD-OG	49.81	55.56	48.88	80.94	61.67	47.13

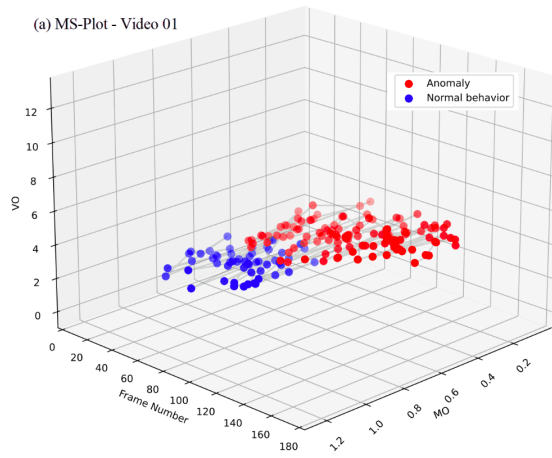
Table 6. Comparison of average performance: MS-Plot vs. univariate methods on UCSD Ped2 test videos.

This comparison reveals the superiority of MS-Plot-based methods over univariate approaches. By analyzing both magnitude and shape outlyingness, MS-Plot effectively detects anomalies in functional residuals data. Unlike univariate methods like FBPlot and ED, which rely on single-dimensional measures, MS-Plot captures nuanced deviations in reconstruction errors, critical for frame-level video anomaly detection. Its ability to handle variability in residuals and assess both the level and direction of outlyingness enables the detection of subtle shifts missed by univariate methods, improving overall detection accuracy.

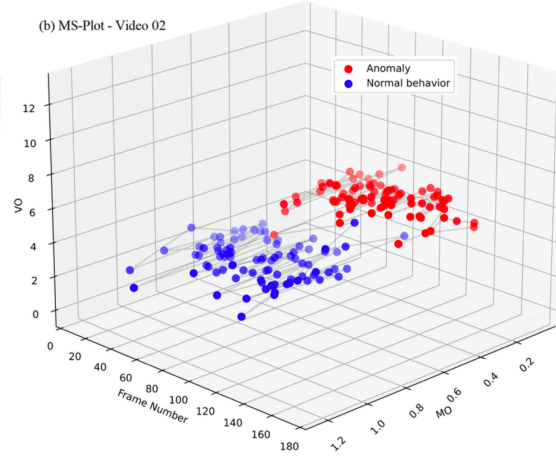
4.4. MS-Plot Visualization Results based on UCSD Ped2 Testing Set

This study utilizes the UCSD Ped2 testing set to demonstrate the MS-Plot's ability to visually represent frame-level anomaly detection, as illustrated in Figure 4. The 3D MS-Plot, with $(|MO|)$ on the x-axis, (VO) on the y-axis, and frame numbers on the z-axis, highlights anomalies (red dots) and normal frames (blue dots) using residuals from the MAMA-based autoencoder. By excluding videos with solely anomalous frames (Table 3), the analysis highlights the MS-Plot's ability to distinguish anomalies from normal behavior in mixed-event scenarios and track their progression over time.

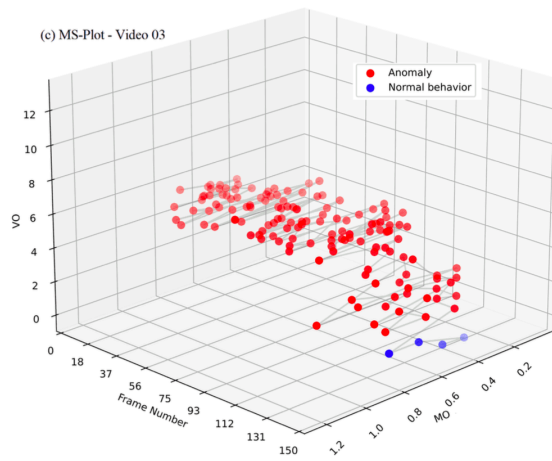
(a) MS-Plot - Video 01



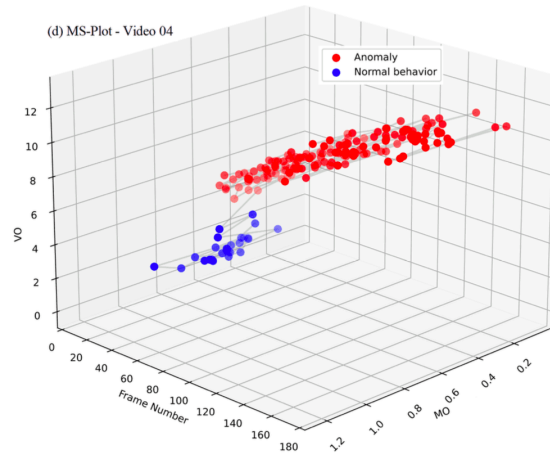
(b) MS-Plot - Video 02



(c) MS-Plot - Video 03



(d) MS-Plot - Video 04



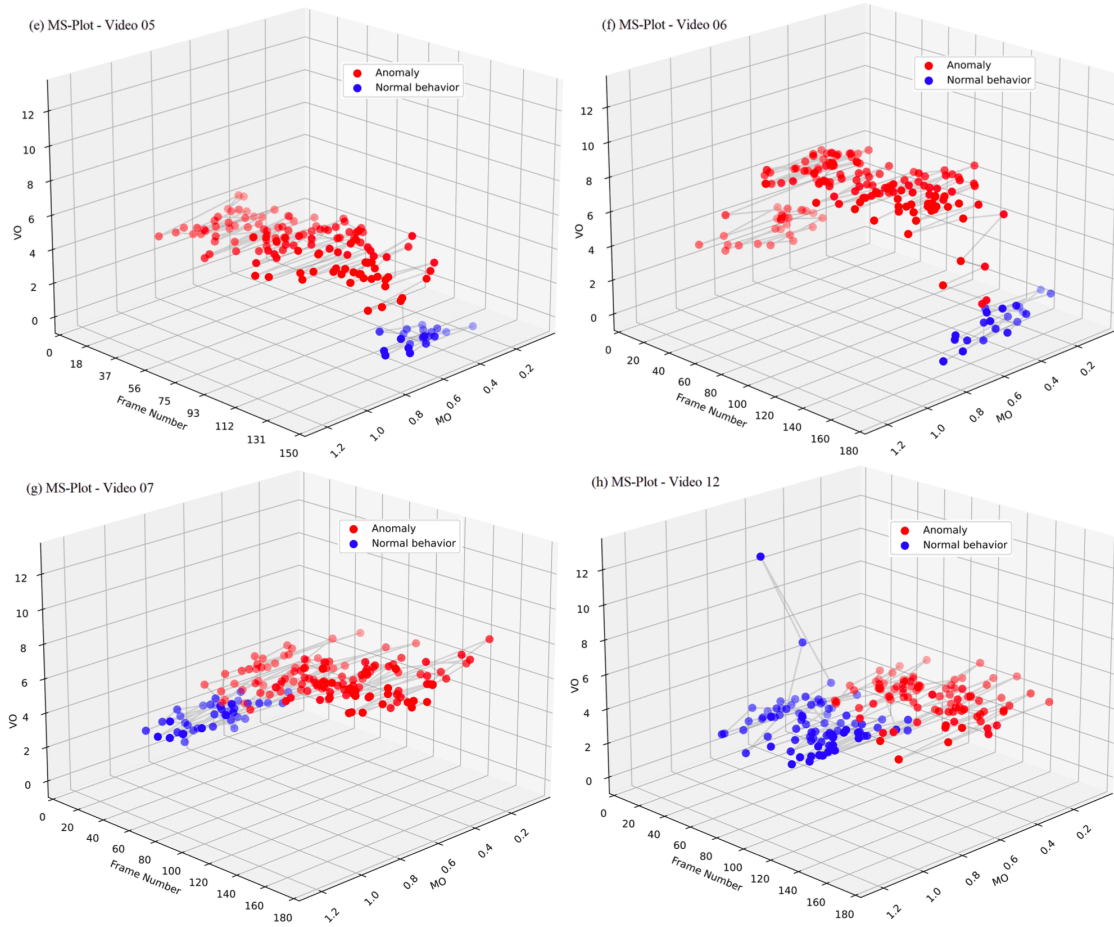


Figure 4. 3D MS-Plot representation for videos in the UCSD Ped2 testing set.

Figure 4 illustrates the MS-Plot's capability to distinguish normal frames (blue points) from anomalies (red points) using magnitude outlyingness ($|MO|$) and shape outlyingness (VO). Normal frames cluster at lower $|MO|$ and VO values, while anomalies show higher values, often forming distinct clusters. Elevated VO values highlight unusual shape variations, such as those caused by cyclists or skaters. The z-axis, representing frame numbers, tracks the temporal progression of anomalies, effectively visualizing their persistence and separation from normal behavior.

4.5. MS-Plot results based on CUHK Avenue data

We extend the evaluation of the two investigated models, the AE-MS-Plot and MAMA-MS-Plot approaches, to the CUHK Avenue dataset to further analyze their performance. The CUHK Avenue dataset is particularly challenging due to its diverse set of anomalous behaviors, such as running, loitering, and

abandoning objects, captured in a dynamic and semi-structured environment. By examining the performance of the MS-Plot-based methods on individual videos, the analysis provides insights into the robustness and sensitivity of these approaches to varying anomaly patterns and complexities within the dataset. Following the same protocol as the UCSD Ped2 dataset, both models are trained exclusively on anomaly-free data to learn typical behavioral patterns and subsequently applied to the test set for anomaly detection. This methodology ensures that the models generalize effectively from normal data without prior exposure to anomalous events. The results from univariate functional detectors on the CUHK dataset were excluded due to their poor performance. The performance of the AE-MS-Plot and MAMA-MS-Plot models on the CUHK Avenue dataset is detailed in Tables 7 and 8. From Tables 7, the AE-MS-Plot performs well on simpler anomalies (e.g., Videos 1, 4, and 12 with AUC > 97%) but struggles with complex or predominantly anomalous frames (e.g., Videos 3, 10, and 19 with AUC of 50%), highlighting its limitations in handling diverse scenarios. The unsatisfactory AE-based results may stem from dataset imbalances, particularly in CUHK Avenue, where the dominance of normal frames amplifies the impact of false negatives. Moreover, frames containing partial abnormal events, such as a small portion of a bike or an object at the start of a video, further complicate detection, revealing the limitations of the AE-based approach.

Video	TPR	FPR	Accuracy	Precision	F1-score	AUC
1	94.90	0.00	98.40	100.00	97.38	97.70
2	40.29	10.01	84.27	34.36	37.09	65.40
3	100.00	100.00	9.35	9.35	17.10	50.00
4	94.62	0.00	99.47	100.00	97.24	97.50
5	92.14	0.00	97.51	100.00	95.91	96.30
6	91.92	0.00	97.27	100.00	95.79	96.20
7	67.62	0.00	94.35	100.00	80.68	84.00
8	100.00	0.00	100.00	100.00	100.00	100.00
9	15.92	0.00	76.11	100.00	27.46	58.20
10	100.00	100.00	19.21	19.21	32.23	50.00
11	100.00	0.00	100.00	100.00	100.00	100.00
12	100.00	0.00	100.00	100.00	100.00	100.00
13	92.59	0.00	98.90	100.00	96.15	96.50
14	72.60	0.00	96.03	100.00	84.13	86.50
15	84.44	20.04	80.36	29.46	43.68	82.40
16	91.92	12.07	88.47	54.17	68.16	90.10
17	79.61	0.00	82.74	100.00	88.65	90.00
18	83.02	0.00	84.54	100.00	90.72	91.70
19	100.00	100.00	53.88	53.88	70.03	50.00
20	44.06	0.00	58.15	100.00	61.17	72.20
21	64.15	10.00	71.23	94.44	76.40	77.30

Table 7. Performance of AE-MS-Plot on CUHK Test Videos.

Video	TPR	FPR	Accuracy	Precision	F1-score	AUC
1	84.92	3.15	93.11	92.51	88.55	91.21
2	84.89	3.18	95.45	77.63	81.10	91.18
3	95.35	3.36	96.52	74.55	83.67	96.32
4	93.55	1.65	97.88	86.14	89.69	96.27
5	82.70	1.02	93.82	97.41	89.46	91.16
6	89.15	1.18	95.55	97.47	93.12	94.30
7	80.95	1.01	95.85	94.44	87.18	90.29
8	90.00	0.00	96.97	100.00	94.74	95.32
9	76.88	0.72	92.92	97.71	86.05	88.40
10	93.79	0.30	98.57	98.69	96.18	97.07
11	81.42	0.00	92.75	100.00	89.76	91.03
12	79.92	0.79	95.43	96.14	87.28	89.89
13	95.06	0.00	99.27	100.00	97.47	97.85
14	90.41	2.09	96.83	88.00	89.19	94.48
15	96.67	2.75	97.19	77.68	86.14	97.28
16	91.92	1.25	97.83	91.92	91.92	95.65
17	77.93	0.00	81.32	100.00	87.60	89.29
18	78.49	0.00	80.41	100.00	87.95	89.57
19	80.30	1.77	88.57	98.15	88.33	89.59
20	78.71	0.00	84.07	100.00	88.09	89.68
21	79.25	0.00	84.93	100.00	88.42	89.94

Table 8. Performance of MAMA-MS-Plot on CUHK Test Videos.

The MAMA-MS-Plot performs well on the CUHK Avenue dataset, achieving high AUC values above 95% in videos like 1, 4, 8, 10, and 13 (Table 8). However, it faces challenges in videos with complex anomalies, such as 9 and 18, where AUC scores drop but remain above 88%. These results indicate the method's overall reliability while highlighting areas for potential improvement in detecting anomalies in highly skewed scenarios.

The overall AUC results show an improvement with the MAMA-MS-Plot (91.43%) over the AE-MS-Plot (81.40%), demonstrating enhanced anomaly detection performance through its advanced residual analysis combined with the MS-Plot framework.

4.6. Comparison with SOTA methods

Table 9 provides an overall comparison of AUC scores for the proposed MS-Plot-based approaches and state-of-the-art (SOTA) methods on the UCSD Ped2 dataset. The MAMA-MS-Plot achieves an impressive AUC of 98.74%, while the AE-MS-Plot records 88.30%, showcasing the overall effectiveness of the MS-Plot framework in accurately identifying video anomalies. Specifically, compared to other detection methods, the MAMA-MS-Plot demonstrated the highest performance, achieving the best AUC score among the evaluated approaches.

Method	AUC (%)
MPPCA ^[10]	77.00
Motion-appearance model ^[69]	90.00
Spatial Temporal CNN ^[33]	86.00
Conv-AE ^[66]	90.00
AMDN ^[70]	90.80
GMFC-VAE ^[71]	92.20
StackRNN ^[72]	92.20
MGFC-AAE ^[44]	91.60
ST-CaAE ^[73]	92.90
MemAE ^[37]	94.10
DAW ^[74]	96.40
MPN ^[75]	96.90
ConvLSTM ^[76]	88.10
spatiotemporal AE (STAE) ^[39]	91.20
Conv2D ^[66]	90.00
TMAE ^[77]	94.10
MAMC ^[78]	96.70
Cascade Reconstruction ^[79]	97.70
MAMA ^[34]	98.20
AE-MS-Plot	88.30
MAMA-MS-Plot	98.74

Table 9. Comparison of AUC (%) with SOTA methods on the UCSD Ped2 dataset.

The MAMA-MS-Plot method demonstrates improved performance over the original MAMA-based approach, achieving an AUC of 98.74% compared to 98.20%. This improvement highlights the benefits of using the MS-Plot framework for anomaly detection, particularly in its ability to enhance the interpretability and precision of detection outcomes. In the original MAMA model, anomaly detection is based on empirically determined thresholds primarily based on abnormal changes in the magnitude of residuals. These thresholds, while functional, lack a systematic foundation, which can lead to inconsistencies when addressing diverse anomaly patterns or capturing more nuanced deviations in behavior. The integration of the MS-Plot framework addresses this limitation by treating the residuals of the MAMA model as multivariate functional data. The MS plot evaluates outlyingness (amplitude deviations) and shape (pattern deviations), providing a multidimensional analysis of residual patterns. This capability is particularly beneficial in video anomaly detection, where anomalies often exhibit subtle or complex deviations that may not be effectively captured by simpler thresholding techniques. Using the MS-Plot framework, the MAMA-MS-Plot approach offers a statistically principled method to analyze residuals, enhancing the consistency and reliability of anomaly detection. The framework's capacity to capture both amplitude and shape deviations allows for a more comprehensive understanding of anomalies, ultimately improving detection performance in complex and dynamic video environments.

We compare the proposed MS-Plot-based methods with several state-of-the-art anomaly detection techniques in the CUHK Avenue dataset, as detailed in Table 10. The MAMA-MS-Plot approach achieves an AUC of 91.43%, demonstrating its strong competitiveness among SOTA methods. In particular, the previous highest reported AUC was 91.30%, achieved by SD-MAE^[56]. The MAMA-MS-Plot exceeds this by a small but meaningful margin of 0.13%, setting a new benchmark for anomaly detection performance in this data set.

Method	AUC (%)
Unmasking ^[80]	80.60
StackRNN ^[72]	81.70
MemAE ^[37]	83.3
FastAno ^[55]	85.30
STemGAN ^[53]	86.00
EVAL ^[81]	86.02
MESDnet ^[82]	86.3
Any-Shot Sequential ^[83]	86.40
AMMC-Net ^[84]	86.60
ASTNet ^[38]	86.70
Context Pre ^[85]	87.10
Learning not to reconstruct ^[86]	87.10
Siamese Net ^[87]	87.20
Object-centric AE ^[88]	87.40
AKD-VAD ^[89]	88.30
AnomalyRuler ^[52]	89.70
Two-stream ^[54]	90.80
MAMA ^[34]	91.20
SD-MAE ^[56]	91.30
AE-MS-Plot	81.40
MAMA-MS-Plot	91.43

Table 10. Comparison of AUC (%) with SOTA methods on the CUHK Avenue dataset.

This improvement, while incremental, underscores the value of integrating the MS-Plot framework, which provides a systematic and statistically principled approach to anomaly detection. By treating residuals as multivariate functional data, the MAMA-MS-Plot captures both magnitude and shape deviations, enabling more precise differentiation between normal and anomalous frames. In contrast, traditional methods often rely on empirically determined thresholds or univariate analyses, which may miss subtle or multidimensional patterns indicative of anomalies. The MS-Plot's ability to enhance the performance of already strong models like MAMA highlights its potential as a generalizable and effective addition to video anomaly detection pipelines. This advantage is particularly significant given the challenging nature of the CUHK Avenue dataset, which features diverse and complex anomalous behaviors.

5. Conclusion

This study demonstrates the efficacy of the Magnitude-Shape (MS) Plot framework for video anomaly detection, combining statistical functional data analysis with reconstruction-based models. By treating reconstruction errors as multivariate functional data, the MS-Plot captures both magnitude and shape deviations, enabling accurate identification of anomalies in complex, crowded video scenes. The integration of autoencoders with the MS-Plot enhances anomaly detection by leveraging their capacity to model normal behavior and identify deviations with statistical rigor. Experimental results on benchmark datasets, including UCSD Ped2 and CUHK Avenue, highlight the advantages of the proposed framework over traditional univariate functional detectors and several state-of-the-art methods. Specifically, the MAMA-MS-Plot approach achieves consistently high AUC scores, showcasing its capability to generalize across diverse scenarios and detect both subtle and pronounced anomalies effectively.

The findings emphasize the potential of MS-Plot-based frameworks to address the challenges of video anomaly detection, such as the variability of anomalies and limited labeled data. An important direction for future work is to focus on developing unsupervised and robust statistical methods that do not rely on anomaly-free training data. Such methods would further enhance the practicality of video anomaly detection by eliminating the dependence on curated normal datasets and addressing the inherent challenges of real-world applications, such as data scarcity and variability in anomaly types.

References

1. [△]Santhosh KK, Dogra DP, Roy PP (2020). "Anomaly detection in road traffic using visual surveillance: A survey". *ACM Computing Surveys (CSUR)*. **53** (6): 1–26.
2. [△]Tripathi RK, Jalal AS, Agrawal SC (2018). "Suspicious human activity recognition: a review". *Artificial Intelligence Review*. **50**: 283–339.
3. [△]Chandola V, Banerjee A, Kumar V (2009). "Anomaly detection: A survey". *ACM computing surveys (CSUR)*. **41** (3): 1–58.
4. [△][▽]Nayak R, Pati UC, Das SK (2021). "A comprehensive review on deep learning-based methods for video anomaly detection". *Image and Vision Computing*. **106**: 104078.
5. [△]Mu H, Sun R, Yuan G, Wang Y (2021). "Abnormal human behavior detection in videos: A review". *Information Technology and Control*. **50** (3): 522–545.
6. [△][▽][◇][◇]Duong HT, Le VT, Hoang VT (2023). "Deep learning-based anomaly detection in video surveillance: A survey". *Sensors*. **23** (11): 5024.
7. [△]Zhu Y, Nayak NM, Roy-Chowdhury AK (2012). "Context-aware activity recognition and anomaly detection in video". *IEEE Journal of Selected Topics in Signal Processing*. **7** (1): 91–101.
8. [△][▽]Pawar K, Attar V (2019). "Deep learning approaches for video-based anomalous activity detection". *World Wide Web*. **22** (2): 571–601.
9. [△]Morris BT, Trivedi MM (2008). "A survey of vision-based trajectory learning and analysis for surveillance". *IEEE transactions on circuits and systems for video technology*. **18** (8): 1114–1127.
10. [△][▽]Kim J, Grauman K. "Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates." In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2009. p. 2921–2928.
11. [△]Mehran R, Oyama A, Shah M. "Abnormal crowd behavior detection using social force model." In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE; 2009. p. 935–942.
12. [△]Basharat A, Gritai A, Shah M. "Learning object motion patterns for anomaly detection and improved object detection." In: *2008 IEEE conference on computer vision and pattern recognition*. IEEE; 2008. p. 1–8.
13. [△]Piciarelli C, Micheloni C, Foresti GL. Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for Video Technology*. **18**(11):1544–1554, 2008.
14. [△]Cof{c{s}}ar S, Donatiello G, Bogorny V, Garate C, Alvares LO, Br{\e}mond F (2016). "Toward abnormal trajectory and event detection in video surveillance". *IEEE Transactions on Circuits and Systems for Video Technology*.

ology. 27 (3): 683--695.

15. [△]Ma X, Wang H, Xue B, Zhou M, Ji B, Li Y (2014). "Depth-based human fall detection via shape features and improved extreme learning machine". *IEEE journal of biomedical and health informatics*. 18 (6): 1915–1922.
16. [△]Li W, Mahadevan V, Vasconcelos N (2013). "Anomaly detection and localization in crowded scenes". *IEEE transactions on pattern analysis and machine intelligence*. 36 (1): 18–32.
17. [△][♢]Lu C, Shi J, Jia J (2013). "Abnormal event detection at 150 fps in matlab". *Proceedings of the IEEE international conference on computer vision*. 2720--2727.
18. [△]Dalal N, Triggs B. "Histograms of oriented gradients for human detection." In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee; 2005. p. 886–893.
19. [△]Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009. p. 1446–1453.
20. [△]Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance. In: *Computer Vision--ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part II 9*. Springer; 2006. p. 428--441.
21. [△]Roy PK, Om H (2018). "Suspicious and violent activity detection of humans using HOG features and SVM classifier in surveillance videos". *Advances in Soft Computing and Machine Learning in Image Processing*. Springer; 2018. p. 277–294.
22. [△]Ahad MAR, Tan JK, Kim H, Ishikawa S (2018). "Activity representation by SURF-based templates". *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 6 (5): 573–583.
23. [△]Cheng KW, Chen YT, Fang WH (2015). "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2909–2917.
24. [△][♢]Sabokrou M, Fayyaz M, Fathy M, Moayed Z, Klette R (2018). "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes". *Computer Vision and Image Understanding*. 172: 88–97.
25. [△]Mansour RF, Escorcia-Gutierrez J, Gamarra M, Villanueva JA, Leal N (2021). "Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model". *Image and Vision Computing*. 112: 104229.
26. [△]Sultani W, Chen C, Shah M. "Real-world anomaly detection in surveillance videos". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. p. 6479–6488.

27. ^aNawaratne R, Alahakoon D, De Silva D, Yu X (2019). "Spatiotemporal anomaly detection using deep learning for real-time video surveillance". *IEEE Transactions on Industrial Informatics*. **16** (1): 393–402.
28. ^aUllah W, Ullah A, Hussain T, Khan ZA, Baik SW (2021). "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos". *Sensors*. **21** (8): 2811.
29. ^aLuo W, Liu W, Lian D, Gao S (2021). "Future frame prediction network for video anomaly detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **44** (11): 7505–7520.
30. ^aMicorek J, Possegger H, Narnhofer D, Bischof H, Kozinski M (2024). "MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 18868–18877.
31. ^aChen Z, Duan J, Kang L, Qiu G (2022). "Supervised anomaly detection via conditional generative adversarial network and ensemble active learning". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **45** (6): 7781–7798.
32. ^aWu P, Pan C, Yan Y, Pang G, Wang P, Zhang Y (2024). "Deep Learning for Video Anomaly Detection: A Review". *arXiv preprint arXiv:2409.05383*.
33. ^{a, b}Zhou S, Shen W, Zeng D, Fang M, Wei Y, Zhang Z (2016). "Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes". *Signal Processing: Image Communication*. **47**: 358–368.
34. ^{a, b, c, d, e, f, g, h}Hong S, Ahn S, Jo Y, Park S (2024). "Making anomalies more anomalous: Video anomaly detection using a novel generator and destroyer". *IEEE Access*. Published by IEEE.
35. ^aDilek E, Dener M (2024). "Enhancement of Video Anomaly Detection Performance Using Transfer Learning and Fine-Tuning". *IEEE Access*. **2024**.
36. ^aChong YS, Tay YH. "Abnormal event detection in videos using spatiotemporal autoencoder." In: *Advances in Neural Networks-ISBNN 2017: 14th International Symposium, ISNN 2017, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, June 21–26, 2017, Proceedings, Part II 14*. Springer; 2017. p. 189–196.
37. ^{a, b, c, d}Gong D, Liu L, Le V, Saha B, Mansour MR, Venkatesh S, van den Hengel A. "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection." In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019. p. 1705–1714.
38. ^{a, b, c, d}Le VT, Kim YG (2023). "Attention-based residual autoencoder for video anomaly detection". *Applied Intelligence*. **53** (3): 3240–3254.
39. ^{a, b, c}Zhao Y, Deng B, Shen C, Liu Y, Lu H, Hua XS. "Spatio-temporal autoencoder for video anomaly detection." In: *Proceedings of the 25th ACM international conference on Multimedia*. 2017. p. 1933–1941.

40. [△]Deepak K, Chandrakala S, Mohan CK (2021). "Residual spatiotemporal autoencoder for unsupervised video anomaly detection". *Signal, Image and Video Processing*. **15** (1): 215–222.
41. [△]Luo W, Liu W, Lian D, Tang J, Duan L, Peng X, Gao S (2019). "Video anomaly detection with sparse coding in spired deep neural networks". *IEEE transactions on pattern analysis and machine intelligence*. **43** (3): 1070–1084.
42. [△]Kumar A, Khari M. Efficient video anomaly detection using residual variational autoencoder. In: *2023 International Conference on Communication System, Computing and IT Applications (CSCITA)*. IEEE; 2023. p. 50–55.
43. [△]Aslam N, Kolekar MH (2024). "DeMAAE: deep multiplicative attention-based autoencoder for identification of peculiarities in video sequences". *The Visual Computer*. **40** (3): 1729–1743.
44. [△]^aLi N, Chang F (2019). "Video anomaly detection and localization via multivariate gaussian fully convolution adversarial autoencoder". *Neurocomputing*. **369**: 92–105.
45. [△]Hu X, Lian J, Zhang D, Gao X, Jiang L, Chen W (2022). "Video anomaly detection based on 3D convolutional auto-encoder". *Signal, Image and Video Processing*. **16** (7): 1885–1893.
46. [△]Duman E, Erdem OA (2019). "Anomaly detection in videos using optical flow and convolutional autoencoder". *IEEE Access*. **7**: 183914–183923.
47. [△]Mishra S, Jabin S (2024). "Anomaly detection in surveillance videos using deep autoencoder". *International Journal of Information Technology*. **16** (2): 1111–1122.
48. ^a^b^c^d^e^f^gDai W, Genton MG (2018). "Multivariate functional data visualization and outlier detection". *Journal of Computational and Graphical Statistics*. **27** (4): 923–934.
49. [△]Rousseeuw PJ, Raymaekers J, Hubert M (2018). "A measure of directional outlyingness with applications to image data and video". *Journal of Computational and Graphical Statistics*. **27** (2): 345–359.
50. [△]Alemán-Gómez Y, Arribas-Gil A, Desco M, Elías A, Romo J (2022). "Depthgram: Visualizing outliers in high-dimensional functional data with application to fMRI data exploration". *Statistics in Medicine*. **41** (11): 2005–2024.
51. [△]Naji Y, Setkov A, Loesch A, Gouiffès M, Audigier R. Spatio-temporal predictive tasks for abnormal event detection in videos. In: *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE; 2022. p. 1–8.
52. ^a^bYang Y, Lee K, Dariush B, Cao Y, Lo S-Y. Follow the rules: reasoning for video anomaly detection with large language models. In: *European Conference on Computer Vision*. Springer; 2025. p. 304–322.

53. ^a Singh R, Saini K, Sethi A, Tiwari A, Saurav S, Singh S. Stemgan: spatio-temporal generative adversarial network for video anomaly detection. *Applied Intelligence*. 53(23):28133–28152, 2023.
54. ^a Cao C, Lu Y, Zhang Y. "Context recovery and knowledge retrieval: A novel two-stream framework for video anomaly detection". *IEEE Transactions on Image Processing*. 2024.
55. ^a Park C, Cho M, Lee M, Lee S (2022). "FastAno: Fast anomaly detection via spatio-temporal patch transformation". *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 2249–2259.
56. ^a Ristea NC, Croitoru FA, Ionescu RT, Popescu M, Khan FS, Shah M, et al. Self-distilled masked auto-encoders are efficient video anomaly detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024. p. 15984–15995.
57. ^a Liu W, Chang H, Ma B, Shan S, Chen X (2023). "Diversity-measurable anomaly detection." In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 12147–12156.
58. ^a Sun Y, Genton MG (2011). "Functional boxplots". *Journal of Computational and Graphical Statistics*. 20(2): 316–334.
59. ^a Narisetty NN, Nair VN (2016). "Extremal depth for functional data and applications". *Journal of the American Statistical Association*. 111 (516): 1705–1714.
60. ^a Arribas-Gil A, Romo J (2014). "Shape outlier detection and visualization for functional data: the outlier gram". *Biostatistics*. 15 (4): 603–619.
61. ^Δ Rousseeuw PJ, Van Driessen K (1999). "A fast algorithm for the minimum covariance determinant estimator". *Technometrics*. 41 (3): 212–223.
62. ^Δ Hardin J, Rocke DM (2005). "The distribution of robust distances". *Journal of Computational and Graphical Statistics*. 14 (4): 928–946.
63. ^Δ Bengio Y, Courville A, Vincent P (2013). "Representation learning: A review and new perspectives". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 35 (8): 1798–1828.
64. ^Δ Hinton GE, Salakhutdinov RR (2006). "Reducing the dimensionality of data with neural networks". *Science*. 313 (5786): 504–507.
65. ^Δ Sabokrou M, Fathy M, Hoseini M (2016). "Video anomaly detection and localisation based on the sparsity and reconstruction error of auto-encoder". *Electronics Letters*. 52 (13): 1122–1124.
66. ^a Hasan M, Choi J, Neumann J, Roy-Chowdhury AK, Davis LS (2016). "Learning temporal regularity in video sequences". *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 733–742.

67. ^a Mahadevan V, Li W, Bhalodia V, Vasconcelos N. "Anomaly detection in crowded scenes." In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE; 2010. p. 1975–1981.
68. ^a Huang H, Sun Y (2019). "A decomposition of total variation depth for understanding functional outliers". *Technometrics*. Taylor & Francis.
69. ^Δ Zhang Y, Lu H, Zhang L, Ruan X (2016). "Combining motion and appearance cues for anomaly detection". *Pattern Recognition*. **51**: 443–452.
70. ^Δ Xu D, Ricci E, Yan Y, Song J, Sebe N (2015). "Learning deep representations of appearance and motion for a nomalous event detection". *arXiv preprint arXiv:1510.01553*.
71. ^Δ Fan Y, Wen G, Li D, Qiu S, Levine MD, Xiao F (2020). "Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder". *Computer Vision and Image Understanding*. **195**: 102920.
72. ^a Luo W, Liu W, Gao S (2017). "A revisit of sparse coding based anomaly detection in stacked rnn framework". *Proceedings of the IEEE international conference on computer vision*. pp. 341–349.
73. ^Δ Li N, Chang F, Liu C (2020). "Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes". *IEEE Transactions on Multimedia*. **23**: 203–215.
74. ^Δ Wang S, Zeng Y, Liu Q, Zhu C, Zhu E, Yin J. Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection. In: *Proceedings of the 26th ACM international conference on Multimedia*. 2018. p. 636–644.
75. ^Δ Lv H, Chen C, Cui Z, Xu C, Li Y, Yang J. "Learning normal dynamics in videos with meta prototype network." In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021. p. 15425–15434.
76. ^Δ Luo W, Liu W, Gao S. "Remembering history with convolutional lstm for anomaly detection." In: *2017 IEEE International conference on multimedia and expo (ICME)*. IEEE; 2017. p. 439–444.
77. ^Δ Hu J, Yu G, Wang S, Zhu E, Cai Z, Zhu X. Detecting anomalous events from unlabeled videos via temporal masked auto-encoding. In: *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE; 2022. p. 1–6.
78. ^Δ Ning Z, Wang Z, Liu Y, Liu J, Song L (2024). "Memory-enhanced appearance-motion consistency framework for video anomaly detection". *Computer Communications*. **216**: 159–167.
79. ^Δ Zhong Y, Chen X, Jiang J, Ren F (2022). "A cascade reconstruction model with generalization ability evaluation for anomaly detection in videos". *Pattern Recognition*. **122**: 108336.

80. [△]Tudor Ionescu R, Smeureanu S, Alexe B, Popescu M (2017). "Unmasking the abnormal events in video". In: *Proceedings of the IEEE international conference on computer vision*. pp. 2895–2903.
81. [△]Singh A, Jones MJ, Learned-Miller EG. "Eval: Explainable video anomaly localization." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023. p. 18717-18726.
82. [△]Fang Z, Zhou JT, Xiao Y, Li Y, Yang F (2020). "Multi-encoder towards effective anomaly detection in videos". *IEEE Transactions on Multimedia*. 23: 4106–4116.
83. [△]Doshi K, Yilmaz Y (2020). "Any-shot sequential anomaly detection in surveillance videos". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. pp. 934–935.
84. [△]Cai R, Zhang H, Liu W, Gao S, Hao Z (2021). "Appearance-motion memory consistency network for video anomaly detection". *Proceedings of the AAAI conference on artificial intelligence*. 35 (2): 938–946.
85. [△]Li D, Nie X, Li X, Zhang Y, Yin Y (2022). "Context-related video anomaly detection via generative adversarial network". *Pattern Recognition Letters*. 156: 183–189.
86. [△]Astrid M, Zaheer MZ, Lee JY, Lee SI (2021). "Learning not to reconstruct anomalies". *arXiv preprint arXiv:2110.09742*. Available from: <https://arxiv.org/abs/2110.09742>.
87. [△]Ramachandra B, Jones M, Vatsavai R (2020). "Learning a distance function with a Siamese network to localize anomalies in videos". *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020: 2598–2607.
88. [△]Ionescu RT, Khan FS, Georgescu MI, Shao L. Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019:7842-7851.
89. [△]Croitoru FA, Ristea NC, Dascalescu D, Ionescu RT, Khan FS, Shah M (2022). "Lightning fast video anomaly detection via adversarial knowledge distillation". *arXiv preprint arXiv:2211.15597*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.