Journal of Agricultural Engineering

https://www.agroengineering.org/

RIBAC-Net: integrative brightness adaptive plant leaf disease classification

Xing Xu,[1] Hongya Ma,[1] Yun Zhao,[1] Xiaoshu Lv[2,3]

[1]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China
[2]Department of Electrical Engineering and Energy Technology, University of Vaasa, Vaasa 65380, Finland
[3]Department of Civil Engineering, Aalto University, Espoo 02130, Finland

Corresponding author: Yun Zhao, School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China. E-mail: zy_super0201@163.com

Please cite this article as doi: 10.4081/jae.2025.1772

Submitted: 27 May 2024
Accepted: 13 Febuary 2025

IBAC-Net: integrative brightness adaptive plant leaf disease classification

Xing Xu,[1] Hongya Ma,[1] Yun Zhao,[1] Xiaoshu Lv[2,3]

[1]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China
[2]Department of Electrical Engineering and Energy Technology, University of Vaasa, Vaasa 65380, Finland
[3]Department of Civil Engineering, Aalto University, Espoo 02130, Finland

Correspondence: Yun Zhao, School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, China. E-mail: zy_super0201@163.com

Keywords: Agricultural diseases, image classification, data processing, brightness adaptation, flexible downsampling.

Abstract

As agricultural technology continues to advance, effective classifications of agricultural diseases are crucial for improving crop yield and quality. This study aims to explore an innovative approach to agricultural disease image classification based on a novel image classification model architecture. First, we design a novel model architecture for image classification that better integrates shallow and deep features. Secondly, to address potential brightness differences in images collected under varying weather conditions, we have introduced an image brightness adaptive block. This block automatically adjusts the brightness of images during the data collection and processing stages, thereby reducing image disparities caused by weather variations. This step is crucial for improving the robustness of the model and ensuring accurate identification of agricultural diseases under different environmental conditions. Additionally, drawing inspiration from the Inception architecture and employing a flexible downsampling strategy, we have designed a custom inception block to integrate shallow and deep features effectively. To validate the effectiveness of our proposed approach, we conducted experiments using an agricultural disease image dataset processed with weather effects. The experimental results demonstrate that our model exhibits higher accuracy and robustness in agricultural disease image classification tasks compared to traditional methods. The code has been uploaded to GitHub at the following address: https://github.com/bettyaya/IBAC-Net.

Introduction

With the global advancement of agriculture, agricultural diseases have become a significant factor restricting crop yield and quality. The rapid spread of diseases incurs substantial economic losses for farmers and directly threatens global food security. Therefore, timely and accurate diagnosis and classification of agricultural diseases are pressing issues in the agricultural sector. Agricultural disease image classification, as a vital research area in modern agriculture, aims to automatically identify and classify plant diseases using computer vision technology. The development of this field is crucial for improving crop yields and alleviating disease pressures in agricultural production. However, due to the complex backgrounds, lighting variations, and diversity of diseases present in images, agricultural disease image classification remains a challenging problem. The rapid advancement of deep learning technology provides us with powerful tools that hold promise for significant breakthroughs in the precise diagnosis and classification of agricultural diseases (Cho et al., 2023).

There are various types of crop diseases. Traditional disease identification methods rely on manual observation and experience to identify disease categories, which are slow, labor-intensive, and subjective. In the initial stages of agricultural disease image classification, researchers primarily employed traditional image processing and machine learning methods. These methods are based on manually designed feature extraction and classifiers. However, due to the complexity and diversity of agricultural diseases, these methods often struggle to achieve satisfactory results, mainly due to low accuracy and challenges in feature extraction.

In recent years, significant progress has been made by researchers in the field of agricultural pest and disease image classification, driven by the rapid advancements in computer vision and deep learning technologies. Traditional machine learning methods such as Support Vector Machine (SVM) (Gao et al., 2016) and Random Forest (Zhang et al., 2018), among others, have been widely used for feature extraction and classification tasks. Conversely, deep learning methods like Convolutional Neural Networks (CNNs) demonstrate advantages in image feature learning and classification accuracy.

With the rise of deep learning, CNNs have significantly advanced agricultural disease image classification, including models like VGG (Khamparia et al., 2019), GoogLeNet (Krizhevsky et al., 2017), ResNet (Zeiler et al., 2014), among others. CNNs, through end-to-end learning, automatically extract features from images, thus enhancing classification accuracy. Many machine learning methods have been specifically proposed for image-based plant disease diagnosis, tailored for crops such as cucumbers (He et al., 2016), bananas (Abdu et al., 2020), cotton (Liu et al., 2013), tomatoes (Narayanan

et al., 2022), wheat, and rice (Zhang et al., 2023). However, challenges such as overfitting and small dataset sizes have arisen, prompting researchers to seek more effective approaches.

To address overfitting and small dataset challenges, researchers have progressively introduced data augmentation and transfer learning techniques. Data augmentation expands the available sample size by applying various transformations to the training set, thus improving the model's generalization performance. Additionally, transfer learning accelerates the training process and enhances performance by fine-tuning pre-trained models.

In recent years, researchers have explored innovative approaches, including the introduction of novel models, architectures, or modules. These methods aim to enhance model accuracy and address practical application issues such as computational costs and model interpretability. This phase of research has propelled agricultural disease image classification to new heights (Raja et al., 2023).

Zhang et al. (2022) proposed the CPRF algorithm to analyze agricultural yield issues, applying it to the agricultural big data domain. By utilizing key dimensions and parallel methods, the algorithm's accuracy was enhanced. Employing a cascade weighted voting method to identify rice varieties with minimal field diseases and maximum yields yielded favorable results. Ashwini et al. (2023) introduced a hybrid model combining an eos-based heuristic method with 3D-DCNN, which also incorporates the idea of the 3DDCNN-EOS algorithm for maize leaf disease classification, mitigating classification errors in the 3D-CNN method. Yao et al. (2024) introduced a novel model, the Generalized Stacked Multi-Output CNN (GSMo-CNN), to enhance plant recognition and disease classification performance, demonstrating the critical importance of backbone CNN selection for achieving good performance. Arshad et al. (2023) proposed a new deep learning framework for efficient and accurate disease prediction, including a hybrid approach integrating deep learning CNN for deep feature extraction and transformer for classification. Raja et al. (2021) presented a simple yet effective CNN regression method for directly predicting multiple plant traits from inputs from multiple sensors.

Currently, the field of agricultural disease image classification still faces challenges such as label imbalance, domain adaptation, and others. Future research directions may include more effective model architecture designs, new data processing methods, and improved model interpretability. In-depth exploration in these areas will contribute to advancing the field of agricultural disease image classification.

Against this backdrop, this study aims to enhance the overall performance of agricultural disease image classification through innovative model architecture and module designs, including a brightness adaptive block designed specifically for practical collection scenarios. This endeavor not only promises more accurate and efficient disease monitoring for agricultural production but also expands new

research directions for the application of deep learning technologies in the agricultural domain. By introducing these innovative aspects, we anticipate driving advancements in the field of agricultural disease image classification, providing robust support for the intelligent and digital transformation of agricultural production.

In existing image classification models, we face a series of challenges. Firstly, these models struggle to overcome the complexities of the environment and factors like lighting variations when dealing with agricultural disease images, resulting in inadequate generalization performance in real-field scenarios. Secondly, current models have limitations in deep-level information mining, particularly concerning agricultural disease images with multi-scale, multi-level features, where traditional convolutional neural networks often fail to fully exploit global attentiveness.

Additionally, issues during the data collection phase cannot be overlooked. Existing models find it challenging to effectively handle fluctuations in image quality under different lighting conditions, which in turn affects the model's classification accuracy. Hence, there is an urgent need for a method that can adaptively adjust image brightness to cope with the complex and changing lighting environments in actual farmland settings.

The primary goal of this research is to enhance the accuracy and robustness of agricultural disease image classification. To achieve this goal, we introduce three key innovations: an image classification model architecture, a module designed to fuse shallow and deep features, and an image brightness adaptive block specifically designed for the data collection phase.

Firstly, addressing the shortcomings of existing image classification models in deep-level information mining, we propose a new model architecture named Integrative Brightness Adaptive Plant Leaf Disease Classification (IBAC-Net). This architecture integrates the strengths of traditional Convolutional Neural Networks (CNNs) and Transformers, effectively combining local perceptiveness and global attentiveness to improve the model's classification performance under different scales and complex backgrounds.

Secondly, we design and introduce an innovative module, the custom inception block. Compared to the traditional stacked convolutional layer structures, the custom inception block possesses stronger feature extraction and representation capabilities, aiding the model in better understanding and learning abstract information in images.

Finally, addressing the specific issues in the data collection phase of agricultural disease image processing, we devise an image brightness adaptive block. This block, by introducing a trainable

brightness adjustment curve, effectively resolves image quality issues caused by insufficient lighting, enhancing the model's adaptability to images under low-light conditions.

urthermore, our research focuses on an agricultural disease image dataset processed with weather effects. This dataset simulates real-world scenarios of image collection under different environmental conditions, including lighting variations and rainy or snowy weather, among others. This specialized processing enhances the robustness of our model, enabling it to adapt to various weather conditions and better cope with the complexities encountered in actual farmland settings.

Materials and Methods

Datasets

This paper utilizes a dataset of plant disease images collected from online sources and public datasets (Bhuiyan et al., 2023; Arman et al., 2023; Kaggle, 2020, 2021a, 2021b, 2023; Sethy et al., 2020; Mwebaze et al., 2020; Singh et al., 2020), depicting complex agricultural backgrounds. Capturing plant disease images under complex weather conditions is challenging; hence, we enhanced the collected image data with weather effects such as rain, snow, fog, insufficient lighting due to overcast or rainy weather, and image blurring caused by camera motion during capture. Our dataset comprises 34 classes with a total of 23,151 images. The plant categories, the number of corresponding disease categories and the number of images included in the dataset are shown in Table 1.

Table 1. Dataset content.

Plant	Number of disease categories	Number of pictures
Apple	2	1260
Banana	4	345
Cacao	3	268
Cassava	6	612
Coffee	5	12392
Cotton	5	3988
Guava	5	402
Rice	4	3884

Below, we will discuss how to simulate image data under complex weather conditions. Employing the imgaug library in Python, we implemented image data augmentation algorithms to simulate environmental changes caused by different conditions. Using the "someof" method within a predefined data augmentation sequence, one enhancement is randomly selected from specified enhancers each time. These enhancers include motion blur, fog, rain, snow, and brightness adjustments. Subsequently, the imported images are enhanced using these enhancers and saved to designated folders. Augmenting images, especially under simulated complex weather conditions, aids in improving the model's robustness and generalization ability. Specifically, operations such as motion blur, rain and snow effects, and fog addition enable the model to better adapt to various weather conditions encountered in real farmland settings. Before treatment is shown in Figure 1. After treatment is shown in Figure 2.

Figure 1. Before processing.

Figure 2. After processing.

Additionally, to further validate the model's generalizability and reliability, we conducted experiments on the publicly available PlantVillage (Hughes and Salathé, 2015) dataset.

IBAC-Net

In this section, we will provide a detailed overview of the proposed IBAC-Net model, an effective and reliable agricultural disease image classification network based on CNN and transformer. Firstly, we will outline the overall framework of IBAC-Net. Subsequently, we will primarily focus on a novel integration of CNN and transformer, the brightness adaptive block, and the custom inception block, followed by providing experimental details.

The innovations detailed in this section are as follows: Firstly, addressing the shortcomings of existing image classification models in deep-level feature extraction, we propose a new model architecture named IBAC-Net. The transformer encoder module in IBAC-Net combines multiple head attention with local window attention, enabling simultaneous integration of global and local information to enhance the model's expression capabilities across different scales and ranges of images. Secondly, we address the impact of environmental factors during agricultural disease image data collection by designing a brightness adaptive block. This block introduces a trainable brightness adjustment curve, effectively resolving image quality issues caused by insufficient lighting and enhancing the model's adaptability to images in low-light conditions. Finally, we introduce a module called the custom inception block. This block incorporates the concept of flexible downsampling,

reducing the spatial dimensions of feature maps without compromising vital information, resulting in lower computational costs.

IBAC-net overview

We propose a novel image classification network architecture for the recognition and classification of plant diseases in complex agricultural environments and diverse weather conditions, aiming to reduce the incidence of plant diseases and protect the agricultural ecosystem. The proposed IBAC-Net model consists of three main components: the transformer encoder module, brightness adaptive block, and custom inception block. The overall framework of IBAC-Net is illustrated in Figure 3.

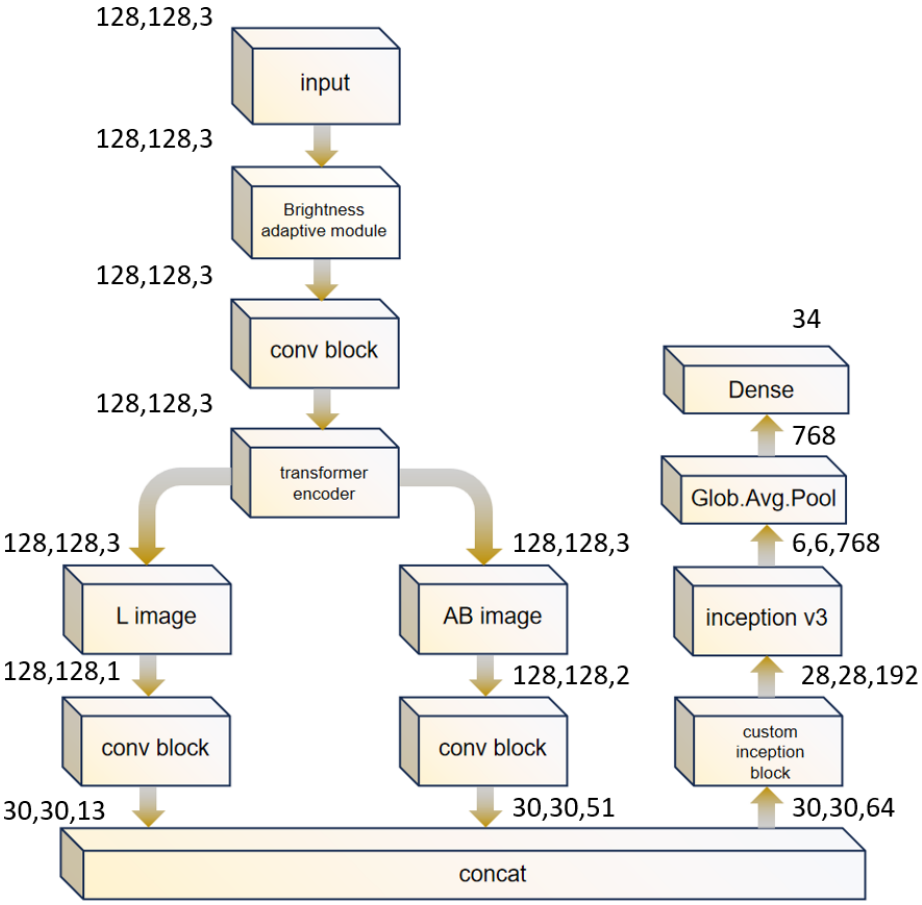

Figure 3. Overall framework of IBAC-Net.

Transformer encoder block

Convolutional Neural Networks (CNNs) and Transformers are two architectures that have achieved tremendous success in the field of deep learning. CNNs excel in image processing due to their

local perception, parameter sharing, and translation invariance. However, they have limitations in capturing global information and handling long-range dependencies (Sariturk and Seker, 2022). On the other hand, Transformers achieve global attention through self-attention mechanisms, making them suitable for sequence data processing with advantages in parallel computation. However, Transformers have a larger number of parameters and are sensitive to sequence length. The necessity of combining CNNs and Transformers lies in leveraging both local and global information in images, enhancing the model's versatility, flexibility, and overall modeling capability. Models that combine these architectures, such as the Vision Transformer (ViT), have shown significant progress in image processing tasks, bringing new ideas and performance improvements to image data processing. This fusion opens up more possibilities for the application of deep learning models across different tasks and domains.

The framework of the Transformer encoder part is illustrated in Fig. 4. Given that the multi-head attention mechanism is adept at capturing global information, while local window attention excels at capturing local information, we parallelize them in the output section. This approach allows us to fully leverage their respective strengths, providing the model with more comprehensive visual information modeling, thereby enhancing its performance and robustness. This proposed method differs from several existing approaches that primarily rely on CNNs for extracting local information and Transformers for extracting global information (Yang et al., 2023; Zhao et al., 2023). Instead, we use the multi-head attention mechanism to capture global relationships in images, which has advantages for overall image semantic understanding. Meanwhile, the local window attention mechanism can focus more finely on local areas, capturing local details and textures. By parallelizing their use, the model can simultaneously integrate global and local information, enhancing its ability to represent images across different scales and ranges. This approach allows for more flexible adjustment of the weights of global and local attention based on specific task requirements. The local window attention mechanism exhibits good computational efficiency when handling large-scale images, as it limits the range of attention at each position. Additionally, due to the introduction of local windows, it exhibits a certain sparsity, aiding in reducing computational complexity. By parallelizing with multi-head attention, the model can improve its ability to process global-local information in images without sacrificing computational efficiency. The multi-head attention mechanism is typically capable of handling multi-scale information, with each head focusing on learning features at different scales. When combined with the local window attention mechanism, it can better integrate multi-scale information at different levels, enhancing the model's perception of objects or structures at different

scales in images. The configuration of different heads and local windows enables the model to adapt to different image feature problems. The parallel use of these mechanisms helps the model better adapt to various scenes and image structures, improving its generalization and robustness.

The following is the detailed computation of the transformer encoder module:

Let the input be X and the output be Y. The following are the symbolic definitions of the main modules:

Multi-Head Attention:

$$MHA(X) \tag{1}$$

Feed Forward Network(FFN):

$$FFN(X) = W_2\big(ReLU(W_1X)\big) \tag{2}$$

Layer Normalization:

$$LN(X) \tag{3}$$

Dropout:

$$Drop(X) \tag{4}$$

Local Window Attention Block:

$$LWA(X) \tag{5}$$

The following is the process of Transformer Encoder:

①Multi-Head Self-Attention:

$$X_1 = LN\left(X + Drop\big(MHA(X)\big)\right) \tag{6}$$

②Position-wise Feed Forward Network:

$$X_2 = LN\left(X_1 = Drop\big(FFN(X_1)\big)\right) \tag{7}$$

③MLP Head:

$$X_3 = LN\left(X_2 = Drop\big(FFN(X_2)\big)\right) \tag{8}$$

④Local Window Attention Block:

$$Y = LN\left(X_3 = Drop\left(LWA(X_3)\right)\right) \tag{9}$$

Summarized in an equation:

$$Y = LN\left(X_3 + Drop\left(LWA\left(LN\left(X_2 + Drop\left(FFN\left(LN\left(X_1 + Drop\left(FFN\left(LN\left(X + Drop\left(MHA(X)\right)\right)\right)\right)\right)\right)\right)\right)\right)\right)\right) \tag{10}$$

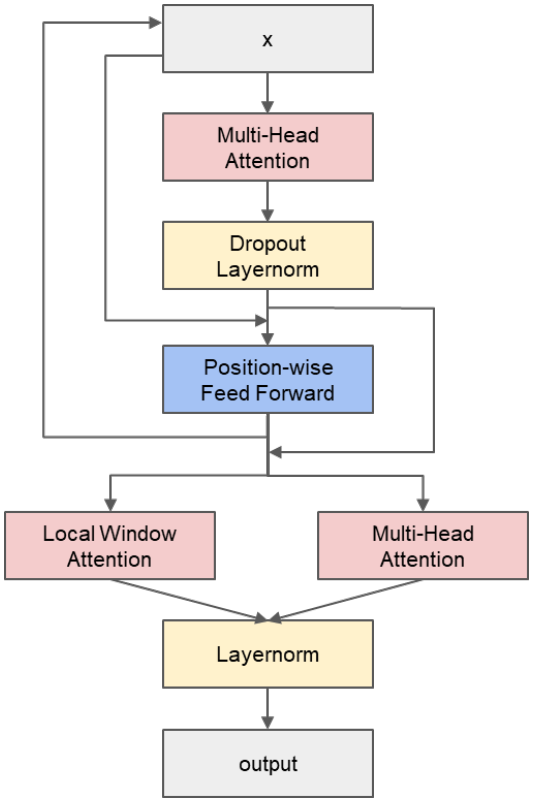

Figure 4 Transformer encoder framework.

Brightness adaptive block

Due to various environmental or technical constraints, many photographs are often captured under suboptimal lighting conditions. These conditions include insufficient or uneven lighting in the environment, improper display of objects under extreme backlighting, and inadequate exposure during the image capture process. The quality of information transmission in these low-light photos significantly decreases. Adverse lighting conditions may lead to the misidentification of plant disease categories. In such scenarios, photos may lack detail, have less vivid colors, and may even appear blurry or contain noise. Therefore, these low-light photos require post-processing or the use of professional equipment to enhance image quality for more accurate information conveyance.

Our goal, based on the methods initially proposed in (Guo et al., 2020; Wen et al., 2023), is to design a solution for the low brightness issue in photos collected during plant disease data acquisition due to insufficient lighting. To adjust the brightness of low-light photos to match normal lighting conditions, we have conceptualized the following considerations for the designed function: To maintain differentiation between adjacent pixels and prevent mutual influence, the function should avoid constant terms within specific segments; to ensure effective adjustment of every pixel in the photo, we normalize pixel values to the [0,1] range, allowing the designed function to adjust pixel brightness within this range and avoid information overflow; Linear functions, typically in the form of y=kx+b, struggle to flexibly adjust brightness solely through parameters k and b, whereas quadratic and higher-order functions can achieve this within the [0,1] range. However, functions of cubic order or higher involve excessive parameters. To minimize computational complexity while adjusting image brightness, we opt for a quadratic adjustment curve.

The adjustment curve can be represented as:

$$SAL\big(I(c,x);\alpha(c,x)\big) = I(c,x) + I(c,x) \otimes \big(1 - I(c,x)\big) \otimes T\big(I(c,x)\big) \otimes \alpha(c,x) \quad (1)$$

$$T\big(I(c,x)\big) = \frac{Tanh(-I(c,x) + 0.65) + 1}{2} \quad (2)$$

$I(c,x)$ represents the normalized pixel brightness value in the range of [0,1], c represents the color channel in the RGB color space, x denotes the pixel coordinate, and $\otimes$ represents element-wise multiplication. $SAL(I(c,x);\alpha(c,x))$ is the enhanced version of the given input $I(c,x)$, $\alpha(c,x) \in (-1,1)$ is a trainable parameter that can adjust the size of the SAL curve.

To further optimize the adjustment curve and achieve more precise adjustments for low-light photos under the influence of the curve, we have decided to introduce a new function into the adjustment curve without increasing parameters. Since pixel values have been normalized to the [0,1] range in the previous step, this new function needs to operate within [0,1]. In order to ensure that the adjustment curve does not excessively adjust pixel brightness, the designed function should suppress high-brightness pixels around 0.5 while enhancing low-brightness pixels. Therefore, a constant of 0.5 can be introduced into the function.

Since both the adjustment curve and the designed function operate within [0,1] and are monotonic, to achieve more precise adjustments without working against the adjustment curve, the designed function should follow the same monotonic direction as the adjustment curve. Additionally, to prevent oscillation in the adjustment curve, the constant 0.5 can be adjusted to allow the adjustment curve to act on pixel values earlier. The specific steps for this operation are as follows:

To improve the training convergence of the curve and prevent excessive adjustments on low-light images, we introduced the $T(I(c,x))$ function. Since the Tanh function outputs values between -1 and 1, to adjust pixel brightness without causing overexposure or artifacts, we need to scale the overall function values to the [0,1] range. By leveraging the monotonically increasing nature of the Tanh function, we achieve the following effects: suppressing pixels with brightness values exceeding 0.65 while simultaneously enhancing pixels with brightness values below 0.65.

Custom inception block

While the Inception structure has shown excellent performance in image classification tasks (Peng et al., 2022), it faces the challenge of high computational costs, limiting its application on resource-constrained devices. One major reason is that the Inception structure contains numerous convolutional layers and parameters, resulting in a large computational workload that requires substantial resources for training and inference. Therefore, to address such situations, we propose using a flexible sampling strategy to improve upon this. By employing a flexible downsampling layer (Han et al., 2021) that is not restricted to integer strides, we can dynamically adjust the size of feature maps without compromising the performance of the Inception structure in image classification tasks. This helps reduce the computational load of the model and enhances its adaptability to small-sized images.

We believe that a simple stack of convolutional layers may not adequately capture deeper-level information (Wang et al., 2022). Inspired by the Inception structure (Chollet, 2017), we initially designed the custom Inception block.

Following the flexible downsampling strategy initially proposed in (Xu et al., 2023), our goal is to optimize this strategy based on the requirements of image classification and integrate it with the original custom inception block to create the final module. The objective of our flexible downsampling approach is to efficiently adjust the spatial dimensions of input feature maps with minimal computational cost, thereby avoiding information loss in the early stages of the network. The core idea of this method is to use non-integer strides, which can flexibly adjust the output size to match the desired shape. Specifically, assuming our input data size is $H_{in} \times W_{in}$, the convolutional kernel size is $k_H \times k_W$, and the stride is $s_H \times s_W$, the output size after the convolution operation can be calculated as follows:

$$H_{\text{out}} = \left\lfloor \frac{H_{in} - k_H}{s_H} \right\rfloor + 1 \tag{3}$$

$$W_{\text{out}} = \left\lfloor \frac{W_{in} - k_W}{s_W} \right\rfloor + 1 \tag{4}$$

Here, $\lfloor \bullet \rfloor$ represents the floor operation. When the stride is an integer, the output size is typically an integer. However, when the stride is non-integer, the output size can be non-integer, allowing for a more precise adjustment of the output size to match the target shape. For example, suppose we need to adjust the input size by $H_{in} \times W_{in}$ to match the target size $H_{target} \times W_{target}$. We can achieve this by adjusting the stride. If the stride is $\frac{H_{\text{in}} - H_{target}}{n} \times \frac{W_{\text{in}} - W_{target}}{m}$, where n and m are positive integers, then the output size after the convolution operation can reach the target size. The choice of such strides ensures that the convolution kernel's sliding distance on the input matches the proportion of the target size, thereby achieving alignment with the desired shape. Specifically, we calculate the downsampling factor between the input and target shapes and then adjust the feature map using multiple bilinear interpolations. The advantage of this flexible downsampling method lies in avoiding the computational burden of evenly sampling the input image, as required in traditional integer-stride convolutions. In traditional convolutions, if the target shape does not match the input shape, additional operations such as zero-padding or cropping are often needed, introducing extra computational overhead. The flexible downsampling method, by selecting non-integer strides, aligns the convolution kernel's sliding steps on the input image with the target shape, thereby avoiding this additional computational burden. The

flexibility of the downsampling method allows for the flexible selection of downsampling factors based on actual needs and the importance of features in the feature map. For instance, for critical details in the feature map, smaller downsampling factors or no downsampling can be chosen to preserve these key details. Conversely, larger downsampling factors can be selected for relatively less important or redundant information, reducing the computational load. To enhance the model's expressive power, we introduce a custom deep convolutional block. This block comprises different types of convolutional layers and is combined with the flexible downsampling method through residual connections. To maintain information integrity and effectively fuse features at different scales, we employ a comprehensive strategy. First, we use the flexible downsampling technique to reduce the image's spatial resolution, focusing more on local detail features. Then, we perform average pooling on the downsampled feature map to capture the average feature representation of local regions. Simultaneously, we use upsampling to restore the pooled feature map to its original resolution and fuse it with the output of the deep convolutional block. Compared to traditional downsampling methods like max pooling and average pooling, our flexible downsampling method can dynamically adjust the feature map's size based on the target shape, thus reducing the feature map's spatial dimensions with lower computational costs while preserving important information. Moreover, our module effectively maintains information flow through residual connections. In the residual block, if the input and output signals have the same dimensions, they are directly added; if the dimensions differ, the input signal is adjusted to the output signal's dimensions through linear transformation (such as 1x1 convolution) before addition. This ensures that even in deep layers of the network, information from the input signal can directly propagate to the output, preserving information integrity. Our module mitigates the problem of vanishing or exploding gradients in deep networks, enhancing model stability and learning capacity.

In the upcoming experiments, we will verify the performance of these two methods in agricultural disease image classification tasks, particularly focusing on their effectiveness in enhancing classification accuracy and model stability. We are confident that these innovative methods will offer more efficient solutions for disease detection in the agricultural domain.

Results

This section focuses on comparative experiments and ablation studies of classification networks, discussing the background of plant disease images, the current status of plant disease symptoms, and the impact of weather conditions on classification results.

The section comprises four subsections: training settings, model evaluation metrics, comparative experiments, and ablation studies. Section 3.1 outlines the training settings of the model, including but not limited to the ratio of training and validation sets, selection of learning rates, batch size configuration, among others. In Section 3.2, evaluation metrics used to assess model performance are introduced, emphasizing the importance of selected metrics for subsequent performance validation and comparison with other methods. Section 3.3 compares the proposed method with existing approaches, including comparisons with other classical image classification methods and the latest research findings. The objective of this part is to demonstrate the superiority of the proposed method and discuss its innovative aspects compared to other methods in the field. Section 3.4 conducts ablation experiments on the brightness adaptive block and custom inception block to verify the contribution of each block to model performance.

Training setting

The experiments are based on TensorFlow and utilize two GPU 3090s. The training settings for the network proposed in this paper, as well as the training settings for comparative experiments and ablation studies, are as follows: input shape, number of epochs, and batch size are set to (128, 128, 3), 30, and 16, respectively.

We partition the entire dataset into training, validation, and testing sets, using 60% of the data for training, 20% for validation, and 20% for testing. The training set consists of 13,891 images, the validation set consists of 4,630 images, and the testing set consists of 4,630 images. This partition ensures that the model adequately learns the features of the data during training and evaluates its generalization ability on the validation set, thereby gaining a better understanding of the model's performance.

Throughout the training process, we record and save the performance metrics of the model on the training, validation, and testing sets for subsequent performance analysis and comparative experiments.

Model evaluation indicators

In this section, we compare and explain the accuracy, loss value, F1 score, and model parameter count of the training, validation, and testing sets. We use accuracy, loss, F1 score, and model parameter count to measure the accuracy of the classification method. Equations (15)-(18) provide the formulas for accuracy (ACC), positive predictive value (PPV), true positive rate (TPR), and F1 score.

$$TPR = \frac{TP}{TP + FN} \tag{15}$$

$$PPV = \frac{TP}{TP + FP} \tag{16}$$

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

$$F1\ Score = \frac{2 * (TPR * PPV)}{(TPR + PPV)} \tag{18}$$

Comparison experiment

To assess the effectiveness of IBAC-Net, it was compared against six widely used classification networks: ResNet50, ResNet101, VGG16, VGG19, ViT (Dosovitskiy et al., 2020), and SwinT(Liu et al., 2021) , Repvit (Wang et al., 2024), EfficientViT (Cai et al., 2022) and EfficientViT (Liu et al., 2023). Such a selection aids in comprehensively evaluating IBAC-Net's relative performance in agricultural disease image classification.

In the comparative experiments, to control variables and accurately depict the models' effects, parameters were uniformly set to: input shape = (128, 128, 3), epoch = 30, batch size = 16. The entire dataset was divided into training, validation, and testing sets, with 60% of the data used for training, 20% for validation, and 20% for testing. The training set comprised 13,891 images, the validation set contained 4,630 images, and the testing set included 4,630 images. The results of the experiments are shown in Table 2.

Table 2. Experimental results of the collected dataset.

Models	train			val			test			Para (M)
	ACC	loss	F1 score	ACC	loss	F1 score	ACC	loss	F1 score	
ResNet101	92.1	0.273	0.918	89.0	0.333	0.887	87.9	0.390	0.875	42.6
ResNet50	93.5	0.258	0.931	90.2	0.327	0.897	89.3	0.351	0.889	23.6

VGG16	89.1	0.338	0.891	85.9	0.439	0.856	84.3	0.460	0.840	134.4
VGG19	86.6	0.430	0.867	83.2	0.500	0.830	81.9	0.671	0.815	139.7
ViT	88.5	0.397	0.883	87.4	0.404	0.874	86.9	0.425	0.861	86.4
Swin_T	92.2	0.258	0.921	89.1	0.341	0.890	88.2	0.356	0.879	86.8
RepViT	93.43	0.221	0.933	91.75	0.280	0.915	**90.53**	**0.318**	**0.903**	7.8
EfficientViT (Han Cai et al.)	93.16	0.214	0.932	91.88	**0.255**	0.916	89.49	0.339	0.893	246.0
EfficientViT (Xinyu Liu et al.)	89.56	0.348	0.893	88.52	0.397	0.883	86.97	0.434	0.865	**2.2**
IBAC-Net	**94.4**	**0.156**	**0.950**	**92.5**	0.264	**0.927**	90.1	0.332	**0.903**	6.0

The training results in Table 2 indicate that IBAC-Net achieved the highest accuracy among all models, reaching 94.4%. This suggests that, compared to other models, IBAC-Net has a strong capability in accurately classifying plant leaf diseases. The loss value for IBAC-Net was significantly lower than that of most other models, at only 0.156, indicating that it effectively reduces prediction errors during training. IBAC-Net also had the highest F1 score, reaching 0.950, which combines precision and recall metrics. This suggests that IBAC-Net not only accurately identifies positive cases but also minimizes false positives and false negatives. The validation results in Table 2 show that IBAC-Net achieved an accuracy of 92.5%, the highest among all models. This demonstrates that, relative to other models, IBAC-Net has a stronger ability to classify plant leaf diseases correctly. The loss value for IBAC-Net was significantly lower than that of most other models, at 0.264, indicating that it effectively reduced prediction errors during training. IBAC-Net's F1 score reached 0.927, reflecting its superior performance in balancing precision and recall. This suggests that it effectively identifies positive cases while minimizing false positives and false negatives. The test results in Table 2 indicate that IBAC-Net achieved the highest accuracy, the lowest loss value, and a relatively small parameter size, leading to the best classification performance.

To further validate the model's generalizability and reliability, we conducted comparative experiments using the publicly available PlantVillage dataset. To control for variables and more accurately assess the model's performance, we standardized the parameters as follows: input shape = (128, 128, 3), epoch = 30, batch size = 16. The dataset was divided into training, validation, and test sets, with 60% of the data allocated to the training set, 20% to the validation set, and 20% to the test set. Specifically, the training set consisted of 32,581 images, the validation set contained 10,861 images, and the test set also comprised 10,861 images.

Table 3. Experimental results of the PlantVillage dataset.

Models	train			val			test			Para (M)
	ACC	loss	F1 score	ACC	loss	F1 score	ACC	loss	F1 score	
ResNet101	98.7	0.106	0.987	97.9	0.110	0.977	96.1	0.199	0.959	42.6
ResNet50	99.0	0.096	0.990	97.7	0.102	0.977	96.9	0.209	0.969	23.6
VGG16	98.4	0.163	0.984	97.3	0.179	0.973	96.7	0.203	0.966	134.4
VGG19	97.7	0.223	0.977	96.7	0.234	0.965	95.3	0.263	0.952	139.7
ViT	97.92	0.066	0.979	96.18	0.111	0.962	95.89	0.136	0.959	86.4
Swin_T	97.93	0.068	0.979	96.64	0.101	0.966	94.94	0.154	0.950	86.8
RepViT	98.41	0.048	0.984	96.60	0.107	0.964	95.43	0.171	0.953	7.8
EfficientViT (Han Cai et al.)	**99.12**	0.031	**0.991**	98.25	0.055	0.982	96.50	0.105	0.965	246.0
EfficientViT (Xinyu Liu et al.)	97.98	0.0657	0.980	95.6	0.143	0.956	94.6	0.169	0.945	**2.2**
IBAC-Net	99.09	**0.030**	**0.991**	**98.43**	**0.050**	**0.984**	**97.17**	**0.062**	**0.972**	6.0

IBAC-Net demonstrates exceptional performance on the PlantVillage dataset, particularly with an accuracy of 97.17% on the test set. This high accuracy indicates that the model effectively classifies plant disease images. The model achieved accuracies of 99.09% on the training set and 98.43% on the validation set, highlighting its strong generalization across different datasets. In terms of loss function performance, IBAC-Net recorded a loss of 0.030 on the training set, 0.050 on the validation set, and 0.062 on the test set. These relatively low values suggest that the model learns effectively during training and reduces misclassifications efficiently. The low loss values during training and validation indicate good convergence and minimal overfitting. Notably, IBAC-Net has a parameter count of 6.0M, significantly lower than other models such as EfficientViT (Han Cai et al.) with 246.0M parameters and VGG19 with 139.7M parameters. This smaller parameter size provides IBAC-Net with a distinct advantage in computational efficiency. The reduced parameter count implies lower computational and storage requirements, making it suitable for deployment in resource-constrained environments, such as edge devices or mobile platforms. Overall, IBAC-Net's high accuracy and low loss values indicate its superior performance in plant disease classification tasks. Its relatively small parameter size further enhances its versatility, allowing it to perform exceptionally well even in environments with limited resources. Considering accuracy, loss values, and parameter count, IBAC-Net represents an excellent model with a strong balance between performance and efficiency.

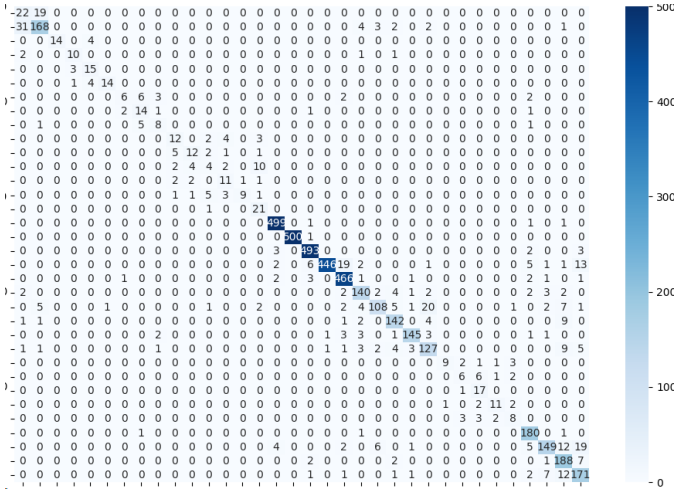

Figure 5. Confusion Matrix for IBAC-Net.

Figure 5 shows the confusion matrix generated by IBAC-Net after training on the collected dataset. From the confusion matrix, it can be seen that the dark blue blocks on the main diagonal represent the correct classification of the model, and the darker the colour, the better the classification. Most of the categories have higher values on the diagonal, indicating that the model has a higher recognition accuracy on these categories. In particular, several categories in the middle part are more effective in classification. However, non-zero values on the off-diagonal indicate misclassification situations. There is confusion between certain categories, which may be due to the similarity of the features of these categories or the uneven distribution of the training data. Although there are not many misclassification cases, the model performs poorly on some categories, especially those with less data or less distinctive features. Overall, the model works well on most of the categories, but there are errors on some categories, which may need to be improved by increasing the amount of data or optimising the model architecture.

IBAC-Net's design takes into account the specific characteristics of agricultural disease images, combining deep learning with traditional image processing techniques to better capture key features in the images. Compared to generic deep learning models, IBAC-Net demonstrates more specificity and adaptability in agricultural image classification tasks. IBAC-Net incorporates an image brightness adaptive block, which automatically adjusts the image brightness during data collection and processing, thereby reducing image variations caused by weather changes. This step enhances the model's robustness, ensuring accurate identification of agricultural diseases under different environmental conditions. IBAC-Net adopts a custom inception block for the fusion of shallow and deep features, which aids in better extracting important information from images, thus enhancing the model's classification capability and accuracy.

IBAC-Net outperforms networks such as ResNet50, ResNet101, VGG16, VGG19, ViT, and SwinT, Repvit, EfficientViT (Cai et al.) and EfficientViT (Liu et al.) in agricultural image classification tasks due to several key advantages:

Specialization and Relevance: IBAC-Net is specifically designed for agricultural image features, leading to higher accuracy and robustness in agricultural disease image classification tasks.

Module Optimization: The introduction of new modules, strategic placement optimization, and the design of the custom inception block enable IBAC-Net to better capture crucial information in images, thereby improving classification accuracy.

Intelligent Data Processing: The application of an image brightness adaptive block allows IBAC-Net to handle variations in image brightness under different weather conditions, ensuring model stability and reliability.

Our proposed network achieves the highest accuracy of 90.1%, the lowest loss value of 0.22, and the highest F1 score of 0.94 on the testing set for plant disease classification. Despite maintaining high performance, the model parameter count for IBAC-Net is only 6M, demonstrating a well-balanced relationship between performance and model complexity. The smaller parameter count contributes to efficient training and inference of the model, making it more feasible in resource-constrained environments.

IBAC-Net demonstrates multiple advantages in agricultural disease image classification tasks, including high accuracy, low loss value, a smaller model parameter count, and comprehensive comparison with other networks. These results further validate IBAC-Net as an effective image classification network with broad potential and reliability in practical applications.

Ablation experiment

In the ablation experiments, we employed the same training, validation, and testing sets, along with identical training configurations, hyperparameters, and evaluation metrics as described in section 3.1. This ensures the comparability of experimental results and eliminates the influence of other factors on the experimental outcomes.

Brightness adaptive block

Below is the composition of the SAL curve-adjusting network:

Initialization Function: Initialize the ReLU activation function to introduce non-linearity into the network; Define seven convolutional layers, named e_conv1 to e_conv7, using 3x3 convolutional kernels with 1-pixel padding. Each convolutional layer has a different number of output channels; Define the max-pooling layer, maxpool, with a 2x2 pooling window and a stride of 2; Define the upsampling layer, upsample, using bilinear interpolation for upsampling.

Forward Propagation Function: Input x represents the image; Pass through the e_conv1 convolutional layer and apply the ReLU activation function to get x1; Pass through the e_conv2, e_conv3, and e_conv4 convolutional layers, applying ReLU activation function to get x2, x3, x4; Concatenate x3 and x4 along the channel dimension, then pass through the e_conv5 convolutional layer

to get x5; Concatenate x2 and x5 along the channel dimension, then pass through the e_conv6 convolutional layer to get x6; Concatenate x1 and x6 along the channel dimension, then pass through the e_conv7 convolutional layer to get x_r.

Feature Segmentation and Enhancement: Split x_r into 8 parts along the channel dimension, labeled r1 to r8.

Apply a series of enhancement operations to the input image x, each step involves element-wise operations between the original image and the previous enhancement results, using an enhancement function in the form of Equation 1. This type of enhancement function enables the network to learn non-linear features of the image.

Return Value: Return two enhanced images and a tensor containing all enhancement coefficients.

By applying the trained SAL curve, we illustrate the before-and-after image effects in Figure 6.

Figure. 6 SAL curve processing image effect.

To thoroughly validate the substantial contribution of the brightness adaptive block to the model's performance, we processed the collected dataset to simulate situations with low brightness without adding weather effects. The purpose of this step is to mimic the conditions of agricultural disease image collection under insufficient lighting, such as cloudy days or post-dusk scenarios. This

design helps us more accurately evaluate the performance improvement of the brightness adaptive block under low-light conditions.

Table 4 .Brightness adaptive block ablation experiment.

Stage	Index	Baseline	brightness adaptive block
train	Train_ACC(%)	96.2	**97.80**
	Train_loss	0.12	**0.0560**
	Train_f1 score	0.95	**0.9771**
validation	Val_ACC(%)	90.1	**90.48**
	Val_loss	0.39	**0.3580**
	Val_f1 score	0.90	**0.9063**
test	Test_ACC(%)	85.6	**90.28**
	Test_loss	0.45	**0.3746**
	Test_f1 score	0.86	**0.9034**

The results of the ablation experiment demonstrate the significant advantage of introducing the brightness adaptive block on this specialized dataset. Specifically, compared to the scenario without the brightness adaptive block, the model's accuracy improved by 4.68% on the dataset enhanced by the module. This indicates that the brightness adaptive block successfully helps the model overcome the challenges of inadequate lighting conditions, enhancing its performance in complex environments. Detailed experimental results are shown in Table 4, listing the performance metrics of the model with and without brightness adaptive processing.

These results further validate the importance of the brightness adaptive block in agricultural disease image classification tasks, especially in real-world scenarios with insufficient lighting conditions. The experiment underscores that employing a brightness adaptive block for images captured in low-light environments effectively boosts model performance, thereby improving classification accuracy and providing robust support for agricultural disease monitoring and identification.

Based on these experimental findings, future efforts can focus on optimizing the design of the brightness adaptive block to enhance its adaptability to different lighting conditions, further improving

model performance. Additionally, exploring the integration of other image enhancement techniques and optimizing model structures can further enhance accuracy and robustness in agricultural disease image classification tasks.

Custom inception block

In order to thoroughly assess and validate the contribution of the custom inception block to model performance, we conducted a series of ablation experiments. This custom inception block is considered a key innovation aimed at overcoming the limitations of simple convolutional stacking structures in capturing deep-level information. In this section, we will introduce the settings, methods, and results of the ablation experiments.

We kept the rest of the model unchanged and only made adjustments to the custom Inception Block. We designed a control group (Baseline) where the entire custom inception block was removed and an experimental group where the custom inception block was retained. This setup allows us to accurately evaluate the impact of this module on model performance. Specific experimental results are presented in Table 5.

Table 5. Custom inception block ablation experiment.

Stage	Index	Baseline	custom inception block
train	Train_ACC(%)	96.2	**96.9**
	Train_loss	0.12	**0.11**
	Train_f1 score	0.95	**0.96**
validation	Val_ACC(%)	90.1	**90.7**
	Val_loss	0.39	**0.36**
	Val_f1 score	0.90	**0.91**
test	Test_ACC(%)	85.6	**91.4**
	Test_loss	0.45	**0.32**
	Test_f1 score	0.86	**0.91**

Based on the ablation experiment results for the custom Inception Block, it significantly improved performance in agricultural disease image classification tasks. Here is a detailed analysis of

the experimental results: After processing with the custom inception block, the model achieved higher classification accuracy in agricultural disease image classification. From the ablation experiment data, we found that under the same dataset and training conditions, the model using the custom Inception Block performed significantly better in the classification task, with an accuracy improvement of 5.8% on the test set.

Based on these experimental results, future work can focus on further optimizing the design of the custom inception block to enhance its adaptability to images and further improve model performance. Additionally, exploring the integration of other image enhancement techniques and model structure optimization strategies can further enhance the accuracy and robustness of agricultural disease image classification tasks.

Discussion

In this chapter, we delve into the entire research process and summarize the key achievements and innovations of our study.

Our research aims to address challenges in agricultural disease image classification, including the accurate identification and classification of plant diseases in complex agricultural environments and varying weather conditions. In the initial stages of our research, we conducted extensive research and analysis of existing image classification models, identifying shortcomings in their ability to extract deep-level information and handle images of different scales.

To address these issues, we propose a novel image classification network architecture, termed IBAC-Net. This network combines the strengths of CNNs and transformers while incorporating a brightness adaptive block and custom inception block to enhance the model's performance and adaptability. The innovations of our study are primarily reflected in the following aspects:

We propose a novel image classification network architecture that integrates CNNs and transformers, combining global and local information processing to improve the model's ability to represent images across different scales and ranges.

We address the issue of image quality due to insufficient lighting by designing a brightness adaptive block. This effectively resolves issues of image information loss and reduced recognition accuracy under low-light conditions.

We design custom inception block and introduce a flexible downsampling strategy, reducing the computational complexity of the model while preserving the integrity of crucial information, thus enhancing the model's expressive power and robustness.

These innovations are validated through experiments, demonstrating improved classification performance and stability, thereby providing a more effective solution for agricultural disease detection. During the research process, we encountered several challenges and issues. These included uneven lighting conditions during data collection and poor image quality, among others, which directly impacted the training and classification performance of the model. To address these challenges, we implemented a series of measures, such as data preprocessing and model optimization, ultimately achieving promising experimental results. In addition to the aforementioned innovations, we also made refinements and optimizations to other parts of the model. For instance, in the dual-branch processing of images into color and grayscale, we switched to using residual connections for the convolutional blocks. This change improved the network's performance and expressive capability, facilitated feature reuse and sharing, and deepened the network's architecture. Additionally, we replaced stacked convolutions with a dual-branch structure comprising stacked convolutions and average pooling. This adjustment enriched the model's feature representation capacity, enhanced its nonlinear representation ability, and improved the model's robustness and generalization capability. These refinements and optimizations resulted in our model demonstrating good performance in experiments and exhibiting strong generalization and robustness. During the experimental validation phase, we utilized a diverse dataset to evaluate and test the performance of our proposed IBAC-Net model in agricultural disease image classification tasks, aiming to demonstrate its effectiveness and superiority. We employed datasets containing various common crop diseases and conducted rigorous data preprocessing and labeling to ensure the accuracy and reliability of our experiments.

Firstly, we compared the IBAC-Net model with traditional image classification models like ResNet and Inception in terms of classification accuracy, model stability, and generalization capability. The experimental results showed that the IBAC-Net model achieved high classification accuracy across different datasets, demonstrating good stability and generalization capability. Particularly in handling uneven lighting conditions and poor image quality, the IBAC-Net model exhibited significant advantages, effectively improving disease image recognition and classification accuracy.

Secondly, we conducted a detailed analysis of the model's training process, including learning rate adjustment, optimizer selection, batch size configuration, and other aspects. By fine-tuning these

training parameters, we further enhanced the model's performance and convergence speed, making it more suitable for agricultural disease image classification tasks.

In addition to classification accuracy, we evaluated other metrics of the model, such as computational complexity, inference speed, and parameter count. The experimental results revealed that while maintaining high accuracy, the IBAC-Net model exhibited relatively low computational complexity and parameter count, with fast inference speed. This indicates the model's practical feasibility and usability in real-world applications.

Finally, we conducted an in-depth analysis of the model's performance across different scenarios and datasets, discussing performance variations and optimization strategies when dealing with specific diseases, different lighting conditions, and image qualities. Through experimental validation, we confirmed the effectiveness and superiority of our proposed approach in agricultural disease image classification tasks, providing reliable support and justification for its practical application.

ConclusionS

In this study, we proposed a novel image classification network architecture, named IBAC-Net, for the recognition and classification of agricultural disease images. This model incorporates a series of innovative designs, including a brightness adaptive block and custom inception block, to address the challenges posed by complex agricultural environments and changing weather conditions. Through detailed experiments and comparative analysis, we draw the following conclusions:

Firstly, the introduction of the brightness adaptive block significantly improves the model's performance under low lighting conditions. Through disintegration experiments conducted on datasets simulating adverse weather conditions, we observed a notable increase in accuracy when handling images with lower brightness, validating the effectiveness of the brightness adaptive block under insufficient lighting conditions.

Secondly, the design of the custom inception block has yielded significant results in exploring deep-level information. By comparing the baseline model with the control group that removed the custom inception block, we found that the addition of the custom inception block resulted in better classification accuracy, precision, and recall on the test set, demonstrating its crucial role in enhancing the model's understanding and learning of deep-level features.

Finally, by comparing it with six widely used classification networks, we validated the effectiveness of IBAC-Net in agricultural disease image classification. IBAC-Net performed excellently across all performance metrics, showcasing its superiority over both classical and emerging models.

Overall, our method has achieved satisfactory performance in addressing the complexities of agricultural disease image environments and variable weather conditions. The innovative design of IBAC-Net provides a reliable solution for the protection of agricultural ecosystems and the early identification of crop diseases. Future research can further explore the model's generalization capabilities and its practical application in real agricultural fields.

References

Abdu, A.M., Mokji, M.M., Sheikh, U.U. (2020). Deep learning for plant disease identification from disease region images. In: Intelligent Robotics and Applications: 13th Int. Conf. vol 12595. Cham, Springer.

Arman, S.E., Bhuiyan, M.A.B., Abdullah, H.M., Islam, S., Chowdhury, T.T., Hossain, M. A. (2023). BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning. Data Brief 50:109608.

Arshad, F., Mateen, M., Hayat, S., Wardah, M., Al-Huda, Z., Gu, Y.H., Al-antari, M.A. (2023). PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction. Alexandria Eng. J. 78:406-418.

Ashwini, C., Sellam, V. (2023). EOS-3D-DCNN: Ebola optimization search-based 3D-dense convolutional neural network for corn leaf disease prediction. Neural Comput. Appl. 35:11125-11139.

Bhuiyan, M.A.B., Abdullah, H.M., Arman, S.E., Rahman, S.S., Al Mahmud, K. (2023). BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases. Smart Agric. Technol. 4:100214.

Cai, H., Li, J., Hu, M., Gan, C., Han, S. (2022). Efficientvit: Multi-scale linear attention for high-resolution dense prediction. arxiv:2205.14756.

Cho, S.B., Jeong, S.H., Yu, J. W., Choi, J.B., Kim, M.K. (2023). Heterogeneous domain adaptation method for tomato leaf disease classification base on CycleGAN. J. Intell. Fuzzy Syst. 45:8859-870.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu. pp. 1800-1807.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929.

Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Seattle. pp. 1780-1789.

Gao, J., Xu, L., Huang, F. (2016). A spectral–textural kernel-based classification method of remotely sensed images. Neural Comp. Appl. 27:431-446.

Han, D., Yun, S., Heo, B., Yoo, Y. (2021). Rethinking channel dimensions for efficient model design. Proc. EEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville. pp. 732-741.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas. pp. 770-778.

Hughes D, Salathé M (2915). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060.

Kaggle.com [Internet]. 2020. Cacao diseases. Available from: https://www.kaggle.com/datasets/zaldyjr/cacao-diseases

Kaggle.com [Internet]. 2021. Coffee plant diseases. Available from: https://www.kaggle.com/datasets/coffeedisease/coffee-plant-disease

Kaggle.com [Internet]. 2021. Guava disease dataset (4 types). Available from: https://www.kaggle.com/datasets/omkarmanohardalvi/guava-disease-dataset-4-types

Kaggle.com [Internet]. 2023. Cotton plant diseases. Available from: https://www.kaggle.com/datasets/dhamur/cotton-plant-disease

Khamparia, A., Saini, G., Gupta, D., Khanna, A., Tiwari, S., de Albuquerque, V.H.C. 2019. Seasonal crops disease prediction and classification using deep convolutional encoder network. Circuit Syst. Signal Process 39:818–836.

Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. Commun. ACM 60:84-89.

Liu, X., Peng, H., Zheng, N., Yang, Y., Hu, H., Yuan, Y. (2023). Efficientvit: Memory efficient vision transformer with cascaded group attention. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Vancouver. pp. 14420-14430.

Liu, Y., Wang, Z., Wang, R., Chen, J., Gao, H. (2023). Flooding-based MobileNet to identify cucumber diseases from leaf images in natural scenes. Comput. Electron. Agr. 213:108166.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al., (2021). Swin transformer: Hierarchical vision transformer using shifted windows. Proc. IEEE/CVF Int. Conf. on Computer Vision, Montreal. pp. 10012-10022.

Mwebaze, E., Mostipak, J., Joyce, Elliott, J., Dane, S. (2020). Cassava leaf disease classification. Kaggle. Available from: https://kaggle.com/competitions/cassava-leaf-disease-classification

Narayanan, K.L., Krishnan, R.S., Robinson, Y.H., Julie, E.G., Vimal, S., Saravanan, V., Kaliappan, M. (2022). Banana plant disease classification using hybrid convolutional neural network. Comput. Intell. Neurosci. 2022:9153699.

Peng, C., Liu, Y., Yuan, X., Chen, Q. (2022). Research of image recognition method based on enhanced inception-ResNet-V2. Multimed. Tools Appl. 81:34345-34365.

Raja, M.R., Jayaraj, V., Shajin, F. H., Devi, E.R. (2023). Radial basis function Neural Network optimized with Salp Swarm algorithm espoused paddy leaf disease classification. Biomed. Signal Proces. 86:105038.

Raja, P., Olenskyj, A., Kamangir, H., Earles, M. (2021). Simultaneously predicting multiple plant traits from multiple sensors via deformable CNN regression. arXiv: 2112.03205.

Sariturk, B., Seker, D.Z. (2022). A residual-inception U-Net (RIU-Net) approach and comparisons with U-shaped CNN and transformer models for building segmentation from high-resolution satellite images. Sensors (Basel) 22:7624.

Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K. (2020). Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agr. 175:105527.

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. Proc. 7th ACM IKDD CoDS and 25th COMAD.(pp. 249-253.

Wang, A., Chen, H., Lin, Z., Han, J., Ding, G. (2024). Repvit: Revisiting mobile cnn from vit perspective. Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition. pp. 15909-15920.

Wang, X., Zhong, M., Cheng, H., Han, J., Zhou, Y., Ren, J., Liu, M. (2022). SpikeGoogle: Spiking Neural Networks with GoogLeNet-like inception module. CAAI T. Intell. Technol. 7:492-502.

Yang, X., Huo, H., Wang, R., Li, C., Liu, X., Li, J. (2023). DGLT-Fusion: A decoupled global–local infrared and visible image fusion transformer. Infrared Phys. Technol. 128:104522.

Yao, J., Tran, S. N., Garg, S., & Sawyer, S. (2024). Deep learning for plant identification and disease classification from leaf images: multi-prediction approaches. ACM Comput. Surv. 56:153.

Zeiler, M.D., Fergus, R. Visualizing and understanding convolutional networks In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds.), Computer Vision – ECCV 2014. Lecture Notes in Computer Science, vol 8689. Cham, Springer. pp. 818-833.

Zhao, F., Li, SJ., Zhang, J.J., Liu, H.Q. (2023). Convolution transformer fusion splicing network for hyperspectral image classification. IEEE Geosci. Remote Sensing Lett. 20:5501005.

Zhang, D., Huang, Y., Wu, C., Ma, M. (2023). Detecting tomato disease types and degrees using multi-branch and destruction learning. Comput. Electron. Agr. 213:108244.

Zhang, L., Xie, L., Wang, Z., Huang, C. (2022). Cascade parallel random forest algorithm for predicting rice diseases in big data analysis. Electronics (Basel) 11; 1079.

Zhang, Y., Cao, G., Li, X., Wang, B. (2018). Cascaded random forest for hyperspectral image classification. IEEE J. Sel. Top. Appl. 11:1082-1094.