Article

Meta Learning Approach for Adaptive Anomaly Detection from Multi Scenario Video Surveillance

Deepak Kumar Singh , Dibakar Raj Pant , Ganesh Gautam , Bhanu Shrestha *

Posted Date: 21 May 2025

doi: 10.20944/preprints202505.1613.v1

Keywords: Meta learning; MSAD Dataset; Swin Transformer; Video Anomaly Detector

Article

Meta Learning Approach for Adaptive Anomaly Detection from Multi Scenario Video Surveillance

Deepak Kumar Singh [1], Dibakar Raj Pant [1], Ganesh Gautam[1] and Bhanu Shrestha[2,*]

[1] Department of Electronics and Computer Engineering, Pulchowk Campus; Tribhuvan University, Nepal; deepak.singh@sagarmatha.edu.np (D.K.S.); drpant@ioe.edu.np (D.R.P.); ganesh@pcampus.edu.np (G. G)

[2] Department of Information Convergence System, Graduate School of Smart Convergence, Kwangwoon University, Seoul, Korea, bnu@kw.ac.kr

* Correspondence: bnu@kw.ac.kr

Abstract: Video surveillance is widely used in different area like road, mall, education, industries, retail, park, Bus stand, Restaurant and smart cities, each presenting unique anomalies requiring specialized detection. However, adapting anomaly detection models to novel viewpoints within the same scenario poses challenges. Extending these models to entirely new scenarios necessitate retraining or fine-tuning, a process that can be time-consuming. To address these challenges, Model named Video Anomaly Detector Model has been proposed, leveraging the meta learning framework for faster adaptation using Swin Transformer for feature extraction to new concepts. In response, The Multi-Scenario Anomaly Detection (MSAD) dataset, featuring 14 scenarios from various camera views, is the first high-resolution anomaly detection dataset for real-world applications. It includes diverse motion patterns and challenging variations like different lighting and weather conditions, providing a strong foundation for training advanced models. Experiments confirm the effectiveness of MAML, particularly when trained on the MSAD dataset. MAML excels under new viewpoints within the same scenario and performs competitively in entirely new scenarios, showcasing its potential for detecting anomalies in diverse and evolving surveillance environments.

Keywords: Meta learning; MSAD Dataset; Swin Transformer; Video Anomaly Detector

1. Introduction

Anomaly detection is essential in fields such as surveillance, healthcare, finance, and industrial monitoring, where identifying deviations from normal patterns is crucial. In video surveillance, anomalies can include unusual movements, unauthorized access, or unexpected incidents, with context playing a significant role in determining what is considered abnormal, for example, walking on a highway instead of a sidewalk. Since obtaining labeled video data for anomalies is challenging, many existing methods approach anomaly detection as a one-class classification problem, training models only on normal data and identifying anomalies as outliers during testing. However, current techniques face several limitations. Many fail to account for environmental variability, such as changes in lighting, weather conditions, and camera motion. Additionally, dataset limitations can lead to inaccurate representations of real-world anomalies, as certain activities, like cycling or running on a street, may be misclassified. Another major issue is the hu-man-centric focus of existing models, which prioritize detecting human-related anomalies while overlooking non-human anomalies, such as falling objects, fires, or water leaks. Furthermore, most datasets capture different camera angles within the same scenario but lack coverage of multiple distinct scenarios, reducing a model's ability to generalize effectively.

To address these challenges, the Multi-Scenario Anomaly Detection (MSAD) dataset [27] was used. This dataset includes high-resolution video recordings from 14 different scenarios, each captured from multiple camera perspectives. It incorporates a wide range of normal activities along with variations in lighting and weather conditions, ensuring better generalization across diverse environments. In addition to the dataset, this study introduces Model-Agnostic Meta-Learning (MAML) [4] as a novel approach to anomaly detection. Meta-learning, or "learning to learn," enables models to quickly adapt to new tasks with minimal data, which is particularly valuable when large-scale data collection is impractical. The meta-learning process consists of two phases: the meta-training phase, where the model is exposed to various tasks (i.e., different scenarios in the MSAD dataset) to develop a meta-model capable of adapting to new tasks, and the meta-testing phase, where the model is evaluated on previously unseen tasks to assess its ability to generalize.

This meta-learning approach offers several advantages for anomaly detection. It allows models to rapidly adapt to new environments with minimal data, improving their ability to function in dynamic settings. Training in diverse tasks enhances the model's ability to generalize, making it more effective in identifying anomalies in unfamiliar scenarios. Additionally, meta-learning increases efficiency by leveraging small amounts of anomalous data, addressing the common issue of data scarcity. Applying MAML to the MSAD dataset enables the model to handle new viewpoints by learning from multiple camera perspectives within the same scenario. It also demonstrates strong generalization capabilities when applied to entirely new scenarios, proving its effectiveness in real-world applications. This approach significantly strengthens anomaly detection in dynamic and evolving environments, making it a promising solution for improving security and monitoring systems.

Previous studies on anomaly detection across multiple scenarios have faced significant challenges in generalization. Traditional models struggle to adapt to different scenarios and viewpoints, and while few-shot learning (FSL) [26] methods have been introduced to address this issue, they remain limited to specific environments and fail to generalize effectively. To overcome these limitations, a Meta-Learning Framework, particularly Model-Agnostic Meta-Learning (MAML), can be employed. MAML optimizes model parameters to enable rapid adaptation to new tasks with only a few examples through minimal gradient updates.

The main contribution of this paper can be summarized as follows:

- MAML technique is designed to enable anomaly detection in 14 different scenarios including highway, shop, front door, office, parking lot, pedestrian street, street Highview, warehouse, road, park, mall, train, restaurant, sidewalk by using small number of normal and anomalous data, where data collection is limited.
- Our method leverages metadata and swin transformers to extract spatial features from frames of video data from different targeted scenes, thereby enhancing adaptability to different scenarios.
- By leveraging anomaly detector models using MAML to acquire the capability to classify between normal and abnormal data, better accuracy in anomaly detection from different scenarios with limited data is realized from targeted domain.

2. Related Works

Over the years, several counter measures have been proposed to detect anomaly for video surveillance in different sectors like education, manufacturing, road etc. Santoro et al., 2016 [19] focused on meta-learning techniques that address this by enabling models to learn from prior tasks and adapt quickly to new ones. Memory-Augmented Neural Networks (MANNs), especially Neural Turing Machines, provided external memory mechanisms that support rapid learning from a few examples. They also demonstrated that MANNs outperform LSTMs in both classification and regression tasks, showing strong potential for efficient, flexible meta-learning.

Finn, Abbeel et al., 2017 [4] concluded that MAML is a powerful and versatile me-ta-learning approach that allows models to adapt quickly to new tasks with minimal data. Its model-agnostic nature makes it applicable across various domains and tasks, ad-dressing key limitations of traditional machine learning methods in dynamic environments. The study demonstrates that MAML significantly enhances few-shot learning capabilities in both supervised and reinforcement learning contexts. Ravi et al., 2017 [17] studied and introduces an LSTM-based meta-learner that learns an optimization strategy tailored for few-shot scenarios by updating a learner model's parameters efficiently.

Munkhdalai et al., 2017 [13] introduced Meta Networks (MetaNet), a novel meta-learning framework designed to enhance rapid generalization and continual learning in neural networks, particularly in scenarios with limited labeled data. It built on previous meta-learning approaches that emphasize two-level learning, where a meta-learner acquired knowledge across tasks to improve a base learner's performance on new tasks.

Similarly, gong et al., 2019 [5] studied reconstruction-based techniques, such as those employing Auto-Encoder methods exclusively train on normal videos and subsequently test on both normal and abnormal videos. Soh and Cho et al., 2020 [21] published a paper that introduces a meta-transfer learning framework that could be adapted for anomaly detection tasks. The method aims to learn a meta-learner that can transfer knowledge from related tasks to new tasks with limited labeled data. By leveraging meta-transfer learning, the model can adapt to new anomaly detection tasks with minimal supervision, potentially improving detection performance.

Zhang et al., 2020 [25] proposed a novel few-shot learning framework leveraging Model-Agnostic Meta-Learning (MAML) which was proposed for bearing anomaly detection. The framework aimed to develop a robust fault classifier with minimal data by adapting quickly to new fault conditions. In a case study involving the introduction of new artificial faults, the method achieved up to 25 percent overall accuracy compared to a Siamese network benchmark. Moreover, when sufficient training data from artificial damages was available, MAML demonstrated competitive generalization abilities compared to state-of-the-art few-shot learning methods in identifying realistic bearing damages.

Smith et al., 2020 [20] addressed the limitations of traditional anomaly detection methods in cybersecurity, emphasizing the potential of meta-learning to enhance detection accuracy and adaptability with limited data. Diverse cyber security datasets were collected. Preprocessing involved normalization, feature extraction, and dimensionality re-duction. Meta-Learning Framework utilized Model-Agnostic Meta-Learning (MAML) to train models on various anomaly detection tasks. Evaluation Metrics assessed using precision, recall, F1-score, and AUC-ROC, compared to baseline models. Meta-learning outperformed traditional models, especially with limited data. Demonstrated robust performance on new, unseen data shows the adaptability of the system. Initial training was resource-intensive, but adaptation to new tasks was efficient.

Bergman et al., 2020 [1] presented a novel approach to anomaly detection through a classification-based method called GOAD, which addressed the limitations of existing techniques by utilizing a semi-supervised learning framework. It was built on previous works that categorize anomaly detection methods into reconstruction, distributional, and classification-based approaches, highlighting the effectiveness of deep learning in this domain.

Chen et al., 2021 [2] studied anomaly detection as a one-class classification problem, relying on the features of normal events where Generative adversarial network aims to re-construct normal events to minimize the reconstruction loss. Wu et al., 2021 [22] introduced an unsupervised transformer-based framework, Meta Former, for anomaly detection without relying on labeled data. The model leverages self-attention mechanisms to learn generalized representations from normal patterns, enabling it to detect anomalies across diverse domains. Meta Former captures both local and global contextual features, enhancing its ability to handle complex data distributions.

Reiss et al., 2022 [18] studied and experimented with anomaly detection methods which were often thought to struggle with understanding normal events, some models can effectively detect anomalies. However, achieving a good reconstruction isn't a guarantee of effective anomaly detection. These methods can be sensitive to object speeds, leading them to misclassify sudden motions as anomalies. Thus, a nuanced understanding of model behavior is necessary to improve anomaly detection accuracy. Natha et al., 2022 [14] conducted a systematic review of anomaly detection methods utilizing machine learning and deep learning techniques across various domains. The study categorizes approaches based on supervised, unsupervised, and semi-supervised learning, highlighting their strengths, limitations, and application areas. Meng et al., 2023 [11] proposed an explainable few-shot learning framework for online anomaly detection in ultrasonic metal welding processes with varying configurations. Their approach combines prototype-based learning with interpretability techniques to enable accurate detection and transparent decision-making using limited labeled data. The model effectively adapts to different welding setups, addressing data scarcity and variability challenges. Moon et al., 2023 [12] proposed an anomaly detection method using a model-agnostic meta-learning-based variational autoencoder (MAVAE) tailored for facility management systems. This approach enables rapid adaptation to new environments using few-shot learning by leveraging MAML to fine-tune VAE parameters with limited normal data.

Joshi et al., 2023 [8] focused on the head pose estimation techniques in this paper named "One, Five, and Ten-Shot-Based Meta-Learning for Computationally Efficient Head Pose Estimation". It compares MAML and FO-MAML methods for efficiency. FO-MAML uses first-order approximation, avoiding Hessian computation. Experiments were conducted in one-shot, five-shot, and ten-shot settings. FO-MAML shows faster optimization than traditional MAML. Navarro et al., 2023 [15] introduced Hydra, a meta-learning framework designed for fast model recommendation in unsupervised multivariate time series anomaly detection. Their approach leverages a small set of meta-features to match datasets with suitable detection models, addressing the challenges of scalability and generalization. Li et al., 2023 [9] proposed a novel zero-shot anomaly detection method leveraging batch normalization statistics to detect anomalies without requiring training data from the target distribution. Their approach models the distribution of batch normalization parameters across diverse tasks, enabling effective generalization to unseen domains.

Duan et al., 2024 [3] introduced a meta-learning framework for efficient anomaly diagnosis in few-shot AIOps scenarios, addressing the challenges of limited labeled data in IT operations. Their approach leverages Model-Agnostic Meta-Learning (MAML) to enable fast adaptation to new fault types with minimal training samples. Wu et al., 2024 [23] provided a comprehensive review of deep learning techniques for video anomaly detection, highlighting advancements across supervised, unsupervised, and self-supervised methods. The paper categorizes key approaches based on learning paradigms, feature representations, and anomaly scoring techniques. Zhang et al., 2024 proposed ADAGENT, a novel anomaly detection framework that leverages multimodal large models to operate effectively in adverse and complex environments. By integrating visual, textual, and contextual data, ADAGENT enhances anomaly detection robustness under challenging conditions such as low visibility or sensor noise.

Zhu, Raj et al., 2024 [26] developed a SAAD (Scenario Adaptive Anomaly Detection) model that takes a unique approach, utilizing a few-shot learning framework for faster adaptation to novel concepts and scenarios. This adaptability, coupled with its competitive performance on diverse scenarios, sets SAAD apart from traditional models. Jeong et al., 2024 [6] presented a novel approach for detecting pipeline leaks using limited normal data, particularly in hazardous environments like power plants. The authors proposed the GAD-PN framework, combining Cycle GAN for generating synthetic anomaly data and Prototypical Networks for anomaly classification based on few-shot learning.

Zhu, Wang et al., 2024 [27] published the paper "Advancing Video Anomaly Detection: A Concise Review and a New Dataset" proposes a novel approach to anomaly detection using the Scenario-Adaptive Anomaly Detection (SA2D) model. This model is designed to address limitations in existing few-shot learning frameworks by enhancing adaptability to multiple scenarios and viewpoints. Unlike traditional methods, SA2D utilizes a meta-learning framework to efficiently adapt to unseen contexts, making it suitable for diverse and dynamic environments. The authors introduce the Multi-Scenario Anomaly Detection (MSAD) dataset, which includes a variety of indoor and outdoor settings, dynamic environments, and multiple camera angles, improving upon the limitations of existing datasets like UCSD Ped1/Ped2 and ShanghaiTech. This dataset adopts frame-level annotations to ensure precise anomaly detection. The SA2D model employs future frame prediction for anomaly detection, supported by backbone architectures like U-Net and ConvLSTM for temporal modeling. It incorporates a generative adversarial network (GAN) for video frame reconstruction, demonstrating strong adaptability and competitive performance across challenging tasks. The paper highlights the importance of scenario diversity and meta-learning in advancing anomaly detection methodologies and datasets

In contrast to existing methods, meta learning framework can be utilized in models to adapt to multiple scenarios even new scenario with more accuracy. It includes MAML (Model-Agnostic Meta-

Learning) that trains a model's parameters in such a way that a small number of gradient updates can quickly adapt to a new task with few examples. This approach is versatile and can be applied to various domains including image classification, regression, and reinforcement learning.

3. Methodology

The methodology starts with data collection. The system model for video anomaly detection using model agnostic meta learning from video surveillance is illustrated in Figure 1, where Frames were extracted and then preprocessed. After preprocessing, feature extraction was done using swin transformer. And the extracted features as feature maps were fed to the anomaly detector model that uses MAML to classify the dataset into normal and anomalous.

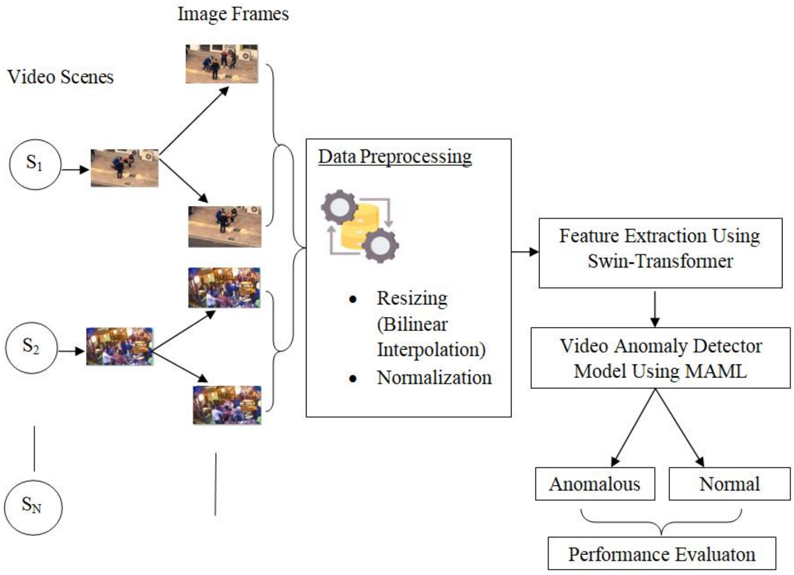


Image Frames
Video Scenes
S1
S2
SN
Data Preprocessing
Resizing (Bilinear Interpolation)
Normalization
Feature Extraction Using Swin-Transformer
Video Anomaly Detector Model Using MAML
Anomalous
Normal
Performance Evaluaton

Figure 1. Block Diagram of Video Anomaly Detection using MAML.

3.1. Dataset

MSAD (Multi Scenario Anomaly Detection) [1] datasets containing anomalies and normal instances have been collected. MSAD dataset has anomaly videos, normal testing and normal training datasets in video form. Dataset also contains annotated csv files that contain frame numbers. In csv file dataset has range values in column from where anomalies have been started. The MSAD dataset, along with any new datasets, serves as input to the system. Here, each dataset represents a different task.

A collection of 720 video datasets has been compiled, categorized as follows: anomaly videos, normal testing videos, and normal training videos. The anomaly video dataset comprises 240 videos, encompassing 11 distinct anomaly types across 13 different scenarios, excluding highway scenes. The normal testing video set includes a total of 120 videos. Finally, the normal training video dataset contains 360 videos, covering 14 different scenarios, including highway scenes.

3.2. Data Preprocessing

Raw data from MSAD dataset undergoes preprocessing to ensure data quality and consistency. This involves video frame resizing using bilinear interpolation method and Normalization using mean and standard deviation. Here Bilinear interpolation for y (height) and x (width) with four coordinate points $A(y_1, x_1)$, $B(y_1, x_2)$, $C(y_2, x_1)$ and $D(y_2, x_2)$ includes formula for computation for X and Y for width dimension is given by:

$$X = A(1 - w_x) + Bw_x \quad (1)$$

$$Y = C\,(1 - w_x) + Dw_x, \tag{2}$$

Then linear interpolation between the two interpolated values X and Y in the height dimension is given by:

$$Z = X\,(1 - w_y) + Yw_y\,, \tag{3}$$

$$= A(1{-}w_x)(1{-}w_y) + Bw_x(1{-}w_y) + C\,(1{-}w_x)w_y + Dw_xw_y \tag{4}$$

where,
$wx = (x - x_1)/(x_2 - x_1)$ and $w_y = (y - y_1)/(y_2 - y_1)$

Using the above equations, images were resized to required dimension i.e. 224×224 RGB image. Here PIL image with [H × W × C] in range [0,255] with unsigned int type were converted to Tensor of float type with shape [C × H × W] in the range [0,1]. In this case, PIL images belong to RGB mode. The normalization of resized images was done to standardize the images based on their mean and standard deviation, ensuring stable and consistent model performance.

For this study, the dataset was initially partitioned into training, validation, and testing sets based on different scenarios. For the initial training phase of the model, nine distinct scenarios were utilized, encompassing shop, front door, office, parking lot, pedestrian street, street Highview, warehouse, road, and park, each containing both anomalous and normal video data. Two different scenarios, mall and train, which also included both anomalous and normal videos, were allocated for the initial validation of the model. Finally, the initial testing of the model employed data from two separate scenarios: restaurant and sidewalk, both containing anomalous and normal video examples.

3.3. Data Preprocessing

For video frames (Image Input) of size 224 × 224 with 3 channels, the Swin Trans-former [10] was used to extract feature vectors from the preprocessed data of the MSAD dataset which is illustrated in Figure 2.

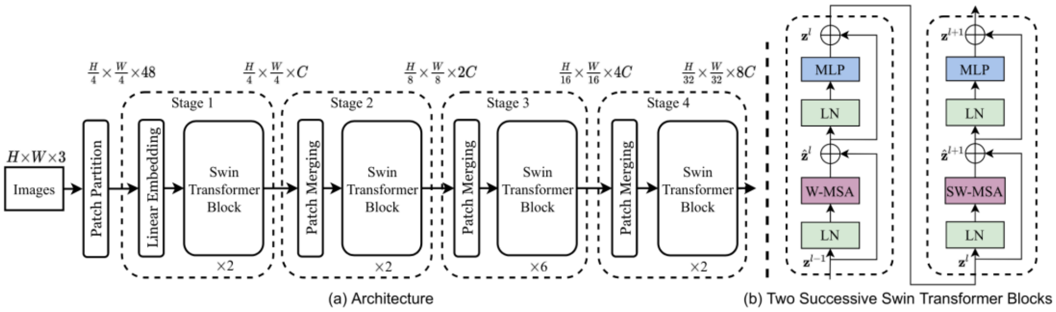


Figure 2. Illustration of feature extraction using swin transformer [10].

Here swin transformers were used for leveraging their hierarchical attention mechanism to capture both local and global spatial patterns. During the forward pass, the model processes the input images, and feature maps were extracted from intermediate layers, providing rich representations for anomaly detection. Mathematical interpretation, for input image I, the feature extraction process using Swin-S can be described as:

$$F = \mathrm{Swin}S(I\,), \tag{5}$$

where, I is the input image of size 224×224×3 and F is the feature vector whose output feature shape is 768 dimensional vectors (feature vectors). In patch embeddings, the image I is divided into patches of size 4×4 and each patch is embedded into a patch token T as

$$T = W_E I + b_E, \tag{6}$$

where, $W_E$is learnable embedding weight and b_E is bias. Then for hierarchical feature representation, swin transformer applies Shifted Windows Multi-Head Self Attention (SW-MSA) in hierarchical stages using:

$$F(l) = SW\text{-}MSA(MLP(F^{(l-1)})), \tag{7}$$

where, $F^{(l)}$ is the feature map at layer l and SW-MSA is the shifted window-based multi-head self-attention. Then feature extraction at final layer was done where final feature representation F can be taken from the last stage of the Swin Transformer using:

$$F = \text{Pooling}(F^{(L)}), \tag{8}$$

where, L is the last layer, and the pooling operation reduces spatial dimensions to a feature vector.

Finally, using input image 224×224×3 was given to swin transformer and 768 dimensional vectors as feature vectors were obtained and these embeddings were then served as inputs to the anomaly detector model, allowing for effective anomaly detection.

3.4. Data Preprocessing

In this phase, the architecture represents a feedforward neural network designed for anomaly detection using image embeddings as input. The model begins with an input layer that processes these embeddings, which are typically extracted from swin model. Figure 3 shows layered architecture where input is image embeddings (dimensional vectors) given to input layer. The Anomaly Detector class takes image embeddings as input, with a default feature size of 1024. The model consists of three sequential layers, each containing a fully connected (Linear) layer, batch normalization (BatchNorm1d), and a ReLU activation function to introduce non-linearity. The first layer reduces the input dimension from 1024 to 512, followed by a second layer that maintains the dimension at 512, and a third layer that reduces it further to 256. The final output layer is a linear trans-formation that maps the 256-dimensional feature representation to a single scalar logit. The forward method processes an input tensor through these layers sequentially, producing an output of shape, where each sample gets a single logit value. This output is used for binary classification, where we can get a probability for determining whether an input is normal or anomalous.

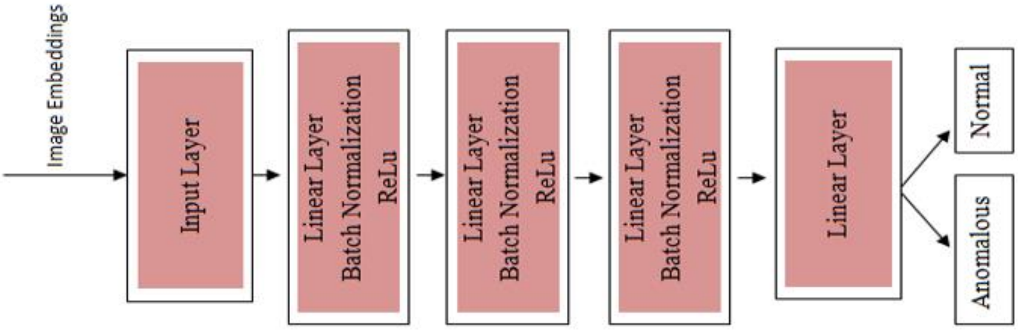


Figure 3. Detector model layered architecture.

3.4.1. Meta Training on Dataset

The meta training process in meta-learning consists of several key steps designed to help a model learn how to quickly adapt to new tasks (Image Embeddings) with limited data and a meta learning approach is illustrated in Figure 4.

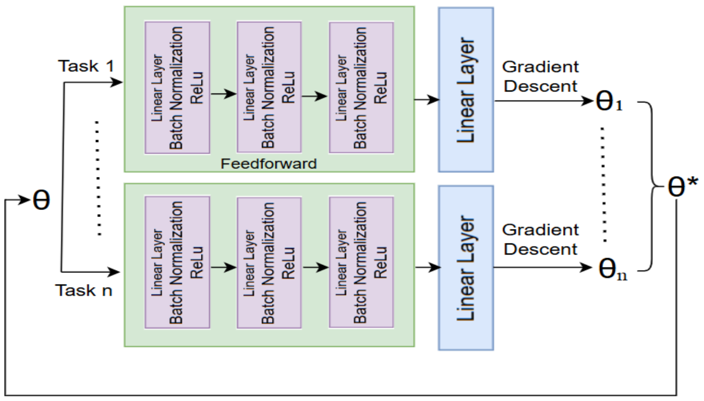


Figure 4. MAML training architecture.

- Task Sampling: The sample task method was designed to generate a dataset sample for anomaly detection task. Here a random scenario from a predefined list was selected. Within that scenario, it identified normal and anomalous images by cross-referencing scenario-specific indices with label-specific indices. Once these images were identified, it randomly selected a fixed number (k_shot + k_query) of normal and anomalous samples. These selected images were then split into two sets: the support set, which consists of k_shot normal and k_shot anomalous images for training, and the query set, which contains k_query normal and k_query anomalous images for evaluation. Each image was assigned to its respective label, either normal or anomalous. To avoid bias, the method shuffled both the support and query sets before returning them. Then the function ultimately ensured a balanced, structured dataset that can be used in meta-learning tasks, particularly for detecting anomalies within specific scenarios.
- Inner Loop: The method began by creating a temporary copy of the model to ensure that updates do not affect the original model. The copied model was then placed in training mode, and a Binary Cross Entropy with Logits Loss function was defined for classification. Then a series of inner-loop updates were performed over self-inner steps iterations. In each iteration, it made predictions using the support set, calculated the loss, and computed gradients with respect to the model's parameters. Instead of using an optimizer, it manually updated each parameter using stochastic gradient descent (SGD) by subtracting the gradient scaled by the inner learning rate. This adapted model, then fine-tuned to the support set, was returned. The purpose of this method was to allow a model to quickly adapt to a new task using only a few examples. This approach maintains the computational graph so that the outer update can backpropagate through the inner adaptation steps.
- Outer Loop: The meta update function implements a meta-learning step based on the Model-Agnostic Meta-Learning (MAML) approach. It processed multiple tasks, where each task consists of a support set (used for inner-loop adaptation) and a query set (used for evaluation). First, it initialized the loss function, optimizer, and storage variables for query labels and predictions. Then, it iterated over tasks, moving data to the appropriate device and performing an inner-loop update, where the model adapts to the task using the support set. After adaptation, the model made predictions on the query set, and the binary cross-entropy loss was computed. The predicted probabilities and true query labels were stored for later analysis. The total meta-loss was averaged across tasks, and outer-loop optimization was performed by backpropagating the loss. Finally, the main model's weights were updated using the adapted model, and the function returned the final loss along with stored query labels and predictions. This process helped the model learn a generalizable initialization that can quickly adapt to new tasks with minimal updates.
- Meta Validate: A meta-learning model was then evaluated by iterating over a set of tasks, each containing a support set for fine-tuning and a query set for evaluation. It first initialized the total loss and prepared storage for true labels and predictions. As it looped through tasks, it moves

data to the appropriate device (CPU/GPU) and fi-ne-tuned a copy of the model on the support set using an inner-loop update function (inner update). After adaptation, it made predictions on the query set, applied a sigmoid activation function to convert logits into probabilities, and stored both true and predicted labels. Then the function computed the binary cross-entropy loss for each task, accumulated the losses, and then averaged them over all tasks. Finally, it invoked an early stop-ping mechanism, which can determine if the model should be saved or training should halt based on the meta-loss trend. Then the function returned the averaged loss, true query labels, predicted labels, and the early stopping decision, ensuring model evaluation in a meta-learning framework.

3.4.2. Meta Testing on Dataset

A new dataset, which the model has not seen during the meta-training phase, was prepared. This dataset contains new tasks from two different scenarios (restaurant and sidewalk) that the model needs to adapt to. Then, like the meta-training phase, tasks were sampled from the novel dataset. Each task was defined with a small number of training examples (support set) and testing examples (query set).

For each sampled task, the model adapted its parameters using the support set. This was usually done using the same adaptation procedure (inner loop) as in the meta-training phase. The key difference here was that the initial model parameters or meta-parameters, obtained from meta-training, were used as the starting point. After adaptation, the model was evaluated on the query set of the same task. The performance was recorded to assess how well the model had adapted to the new task.

3.5. Data Preprocessing

In the context of video anomaly detection, each task T_i corresponds to a different set of video frames, representing a unique scenario or environment. The collection of all such tasks is denoted by $T = \{T_1, T_2, \ldots, T_n\}$. For each task T_i, there exists a dedicated dataset D_i={(x_j,y_j)}mi for j=1, where x_j is the input and y_j is the corresponding level. The model was used for anomaly detection is parameterized by θ, which represents the set of all learnable parameters. During meta-training, the goal is to find an optimal initialization of θ such that, after a few gradient updates using data from any task T_i, the model quickly adapts and performs well.

3.5.1. Dataset Meta Training Phase

The meta training phase aims to find an initial set of model parameters, denoted as θ, that can rapidly adapt to new tasks. Within this phase, the inner loop involves task-specific adaptation: for each task T_i, the model performs a few steps of gradient descent to adapt the initial parameters θ, resulting in task-specific parameters θ'_i.

$$\theta'_i = \Theta - \alpha \Delta_\theta L_{Ti}(\theta) \tag{9}$$

where α is the learning rate and $L_{Ti}(\theta)$ is the loss function for task T_i.

Following this, the outer loop performs a meta-update by adjusting the initial parameters θ based on the performance of the adapted parameters θ'_i across all tasks, enabling better generalization to unseen tasks.

$$\theta \ \ <\!-\theta - \beta \Delta_\theta \textstyle\sum_i L_{Ti}(\theta'_i) \tag{10}$$

where, β is meta learning rate.

3.5.2. Meta Testing Phase

During the meta testing phase, the learned initial parameters θ are fine-tuned on a new dataset corresponding to a previously unseen task using a few gradient descent steps. Given a new task T_{new} with its dataset D_{new}, the model adapts its parameters according to the following update equation:

$$\theta'_{\mathrm{new}} = \theta - \alpha \Delta_{\theta} L_{\mathrm{Tnew}}(\theta) \tag{11}$$

where, α is the learning rate and $L_{Tnew}(\theta)$ represents the loss on the new task. This adaptation allows the model to quickly generalize new tasks using minimal training data.

3.5.3. Loss Function for Anomaly Detection

For anomaly detection, the loss function $L_{\mathtt{ti}}(\theta)$ can be designed based on either reconstruction error or the output of a discriminator. In a reconstruction-based approach, the loss function measures how well the model can reconstruct the input data and, it can be defined as:

$$L_{Ti}(\theta) = \Sigma_{(x,y) \in Di} \| x - f_{\theta}(x) \|^{2} \tag{12}$$

where $f_{\theta}(x)$ is the reconstructed output of the model. This formulation penalizes large deviations between the input x and its reconstruction, thereby helping the model learn normal patterns and identify anomalies as instances with high reconstruction errors.

3.5.4. Model Agnostic Meta Learning Architecture

This figure 5 illustrates the core concept of Model-Agnostic Meta-Learning (MAML), where the objective is to find an optimal set of initial parameters θ that can quickly adapt to a variety of tasks with minimal training. The solid curve represents the trajectory of meta-learning, which optimizes the initial parameters across multiple tasks. Each red arrow (∇L_1, ∇L_2, ∇L_3) indicates the task-specific gradient computed during the inner loop of training for different tasks. These gradients guide the adaptation of the initial parameters θ toward task-specific optima (θ_1*, θ_2*, θ_3*), shown as blue dashed arrows labeled as "Learning/Adaptation." The goal of MAML is to optimize θ such that a small number of gradient steps can lead to effective task-specific models, enabling rapid learning on new tasks. The figure highlights how meta-learning integrates feedback from multiple tasks to update the shared initialization in a way that supports efficient fine-tuning.

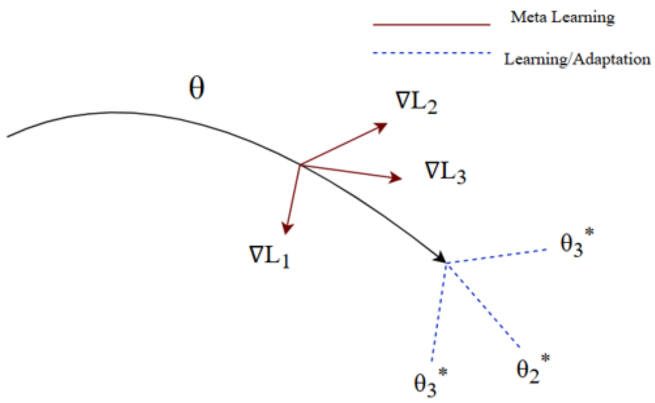


Figure 5. The architecture of MAML.

To implement the meta-learning approach in a task-agnostic manner, the following algorithm outlines the Generic Model-Agnostic Meta-Learning (MAML) procedure. This algorithm is designed to optimize an initial model that can quickly adapt to new video frame-based anomaly detection tasks with only a few gradient steps. The steps below detail the inner and outer learning loops required to achieve this fast adaptation:

```
Algorithm: Generic Model Agnostic Meta learning algorithm
Require: $D(T)$: Distribution over tasks (Video frames)
Require: $\alpha$ and $\beta$: step size hyperparameters
Step1. Randomly initialize parameter $\theta$
Step2. While not done do
Step3. Sample batch of task $Ti \sim D(T)$
Step4. For all $Ti$ do
```

```
Step5. Evaluate $\nabla_\theta L_{T_i}(f_\theta)$ with respect to k samples
Step6. Calculate adapted parameters using gradient descent:
$\theta'_i = \theta - \alpha\nabla_\theta L_{T_i}(f_\theta)$
Step7. end for
Step8. Update $\theta \leftarrow \theta - \beta\nabla_\theta\sum_{T_i\sim D(T)} L_{T_i}(f'_\theta)$
Step9. end while
```

4. Result and Analysis

In this study, MSAD videos were extracted into frames. Then these frames were preprocessed by resizing to 224×224. And normalization using mean and standard deviation is used. After normalization, input images were given to swin transformers that gives 768 feature vectors. These vectors were then given to detector models for classification into anomalous or normal. The output of the study is to classify input video frames into anomalous and normal types illustrated in Figure 6.

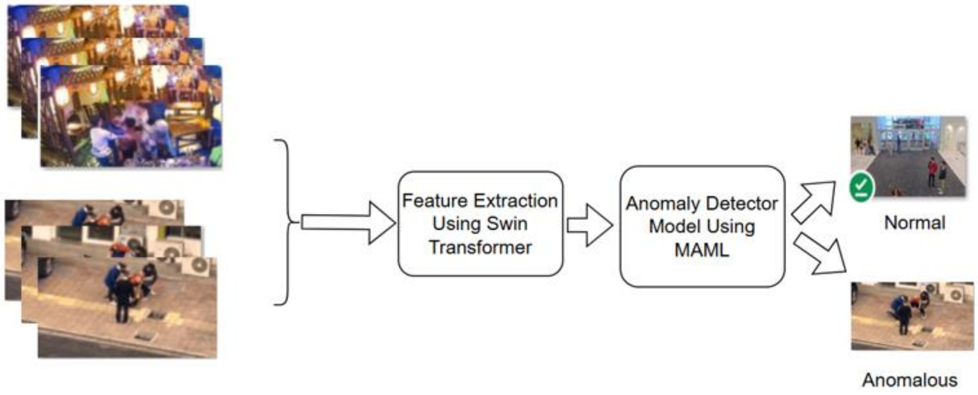

Feature Extraction Using Swin Transformer
Anomaly Detector Model Using MAML
Normal
Anomalous

Figure 6. Output of video anomaly detection.

The experimental setup contains given MAML parameters as illustrated in given Table 1. Using these parameters, the model was trained, validated and tested. Ten shot one query approach has been implemented

Table 1. The MAML parameters and hyperparameters used in the implementation.

Metrics	Values
Epoch	100
K-Shot	10
K-Query	1
Optimizer	Adam
N-Way	5
Inner Learning Rate	0.1
Outer Learning Rate	0.01
Batchsz	2000
Input Image Size	224x224
No. of Layers	3
Feature Embedding Size	768
Device	Cuda

4.1. Confusion Matrix

The confusion matrix for model evaluation is given in Figure 7, where the matrix consists of four key values: true positives (1668), where the model correctly identified class "1"; true negatives (1671), where it correctly identified class "0"; false positives (329), where it mistakenly classified a "0" as "1;

and false negatives (332), where it incorrectly classified a "1" as "0". From these values, key performance metrics were calculated.

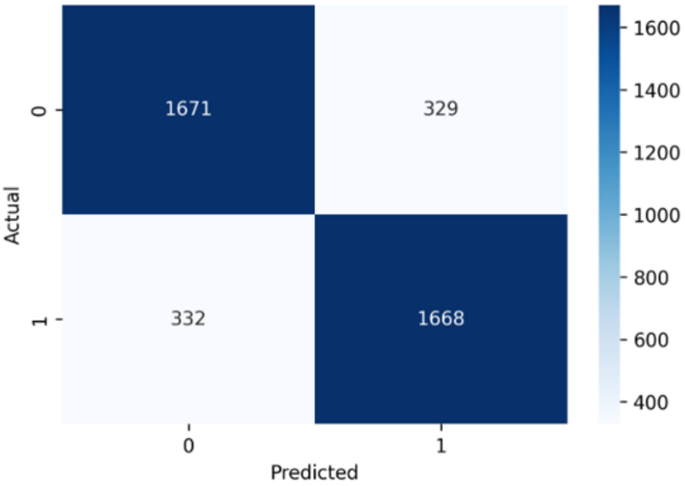


Figure 7. Confusion Matrix for Model Evaluation.

The performance matrix can be expressed in terms of accuracy, precision, recall and F1-score which are calculated as follows:

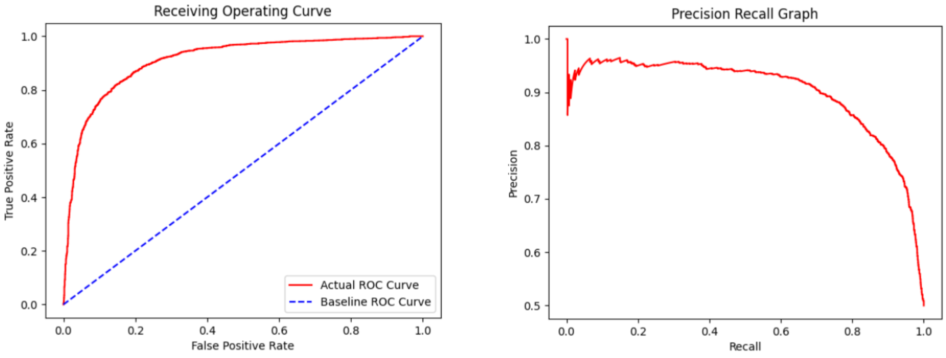
- Accuracy = (T P + T N)/(T P + T N + F P + F N) = 3339/4000 = 0.834
- Precision = T P/(T P + F P) = 1668/(1668 + 329) = 1668/1997 = 0.835
- Recall = T P/(T P + F N) = 1668/(1668 + 322) = 1668/2000 = 0.834
- F1-Score = 2 ∗ (Precision ∗ Recall)/(Precision + Recall) = 0.834

For video anomaly detection, this confusion matrix indicates that the model performs reasonably well, correctly classifying 83.5% of cases. The F1-score, a balance between precision and recall, is 83.4%, showing consistent performance. However, the presence of 332 false negatives suggests that some anomalies are being missed, which could be critical in applications where detecting anomalies is the priority. The false positives (329) indicate some normal events are beingflagged as anomalies, which may lead to unnecessary alerts. Given the nature of anomaly detection, improving recall (currently 83.4%) is crucial to minimize missed anomalies. Fine-tuning the model, adjusting thresholds, or incorporating additional features could help reduce false negatives and improve overall detection performance.

4.2. Learning Curve

Figure 8 shows the two kinds of model evaluation, the Receiver Operating Characteristic (ROC) and the Precision-Recall (PR) curve. The ROC curve shown in Figure 8(a), evaluates the performance of the video anomaly detection model by illustrating the trade-off between the true positive rate (TPR) and the false positive rate (FPR) at different classification thresholds. The red curve represents the actual model's performance, while the blue dashed line serves as a baseline, indicating random guessing. A well-performing model should have its ROC curve significantly above this baseline, which is the case here. The steep initial rise of the curve suggests that the model can detect a substantial portion of anomalies with a low false positive rate, which is crucial for anomaly detection. However, as the curve flattens, increasing sensitivity results in a higher number of false positives, potentially leading to unnecessary alerts. The Precision-Recall (PR) curve shown in Figure 8(b) for video anomaly detection provides insight into the model's ability to distinguish anomalies from normal events. Initially, at very low recall values, the model achieves extremely high precision, indicating that the first few detected anomalies are highly accurate with minimal false positives. As recall increases, meaning the model detects more anomalies, precision remains relatively stable before gradually declining. This suggests that while the model can correctly identify many anomalies with high precision, its performance degrades as it attempts to detect a larger number of anomalies. The sharp decline in precision at higher recall values indicates that an increasing number of normal events are being misclassified as anomalies, leading to a rise in false positives. This trade-off is common in anomaly detection systems, where optimizing high recall often results in reduced precision.

To improve the model's effectiveness, techniques such as better feature extraction, threshold tuning, or post-processing methods like spatial smoothing can help reduce false positives while maintaining a balance between precision and recall. The optimal operating point for the model depends on the application—whether prioritizing precision to minimize false alarms or maximizing recall to ensure all potential anomalies are detected.



(a) (b)

Figure 8. The model evaluation: (**a**) ROC curve for model evaluation; (**b**) PR curve for model evaluation.

4.3. Model Comparison

Table 2 represents the comparative analysis of 4 different models including SAAD, RFTM, MGFN and Anomaly Detector Model (MLP+MAML+SWIN). The MLP+MAML+SWIN model outperformed all other models in comparison due to its unique combination of meta-learning and advanced feature extraction. Unlike traditional models like SAAD, which achieved a lower AUC of 0.69, or RFTM and MGFN, which reached AUCs of 0.867 and 0.85 respectively, implemented model achieved a remarkable AUC of 0.91. This significant improvement is largely attributed to the integration of MAML (Model-Agnostic Meta-Learning), which allows the model to adapt quickly to new scenarios with minimal data, a crucial feature in video anomaly detection, where anomalies vary across different contexts. The Swin Transformer backbone enhanced feature extraction by effectively capturing spatial dependencies in the video frames through its attention mechanism. This enabled the model to accurately identify even subtle anomalies in complex video data. Furthermore, the combination of MLP with MAML and Swin Transformer leveraged the strengths of both powerful feature extraction and robust decision-making, ensuring optimal performance.

Table 2. Model comparison.

Model	AUC
SAAD	0.69
RFTM	0.867
MGFN	0.85
MLP+MAML+SWIN(Our Model)	**0.91**

5. Conclusions

In this study, a comprehensive approach was presented to video anomaly detection by leveraging a multi-scenario anomaly detection dataset. The dataset, containing both normal and anomalous video sequences, served as the foundation for extracting frames, which were then resized and normalized to ensure consistency and prepare them for effective processing. The preprocessing steps ensured that the input data was in an optimal format for feature extraction, maintaining size 224×224 with 3 channels. Then the anomaly detection approach demonstrated the effectiveness of

Swin Transformer in capturing spatial dependencies in video frames while detector model using MAML (10 shot 1 Query approach) was employed to feature map for classifying video frames into normal and anomalous.

The results demonstrated significant improvement in anomaly detection performance, showing the power of combining Swin Transformer for feature extraction with MAML for meta-learning. This approach not only improved classification accuracy but also ensured better generalization to novel and complex datasets.

Author Contributions: Conceptualization, D.K.S., D.R.P. and G.G.; methodology, B. S., D.K.S. and D.R.P.; software, D.K.S.; validation, D.R.P, G.G. and B.S.; formal analysis, D.K.S.; investigation, D.K.S.; resources, D.K.S.; data curation, D.K.S.; writing—original draft preparation, D.K.S. and G.G.; writing—review and editing, D.K.S., G.G. and B.S.; visualization, D.K.S. and D.R.P.; supervision, D.R.P.; project administration, B.S., D.R.P,. and G.G.; funding acquisition, G.G. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: “This research was conducted by the grant of Kwangwoon University in 2025”.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bergman, L.; Hoshen, Y. Classification-Based Anomaly Detection for General Data. arXiv 2020, arXiv:2005.02359.
2. Chen, D.; Yue, L.; Chang, X.; Xu, M.; Jia, T. NM-GAN: Noise-Modulated Generative Adversarial Network for Video Anomaly Detection. Pattern Recognit. 2021, 116, 107969.
3. Duan, Y.; Bao, H.; Bai, G.; Wei, Y.; Xue, K.; You, Z.; Ou, Z. Learning to Diagnose: Meta-Learning for Efficient Adaptation in Few-Shot AIOps Scenarios. Electronics 2024, 13(11), 2102.
4. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp 1126–1135.
5. Gong, D.; Liu, L.; Le, V.; Saha, B.; Mansour, M. R.; 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 (ICCV), Seoul, Korea, 27–28 October 2019; pp 1705–1714.
6. Jeong, J.; Yeo, D.; Roh, S.; Jo, Y.; Kim, M. Leak Event Diagnosis for Power Plants: Generative Anomaly Detection Using Prototypical Networks. Sensors 2024, 24(15), 4991.
7. Ji, C. Bilinear Resize. https://chao-ji.github.io/jekyll/update/2018/07/19/BilinearResize.html (accessed Jan 28, 2025).
8. Joshi, M.; Pant, D. R.; Heikkonen, J.; Kanth, R. One, Five, and Ten-Shot-Based Meta-Learning for Computationally Efficient Head Pose Estimation. Int. J. Embed. Real-Time Commun. Syst. 2023, 14(1), 1–24.
9. Li, A.; Qiu, C.; Kloft, M.; Smyth, P.; Rudolph, M.; Mandt, S. Zero-Shot Anomaly Detection via Batch Normalization. Adv. Neural Inf. Process. Syst. 2023, 36, 40963–40993.
10. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 10–17 October 2021; pp 10012–10022.
11. Meng, Y.; Lu, K. C.; Dong, Z.; Li, S.; Shao, C. Explainable Few-Shot Learning for Online Anomaly Detection in Ultrasonic Metal Welding with Varying Configurations. J. Manuf. Process. 2023, 107, 345–355.
12. Moon, J.; Noh, Y.; Jung, S.; Lee, J.; Hwang, E. Anomaly Detection Using a Model-Agnostic Meta-Learning-Based Variational Auto-Encoder for Facility Management. J. Build. Eng. 2023, 68, 106099.
13. Munkhdalai, T.; Yu, H. Meta Networks. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp 2554–2563.
14. Natha, S.; Leghari, M.; Rajput, M. A.; Zia, S. S.; Shabir, J. A Systematic Review of Anomaly Detection Using Machine and Deep Learning Techniques. Quaid-e-Awam Univ. Res. J. Eng. Sci. Technol. 2022, 20(1), 83–94.

15. Navarro, J. M.; Huet, A.; Rossi, D. Meta-Learning for Fast Model Recommendation in Unsupervised Multivariate Time Series Anomaly Detection. In Proceedings of the International Conference on Automated Machine Learning (AutoML), New Orleans, LA, USA, December 2023; pp 24-1.
16. Parnami, A.; Lee, M. Learning from Few Examples: A Summary of Approaches to Few-Shot Learning. arXiv 2022, arXiv:2203.04291.
17. Ravi, S.; Larochelle, H. Optimization as a Model for Few-Shot Learning. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, April 2017.
18. Reiss, T.; Hoshen, Y. Attribute-Based Representations for Accurate and Interpretable Video Anomaly Detection. arXiv 2022, arXiv:2212.00789.
19. Santoro, A.; Bartunov, S.; Botvinick, M.; Wierstra, D.; Lillicrap, T. Meta-Learning with Memory-Augmented Neural Networks. In Proceedings of the International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; pp 1842–1850.
20. Smith, J.; Doe, A.; Lee, M. Investigating the Impact of Meta-Learning on Anomaly Detection in Cybersecurity. J. Cybersecur. Res. 2020, 10(3), 145–160.
21. Soh, J. W.; Cho, S.; Cho, N. I. Meta-Transfer Learning for Zero-Shot Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp 3516–3525.
22. Wu, J. C.; Chen, D. J.; Fuh, C. S.; Liu, T. L. Learning Unsupervised MetaFormer for Anomaly Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 10–17 October 2021; pp 4369–4378.
23. Wu, P.; Pan, C.; Yan, Y.; Pang, G.; Wang, P.; Zhang, Y. Deep Learning for Video Anomaly Detection: A Review. arXiv 2024, arXiv:2409.05383.
24. Zhang, M.; Shen, Y.; Yin, J.; Lu, S.; Wang, X. ADAGENT: Anomaly Detection Agent with Multimodal Large Models in Adverse Environments. IEEE Access 2024.
25. Zhang, S.; Ye, F.; Wang, B.; Habetler, T. G. Few-Shot Bearing Anomaly Detection via Model-Agnostic Meta-Learning. In Proceedings of the 2020 23rd International Conference on Electrical Machines and Systems (ICEMS), Hamamatsu, Japan, 24–27 November 2020; pp 1341–1346.
26. Zhu, L.; Raj, A.; Wang, L. Advancing Anomaly Detection: An Adaptation Model and a New Dataset. arXiv 2024, arXiv:2402.
27. Zhu, L.; Wang, L.; Raj, A.; Gedeon, T.; Chen, C. Advancing Video Anomaly Detection: A Concise Review and a New Dataset. arXiv 2024, arXiv:2402.04857.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.