# Dual-path model for irrigation facility detection via UAV remote sensing

Fahu Xu[1,2,3], Jiawei Jiang[1,2,3], Yihui Wang[1,2,3], Shunhang Zhou[1,2,3], Baofeng Su[1,2,3*]

(1. *College of Mechanical and Electronic Engineering, Northwest A & F University, Yangling 712100, Shaanxi, China*;
2. *Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling 712100, Shaanxi, China*;
3. *Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling 712100, Shaanxi, China*)

**Abstract:** Accurate and efficient condition detection of irrigation facilities is crucial for ensuring stable crop production and enhancing food security. However, traditional detection methods often fail to meet the demands for refined monitoring in large-scale irrigation areas. To address this issue, this study proposed an irrigation facility condition detection system based on UAV low-altitude remote sensing and the DP-ResNet50 model to identify four common states of irrigation canal systems. First, with ResNet-50 as the backbone network, the CalibratedFocal Loss function was adopted to mitigate biases arising from imbalanced class distributions and reduce overconfident predictions in ambiguous scenarios. Second, the Dual-Path and Fusion Collaborative Module (DP-FCM) was designed to enhance the ability to capture differentiated local features, such as silt and gravel, through the complementary combination of a general backbone and a lightweight discriminative path. Finally, to further address the high-dimensional redundancy and low-dimensional confusion in irrigation canal state features, the Enhancement and Classifier Integrated Module (EC-IM) was proposed to realize more effective feature mapping and state discrimination. Experimental results demonstrate that the DP-ResNet50 model achieved favorable performance in irrigation canal condition detection, with an accuracy of 0.9413, alongside macro precision, macro recall, and macro F1-score of 0.9250, 0.9195, and 0.9211, respectively. Furthermore, compared to other classic models, the DP-ResNet50 model demonstrates higher and more balanced recall rates for individual classes. The results of this study can provide a promising solution for the intelligent monitoring of smart agricultural irrigation systems.

**Keywords:** irrigation facilities, condition detection, UAV, DP-ResNet50

**DOI:** 10.25165/j.ijabe.20261904.10532

**Citation:** Xu F H, Jiang J W, Wang Y H, Zhou S H, Su B F. Dual-path model for irrigation facility detection via UAV remote sensing. Int J Agric & Biol Eng, 2026; 19(4): 174–183.

## 1 Introduction

In recent years, the frequent occurrence of extreme drought events has severely impacted agricultural production. As critical infrastructure for ensuring stable crop yields and food security, the operational status of irrigation facilities directly influences agricultural water use efficiency, irrigation uniformity, and crop health[1,2]. However, due to long-term lack of maintenance and inadequate supervision, irrigation canals commonly suffer from problems such as garbage blockages and silt accumulation, which significantly impair the rational allocation and efficient use of water resources[3,4]. Traditional manual inspection is constrained by complex terrain, low efficiency, and subjective evaluation, making it unsuitable for large-scale irrigation district management[5]. Satellite remote sensing technology supports large-scale monitoring, yet the relatively low resolution and high cloud-cover susceptibility of satellite imagery impede compliance with the requirements for refined monitoring of irrigation canals[6]. The deployment of IoT-based sensor monitoring networks faces challenges such as high deployment costs, maintenance difficulties, and poor signal stability in complex farmland environments. As a result, such sensor monitoring networks are primarily suitable for critical infrastructure such as urban water supply canals[7,8].

Unmanned Aerial Vehicles (UAVs) possess high maneuverability, wide coverage, and flexible operational capabilities, constituting efficient technological carriers in the field of agricultural monitoring. Relevant studies corroborate the application potential of UAVs in irrigation canal condition detection. Traore et al.[9] utilized a multi-sensor UAV system for irrigation canal leakage detection, successfully achieving precise identification of leakage points and exploration of unknown leakage areas. Jalajamony et al.[10] employed UAV-based image recognition technology integrated with GPS positioning to accomplish accurate geolocation of localized drought-prone areas in farmland, further demonstrating the adaptability of this technology in agricultural contexts. Leveraging the inherent advantages of UAVs, UAV-based image recognition technology enables rapid coverage of extensive irrigation areas, efficient detection of target conditions, and significant reduction in labor and time costs[11]. However, existing research exhibits notable shortcomings. This body of research relies heavily on multi-sensor fusion schemes, which give rise to elevated hardware and data processing costs. In addition, focus is limited to drought monitoring within farmland environments, thereby neglecting the detection of the operational status of irrigation canals. Both limitations hinder their applicability to the practical needs of large-scale irrigation canal condition monitoring. Therefore, this study leverages the core strengths of UAVs, adopts UAV-based image recognition as the primary method for irrigation canal condition detection, and aims to enhance the performance of UAV-based image recognition for this specific scenario through

**Received date:** 2026-02-24 **Accepted date:** 2026-06-08

**Biographies: Fahu Xu**, Lecturer, research interest: remote sensing, Email: fahu.xu@nwafu.edu.cn; **Jiawei Jiang**, MS candidate, research interest: agricultural engineering, Email: jiawei.jiang@nwafu.edu.cn; **Yihui Wang**, MS candidate, research interest: agricultural machinery equipment and technology, Email: 2296307988@qq.com; **Shunhang Zhou**, MS candidate, research interest: agricultural engineering, Email: shunhang.zhou@nwafu.edu.cn.

***Corresponding author: Baofeng Su**, PhD, Professor, research interest: smart agriculture. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, Shaanxi, China. Tel: +86-15609279711, Email: bfs@nwafu.edu.cn.

algorithmic optimization.

Early image recognition predominantly relied on traditional machine learning methods to extract morphological features of irrigation canals, including the color, texture, and shape of surface defects[12,13], as well as grayscale spectral information[14,15]. Typical algorithms include Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN)[16-18]. However, traditional machine learning is heavily reliant on manual feature extraction, a process that requires labor-intensive feature screening and is susceptible to environmental factors, resulting in insufficient feature robustness and difficulty in guaranteeing detection and classification accuracy[19]. Deep learning with end-to-end automatic feature extraction capability has been widely applied in agricultural image recognition, and CNNs have become the mainstream for classification and detection due to their superior spatial feature learning ability[20-23]. Among such architectures, the Residual Network (ResNet) effectively addresses the gradient vanishing problem in deep networks by introducing residual connections, significantly enhancing the model's feature learning capability[24]. Upadhyay et al.[25] proposed an enhanced ResNet-50, which achieved 95% accuracy in tomato leaf disease detection by integrating advanced data augmentation techniques and transfer learning strategies and demonstrated robustness against environmental variations such as illumination changes and viewing angles. Ahmed et al.[26], aiming at the disease classification of date palm images, designed a basic ResNet architecture and an Inception ResNet transfer learning model to build a disease recognition system, effectively accomplishing the classification of different date palm disease types. Haggab et al.[27] utilized YOLOv7 models to detect pipeline and tower defects in center pivot irrigation systems. The YOLOv7x model combined with the RGB system achieved a 95% detection rate with a precision of 0.798 and an F1-score of 0.954. However, this study only realized binary classification between safe and dangerous states. Consequently, manual verification remains necessary, making it difficult to meet the practical demand for precise fault classification of irrigation facilities. Although the aforementioned studies have made progress in irrigation canal detection, research on the integration of deep learning with irrigation canal condition detection remains relatively scarce. For irrigation canal scenarios, challenges including complex background interference, imbalanced sample distribution, and feature complexity still require in-depth investigation.

To address the aforementioned challenges, this study proposes an irrigation canal condition detection system based on UAV low-altitude remote sensing images. Specifically, a dual-path model named DP-ResNet50 is proposed. The model adopts CalibratedFocal Loss to alleviate distribution imbalance and overfitting, and integrates the Dual-Path and Fusion Collaborative Module (DP-FCM) and the Enhancement and Classifier Integrated Module (EC-IM) to capture complex features, thereby improving canal condition detection performance. A corresponding dataset is constructed from UAV-captured low-altitude remote sensing images of irrigation canals for model training and evaluation.

## 2 Materials and methods

### 2.1 Dataset preparation

The operational status of irrigation facilities is complex and diverse, with classification standards varying to some extent across different application scenarios. To enhance the adaptability and robustness of operational status recognition algorithms in multi-type and multi-scale practical scenarios, this study selected the Jiaokou Chouwei Irrigation District in Weinan City, Shaanxi Province, China, as the experimental area and conducted field data collection on the operational status of outdoor farmland irrigation canal systems. As illustrated in Figure 1, a DJI Phantom 4 RTK UAV was employed for low-altitude image acquisition. Since this study focuses on irrigation canal condition detection, a flight altitude of 1.2-1.5 m, a flight speed of 0.3-0.5 m/s, and a gimbal angle of 57° downward were configured. Low-altitude and low-speed flight can effectively prevent the loss of subtle features such as texture and edge morphology and ensure consistency in dataset collection. As listed in Table 1, six representative sampling sites with diverse canal structures were selected to ensure dataset diversity. After filtering out blurred and unrecognizable images, a total of 2247 clear irrigation canal status images were obtained. These images cover four conditions, including Normal Circulation (NLC), Silt Clogging (STC), Gravel Clogging (GLC), and Garbage Clogging (GEC), with representative images for each category shown in Figure 2. On this basis, the original images were first cropped and then augmented through data enhancement techniques such as rotation, brightness, and contrast adjustment, resulting in a final total of 3749 clear irrigation canal status images. The dataset was randomly partitioned into training and test sets at a 7:3 ratio, comprising 2590 and 1159 images, respectively. Figure 3 presents the distribution of irrigation canal status categories across both sets.

### 2.2 DP-ResNet50 model construction

ResNet-50 was selected as the baseline, yet this baseline architecture presents limitations in adapting to class imbalance and local feature variations in irrigation canal detection. To address these limitations, this study proposed the DP-ResNet50 model as illustrated in Figure 4, which incorporates optimization strategies for multi-class scenarios and fine-grained feature extraction.

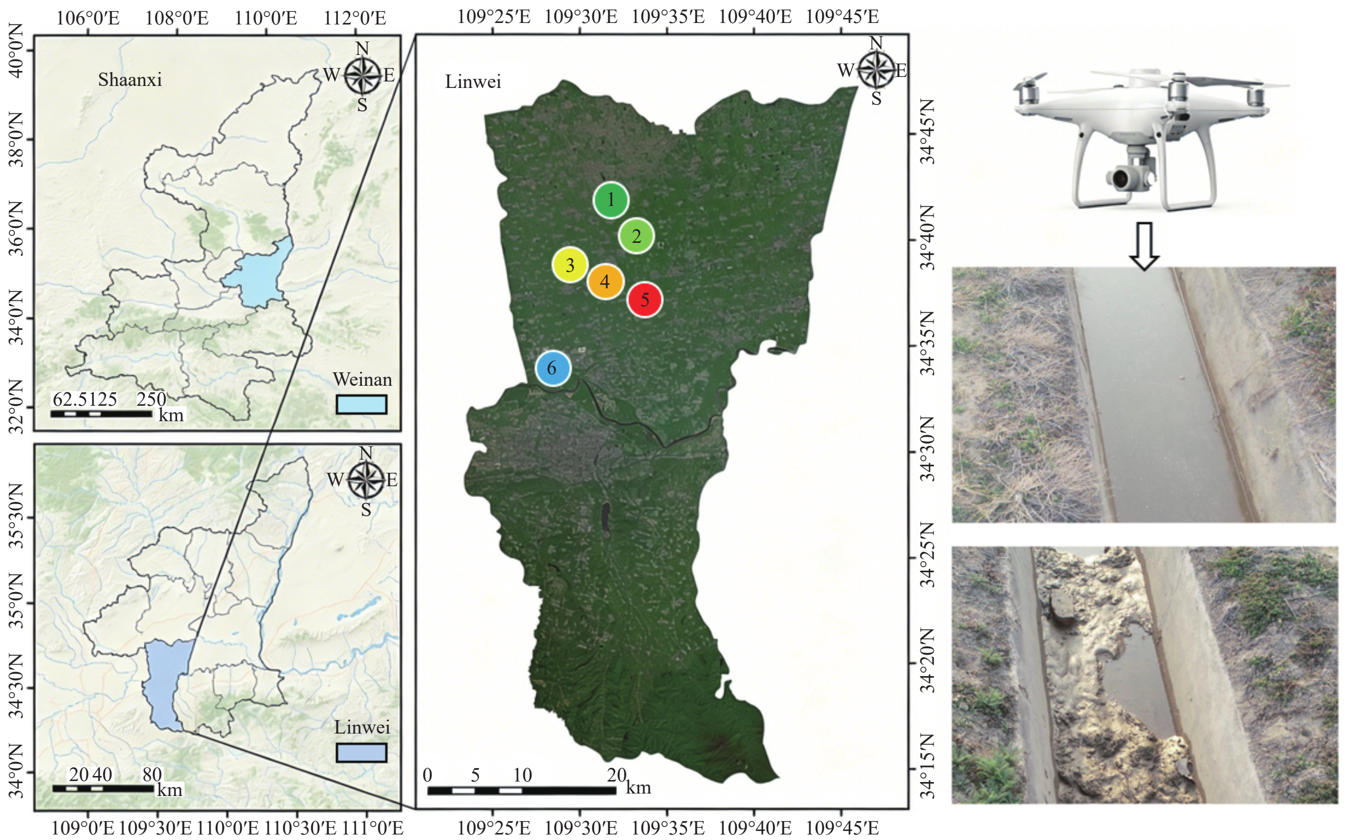
At the loss function level, the CalibratedFocal Loss function was introduced to mitigate class imbalance and overfitting. By integrating class weight optimization, label smoothing, and a Focal Loss modulation factor, this function overcomes the limitations of traditional cross-entropy loss. Consequently, this loss function improves numerical stability and strengthens the model's capacity to learn from difficult samples.

At the feature extraction level, a DP-FCM was designed to enhance feature representation. This module constructs a dual-branch architecture combining a general backbone path with a lightweight discriminative path. By integrating an attention mechanism, this dual-path structure enables the complementary extraction of global semantic and local fine-grained features. The collaborative design resolves the limitations of single-path networks in capturing differentiated local details, such as silt and gravel blockages.

At the classification level, the EC-IM was employed to perform feature refinement and state classification. This module utilizes feature dimensionality reduction, channel attention calibration, and residual enhancement to compress high-dimensional features effectively. Furthermore, a hierarchical classifier head design ensures precise discrimination of confusing low-dimensional features.

### 2.3 CalibratedFocal loss

Traditional Cross-Entropy loss cannot effectively address class imbalance, sample difficulty differentiation, and numerical stability[28]. To overcome these challenges, this study introduces CalibratedFocal Loss as the loss function. By integrating Weighted Focal Loss with Label Smoothing techniques, this approach aims to systematically enhance the model's recognition accuracy.



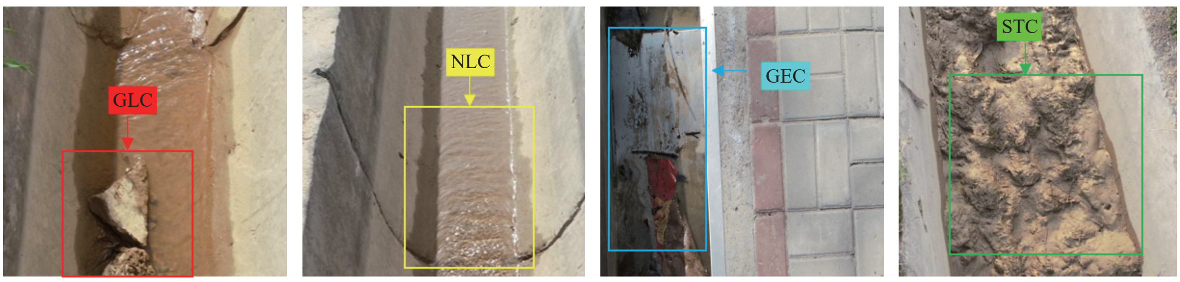

a. Data collection area in Linwei District, Weinan City, Shaanxi Province, China (the detailed geographic locations corresponding to each colored serial number are listed in Table 1)

b. Schematic diagram of UAV-based data collection

Figure 1 Study area and UAV remote sensing images

**Table 1 Data collection site type**

| No. | Location | Canal system type |
|---|---|---|
| 1 | Hanggu Road, Guandao Town, Linwei District (Xizhao Village Section) | Narrow branch canal with varying slopes. |
| 2 | Hanggu Road, Guandao Town, Linwei District (Chengguo Village Section) | Linear open canal with a single structure. |
| 3 | Liangjia Village Experimental Field, Lusheng Agricultural Services | Secondary branch canal system distributed in an "H" shape, consisting of a composite structure of open canals and buried pipes. |
| 4 | Xitun Village, Guandao Town, Linwei District | Micro branch canal with a width of less than 1 meter. |
| 5 | Main trunk canal on the east side of Tunnan Village | Primary trunk canal network with a "T" shaped junction structure. |
| 6 | Qinchuan Avenue, Linwei District (Nianyang Village Section) | Small-scale irrigation and drainage canal in the urban-rural fringe area. |

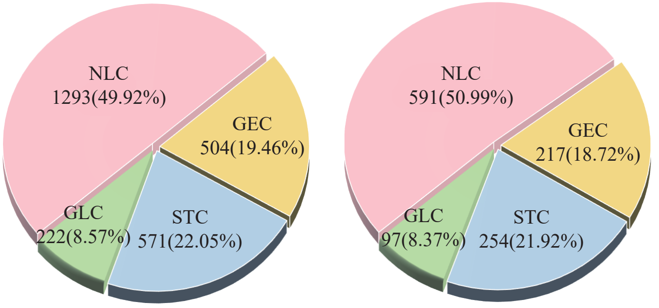



a. Gravel clogging (GLC) b. Normal circulation (NLC) c. Garbage clogging (GEC) d. Silt clogging (STC)

Figure 2 Labeled images of different irrigation canal statuses

The CalibratedFocal Loss integrates multi-dimensional optimization strategies to enhance model performance. First, class weights $\omega_{ytrue}$ are normalized and clipped within the range of 0.2 to 2.0 to mitigate training bias, as shown in Equation (1). The clipping range is determined based on the computed class weight distribution $\omega_{ytrue} = [1.28, 2.92, 0.5, 1.13]$. The upper bound of 2.0 limits the gravel class weight from 2.92 to prevent gradient domination by a single class, while the lower bound of 0.2 ensures that no class is excessively down-weighted, maintaining a dynamic range factor of 10-fold commonly adopted in weighted loss practices. Second, label smoothing is adopted to reduce model over-confidence as defined in Equation (2). Finally, the total weighted loss is calculated by combining the smoothed cross-entropy with the Focal Loss modulation factor as presented in Equations (3) and (4). This computation ensures numerical stability and directs model optimization toward hard samples.

$$\omega_i = \frac{N}{C \times n_i} \tag{1}$$




a. Distribution of sample counts in the training set b. Distribution of sample counts in the test set

Figure 3 Distribution of sample counts for different irrigation canal statuses

where $\omega_i$ denotes the weight of the $i$-th class, with $i$=1, 2, 3, 4. $N$ represents the total number of samples across all classes, $n_i$ represents the number of samples in the $i$-th class, and $C$ is the total number of classes.

$$\hat{y}_c = (1-\alpha)\cdot y_c + \frac{\alpha}{C} \tag{2}$$

where $\hat{y}_c$ is the smoothed label value for the $c$-th class, $y_c$ is the one-hot hard label, and $\alpha$ is the smoothing factor with a value of 0.05, a conventional setting for label smoothing.

$$L_{focal} = -(1-p_t)^{\gamma}\cdot\sum_{c=1}^{C}\hat{y}_c\cdot\log(p_c) \tag{3}$$

where $L_{focal}$ denotes the focal loss. The term $p_t$ denotes the predicted probability of the true class. This value is clipped to $[\epsilon, 1-\epsilon]$ with $\epsilon = 10^{-7}$. The parameter $\gamma$ represents the focusing factor and is set to 1, a commonly used default for balanced classification scenarios. The term $p_c$ refers to the predicted probability for the $c$-th class.

$$L_{final} = L_{focal}\cdot\omega_{ytrue1} \tag{4}$$

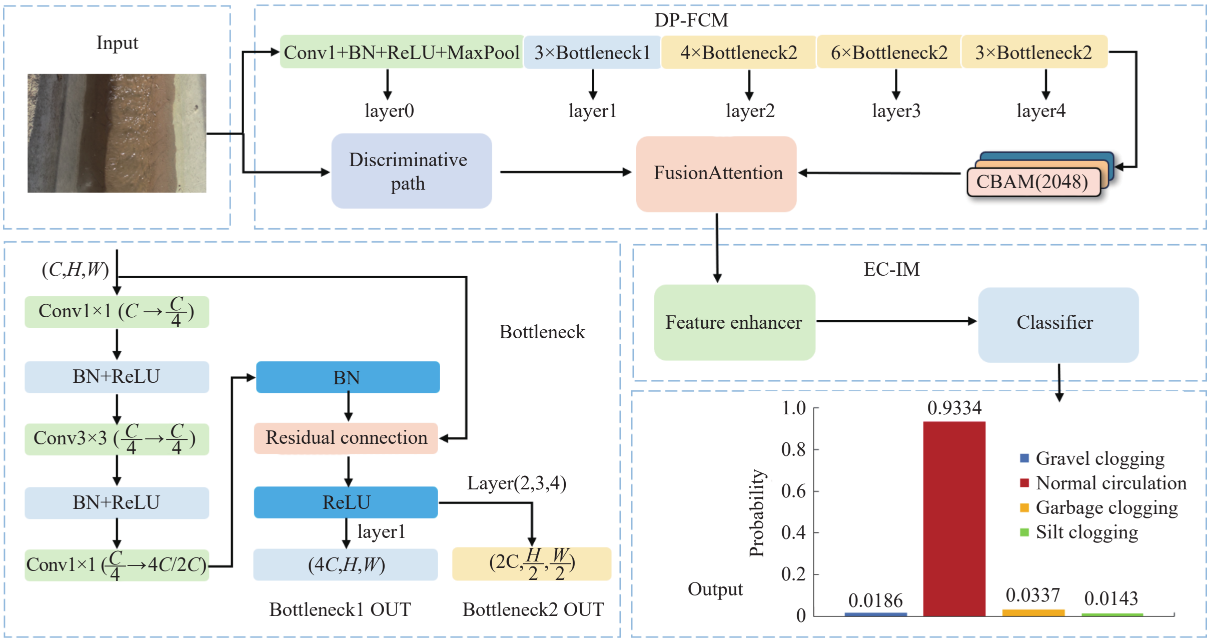
Finally, $L_{final}$ denotes the final weighted loss, and $\omega_{ytrue1}$ represents the preprocessed set of class weights.

### 2.4 Dual-Path and Fusion Collaborative Module (DP-FCM)

In irrigation canal condition detection, feature variations manifest a globally similar yet locally heterogeneous pattern. ResNet-based models rely on a single backbone path for feature extraction, which limits the discrimination of categories with diverse morphological traits, particularly the paste-like texture of silt, the blocky edges of gravel, and the irregular contours of debris. Simply increasing model capacity tends to cause overfitting on small-scale datasets.

To overcome these limitations, this study proposed the DP-FCM module, which adopts a dual-path collaborative paradigm that unifies a backbone General Path and a lightweight Discriminative Path to complement fine-grained features. The architecture is illustrated in Figure 5.

The general path is built upon the ResNet-50 architecture initialized with ImageNet pre-trained weights. A CBAM[29] is embedded after layer4 to enhance the feature response of key regions.




Note: $C$, $H$, $W$ denote the number of channels, height, and width of the input feature map, respectively. BN, Batch Normalization. ReLU, Rectified Linear Unit. CBAM, Convolutional Block Attention Module. Conv, Convolution.

Figure 4 DP-ResNet50 model structure diagram

The discriminative path accepts the original input image as an independent input and employs a hierarchical cascade of convolution, pooling, and attention to extract low-dimensional fine-grained features. It starts with a standard 3×3 convolution-pooling block with 64 output channels. The channel dimensions are progressively expanded to 128 and 256 through successive convolution-pooling blocks, with CBAM embedded at each stage to suppress background interference and highlight task-relevant fine-grained features. Finally, adaptive average pooling yields the discriminative path output.

To fuse the discriminative path features with the backbone features, global average pooling and flattening compress both features into vectors, which are concatenated and processed by a lightweight attention network to generate channel-wise dynamic weights, as illustrated in Figure 6. These weights recalibrate the fused features, strengthening the most discriminative components for the downstream classification task.

## 2.5 Enhancement and Classifier Integrated Module (EC-IM)

High-dimensional features of irrigation canals contain redundant background information, while low-dimensional embedded features of key categories such as silt and gravel are prone to confusion. High-dimensional redundancy and low-dimensional confusion render traditional single-level fully connected classification heads unable to achieve effective feature mapping. To address this problem, the EC-IM module was designed. The module adopts an end-to-end architecture consisting of feature dimensionality reduction, attention optimization, residual enhancement, and hierarchical classification to improve discriminative accuracy. The structure of the module is shown in Figure 7.

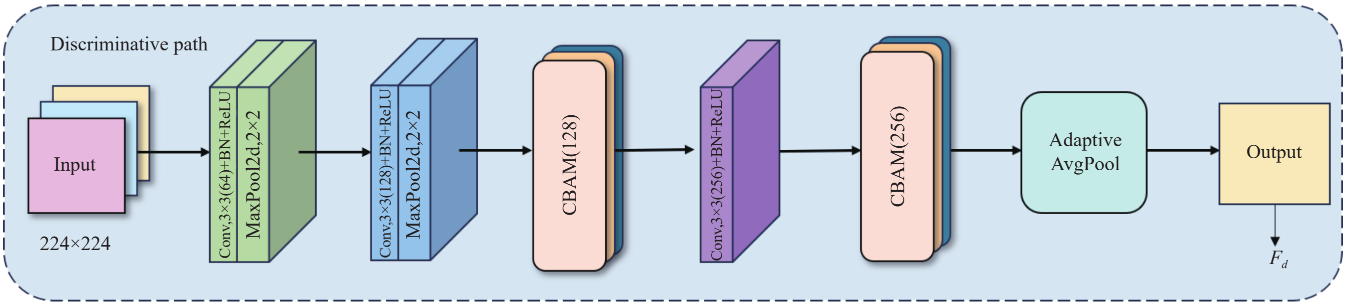
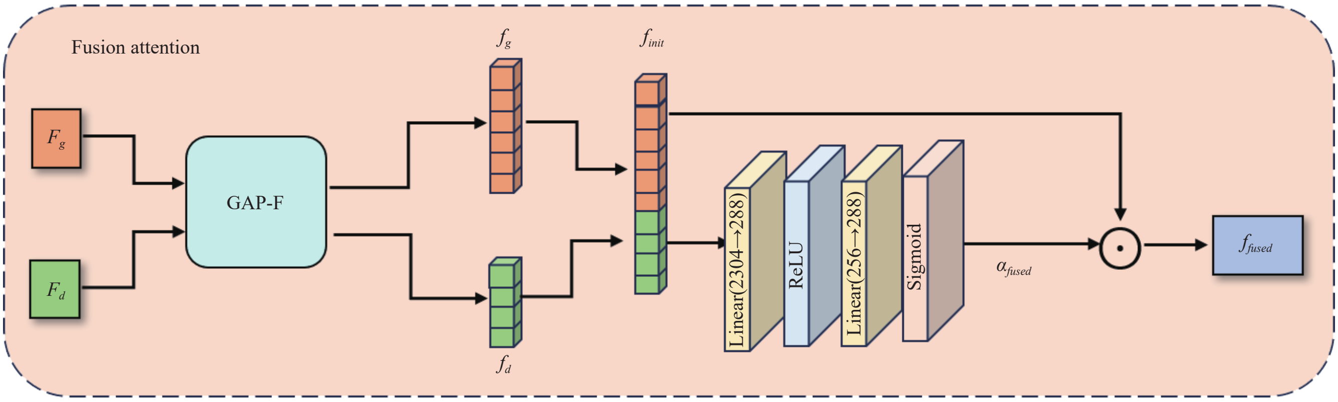



Figure 5 Structure diagram of the lightweight discriminative path




Note: $F_g$ and $F_d$ denote the raw output features of the backbone General Path and the Discriminative Path, respectively. These features are transformed into one-dimensional feature vectors $f_g$ and $f_d$ via global average pooling and flattening (GAP-F). The vector $f_{init}$ denotes the fused feature. The vector $\alpha_{fused}$ denotes the dynamic weight. The vector $f_{fused}$ denotes the enhanced fused feature.

Figure 6 Structure diagram of feature fusion

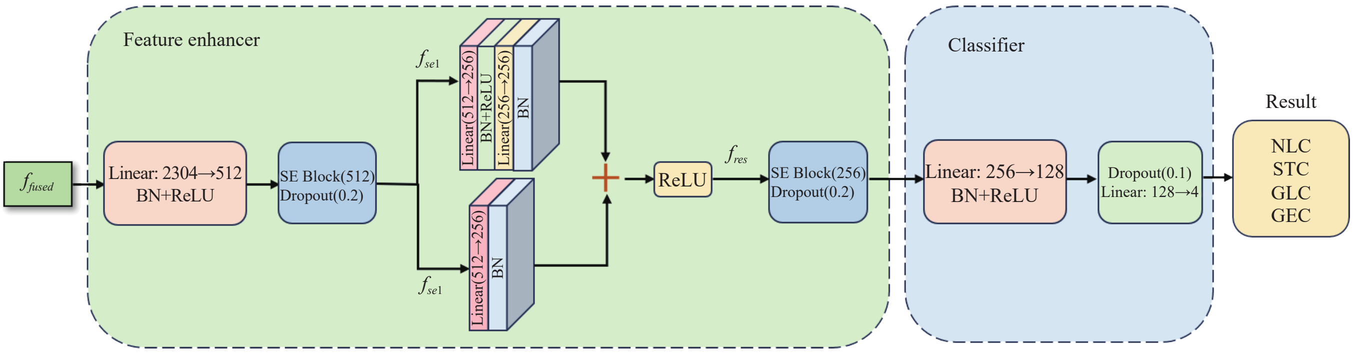



Figure 7 Structure diagram of the enhancement and classifier integrated module

The module first compresses high-dimensional fused features into a 512-dimensional representation via a fully connected layer to alleviate background noise interference and reduce computational burden. This representation undergoes channel-wise attention recalibration using a Squeeze-and-Excitation (SE) block, which adaptively emphasizes informative feature maps, as illustrated in Figure 8. A Dropout layer with a drop rate of 0.2 was introduced to improve the model's generalization capability, yielding the first-stage enhanced feature $f_{se1}$.

Subsequently, a residual enhancement stage was adopted to deepen feature processing while avoiding feature degradation. The main branch performs a non-linear transformation to capture complex feature patterns, while the parallel shortcut connection ensures stable gradient flow and feature integrity. The residual output $f_{res}$, as formulated in Equation (5), was obtained, followed by further feature refinement via another SE block.

Finally, a hierarchical two-level fully connected classification head is deployed to address the issue of low-dimensional feature confusion. This classification head maps the multi-stage optimized features to a 4-dimensional output space, enabling accurate classification of irrigation canal conditions.

$$f_{res} = ReLU(MainPath(f_{se1}) + ShortcutPath(f_{se1})) \tag{5}$$

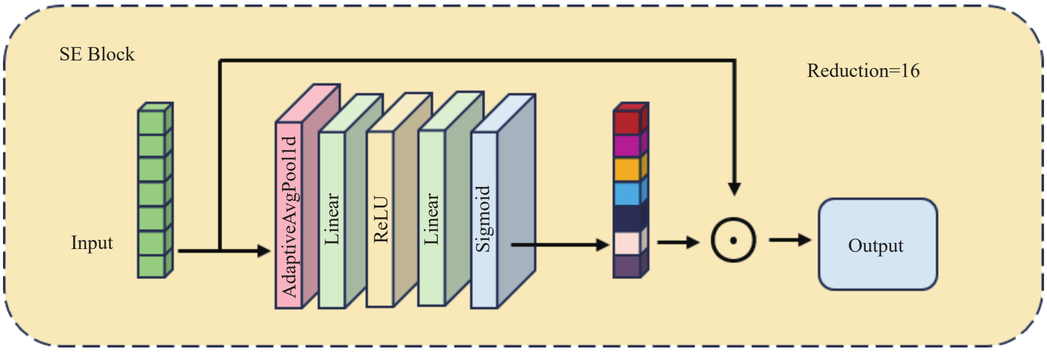



Figure 8 Structure diagram of SE block

## 3 Experiments and evaluation

### 3.1 Experimental parameters

Experiments were conducted on a Windows system with an NVIDIA GeForce RTX 3060 GPU, Python 3.12.7, and PyTorch 2.3.1. All algorithms were executed under this identical setup to ensure comparability.

For model training, input images were resized to a spatial resolution of 224×224 pixels. The Adam optimizer was used with an initial learning rate of $1e^{-4}$ and weight decay of $1e^{-4}$ to optimize convergence. A StepLR scheduler reduced the learning rate by 30% every 10 epochs. Training proceeded for a maximum of 100 epochs with early stopping, where the patience parameter for early stopping was 12 epochs. The same hyperparameters were maintained during fine-tuning.

### 3.2 Evaluation metrics

In this study, four evaluation metrics were selected to assess the model performance, namely Accuracy, Macro Precision, Macro Recall, and Macro F1-score. These metrics aim to comprehensively measure the classification accuracy and generalization ability of the improved algorithm, with the specific formulas provided in Equations (6)-(9).

$$\text{Accuracy} = \frac{1}{M}\sum_{i=1}^{M} 1(\hat{y}_i = y_i) \tag{6}$$

where, $M$ denotes the total number of test samples, $\hat{y}_i$ represents the predicted result of the model for the $i$-th sample, and $y_i$ is its true label.

$$\text{Macro Precision}=\frac{1}{C}\sum_{c=1}^{C}\frac{\text{TP}_c}{\text{TP}_c+\text{FP}_c} \tag{7}$$

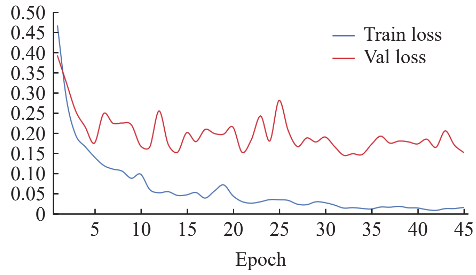
$$\text{Macro Recall}=\frac{1}{C}\sum_{c=1}^{C}\frac{\text{TP}_c}{\text{TP}_c+\text{FN}_c} \tag{8}$$

$$\text{Macro F1} = \frac{1}{C}\sum_{c=1}^{C}\frac{2\cdot\text{Precision}_c\cdot\text{Recall}_c}{\text{Precision}_c+\text{Recall}_c} \tag{9}$$

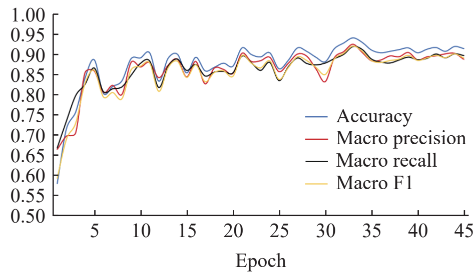
where $\text{TP}_c$, $\text{FP}_c$, and $\text{FN}_c$ denote the number of true positives, false positives, and false negatives for class $C$ respectively, and $C$ is the total number of classes. In the classification model designed in this study, $C = 4$.

### 3.3 Performance of the DP-ResNet50 model

The training dynamics of the DP-ResNet50 model are summarized in Figure 9. As shown in Figure 9a, both training and validation losses decrease rapidly during early epochs and then stabilize, indicating effective convergence. The training loss drops sharply from 0.4653 to 0.1412 in the first five epochs and eventually reaches around 0.0160 with minor fluctuations, demonstrating fine-grained feature fitting. The validation loss follows a similar trend, attaining its minimum of 0.1452 at epoch 32 and recovering robustly after slight variations.




a. Loss curve




b. Performance metric curves

Figure 9 Training curves of the DP-ResNet50 model

Concurrently, the core classification metrics show synchronized and sustained improvement throughout training, as illustrated in Figure 9b. All metrics rise significantly within the first 10 epochs and continue to climb until reaching their peak at epoch 33. At this optimum, the model achieves its best performance with an accuracy of 0.9413, a macro precision of 0.9250, a macro recall of 0.9195, and a macro F1-score of 0.9211. The subsequent stability of these metrics at high levels confirms the training process's effectiveness and the model's reliable convergence.

To evaluate the classification performance of the proposed model across various irrigation canal conditions, an assessment was conducted on the test set. Figure 10 presents the confusion matrix at the optimal macro F1-score, where the recall values for the four categories are 0.9814 for $\text{RE}_{\text{NLC}}$, 0.9724 for $\text{RE}_{\text{STC}}$, 0.9175 for $\text{RE}_{\text{GLC}}$, and 0.8065 for $\text{RE}_{\text{GEC}}$. These findings demonstrate that the DP-ResNet50 model possesses a robust detection capability and maintains high performance even in scenarios characterized by imbalanced sample distributions.

### 3.4 Ablation experiments

To evaluate the contributions of CalibratedFocal Loss, DP-

FCM, and EC-IM to the ResNet-50 baseline, the study conducts ablation experiments under consistent training settings, where rows 1-4 in Tables 2 and 3 correspond to progressively stacking each component onto the baseline. The results are summarized as follows:

1) CalibratedFocal Loss effectively addresses class imbalance. After its introduction, validation loss drops from 0.3341 to 0.1583, macro F1-score rises to 0.9106, and the recall of silt clogging improves by 5.9 percentage points. Moreover, the recall distribution becomes the most balanced across all categories, confirming that the loss not only directs attention to hard samples but also maintains inter-class equilibrium.

2) Adding DP-FCM further enhances discriminability for categories with concentrated feature patterns, raising the recall of silt clogging to 0.9724. However, the recall of garbage clogging declines significantly. This drop in recall rate stems from the high heterogeneity and amorphous morphology of garbage; during deep feature extraction, DP-FCM tends to suppress such variable patterns as noise, impairing effective feature aggregation for this category.

3) Incorporating EC-IM mitigates the above feature conflict. The recall of garbage clogging rebounds to 0.8065, while silt clogging and gravel clogging maintain high recalls of 0.9724 and 0.9175, respectively. Meanwhile, the overall performance achieves the optimum, where the validation loss drops to 0.1490, accuracy reaches 0.9413, and the macro F1-score stands at 0.9211. In addition, the V5-F1 value is defined as the variance of the macro F1-score across five consecutive training epochs after the model reaches its optimal state. The V5-F1 value of the model integrated with the EC-IM mechanism is the lowest among all experimental groups, at only 0.000116. This result fully confirms that the EC-IM mechanism plays a critical role in improving model robustness and training stability.

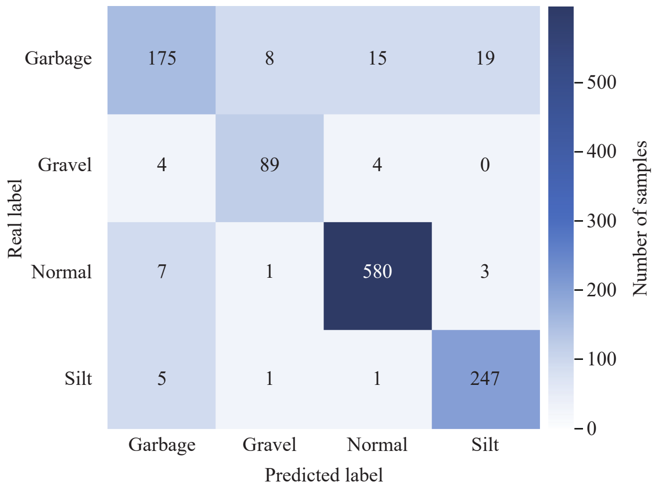



Figure 10 Confusion matrix of the best results on the test set

**Table 2 Various improved methods on the detection performance of irrigation canal states**

| No. | ResNet-50 | CalibratedFocal Loss | DP-FCM | EC-IM | Validation Loss | Accuracy | Macro Precision | Macro Recall | Macro F1-score |
|---|---|---|---|---|---|---|---|---|---|
| 1 | √ | | | | 0.3341 | 0.9249 | 0.9163 | 0.8991 | 0.9070 |
| 2 | √ | √ | | | 0.1583 | 0.9258 | 0.9064 | 0.9151 | 0.9106 |
| 3 | √ | √ | √ | | 0.1918 | 0.9327 | **0.9278** | 0.9015 | 0.9113 |
| 4 | √ | √ | √ | √ | **0.1490** | **0.9413** | 0.9250 | **0.9195** | **0.9211** |

**Table 3 Impact of various improved methods on recall and macro F1-score variance of different irrigation canal states**

| No. | $RE_{NLC}$ | $RE_{STC}$ | $RE_{GLC}$ | $RE_{GEC}$ | V5-F1 |
|---|---|---|---|---|---|
| 1 | 0.9780 | 0.8701 | 0.8866 | 0.8618 | 0.000 158 |
| 2 | 0.9509 | 0.9291 | 0.9278 | 0.8525 | 0.000 167 |
| 3 | 0.9865 | 0.9724 | 0.8866 | 0.7604 | 0.000 245 |
| 4 | 0.9814 | 0.9724 | 0.9175 | 0.8065 | 0.000 116 |

In summary, this ablation experiment demonstrates that CalibratedFocal Loss lays a solid foundation for classification balance. While DP-FCM greatly enhances key feature extraction, the module may introduce feature conflicts for specific categories. To address this, the EC-IM mechanism alleviates these conflicts through its feature enhancement capabilities, thereby achieving a joint improvement in both overall performance and inter-category balance. Consequently, the final model exhibits favorable comprehensive performance in scenarios where sample imbalance and feature complexity coexist.

### 3.5 Results comparison of different classification and detection algorithms

To further verify the performance of the DP-ResNet50 model in irrigation canal condition detection, this experiment adopted the overall metrics such as accuracy and the per-category recall rates described in Section 3.2 for evaluation. Under the same training parameters and using the image dataset of four irrigation canal states collected from the Jiaokou Chouwei Irrigation District in Weinan, Shaanxi, China, the performance of this model is compared with that of six other typical models. These benchmarks include ConvNeXtV2-Tiny[30], Vision Transformer[31], ResNet-34[24], ResNet-50[24], MaxViT-Tiny[32], and EfficientNet-B3[33].

The experimental results are illustrated in Figure 11. Specifically, the Vision Transformer exhibits consistently low values across all evaluation metrics. While models such as ConvNeXtV2-Tiny and EfficientNet-B3 show relatively competitive performance, they still have obvious gaps in multiple metrics compared with DP-ResNet50. The ResNet-34 and ResNet-50 models from the same series also fail to reach the performance level of DP-ResNet50 in most dimensions. Although the recall for garbage clogging does not reach the optimal value among all models, a result primarily driven by the inherent complexity of the garbage clogging category, the DP-ResNet50 model performs favorably. The model achieves the best performance in all overall metrics and the remaining three per-category recall rates.

In summary, DP-ResNet50 demonstrates favorable overall performance and classification accuracy in the irrigation canal detection task compared with the other models.

### 3.6 Heatmap visualization analysis

To more intuitively evaluate the performance of the DP-ResNet50 model, this study visualized the activation maps at different layers of the model to analyze its attention region capture and feature representation in the classification task. For the ResNet-50 model, spatial features from layer4 were directly extracted, and heatmaps were generated through channel-wise averaging, normalization, and Gaussian blurring. For the DP-ResNet50 model, channel weights from the attention-guided feature fusion mechanism were first extracted. These weights were then fused with the spatial features of layer4 via channel-wise weighting, after which the same preprocessing steps as above were applied to generate heatmaps. In the experiment, two typical irrigation canal states, namely normal flow and silt clogging, were selected as test scenarios.

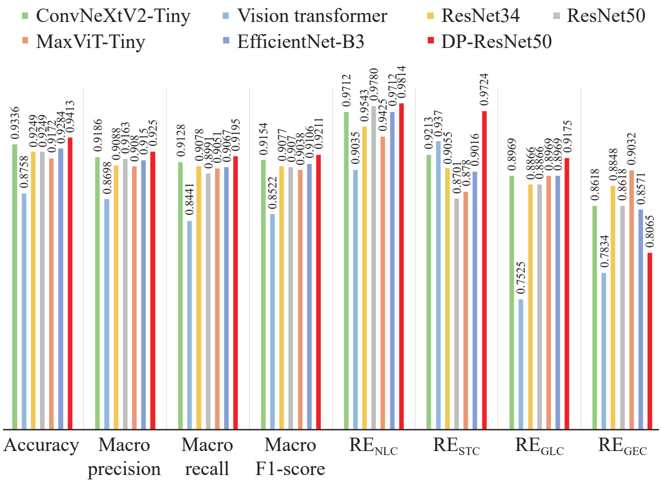
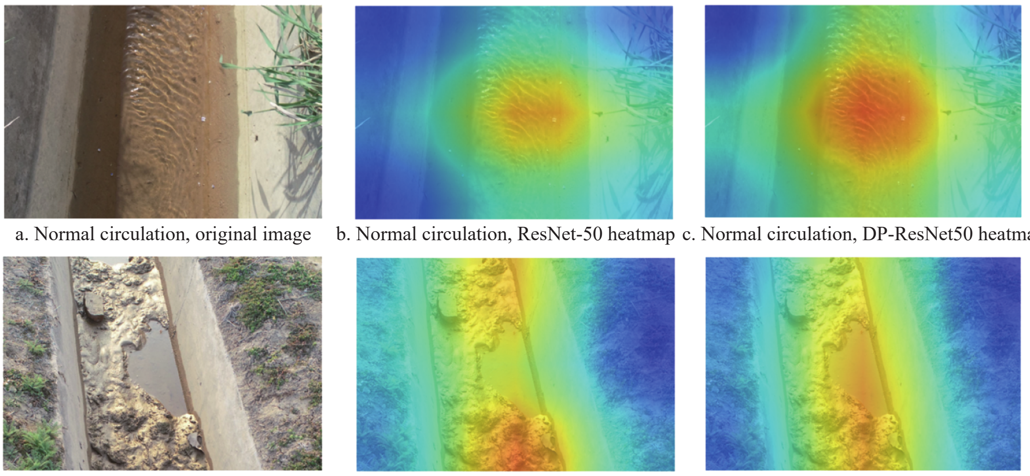



Figure 11　Performance comparison of different models

Figures 12a-12c show the original image and corresponding heatmap of a normal-flow irrigation canal. The heatmap of the ResNet-50 model covers the water flow region, but the activation boundaries are blurred. In contrast, the DP-ResNet50 model focuses its activation on the core water flow region, with a significant increase in activation intensity. Irrelevant areas at the canal edges are suppressed, demonstrating more favorable feature focusing capability.

Figures 12d-12f display the original image and corresponding heatmap of a silt-clogged irrigation canal. The heatmap of the ResNet-50 model mainly covers the silt accumulation area, but the activation intensity in the core region of the water pit is weak with blurred boundaries. Meanwhile, meaningless weak activations appear in weed-interfered regions, indicating insufficient feature focusing and vulnerability to irrelevant interference. In contrast, the main activation region of the DP-ResNet50 model covers both the silt accumulation area and the water pit, with high and continuously distributed activation intensity. Moreover, its activation boundaries closely follow the connecting contours of silt, water accumulation, and canal walls, effectively avoiding feature interference from weed areas.

In summary, due to the fusion mechanism, the DP-ResNet50 model demonstrates more precise feature focusing capability. The model's decision attention aligns more accurately with the key regions of the irrigation canal state, further verifying the effectiveness of this fusion architecture design.



a. Normal circulation, original image　b. Normal circulation, ResNet-50 heatmap　c. Normal circulation, DP-ResNet50 heatmap

d. Silt clogging, original image　e. Silt clogging, ResNet-50 heatmap　f. Silt clogging, DP-ResNet50 heatmap

Note: The color intensity in the heatmaps indicates the relative importance of image regions for the model's decision (warmer colors denote higher importance).

Figure 12　Original images and heatmaps

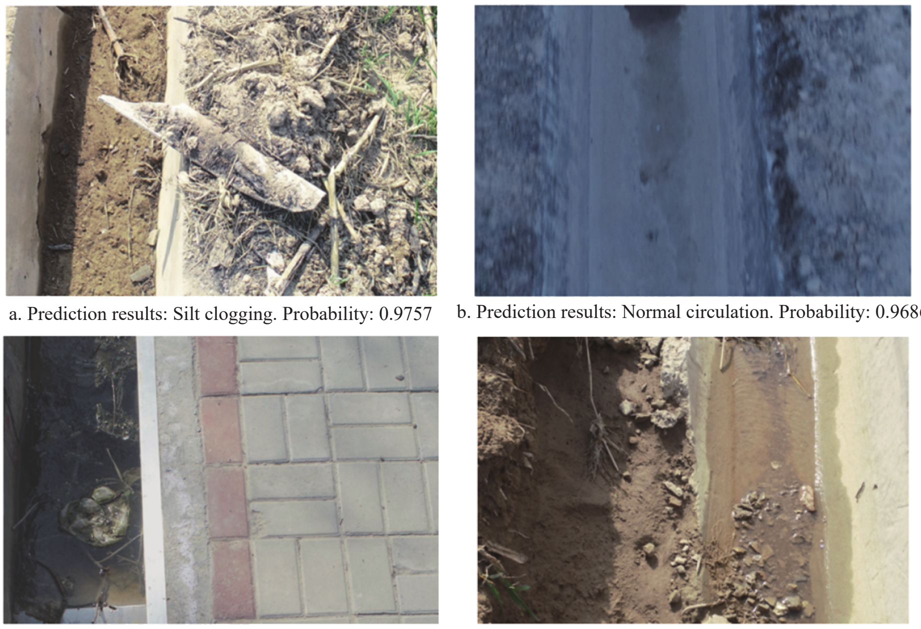
### 3.7　Impact of complex images

In low-altitude remote sensing images captured by UAVs, complex environments can affect the recognition performance of the model. To evaluate the performance of the DP-ResNet50 model in the complex irrigation canal image detection task, this study conducts a targeted analysis from the perspective of multi-dimensional interference scenarios.

Figure 13a shows a silt clogging scenario with complex impurities such as entangled weeds and loose soil. Figure 13b covers environmental interference conditions such as image blurring and brightness variation, along with the normal circulation state of irrigation canals under non-water-flow conditions. Figure 13c presents a garbage clogging scenario integrated with road pavement and water accumulation environments. The results demonstrate that the model exhibits strong robustness in the above diverse scenarios and can accurately identify the operating states of irrigation canals.

Figure 13d focuses on a more extreme gravel clogging scenario where the canal walls are damaged and the surrounding soil is exposed. The model achieves a recognition confidence of merely 0.4138, reflecting low prediction certainty for such samples.

Overall, the DP-ResNet50 model exhibits favorable recognition capability under complex environmental conditions. However, in practical applications, optimizing image acquisition strategies such as UAV flight altitude and speed remains necessary to further improve the accuracy and reliability of irrigation canal condition detection.



a. Prediction results: Silt clogging. Probability: 0.9757 b. Prediction results: Normal circulation. Probability: 0.9686

c. Prediction results: Garbage clogging. Probability: 0.9854 d. Prediction results: Gravel clogging. Probability: 0.4138

Note: Subfigures a-d correspond to silt clogging, normal circulation, garbage clogging, and gravel clogging scenarios, respectively.

Figure 13 Complex image detection results

## 4 Discussion

This study demonstrates that the DP-ResNet50 model, incorporating CalibratedFocal Loss, DP-FCM, and EC-IM modules, effectively addresses sample imbalance and complex local feature variations in irrigation canal condition detection. However, several limitations remain.

Despite the method's favorable performance, it still has several limitations. Geographically, the dataset only covers Weinan City, Shaanxi Province, China, which restricts the cross-regional generalization ability of the model. In terms of practical application, UAV data collection is susceptible to weather conditions, which leads to a scarcity of data samples in extreme scenarios such as strong specular reflection.

In addition, garbage clogging targets exhibit high feature heterogeneity in material, shape, color, and size, causing the recall rate to drop from the baseline 0.8618 to 0.8065. A representative misclassification example is shown in Figure 14, where a garbage clogging sample is incorrectly predicted as silt clogging with a confidence of 0.549. This confusion pattern indicates that category ambiguity, rather than annotation quality, is the primary cause of the performance gap. Specifically, garbage exhibits extreme intra-class heterogeneity, and diverse objects such as tree branches, plastic bags, and water accumulation produce fragmented feature patterns that overlap with the irregular boundaries of silt deposits and the scattered texture of gravel aggregates in the feature space, while sparse garbage in flowing water visually resembles normal canal conditions. In addition, the DP-FCM module prioritizes dominant feature patterns at the expense of fine-grained discriminative cues. Although EC-IM partially mitigates this issue, the garbage clogging category remains disproportionately affected.

Notably, compared with the baseline ResNet-50 model, the DP-ResNet50 architecture shows a 16.2% increase in parameters, a 47.7% increase in FLOPs, and a 25.1% decrease in inference speed, which poses challenges for deployment on edge devices.

Future research can be advanced in the following aspects. First, expand the coverage of the dataset by incorporating irrigation canal images from multiple regions and under diverse environmental conditions to enhance the model's environmental adaptability and generalization ability. Second, pursue model lightweighting through techniques such as network pruning, knowledge distillation, or lightweight convolution design to reduce computational overhead and improve feasibility for deployment on edge devices. Third, optimize data acquisition and training strategies by leveraging methods such as adversarial training and multimodal data fusion to strengthen the model's robustness in extreme scenarios. Fourth, explore targeted data augmentation and dedicated loss functions for heterogeneous targets like garbage clogging to further improve recall, thereby enhancing the practicality and generalizability of the system in real-world complex farmland environments.

Note: Garbage clogging misclassified as silt clogging, with a confidence score of 0.5490 for the silt category.

Figure 14 Misclassification of garbage clogging

## 5 Conclusions

Against the backdrop of frequent extreme weather events and the need for refined monitoring of large-scale irrigation canals, this study proposes a condition detection system for irrigation facilities based on UAV low-altitude remote sensing and deep learning models. By using UAVs to collect multi-type canal images and constructing a diverse dataset, the proposed DP-ResNet50 model effectively alleviates problems including class imbalance and feature confusion in canal condition recognition and achieves favorable performance on the test set.

This study provides an efficient and accurate technical pathway for the automated monitoring of smart agricultural irrigation facilities. In future work, the system's adaptability and practical value can be further improved by expanding the dataset, lightweighting the model, and optimizing data acquisition strategies.

## Acknowledgements

This work was financially supported by the National Key R&D Program of China (Grant No. 2024YFD1500602).

## [References]

[1] Liu M H, Xu X, Jiang Y, Huang Q Z, Huo Z L, Liu L, et al. Responses of crop growth and water productivity to climate change and agricultural water-saving in arid region. Science of the Total Environment, 2020; 703: 134621.

[2] Gohar A A, Amer S A, Ward F A. Irrigation infrastructure and water appropriation rules for food security. Journal of Hydrology, 2015; 520: 85–100.

[3] Awais M, Li W, Cheema M J M, Zaman Q U, Shaheen A, Aslam B, et al. UAV-based remote sensing in plant stress imagine using high-resolution thermal sensor for digital agriculture practices: A meta-review. International Journal of Environmental Science and Technology, 2023; 20(1): 1135–1152.

[4] Bhatti M T, Ashraf M, Anwar A A. Soil erosion and sediment load management strategies for sustainable irrigation in arid regions. Sustainability, 2021; 13(6): 3547.

[5] Pádua L, Marques P, Dinis L-T, Moutinho-Pereira J, Sousa J J, Morais R, et al. Detection of leak areas in vineyard irrigation systems using UAV-based data. Drones, 2024; 8(5): 187.

[6] Colomina I, Molina P. Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 2014; 92: 79–97.

[7] Hnatiuc M, Paun M, Kapsamun D. IoT sensors system for vineyard monitoring. In: 2022 7th International Conference on Information and Network Technologies (ICINT), 2022; pp.46–51. doi: 10.1109/ICINT55083.2022.00015.

[8] Jabbar W A, Ting T M, Hamidun M F I, Kamarudin A H C, Wu W, Sultan J, et al. Development of LoRaWAN-based IoT system for water quality monitoring in rural areas. Expert Systems with Applications, 2024; 242: 122862.

[9] Traore S, Fipp G, Swanson C, Dufek J, Robin M. Unmanned aerial vehicle system integrating high-resolution sensors for detecting leaks in irrigation canals—Proof of concept. Irrigation Drainage, 2021; 71(1): 206–217.

[10] Jalajamony H M, Nair M, Mead P F, Fernandez R E. Drone aided thermal mapping for selective irrigation of localized dry spots. IEEE Access, 2023; 11: 7320–7335.

[11] Ferro M V, Catania P, Miccichè D, Pisciotta A, Vallone M, Orlando S. Assessment of vineyard vigour and yield spatio-temporal variability based on UAV high resolution multispectral images. Biosystems Engineering, 2023; 231: 36–56.

[12] Zheng Y, Zhu Q B, Huang M, Guo Y, Qin J. Maize and weed classification using color indices with support vector data description in outdoor fields. Computers and Electronics in Agriculture, 2017; 141: 215–222.

[13] Liu T, Jin X J, Zhang L Y, Wang J, Chen Y, Hu C S, et al. Semi-supervised learning and attention mechanism for weed detection in wheat. Crop Protection, 2023; 174: 106389.

[14] Castellano G, De Marinis P, Vessio G. Weed mapping in multispectral drone imagery using lightweight vision transformers. Neurocomputing, 2023; 562: 126914.

[15] Gao J F, Nuyttens D, Lootens P, He Y, Pieters J G. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. Biosystems Engineering, 2018; 170: 39–50.

[16] Bakhshipour A, Jafari A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. Computers and Electronics in Agriculture, 2018; 145: 153–160.

[17] Saad El Imanni H, El Harti A, Bachaoui E M, Mouncif H, Eddassouqui F, Hasnai M A, et al. Multispectral UAV data for detection of weeds in a citrus farm using machine learning and Google Earth Engine: Case study of Morocco. Remote Sensing Applications: Society and Environment, 2023; 30: 100941.

[18] Su J Y, Yi D W, Coombes M, Liu C J, Zhai X J, McDonald-Maier K, et al. Spectral analysis and mapping of blackgrass weed by leveraging machine learning and UAV multispectral imagery. Computers and Electronics in Agriculture, 2022; 192: 106621.

[19] Zhang K, Qin L, Zhu L. PDS-YOLO: A real-time detection algorithm for pipeline defect detection. Electronics, 2025; 14(1): 208.

[20] Christiansen P, Nielsen L N, Steen K A, Jørgensen R N, Karstoft H. DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. Sensors, 2016; 16(11): 1904.

[21] Islam M M, Talukder M A, Sarker M R A, Uddin M A, Akhter A, Sharmin S, et al. A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture. Intelligent Systems with Applications, 2023; 20: 200278.

[22] Aloysius N, Geetha M. A review on deep convolutional neural networks. In: 2017 International Conference on Communication and Signal Processing (ICCSP), 2017; pp.0588–0592. doi: 10.1109/ICCSP.2017.8286426

[23] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems, 2022; 33(12): 6999–7019.

[24] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; pp.770-778. doi: 10.1109/CVPR.2016.90.

[25] Upadhyay L, Saxena A. Evaluation of enhanced Resnet-50 based deep learning classifier for tomato leaf disease detection and classification. Journal of Electrical Systems, 2024; 20(3s): 1–14.

[26] Ahmed M, Ahmed A. Palm tree disease detection and classification using residual network and transfer learning of inception ResNet. PLOS One, 2023; 18(3): e0282250.

[27] Haggab O N, Al-Qaysi Z T. Detecting defect in central pivot irrigation system using YOLOv7 algorithms. Al-Salam Journal for Engineering and Technology, 2024; 3(2): 38–49.

[28] Mukhoti J, Kulharia V, Sanyal A, Golodetz S, Torr P H S, Dokania P K. Calibrating deep neural networks using focal loss. In: Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS '20), 2020; Article 1282, pp.15288–15299.

[29] Woo S, Park J, Lee J-Y, Kweon I S. CBAM: Convolutional block attention module. In: Computer Vision -- ECCV 2018, 2018; pp.3–19. doi: 10.1007/978-3-030-01234-2_1.

[30] Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon I S, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023; pp.16133–16142. doi: 10.1109/CVPR52729.2023.01548.

[31] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint, 2021; arXiv: 2010.11929. doi: 10.48550/arXiv.2010.11929.

[32] Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A, et al. MaxViT: Multi-axis vision transformer. In: European Conference on Computer Vision (ECCV), 2022; pp.459–479. doi: 10.1007/978-3-031-20053-3_27.

[33] Tan M, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning (ICML), 2019; arXiv: 195.11946. doi: 10.48550/arXiv.1905.11946.