**Journal of Organizational and End User Computing**
Volume 36 • Issue 1 • January-December 2024

# A Deep Learning-Based Animation Video Image Data Anomaly Detection and Recognition Algorithm

Cheng Li
*Luxun Academy of Fine Arts, China*

Qiguang Qian
*Luxun Academy of Fine Arts, China*

## ABSTRACT

Anomaly detection plays a crucial role in the field of machine learning, as it involves constructing detection models capable of identifying abnormal samples that deviate from expected patterns, using unlabeled or normal samples. In recent years, there has been a growing interest in integrating anomaly detection into image processing to tackle challenges related to target detection, particularly when dealing with limited sample availability. This paper introduces a novel fully connected network model enhanced with a memory augmentation mechanism. By harnessing the comprehensive feature capabilities of the fully connected network, this model effectively complements the representation capabilities of convolutional neural networks. Additionally, it incorporates a memory module to retain knowledge of normal patterns, thereby enhancing the performance of existing models for video anomaly detection. Furthermore, we present a video anomaly detection system designed to identify abnormal image data within surveillance videos, leveraging the innovative network architecture described above.

## KEYWORDS

Anomaly Detection, Autoencoder, Deep Learning, Fully Connected Network, Image Anomaly Detection, Neural Network

## INTRODUCTION

Anomaly detection (Du et al., 2022) is an essential area of research in machine learning. It is a method of constructing a model using unlabeled or normal samples to detect anomalous samples that differ from the desired pattern. Anomaly detection has a wide range of applications in various fields, such as defect detection, medical image analysis, hyperspectral image processing, abnormal behavior detection, and image and video processing. Table 1 shows that many methods have been applied to image anomaly detection in various fields.

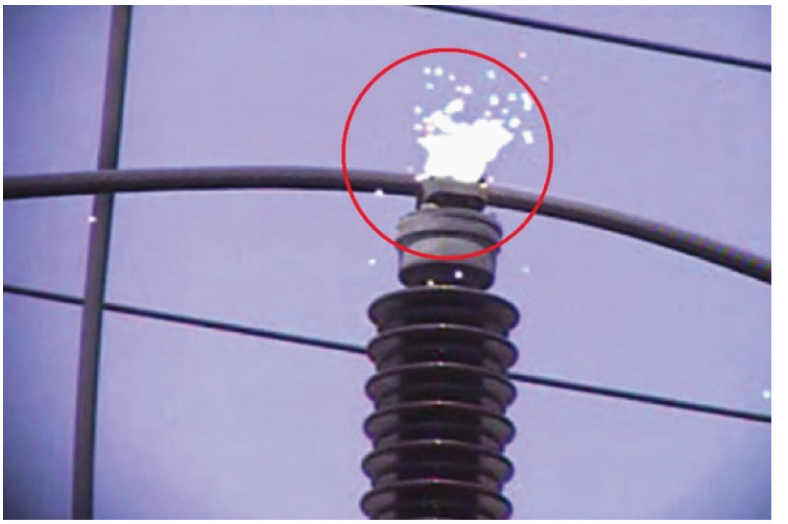
Early anomaly detection algorithms were primarily applied in the field of data mining (Pasini, 2021). In recent years, with the development of computer vision, deep learning, and related technologies (Gao et al., 2023; Kaur et al., 2022; Luo & Pundlik, 2022; You et al., 2023), many studies have introduced anomaly detection into image processing to solve the problem of target detection in the case of sample scarcity. Figure 1 shows the application of image anomaly detection in power equipment protection.

DOI: 10.4018/JOEUC.345929

**Table 1. Applications of Image Anomaly Detection**

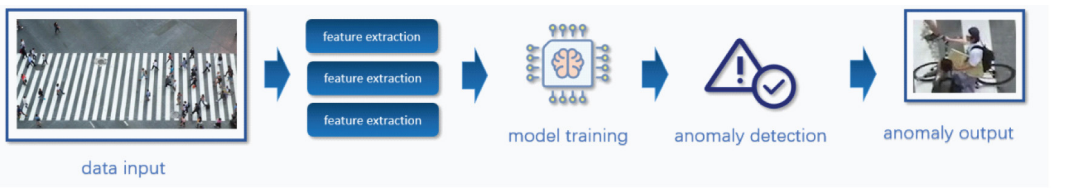
| Areas of application | Specific application |
|---|---|
| Defect detection (Wu et al., 2023) | Inspection of surface defects on a variety of products including textured surfaces, such as cloth (Zhao et al., 2010), glass (Hu et al., 2022), steel (Zheng et al., 2022), cement (Meng & Zhu, 2023), and objects such as printed circuit boards (Alghassab, 2022) and wine bottles (Napoletano et al., 2021). |
| Medical image analysis (Ramaraj et al., 2020; Tschuchnig & Gadermayr, 2022) | Detection of possible lesion areas in medical images, such as MRI images (Ramaraj et al., 2020), iris images (Modwel et al., 2021), and fundus retinal images (Hussain & Holambe, 2015). |
| Hyperspectral image processing (Yan et al., 2021) | Marine vessel detection (Yu et al., 2022), ground anomaly area detection (Ning et al., 2023), etc. |
| Power image analysis | Defective power equipment classification (Wang et al., 2022), etc. |

**Figure 1. Purpose of Image Anomaly Detection Algorithm**



The general computational framework of the image anomaly detection algorithm is shown in Figure 2. After the original image is input into the system, the image features are first extracted. Then, the image features are classified, and an image anomaly detection model is trained using machine learning algorithms and the input pre-labeled image feature data. It can then be used to detect and identify image data anomalies in a new image dataset (Ye & Zhao, 2023).

**Figure 2. General Flow of Image Anomaly Detection Algorithm**



In this paper, we first introduce the background and significance of video anomaly detection and then propose a fully connected network (FCN) model based on the memory enhancement method (Mem-FCN). In designing this model, we made three contributions:

1. We designed a memory-enhanced network model for video anomaly detection. This model utilizes the comprehensive feature capability of the FCN, effectively reduces the representation capability of the convolutional neural network (CNN), and incorporates a memory module to memorize all normal patterns.
2. Additionally, we designed a video anomaly detection system to detect abnormal image data of surveillance video and realized the video anomaly detection function by using the novel network.
3. Finally, we designed several comparison experiments and an ablation experiment.

We analyzed the Mem-FCN model and compared it with other baseline models to verify the superiority of the algorithm proposed in this paper. We also compared the Mem-FCN model with the FCN model to ascertain the effectiveness of the algorithm improvement proposed in this paper. The structure of this paper is shown in Figure 3.

## RELATED WORK

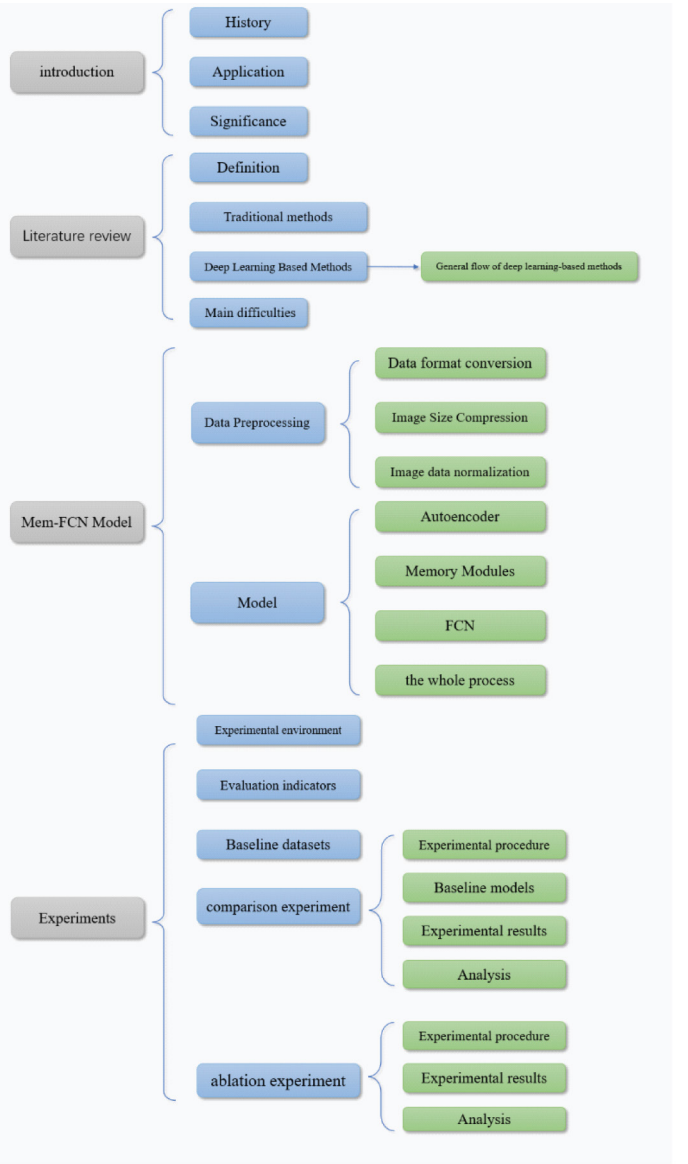
Existing video anomaly detection techniques are broadly classified into two types: the traditional method (Ning et al., 2023), based on extracting manual features to construct models for behavioral patterns, and the detection method, using deep neural networks (Pang et al., 2021) for self-learning. The model built using the traditional method is unstable and needs to re-extract features for different scenes. For newly added samples, it needs to be re-modeled after fusing the original data with new data, which consumes significant time and computational power. In contrast, models based on deep learning can autonomously learn and extract high-level features of the video for the training process. The performance of the deep neural network will improve as the training data and the number of iterations continue to increase. Accordingly, the model's ability to represent normal behavioral patterns grows exponentially in a short period.

### Definition of Image Anomalies

An anomaly (Deecke et al., 2019), also known as an outlier (Murat, 2023), is a common concept in the field of data mining, and quite a few works have tried to define anomalous data (Mrdovic, 2006; Talagala et al., 2021; Yepmo et al., 2022). Generally, typical anomaly samples are categorized into point, contextual, and cluster anomalies (Wang et al., 2024). Point anomalies are usually characterized by observations that deviate significantly from the normal range of data distribution. Contextual anomalies are observations that are within the normal data distribution but, when analyzed with the surrounding data, show significant anomalies. Cluster anomalies, also known as pattern anomalies, result from the aggregation of a series of observations and differ from the normal data. In this type of anomaly, any of the points may not be an anomaly when viewed individually, but it is an anomaly when a series of points appear together. The three anomalies are shown in Figure 4.

The pixel value at each pixel point in the image data corresponds to an observation. The diversity of pixel values in an image makes it difficult to analyze only the pixel value of one point to determine whether it is an anomaly. Therefore, in most image anomaly detection tasks, we must jointly analyze the image background and the surrounding pixel information to classify the anomalies. Most of the detected anomalies belong to contextual or pattern anomalies. Figure 5 shows the classification of commonly used image anomaly detection techniques.

**Figure 3. Structure of the Paper**




## Literature Review

In recent years, computer vision scholars have paid significant attention to video image anomaly detection, proving pivotal in domains such as smart homes and security. With the evolution of deep learning, methods utilizing CNNs have made considerable strides across diverse applications within computer vision. This section provides a comprehensive overview of recent advancements in video anomaly detection, encompassing traditional methods reliant on manual feature extraction and deep learning-based approaches. Specifically, we delve into related works based on supervised learning, semi-supervised learning, and unsupervised learning methods, spanning various levels of complexity and performance.

**Figure 4. Examples of Three Kinds of Image Data Anomalies**

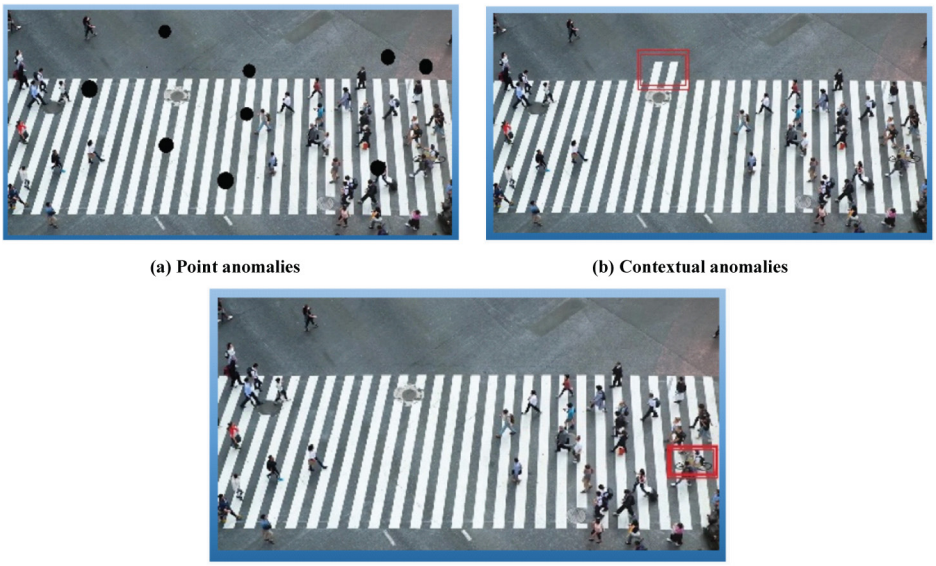
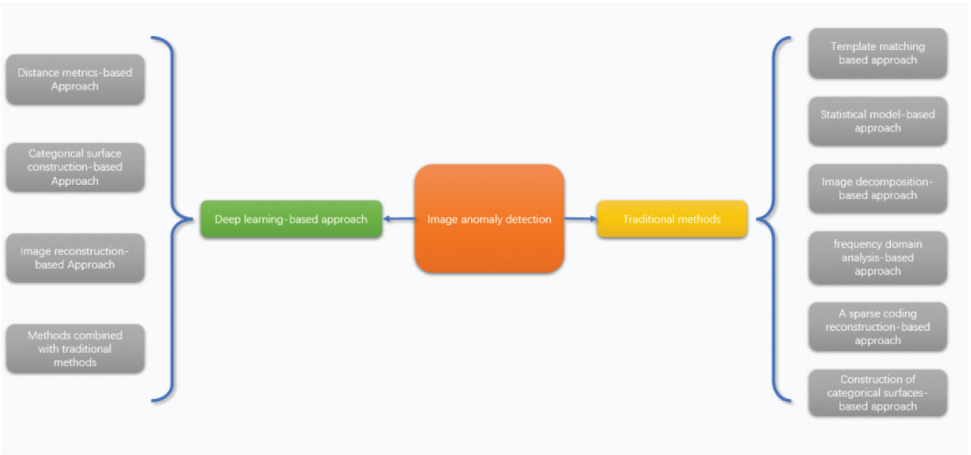


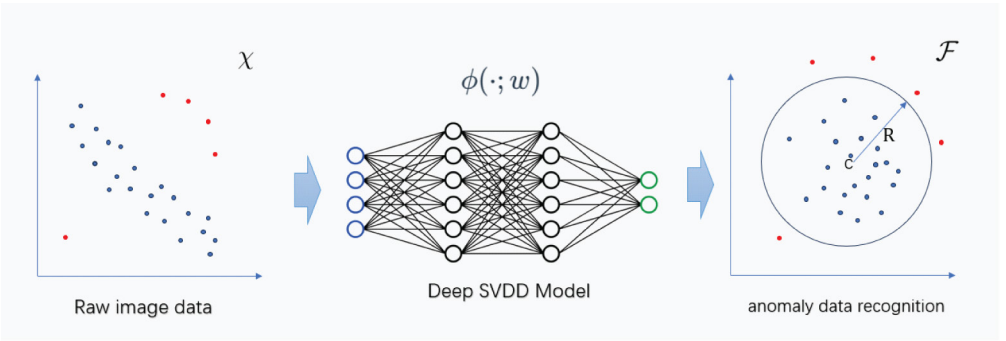

**Figure 5. Classification Diagram of Image Anomaly Detection Techniques**




## *Traditional Methods*

A large part of traditional target detection algorithms belongs to the category of supervised learning. This requires collecting enough samples of target categories and performing accurate labeling, such as the image category, the location of the target in the image, and the category information of each pixel point. Most traditional image anomaly detection algorithms learn a model to describe a normal image and subsequently perform anomaly detection in the detection phase based on the degree of match between the image to be examined and the existing model. According to the principle of the detection algorithm, the traditional image anomaly detection methods can be classified into the following categories: template matching–based, statistical model–based, image decomposition–based,

**Figure 6. Processing Flow of Deep Learning-Based Image Anomaly Detection Algorithm Deep SVDD**




frequency domain analysis–based, sparse coding reconstruction–based and anomaly detection methods based on categorical surface construction. In a study on template matching–based detection methods, Chen et al. (2019) extracted the Hu moments of the to-be-measured image and the template image as features to classify three-dimensional (3D) printed parts as abnormal or not. Reed and Yu (1990) used statistical modeling-based detection methods. The proposed the Reed-Xiao (RX) algorithm utilizes a Gaussian distribution function to describe the distribution of information within a pixel point in a hyperspectral image. Li et al. (2019) used the detection method based on image decomposition. In their research, by taking advantage of the low-rank nature of the periodic background texture, the original image to be detected is decomposed into a low-rank matrix representing the background and a sparse matrix representing the anomalous regions using the low-rank decomposition. Liu et al. (2010) used frequency domain analysis, performing the anomaly detection by differing the spectrum of the image to be examined from that of the normal image. (Liang et al. (2016) used the detection method based on sparse coding reconstruction. They reconstructed the image with the help of sparse coding and learned a dictionary to represent the normal image in the process. They then detected anomalies in terms of reconstruction differences and sparsity in the test phase. Schölkopf et al. (1999) used the detection method based on classification surface construction, specifically the one-class support vector machines (OC-SVMs) method.

### *Deep Learning–Based Methods*

Compared with traditional methods, deep learning has been widely introduced into the image anomaly detection task because it does not need the manual design of features and a higher versatility of algorithms. Existing methods can be roughly divided into the following categories: methods based on distance metrics, methods based on the construction of classification surfaces, methods based on image reconstruction, and methods combined with traditional methods. Ruff et al. (2019) used a method based on distance metrics, proposing deep support vector data description (Deep SVDD). Hendrycks and Gimpel (2016) used a method based on categorization surface construction. They geometrically transformed the original single-category samples to obtain multi-category samples and combined them with the more common confidence-based methods for out-of-distribution (OOD) tasks to detect anomalies. Hinton and Zemel (1993) used a method based on image reconstruction. Their self-encoder, trained using only normal samples, can reconstruct normal images well in the test phase. For images with abnormalities, the image encoding and subsequent reconstruction process will significantly differ from the normal image to realize the recognition function. Gupta et al. (2019) used a method combined with the traditional method. In particular, they used a network pre-trained with ImageNet as a feature extractor and then used OC-SVM to classify the anomalous images. The processing flow of the deep learning–based image anomaly detection algorithm Deep SVDD is shown in Figure 6.

Overall, deep learning-based methods can be applied to detect anomalies in images with various textures and structures thanks to the powerful learning ability of neural networks. Thus, they are significantly better than traditional methods in terms of detection accuracy and generality. However, these methods are also more complex and require the design of various strategies to ensure the smooth training of the network.

### *Main Difficulties*

The anomaly detection of video images has become more widely used in many fields, attracting the interest of researchers worldwide. However, the work still has many difficulties. The main ones are as follows:

1. The probability of anomalous events is low, the anomalous events are challenging to collect, and the vast majority of surveillance videos have normal segments, which results in the inability to train the network with supervised learning methods. We can only use unsupervised learning methods but cannot collect all the normal events to train the model. Thus, the models used to recognize the untrained networks are abnormal, increasing the difficulty of training the network model.
2. The definitions of normal and abnormal are fuzzy. In different scenes, the same behavior may be defined as two different situations. In the same scene, abnormal ways of behaving also vary. For example, running in the playground is normal, but running in the shopping mall suddenly can be recognized as abnormal. In a highway scene, the abnormality might be speeding, going against the flow of traffic, or pedestrians not using the crosswalk.
3. Video clips can have an occluded target, and as the target moves away, the number of detectable pixel points becomes less, which is not conducive to analyzing the anomaly.
4. Besides the above, problems such as lighting changes and high and low image resolution can affect the accuracy of video anomaly detection and increase the possibility of misjudgment.

## METHODS

This section introduces Mem-FCN, the video image anomaly detection model developed in this research. Firstly, we discuss video data preprocessing, comprising three key steps: data format conversion, video image size compression, and standardization of video image data. Subsequently, we introduce a deep learning-based algorithm for detecting anomalies in image data. The algorithm is built upon a fully connected neural network (He et al., 2022), augmented with a memory module. This module enhances the model's capacity for information storage, thereby improving the efficacy of model training.

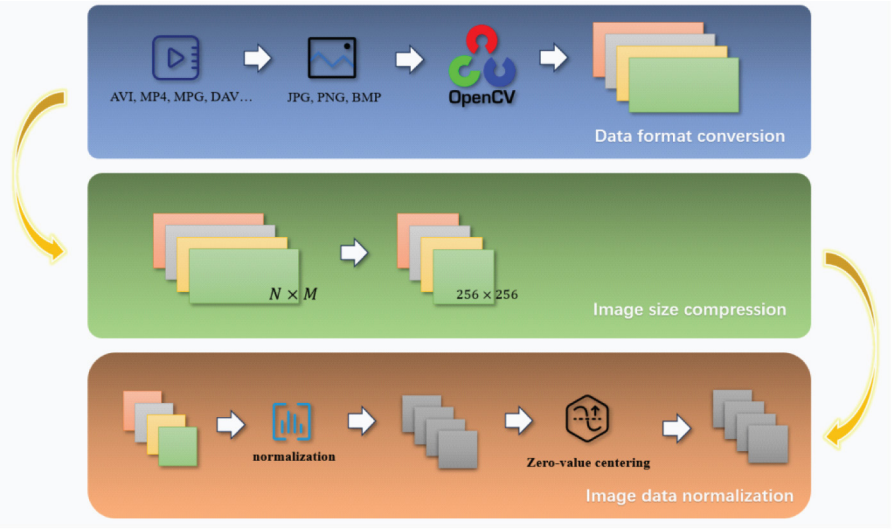
### Data Preprocessing

Data preprocessing is usually required to adjust the raw data into a unified format. Deep learning model training and parameter adjustment is usually a long process. If the model training due to data format or size and other factors leads to a system crash, all the training tasks need to restart, wasting significant time, so data preprocessing is also critical (Mukherjee et al., 2018). The preprocessing of image data generally includes grayscale, resampling, normalization, standardization, and zero-valued centering. This study summarizes the preprocessing into three steps: video data format conversion, image size compression, and image data normalization (Yuan et al., 2022).

### *Data Format Conversion*

Standard video formats include AVI, MP4, MPG, and DAV. Similarly, video frames extracted from videos come in various formats, such as JPG, PNG, and BMP. These formats are not directly

**Figure 7. Image Data Preprocessing Process**




compatible with the program. Therefore, we utilized the OpenCV (Sharmila et al., 2015) image processing library to handle the data. Specifically, we employed OpenCV to convert the video into a 3D matrix format, facilitating subsequent operations.

### *Image Size Compression*

Due to the variance in video frame resolutions across different datasets, it is imperative to standardize all image sizes before inputting them into the model. Theoretically, retaining larger sizes yields better results. However, excessively large sizes lead to an exponential increase in model parameters, resulting in computational burdens that may exceed the capacity of the equipment. Hence, a judiciously chosen size is essential for model input. In this study, we uniformly adjusted the original dataset resolution to 256×256 pixels.

### *Image Data Normalization*

Image data normalization is divided into two steps: normalization and zero-value centering. Normalization aims to map the image's grayscale to the standard grayscale value. The core idea is to perform a segmentation operation on the image grayscale and then linearly map it to the reference grayscale value. The formula for performing normalization is shown in Equation (1).

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

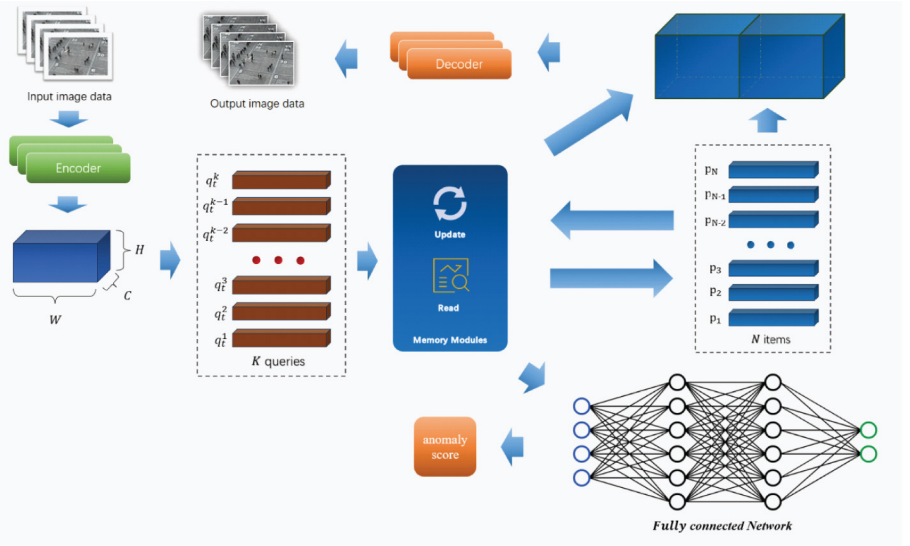
where $x$ is the gray value of the pixels, $\mu$ is the mean of the gray values of all pixels in the image, and $\mu$ is the standard deviation of the gray values of all pixels in the image.

Generally speaking, if the parameters such as contrast and brightness are appropriate, zero-value centering can be dispensed with. The flow of data preprocessing is shown in Figure 7.

## Deep Learning-Based Anomaly Detection and Recognition of Image Data

In this research, we designed a Mem-FCN model to realize the anomaly detection and recognition of video images. The structure of the model is shown in Figure 8. The details of the model are described in the following sections.

**Figure 8. The Structure of the Mem-FCN Model**




## *Autoencoder*

In this study, we constructed a new autoencoder (Pandey et al., 2022) that learns feature representations from the input frames and utilizes these features for reconstruction. The autoencoder consists of two parts, the encoder and the decoder, where the encoder uses convolution to extract shallow features and dimensionality reduction of the extracted features by pooling operation. First, a video frame is input to obtain a feature map of size $H \times W \times C$, where $H$, $W$, and $C$ are the feature map's height, width, and number of channels, respectively. We used a sequence of size $I \times 1 \times C$ in the query sequence to update the memory unit so that the memory module learns the prototype features of normal samples. The decoder restores the video frames to the original size image using inverse convolution, which is used to compute the pixel-level loss from the original image. We chose to minimize the mean square error (MSE) to optimize the above autoencoder model. The formula we used to calculate the loss between the reconstructed frame and the original frame is shown in Equation (2).

$$L_{mse} = \frac{1}{N}\sum_{i=1}^{n} \| \hat{T}_i - T_i \|_2^2 \tag{2}$$

where $N$ is the number of pixels in the image, $\hat{T}_i$ is a pixel in the reconstructed image for the input image, and $T_i$ is a pixel in the original image.

## *Memory Modules*

To facilitate the storage of image features, we define the number of memory units (M) and group them into memory modules. These modules possess identical sizes and lengths as the number of channels in the feature map output during the final stage of encoder down sampling. We opted for a conservative approach to mitigate computational overhead by employing ten memory units within each memory module.

The process of reading and updating the memory cells is shown in Figure 6. The left side is the schematic diagram of reading memory cells, and the right side is the schematic diagram of updating memory cells. The reading process is as follows: by calculating the cosine similarity between each

memory unit and the query sequence, the similarity is defined as the matching probability after *softmax* (·) function calculation, and the matching probability is used as the weight, which is combined with the memory units to form the prediction sequence and perform the next calculation. The updating process is as follows: we calculate the cosine similarity between all query sequences and memory units and get the probability set $X$ after *softmax*(·) function calculation, and the probability set is combined with the normalized query sequences to generate the updating sequence.

These memory units need to be read when up-sampling and training the Mem-FCN model, and we compute the cosine similarity between each query $d$ and all memory items $p$ as shown in Equation (3).

$$m_i^{k,n} = \frac{\exp\left((p_n)^T q_t^k\right)}{\sum_{j=1}^{M} \exp\left((p_j)^T q_t^k\right)} \tag{3}$$

Then, the match probability $m$ is obtained using the *softmax*(·) function, and the items are read using the match probability as a weighted average. Finally, we merge $q$ and the memory cells according to the weights to obtain a new predicted sequence $\hat{q}_t^k$, which is used as an input to the decoder or input to the Mem-FCN module for the next computation, and the merging formula is shown in Equation (4).

$$\hat{q}_t^k = \sum_{i=1}^{N} m_t^{k,i} p_i \tag{4}$$

where $N$ is the number of memory items, $p_i$ is the memory unit, and $m$ is the matching probability. For each memory unit, we aim to maximize the memory capacity of a single memory item to recognize a scene accurately.

When down sampling ends, we use the feature map obtained from the last convolution to update the memory units. We use Equation (4) to compute the cosine similarity between the memory unit and each sequence in the feature map. Then, we obtain the probability set $X$ by computing the *softmax* (·) function. The probability set is used as the weights, which are combined with the normalized sequences to generate an update sequence to update the corresponding memory unit. The schematic diagram of the memory module is shown in Figure 9.

### *Memory Enhancement Based Model Mem-FCN*

The fully connected neural network is derived from the CNN model (Turuk et al., 2022), renowned for its robust capacity to synthesize feature information from the input layer. Through multiple convolutions, it effectively amalgamates abstract features to achieve classification objectives. Leveraging the strengths of fully connected neural networks (Reddy et al., 2022), we devised a memory-based FCN to facilitate feature integration from normal samples. This enhancement aims to bolster the storage accuracy of the memory module by computing the loss between the output layer and the original feature map (Xie et al., 2023).

After the first stage of training, the memory module has learned the normal high-dimensional behavioral features. Then, its parameters are fixed before the Mem-FCN module is trained, which can maximize the saving of training pressure (Peng et al., 2023), and the Mem-FCN module can be trained to the best effect quickly by using the trained memory module. As shown in Figure 10, we extract the normal feature information of the predicted sequence (256 dimensions) through the FCN. At the same time, we use two intermediate layers to downsize the predicted sequence, and finally, the dimensionality is fixed at 64. We compress the original sequence of the feature maps obtained by the convolutional operation and calculate the cosine similarity with the former.

The formula used for the fully connected module is shown in Equation (5).

$$f(x) = G_2\left(b^{(2)} + W^{(2)}\left(G_1(b^{(1)} + W^{(1)}x)\right)\right) \tag{5}$$

**Figure 9. Schematic Diagram of the Memory Module**

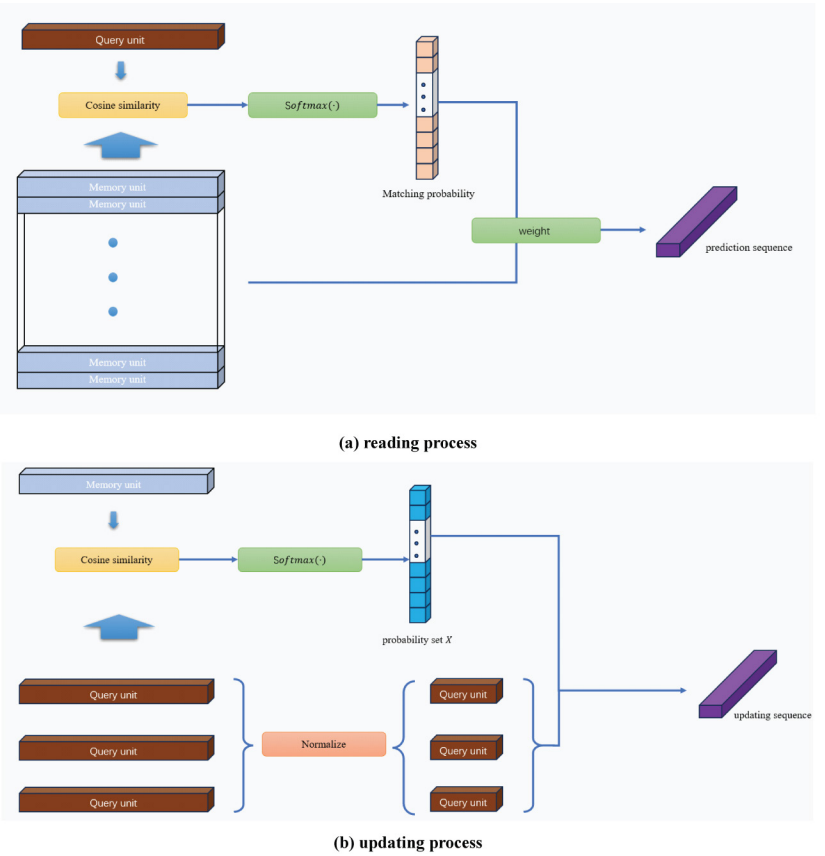



**Figure 10. Schematic Diagram of the Fully Connected Module**

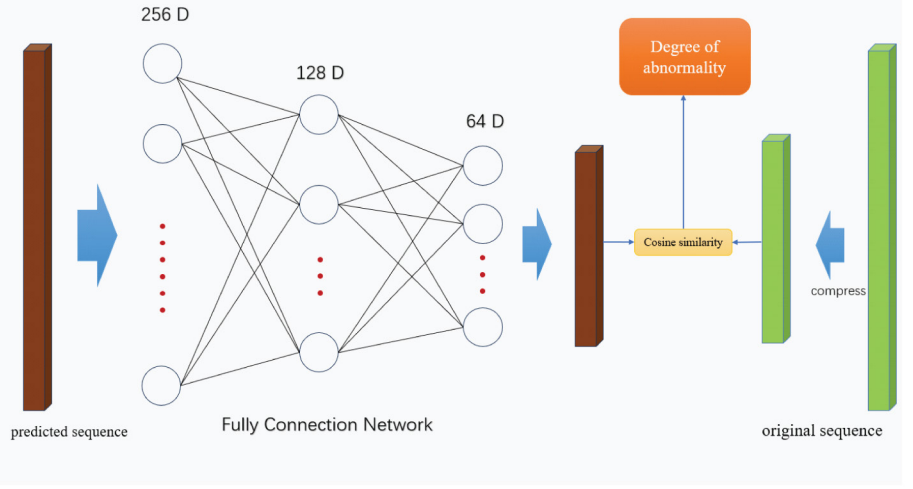
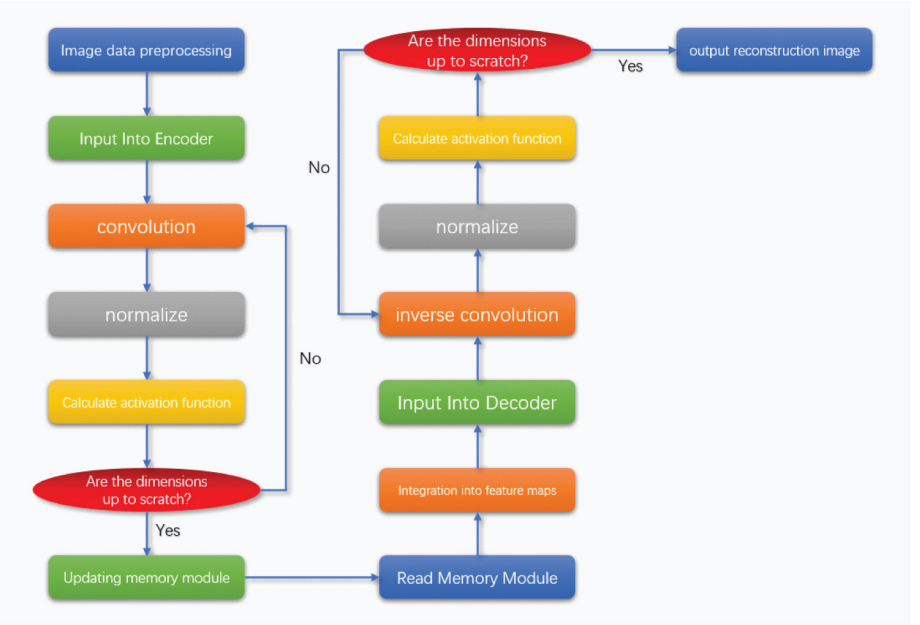


**Figure 11. First Stage of Model Calculation**




where $W$ denotes the weights and $b$ denotes the bias.

### *The Process of Computing Video Image Anomaly Detection*

The full process of model computation is described below. In the first stage, the input to the encoder is a training set of video frames containing only normal events, and the potential feature representations are output as a specific query sequence. The memory module learns the prototype patterns of the normal samples, and multiple memory units learn different prototypes individually. We aggregate the learned features into a decoder to reconstruct the input video frames.

In the second stage, the feature sequences are also fed into the Mem-FCN module, which learns the features that represent normal samples by recording the general representation of normal samples. In testing, the model calculates the loss between the reconstructed and original video frames and the loss calculated by the Mem-FCN module and employs a weighting rule to balance the two, resulting in an anomaly score as a basis for determining whether both are anomalous.

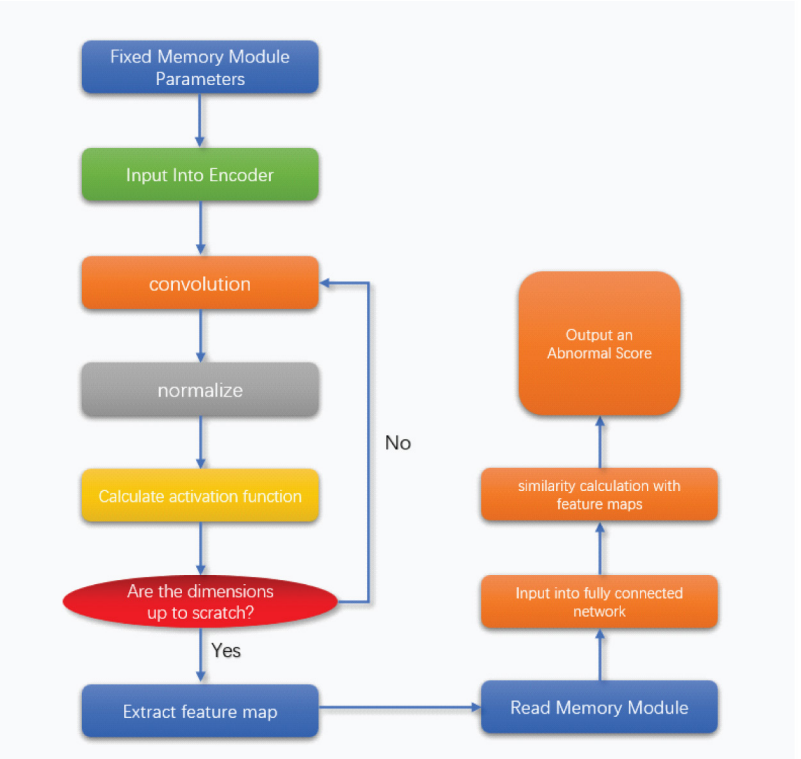
## EXPERIMENTS

This section describes the video anomaly detection simulation system we designed to detect abnormal image data of surveillance video. The above novel network is used to realize the video anomaly detection function, and the effectiveness of the algorithm proposed in this paper is verified by comparing and analyzing it with other baseline models, as shown in Figures 11 and 12.

### Experimental Procedure

The Mem-FCN model proposed in this paper is used to train on two datasets, with 80% as the training set and 20% as the test set. The parameters are constantly updated by iteration on the training set, and the optimal solution is calculated; on the test set, the trained model predicts the data. Then, the values of the five evaluation indexes are calculated according to the prediction results on the

**Figure 12. Second Stage of Model Calculation**




test set. To reduce the overhead cost of computation, interval sampling is utilized to draw one video frame in both adjacent video frames when calculating the evaluation metrics. Since the sample states are close to continuous, this results in no change in the probability of each anomalous image being sampled, and the sampled samples can represent the characteristics of the entire data space.

## Baseline Datasets

This section uses two public datasets widely used in the video anomaly detection problem: the University of California, San Diego Pedestrian 2 (UCSDPed2) dataset (Wang & Miao, 2010) and the Chinese University of Hong Kong Avenue (CUHKAvenue) dataset (Liu et al., 2018).

The UCSDPed2 anomaly detection dataset, curated by the University of California, San Diego, features footage from a stationary camera positioned high above the ground, surveilling a pedestrian sidewalk and vehicle traffic. The density of the crowd on the sidewalk varies from sparse to heavily congested. The dataset's resolution is 360×240 pixels. Normal events involve pedestrians moving in a typical manner, while abnormal events include bicyclists, skateboarders, vehicles, and pedestrians walking in uncommon areas. Figure 13 illustrates a sample video frame from the UCSDPed2 dataset, depicting a typical scene on the left side and an anomaly on the right.

The CUHKAvenue dataset boasts a resolution of 640×360 pixels and comprises videos captured by a fixed camera at the Chinese University of Hong Kong. Anomalous events in this dataset include individuals littering or throwing objects, loitering, approaching the camera, walking on grassy areas, and tossing objects. Examples of video frames from the CUHKAvenue dataset are depicted in Figure 14, with ordinary scenes featured on the left and anomalies on the right.

Figure 13. Examples of Video Frames from the UCSDPed2 Dataset

(a) Normal sample from train dataset (b) Anomaly sample from test dataset

Figure 14. Examples of Video Frames from the CUHKAvenue Dataset

(a) Normal sample from train dataset (b) Anomaly sample from test dataset

## Baseline Models

The 13 baseline models used in this paper are listed in Table 2. Five of these models—AESc, FCDD, TSN-HAD, MPI-VAD, and CVAD—were published in the last three years.

## Experimental Environment

This experiment was conducted in the experimental environment shown in Table 3.

## Evaluation Indicators

In machine learning, for a binary classification problem, we classify the test instances into positive or negative classes. However, when classifying image anomalies based on neural networks, the following four situations occur:

1. If an image is normal and is judged to be normal, it is a true positive (TP).
2. If an image is normal but is predicted to be abnormal, it is a false negative (FN).
3. If an image is abnormal but is determined to be normal, it is a false positive (FP).
4. If an image is anomalous and is judged to be anomalous, it is a true negative (TN).

The above four indicators form a confusion matrix, as shown in Figure 15.
With these data, the main indicators used to judge the model are as follows.

**Table 2. Baseline Model List**

| NO. | Method | Published year | Literature | Modeling ideas |
|---|---|---|---|---|
| 1 | AE | 2019 | Bergmann et al. (2019) | Utilizes a self-encoder for image reconstruction |
| 2 | AnoGAN | 2017 | Schlegl et al. (2017) | Utilizes the generator in GAN for image reconstruction |
| 3 | Iterative Projection | 2020 | Dehaene et al. (2020) | Uses iterative optimization to find the optimal normal image based on image reconstruction |
| 4 | AESc | 2021 | Collin, etc. (Dehaene et al., 2020) | Uses Monte Carlo to dropout the reconstructed network and use prediction uncertainty for anomaly localization |
| 5 | P-Net | 2020 | Zhou et al. (2020) | Adds constraints on the texture structure in the image reconstruction process; Uninformed Students, jointly considers the distance and variance between the to-be-measured map features to the target features for anomaly localization |
| 6 | CAVGA | 2019 | Venkataramanan et al. (2020) | Designed to localize anomalous regions by using an attention map based on image reconstruction |
| 7 | FCDD | 2021 | Liznerski et al. (2020) | Uses the full convolutional network to extract features and use the bias term as the center of feature mapping |
| 8 | Patch SVDD | 2020 | Yi and Yoon (2020) | Uses the bias term as the center of feature mapping |
| 9 | PaDiM | 2020 | Defard et al. (2021) | Uses pre-trained network for feature extraction and multi-dimensional Gaussian model for anomaly localization |
| 10 | SPADE | 2020 | Cohen and Hoshen (2020) | Designed to find the K-nearest neighbor normal images of the samples to be tested as a reference, and then detect the anomalies by the distance metric |
| 11 | TSN-HAD | 2022 | Rao et al. (2022) | Addresses the anomaly detection problem through similarity metric evaluation |
| 12 | MPI-VAD | 2022 | Xu et al. (2022) | Utilizes multiple probabilistic models' inference to detect as many different kinds of abnormal events as possible |
| 13 | CVAD | 2022 | Guo et al. (2022) | This open source tool models the latent embeddings of in-distribution, which formulates the "normality" features and provides the basis of excluding the outliers that show dissimilarity. |

**Table 3. Experimental Environment Table**

| **Hardware environment** | |
|---|---|
| Memory capacity | 32GB |
| Display card | GeForceRTX308012G |
| CPU | i7 |
| Solid state drive | 2T |
| **Software environment** | |
| Operating system | Windows 11 |
| CUDA | 10 |
| Python | 3.7 |
| PyTorch | 1.1.0 |
| **Model parameters** | |
| Learning rate | 0.0001 |
| Loss function | mean square loss function |
| Epoch | 60 |

**Figure 15. Confusion Matrix**

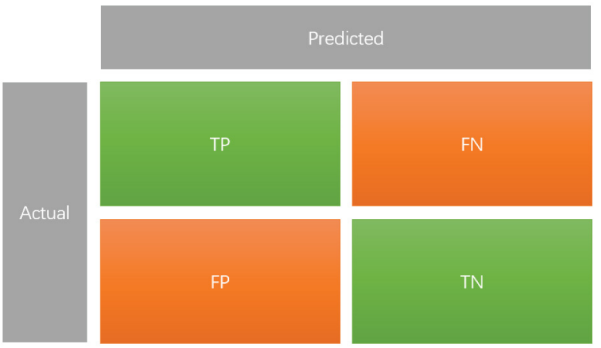



**Table 4. Table of Experimental Results**

| AUC | Accuracy | Recall |
|---|---|---|
| 96.80% | 96.40% | 90.20% |

The precision rate (precision) refers to the proportion of TPs set among the TP and FP sets. The formula is shown in Equation (6).

$$TPR = TP/(TP + FP) \tag{6}$$

The false positive rate (FPR) refers to the proportion of FPs in the TP and FP sets. The formula is shown in Equation (7).

$$FPR = FP/(TN + FP) \tag{7}$$

Accuracy refers to the number of image abnormalities in the whole data set that are determined correctly. The formula is shown in Equation (8).

$$ACC = (TP + TN)/(P + N) \tag{8}$$

Recall refers to how many samples in which the image is normal are determined to be normal.

The receiver operating characteristic (ROC) curve is a coordinate graph combining FPR and TPR. It is the resulting curve drawn using different evaluation methods for measuring the data under different conditions.

The area under curve (AUC) refers to the ROC curve; the value range generally falls between 0.5 and 1. As the ROC curve cannot be visualized to see if the result of the classifier is good or bad, and the AUC is a numerical value, the larger the AUC value, the better the classification effect. The value of the AUC can be viewed as two randomly removed positive and negative samples, and the positive sample score is greater than the probability of the negative samples.

## Experimental Results

As shown in Table 4, the Mem-FCN model performs excellently on the test datasets, with high values obtained for each metric, indicating the model has strong generalization ability.

**Table 5. Comparison of Average AUC Value on Two Datasets for Each Model**

| Model | AE | AnoGAN | Iterative Projection | AESc | P-Net |
|---|---|---|---|---|---|
| AUC value | 0.817 | 0.743 | 0.893 | 0.86 | 0.89 |
| Model | CAVGA | FCDD | Patch SVDD | PaDiM | SPADE |
| AUC value | 0.781 | 0.93 | 0.957 | 0.965 | 0.964 |
| Model | TSN-HAD | MPI-VAD | MPI-VAD | Mem-FCN | |
| AUC value | 0.956 | 0.964 | 0.963 | 0.968 | |

**Table 6. Comparison of Average Precision Rate Value on Two Datasets for Each Model**

| Model | AE | AnoGAN | Iterative Projection | AESc | P-Net |
|---|---|---|---|---|---|
| Precision rate value | 0.728 | 0.727 | 0.834 | 0.782 | 0.891 |
| Model | CAVGA | FCDD | Patch SVDD | PaDiM | SPADE |
| Precision rate value | 0.812 | 0.963 | 0.921 | 0.901 | 0.814 |
| Model | TSN-HAD | MPI-VAD | MPI-VAD | Mem-FCN | |
| Precision rate value | 0.912 | 0.923 | 0.941 | 0.965 | |

**Table 7. Comparison of Average False Positive Rate Value on Two Datasets for Each Model**

| Model | AE | AnoGAN | Iterative Projection | AESc | P-Net |
|---|---|---|---|---|---|
| FPR value | 0.12 | 0.12 | 0.13 | 0.09 | 0.16 |
| Model | CAVGA | FCDD | Patch SVDD | PaDiM | SPADE |
| FPR value | 0.07 | 0.08 | 0.04 | 0.09 | 0.05 |
| Model | TSN-HAD | MPI-VAD | MPI-VAD | Mem-FCN | |
| FPR value | 0.06 | 0.12 | 0.15 | 0.08 | |

## Comparative Experimental Results and Analysis

As shown in Tables 5–9, this experiment compares the detection results of the Mem-FCN model on the UCSDPed2 and CUHKAvenue datasets with several other typical detection methods. As can be seen in Tables 5–9, the Mem-FCN model has the best results on UCSDPed2 and CUHKAvenue, with average AUC values of 96.8% and 87.5%, respectively. Moreover, the Mem-FCN model significantly improves the AUC, precision, false positive rate, accuracy, and recall values for video image anomaly detection in both datasets.

Figure 16 shows the data in the previous tables plotted as a graph.

**Table 8. Comparison of Average Accuracy Rate Value on Two Datasets for Each Model**

| Model | AE | AnoGAN | Iterative Projection | AESc | P-Net |
|---|---|---|---|---|---|
| Accuracy rate value | 0.792 | 0.693 | 0.843 | 0.871 | 0.901 |
| Model | CAVGA | FCDD | Patch SVDD | PaDiM | SPADE |
| Accuracy rate value | 0.722 | 0.921 | 0.945 | 0.921 | 0.972 |
| Model | TSN-HAD | MPI-VAD | MPI-VAD | Mem-FCN | |
| Accuracy rate value | 0.912 | 0.901 | 0.935 | 0.964 | |

**Table 9. Comparison of Average Recall Rate Value on Two Datasets for Each Model**

| Model | AE | AnoGAN | Iterative Projection | AESc | P-Net |
|---|---|---|---|---|---|
| Recall rate value | 0.712 | 0.601 | 0.821 | 0.812 | 0.881 |
| Model | CAVGA | FCDD | Patch SVDD | PaDiM | SPADE |
| Recall rate value | 0.792 | 0.912 | 0.901 | 0.871 | 0.892 |
| Model | TSN-HAD | MPI-VAD | MPI-VAD | Mem-FCN | |
| Recall rate value | 0.894 | 0.904 | 0.911 | 0.902 | |

**Figure 16. Schematic Diagram of Comparison of Model Calculation Results**


**(a) Comparison of AUC** **(b) Comparison of TPR**

**(c)Comparison of average FPR** **(d) Comparison of accuracy**

**(e) Comparison of recall**

Comparing the results of the experiments, it can be seen that the AUC value of the Mem-FCN model is larger than the other models, at 0.003% larger than the second-largest model. The accuracy of the Mem-FCN model is also larger than the other models, at 0.003% larger than the second-largest model. Finally, the recall of the Mem-FCN model is larger than the other models, at 0.001% larger than the second largest model, which indicates that the Mem-FCN model possesses better predictive ability than the other models.

Carefully analyzing these experimental results gives us the following phenomena:

- **Phenomena (1):** The recall rate of the Mem-FCN model is slightly lower than that of algorithms proposed in recent years, such as FCDD, MPI-VAD, and MPI-VAD. This might be because after adding the memory module, the model's ability to memorize normal samples and its detection

**Table 10. Comparison of Average Evaluation Indicator Values on Two Datasets for Two Models**

| | AUC | Precision | FPR | Accuracy | Recall |
|---|---|---|---|---|---|
| FCN | 89.3% | 89.4% | 18% | 83.1% | 88.1% |
| Mem-FCN | 96.80% | 96.50% | 4.00% | 96.40% | 90.20% |

effect on normal samples are improved. However, the detection effect on abnormal samples is negatively affected.

- **Phenomena (2):** Mem-FCN model training is a bit slower than some of the baseline algorithms. This may be caused by the model's frequent reading and writing of data on the hard disk during training.
- **Phenomena (3):** The AUC value of Mem-FCN is higher than all other models. This result proves that the Mem-FCN model has significant advantages over other algorithms.
- **Phenomena (4):** The Mem-FCN model does not have the lowest FPR value among all the models. CAVGA, Patch SVDD, SPADE, and TSN-HAD have lower FPR values than the Mem-FCN model. FPR is slightly higher than some algorithms because the FCN has this problem, but the model presented in this paper does not suggest improvements in this regard.

### Ablation Experiments

To verify whether the memory module is helpful for image anomaly detection, this research compares the image anomaly detection effect of the FCN with a memory module and a single FCN without a memory module, and the results of the ablation experiments are shown in Table 10. The experimental results show that the FCN with the added memory module performs better in the image anomaly detection task.

As shown in Table 10, adding the memory module significantly improved all the indicators of model anomaly detection. The improvement in the accuracy rate is especially pronounced, probably due to the addition of the memory module, which significantly increased the probability of positive samples being recognized.

### New Model-Based Prediction Experiment on Another Dataset

The trained model and model parameters are saved in the training phase, and the prediction results can be obtained in the prediction phase by simply passing brand-new data into the model, as shown in Figures 17 and 18. This experiment runs the trained model on the University of California, San Diego Pedestrian 1 (UCSD Ped1) dataset (Wang & Miao, 2010) and obtains relatively accurate image anomaly prediction results.

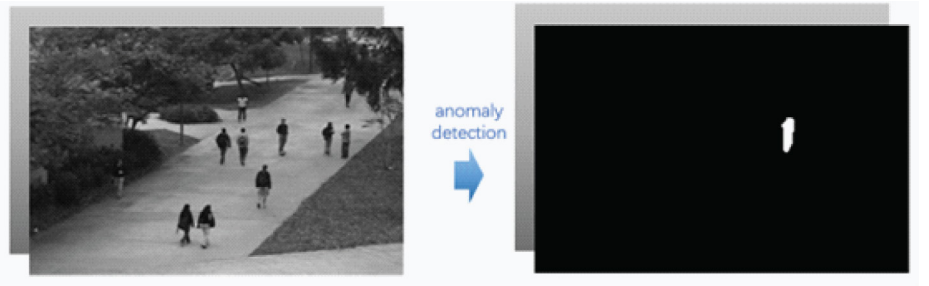
## CONCLUSION

### Summary

Unlike the goal of traditional image detection algorithms, the goal of video image anomaly detection is to construct classification models for anomaly detection of time series images. Many images that would not otherwise be considered anomalous are defined as anomalies in video image anomaly detection. In this paper, we first introduced the definition and common types of anomalies. Then, we reviewed and analyzed the development status of traditional and deep learning–based image anomaly detection methods according to the specific detection ideas. Next, we introduced a fully connected neural network model, Mem-FCN, based on memory augmentation, which effectively reduces the representation ability of CNNs by using the comprehensive feature capability of the FCN

Figure 17. First Example of Prediction Results of Mem-FCN Model on UCSD Ped1 Dataset


Figure 18. Second Example of Prediction Results of Mem-FCN Model on UCSD Ped1 Dataset




and realizes the purpose of remembering all the normal patterns by using the memory module, which improves the previous video anomaly detection model patterns, enhancing the operation of prior models used for video anomaly detection. Subsequently, we designed a video anomaly detection system that can detect abnormal image data from surveillance video and use the above novel network to implement the video anomaly detection function. We verified the effectiveness of the algorithm proposed in this paper by comparing and analyzing it with ten other baseline models (Zhong & Zhao, 2024).

## Future Work

The algorithm presented in this paper shows specific performance improvements compared to traditional detection methods and existing video anomaly detection algorithms. However, due to numerous influencing factors, the detection results are not optimal. We have identified three directions for improvement:

1. The evaluation metrics employed in this paper are relatively limited. Future research can explore a variety of evaluation metrics. By using multiple evaluation metrics comprehensively to derive anomaly scores for video frames without affecting the existing evaluation results, the detection outcomes can be more persuasive and credible, effectively reducing errors and enhancing detection performance.
2. Due to the complexity of the codec and memory module methods, integrating a background extraction module is currently not feasible. Streamlining the model and incorporating a background extraction module, analyzing only the foreground portion to determine video anomalies, would significantly enhance the overall detection accuracy of the model.

3. Additionally, we plan to experiment with different datasets and surveillance videos from our institution to continually refine the current model for broader applicability and improved performance.

## CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## FUNDING STATEMENT

No funding was received for this work.

## PROCESS DATES

Received: February 14, 2024, Revision: April 26, 2024, Accepted: April 16, 2024

## CORRESPONDING AUTHOR

Correspondence should be addressed to Qiguang Qian (China, qianqiguang@lumei.edu.cn)

## REFERENCES

Alghassab, M. A. (2022). Defect detection in printed circuit boards with pre-trained feature extraction methodology with convolution neural networks. *Computers, Materials & Continua*, *70*(1), 637–652. 10.32604/cmc.2022.019527

Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. 10.1109/CVPR.2019.00982

Chen, W. J., Ho, J.-H., Mustapha, K. B., & Chai, T.-Y. (2019). A vision based system for anomaly detection and classification in additive manufacturing. In *Proceedings of the 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET)*. IEEE. 10.1109/CSUDET47057.2019.9214635

Deecke, L., Vandermeulen, R., Ruff, L., Mandt, S., & Kloft, M. (2019). Image anomaly detection with generative adversarial networks. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases:European Conference*, pp. 3–17. ECML PKDD. 10.1007/978-3-030-10925-7_1

Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021). Padim: A patch distribution modeling framework for anomaly detection and localization. In *Proceedings of the International Conference on Pattern Recognition*, pp. 475–489. 10.1007/978-3-030-68799-1_35

Du, N., Huo, Y., & Wang, D. (2022). A video anomaly detection method based on percentile loss training and attention mechanism. *Displays*, *75*, 102327. 10.1016/j.displa.2022.102327

Gao, T., Wang, C., Zheng, J., Wu, G., Ning, X., Bai, X., Yang, J., & Wang, J. (2023). A smoothing group lasso based interval type-2 fuzzy neural network for simultaneous feature selection and system identification. *Knowledge-Based Systems*, *280*, 111028. 10.1016/j.knosys.2023.111028

Guo, X., Gichoya, J. W., Purkayastha, S., & Banerjee, I. (2022). CVAD—An unsupervised image anomaly detector. *Software Impacts*, *11*, 100195. 10.1016/j.simpa.2021.100195

Gupta, K., Bhavsar, A., & Sao, A. K. (2019). Detecting mitotic cells in HEp-2 images as anomalies via one class classifier. *Computers in Biology and Medicine*, *111*, 103328. 10.1016/j.compbiomed.2019.10332831326866

He, H.-W., Wang, Z.-H., Zou, Z.-J., & Liu, Y. (2022). Nonparametric modeling of ship maneuvering motion based on self-designed fully connected neural network. *Ocean Engineering*, *251*, 111113. 10.1016/j.oceaneng.2022.111113

Hinton, G. E., & Zemel, R. (1993). Autoencoders, minimum description length and Helmholtz free energy. *Advances in Neural Information Processing Systems*, ●●●, 6.

Hu, D., Liu, X., & Wang, L. (2022). An novel anomaly detection method for tiny defects on translucent glass. In *Proceedings of the IEEE 16th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*. IEEE. 10.1109/ASID56930.2022.9995999

Hussain, S. A., & Holambe, A. (2015). Automated detection and classification of glaucoma from eye fundus images: A survey. *International Journal of Computer Science and Information Technologies*, *6*(2), 1217–1224.

Kaur, H., Kaur, N., & Neeru, N. (2022). Evolution of multiorgan segmentation techniques from traditional to deep learning in abdominal CT images–A systematic review. *Displays*, *73*, 102223. 10.1016/j.displa.2022.102223

Li, C., Gao, G., Liu, Z., Huang, D., & Xi, J. (2019). Defect detection for patterned fabric images based on GHOG and low-rank decomposition. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 83962–83973. 10.1109/ACCESS.2019.2925196

Liang, L.-Q., Li, D., Fu, X., & Zhang, W.-J. (2016). Touch screen defect inspection based on sparse representation in low resolution images. *Multimedia Tools and Applications*, *75*(5), 2655–2666. 10.1007/s11042-015-2559-8

Liu, H., Zhou, W., Kuang, Q., Cao, L., & Gao, B. (2010). Defect detection of IC wafer based on spectral subtraction. *IEEE Transactions on Semiconductor Manufacturing*, *23*(1), 141–147. 10.1109/TSM.2009.2039185

Liu, W., Luo, W., Lian, D., & Gao, S. (2018). Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 10.1109/CVPR.2018.00684

Luo, G., & Pundlik, S. (2022). Usage patterns of mead-mounted vision assistance app as compared to handheld video magnifier. *Displays*, *75*, 102303. 10.1016/j.displa.2022.102303

Meng, Q., & Zhu, S. (2023). Anomaly detection for construction vibration signals using unsupervised deep learning and cloud computing. *Advanced Engineering Informatics*, *55*, 101907. 10.1016/j.aei.2023.101907

Modwel, G., Mehra, A., Rakesh, N., & Mishra, K. (2021). Advanced object detection in bio-medical X-ray images for anomaly detection and recognition. [IJEHMC]. *International Journal of E-Health and Medical Communications*, *12*(2), 93–110. 10.4018/IJEHMC.2021030106

Mrdovic, S. (2006). *Data mining for anomalous network payload detection. University of Sarajevo, Faculty of Electrical Engineering*. Bosnia and Herzegovina.

Mukherjee, S., Zhang, Y., Fan, J., Seelig, G., & Kannan, S. (2018). Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge. *Bioinformatics (Oxford, England)*, *34*(13), i124–i132. 10.1093/bioinformatics/bty29329949988

Murat, N. (2023). Outlier detection in statistical modeling via multivariate adaptive regression splines. *Communications in Statistics. Simulation and Computation*, *52*(7), 3379–3390. 10.1080/03610918.2021.2007400

Napoletano, P., Piccoli, F., & Schettini, R. (2021). Semi-supervised anomaly detection for visual quality inspection. *Expert Systems with Applications*, *183*, 115275. 10.1016/j.eswa.2021.115275

Ning, E., Wang, C., Zhang, H., Ning, X., & Tiwari, P. (2023). Occluded person re-identification with deep learning: A survey and perspectives. *Expert Systems with Applications*, ●●●, 122419.

Pandey, J. K., Kumar, S., Lamin, M., Gupta, S., Dubey, R. K., & Sammy, F. (2022). A metaheuristic autoencoder deep learning model for intrusion detector system. *Mathematical Problems in Engineering*, *2022*, 1–11. 10.1155/2022/3859155

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, *54*(2), 1–38. 10.1145/3439950

Pasini, K. (2021). *Forecast and anomaly detection on time series with dynamic context: Application to the mining of transit ridership data* [Doctoral thesis, Université Gustave Eiffel].

Ramaraj, K., Govindaraj, V., Murugan, P. R., Zhang, Y., & Wang, S. (2020). Safe engineering application for anomaly identification and outlier detection in human brain MRI. *Journal of Green Engineering*, *10*(10), 9087–9099.

Rao, W., Qu, Y., Gao, L., Sun, X., Wu, Y., & Zhang, B. (2022). Transferable network with Siamese architecture for anomaly detection in hyperspectral images. *International Journal of Applied Earth Observation and Geoinformation*, *106*, 102669. 10.1016/j.jag.2021.102669

Reddy, H., Kar, A., & Østergaard, J. (2022). Performance analysis of low complexity fully connected neural networks for monaural speech enhancement. *Applied Acoustics*, *190*, 108627. 10.1016/j.apacoust.2022.108627

Reed, I. S., & Yu, X. (1990). Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *38*(10), 1760–1770. 10.1109/29.60107

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Proceedings of the International Conference on Information Processing in Medical Imaging*. 10.1007/978-3-319-59050-9_12

Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., & Platt, J. (1999). Support vector method for novelty detection. *Advances in Neural Information Processing Systems*, ●●●, 12.

Sharmila, B., Karalan, N., & Nedumaran, D. (2015). Image processing on DSP environment using OpenCV. *International Journal of Advanced Research in Computer Science and Software Engineering*, *5*(2), 489–493.

Talagala, P. D., Hyndman, R. J., & Smith-Miles, K. (2021). Anomaly detection in high-dimensional data. *Journal of Computational and Graphical Statistics*, *30*(2), 360–374. 10.1080/10618600.2020.1807997

Tschuchnig, M. E., & Gadermayr, M. (2022). Anomaly detection in medical imaging-a mini review. In D*ata Science–Analytics and Applications: Proceedings of the 4th International Data Science Conference–iDSC2021*. 10.1007/978-3-658-36295-9_5

Turuk, M., Sreemathy, R., Kadiyala, S., Kotecha, S., & Kulkarni, V. (2022). CNN based deep learning approach for automatic malaria parasite detection. *IAENG International Journal of Computer Science*, *49*(3).

Venkataramanan, S., Peng, K.-C., Singh, R. V., & Mahalanobis, A. (2020). Attention guided anomaly localization in images. In *Proceedings of the European Conference on Computer Vision*.

Wang, J., Li, F., An, Y., Zhang, X., & Sun, H. (2024). Towards robust LiDAR-camera fusion in BEV space via mutual deformable attention and temporal aggregation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1. 10.1109/TCSVT.2024.3366664

Wang, K., Yuan, S., Yao, Z., Gao, J., & Feng, J. (2022). Design and implementation of infrared image classification algorithm for defective power equipment based on deep learning. *Nonlinear Optics, Quantum Optics: Concepts in Modern Optics, 56*.

Wang, S., & Miao, Z. (2010). Anomaly detection in crowd scene. In *Proceedings of the IEEE 10th International Conference on Signal Processing*. IEEE.

Wu, B., Liu, P., Wu, H., Liu, S., He, S., & Lv, G. (2023). An effective machine-learning based feature extraction/recognitionmodel for fetal heart defect detection from 2D ultrasonic imageries. *CMES-Computer Modeling in Engineering & Sciences, 134*(2).

Xie, X., Peng, H., Hasan, A., Huang, S., Zhao, J., Fang, H., Zhang, W., Geng, T., Khan, O., & Ding, C. (2023). Accel-gcn: High-performance gpu accelerator design for graph convolution networks. In *Proceedings of the IEEE/ACM International Conference on Computer Aided Design (ICCAD)*. IEEE/ACM. 10.1109/ICCAD57390.2023.10323722

Xu, Z., Zeng, X., Ji, G., & Sheng, B. (2022). Improved anomaly detection in surveillance videos with multiple probabilistic models inference. *Intelligent Automation & Soft Computing*, *31*(3), 1703–1717. 10.32604/iasc.2022.016919

Yan, H., He, M., & Mei, H. (2021). Adaptive multi-layer structure with spatial-spectrum combination for hyperspectral image anomaly detection. *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University, 39*(3), 484–491.

Ye, S., & Zhao, T. (2023). Team knowledge management: How leaders' expertise recognition influences expertise utilization. *Management Decision*, *61*(1), 77–96. 10.1108/MD-09-2021-1166

Yepmo, V., Smits, G., & Pivert, O. (2022). Anomaly explanation: A review. *Data & Knowledge Engineering*, *137*, 101946. 10.1016/j.datak.2021.101946

Yi, J., & Yoon, S. (2020). Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*.

You, H., Zhao, Y., Sun, Q., Wu, W., Lv, X., Chen, Y., Zhang, H., & Li, Z.-C. (2023). Deep learning MRI signature to predict survival and treatment benefit from temozolomide in IDH-wildtype glioblastoma. *Displays*, *77*, 102399. 10.1016/j.displa.2023.102399

Yuan, L., Li, H., Fu, S., & Zhang, Z. (2022). Learning behavior evaluation model and teaching strategy innovation by social media network following learning psychology. *Frontiers in Psychology*, *13*, 843428. Advance online publication. 10.3389/fpsyg.2022.84342835936300

Zhao, Z., Yang, W., Wang, N., & Xu, Z. (2010). A hardware design of defect detection system in cloth belt based on machine vision. In *Proceedings of the International Conference on Electrical and Control Engineering*. 10.1109/iCECE.2010.23

Zheng, Z., Hu, Y., Zhang, Y., Yang, H., Qiao, Y., Qu, Z., & Huang, Y. (2022). CASPPNet: A chained atrous spatial pyramid pooling network for steel defect detection. *Measurement Science & Technology*, *33*(8), 085403. 10.1088/1361-6501/ac68d2

Zhong, Z., & Zhao, Y. (2024). Collaborative driving mode of sustainable marketing and supply chain management supported by metaverse technology. *IEEE Transactions on Engineering Management*, *71*, 1642–1654. 10.1109/TEM.2023.3337346

Zhou, K., Xiao, Y., Yang, J., Cheng, J., Liu, W., Luo, W., Gu, Z., Liu, J., & Gao, S. (2020). Encoding structure-texture relation with p-net for anomaly detection in retinal images. In *Computer Vision–ECCV 2020:16th European Conference*. 10.1007/978-3-030-58565-5_22