

Infrared Defect Detection for Power System Photovoltaic Panels in UAV Inspection Using Dynamic Convolution

Xiaoyang Hu^{1,*}, Donghui Xi¹, Chuande Hou¹, Shuhan Chen¹ and Zehua Zhang¹

¹School of Equipment Engineering, Shenyang Ligong University, Shenyang 110159, Liaoning, China

Abstract

INTRODUCTION: This paper addresses the significant demand for non-contact, efficient, and large-area inspections in photovoltaic power system drone infrared inspections. It also improves the performance bottlenecks of existing lightweight models in handling the low-resolution, high-noise, and complex features characteristic of real-world infrared images from power plants.

OBJECTIVES: The objective is to propose an Infrared Adaptive Dynamic Perceptron (IADP) that enhances defect recognition accuracy and hot spot detection precision for drone-based infrared inspections, while maintaining low computational complexity and achieving real-time performance on resource-constrained embedded platforms.

METHODS: The core method employs a dynamic convolution mechanism that aggregates multiple heterogeneous kernels based on input-dependent attention. The key innovations focus on the resource limitations of drones: a global average pooling method tailored to infrared images optimizes the attention mechanism, and the introduction of the Softplus activation function mitigates the loss of low-contrast features. Dynamic convolution kernels (DK-conv) are used to effectively adapt to the regular textures and localized defects of photovoltaic panels. A hierarchical temperature annealing training strategy is introduced to address training stability issues.

RESULTS: Experiments on a dataset of 11,608 drone infrared images show that the IADP-MobileNetV3-Small model improves defect recognition accuracy (TOP-1) by 4.0% and hot spot detection precision (mAP@0.5) by 4.7%, with only a 3.5% increase in computational complexity. It achieves 40.1 FPS on an embedded platform, exceeding the 20 FPS engineering requirement.

CONCLUSION: The proposed IADP model effectively meets the demands for real-time, accurate, and efficient infrared inspection in photovoltaic power plants using drones, offering a good balance between performance and computational cost for embedded deployment.

Keywords: Dynamic Convolution, Infrared Image Detection, Lightweight Models, Multi-core Aggregation Strategy, Photovoltaic Power Systems

Received on 15 February 2026, accepted on 19 May 2026, published on 12 October 2026

Copyright © 2026 Xiaoyang Hu *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.14715

*Corresponding author. Email: xiaoyanghu@163.com

1. Introduction

As the main energy conversion equipment in modern photovoltaic power stations, photovoltaic panels are the foundation and lifeline of the power system's stable and efficient operation, as well as the basis for clean energy output and the economic profit of the system. Driven by global carbon neutrality policies, photovoltaic installed

capacity in various countries has maintained rapid and sustained growth, and the overall scale and spatial complexity of photovoltaic power stations keep rising year by year [1]. PV arrays directly face the important role in converting solar energy into electrical energy in the power system, and the performance of PV arrays can directly influence the output power, operating reliability, and economic benefit of the power plant. But, in complex environments, defects such as microcracks and hot spots are potential issues with long-term

PV operation. These defects can be very harmful: cracks may destroy the cell structure, lead to a premature ageing of the components, and result in a permanent reduction of the power output, whilst hot spots are caused by partial shading or failure and lead to high temperatures in local areas. The main principle in this effect is the loss of focused energy at defect sites or bypass paths between defect sites, and the subsequent increase of local temperature as a consequence of the reverse current. This not only causes severe damage to the PV cells themselves, but can be dangerous, e.g., burning the backsheet or burning the soldering strips.

At a macro scale, PV components of failure will greatly decrease the power generation efficiency of the PV array and affect the revenue of the power plant. The biggest problem is the extremely high temperatures resulting from the increasing development of hot spots that can be easily ignited by surrounding materials, thus firing power plants and jeopardizing the safety of personnel, power plant equipment, and the stability of the grid-connected power system. As these defects are usually highly hidden and not easily visible in the early stages, but their impact can lead to catastrophic power losses and safety accidents, the development and application of new intelligent detection technologies for early identification and precise localization of potential risks are irreplaceable for the sustainable development of the photovoltaic industry and the safe and stable operation of power systems. At present, inspection of photovoltaic panels with the help of infrared drones has become a common detection method in power plants, which is an important step to ensure the safe and efficient operation of power plants. The advantages of the drone platforms are their flexibility in deployment, ability to cover a wide area, and being a non-contact technology, while infrared imaging technology can accurately detect temperature abnormalities on the surface of photovoltaic panels (hot spots and localized overheating due to poor contact), which can effectively locate electrical or structural abnormalities. Furthermore, it is not sensitive to light sources and is the main approach used in PV defect detection using drones.

However, the infrared images are often of low grayscale contrast, with a high amount of noise interference, and with blurred textures. Moreover, the computational power of drone platforms (such as weak CPU/GPU power and limited memory) is a major obstacle to the accuracy and real-time capabilities of defect detection algorithms. Although lightweight models, such as the Mobilenet [2], [3] series with static convolution kernel, have certain advantages of lightweight, it is still difficult to use them to deal with multi-scale feature fusion and low-resolution and high-noise infrared images. When considering lightweight models and multi-scale feature fusion, some previous lightweight models based on static convolution (such as Mobilenet series) have certain advantages in computational speed, but it is difficult to meet the requirements of low resolution and high noise of infrared images due to the limitation of a fixed convolution kernel. Additionally, the computing power of the drone platform is extremely limited, making the increase of the model's depth and width extremely challenging, which are key to the performance of the model. Limited scalability

makes the performance bottleneck in detection. In response to these challenges, we are presenting an Infrared-Adaptive Dynamic Perceptor (IADP), a Dynamic Feature Extraction Module for Drone-based Infrared Images to detect defects of photovoltaic panels in the power system. It is built on the basis of a dynamic kernel aggregation mechanism and a lightweight multi-scale perception architecture. To accomplish this, the proposed method takes a convolutional approach that dynamically fuses kernels and is inspired by dynamic convolution, offering state-of-the-art complex feature extraction capabilities, in addition to low overhead requirements for the edge computing devices. The key features of the design are:

- (i) In order to solve the image with low contrast and noise interference in infrared images, a dynamic kernel fusion mechanism is proposed. A lightweight attention network is used to generate weights $\pi_k(x)$ to dynamically combine multiple sets of heterogeneous convolution kernels $\{\widehat{W}_k, \widehat{b}_k\}$.
- (ii) The IADP module uses a global average pooling (GAP) feature extraction method that is especially suitable for infrared images to lower the computational cost while efficiently extracting the temperature features of photovoltaic panels, and introduces an attention mechanism module (IADP-Attention Mechanism, IADP-AM). The Softplus function is used as the activation function.
- (iii) The dynamic convolutional kernel (DK-conv) design is used: it is a combination of 3×3 depthwise convolution (DWConv) and 1×1 pointwise convolution (PWConv) to suit the regular pattern of the PVs and the localized defect.

These designs help the IADP to see the temperature of each PV panel in the PV power system, as well as to get an understanding of the overall temperature distribution and local temperature anomalies among the PV panels. They successfully alleviate the problem of feature loss in infrared image scenes with low contrast and greatly decrease the amount of calculation. This design is suitable for edge devices and can be used for defect recognition of photovoltaic panel arrays in an infrared imaging system under a drone to ensure the reliable operation of photovoltaic power plants.

2. Related Work

2.1 The Evolution of Photovoltaic Power System Inspection Technologies

The power inspection technology of photovoltaic (PV) power stations has been developed in terms of inspection carriers. Manual inspection was mainly used in the early stages, where inspection personnel had to climb to the PV array to perform a close-up visual inspection or use handheld devices to perform the inspection. This method was inefficient, untimely, and risked safety while operating at high altitudes

or in difficult terrain. To improve safety and efficiency, ground-based robotic inspection technology emerged. Reference [4] proposed a PV inspection unmanned vehicle that can move along a predefined track based on path planning algorithms to perform some detection tasks, thus reducing personnel risks. However, its core limitation lies in the inability to effectively reach elevated areas or cross complex obstacles for inspections, rendering it ineffective for inspecting large-scale, multi-layered, or rooftop PV arrays. The advent of unmanned aerial vehicles (UAVs) for inspections overcame the limitation of height accessibility. UAV-based PV panel inspection technology, as proposed in Reference [5], can quickly cover large areas due to its flexible maneuverability, making it particularly suitable for large ground-based and rooftop PV stations. Early UAV inspections primarily mounted visible light cameras, capable of identifying obvious physical damage such as cracks, shading, or dirt, but were unable to detect key indicators related to internal faults or electrical connection issues—thermal anomalies (hotspots caused by localized overheating). Therefore, the integration of infrared thermal imaging technology into UAV platforms became a critical breakthrough. UAVs equipped with infrared cameras can non-contactly capture thermal images of the PV panel surface, accurately focus on and locate hotspots caused by issues such as defects in the cells, bypass diode failure, series mismatch, or shading, thus enabling efficient, safe, and comprehensive diagnostics of potential faults and performance degradation in PV modules. Reference [6] specifically focuses on the use of UAV platforms to detect infrared images of PV panels, identifying hotspots and allowing for the early detection and prevention of potential safety hazards.

2.2 The Evolution of Defect Detection Methods for Photovoltaic Panels in Power Systems

The classic infrared image defect recognition methods are primarily based on image processing techniques, such as threshold segmentation [7], morphological operations [8], and template matching [9]. Reference [10] adopts the Otsu thresholding algorithm to realize hotspot segmentation detection for photovoltaic module infrared images. However, these methods rely on manually designed features and are sensitive to image noise and background interference, showing poor robustness in complex scenarios. In recent years, deep learning-based methods have significantly improved defect detection accuracy, with mainstream frameworks including Faster R-CNN [11],[12] and the YOLO series [13],[14] of object detection models. Reference [15] uses an improved YOLOv7 model for defect detection in photovoltaic infrared images, which effectively captures defect features and improves detection accuracy. However, such detection models are often computationally expensive, and for tasks that require deploying models on drones for ground target detection, these mainstream frameworks generally do not adapt well. As a result, lightweight network

models (e.g., MobileNet, ShuffleNet [16]) have been proposed. Reference [17] adopts MobileNetV3 as the backbone to construct a lightweight real-time object detection model applicable to UAV onboard deployment. However, to meet the light-weight deployment requirements on UAVs, using only these traditional light-weight models leads to limited performance on infrared images: on the one hand, the low grayscale contrast and blurred textures in infrared images result in insufficient shallow feature extraction; on the other hand, lightweight models compress computation, often sacrificing network depth and width, which makes them prone to underfitting in complex noisy environments, leading to higher miss detection and false detection rates. To improve the performance of lightweight models in infrared image recognition, dynamic convolutional neural networks (CNNs) [18],[19], which have become a recent research hotspot, are introduced. The core idea of this network is to use multiple parallel convolution kernels and dynamically aggregate kernel weights through an attention mechanism. Reference [20] adopts multi-kernel attention-weighted dynamic convolution to optimize the detection model for UAV-captured photovoltaic infrared thermal images. This method adaptively adjusts convolution kernel weights according to diverse thermal features, effectively extracting multi-scale hot spot features and suppressing background noise. Dynamic convolutional neural networks, with the same computational cost as static convolutions, provide a notable improvement in model representation capabilities. However, existing dynamic convolution methods have not fully considered the computational limitations of edge devices, such as the use of high kernel counts ($K \geq 8$) or complex attention modules, which increase inference latency and are difficult to adapt to embedded platforms like drones. To address this limitation, this paper proposes a low-kernel design ($K = 4$) and optimizes the attention computation process, maintaining dynamic feature adaptability while controlling the additional computational overhead. Additionally, a progressive temperature annealing strategy and kernel diversity regularization are employed to solve the instability issues in training dynamic convolution for lightweight scenarios, providing a new paradigm for defect detection of photovoltaic panels in power systems using UAV infrared images.

3. The Model for Defect Detection of Photovoltaic Panels in Power Systems Based on Dynamic Convolution

3.1 The Infrared Inspection Scene Modeling and Problem Definition of Photovoltaic Power Systems

This section proposes a lightweight multi-task recognition model based on dynamic convolution to address the complex requirements of photovoltaic panel defect detection in UAV infrared images. The problem definition and scene modeling focus on infrared characteristics. The input consists of single-

channel infrared images (resolution $H \times W \times 1$) captured by UAVs, with pixel values normalized to the $[0, 1]$ range to represent the temperature distribution on the surface of the photovoltaic panels. Due to the low grayscale contrast and significant noise interference (such as cloud reflection and rain or fog scattering), and the fact that defects manifest as localized temperature anomalies (hotspots), traditional static convolution models struggle to effectively extract such features. Therefore, the model needs to precisely locate the photo-voltaic panel array in the preprocessing stage: this is achieved by image registration to eliminate displacement errors caused by UAV flight jitter, and by using a lightweight semantic segmentation network (such as a modified version of Mo-bileNetV3-Small) to extract regions of interest (ROI), thereby filtering out back-ground interference (such as vegetation and supports) and focusing the input on the photovoltaic panel body.

- (i) Classification Task: Outputs the probability distribution of defect types, enhancing the ability to distinguish low-contrast infrared features through dynamic convolution.
- (ii) Localization Task: Regresses the coordinates of defect bounding boxes $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$. By integrating the self-adaptive characteristics of dynamic kernels, it accurately captures the spatial distribution of temperature anomalies.

3.2 Design of Dynamic Convolution Module for Photovoltaic Power Systems

The Infrared Feature Perception Attention Mechanism for Photovoltaic Power Systems

In the task of detecting defects in photovoltaic panels using infrared drone images, the core value of infrared images lies in the temperature information they contain, which directly serves as the basis for detecting key defects such as hot spot effects. However, traditional Convolutional Neural Networks (CNNs) may dilute or weaken the global temperature information (particularly the overall or large-area temperature statistical characteristics of the image) during the feature extraction process, due to the spatial down-sampling operations in each layer. This could impact the model's sensitivity to temperature anomalies. To address this issue, an attention mechanism module is introduced in the IADP (IADP-Attention Mechanism, IADP-AM).

IADP-AM utilizes a Global Average Pooling (GAP) operation at the initial stage. The core function of GAP is to globally compress the spatial dimensions and extract global statistics for each channel. Its operation can be represented as in equation (1):

$$w_c = \text{GAP}(F_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c(i, j) \quad (1)$$

Let F_c represent the c -th channel of the input feature map, and H and W denote the spatial height and width of the feature map. w_c is the output channel descriptor (scalar value), and in the context of infrared imagery, w_c directly represents the

average temperature response intensity of the corresponding mode for the c -th channel across the entire image (i.e., the global temperature statistics).

By using the feature vector generated by the Global Average Pooling (GAP), which contains the global temperature statistics, as the starting point for the IADP-AM, it ensures that the subsequent attention weight computations (whether for channels or spatial regions) are fully integrated with the global temperature distribution information from the very beginning. This enables the IADP-AM to explicitly refer to and utilize the global temperature context when determining which channels or spatial areas to emphasize. As a result, the network can be guided to focus more on local regions that are anomalous in terms of overall temperature distribution (such as thermal spots significantly higher than the average temperature). This method, while retaining key temperature information, incurs minimal computational overhead, has a simple structure, and effectively enhances the model's accuracy and robustness in detecting defects, especially temperature-related defects, in photovoltaic panels in unmanned aerial vehicle (UAV) infrared inspection scenarios.

Subsequently, the output from the GAP is passed through a multi-layer perceptron (MLP) along the channel dimension. To reduce computational overhead, the IADP sets the input channel compression ratio of the MLP to 4, reducing the number of channels to $C/4$, and uses the Softplus function as the activation function, as expressed in Equation (2):

$$\text{Softplus}(x) = \log(1 + e^x) \quad (2)$$

The smooth gradient characteristics of Softplus can alleviate the feature loss problem in low-contrast scenarios. Compared to ReLU, Softplus still transmits subtle temperature difference signals in regions close to zero, avoiding the loss of details caused by hard truncation.

At the end of the MLP, the final attention weights $\{\pi_k(x)\}$ are generated via the softmax function. The structure of the IADP-AM is shown in Figure 1.

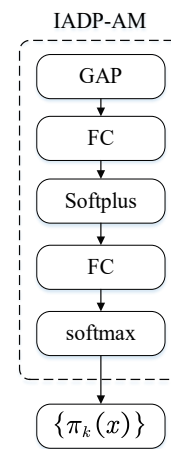


Figure 1. IADP-AM structure diagram

The IADP-AM combines GAP operations and the Softplus activation function. The overall module structure is simple and computationally efficient, effectively enhancing the model's sensitivity to temperature-related defects (particularly thermal spots) in infrared images, as well as its recognition accuracy and overall robustness. This provides a more reliable automated analysis tool for drone-based photovoltaic power station inspections without introducing significant overhead.

Optimization of Multi-Core Aggregation Strategy for Photovoltaic Power Systems

In the context of defect detection for drone-based infrared photovoltaic panels, this paper proposes a dynamic feature extraction module, IADP, which breaks through the limitations of traditional static convolution by introducing an adaptive kernel aggregation mechanism. Traditional static convolution layers employ a fixed-weight single-kernel operation, as shown in Equation (3):

$$y = g(W^T x + b) \quad (3)$$

The operation is difficult to adapt to the feature distribution of infrared images, where low-frequency background and high-frequency temperature anomalies coexist, especially when noise interference leads to missed detections.

To address this, IADP introduces multiple heterogeneous convolution kernels $\{\widehat{W}_k, \widehat{b}_k\}$, and generates input-related weights $\pi_k(x)$ through a lightweight attention network, dynamically aggregating them into adaptive convolution kernels, as shown in Equation (4):

$$\widehat{W}(x) = \sum_{k=1}^K \pi_k(x) \widehat{W}_k, \widehat{b}(x) = \sum_{k=1}^K \pi_k(x) \widehat{b}_k \quad (4)$$

The constraint $0 \leq \pi_k(x) \leq 1$ and $\sum_{k=1}^K \pi_k(x) = 1$ ensures that the aggregation kernel lies within the convex hull of the kernel space, preventing weight divergence.

To accommodate the extreme computational limitations of unmanned aerial vehicle (UAV) platforms, such as the Jetson Nano embedded GPU, the design of IADP strictly follows the principle of minimizing computational overhead. Compared to static perceptrons, IADP outputs the same number of channels, yet its model scale is larger. This is primarily due to the introduction of two additional computations: First, the generation of attention weight distributions $\{\pi_k(x)\}$ based on the input. Second, the dynamic aggregation of multiple parameter sets based on these weights, i.e., the weighted sum to generate dynamic parameters $\sum_{k=1}^K \pi_k \widehat{W}_k$ and $\sum_{k=1}^K \pi_k \widehat{b}_k$. To ensure the practicality of IADP, the total cost of these additional computations must be significantly lower than the core computational cost of the static perceptron, i.e., the floating-point operations of $\widehat{W}^T x + \widehat{b}$. This constraint can be formalized as Equation (5):

$$O(\widehat{W}^T x + \widehat{b}) \gg O\left(\sum_{k=1}^K \pi_k \widehat{W}_k\right) + O\left(\sum_{k=1}^K \pi_k \widehat{b}_k\right) + O(\pi(x)) \quad (5)$$

The $O(\cdot)$ notation denotes computational complexity. It is worth noting that fully connected layers, due to their dense parameter matrices, incur excessively high core computational costs, which cannot meet the constraints of this design. In contrast, convolutional layers, benefiting from sparse connectivity, exhibit a core computational complexity that is significantly higher than the overhead of dynamic parameter aggregation, thus effectively adapting to the design requirements of IADP. This trade-off allows IADP to enhance the model's expressive capability while avoiding excessive consumption of computational resources.

Regarding the selection of the number of kernels, K , consideration must also be given to the computational limitations of the drone. IADP, based on a multi-core dynamic aggregation mechanism, adaptively aggregates K convolution kernels (with size $D_k \times D_k$ and input/output channel numbers of C_{in} and C_{out} , respectively) according to the input features. The additional computational overhead is limited to $K C_{in} C_{out} D_k^2 + K C_{out}$ multiplication and accumulation operations. Compared to the core computational volume of standard convolution operations, $H W C_{in} C_{out} D_k^2$, when the number of kernels, K , is much smaller than the feature map dimensions $H \times W$ (e.g., in the case of drone infrared images with a typical resolution of 512×512 and $K=4$), the additional computational cost becomes almost negligible. As shown in Table 1, taking MobileNetV2 ($\times 1.0$) as an example, when $K=4$, the overall computational load increases by only about 4%, thus validating the lightweight advantage of this method. This design is particularly suitable for embedded platforms in drones: by maintaining small convolution kernels and a low number of kernels ($K=4$), the model can perform real-time infrared image processing on edge devices such as the Jetson Nano, enabling precise detection of defects such as hot spots on photovoltaic panels. It also prevents the risk of memory overflow caused by model size expansion, providing a cost-effective technological solution for low-altitude inspections in complex environments. Therefore, considering the overall performance and computational cost, the dynamic feature extraction module IADP in this study adopts a kernel configuration of $K=4$.

Table 1. The multiplication and accumulation operations of static convolution and dynamic convolution in MobileNetV2 (with four different width multipliers: $\times 1.0$, $\times 0.75$, $\times 0.5$, and $\times 0.35$)

	$\times 1.0$	$\times 0.75$	$\times 0.5$	$\times 0.35$
static	300.0M	209.0M	97.0M	59.2M
K=2	308M	215.4M	101.7M	62.3M
K=4	310.2M	216.5M	102.0M	62.5M
K=6	316.5M	219.5M	102.3M	62.5M
K=8	320.2M	221.3M	103.4M	63.1M

For the selection of convolutional kernels, in response to the regular textures and local defect features presented by photovoltaic panel arrays in infrared images, IADP innovatively employs a dynamic convolutional kernel (DK-conv) design that combines 3×3 depthwise convolution (DWConv) with 1×1 pointwise convolution (PWConv). This design achieves efficient detection through a collaborative mechanism of spatial feature modeling and cross-channel information fusion. The depthwise convolution independently processes each input channel's 3×3 spatial convolution in a sparse connection manner, focusing on extracting structural texture features such as photovoltaic panel grid lines and edge arrangements. Its local receptive field effectively suppresses the interference of cloud reflection and sensor noise while retaining low-frequency temperature distribution information of the photovoltaic panel background. The pointwise convolution, on the other hand, implements cross-channel feature linear combinations with a 1×1 kernel, dynamically fusing heterogeneous information such as high-temperature responses and texture fractures encoded in different channels. This enhances the ability to distinguish local anomalies, such as hot spots. The two convolution types are cascaded to form a heterogeneous convolution kernel group. The multi-kernel responses are dynamically aggregated using input-adaptive attention weights. When processing uniform backgrounds, depth-wise convolution dominates the spatial texture modeling to reduce false positives. In defect regions, pointwise convolution enhances the prominence of local anomalies through cross-channel feature fusion.

Finally, the paper proposes an optimal multi-aggregation strategy, where four parallel DK-conv kernels are dynamically aggregated using attention weight ($\{\pi_k(x)\}$) generated by IADP-AM, forming the IADP module. The structure of this module is illustrated in Figure 2:

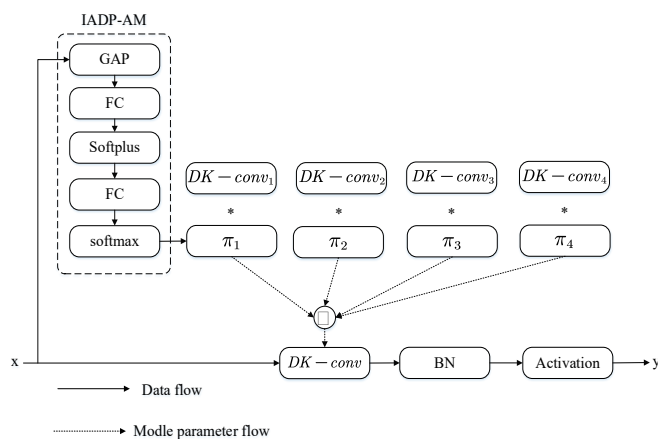


Figure 2. IADP structural diagram

The design of DK-conv effectively adapts to the regular textures and localized defect features of photovoltaic panel arrays. By employing $\{\pi_k(x)\}$ as attention weights, the

dynamic aggregation strategy of four DK-conv groups efficiently enhances both model performance and real-time capability. This approach is particularly suitable for resource-constrained scenarios such as edge devices, thereby providing advantages for defect detection in photovoltaic panel arrays under infrared imaging by unmanned aerial vehicles.

3.3 The architecture of the IADP-MobileNetV3-Small model for photovoltaic power systems

In this section, the Dynamic Feature Extraction Module (IADP) is integrated into the traditional lightweight model MobileNetV3-Small, resulting in a novel lightweight architecture, IADP-MobileNetV3-Small, designed for defect recognition of photovoltaic panel arrays under infrared imaging. Based on the core concept of Dynamic Convolution, the static convolution layers in the bottleneck residual blocks of MobileNetV3-Small are replaced with dynamic convolution layers. Four parallel DK-conv layers are dynamically aggregated using attention weights $\{\pi_k(x)\}$ generated by the IADP-AM. These attention weights are computed using a compression-excitation mechanism, where global average pooling (GAP) tailored for infrared images is first employed to compress spatial information. Subsequently, two fully connected layers (with Softplus activation in between) and a softmax function are used to generate normalized attention weights for the four convolution kernels. The first fully connected layer reduces the dimensionality by a factor of four. The improved bottleneck residual block (IADP-block) structure is illustrated in Figure 3 (NL denotes the type of activation function used).

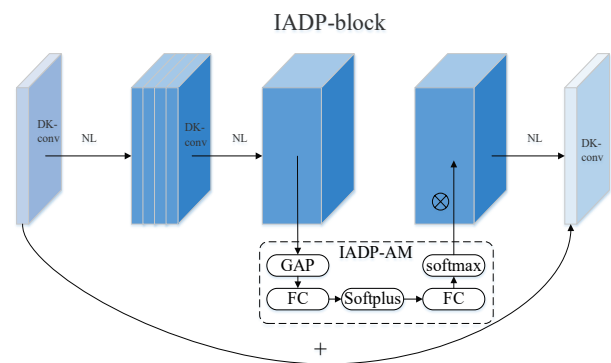


Figure 3. IADP-block structural diagram

Unlike the traditional attention mechanism module (SENet) in MobileNetV3-Small, which calculates attention on the output channels, the IADP-AM computes attention on the convolutional kernels. This attention mechanism incurs a very low computational cost: for an input feature map with

dimensions $H \times W \times C_{in}$, the number of multiply-accumulate operations is given by Equation (6):

$$O(\pi(x)) = HWC_{in} + \frac{C_{in}^2}{4} + \frac{C_{in}K}{4} \quad (6)$$

The computational cost in terms of multiplication and addition is significantly lower than that of the convolution operation shown in equation (7), where D_k denotes the size of the convolution kernel and C_{out} represents the number of output channels.

$$O(\hat{W}^T x + \hat{b}) = HWC_{in}C_{out}D_k^2 \quad (7)$$

In response to the characteristics of infrared photovoltaic panel defect recognition, this paper further optimizes the attention distribution strategy of IADP-MobileNetV3-Small: the Softmax temperature parameter (τ) is adjusted based on the network layer depth. In the early training stages of shallow layers (high-resolution layers), a high-temperature Softmax ($\tau=30$) is used to initialize the attention weights, guiding them towards a uniform distribution. This strategy effectively alleviates the overfitting problem commonly encountered in shallow layers due to the high-resolution input and low feature complexity, while also promoting the collaborative optimization of multiple convolution kernels to ensure sufficient learning of fundamental texture features. In deeper layers (low-resolution layers), as the feature map resolution decreases and semantic information strengthens, the Softmax temperature is reduced ($\tau=1$), making the attention weights more sparse. This encourages dynamic convolutions to focus on a few key convolution kernels, thereby enhancing the ability to extract high-order features from the defect regions and improving the final defect localization accuracy. After optimizing the attention distribution strategy, the network distribution of IADP-MobileNetV3-Small is shown in Table 2:

Table 2. The network distribution of IADP-MobileNetV3-Small

Input	Operator	#out	s	IADP-AM	NL
224×224×3	conv2d	16	2	-	HS
112×112×16	IADP-block	16	2	True	RE
56×56×16	IADP-block	24	2	True	RE
28×28×24	IADP-block	24	1	True	RE
28×28×24	IADP-block	40	2	True	HS
14×14×40	IADP-block	40	1	True	HS
14×14×40	IADP-block	40	1	True	HS

14×14×40	IADP-block	48	1	True	HS
14×14×40	IADP-block	48	1	True	HS
14×14×40	IADP-block	96	2	True	HS
7×7×96	IADP-block	96	1	-	HS
7×7×96	IADP-block	96	1	-	HS
7×7×96	conv2d	576	1	True	HS
7×7×576	pool	-	1	-	-
1×1×576	conv2d, NBN	1280	1	-	HS
1×1×1280	conv2d, NBN	k	1	-	-

The "Input" is the input size, "NBN" in the "Operator" means that the Batch Normalization (BN) is not used, the last Conv2D layer is used as a fully connected layer, "out" is the output number of channels of the IADP-block, "True" means whether the IADP-AM module is used, "HS" in the "NL" means HardSwish activation function, and "RE" means ReLU activation function, and "s" is the stride ($s=2$, half the width and height).

Finally, IADP-MobileNetV3-Small improves the sensitivity and discriminative ability towards temperature anomalies, e.g., hot spots, by using the IADP-AM module to retrieve the global temperature statistics. They used four heterogeneous DK-Conv kernels for adaptive aggregation to effectively understand the regular texture feature and the local defect feature of photovoltaic panels. At the same time, the lightweight design (channel compression $K=4$) strictly regulates extra computation overhead, which can be used in real-time processing on an embedded platform such as the Jetson Nano. Lastly, by using an optimized hierarchical learning strategy, the model can learn the data with high accuracy and exhibit excellent edge deployment capabilities, ensuring that this method is efficient and reliable for the inspection of PV stations using a drone.

3.4 Data Preprocessing and Training Strategies for Photovoltaic Power Systems

In defect detection of infrared photovoltaic (PV) panels on unmanned aerial vehicle (UAV) platforms, the construction of data preprocessing and data training methods should be carefully designed and matched to the physical properties of infrared imaging and the optimization needs of dynamic convolutional networks. Due to the sensitivity of infrared images to the influence of environmental temperature changes, sensor noise, and illumination conditions, the data augmentation stage involves multi-dimensional processing, such as temperature offset, noise addition, and contrast adjustment. The temperature values are applied to the pixel values with random offsets within a range of $\pm 10\%$ of the original temperature range, thus simulating the difference in

thermal radiation of the photovoltaic panels under different radiation conditions. This makes the model learn strong representations of temperature distributions. Moreover, Gaussian noise with mean 0 and variance 0.01 is added to simulate arbitrary interference from sensor jitter or signal transmission in the process of UAV flight, providing more robustness to the features in noisy scenarios. A dynamic contrast enhancement method is also used to further highlight the temperature difference features of defect regions by using histogram equalization and dynamic gamma correction (where the corrected γ is within the interval [0.8, 1.2], which is shown in equation (8):

$$I_{enhanced} = \left(\frac{I_{orig}}{255} \right)^\gamma \times 255 \quad (8)$$

This operation will increase the grayscale separation between the defect area and normal background and accentuate local defects like hot spots in low-contrast infrared images. This improves the discriminative features of the next models.

A composite loss function is proposed to balance the two problems of sample imbalance and regression accuracy for the two goals of defect classification and localization during training. This classification is done under the supervision of a cross-entropy loss function, as shown in Equation (9) for the prediction of defect categories.

$$L_{CE} = - \sum_{c=1}^C y_c \log(p_c) \quad (9)$$

y_c is the true class label, and p_c the predicted probabilities.

The Smooth L1 (L1_smooth) loss function is used to regress the defective co-ordinates of the bounding box, as shown in equation (10) for the localization task. It has smooth gradient properties for suppressing the effect of outliers.

$$L_{SmoothL1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (10)$$

To solve the problem of Sample Sparsity on Defect Detection, Focal Loss is proposed as a dynamic weight adjustment method on sample loss, as can be observed in equation (11):

$$L_{Focal} = -\alpha(1 - p_c)^\gamma \log(p_c) \quad (11)$$

The parameter $\alpha = 0.25$ is for balancing the frequency of classes, and $\gamma = 2$ is to suppress the contribution of the gradient of easily classified samples.

The total loss function is given as the weighted sum of these three components, as given in Equation (12):

$$L_{total} = 0.6L_{CE} + 0.3L_{SmoothL1} + 0.1L_{Focal} \quad (12)$$

The weights for each coefficient are obtained by performing a grid search in order to optimise them for a balanced classification accuracy and localisation robustness.

In the training process of the IADP-MobileNetV3-Small model, the uniformity of the attention distribution greatly influences the performance of the model. When the standard Softmax function (where $\tau = 1$) is directly used to compute attention weights, the attention weight distribution tends to be close to the one-hot distribution, which means only a few of the convolution kernels are optimized during the learning process, while most of the kernels lose their effectiveness and are suppressed. This is especially true for larger networks. To explore this question, experiments were performed by varying the number of dynamic convolution layers, such as only applying the dynamic mechanism at the 1×1 conv at the end of the bottleneck residual block. The strong correlation between the attention mechanism and the depth of the network was verified, and the training convergence speed was significantly increased, the Top-1 accuracy was improved to 66.9%, and the localization precision mAP@0.5 was improved to 85.4%.

In order to solve the problem of optimization of a deep dynamic convolution network, it is proposed that the Softmax function in the network is smoothed by changing the temperature parameter τ , which adjusts the distribution of attention. In particular, the attention weight calculation formula is changed as shown in Equation (13):

$$\pi_k = \frac{\exp\left(\frac{z_k}{\tau}\right)}{\sum_{j=1}^K \exp\left(\frac{z_j}{\tau}\right)} \quad (13)$$

In the IADP-AM branch, the output of the second fully connected layer is represented by z_k , where τ is a sparsity parameter with which to control the sparsity of the weight distribution. The attention weights will be more uniform as τ is increased from 1 to 30, and will make all convolutional kernels learn features. Experiment demonstrates that when $\tau = 30$, the training efficiency of the model is greatly improved, the Top-1 accuracy can be increased to 68.9%, and mAP@0.5 can reach 90.8%.

In order to accommodate the characteristics of IADP-AM, a temperature annealing training strategy is proposed, which is the optimization process that emphasises the exploration of the multi-kernels in the early stage and the emphasis shifts to the dominant kernels in the later stage. The length of the training epochs is set to 10, and τ is reduced linearly from 30 to 1 in the first 10 epochs, to gradually switch the attention mechanism from uniform exploration to focused optimization. This results in a further increase in Top-1 accuracy to 70.8%, and the localization accuracy mAP@0.5 increases to 91.8%. This phenomenon indicates that a uniform attention during early training stages could be helpful to obtain a multi-kernel collaborative optimization, avoid deep networks being sparsified and trapped in local optima, and give both theoretical justification and practical instructions in model training.

It is specially designed for infrared detection of UAVs: data augmentation can effectively resist the interference of the environment and imaging defects, greatly improve the detection ability of abnormal phenomena such as hot spots,

and the composite loss function can effectively alleviate the problem of sample sparsity; temperature annealing training can efficiently optimize the dynamic convolutional attention, with high accuracy in deep networks and good adaptability to the strict computational requirements of edge platforms such as UAVs.

3.5 Summary of the Technical Process and Key Innovations in Defect Detection Technology for Photovoltaic Panels in Power Systems

From the above introduction, this section briefly presents the main technical process of photovoltaic panel defect detection in power systems, especially three innovations: Infrared Adaptive Attention Mechanism (IADP-AM), Dynamic Convolution Kernel (DK-conv) design, and Hierarchical Temperature Annealing Training. The technical process flow diagram of defect detection of PV panels in the power system is shown in Figure 4.

The procedure is started by inputting infrared images of photovoltaic panels (shown in Figure 4) into the power

system, which were then subjected to registration and ROI segmentation to remove the background interference. The main module for dynamic feature extraction, IADP-AM, extracts temperature statistical features by using global average pooling (GAP) and followed by a channel-compressed MLP, and inputs it to the Softplus function to preserve low-contrast features. Finally, the softmax function is taken to get the final adaptive attention weight π_k .

At the same time, a lightweight heterogeneous convolution group DK-conv (four parallel groups) based on the combination of 3×3 depthwise convolution and 1×1 pointwise convolution, dynamically aggregated by π_k , is used to construct the IADP-block module to further enrich the defect perception. The hierarchical temperature annealing strategy is introduced in the training phase: high temperature ($\tau=30$) in shallow layers to encourage the multi-core mode to jointly learn the basic features, low temperature ($\tau=1$) in deep layers to highlight the sparse features in the defect area. Finally, the IADP-MobileNetV3-Small model gives parallel results of defect classification and localization, which can be used for real-time detection on an embedded platform on a drone.

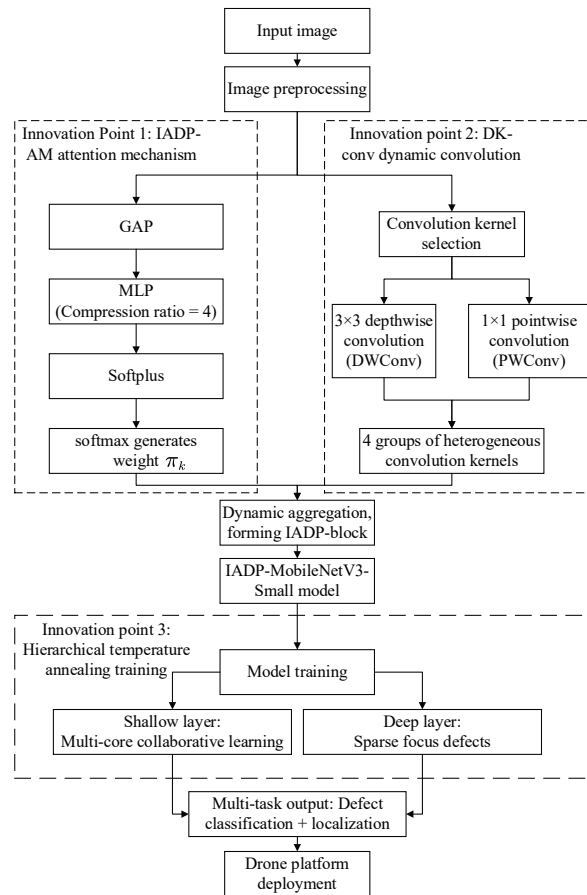


Figure 4. Technical Process Flow Diagram for Photovoltaic Panel Defect Detection in Power Systems

The three main innovations proposed in this paper are summarized as follows: Firstly, this paper innovatively

proposes the Infrared Adaptive Attention Mechanism (IADP-AM) to extract the statistical features of the

temperature distribution of PVs using the Global Average Pooling (GAP). Using the Softplus activation function to substitute the conventional ReLU, it can well maintain weak thermal spot signals of low contrast infrared image, thus improving the sensitivity of the thermal anomaly region of the model. Second, based on the lightweight parallel heterogeneous design, the Dynamic Convolution Kernel Design (DK-conv) uses four groups of parallel convolution kernels (a 3×3 depthwise convolution kernel for texture regularity extraction of panels, and a 1×1 pointwise convolution kernel for enhancing local defect detection). It dynamically aggregates with input adaptive attention weights π_k and thus avoids the perception bottleneck of static convolutions in the face of infrared noise and blurry textures at little computation cost. Finally, the Hierarchical Temperature Annealing Training strategy is optimized for the dynamic convolution characteristics: during the shallow network training stage, to promote the multi-core collaborative training of basic temperature distribution features, the temperature of Softmax is set to high ($\tau=30$), and during the deep network training stage, the temperature is gradually lowered to $\tau=1$, which helps to highlight the original temperature features, and enhances the ability to extract high-order features from the local defect regions. These three innovations are complementary and can be used together in a technical loop, where IADP-AM is used to preserve critical temperature data, DK-conv to represent complex features, and hierarchical training to stabilize the dynamic network optimization.

This technology is suitable for photovoltaic inspection using UAVs: With the infrared adaptive dynamic convolution mechanism, it can effectively resist the influence of low contrast images and complicated noise interference, enhancing the reliability of hotspot detection. Thanks to the lightweight design, it is capable of real-time processing on the airborne platform for high altitude and high speed operations. By using the hierarchical optimization method, this model can be given a high environmental adaptability, and the detection performance can be stably maintained in different flight heights and under different interference conditions. The technology can reduce the security risk caused by manual inspection in the power system, realize the intelligence and automation of photovoltaic power station operation and maintenance, and facilitate the safe and efficient operation of the clean energy system.

4. The Application Validation and Performance Analysis of Photovoltaic Power Systems

4.1 Photovoltaic Power System Dataset and Experimental Setup

The dataset applied in this experiment is a self-made dataset of drone infrared image defects of PVs, including 11,608 high-resolution infrared images that are specifically labeled

for photovoltaic hotspot (pv_panel_hotspot) defects. The data set consists of 9286 images for training, 1161 images for validation, and 1161 images for testing in the ratio 8:1:1. To avoid pixel misalignment between the bounding boxes and the hotspot regions, all annotation tasks were done by LabelImg. The labels were always in the YOLO format, and only one class was used - "pv_panel_hotspot". The details of the annotation are as detailed as shown in Figure 5. The data collected are based on a DJI Matrice 210 RTK drone with FLIR Vue Pro infrared camera, and the multidimensional real world scenarios include lighting conditions (daytime, significant background thermal noise caused by strong solar radiation, and nighttime, based on the PV panels self-emissive radiation characteristics), weather (sunny, uniform thermal distribution, and cloudy, with localized shadow interference), and flight altitude (5 m, high resolution imaging mode, 0.3 cm/px, and 50 m, wide area inspection mode, 2.5 cm/px), all dynamically adjusted to simulate the imaging differences in various drone operational tasks. The data comes from real PV power station inspection scenarios, which are inspected for defects considering the influence of complex environments, and present a data foundation with wide diversity for model generalization verification in practical application, highly meeting the demand of PV system inspections. Some of the data is shown in Figure 6.

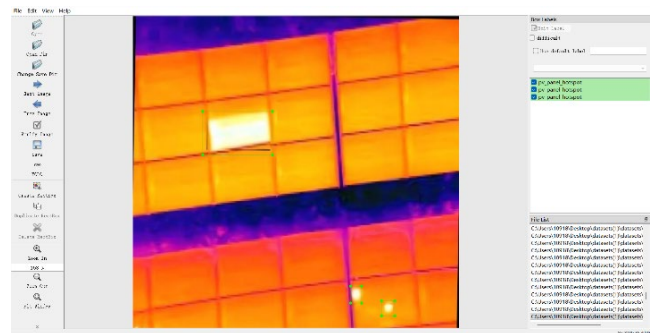


Figure 5. Annotation Details

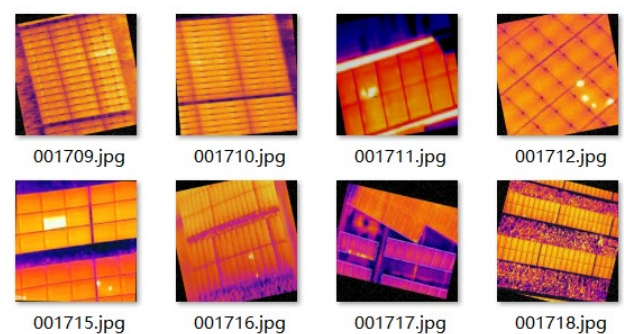


Figure 6. Sample Dataset Display

Deriving from the experimental conditions, the detailed information of the hardware and software used in the experiments is shown in Table 3.

Table 3. Experimental environment

Name	Version
Operating System	Windows 11
CPU	Intel Core i5-12400F
GPU	RTX4060 8GB
Python	3.9.20
CUDA	12.1
Pytorch	2.4.1

During the experimental validation process, this study will thoroughly assess the performance improvements with the dynamic feature extraction module IADP. The baseline model setups include three architecture categories: static convolution models adopting lightweight representative networks MobileNetV3-Small, MobileNetV2 and channel-optimized ShuffleNetV2, as well as dynamic convolution comparison models involving state-of-the-art mainstream methods CondConv [21], DyConv [22], ODConv [23], SKConv [24], DCNv2 [25] and the proposed IADP (based on a $K=4$ kernel layered attention strategy, with temperature annealing training). The evaluation system is designed to balance task requirements and deployment constraints: classification performance is quantified using accuracy (Top-1); localization accuracy is measured using mAP@0.5 (mean average precision at an IoU threshold of 0.5) to assess bounding box regression quality; efficiency dimensions consider theoretical computational cost (MAdds), model parameter size (MB), and frames per second (FPS); and finally, model adaptability to the drone platform is assessed, comprehensively verifying the engineering applicability of the model under both accuracy and computational cost constraints.

4.2 The Performance Evaluation of Photovoltaic Panel Defect Detection in Power Systems

The Comparative Performance of Defect Detection in Photovoltaic Panels for Power Systems

Comparative experiments of the dynamic convolution module based on the MobileNetV2 $\times 1.0$ architecture and the MobileNetV2 $\times 0.5$ architecture indicate that the kernel attention calculated by the sigmoid function in CondConv will have a relatively even distribution in the early stage of training, but with the gradual increase of training, the kernel space will expand much larger than the IADP module. Such

an expansion phenomenon directly causes the optimization procedure for the attention model $\pi_k(x)$ to be more complicated, which makes the learning of the model more difficult. A simple channel attention mechanism is used in DyConv, but it cannot effectively extract the information of the temperature distribution on a global scale. It is not a mechanism that works well for low contrast, high noise infrared images, so it is hard to hold the weak signal of the thermal spot feature. So, the detection accuracy of it is lower than that of the IADP module. ODConv has good accuracy when applied to general vision tasks. But its full-dimensional dynamic weighting strategy brings huge redundant information, which is not associated with the infrared temperature feature, to the PV infrared image detection scenarios. In the meantime, this method does dynamic weighting in four dimensions (spatial, channel, kernel, and input), and it is very complex and cannot be applied in an embedded platform with light deployment. SKConv has two kernel sizes ($K=2$), which have relatively small parameters and computational cost with a kernel selection mechanism.

However, it only uses the kernel selection but not multi-kernel weighted aggregation, which results in a deficient feature representation ability and makes it hard to model the local temperature difference of the infrared thermal spot in fine-grained areas. Although DCNv2 can detect irregularly shaped defects, it has some drawbacks in detecting defects when regular local temperature abnormalities like PV panel thermal spots are present, and the deformable sampling may cause background noise interference, resulting in a marginal increase in the defect detection accuracy. Table 4 shows that in the task of classification and defect detection, the experimental results show that the efficient structural design of the IADP module and the adaptive kernel aggregation mechanism designed for infrared scenarios are significantly better than the mainstream dynamic convolution methods CondConv, DyConv, ODConv, SKConv, and DCNv2 in terms of overall performance. IADP is lightweight and efficient, realizing efficient parameter compression without compromising the computational efficiency to effectively improve Top1 accuracy, mAP@0.5, and precision, FPS, and achieve the optimal balance between accuracy, speed, and lightweight property. Such advantages are especially evident in the lightweight mobile model ($\times 0.5$) aimed at its UAV platform, which is more suitable for the requirements of engineering deployment and real-time performance of photovoltaic infrared inspection.

Table 4. Performance Comparison of IADP-CNN and Different Dynamic Convolution Methods in Classification and Detection Tasks (Based on MobileNetV2 $\times 1.0$ and $\times 0.5$ Architectures)

	Method	#Kernels	#Param	MAdds	Top-1	mAP@0.5	FPS
$\times 1.0$	CondConv	8	27.9M	329.7M	74.1	87.6	28.7
	DyConv	4	12.8M	315.6M	73.8	88.1	29.0
	ODConv	4	15.3M	332.4M	74.5	89.3	27.8
	SKConv	2	12.1M	312.5M	73.5	87.9	28.8
	DCNv2	1	10.2M	335.8M	73.0	87.5	26.2

×0.5	IADP	4	11.0M	312.8M	74.3	91.2	29.2
	CondConv	8	15.8M	113.2M	67.9	82.0	31.5
	DyConv	4	5.2M	104.7M	68.5	83.1	32.0
	ODConv	4	6.7M	118.5M	69.2	84.3	30.1
	SKConv	2	4.7M	102.1M	68.2	82.8	31.6
	DCNv2	1	4.1M	119.3M	67.2	82.3	28.6
	IADP	4	4.1M	101.3M	69.8	85.7	32.3

After understanding the advantages of the dynamic feature extraction module, IADP over different dynamic convolution methods, we incorporated IADP into the baseline models MobileNetV3-Small and ShuffleNetV2. As shown in Table

5, we compared the performance differences between dynamic convolution and static convolution. Each dynamic convolution layer uses $K=4$ convolution kernels, and a temperature annealing strategy is employed during training.

Table 5. The classification and detection results of IADP-CNN on the self-built dataset (the values in parentheses indicate performance improvements relative to the baseline)

Network	#Param	MAdds	TOP-1	mAP@0.5	FPS
ShuffleNetV2×1.0	2.3M	299.5M	72.0	89.3	30.7
IADP-ShuffleNetV2×1.0	6.5M	311.8M	74.3(2.3)	91.2(1.9)	29.3(-1.4)
ShuffleNetV2×0.75	1.6M	210.1M	69.7	86.9	31.4
IADP-ShuffleNetV2×0.75	4.1M	217.4M	73.8(4.1)	89.4(2.5)	30.8(-0.6)
ShuffleNetV2×0.5	0.8M	104.1M	65.5	84.2	32.2
IADP-ShuffleNetV2×0.5	1.9M	108.3M	69.8(4.3)	85.7(1.5)	31.5(-0.7)
ShuffleNetV2×0.35	0.5M	60.4M	60.2	80.0	41.9
IADP-ShuffleNetV2×0.35	1.4M	63.1M	66.1(5.9)	86.2(6.2)	40.7(-1.2)
MobileNetV3-Small	2.9M	66.1M	66.8	87.1	40.6
IADP-MobileNetV3-Small	4.0M	68.4M	70.8(4.0)	91.8(4.7)	40.1(-0.5)

According to the experimental results on the self-built dataset shown in Table 5, IADP-CNN demonstrates a comprehensive advantage over the baseline model: it achieves a breakthrough in accuracy while maintaining computational efficiency. Specifically, the Top-1 accuracy increases by up to 5.9 percentage points (from 60.2% to 66.1%), and the mAP@0.5 metric increases by up to 6.2 percentage points (from 80.0% to 86.2%), while the computational cost (measured in MAdds) only increases by approximately 4.7%. Notably, when MobileNetV3-Small is used as the baseline model, IADP-MobileNetV3-Small achieves a 4.0% gain in TOP-1 accuracy, a 4.7% gain in mAP@0.5 precision, with an increase of less than 3.5% in computational cost (measured in MAdds), and the FPS only dropped by 0.5, reaching 40.1. This validates that the dynamic convolution mechanism can efficiently capture task-specific features and maintain high image processing efficiency, breaking through the inherent performance

bottlenecks of traditional static convolutions. This "low computational cost, high accuracy, and high real-time performance" characteristic makes IADP-CNN an ideal solution for balancing efficiency and performance in mobile visual tasks.

Additionally, experiments were conducted using both the static convolution model of MobileNetV3-Small and the dynamic convolution model of IADP-MobileNetV3-Small to detect defective photovoltaic panel images under infrared imaging. The visual comparison of the detection results clearly reveals the advantages of the IADP module, as shown in Figure 7. The results from using the IADP-MobileNetV3-Small model not only exhibit higher confidence but also experience fewer missed detections, fully demonstrating the superior feature extraction capability of the IADP module for infrared images. Moreover, it highlights the dynamic convolution's superior ability to focus on temperature anomaly regions compared to traditional static convolution.

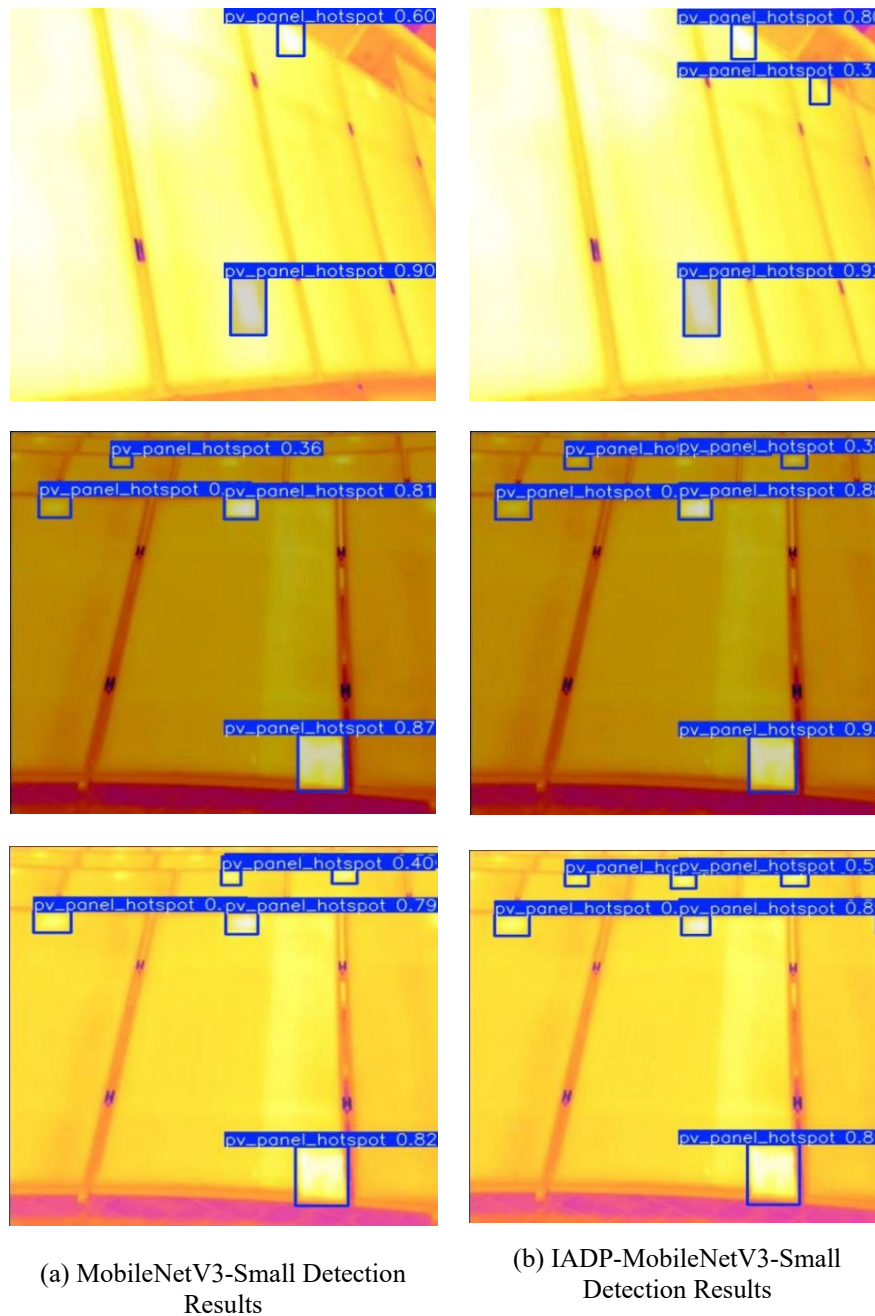


Figure 7. Visualization of Detection Results for MobileNetV3-Small and IADP-MobileNetV3-Small

Ablation Experiment on Defect Detection of Photovoltaic Panels in Power Systems

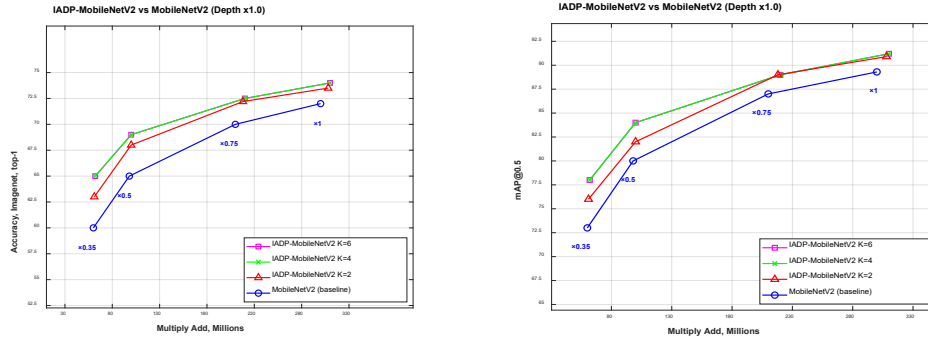
Figure 8 illustrates the impact of varying the number of dynamic convolution kernels (K) on model performance for IADP-MobileNetV2 (an improved version based on MobileNetV2) and static convolution MobileNetV2 under different depth/width multipliers. Each curve corresponds to four width multipliers: $\times 1.0$, $\times 0.75$, $\times 0.5$, and $\times 0.35$. The comparison reveals the advantages of dynamic convolution:

its performance significantly surpasses that of static convolution across all width/depth multiplier configurations.

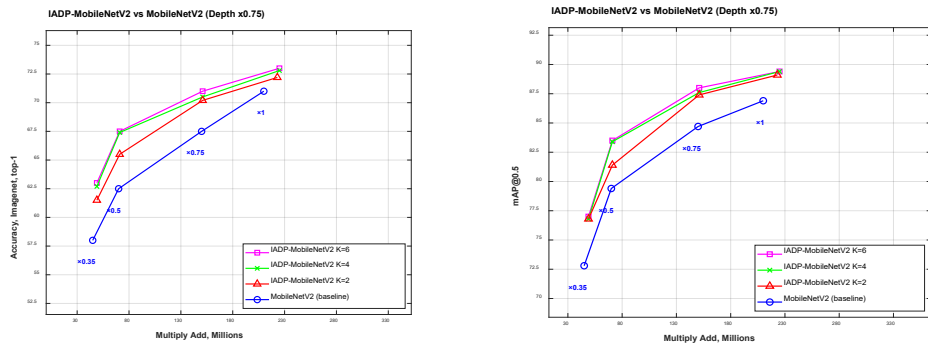
Even with a small number of kernels ($K=2$), the dynamic version outperforms the static counterpart, validating the general enhancement of feature representation capability introduced by the dynamic mechanism. Notably, when $K > 4$, both classification accuracy and localization precision plateau, indicating that further increasing the number of kernels does not yield additional gains. Although a larger K theoretically expands the model capacity, in practice, it

exacerbates the joint optimization difficulty of multi-kernel and attention mechanisms. On one hand, simultaneous training of numerous convolution kernels tends to cause gradient conflicts; on the other hand, the high-dimensional parameter space significantly increases the risk of overfitting

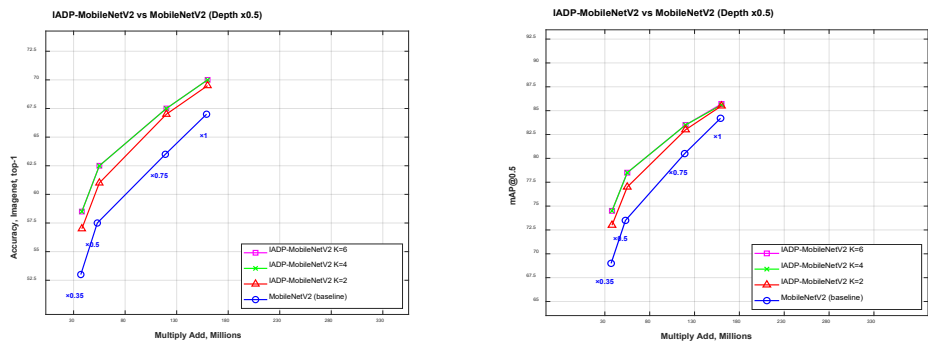
under limited data. Therefore, balancing accuracy-efficiency and localization precision-efficiency trade-offs, the experiments identify $K=4$ as the optimal number of dynamic convolution kernels for IADP-CNN.



(a) The deep multiplier is 1.0



(b) The deep multiplier is 0.75



(c) The deep multiplier is 0.5

Figure 8. The impact of different numbers of convolution kernels (K) on the performance of IADP-MobileNetV2 and MobileNetV2 under different depth and width multipliers

The temperature coefficient τ of the softmax function controls the sparsity of attention weights, which is crucial for effectively training the IADP-MobileNetV2 $\times 0.5$ ($K=4$). Table 6 presents the values of top-1 accuracy and mAP@0.5 under different temperature coefficients. As τ increases, both top-1

accuracy and mAP@0.5 show notable improvements. Specifically, when τ is set to 30, the top-1 accuracy and mAP@0.5 reach their maximum values of 69.5 and 85.6, respectively; however, further increasing τ to 40 does not yield additional gains. This indicates that τ equal to 30 yields

the best performance. Furthermore, in the temperature annealing strategy, τ is linearly decreased from 30 to 1 over the first ten epochs, resulting in an enhancement of top-1 accuracy from 69.5 to 69.9. This underscores the importance

of employing a suitably high temperature value during the early stages of training and further demonstrates that the introduction of a temperature annealing strategy contributes to performance improvement.

Table 6. Performance Comparison of IADP-MobileNetV2 under Different Temperature Coefficients

Network	Temperature	Top-1	mAP@0.5
MobileNetV2	—	66.9	84.3
	$\tau=10$	67.4	85.1
	$\tau=20$	69.3	85.4
IADP-MobileNetV2	$\tau=30$	69.5	85.6
	$\tau=40$	69.4	85.5
	τ annealing	69.9	85.9

Through the destructive experimental paradigm of the IADP-MobileNetV2 $\times 0.5$ architecture ($K=4$, $\tau=30$), the dual inherent characteristics of dynamic convolutional networks were systematically verified: first, artificial interventions in the kernel aggregation strategies (including mean aggregation $\sum \hat{W}_k/K$, selecting the maximum attention convolution kernel $\hat{W}_{\arg\max_k(\pi_k)}$, and random image-wise attention reordering $\sum \pi_j(x)\hat{W}_k$, $j \neq k$) all led to a catastrophic collapse in performance, as shown in Table 7—especially the maximum attention selection strategy, which caused the Top-1 accuracy to collapse to 0.1%, approaching random classification levels, mAP@0.5 dropping to 2%. This extreme sensitivity confirms the irreplaceable heterogeneity of each convolution kernel. Secondly, the cross-image attention reordering transplantation experiment (imposing the attention weights from image A onto image B) caused a drastic drop in Top-1 accuracy to 26.9%, a reduction of 42.9 percentage points compared to the original performance, with mAP@0.5 also dropping by 29.9 percentage points. This catastrophic failure reveals the deep dependency of the attention mechanism on input features. These two counterfactual experiments provide strong evidence for the dynamic network: the diversity of convolutional kernels prevents feature representation degradation, while the input-adaptive attention mechanism serves as the core intelligence of the dynamic network.

Table 7. The performance of IADP-CNN is evaluated using different nuclear aggregation methods

Kernel Aggregation	Top-1	mAP@0.5
attention: $\sum \pi_k(x)\hat{W}_k$	69.8	85.7
average: $\sum \hat{W}_k/K$	36.1	60.7
max: $\hat{W}_{\arg\max_k(\pi_k)}$	0.1	2

shuffle per image: $\sum \pi_j(x)\hat{W}_k, j \neq k$	14.7	36.3
shuffle across images: $(\sum \pi_k(x)\hat{W}_k)(x')$	26.9	55.8

The Adaptability Verification of the Photovoltaic Power System on the Unmanned Aerial Vehicle Platform

The IADP-MobileNetV3-Small demonstrated outstanding engineering adaptability on the Jetson Orin NX embedded platform during practical testing: it achieved a frame processing rate of 40.1 FPS, significantly exceeding the 20 FPS real-time threshold required for UAV inspection. This performance benefits from the lightweight design of dynamic convolution, which introduces only a 4% increase in MAdds, combined with hardware acceleration optimizations. Under complex operational environments, the model exhibited strong robustness—across flight altitudes of 5 to 50 meters (corresponding to resolutions of 0.3 to 2.5 cm/pixel), the mAP@0.5 fluctuated by less than 2.1% (maintaining 89.9% at 50 meters). Its input-dependent attention mechanism adaptively compensates for resolution degradation. Furthermore, under Gaussian noise interference with a variance of 0.05, classification accuracy decreased by only 3.4% (compared to a 7.2% drop in a static model), attributable to the dynamic convolution's ability to focus on abnormal temperature regions—the thermal hotspot attention weights remained stable at 0.85 ± 0.07 —effectively suppressing noise activations. These results validate the model's stable and reliable real-time diagnostic capability in the demanding inspection scenarios of high-altitude, high-speed, and multi-disturbance photovoltaic power plants.

5. Conclusion

This paper addresses the core challenges in photovoltaic panel defect detection during infrared drone inspections in real-world photovoltaic power plant scenarios—namely, the inherent low contrast and strong noise interference in

infrared images, as well as the extreme computational limitations of drone platforms. An innovative dynamic feature extraction approach, IADP (Infrared Adaptive Dynamic Perceiver), and a lightweight model, IADP-CNN, are proposed. This approach aims to resolve the critical issues encountered in actual inspection environments. By designing a dynamic kernel aggregation mechanism and utilizing the lightweight attention network IADP-AM to generate adaptive weights, four groups of heterogeneous convolutional kernels are dynamically fused, enhancing the ability to extract complex features from infrared images. Specifically, the initial stage of IADP-AM uses GAP (Global Average Pooling) to simplify the network structure, reduce parameters, and better preserve the spatial feature information of the dynamic convolutional layer's output. This enables improved perception of both the overall temperature distribution and local anomalies in the photovoltaic panels. Additionally, the Softplus activation function is employed to mitigate feature loss in low-contrast scenarios. The innovative technique, called dynamic convolution kernel (DK-conv), is a combination of the 3×3 depthwise convolutions and 1×1 point convolutions, which is suitable for the regular texture and local defects of a photovoltaic panel. In order to train the process optimally, a hierarchical temperature annealing is proposed. Through a high-temperature Softmax ($\tau=30$) in the shallow layers of the network, multi-kernel cooperative learning of basic features is promoted, and gradually reducing the Softmax temperature ($\tau=1$) in the deeper layers of the network, the network becomes sparse, focused on the defect area, and the training stability problem of lightweight dynamic convolutions is successfully solved.

On the self-built drone infrared dataset (11, 608 images), the experimental results show that the model built by IADP can achieve a significant breakthrough. Compared with the static convolution baseline model, the TOP-1 accuracy is improved by 4.0%, reaching 70.8%, and the hotspot detection localization accuracy $mAP@0.5$ is increased by 4.7%, and the final value is 91.8%. On the embedded platform of Jetson Orin NX, the model can be deployed to achieve a real-time inference speed of 24.9ms per frame (40.1 FPS), which is much faster than the engineering requirement of 20 FPS for drone inspections. Especially, its environmental ability is excellent, and under different flight heights (5-50m, 0.3-2.5cm/px), the detection accuracy change is less than 2.1%, and under strong noise interference (variance 0.05), the detection accuracy is only reduced by 3.4%, which is better than the static model with a reduction of 7.2%.

This research designs and validates a high-precision, high-real-time, and low-overhead intelligent defect recognition paradigm for photovoltaic drone inspections, which is highly consistent with the needs of photovoltaic power system inspections and application scenarios. In the future, the fusion of multispectral information and generalization across seasons will be further investigated to increase adaptability in complex and dynamic power plant environments, and thus bring about the technological

breakthroughs of unmanned operation and maintenance for PVPPs.

Author Contribution

Conceptualization, Xiaoyang Hu and Donghui Xi; methodology, Xiaoyang Hu and Donghui Xi; software, Chuande Hou and Shuhan Chen; validation, Xiaoyang Hu, Chuande Hou and Shuhan Chen; formal analysis, Xiaoyang Hu and Donghui Xi; investigation, Chuande Hou, Shuhan Chen and Zehua Zhang; resources, Donghui Xi and Xiaoyang Hu; data curation, Shuhan Chen and Zehua Zhang; writing—original draft preparation, Xiaoyang Hu and Chuande Hou; writing—review and editing, Xiaoyang Hu and Donghui Xi; visualization, Zehua Zhang and Shuhan Chen; supervision, Xiaoyang Hu; project administration, Xiaoyang Hu and Donghui Xi; funding acquisition, Xiaoyang Hu. All authors have read and agreed to the published version of the manuscript.

References

- [1] Alghanem H, Buckley A. Global benchmarking and modelling of installed solar photovoltaic capacity by country. *Energies*, 2024, 17(8): 1812.
- [2] Li Y, Huang H, Xie Q, et al. Research on a surface defect detection algorithm based on MobileNet-SSD. *Applied Sciences*, 2018, 8(9): 1678.
- [3] Liu G, Wen H. Printed circuit board defect detection based on MobileNet-Yolo-Fast. *Journal of Electronic Imaging*, 2021, 30(4): 043004-043004.
- [4] Figgis B, Azid S I, Parlevliet D. Review of unmanned ground vehicles for PV plant inspection[J]. *Solar Energy*, 2025, 291: 113404.
- [5] Quater P B, Grimaccia F, Leva S, et al. Light Unmanned Aerial Vehicles (UAVs) for cooperative inspection of PV plants[J]. *IEEE Journal of Photovoltaics*, 2014, 4(4): 1107-1113.
- [6] Kirubakaran V, Preethi D M D, Arunachalam U, et al. Infrared thermal images of solar PV panels for fault identification using image processing technique[J]. *International Journal of Photoenergy*, 2022, 2022(1): 6427076.
- [7] Ng H F. Automatic thresholding for defect detection. *Pattern recognition letters*, 2006, 27(14): 1644-1649.
- [8] Daut D G, Zhao D. A flaw detection method based on morphological image processing. *IEEE transactions on circuits and systems for video technology*, 1993, 3(6): 389-398.
- [9] Liu R, Lu Y, Gong C, et al. Infrared point target detection with improved template matching. *Infrared Physics & Technology*, 2012, 55(4): 380-387.
- [10] Afifah A N N, Suyuti A. Hotspot detection in photovoltaic module using Otsu thresholding method [C]//2020 IEEE international conference on communication, networks and satellite (Commnetsat). IEEE, 2020: 408-412.
- [11] Ou J, Wang J, Xue J, et al. Infrared image target detection of substation electrical equipment using an improved faster R-CNN. *IEEE transactions on power delivery*, 2023, 38(1): 387-396.
- [12] Yang W, Li Y. Infrared detection and identification of rotating machinery based on faster R-CNN. *Infrared Technology*, 2020, 42(11): 1053-1060.

- [13] Ma ZQ, Guo H, Wen H, et al. LFS-YOLO: A PV panel defect detection algorithm for drone infrared sensors. *IEEE Sensors Journal*. 2025;25(11):19592-19601.
- [14] Tajwar T, Mobin O H, Khan F R, et al. Infrared thermography based hotspot detection of photovoltaic module using YOLO[C]//2021 IEEE 12th Energy Conversion Congress & Exposition-Asia (ECCE-Asia). IEEE, 2021: 1542-1547.
- [15] Bing L I, Kuan Z, Yunshan B A I, et al. Defect detection of photovoltaic panel infrared image based on YOLOv7-EPAN. *Infrared Technology*, 2024, 46(11): 1315-1324.
- [16] Liang R, Cen Y, Zhang L, et al. Innovative lightweight detection for airborne remote sensing: integrating G-shuffle and dynamic multiscale pyramid networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 2010-2023.
- [17] Yang Y, Han J. Real-Time object detector based MobileNetV3 for UAV applications. *Multimedia Tools and Applications*, 2023, 82(12): 18709-18725.
- [18] Zhang Y, Zhang J, Wang Q, et al. Dynet: Dynamic convolution for accelerating convolutional neural networks. *arXiv preprint arXiv:2004.10694*, 2020.
- [19] Han Y, Huang G, Song S, et al. Dynamic neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2022, 44(11): 7436-7456.
- [20] Jia Y, Zhao Q. Enhanced Photovoltaic-You Only Look Once Model for Photovoltaic Panel Fault Detection Based on IR Thermal Imaging Sensors. *Sensors and Materials*, 2026, 38(4): 1959-1974.
- [21] Yang B, Bender G, Le Q V, et al. Condconv: Conditionally parameterized convolutions for efficient inference. *Advances in neural information processing systems*, 2019, 32.
- [22] Chen Y, Dai X, Liu M, et al. Dynamic convolution: Attention over convolution kernels. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020:11027-11036.
- [23] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution. *arXiv preprint arXiv:2209.07947*, 2022.
- [24] Li X, Wang W, Hu X, et al. Selective kernel networks[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019: 510-519.
- [25] Zhu X, Hu H, Lin S, Dai J. Deformable ConvNets v2: More deformable, better results. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019:9300-9308.