Gate-Conv SVDD: An Anomaly Detection Framework for Fault Inspection of Photovoltaic Panels Using UAVs

Lei Wang, *Student Member, IEEE*, Huaming Wu, *Senior Member, IEEE*, Yingfang Yu, Wei Yu, Shuo Chen, and Jun Wang

***Abstract*—Anomaly detection in solar photovoltaic panels using uncrewed aerial vehicles (UAVs) faces challenges due to minimal texture variations from surface anomalies (e.g., shadows, eddy currents) in UAV-captured imagery, which constrain both detection precision and real-time performance. Existing approaches often lack quantitative anomaly analysis that integrates aerial imagery with operational data, thereby limiting their practical value and impeding sustainable industry advancement. To address these limitations, we propose gate convolution support vector data description (Gate-conv SVDD), a novel framework that enhances the efficiency and accuracy (ACC) of anomaly detection through rapid parallel gated feature compression and hypersphere-based support vector data description (SVDD) clustering. This approach enables precise anomaly localization in high-resolution UAV imagery, as validated through simulations and controlled experimental datasets. Gate-conv SVDD further supports quantitative assessments of anomaly severity, thereby bridging the gap between image-based detection and actionable analysis. Designed for computational efficiency, the framework demonstrates strong potential for fast inference in support of real-time UAV imaging, subject to further hardware integration and field validation. Extensive experiments demonstrate that Gate-conv SVDD outperforms state-of-the-art methods, offering superior ACC and robustness in controlled settings.**

***Index Terms*—Anomaly detection, intelligent inspection, uncrewed aerial vehicle (UAV), urban infrastructure.**

I. Introduction

IN RECENT years, with rapid global economic growth, nations have prioritized sustainable clean energy, and the solar photovoltaic industry [1] has garnered significant attention due to its environmental benefits, high efficiency, and cost-effectiveness. New photovoltaic power stations are often established in remote areas to enhance efficiency and mitigate light pollution, yet they are susceptible to dust accumulation, debris obstruction, and component damage, leading to reduced efficiency and safety hazards [2]. Traditional manual inspections are constrained by weather conditions, high-altitude risks, and limited coverage, making them inadequate for these challenges [3]. Advances in uncrewed aerial vehicle (UAV) technology have driven the development of automated inspections, offering low cost, strong adaptability, and ease of operation, with the capability to rapidly cover large areas and transmit high-resolution images in real time. High-efficiency, high-precision intelligent analysis algorithms for UAV-captured images have become a research focus in photovoltaic panel anomaly detection, serving as a critical technology for ensuring the safe operation of photovoltaic systems. Developing an efficient and accurate anomaly detection framework for UAV-captured images is urgently needed to satisfy practical demands.

Recently, research on anomaly detection methods for solar photovoltaic panels using drone-captured imagery has surged, yet significant challenges persist. We analyze current methods and their associated challenges as follows.

1) *Diversity and Complexity of Datasets:* Photovoltaic panel performance degradation and anomalies manifest in various forms (e.g., hot spots and micro-cracks), influenced by environmental factors, such as illumination, temperature, and shading, resulting in noisy, highly nonlinear sensor data [4]. Moreover, datasets used for training and testing anomaly detection models are inherently complex and diverse, encompassing different photovoltaic panel types, tilt angles, flight conditions (vertical, horizontal, oblique), and weather conditions (e.g., sunlight reflection, shadows). Existing models, such as Region-Based convolutional neural network (CNN) for Anomaly Detection, combine multilevel CNN features to enhance sensitivity to small-scale anomalies. However, practical applications require consistent

Received 3 March 2025; revised 5 May 2025 and 7 July 2025; accepted 19 July 2025. Date of publication 22 July 2025; date of current version 25 September 2025. This work was supported in part by the Tianjin Intelligent Manufacturing Special Fund Project under Grant 20211093; in part by the National Natural Science Foundation of China under Grant 62071327 and Grant 62102262; and in part by the Tianjin Science and Technology Planning Project under Grant 22ZYYYJC00020. *(Corresponding author: Huaming Wu.)*

Lei Wang is with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China, and also with the Tianjin Shenglian Intelligent Technology Development Company Ltd., Tianjin 300480, China (e-mail: wanglei2019@tju.edu.cn).

Huaming Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China (e-mail: whming@tju.edu.cn).

Yingfang Yu is with the Sino-Singapore Tianjin Eco-City, Tianjin 300480, China (e-mail: yuyingfang@eco-city.gov.cn).

Wei Yu is with the College of International Business, Zhejiang Yuexiu University of Foreign Languages, Shaoxing 312030, China (e-mail: weiyu@zyufl.edu.cn).

Shuo Chen is with the Tianjin Shenglian Intelligent Technology Development Company Ltd., Tianjin 300480, China (e-mail: 18630631289@163.com).

Jun Wang is with the School of Mathematics, Tianjin University, Tianjin 300072, China (e-mail: jun.wang@tju.edu.cn).

Digital Object Identifier 10.1109/JIOT.2025.3591635

model performance across diverse scenarios, and current algorithms struggle to distinguish between degradation-related anomalies and normal fluctuations accurately.

2) *Processing Demands for High-Speed Flight Imagery:* Images captured by high-speed drones are relatively sparse, necessitating high-precision feature extraction and anomaly detection for subtle texture changes caused by surface anomalies (e.g., shadows, turbulence). Additionally, real-time requirements impose stringent demands on algorithm inference speed, computational efficiency, and operational consistency of drones. Existing CNN-based anomaly detection techniques [5] exhibit precision and real-time performance limitations. Furthermore, there is a lack of quantitative analysis for anomaly detection under real-world operating conditions, underscoring the urgent need for efficient processing techniques that balance accuracy (ACC) and speed.
3) *Bottlenecks Due to Insufficient Datasets and Resources:* The application of drone-captured imagery for photovoltaic panel anomaly detection is hindered by a scarcity of dataset resources, particularly the lack of publicly available annotated datasets [4]. This severely limits the training and validation of algorithmic models, resulting in poor generalization. Due to the absence of sufficiently large annotated public datasets, detecting complex photovoltaic panel anomalies still relies heavily on expert manual intervention to ensure ACC, significantly increasing detection costs and time.

Recent research to improve anomaly detection in photovoltaic panels has leveraged supervised learning models trained on labeled abnormal instances. For instance, a Transformer-based self-supervised model has enhanced image anomaly recognition ACC and efficiency [6], and a novel segmentation approach has effectively handled complex anatomical structures with minimal labeled data [7]. Supervised anomaly detection methods in the photovoltaic domain are hindered by scarce labeled abnormal samples, which limits their practical utility. Consequently, unsupervised anomaly detection methods require less labeled data and are gaining traction. Such as Siamese networks [8] differentiate features between normal and synthetic defective samples for anomaly identification, while local and global network branches [9] address structural and logical surface defects. Graph neural networks and attention mechanisms [10], [11] further integrate logical constraints into unsupervised frameworks. For UAV-based photovoltaic panel anomaly detection [12], existing unsupervised methods often demonstrate poor classification ACC and limited generalization capability, making them unsuitable for precise solar panel fault detection.

Anomaly detection in solar photovoltaic panels using UAVs is crucial for ensuring operational reliability, but it is challenged by subtle surface anomalies and the requirement for real-time processing. Existing methods struggle with precision, computational efficiency, and quantitative analysis of anomalies under operational conditions. To address these issues, we introduce gate convolution support vector data description (Gate-conv SVDD), a framework combining gated convolutions with support vector data description (SVDD) for precise and efficient anomaly detection. The approach systematically tackles the challenges of anomaly analysis in UAV-captured imagery. The main contributions of this article are as follows.

1) The proposed method employs gate convolution (Gate-conv) Blocks for robust feature extraction and integrates SVDD for hypersphere-based feature clustering, achieving precise anomaly detection while enabling quantitative assessment of anomaly severity.
2) This study develops a parallel gated convolutional feature extraction component addressed for diverse, sparse imagery captured by UAVs at high altitudes. The component integrates Gated Convolution for adaptive high-fidelity feature extraction, Causal Convolution to preserve temporal dependencies through strict input-output causality, and Dilated Convolution with adjustable rates for multiscale feature aggregation. This parallel architecture not only enables lightweight and efficient feature extraction but also achieves high-efficiency feature extraction for complex, diverse image data.
3) To accelerate progress in UAV-based defect detection for photovoltaic panels and address model ACC and generalization deficiencies due to limited datasets and resources, we publicly release the *Battery-SL* dataset. This significant contribution expands available resources and promotes reproducible research in this critical field.
4) Extensive experiments confirm that Gate-conv SVDD outperforms state-of-the-art methods in real-world photovoltaic anomaly detection while maintaining strong generalization. Real-time validation further demonstrates its efficient live data processing capability, verifying practical applicability.

Our approach has been validated through simulations and controlled experiments using UAV-captured imagery, demonstrating superior ACC and potential for real-time applications. However, further hardware tests and field validations are needed to confirm its performance in operational UAV deployments. This article presents the methodology, experimental results, and future directions for enhancing the framework's real-world applicability. The remainder of the article is organized as follows. We review the related work in Section II. Section III describes the proposed Gate-conv SVDD in detail. Section IV presents the experimental results and their analysis. Finally, Section V concludes this article.

II. Related Work

Current mainstream solutions for abnormal detection in photovoltaic panels are primarily focused on image-based detection techniques. These techniques mostly utilize deep neural network models to identify abnormal images or areas deviating from normal image structures.

This is typically accomplished through the application of unsupervised or semi-supervised learning methodologies. These models can be broadly classified into four distinct categories based on their algorithmic design principles.

Specifically, they are engineered to detect anomalous images or localized abnormal regions that diverge from the expected normative image structure, relying on unsupervised or semi-supervised learning frameworks [13]. According to their conceptual foundations, they can be systematically grouped into the following four types.

A. Feature Distance Measure-Based Methods

The traditional SVDD model [14] incorporates deep learning techniques to preserve data structure for unsupervised anomaly detection. Their deep SVDD framework leverages a neural network to learn feature representations end-to-end, reducing the reliance on manual feature engineering. However, it struggles with handling high-dimensional feature spaces, and complex nonlinear data distributions, and lacks mechanisms to focus on critical information, limiting its robustness in detecting subtle anomalies.

Additionally, it faces challenges related to the curse of dimensionality and computational inefficiencies when scaling to large datasets. Similarly, Fan et al. [15] proposed a semi-supervised video anomaly detection method using a Gaussian mixture fully convolutional variational autoencoder (GM-FCVAE), which is efficient when labeled abnormal data is scarce. However, the model's reliance on the assumption that anomalies are Gaussian outliers limits its ability to capture subtle or dynamic anomalies, particularly in high-dimensional data. Hassanat et al. [16] introduced a novel outlier-robust ACC measure for regression models based on the Hassanat distance metric, which is invariant to outliers and offers an easily interpretable ACC-like value. While this approach demonstrates robustness across multiple benchmarks, it is focused on regression models, making it less applicable to anomaly detection tasks or complex, high-speed data.

B. Feature-Surface Construction for Detection Methods

A deep learning-based framework for detecting out-of-distribution (OOD) [17] and adversarial samples in videos has been developed using unsupervised learning, where a neural network is trained solely on normal data to identify anomalies by comparing new frames to learned patterns [18]. This mitigates data imbalance by not requiring abnormal samples, but its dependence on a single normal distribution limits the detection of complex or multiscale anomalies, risking false positives from normal variations. An energy-based method enhances reliability in safety-critical applications using energy scores, which align more closely with input probability density than softmax confidence, yet it focuses narrowly on anomalies, struggling with subtle or high-speed image cases and lacking the ability to quantitatively analyze anomaly severity.

A spatiotemporal transform network-based framework [19] is proposed for anomaly detection and localization in distributed parameter systems, such as lithium-ion battery short circuits, using nonlinear space-time separation, Gaussian process regression, and combined statistics. Though effective for its target application, its complex models struggle to generalize across domains or process high-dimensional image data. PatchCore [20], on the other hand, uses a pretrained CNN to build a patch-based feature memory, achieving high area under the curve (AUC) and localization ACC on datasets like MVTec AD with only normal samples, yet it is limited by sensitivity to positional changes, dependence on the pretrained model, and moderate computational demands.

C. Image Reconstruction-Based Methods

A road anomaly detection method [21] for autonomous vehicles employs partial image reconstruction coupled with semantic segmentation to identify anomalies by detecting reconstruction failures in regions outside known categories, such as roads, in an unsupervised setting. Although effective for distinguishing unknown anomalies, it struggles with subtle anomalies and complex, multiscale patterns, and its reliance on reconstruction may hinder performance in dynamic or cluttered environments.

Similarly, a self-supervised reconstruction-by-inpainting (RIAD) method [22] trains an auto-encoder on normal images and detects anomalies by reconstructing partial images through random inpainting. While promising, RIAD has limitations with large, complex anomalies and scenarios where inpainting cannot adequately represent missing parts, assuming anomalies are sufficiently distinct for inpainting to fail. The dual deep reconstruction networks-based image decomposition (DDR-ID) approach decomposes unseen images [23] into normal class and residual components to calculate two anomaly scores. Despite strong performance on benchmarks, DDR-ID's reliance on reconstruction errors and lack of normal-class-specific information limits its ability to detect subtle anomalies in high-speed datasets.

D. Hybrid Approaches Combining Traditional Methods With Deep Neural Networks

A range of sophisticated approaches has been devised to address anomaly detection across various domains, effectively integrating traditional methodologies with state-of-the-art deep neural networks. One such method leverages self-supervised learning to identify defects in high-resolution images, constructing models solely from normal data and extracting feature representations through carefully designed proxy tasks [24]. This semi-supervised technique demonstrates particular efficacy in scenarios with imbalanced datasets or scarce anomaly instances. Nevertheless, its performance may diminish when subtle irregularities elude detection due to insufficiently representative proxy tasks, potentially undermining its precision in complex, real-world applications.

A GAN-based technique [25] identifies anomalies in high-dimensional images by seeking effective latent representations, flagging those without as anomalous and delivering strong benchmark results, though it grapples with training instability, misses subtle anomalies in complex data, and demands high computational effort and tuning. In contrast, AdaCLIP [26], a zero-shot approach, adapts the pretrained CLIP model with hybrid static and dynamic prompts to enhance generalization, excelling across industrial and medical datasets without task-specific training, yet it relies on CLIP's quality, faces

computational costs from prompt generation, and struggles with noisy or highly variable data.

In the realm of real-time crowd anomaly detection, a comprehensive survey highlights deep learning methods integrated with the Web of Things (WoT) platform for secure video surveillance [27]. It categorizes approaches into tracking, feature-based, deep learning, and hybrid methods, with the latter two showing superior performance. Yet, issues like pixel occlusion and computational demands limit their adaptability beyond crowded scenes to broader, high-dimensional anomaly tasks.

Focusing on sustainable healthcare systems, a hybrid quantum-classical optimization method enhances energy efficiency in AIoT networks using deep learning and edge computing [28]. While achieving high ACC and low carbon emissions, its specificity to healthcare and energy optimization restricts its relevance to domains like image anomaly detection.

Similarly, an interference hypergraph-based algorithm optimizes energy harvesting for UAV-assisted TinyML in IoT networks, improving throughput and UAV lifespan. However, its emphasis on resource allocation in communication systems limits its applicability to image-based anomaly detection challenges.

To meet the specific demands of quantitatively analyzing anomalies in high-speed UAV-captured imagery and overcome the shortcomings of existing methods, we propose the Gated-conv SVDD framework. This approach integrates traditional techniques with deep learning, leveraging their complementary strengths to deliver a rapid and precise anomaly detection solution tailored to real-world UAV applications.

TABLE I
SYMBOLS AND THEIR MEANINGS

Symbol	Meaning
D	Photovoltaic panel image dataset
n	Sample size
P_{ij}	Pixel values of column j in row i of the photovoltaic panel image before whitening
x_{ij}	Pixel value of column j, row i, of the photovoltaic panel image after whitening
μ	Mean
σ	Variance
x_i	Image tensor data of photovoltaic panels
C	Image channel
H	Image height
W	Image width
X	Sample space
F	Eigenspace
$f(\cdot)$	Overall model network
θ	Overall model parameters
$L(x,c,l,g)$	SSD target loss
$L_{loc}(x,l,g)$	SSD positioning loss
$L_{conf}(x,c)$	SSD classification loss
N	Default number of boxes that the label box matches
α	Balancing weighting parameters for positioning loss and confidence loss
$x^{(k-1)}$	Feature tensor data passed through layer $k-1$ in the layer k network
$y^{(k)}$	Feature tensor data output by the k-layer network
$f_{Filter}(\cdot)$	Filter-Inception network
$f_{gate}(\cdot)$	Gate-Inception network
$\theta_{1\times3,3\times1}$	Stacked $1\times3, 3\times1$ convolutional layer parameters
$\theta_{1\times1}$	1×1 convolutional layer parameters
$\Phi(\cdot)$	Feature mapping function of the model
W	Parameters for each layer of the network
c	Center of the hypersphere
R	Radius of the hypersphere
λ	Network weight regularization coefficient
$score$	Anomaly score

III. METHODOLOGY

A. Problem Definition

The objective of image anomaly detection for photovoltaic panels is to evaluate the presence of abnormalities within a given photovoltaic panel image, quantify the degree of abnormality, and provide actionable insights into their severity and impact on operational performance. This process aims to enhance maintenance efficiency and ensure the reliability of solar energy systems by identifying and assessing defects, such as dust accumulation, component damage, or other irregularities in UAV-captured imagery. This process can be expressed mathematically as follows:

$$\text{score} = f(\mathscr{D}(x_i);\theta) \in [0,1] \tag{1}$$

where $D = \{x_1, x_2, \ldots, x_n\}$, and $x_i \in R^{C\times H\times W}$ is the image tensor data of the photovoltaic panel. Then, anomaly detection for the photovoltaic panel is defined as the mapping function: $score = f(x_i;\theta) \in [0,1]$. Through this definition, it can be seen that the design of an effective and reasonable neural network $f(\cdot)$ is the key to abnormality detection. It is also the research priority of this paper. The symbols used in this article are described in Table I.

B. Proposed Framework

This article presents an anomaly detection framework based on Gate-conv SVDD, the structure of which is illustrated in Fig. 1. The framework first applies the single shot multibox detector (SSD) algorithm [29] to a series of preprocessing operations on aerial images of photovoltaic power generation systems collected by UAVs, including image cropping, anchor frame positioning, perspective correction, etc., to generate the sample dataset for training. Subsequently, it employs gated convolutional modules (Gate-conv) for feature extraction, followed by SVDD for hypersphere-based anomaly clustering. The Gate-conv mechanism enhances feature extraction by modulating hierarchical features, while SVDD constructs a compact hypersphere to distinguish normal and anomalous samples efficiently.

SVDD was selected as the primary module for extracting anomaly features due to its enhanced suitability for UAV-based anomaly detection. In contrast to alternative deep learning-based one-class methods like Deep One-Class (DeepOC) or autoencoder-based approaches, SVDD constructs a hypersphere that encloses normal data in a high-dimensional feature space, minimizing the radius while maximizing the separation of anomalies. This approach excels in photovoltaic panel inspection, where normal samples dominate and anomalies are rare, demonstrating superior robustness.

Compared to DeepOC, which requires extensive training data to learn a robust decision boundary, SVDD is less data-intensive, making it suitable for scenarios with limited labeled anomalies. Autoencoder-based methods, while effective for reconstruction-based anomaly detection, often incur higher

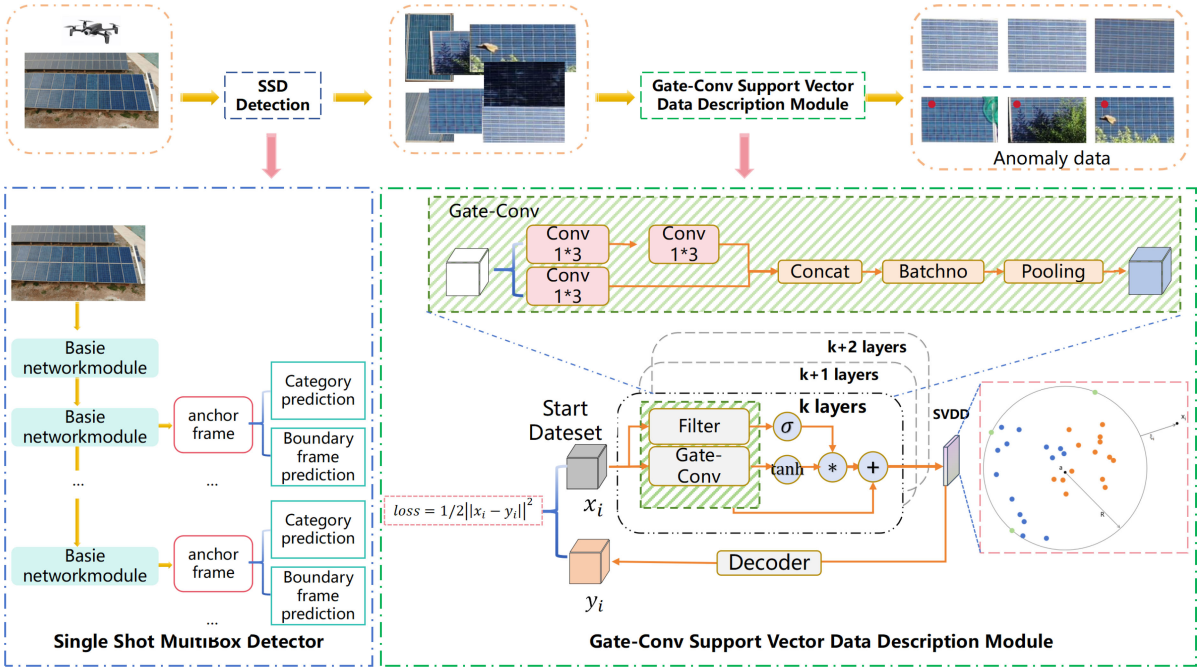

Fig. 1. Anomaly detection framework based on Gate-conv SVDD. This framework preprocesses photovoltaic images collected by UAVs using the SSD algorithm, then extracts features using Gate-conv modules, and performs anomaly analysis within the hypersphere constructed by SVDD, including two key steps: sample generation and anomaly detection.

computational costs due to their encoder-decoder architecture, which is less feasible for real-time UAV applications. SVDD's computational efficiency, with a complexity of $O(n^2)$ for training and $O(n)$ for inference, aligns with the real-time requirements of high-speed UAV imaging. Extensive experiments (Section IV) demonstrate that SVDD, when combined with Gate-conv, outperforms these alternatives in ACC and speed on the Battery-SL10000 dataset.

C. Anomaly Sample Dataset of Photovoltaic Panel

1) Sample Dataset Generation: The advancement of deep learning technology has promoted the widespread application of intelligent inspection in photovoltaic power plants. UAVs have shown significant advantages in power inspection due to their low cost, small size, high adaptability, and ease of operation. Equipped with multiple sensors, they can capture high-resolution images in a short period, which is crucial for identifying anomalies in photovoltaic systems. The flexibility of UAVs also allows them to cover areas that are difficult to reach with traditional methods, reducing costs and improving inspection efficiency. In the sample generation phase of this study, we used UAV technology to collect image data of photovoltaic panels. With this technology, we can quickly and comprehensively collect critical visual information on photovoltaic power generation systems. To further optimize the quality and usability of the data, we input the collected image data into the SSD algorithm for preprocessing. This step involves image cropping, anchoring, perspective correction, etc., aiming to enhance the ACC and consistency of the image data, laying a solid foundation for subsequent anomaly detection analysis.

Fig. 2. Normal samples in the Battery-SL dataset.

Fig. 3. Anomaly samples in the Battery-SL dataset.

We manipulate the UAVs to fly at low altitudes of 20, 10, and 5 km/h with a uniform speed to images shot the photovoltaic power generation systems in a certain area, and sample the collected image data evenly in time without repeatedly capturing the overall image of the photovoltaic panel. Then, the edge data of each photovoltaic panel is labeled. SSD is used in multiscale target detection for image data. As shown in Figs. 2 and 3, the frontal image of each photovoltaic panel is obtained after a perspective correction, and is stored in the dataset. The SSD objective function is expressed as follows:

TABLE II
BATTERY-SL DATASETS

Data Set	Training datasets		Test dataset	
	Normal	Abnormal	Normal	Abnormal
Battery-SL2000	2000	–	500	500
Battery-SL5000	5000	–	500	500
Battery-SL10000	10000	–	1000	1000

$$L(x, c, l, g) = \frac{1}{N}(L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \tag{2}$$

where $L_{\text{conf}}(x, c)$ and $L_{\text{loc}}(x, l, g)$ are the classification loss and positioning loss of SSD.

2) Whitening Treatment: In the field of image detection, environmental factors, such as lighting intensity, object reflectance, and camera characteristics may all interfere with the detection results. In order to extract stable information from images that are not affected by these external factors, image whitening processing is particularly important [30]. Image whitening can be used to process overexposed or underexposed pictures by changing the average pixel value of the image to 0 and changing the square deviation of the image to unit variance 1. First, the mean and variance of the original image are calculated, and then each pixel value of the original image is transformed. Suppose the image P has I rows and J columns, namely, the value of each pixel is P_{ij}. The formula is expressed as follows:

$$\mu = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} P_{ij}}{IJ} \tag{3}$$

$$\sigma = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} \left(P_{ij} - \mu\right)^2}{IJ} \tag{4}$$

$$x_{ij} = \frac{p_{ij} - \mu}{\sigma} \tag{5}$$

where μ is the mean value, σ is the variance and X_{ij} is the pixel value of column j and row i. After whitening stabilizes image features, the Gate-conv SVDD model leverages these processed inputs for robust anomaly detection in photovoltaic panels.

3) Battery-SL Dataset: The Battery-SL dataset we constructed is divided into three different scales, as detailed in Table II. The dataset contains images with a wide range of resolutions, ranging from lower resolutions of 64 × 32 pixels to higher resolutions of 1280 × 640 pixels. These images mainly focus on the detection of photovoltaic panels, covering both normal and abnormal states. The abnormal categories specifically include phenomena, such as cracking, blistering, foreign object intrusion, and blade obstruction on photovoltaic panels, all based on the classification criteria from previous studies [31].

In this study, we particularly emphasize the detection of abnormal obstruction, with solar panels as the primary test object. The experimental dataset used in this study is characterized by three key dimensions: the number of 1) image channels; 2) image width; and 3) image height.[1] The comprehensive consideration of these dimensions provides rich feature information for the abnormal detection of photovoltaic panels, contributing to the improvement of detection algorithm performance and ACC. Through the meticulously designed and constructed Battery-SL dataset, we aim to provide high-quality data support for the intelligent inspection and abnormal detection of photovoltaic panels, thereby promoting research progress in this field.

[1]https://github.com/Wanllence-tju/Battery-dataset.

D. Core Module of Anomaly Detection

1) Gated Convolution Network: To capture the dependencies between pixels with relatively long spans and avoid the problems of gradient disappearance and gradient explosion, the proposed model introduces a Gate-conv module. This module is designed to selectively propagate critical features through a bidirectional gating mechanism, filtering out key information and suppressing invalid signals. As shown in Fig. 1, for an input feature map $x \in \mathbb{R}^{C\times H\times W}$, the Gate-conv operation is defined as

$$z = \sigma\left(W_g * x\right) \cdot \left(W_f * x\right) \tag{6}$$

where W_g and W_f are convolutional kernels for the gating and feature transformation branches, respectively, σ is the sigmoid activation, and $*$ denotes convolution. The gating branch $\sigma(W_g * x)$ generates a spatial attention map that modulates the feature branch $W_f * x$, emphasizing anomaly-relevant features. Unlike attention modules like CBAM, which focus on global context modeling, Gate-conv operates within convolutional layers to hierarchically refine local features, reducing computational overhead while preserving critical details for anomaly detection.

In detail, the Gated-conv module leverages gated convolution to process features extracted from high-speed UAV imagery, effectively minimizing errors and improving precision. To capture temporal dependencies within the image sequences, causal convolution is employed, allowing the model to analyze the sequential relationships between past and future frames. Furthermore, dilated convolution is incorporated to compress feature dimensions while enabling multiscale temporal feature extraction. To facilitate efficient training and enhance convergence speed, residual connections are introduced, with dimension-matching shortcuts ensuring compatibility with the depth of subsequent residual layers, as illustrated

$$y^{(k)} = \tanh\left(f_{\text{Filter}}(x^{k-1})\right) * \sigma\left(f_{\text{Gate}}(x^{k-1})\right) + x^{k-1} \tag{7}$$

$$x^{(t)} = \tanh\left(w_t * x^{(t-1)}\right) \tag{8}$$

$$RF_{n+1} = RF_n + (k + (k-1)(d-1)\ -1)S \tag{9}$$

where f_{Filter} is the filter-inception network layer, and f_{Gate} is the gate-inception network layer. Furthermore, $x^{(t)}$ is the output at time step t, w, and b are convolution and bias parameters, RF is the receptive field size, k is the kernel size, d is the dilation rate, and S is the stride.

The Gated-conv module, as defined in (7), adaptively regulates the flow of information through a set of gating mechanisms. It effectively captures fine-grained dependencies, extracts image features with high efficiency, and enhances

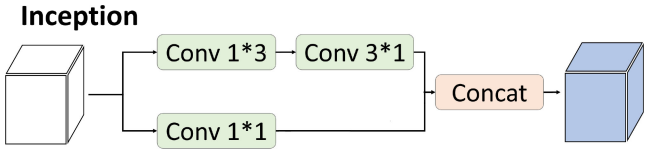

Fig. 4. Schematic representation of the Inception structure.

the model's representational capacity while suppressing noise and irrelevant signals for rapid and accurate feature extraction. Specifically, the Gated-conv module processes input image features through multiple parallel gated convolutional branches, where each branch employs distinct convolutional filters and gating units. This design enables the compression of high-dimensional representations into a lower-dimensional space while retaining anomaly-relevant information, a process referred to as *parallel gated feature compression*. Mathematically, the output across M parallel branches is given by

$$z_{\text{out}} = \sum_{m=1}^{M} \sigma\left(W_{g,m} * x\right) \cdot \left(W_{f,m} * x\right) \tag{10}$$

where $W_{g,m}$ and $W_{f,m}$ are the gating and feature kernels for the mth branch. This parallel structure enhances robustness by capturing diverse feature representations, which are then fused via dot-product concatenation for SVDD clustering.

Subsequently, causal Convolution shown in (8) ensures outputs depend solely on current and past inputs, capturing high-order dependencies by extracting contextual information from image data features. Dilated Convolution presented in (9) expands the receptive field in the temporal dimension by systematically adjusting the dilation rate d, enabling rapid expansion in a multilayer causal network, performing progressive information compression, and extracting multiscale features to capture global temporal information across all frequencies without increasing parameters or computational costs.

2) Inception Structure: We adopted the optimized Inception structure [32] to extract high-dimensional features. Considering the potential issues of parameter expansion and overfitting caused by deep convolutions, we selectively retained the 3×3 and 1×1 convolutional units that are most sensitive to photovoltaic panel anomaly detection. This strategy not only reduces the model's parameter count but also improves computational efficiency.

As shown in Fig. 4, we simulated the biological neural system through sparse connections, effectively expanding the network's perceptual capabilities. The multipath design of the Inception structure allows the network to extract features at different scales, enriching feature representation and enhancing the ability to capture details. With this design, we can achieve precise detection of anomalies in photovoltaic panels while keeping the model concise. For each Inception, the model proposed in this article copies the input in two copies for different path processing, one branch is 1×3 convolution and 3×1 convolution stacked, and the other branch is 1×1 convolution. Finally, the visual information is aggregated and output at different sizes, and the structure enables the Gate-conv SVDD to make more efficient use of parameters. The formula is expressed as follows:

$$f_{\text{Filter}}\left(x^{(k-1)}\right) = f\left(x^{(k-1)}; \theta_{1\times3,3\times1}\right) \oplus f\left(x^{(k-1)}; \theta_{1\times1}\right) \tag{11}$$

$$f_{\text{Gate}}\left(x^{(k-1)}\right) = f\left(x^{(k-1)}; \theta_{1\times3,3\times1}\right) \oplus f\left(x^{(k-1)}; \theta_{1\times1}\right) \tag{12}$$

where $\theta_{1\times3,3\times1}$ is 1×3, 3×1 convolutional layer parameters, and $\theta_{1\times1}$ is 1×1 convolutional layer parameters.

3) Normalized Output: As network depth increases, data feature distributions often saturate nonlinear activation boundaries, causing vanishing gradients and slowing convergence. This study employs batch normalization (BN) to normalize neuron inputs to a standard distribution (mean 0, variance 1) [33], mitigating vanishing gradients, enhancing convergence, and stabilizing training, while its decorrelation property [34] reduces interdata dependencies for better classification. BN is thus applied within gated convolutional layers, fusing shallow and deep features via skip connections, as shown

$$x^{(k)} = \text{batchnorm}\left(x^{(k)}\right) \tag{13}$$

$$y = \text{concat}\left[\text{conv}\left(x^{(1)}, \ldots x^{(k)}\right)\right]. \tag{14}$$

4) Feature Dimension Reduction: Excessive feature dimensions during model training can trigger the curse of dimensionality, leading to overfitting, while reducing dimensions may sparsify data, impairing generalization. To counter this, we use max-pooling to lower dimensionality and noise while preserving key information, followed by a 3×3 convolutional kernel to broaden the receptive field and capture diverse features, with ReLU activation enhancing dimensionality reduction efficacy.

5) Hypersphere Model: During the training initialization phase, the model is linked to a decoder comprising multiple layers of deconvolution and upsampling, with training initialization achieved by preprocessing the features of input image data. This step aims to reconstruct an image matching the input size while minimizing information loss, enabling the identification of the centroid of normal sample features in a high-dimensional space.

In the subsequent formal training phase, this pretrained centroid is utilized to perform high-dimensional compression on the hyperspherical model, progressively reducing the hypersphere's volume. This process encourages the aggregation of similar features while distancing anomalous features from the cluster center, thereby achieving efficient and rapid anomaly detection.

The detection framework proposed in this article is based on the SVDD model, as shown in Fig. 5. The feature representation function $\Phi(\cdot)$ maps image data sample features $x \in D$ to a high-dimensional feature space. The goal of the detection module is to construct a feature hypersphere with minimal volume in the high-dimensional feature space, defined by its center c and radius R, enveloping the high-dimensional features mapped from the image data and iteratively reducing the vector radius to continuously compress the hypersphere's volume.

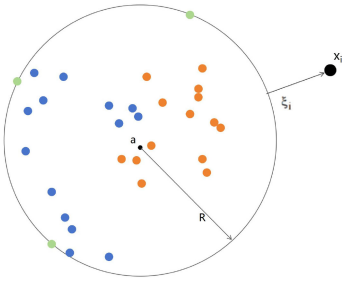

Fig. 5. Schematic of SVDD-based hyperspherical abnormal detection.

In unsupervised anomaly detection, where training data predominantly consists of normal samples, the focus lies on minimizing the distance from sample points to the hypersphere's center c, rather than the radius R, thereby compressing the hypersphere's volume to cluster high-dimensional image data features toward the center. For a given set of training sample points, $D = \{x_1, \ldots, x_n\}$, the defined objective function is expressed as follows:

$$\min_W = \frac{1}{n}\sum_{i=1}^{n} \|\Phi(x_i, W) - c\|^2 + \frac{\lambda}{2}\sum_{l=1}^{L} \|W^l\|_F^2 \quad (15)$$

$$\text{s.t. } \|\Phi_k(x_i, W) - c\|^2 \le R^2 + \xi_i,\ \xi_i \ge 0 \quad \forall i. \quad (16)$$

where $\Phi(\cdot)$ maps input samples x_i to a high-dimensional feature space using network weights W. The objective minimizes the average squared distance from mapped features $\Phi(x_i, W)$ to the hypersphere center c (with n samples), while $(\lambda/2)\sum_{l=1}^{L} \|W^l\|_F^2$ regularizes the weights W^l across L layers using the Frobenius norm and hyperparameter λ. The constraint limits the distance of each feature $\Phi_k(x_i, W)$ from c to the radius R^2, with slack variables $\xi_i \ge 0$ for flexibility.

For an image data test sample, the anomaly score is defined as the distance from the sample point to the hyperspherical center in the output space. After clustering high-dimensional image data features near the center, points whose high-dimensional features deviate significantly from the compressed hypersphere's center are identified as anomalous data features. Furthermore, the high-dimensional distance of these outlier features from the hypersphere center serves as a quantitative metric of their anomaly severity. This metric is subsequently used to assess the extent of anomalies, enabling the evaluation of the severity of irregular information in UAV-captured photovoltaic panel imagery.

E. Complexity Analysis

To evaluate the Gate-conv SVDD framework's suitability for real-time UAV anomaly detection, we analyze its computational complexity across key components: SSD preprocessing, Gate-conv feature extraction, and SVDD clustering.

1) SSD Preprocessing: The SSD algorithm handles image cropping, anchoring, and perspective correction. For an input image $x_i \in \mathbb{R}^{C\times H\times W}$, SSD uses a CNN backbone (e.g., VGG-16) with complexity $O(C \cdot H \cdot W \cdot F)$, where F is the filter count, and detection layers with complexity $O(B \cdot S \cdot (H \cdot W))$. With $C = 3$, $H = 128$, $W = 64$, $F = 512$, $B \cdot S \approx 10^4$, the complexity is $O(H \cdot W \cdot F) \approx O(10^6)$ per image, efficient for high-resolution UAV images.

2) Gate-Conv Feature Extraction: The Gate-conv module uses M parallel branches, each with a convolution of complexity $O(C \cdot H \cdot W \cdot k^2 \cdot C_{\text{out}})$. For M branches and K layers, total complexity is $O(M \cdot K \cdot C \cdot H \cdot W \cdot k^2 \cdot C_{\text{out}})$. Using $M = 4$, $K = 4$, $C = 3$, $C_{\text{out}} = 64$, $H = 128$, $W = 64$, $k = 3$, the complexity is $O(4 \cdot 4 \cdot 3 \cdot 128 \cdot 64 \cdot 9 \cdot 64) \approx O(10^6)$ per image, optimized by GPU parallelism (Section IV).

3) SVDD Clustering: SVDD constructs a hypersphere for anomaly detection. For a dataset $D = \{x_1, x_2, \ldots, x_n\}$, training complexity is $O(n^2)$, and inference complexity is $O(n \cdot d)$, where d is the feature dimension ($d = 64$). For Battery-SL10000 ($n = 10000$), training is $O(10^6)$, inference is $O(10^4 \cdot 64) \approx O(10^6)$ per image, supporting real-time use.

4) Overall Complexity: Gate-conv dominates with $O(10^6)$ per image, followed by SSD and SVDD at $O(10^6)$. Training ($O(n^2)$) is offline, while inference ($O(10^8)$) supports high-speed UAV imaging, as shown in Section IV, with potential for hardware optimization.

This analysis confirms Gate-conv SVDD's efficiency for real-time UAV anomaly detection, aligned with Section IV performance metrics.

F. Model Optimization Algorithm

We have improved the SVDD model by employing bidirectional diffusion convolution to perceive data features in both forward and backward directions, implementing an information expansion perception strategy that enhances computational efficiency and ACC. The parallel gating mechanism enhances feature precision by selectively filtering key information, followed by dilated convolution to compress features and expand the receptive field, refining feature representation across diverse tasks.

Thanks to the minimal computational overhead introduced by the gating structure, the optimized convolutional parameters efficiently manage model complexity, enabling the Gated-conv SVDD model to achieve both high ACC and fast processing speed. This makes it particularly well-suited for large-scale datasets and real-time applications, such as anomaly detection in UAV-captured photovoltaic panel imagery. By compressing high-dimensional feature vectors into a hypersphere, the model clusters normal image features near the center while effectively isolating anomalous features. This compact representation enhances the model's ability to identify and distinguish anomalies with greater precision and efficiency.

IV. Experimental Results and Analysis

A. Experimental Setup

The experiments in this study were conducted on a Linux server equipped with a GeForce RTX 2080 Ti graphics card, which provides 12 GB of graphics memory. To ensure rapid convergence, the Adam optimizer was used during training. The proposed Gate-conv SVDD model efficiently extracts image features through gated convolutions and constructs a hypersphere in high-dimensional feature space for anomaly detection.

Algorithm 1 Gate-Conv SVDD

Require: Training dataset $D = \{x_1, x_2, \ldots, x_n\}, x_i \in R^{C\times H\times W}$, maximum number of iterations T, number of convolutional network layers K

Ensure: Optimal model parameters

procedure GATE-CONV-SVDD(D, T, K)
  Unified image data for: $x_i \in R^{3\times 128\times 64}$
  Initialize model parameters
  Train encoders, decoders
  for $t \leftarrow 1$ **to** T **do**
    for $k \leftarrow 1$ **to** K **do**
      $Z = encoder(x)$
      $Z = encoder(x)$
      $loss = \frac{(y-x)^2}{n}$
    end for
  end for
  Set the feature center
  Save the pre-trained model
  for $t \leftarrow 1$ **to** $\frac{T}{2}$ **do**
    for $k \leftarrow 1$ **to** K **do**
      $Z = encoder(x)$
      Use Eq. (2) to get the loss
    end for
  end for
end procedure

1) Data Capture Conditions: The data for the experiments were captured using a drone equipped with a high-resolution camera. The drone was flown at an average altitude of 50 m above the photovoltaic panels, ensuring a clear and detailed view of the panel surfaces. The drone's speed during image capture was approximately 5–20 m per second, which allowed for consistent image coverage without introducing motion blur. The photovoltaic panel data were collected via drone and will be made publicly available as open-source data for further research by other scholars. The dataset includes three sample sets of 2000, 5000, and 10 000 images, corresponding to drone speeds of 20, 10, and 5 m/s, respectively, as previously described.

The images were collected under clear and cloudy weather conditions to simulate real-world scenarios. The experiments were conducted on days with moderate winds (below 15 km/h), and temperatures ranging from 15 °C to 30 °C. These conditions provided a suitable environment for image capture, minimizing potential interference from extreme weather while ensuring that typical operational conditions for UAV-based inspections were represented. These conditions ensured clear and detailed imagery without motion blur. The experiments processed simulated UAV-captured data on the server, and no hardware-specific tests (e.g., on embedded UAV systems) were conducted in this study. Future work will focus on integrating the framework with UAV hardware and validating its performance in real-world settings.

2) Evaluation Methodology: The performance of the Gate-conv SVDD model was evaluated from five key perspectives: baseline comparison, ablation and component study, parameter sensitivity, flight speed sensitivity, and computational scale and efficiency. These evaluations comprehensively demonstrate the effectiveness of the proposed model. The five baseline models used for comparison are as follows.

1) *Deep SVDD:* This model is used for image-level anomaly detection. A neural network extracts features from the input data, and during training, normal samples are contracted within a minimized hypersphere in the feature space. Abnormal samples are placed farther from the hypersphere. The model determines whether a test sample is anomalous by measuring the distance between the sample point and the center of the hypersphere.
2) *CNN Autoencoder [35]:* This model is based on an unsupervised learning approach where the goal is to reconstruct the input data without relying on labeled samples. The architecture combines multiple layers of convolutions with encoder-decoder structures to learn a compact representation of the input and reconstruct it as accurately as possible.
3) *Region CNN [4]:* This model utilizes region-based CNNs to detect anomalies in photovoltaic panels. The approach focuses on identifying localized defects or irregularities within specific regions of the solar panels.
4) *PatchCore [20]:* PatchCore, which leverages pretrained CNNs to create a feature patch memory bank, achieving high AUROC and localization ACC on benchmarks like MVTec AD using only normal data, yet it faces challenges with positional variations, pretrained model dependency, and moderate computational demands that constrain its broader applicability.
5) *AdaCLIP [26]:* AdaCLIP adapts the pretrained CLIP model with hybrid learnable prompts for zero-shot anomaly detection, excelling across diverse domains like industrial and medical datasets. However, it depends on CLIP's quality, incurs computational costs from prompt generation, and struggles with highly variable or noisy data.

B. Evaluating Metrics

In the context of anomaly detection, the primary focus is on the model's ability to effectively distinguish between normal and abnormal samples, framing it as a classification task. The model clusters feature information through hypersphere compression, and the anomaly confidence score for each sample is defined based on its deviation from the cluster center, with outliers assigned higher scores according to their distance from the center. Samples with relatively low confidence scores are considered anomalous.

To objectively and comprehensively measure the performance of the proposed method, we select two commonly used evaluation indicators in image detection tasks: the ACC [31] and the AUC [36]. These metrics are defined as follows.

1) Definition of Accuracy: Considering the true normal sample as a negative case and treating anomalies as positive cases, the evaluation metrics include the following classifications: a successfully detected anomaly is labeled as a true

positive case (TP), correctly identifying a normal sample is a true negative case (TN), a normal sample incorrectly classified as abnormal is a false-positive case (FP), and an abnormal sample incorrectly classified as normal is a false-negative case (FN) [37]. The calculation formula is as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{17}$$

2) Definition of AUC (Area Under the Receiver Operating Characteristic Curve): Based on the results of the anomaly score for the predicted data, we rank the samples and regard those samples whose anomaly score is less than or equal to the sample selected as positive examples to do predictions. A series of true case rates (TPR) and FP rates (FPR) are generated, from which we can plot the receiver operating characteristic (ROC) curve. This curve is an analysis tool for the classification effect of binary classification models and is suitable for anomaly detection tasks. The formulas for TPR and FPR are as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = \frac{\mathrm{TP}}{P} \tag{18}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} = \frac{\mathrm{FP}}{N}. \tag{19}$$

In a coordinate system, the diagonal line from (0, 0) to (1, 1) divides the ROC space into two regions: the upper left and the lower right. Points above this diagonal line represent good classification results (better than random classification), while points below this line represent poor classification results (worse than random classification) [38]. The formula for calculating AUC by approximating the area under the ROC curve as the sum of areas of multiple trapezoids is

$$\mathrm{AUC} = \sum_{i=1}^{n-1} \frac{(\mathrm{TPR}_{i+1} - \mathrm{TPR}_i) \times (\mathrm{FPR}_{i+1} + \mathrm{FPR}_i)}{2} \tag{20}$$

where n denotes the number of points on the ROC curve. This formula computes the Area Under the AUC by summing the areas of trapezoids formed between successive points on the curve, with each trapezoid's area determined by the differences in TPR and FPR values.

AUC measures the likelihood that the anomaly detection model will rank a randomly chosen anomalous sample higher than a randomly chosen normal sample. A higher AUC value indicates a better model performance in distinguishing between normal and anomalous samples.

While both ACC and AUC are important evaluation metrics, in this study, the AUC is of greater significance. This is because AUC provides a more comprehensive measure of model performance across various decision thresholds, whereas ACC can be sensitive to class imbalances, especially in anomaly detection tasks where the number of normal samples typically far exceeds the number of anomalous samples. Therefore, AUC is a more robust indicator of the model's ability to distinguish between normal and abnormal instances across different decision thresholds, making it the preferred metric for evaluating the effectiveness of our anomaly detection model.

TABLE III
BASELINE COMPARING THE EXPERIMENTAL RESULTS ON THE BATTERY-SL10000 DATASET

Method	AUC	ACC
Deep SVDD	68.1467 (± 5.6902)	66.7088 (± 4.1581)
CNN Autoencoder	67.1395 (± 0.3511)	66.3509 (± 0.4582)
Region CNN	72.1301 (± 0.7873)	68.8529 (± 0.8231)
PatchCore	77.6088 (± 0.6781)	71.8961 (± 0.3577)
CFlow-AD	78.9655 (± 0.4772)	72.0367 (± 0.4591)
Gate-conv SVDD	**79.5836** (± 0.5198)	**72.2987** (± 0.3353)

C. Baseline Comparisons

To evaluate the proposed Gate Convolutional SVDD model in this article, we conducted experiments on the Battery-SL10000 dataset. The experimental dataset consists of 15 000 UAV captured images, with 9968 normal samples included in the training set. The test set comprises an equal number of 1000 normal and abnormal samples, all carefully selected from the remaining images.

We selected Deep SVDD, CNN autoencoder, and Region CNN as comparative models to ensure the comprehensiveness and fairness of the experimental results. All experimental results are presented in Table III.

The results in the table reveal that Gate-conv SVDD outperforms all other models in the comprehensive evaluation of the Battery-SL10000 dataset, with the highest AUC (79.5836) and ACC (72.2987). This superior performance is primarily attributed to the innovative gating mechanism, which significantly enhances the feature extraction process within the SVDD framework. By optimizing the information flow and leveraging the hypersphere structure, Gate-conv SVDD excels in anomaly detection tasks, demonstrating its effectiveness in accurately identifying outliers.

While other methods like CFlow-AD and Anomalib show competitive results, Gate-conv SVDD stands out due to its advanced feature propagation strategy, which enables it to better capture complex patterns and anomalies. These findings underscore the professional capabilities and potential of Gate-conv SVDD as a state-of-the-art solution for anomaly detection, highlighting its edge over traditional and newer models.

D. Ablation Study

In this section, we conduct a comprehensive ablation study to investigate the effectiveness of various architectural choices and structural components in improving the performance of the proposed anomaly detection model. The study focuses on several key aspects, including the comparison of different gating mechanisms, convolutional output structures, and dimensionality reduction techniques. Through a series of controlled experiments on the Battery-SL10000 dataset, we aim to evaluate the impact of each design decision on model performance, measured by AUC and ACC metrics.

The experiments are designed to highlight the importance of bidirectional gating, the placement of normalization layers,

TABLE IV
EXPERIMENTAL RESULTS COMPARING THE GATING STRUCTURES ON THE BATTERY-SL10000 DATASET

Method	AUC	ACC
Bilateral Gating Splicing	**78.0487** (± 0.5198)	**71.7358** (± 0.3353)
Bilateral Gating Hadamard-product	74.2167 (± 1.5312)	69.9000 (± 1.0440)
Unilateral Gating Splicing	75.0567 (± 0.6092)	70.5667 (± 0.8020)

TABLE V
EXPERIMENTAL RESULTS COMPARISON OF DIFFERENT CONVOLUTION OUTPUT STRUCTURES ON THE BATTERY-SL10000 DATASET

Method	AUC	ACC
Serial output Wo. BN	59.5908 (± 8.5752)	58.4167 (± 7.3014)
Parallel output Wo. BN	68.9273 (± 5.8967)	65.3482 (± 4.0250)
Serial output + In-layer BN	73.0333 (± 5.1083)	67.8551 (± 4.5664)
Parallel output + In-layer BN	**78.0487** (± 0.5198)	**71.7358** (± 0.3353)
Serial output + Out-layer BN	62.1467 (± 10.5594)	59.2259 (± 12.4936)
Parallel output + Out-layer BN	58.1373 (± 7.3364)	60.7209 (± 4.8094)

the type of convolutional output structures, and the use of dimensionality reduction techniques. By systematically testing these configurations, we provide valuable insights into how each component contributes to the overall model performance and stability.

1) Comparison of Different Gating Structures: Multilayer convolutional networks effectively extract high-dimensional features containing rich texture and color information through a deep hierarchical structure. These features play a key role in improving model performance without increasing the amount of data [39]. This study aims to verify the effectiveness of the proposed structure in achieving this goal through the following experimental settings. The experiment first compared the settings of different numbers of parallel gated convolutions to evaluate the difference in feature extraction between bidirectional and unidirectional gating mechanisms. Furthermore, we explored the fusion strategy of bidirectional gating, comparing two information fusion methods: concatenation dot product and dot product concatenation, in order to evaluate the comprehensive performance of the three gating structures. The experimental results are shown in Table IV.

For the multiconvolutional framework designed for image feature extraction, the ablation experiments first compared different numbers of parallel gated convolution settings to evaluate the differences between bidirectional and unidirectional gating mechanisms in feature extraction. The role of the gating structure is to eliminate errors by processing similar data through different activation and convolution forms. This enables the model to reduce errors and improve ACC.

The combination of the bidirectional gating mechanism and dot-product fusion ensures superior feature integration and transmission, contributing to better performance stability. The bidirectional gated convolution design further enhances model performance by leveraging smaller 1×3 and 3×1 convolutional kernels along with a 1×1 kernel. This design reduces parameter complexity while improving the model's ability to learn global features, which is crucial for effective anomaly detection. Therefore, we explored fusion strategies for bidirectional gating and compared two information fusion methods: direct concatenation and Hadamard product fusion, to evaluate the overall performance of the three gating structures. The experimental results are presented in Table IV.

The experimental results of the "Comparison of Different Gating Structures" on the Battery-SL10000 dataset clearly demonstrate the effectiveness of the proposed dual-gating mechanism with Hadamard product fusion in anomaly detection tasks, achieving an AUC of 78.0487 and an ACC of 71.7358%, which represents a significant improvement over other gating mechanisms. The results indicate that this structure outperforms all other gating configurations in both AUC and ACC, highlighting its superior ability to capture and utilize key features for detection. This significantly enhances the model's ACC, stability, and overall anomaly detection capability.

2) Comparison of Different Convolution Output Structures: After each iteration of the parameters of the training cycle, the distribution of the data features will change through the convolutional layer, which will make subsequent learning difficult [40]. To this end, we experiment with the insertion position of the BatchNorm layer relative to the gated convolutional layer. At the same time, the input-output relationship between the multilayer networks will directly affect the application effect of the model, so the output structure of serial and parallel outputs is set up in this experiment. On the training set of 10 000 normal samples, six output methods are tested to verify the rationality of the structure selected in this article. The experimental results are shown in Table V.

The experimental results on the Battery-SL10000 dataset demonstrate the significant impact of the model's decoding output structure on enhancing anomaly detection performance. Among the tested configurations, the combination of intralayer normalization and parallel output achieved the best results, with an AUC of 78.0487 and an ACC of 71.7358. This structure outperforms all other configurations because normalization effectively removes data noise and standardizes data scaling. In contrast, serial processing in the decoding output module leads to greater feature loss for image data due to the weak sequential correlation in images. Therefore, a parallel processing structure for decoding and outputting image data features yields the best performance.

Moreover, configurations without normalization layers result in significantly lower AUC values, regardless of whether the decoding output module processes data sequentially or in parallel. This confirms the critical role of normalization in enhancing model performance. Specifically, BatchNorm helps stabilize and optimize the information flow within convolutional layers, preserving the distribution characteristics of the data and improving the model's ability to detect anomalies.

Furthermore, the parallel output structure consistently outperforms the sequential output configuration, with AUC values of 68.92 and 59.59, respectively. This indicates that the parallel output structure is more effective at extracting image data features, as sequential structures are better suited for capturing sequential data characteristics. In contrast, parallel structures

TABLE VI
RESULTS OF THE EXPERIMENTAL COMPARISON OF DIFFERENT DIMENSIONALITY REDUCTION COMPONENTS ON THE BATTERY-SL10000 DATASET

Method	AUC	ACC
Single pooling + Conv	**79.5836** (± 0.5198)	**72.2987** (± 0.3353)
Single Pooling Only	72.9467 (± 1.5312)	69.8022 (± 1.0440)
Double Conv	75.0567 (± 0.6092)	70.5667 (± 0.8020)
Double Pre-pooling + Conv	76.0433 (± 0.5204)	70.7354 (± 4.1057)
Single Pooling + Bilayer Conv	73.5438 (± 1.8508)	68.9373 (± 4.0632)

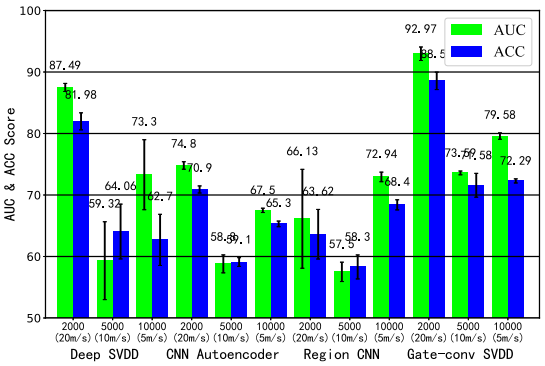

Fig. 6. Comparative experimental results of the real data collected by UAV.

excel in anomaly detection tasks by capturing both local and global features in image data. By identifying differences in texture and color, as well as normalizing features, the model significantly enhances its ability to distinguish between normal and anomalous patterns.

In conclusion, the experimental results highlight the crucial role of intralayer normalization and the parallel output structure in improving model ACC and robustness. These elements play a vital role in enhancing feature extraction, ensuring that the model can effectively detect anomalies.

3) Comparison of Different Dimensionality Reduction Structures: Although multilayer convolution can effectively map data features into high-dimensional space, directly partitioning hyperspheres in this space encounters challenges due to high computational complexity and the inability to meet real-time detection requirements [41]. Therefore, we explored various methods to reduce the data dimensions while maintaining the model's performance.

This study examined five dimensionality reduction techniques, and designed corresponding experiments. In these experiments, we ensured the consistency of dimensions before and after data reduction. We modified the position of the pooling layer, creating post-pooling single-layer convolution and prepooling single-layer convolution structures. Further modifications included adjusting the number of convolution layers to form double-layer convolution and prepooling double-layer convolution structures. Additionally, a control group without pooling was also implemented.

As presented in Table VI, the experimental results on the Battery-SL10000 dataset highlight the importance of dimensionality reduction in balancing computational efficiency and model ACC. Among the five methods tested, the structure that incorporates both pooling and convolutional layers achieved the best performance, with an AUC of 79.5833 and an ACC of 72.2933. The experimental results highlight the importance of pooling before dimensionality reduction, as it helps to filter out noise and preserve key features that are crucial for effective anomaly detection. This method effectively combines feature compression with the ResNet embeddings at each layer, ensuring that the original information is gradually integrated into the feature vectors during the dimensionality reduction process, thereby enhancing both efficiency and ACC. In contrast, the performance of the prepooling dual-layer convolutional structure shows a slight decline, with an AUC of 76.0433 and an ACC of 70.7354. The dual-layer convolutional structure not only increases the computational burden but also makes the model more prone to overfitting, leading to a decline in ACC. Nonetheless, it still outperforms other configurations, such as the single-layer convolution with post-pooling and the dual-layer convolution without pooling.

Moreover, the design of multiple optimal dimensionality reduction modules integrates layer-wise feature compression with the continuous incorporation of ResNet embeddings, successfully balancing feature retention and computational complexity reduction. This approach enhances feature extraction while minimizing the number of parameters.

In conclusion, the experimental results confirm the necessity of a well-structured dimensionality reduction strategy, which integrates the placement of pooling layers and ResNet embeddings. These elements enable the model to retain critical features while optimizing efficiency and ACC. The combination of these strategies significantly enhances the model's anomaly detection capability while ensuring computational feasibility for real-time applications.

E. Impact of Different Speeds and Datasets

In the experimental process, we simulated a real-world inspection scenario and conducted anomaly analysis and detection on images captured by drones at different flight speeds, yielding some interesting results, as shown in Fig. 6. Specifically, we set three distinct flight speeds: 5, 10, and 20 m/s. As the speed increased, the number of images captured per unit of time correspondingly decreased. Specifically, at a speed of 5 m/s, we obtained 10 000 images; at 10 m/s, 5000 images were collected; and at 20 m/s, only 2000 images were captured. These images formed three datasets of varying scales, which were utilized to evaluate the performance of the anomaly detection model under different speed conditions.

During the optimization process, in order to balance rapid progress and precise tuning, we initially adopted a higher learning rate and then gradually decreased it, ensuring the algorithm converged to the optimal solution [42]. By configuring the parameters of the Adam optimizer, including the decay factors β_1, β_2 and the multistep decay parameter γ, we effectively regulate the minimization of the hypersphere radius [43]. We compared three inspection speeds—20, 10, and 5 m/s—across differently scaled datasets and summarize the main findings as follows.

1) Comparison of Detection Accuracy Across Different Methods and Speeds: From the AUC and ACC metrics in the experimental results, it is apparent that under speeds of 20 m/s, 10 m/s, and 5 m/s, the Gate-conv SVDD model outperforms the other three methods (Deep SVDD, CNN Autoencoder, and Region CNN). For instance, when operating at 20 m/s on a small dataset of approximately 2000 images, Gate-conv SVDD achieves an AUC of 92.97% and an ACC of 88.56%, which is significantly higher than the other approaches. Deep SVDD, though it obtains comparatively good results at 20 m/s (AUC 87.49%, ACC 81.98%), exhibits a notable performance drop when the speed decreases to 10 m/s or 5 m/s, accompanied by an increase in dataset size. Similarly, CNN Autoencoder and Region CNN show weaker stability under all speeds, particularly for medium-sized datasets (about 5,000 images) collected at 10 m/s.

2) Performance Comparison Across Three Speeds: We conduct the performance comparison across three different speeds as follows.

1) *Speed 20 m/s (Small Dataset, 2000 Images):* The faster flight speed covers a larger area in a single inspection, but yields fewer images. Nevertheless, Gate-conv SVDD and other deep models with robust feature extraction and high-dimensional mapping capabilities can still achieve high ACC even with a relatively limited amount of data.
2) *Speed 10 m/s (Medium Dataset, 5000 Images):* More images are captured, but certain methods, such as Deep SVDD, CNN Autoencoder, and Region CNN, are more sensitive to shifts in data distribution or may be insufficiently tuned for this volume of data, leading to underwhelming performance. By contrast, Gate-conv SVDD maintains robust ACC and stability at this speed.
3) *Speed 5 m/s (Large Dataset, 10000 Images):* This slower speed produces the largest dataset, favoring deep models that need extensive data for capturing subtle fault features. Although Region CNN shows some improvements in AUC under these conditions, Gate-conv SVDD still stands out with the highest overall metrics, reflecting its strong generalization on large-scale data.

3) Balances and Optimizations Among Speed and Data Volume: The experiments indicate that at higher speeds (e.g., 20 m/s), it is possible to inspect a wider area more efficiently, but the total number of collected images is relatively small. Thus, to further improve ACC, it may be necessary to employ more powerful feature extraction or incremental data gathering in subsequent training. Conversely, at slower speeds (e.g., 5 m/s), although the model benefits from a significantly larger dataset for more thorough learning, the inspection efficiency is reduced. For extensive solar farms, prolonged slow-speed inspections may increase operational costs. Therefore, to balance detection ACC and operational efficiency, one may rely on *larger datasets* for model training and optimization while considering real-world constraints, such as the solar farm's acreage and photovoltaic array arrangement to determine the appropriate flight speed and data quality.

4) Conclusion and Practical Recommendations: From the standpoint of balancing speed and ACC, it is advisable to employ different strategies based on the scale of the photovoltaic system.

TABLE VII
COMPARISON OF DIFFERENT METHODS ON THE BATTERY-SL10000 DATASET

Methods	Model size	Training time	Inference time	AUC	ACC
Deep SVDD	520544	66.05	0.01100	68.1467	66.7088
CNN Autoencoder	35790	43.83	0.01635	67.1395	66.3509
Region CNN	12006	22.03	0.00210	72.1301	68.8529
Gate-conv SVDD	629807	57.30	0.01085	79.5836	72.2987

1) *Small Photovoltaic Installations:* Since the inspection area is limited, it is preferable to operate at slower speeds (e.g., 5 m/s), allowing the collection of higher-resolution images and improving fault detection ACC.
2) *Medium and Large Photovoltaic Farms:* To enhance operational efficiency, faster speeds (e.g., 20 m/s) are recommended for swift data collection. Coupled with efficient deep anomaly detection models such as Gate-conv SVDD, adequate ACC can still be maintained, striking a balance between inspection speed and detection precision.

F. Efficiency Experiment

To evaluate the tradeoff between computational cost and detection performance, we conducted an efficiency experiment involving four representative methods: 1) Deep SVDD; 2) CNN Autoencoder; 3) Region CNN; and 4) Gate-conv SVDD. As shown in Table VII, each method was assessed based on model size, training time, inference time, and detection metrics (AUC and ACC).

From the table, we observe that Gate-conv SVDD achieves the highest anomaly detection performance, with an AUC of 79.5836 and an ACC of 72.2987, albeit at the cost of a relatively large model size (629 807 parameters). Region CNN, Deep SVDD, and CNN Autoencoder demonstrate a lower parameter count compared to Gate-conv SVDD. These observations highlight a fundamental tradeoff in model design, where more compact architectures, despite reducing computational demands, exhibit lower detection ACC compared to the parametrically richer Gate-conv SVDD.

Furthermore, the training time and inference time reported in Table VII highlight differing suitability for real-world deployment scenarios. While Region CNN achieves faster inference at 0.00210 s, Gate-conv SVDD delivers superior detection ACC, rendering it a more suitable option when performance is the priority. CNN Autoencoder and Deep SVDD achieve shorter training times of 52.83 and 66.05 s, respectively, while exhibiting reduced detection performance. Overall, this efficiency experiment illustrates Gate-conv SVDD effectively achieving high performance by optimizing ACC while maintaining efficiency despite increased inference time and parameter count.

G. Parametric Sensitivity Analysis

In order to examine the impact of different hyperparameters on anomaly detection performance, we followed the methodology described in the original experiment. Specifically, we selected seven equally spaced values for β_1, β_2, and γ, then employed stepwise optimization and visual

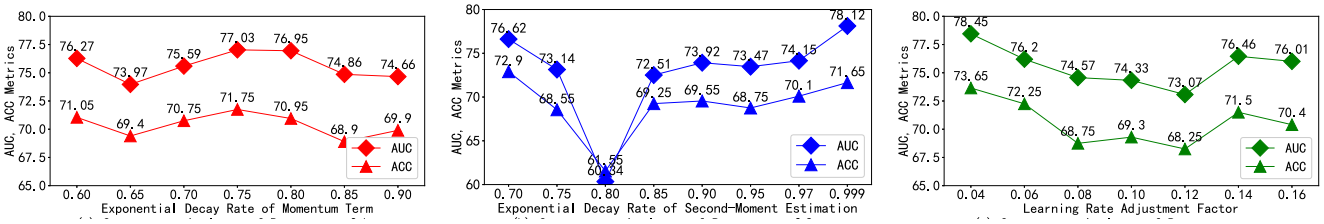

Fig. 7. Parametric sensitivity analysis curves. (a) Sensitivity analysis of parameter β_1. (b) Sensitivity analysis of parameter β_2. (c) Sensitivity analysis of parameter β_3.

comparison (see Fig. 7) to ascertain the influence of each parameter on the model's learning dynamics. This systematic approach ensured a thorough exploration of the parameter space and provided a reliable basis for subsequent conclusions.

The first stage of the sensitivity analysis focused on the momentum term, β_1. When β_1 varied within $[0.60, 0.90]$, both AUC and ACC metrics underwent periodic fluctuations. Notably, performance peaked at $\beta_1 = 0.75$. This somewhat lower momentum decay rate, relative to typical defaults in Adam, enables more rapid adjustments in the early training phases to better capture anomalous patterns. At the same time, it avoids excessive gradient smoothing that can overlook transient but critical features of the data.

Subsequent experiments investigated the second-moment estimation parameter, β_2. Varying β_2 in the range $[0.70, 0.999]$ revealed a distinct U-shaped curve in the AUC metrics. Significantly lower settings (e.g., 0.70 or 0.75) compromised performance, indicating that an insufficiently smoothed second-moment estimate hinders stable convergence. By contrast, a larger decay rate yielded smoother gradient updates. The best anomaly detection results emerged at $\beta_2 = 0.999$, demonstrating that a highly smoothed variance accumulation is advantageous for this task.

Finally, with β_1 and β_2 fixed at their optimal values, we analyzed the effect of the learning rate adjustment factor, γ. As γ increased from 0.04 to 0.16, both AUC and ACC exhibited an overall decline. The optimal performance was achieved at $\gamma = 0.04$, confirming that a relatively small adjustment factor, combined with a carefully designed multistep learning rate decay strategy, effectively refines model parameters in later training stages without overshooting or converging too rapidly.

Overall, these findings support the efficacy of the Adam algorithm's adaptive properties [44], particularly its ability to escape saddle points and dynamically adjust learning rates. Moreover, employing a multistep decay strategy further enhances convergence and stability as training progresses. In practice, the best detection performance in these experiments was reached by setting β_1 near 0.75, β_2 close to 0.999, and γ at 0.04. These results provide valuable guidance for fine-tuning model hyperparameters in anomaly detection scenarios, helping to achieve both efficiency and ACC.

H. Anomaly Detection Case Study

This study adopts a case study approach to analyze surface anomalies in photovoltaic panels recorded in the Battery-SL dataset. These anomalies directly lead to a reduction in power generation and even pose a risk of failure. In real-world operations, such anomalies are difficult to detect due to the remote locations of photovoltaic panels and their random occurrence. In the case study, the Gated Conv-SVDD model accurately identified anomalous photovoltaic panels in the image data, as indicated by the red-boxed areas in Fig. 8. The high label scores of these panels suggest the presence of potential issues.

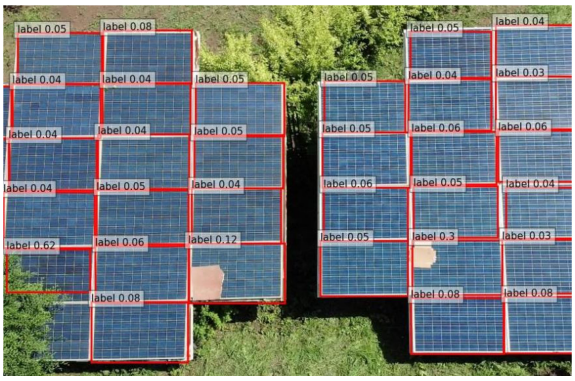

Fig. 8. Abnormal detection effect.

This case study validates the effectiveness of the Gated Conv-SVDD model in real-world photovoltaic power station monitoring and provides quantitative anomaly scores. The findings offer crucial support for the automatic inspection and anomaly detection of photovoltaic panels, further contributing to intelligent maintenance in photovoltaic power stations and enhancing smart applications in similar scenarios.

V. Conclusion and Discussion

This article introduces Gate-conv SVDD, a novel method for anomaly detection on photovoltaic panel surfaces using drone-captured images. It combines gated convolutional feature extraction with SVDD clustering to enhance detection ACC and computational efficiency. The approach excels in rapid, high-precision anomaly detection for large-scale solar panels under high-speed drone motion, advancing scalable and generalizable drone-based image analysis. However, the current study is limited to server-based experiments, and claims of deployment-readiness require further validation. Future work will involve hardware integration tests on UAV platforms and field validations under diverse environmental conditions to

ensure robust performance in real-world applications. These steps will bridge the gap between controlled experiments and operational deployment, enhancing the framework's practical utility for sustainable solar energy systems.

In addition, despite the Gate-conv SVDD model demonstrating effective quantitative anomaly detection, its performance in scenarios with limited labeled data or without labeled data remains suboptimal. Future work will investigate unsupervised, rapid, and efficient quantitative anomaly detection methods using unlabeled data to improve the algorithm's efficiency and quality in large-scale system applications.

REFERENCES

[1] A. Rahman, O. Farrok, and M. M. Haque, "Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic," *Renew. Sustain. Energy Rev.*, vol. 161, Jun. 2022, Art. no. 112279.

[2] A. A. Hasan et al., "Delamination-and electromigration-related failures in solar panels a review," *Sustainability*, vol. 13, no. 12, p. 6882, 2021.

[3] M. Köntges, S. Kajari-Schröder, I. Kunze, and U. Jahn, "Crack statistic of crystalline silicon photovoltaic modules," in *Proc. 26th Eur. Photovoltaic Solar Energy Conf. Exhibit.*, 2011, pp. 3290–3294.

[4] M. Vlaminck, R. Heidbuchel, W. Philips, and H. Luong, "Region-based cnn for anomaly detection in PV power plants using aerial imagery," *Sensors*, vol. 22, no. 3, p. 1244, 2022.

[5] A. Oulefki et al., "Detection and analysis of deteriorated areas in solar pv modules using unsupervised sensing algorithms and 3D augmented reality," *Heliyon*, vol. 10, no. 6, 2024, Art. no. e27973.

[6] F. A. Almalki, A. A. Albraikan, B. O. Soufiene, O. Ali, and T. Gaber, "Utilizing artificial intelligence and lotus effect in an emerging intelligent drone for persevering solar panel efficiency," *Wireless Commun. Mobile Comput.*, 2022, to be published.

[7] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[8] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2909–2917.

[9] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient gan-based anomaly detection," 2018, *arXiv preprint*.

[10] L. Wang, D. Guo, H. Wu, K. Li, and W. Yu, "TC-GCN: Triple cross-attention and graph convolutional network for traffic forecasting," *Inf. Fusion*, vol. 105, May 2024, Art. no. 102229.

[11] J. Wang et al., "STHGCN: A spatiotemporal prediction framework based on higher-order graph convolution networks," *Knowl.-Based Syst.*, vol. 258, Dec. 2022, Art. no. 109985.

[12] A. K. Sangaiah, F.-N. Yu, Y.-B. Lin, W.-C. Shen, and A. Sharma, "UAV T-YOLO-Rice: An enhanced tiny yolo networks for rice leaves diseases detection in paddy agronomy," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 6, pp. 5201–5216, Nov./Dec. 2024.

[13] B. Mohammadi, M. Fathy, and M. Sabokrou, "Image/video deep anomaly detection: A survey," 2021, *arXiv:2103.01739*.

[14] Z. Zhang and X. Deng, "Anomaly detection using improved deep svdd model with data structure preservation," *Pattern Recognit. Lett.*, vol. 148, pp. 1–6, Aug. 2021.

[15] Y. Fan, G. Wen, D. Li, S. Qiu, M. D. Levine, and F. Xiao, "Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder," *Comput. Vis. Image Understand.*, vol. 195, Jun. 2020, Art. no. 102920.

[16] A. B. Hassanat et al., "A novel outlier-robust accuracy measure for machine learning regression using a non-convex distance metric," *Mathematics*, vol. 12, no. 22, p. 3623, 2024.

[17] K. Lee, K. Lee, H. Lee, and J. Shin, "A simple unified framework for detecting out-of-distribution samples and adversarial attacks," in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.

[18] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," in *Proc. 34th Conf. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21464–21475.

[19] P. Wei, W. Zhu, Y. Yang, Z. Fei, and C. Xie, "Spatiotemporal transform network-based anomaly detection and localization of distributed parameter systems," *IEEE Trans. Ind. Informat.*, vol. 21, no. 4, pp. 2770–2778, Apr. 2025.

[20] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 14298–14308.

[21] T. Vojir, T. Šipka, R. Aljundi, N. Chumerin, D. O. Reino, and J. Matas, "Road anomaly detection by partial image reconstruction with segmentation coupling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15651–15660.

[22] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107706.

[23] D. Lin, Y. Li, S. Xie, T. L. Nwe, and S. Dong, "DDR-ID: Dual deep reconstruction networks based image decomposition for anomaly detection," *J. Ambient Intell. Human. Comput.*, vol. 14, no. 3, pp. 2125–2139, 2023.

[24] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9664–9674.

[25] L. Deecke, R. Vandermeulen, L. Ruff, S. Mandt, and M. Kloft, "Image anomaly detection with generative adversarial networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, Dublin, Ireland, 2019, pp. 3–17.

[26] Y. Cao, J. Zhang, L. Frittoli, Y. Cheng, W. Shen, and G. Boracchi, "Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 55–72.

[27] K. Rezaee, S. M. Rezakhani, M. R. Khosravi, and M. K. Moghimi, "A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance," *Pers. Ubiquitous Comput.*, vol. 28, no. 1, pp. 135–151, 2024.

[28] K. Yu et al., "Hybrid quantum classical optimization for low-carbon sustainable edge architecture in RIS-assisted AIoT healthcare systems," *IEEE Internet Things J.*, vol. 11, no. 24, pp. 38987–38998, Dec. 2024.

[29] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vi (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[30] S. Choi, S. Jung, H. Yun, J. T. Kim, S. Kim, and J. Choo, "RobustNet: Improving domain generalization in urban-scene segmentation via instance selective whitening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 11580–11590.

[31] X. Hu, B. Zhao, T. Sun, S. Cao, S. Fan, and H. Xing, "A detecting algorithm for occlusion on the surface of photovoltaic modules based on improved YOLOv5," in *Proc. China Autom. Congr. (CAC)*, 2021, pp. 3131–3135.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[33] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. 32nd Conf. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.

[34] L. Huang, D. Yang, B. Lang, and J. Deng, "Decorrelated batch normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 791–800.

[35] Y. Pu et al., "Variational autoencoder for deep learning of images, labels and captions," in *Proc. 30th Conf. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–11.

[36] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[37] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thoracic Oncol.*, vol. 5, no. 9, pp. 1315–1316, 2010.

[38] C.-I. Chang, "Comprehensive analysis of receiver operating characteristic (ROC) curves for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5541124.

[39] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning rich features for image manipulation detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1053–1061.

[40] S.-H. Gao, Q. Han, D. Li, M.-M. Cheng, and P. Peng, "Representative batch normalization with feature calibration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8669–8679.

[41] Y. Aytar and A. Zisserman, "Immediate, scalable object category detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2377–2384.

[42] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of deep learning models for multi-step ahead time series prediction," *IEEE Access*, vol. 9, pp. 83105–83123, 2021.

[43] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, 2023.

[44] C. Jin, R. Ge, P. Netrapalli, S. M. Kakade, and M. I. Jordan, "How to escape saddle points efficiently," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1724–1732.