doi:10.13682/j.issn.2095-6533.2025.06.013

面向害虫识别的互补特征融合双流网络

李大湘[1,2],孙家宁[1],刘 颖[1,2]

(1.西安邮电大学 通信与信息工程学院,陕西 西安 710121;
2.西安市公共安全图像处理技术及应用重点实验室,陕西 西安 710121)

摘 要:针对害虫类别间细节差异小、田间背景干扰严重及样本分布不均衡等问题,提出一种面向害虫识别的互补特征融合双流网络。该网络结合卷积神经网络的局部感知能力与 Mamba 模型的全局建模能力,捕获并融合害虫图像的全局与局部信息。设计层次化多尺度感知模块,通过分组层次化卷积提取多尺度图像特征,并采用细节强化感知策略增强害虫细节信息;设计自适应聚焦 Mamba 模块,利用动态卷积算子定位害虫关键区域,减少复杂背景干扰;设计注意力加权融合模块,通过交叉注意力机制实现全局和局部特征的自适应交互优化,进一步提升语义表达的准确性。最后,构建均衡损失函数,缓解数据集类别不平衡的影响。实验结果表明,该网络在大规模害虫数据集 IP102 上的准确率达到 71.19%,在 D0 数据集上的准确率为 99.36%,能够有效识别害虫种类。

关键词:害虫识别;Mamba;多尺度特征提取;特征融合;注意力机制

中图分类号:TP18;S435 **文献标志码**:A **文章编号**:2095-6533(2025)06-0113-10

A dual-stream network with complementary feature fusion for pest identification

LI Daxiang[1,2], SUN Jianing[1], LIU Ying[1,2]

(1. School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Xi'an Key Laboratory of Image Processing Technology and Applications for Public Security, Xi'an 710121, China)

Abstract: To address the challenges of small inter-class differences in pest details, severe field background interference, and imbalanced sample distribution, a complementary feature fusion dual-stream network for pest recognition is proposed. This network combines the local perception capability of convolutional neural networks with the global modeling ability of the Mamba model, capturing and integrating the global and local information of pest images. A hierarchical multi-scale perception module is designed to extract multi-scale image features through grouped hierarchical convolution and enhance pest detail information with a detail enhancement perception strategy. An adaptive focusing Mamba module is designed to locate key pest regions using dynamic convolution operators and reduce complex background interference. Additionally, an attention-weighted fusion module is designed to achieve adaptive interaction and optimization of global and local features through a cross-attention mechanism, further improving the accuracy of semantic expression. A balanced loss function is constructed to mitigate the effects of class imbalance in the dataset. The experimental results show that the network achieves an accuracy of 71.19% on the large-scale pest dataset IP102, and an accuracy of 99.36% on the D0 dataset, demonstrating its ability to effectively identify pest species.

收稿日期:2024-12-20

基金项目:国家自然科学基金项目(62071379);陕西省自然科学基金项目(2025JC-YBMS-663)

引文格式:李大湘,孙家宁,刘颖.面向害虫识别的互补特征融合双流网络[J].西安邮电大学学报,2025,30(6):113-122.
LI D X,SUN J N,LIU Y. A dual-stream network with complementary feature fusion for pest identification [J]. Journal of Xi'an University of Posts and Telecommunications,2025,30(6):113-122.

Keywords: pest recognition; Mamba; multi-scale feature extraction; feature fusion; attention mechanism

害虫的广泛传播和快速演变已对全球农业构成了严重威胁[1]。由于害虫种类繁多且具有一定的相似性,传统人工识别方法不仅耗时费力,且需要高度的专业知识与实践经验,在大规模农业生产中难以高效实施[2]。因此,利用图像处理技术研究害虫识别算法[3]快速且准确地识别害虫种类,并及时采取精细化防治措施,已成为保障农业可持续发展的关键[4]。

近年来,深度学习技术凭借其自动提取深层特征的能力,在复杂图像识别任务中展现了优越性能[5],逐渐成为害虫识别领域的研究热点[6]。现有的害虫识别网络主要可分为以下 3 类。第一类,卷积神经网络[7](Convolutional Neural Network, CNN):Liu 等[8]基于 Resnet 在每个残差块中使用三分支提取图像的多尺度特征,设计了改进的深度多分支融合残差(Deep Multibranch Fusion Residual Network, DMF-ResNet)网络,提高了害虫识别的准确率;Nandhini 等[9]设计视觉再生融合(Visual Regenerative Fusion Network, VRFNet)网络,利用数据增强技术和特征融合技术有效缓解了害虫识别任务中扭曲图像和遮挡图像的影响。然而,尽管 CNN 在提取局部特征方面表现优异,但受限于局部感受野,难以捕获复杂的全局特征。第二类,视觉变换器[10-11](Vision Transformer, ViT):Dosovitskiy 等[12]通过 ViT 建模图像中的远程依赖关系,捕捉更为全面的上下文信息。Ishak[13]将 ViT 引入植物病虫害识别任务中,通过将主干网络中的传统卷积结构替换为 SE-Net(Squeeze-Excitation Networks)提高精度,在 4 类玉米叶片病害数据集上达到了 99.24%的准确率。然而,由于 ViT 缺乏归纳偏置且具有二次计算复杂度,导致其训练缓慢且难以收敛,需要更大的数据集和更长的训练时间。第三类,集成学习(Ensemble Learning, EL)方法[14-15]:Ayan 等[16]对 7 个不同的预训练 CNN 模型进行微调和再训练,选择效果最好的 3 个模型进行集成,并使用遗传算法获得优化权重。EL 方法通过训练多个基分类器,选择合适的策略组合多个模型的预测结果,获得比单个模型更好的预测效果。但是,该方法需要训练和维护多个模型,计算和存储成本较高,难以适用于实际的农业场景。

尽管上述网络表现良好,但面对背景复杂、种类繁多且类别不平衡的害虫图像时,现有研究存在分类精度低、泛化能力较弱的问题。因此,为了提高真实农业场景中的大规模害虫图像的识别准确性,提出一种面向害虫识别的互补特征融合双流网络(Complementary Feature Fusion Dual-Stream Network, CFFDS-Net)。针对害虫类别间细节差异较小的问题,CFFDS-Net 在局部分支采用分层残差连接和细节强化感知的联合策略,捕捉多尺度的图像特征并增强网络对害虫微小差异的辨别能力。在全局分支通过远距离依赖建模和动态区域定位机制,以线性复杂度捕获全局语义信息,有效定位害虫关键区域,抑制背景干扰,并设计双向交叉注意力融合机制,实现全局与局部特征的自适应优化与协同整合。最后,构建均衡损失函数,缓解类别不平衡的负面影响,提升网络对少数类别害虫的识别精度。

1 CFFDS-Net 整体架构

CFFDS-Net 采用双流主干结构并行处理输入图像,旨在高效提取与融合害虫图像的局部细粒度特征与全局上下文特征,其主要由 4 个核心模块组成,即层次化多尺度感知(Hierarchical Multi-Scale Perception, HMSP)模块、自适应聚焦 Mamba(Adaptive Focus Mamba, AF-Mamba)模块、注意力加权融合(Attention-Weighted Fusion, AWF)模块和均衡损失函数 L_{EQ},整体架构示意图如图 1 所示。首先,对输入图像 $I \in R^{224\times224\times3}$ 进行初步特征处理。在 HMSP 分支中,采用由步长为 2 的 7×7 卷积层构成的 CNN Stem 处理输入图像,提取浅层特征的同时将图像分辨率减半,减少计算开销。AF-Mamba 分支则通过 Patch Partition 操作进行图像分块,生成分辨率为 56×56×64 初始特征图。在后续的特征提取过程中,网络分别对两条分支进行 4 个阶段(Stage1 至 Stage4)的特征下采样与表示学习。其中,HMSP 分支在每个阶段堆叠不同数量的 HMSP 模块,逐步生成 56×56×64、28×28×128、14×14×256 和 7×7×512 的多层次特征表示。为了提高网络对害虫局部细节的敏感度,在 HMSP 模块中设计细节强化感知策略,提取更加细致的局部细节特征 F_{local}。AF-Mamba 分支则在 Stage2 至

Stage4 阶段对特征图进行 Patch Merging 下采样操作，同时利用 AF-Mamba 模块的线性计算复杂度优势和长距离建模能力学习害虫图像的全局上下文关系，捕获全局聚焦特征 F_{global}。然后，利用 AWF 模块对 F_{local} 与 F_{global} 进行动态协同融合，自适应平衡 F_{local} 与 F_{global} 的贡献度，最终得到互补特征 Z。融合后的特征 Z 通过分类头与 Softmax 激活函数完成害虫分类任务，并在均衡损失函数 L_{EQ}（Equalization Loss，L_{EQ}）的作用下，改善类别不平衡问题，进一步提升少数类别的识别性能。

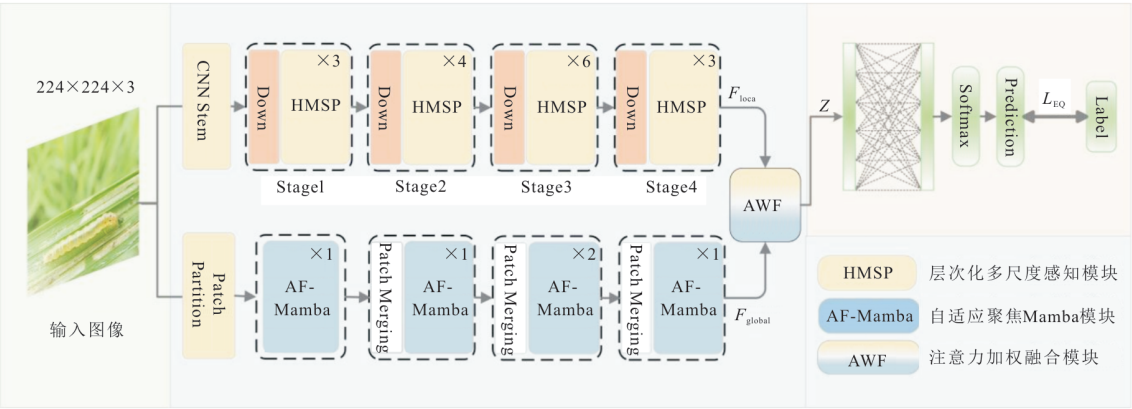


图 1　CFFDS-Net 整体架构示意图

2　模块设计

2.1　层次化多尺度感知模块

基于深度神经网络中感受野随层数增加而扩大的特点[17]，设计 HMSP 模块。与传统方法不同，HMSP 模块可以直观地分成 4 个操作步骤，即划分（Split）、分组层次化卷积、细节强化感知（Detail Enhancement Perception，DEP）与特征拼接（Concat）。总结起来，HMSP 模块主要是采用分组层次化卷积与 DEP 操作，且结合残差连接而不断累积扩展感受野，使模型能够在不同尺度上捕捉更加丰富的害虫图像细节信息，从而显著提升特征表达的全面性和精细度。HMSP 模块结构示意图如图 2 所示。

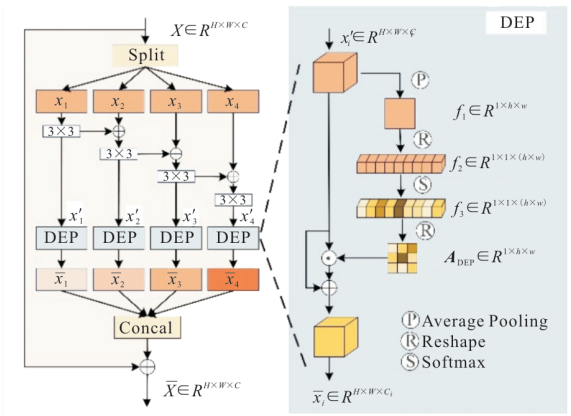


图 2　HMSP 模块结构示意图

给定 $X\in R^{H\times W\times C}$ 作为输入 HMSP 模块的特征图谱，其中 W、H 和 C 分别表示 X 的宽度、高度与通道数。首先，沿通道维度将 X 均匀划分为 4 个特征图子集 $\{x_i\}_{i=1}^{4}$，其中 $x_i\in R^{H\times W\times C_i}$ 且 C_i 满足 $C=4\times C_i$。随后，对每个子特征 x_i 依次执行一组卷积操作，表示为 $\Psi_i(\cdot)$，其中包括 3×3 标准卷积和批量归一化，生成对应的特征图子集 $x_i'\in R^{H\times W\times C_i}$。HMSP 模块最核心的设计是对每组 $\Psi_i(\cdot)$ 采用层次化的残差连接。这种设计使 $\Psi_i(\cdot)$ 不仅能处理当前层的特征图子集，还能整合前一层提取的特征信息，提升特征表达的丰富性。具体而言，第一个特征图子集 x_1 直接通过 $\Psi_i(\cdot)$ 操作生成 x_1'，然后第二个特征图子集 x_2 与前一层的输出 x_1' 相加，再经卷积层 $\Psi_i(\cdot)$ 操作生成 x_2'，重复该过程直到处理完所有特征图子集。上述计算过程可表示为

$$x_i'=\begin{cases}\Psi_i(x_i), & i=1\\ \Psi_i(x_i\oplus x_{i-1}'), & 1<i\leqslant 4\end{cases}\tag{1}$$

式中：⊕表示逐元素求和。

为了增强网络对害虫细节特征的感知能力，设计 DEP 策略进一步处理 x_i'，DEP 的具体过程如图 2 右侧部分所示。首先，对输入特征图 $x_i'\in R^{H\times W\times C_i}$ 进行全局平均池化（Global Average Pooling，GAP），得到 $f_1\in R^{1\times h\times w}$。接着将 f_1 重塑为 $f_2\in R^{1\times 1\times (h\times w)}$，并将 f_2 输入到 Softmax 函数中，生成权重矩阵。然后，将权重矩阵 $\boldsymbol{A}_{DEP}$ 与输入特征 x_i' 相乘，实现内容感知的注意力增强，得到增强后的特征图 $\bar{x}_i\in R^{H\times W\times C_i}$。这个过程可表示为

$$\begin{cases}f_2=\mathrm{Reshape}(G(x_i'))\\ \boldsymbol{A}_{DEP}=\mathrm{Reshape}^{-1}(\mathrm{Softmax}(f_2))\\ \bar{x}_i=\delta(x_i'\odot\boldsymbol{A}_{DEP}+x_i')\end{cases}\tag{2}$$

式中：Reshape 和 $\mathrm{Reshape}^{-1}$ 分别表示重塑操作和

其反向操作；$G(\cdot)$表示全局平均池化操作；δ 表示 Softmax 激活函数；$\odot$表示元素乘法。

最后，将 DEP 处理后的增强特征 $\bar{x}_i$ 拼接起来，并在 HMSP 中使用残差结构保留原始特征，最终得到输出特征图 $\overline{X}\in R^{H\times W\times C}$，其计算过程可表示为

$$\overline{X}=\mathrm{Concat}([\bar{x}_1,\bar{x}_2,\bar{x}_3,\bar{x}_4])+X \tag{3}$$

2.2 自适应聚焦 Mamba 模块

2.2.1 预备知识

状态空间模型[18]（State Space Model，SSM）源于现代控制理论中的线性时不变系统，具有线性复杂度，其核心是在连续时间 t 上，通过隐藏状态 $h(t)\in R^N$ 将输入信号 $x(t)\in R$ 映射为输出信号 $y(t)\in R$，该过程可用线性常微分方程表示为

$$\begin{cases}h'(t)=\boldsymbol{A}h(t)+\boldsymbol{B}x(t)\\ y(t)=\boldsymbol{C}h(t)\end{cases} \tag{4}$$

式中：$\boldsymbol{A}\in R^{N\times N}$、$\boldsymbol{B}\in R^{N\times 1}$ 和 $\boldsymbol{C}\in R^{1\times N}$ 分别为状态矩阵、输入矩阵和输出矩阵。

在深度学习中，为了更好地处理文本和图像等离散输入，使用零阶保持技术将连续 SSM 转换为离散 SSM，利用可学习的时间尺度参数 Δ 对 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 进行离散化，具体过程为

$$\begin{cases}\bar{\boldsymbol{A}}=\exp(\Delta\boldsymbol{A})\\ \bar{\boldsymbol{B}}=(\Delta\boldsymbol{A})^{-1}(\exp(\boldsymbol{A})-\boldsymbol{I})\cdot\Delta\boldsymbol{B}\end{cases} \tag{5}$$

式中：$\bar{\boldsymbol{A}}$ 和 $\bar{\boldsymbol{B}}$ 分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的离散化版本；$\boldsymbol{I}$ 为单位矩阵。

经式(5)离散化处理后，式(4)可转换为

$$\begin{cases}h_n=\bar{\boldsymbol{A}}h_{n-1}+\bar{\boldsymbol{B}}x_n\\ y_n=\bar{\boldsymbol{C}}h_n\end{cases} \tag{6}$$

式中：h_{n-1} 和 h_n 分别代表 $n-1$ 时刻与 n 时刻的状态信息。

在选择性状态空间模型 S6，也称为 Mamba[19] 中，矩阵 $\boldsymbol{B}$、$\boldsymbol{C}$ 和时间尺度参数 Δ 均由输入数据产生，使模型能够根据输入数据自适应调整参数，增强对输入上下文的感知能力。

2.2.2 AF-Mamba 模块设计

针对复杂背景下害虫识别任务中的全局依赖关系建模与背景噪声抑制的协同需求，设计 AF-Mamba 模块，该模块包括关键区域自适应聚焦器（Adaptive Key Region Focuser，AKRF）和 2D 选择扫描（2D-Selective-Scan，SS2D）机制[20]，结构示意图如图 3 所示。AKRF 采用与 Transformer[21] 相似的结构，结合可变形卷积网络[22]（Deformable Convolutional Networks v4，DCNv4）和混合前馈网络[23]（Mix-Feed Forward Network，Mix-FFN），通过残差连接将特征进行逐步聚合和传递，增强网络对复杂图像信息的捕捉能力。

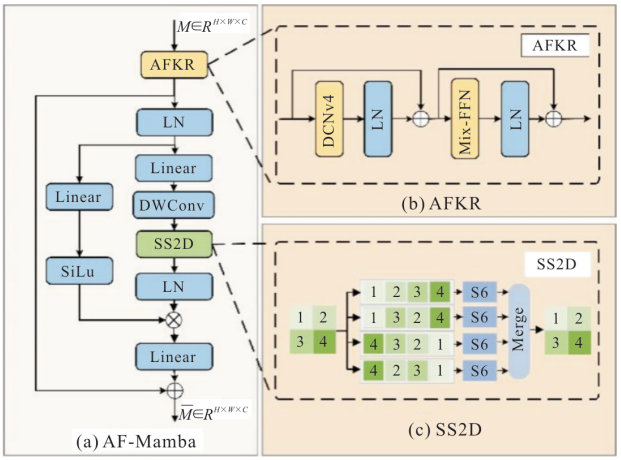


图 3 AF-Mamba 模块结构示意图

给定输入特征图 $M\in R^{H\times W\times C}$，其中 H、W 和 C 分别表示特征的高度、宽度与通道数。DCNv4 通过动态学习的偏移量 Δp_k 和调制权重 Δm_k，从输入特征图 M 中采样，生成具有更强空间感知能力的输出特征图 M_1，其计算过程可表示为

$$M_1(p)=\sum_{k=1}^{K}w_k\Delta m_k M(p+\Delta p_k) \tag{7}$$

式中：$M_1(p)$表示在每个位置 p 上聚合的动态特征；K 为卷积核大小，例如 3×3 时 $K=9$；Δm_k 和 Δp_k 分别为网络动态学习到的调制权重和偏移量，且 $\Delta m_k\in[0,1]$；$M(p+\Delta p_k)$表示输入特征图 M 在动态采样点 $p+\Delta p_k$ 处的值。

DCNv4 的输出特征 M_1 经过层归一化与残差连接，得到的特征表示为

$$M_2=M+\mathrm{LN}(M_1) \tag{8}$$

式中：LN$(\cdot)$表示层归一化操作。

将 M_2 输入到 Mix-FFN 中并使用残差连接，进行特征聚合与非线性变换，旨在更好地捕捉上下文特征之间的复杂关系，得到增强特征 $M_3\in R^{H\times W\times C}$，其计算过程为

$$M_3=\mathrm{LN}(\boldsymbol{W}_2\cdot\mathrm{SiLU}(\boldsymbol{W}_1\cdot M_2))+M_2 \tag{9}$$

式中：$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 为 Mix-FFN 中全连接层的权重矩阵；SiLU$(\cdot)$为激活函数。

对 M_3 进行线性投影和深度卷积处理，将其输入到 SS2D 中。SS2D 将形状为 $R^{H\times W\times C}$的特征图沿 4 个方向（左上到右下、右下到左上、右上到左下、左下到右上）展开，生成 4 个大小为 $R^{H\times W\times C}$的独立序列。这些序列通过 SSM 提取多方向的长距离依赖关系，并将 4 个方向的特征合并，从而生成完整的 2D 特征图 $\bar{M}\in R^{H\times W\times C}$。

2.3　注意力加权融合模块

有效融合不同特征对提升网络的表征能力至关重要，传统的静态融合策略，如特征相加或特征拼接操作，因缺乏两种特征之间的交互感知能力，导致无法充分利用局部特征与全局特征的互补优势，因此设计 AWF 模块。该模块通过动态平衡 HMSP 分支和 AF-Mamba 分支的贡献，进行交互互补融合，使全局特征提供整体语义信息，局部特征强化细节表达，从而滤除背景干扰，获得更加全面且精准的特征表示。AWF 模块结构示意图如图 4 所示。

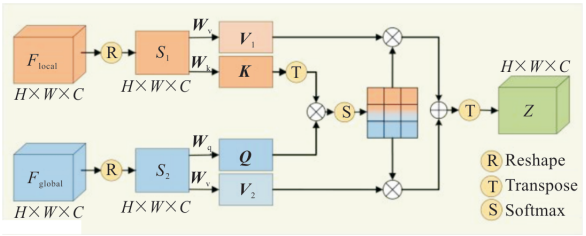


图 4　AWF 模块结构示意图

设 F_{local} 与 $F_{\text{glocal}}\in R^{H\times W\times C}$ 分别表示来自 HMSP 分支的局部细节特征和来自 AF-Mamba 分支的全局聚焦特征。首先，对 F_{local} 和 F_{glocal} 分别进行展平，得到相同维度的特征图谱 S_1 和 S_2，S_1，$S_2\in R^{H\times W\times C}$。然后对 S_1 和 S_2 进行线性投影，生成查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 矩阵，具体表达式为

$$\begin{cases}\boldsymbol{K}=S_1\boldsymbol{W}_{\boldsymbol{K}}\\ \boldsymbol{Q}=S_2\boldsymbol{W}_{\boldsymbol{Q}}\\ \boldsymbol{V}_1=S_1\boldsymbol{W}_{\boldsymbol{V}_1}\\ \boldsymbol{V}_2=S_2\boldsymbol{W}_{\boldsymbol{V}_2}\end{cases}\tag{10}$$

式中：$\boldsymbol{W}_{\boldsymbol{Q}}$、$\boldsymbol{W}_{\boldsymbol{K}}$、$\boldsymbol{W}_{\boldsymbol{V}_1}$ 和 $\boldsymbol{W}_{\boldsymbol{V}_2}$ 分别表示与 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 相对应的可学习的线性变换矩阵。

不同于自注意力机制，AWF 模块中的 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 分别来自 F_{glocal} 和 F_{local}，这使得网络能够在特征空间中引入显示的全局和局部信息的交互关系。对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行缩放点积运算，并使用 SoftMax 函数进行归一化，从而得到注意力权重矩阵为

$$\boldsymbol{A}_{\text{AWF}}=\text{Softmas}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d}}\right)\tag{11}$$

式中：d 为缩放因子，用于平衡点积的数值范围，确保梯度稳定性。

使用注意力权重 $\boldsymbol{A}_{\text{AWF}}$ 对 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 进行加权融合，实现 F_{local} 和 F_{glocal} 的动态交互与整合，得到融合特征为

$$Z=\boldsymbol{A}_{\text{AWF}}(\boldsymbol{V}_1+\boldsymbol{V}_2)\tag{12}$$

2.4　均衡损失函数

在自然环境中，难以获取害虫图像并进行标注，导致害虫数据集的类别分布极不均衡，传统交叉熵损失因梯度更新方向被多数类主导，导致分类决策边界严重偏向高频类别。因此，设计均衡损失函数 L_{EQ}。L_{EQ} 根据不同类别的样本数量差异，动态调整损失计算，从而对每个类别给予更加公平的关注，避免多数类主导训练过程。

具体来说，设 $D=\{(I_l,y_l)\}_{l=1}^{N}$ 为由 N 张图像组成的任意一个批次的训练样本集合，I_l 表示第 l 幅图像，y_l 表示该图像的类别标签。L_{EQ} 引入一个权重系数调整每个类别的损失，具体计算方式为

$$L_{\text{EQ}}(D)=-\sum_{l=1}^{N}\alpha_l\log(\bar{y}_l)y_l\tag{13}$$

其中，

$$\alpha_l=1-\beta T_\lambda(p_l)(1-y_l)。$$

式中：y_l 与 $\bar{y}_l$ 分别表示第 l 幅图像的真实标签与预测标签；α_j 表示 L_{EQ} 的权重系数，用来抑制多数类的梯度更新，使网络能够更关注少数类的学习；β 为随机变量，用于决定是否保持负样本的梯度，避免在少数类别训练过程中完全忽视负样本；p_i 表示类别 i 在数据集中的频率，由类别 i 的图像数量除以整个数据集的图像总数得到；$T_\lambda(\cdot)$ 表示阈值函数，用于判断 l 是否属于该数据集中的少数类别，当 $p_l<\lambda$ 时输出 1，即判断 l 为少数类别，在训练过程中增强该类的梯度更新，否则输出 0。

3　实验结果及分析

3.1　实验设置

在实验过程中，采用的软硬件平台配置如表 1 所示。所有输入图片的分辨率均被统一调整为 224×224，在训练与测试过程中，选择 Adam 优化器，初始学习率设置为 0.000 03，且采用余弦退火衰减策略进行更新，批处理大小设置为 16，训练迭代次数 Epochs 设置为 100。

表 1　实验环境

类别	名称	型号和参数
硬件	中央处理器	AMD EPYCTM 9754
	图像处理器	NVIDIARTX 4090D(24G)
	操作系统	Windows10
软件	框架	Pytorch 2.0-cuda 11.8
	编程语言	Python 3.8
	环境	Anaconda 4.12

3.2　数据集选取

选取 IP102[24] 和 D0[25] 两个公开且常用的害虫数据集评估 CFFDS-Net 的有效性。IP102 是一个

用于害虫识别和检测任务的大型农作物害虫数据集，涵盖 102 种常见的农作物害虫，共计 75 222 张图片，分辨率在 150～400 像素之间，按 6∶1∶3 的比例将数据集划分为训练集、验证集和测试集，其样本示例如图 5 所示。

(a) 虫卵 (b) 幼虫 (c) 成虫 (d) 蛹

图 5 IP102 数据集中的害虫样本示例

IP102 最大程度保留了真实农田场景的复杂性，具体表现为：1）种内差异突出，种间相似性高。图像覆盖了害虫的整个生命周期，包括卵、幼虫、蛹和成虫 4 个阶段，导致同一种害虫在不同阶段的外观特征差异显著同时，不同种类害虫在相同的生长阶段具有极大的相似性；2）数据集呈现严重的类别不平衡问题，其中最大类别包含 3 444 张图像，最小类别仅包含 42 张图像；3）图像中包含大量复杂的背景信息，如叶片、杂草等，导致网络难以捕捉关键的害虫特征。

D0 数据集涵盖了在野外捕获的 40 种害虫，总共 4 500 张害虫图像，分辨率均为 200×200，按照 7∶2∶1 的比例将 D0 数据集划分为训练、测试和验证 3 个子集，D0 数据集同样存在数据不平衡问题，其中最大类包含 238 张图片，最小类包含 50 张图片。

3.3 评价指标

选取准确率、精确率、召回率、F1 分数及模型参数量等 5 个指标作为模型的综合评价指标。准确率是指预测正确的样本数占总样本数的比例，反映模型的总体预测能力；精确率是指模型预测为正且实际为正的样本占模型预测为正类样本总数的比例；召回率是指模型预测为正的样本占实际为正的样本的比例；F1 分数则为精确率和召回率的加权调和平均值；模型参数量用来评估各模型的复杂度和计算开销，验证其在实际农业生产应用中的可行性。

3.4 综合对比实验

为全面评估 CFFDS-Net 的性能，选取经典网络和最新害虫识别网络进行对比分析，分别为基于 CNN 的网络：DMF-ResNet、Res-Net50[24]、Mobilent-V3[26]、EfficientNet-B7[27]、ConvNeXt[28] 和 FasterNet[29]；基于 Transformer 的网络：ViT、SwinViT[30]；基于集成学习方法的网络：GAEnsemble 和 SAEnsemble[31]。此外，Mamba 作为图像处理领域内新兴且具有影响力的模型，已在其他下游任务中得到广泛应用。为进一步验证 CFFDS-Net 网络的性能，将其与 Vmamba 和 ViM[32] 进行比较。最后，将 CFFDS-Net 与当前最新的双流网络 VRFNet 进行对比，以验证其在害虫识别任务中的先进性。不同网络在 IP102 数据集上的实验结果如表 2 所示。

表 2 不同网络在 IP102 数据集上的性能对比

网络	准确率/%	召回率/%	F1 分数/%	模型参数量/M
DMF-ResNet	59.22	—	58.37	29.70
ResNet-50	49.50	—	40.10	25.64
Mobilent-V3	63.20	58.70	60.90	4.30
EfficientNet-B7	65.70	60.90	65.20	64.70
ConvNeXt	68.40	67.20	67.80	27.90
FasterNet	66.10	—	—	30.00
ViT	65.50	57.70	61.40	49.30
SwinViT	70.20	69.70	69.90	87.70
GAEnsemble	67.13	67.13	65.76	—
SAEnsemble	66.30	—	—	—
Vmamba	66.41	65.30	65.11	30.00
ViM	62.42	56.68	58.62	26.00
VRFNet	68.34	68.33	68.34	—
CFFDS-Net	71.19	71.01	71.08	29.22

根据表 2 的数据可知，CFFDS-Net 在 IP102 数据集上展现了良好的识别效果，其准确率、召回率和 F1 分数分别达到了 71.19%、71.01%和 71.08%，达到最优值，且 CFFDS-Net 的参数量为 29.22 M，在性能与模型复杂度之间实现了较好的平衡，比较适合实际的模型部署。在基于 CNN 的网络中，与 IP102 数据集创建者在 ResNet-50 网络的测试结果相比，CFFDS-Net 的准确率提高了 21.69 个百分点，同时与其他 CNN 网络相比，CFFDS-Net 的准确率均可以提高 2.79～11.97 个百分点；在基于 Transformer 和 Mamba 的网络中，相比 SwinViT 和 Vmamba，CFFDS-Net 的准确率分别提高 0.99 个百分点和 4.78 个百分点，表明 CFFDS-Net 在捕捉远距离特征依赖上具有优势；最后，与最新的双流网络 VRFNet 相比，CFFDS-Net 表现出显著优势，表明 CFFDS-Net 能够更加有效地提取并融合互补特征。综上所述，在大型复杂害虫数据集 IP102 上的实验结果表明，CFFDS-Net 能够在包含多类害虫和干扰背景的场景下有效地识别和区分不同的害虫种类，展现出了较高的鲁棒性和良好的泛化性能。

为了更全面地评估 CFFDS-Net 的性能，使用 D0 数据集进行对比实验，结果如表 3 所示。由表 3 数据可知，现有模型在 D0 数据集上的识别精度均较高，但基于 Transformer 的网络在 D0 数据集上表现不佳，这是因为 Transformer 通常需要大量数据来充分捕捉和学习数据中的复杂特征和模式。相比之下，CFFDS-Net 在 D0 数据集上获得了 99.36%的准确率，优于所有模型，表明 CFFDS-Net 不仅在大型数据集上具备出色的识别能力，在较小的数据集上也展现了强大的泛化性能。

表 3　不同网络在 D0 数据集上的性能对比

网络	准确率/%	召回率/%	F1 分数/%	模型参数量/M
VRFNet	99.12	99.12	99.13	—
ResNet-50	97.20	94.80	96.00	23.70
MobilentV3	97.77	97.34	97.30	4.30
EfficientNet-B7	96.20	95.20	95.70	64.70
ConvNeXt	97.10	94.00	95.50	27.90
ViT	95.20	93.60	94.40	49.30
SwinViT	97.60	96.80	97.30	87.70
GAEnsemble	98.81	98.81	98.81	—
SAEnsemble	98.50	—	—	—
Vmamba	98.09	97.51	97.87	30.00
ViM	97.07	96.71	96.95	26.00
VRFNet	99.12	99.12	99.13	—
CFFDS-Net	99.36	99.28	99.27	29.22

3.5　消融实验

为了验证在 CFFDS-Net 中的关键模块（即 HMSP 模块、AF-Mamba 模块和 AWF 模块）以及均衡损失函数 L_{EQ} 在害虫识别网络中的有效性，基于 IP102 数据集进行消融实验。实验中，使用交叉熵损失函数作为对比基线，之后使用 L_{EQ} 代替交叉熵损失函数，以验证 L_{EQ} 的有效性。在进行双分支网络的模块消融实验时，为验证 AWF 模块的有效性，使用简单的特征相加操作代替 AWF 模块融合 HMSP 模块和 AF-Mamba 模块提取的特征。消融实验结果如表 4 所示。

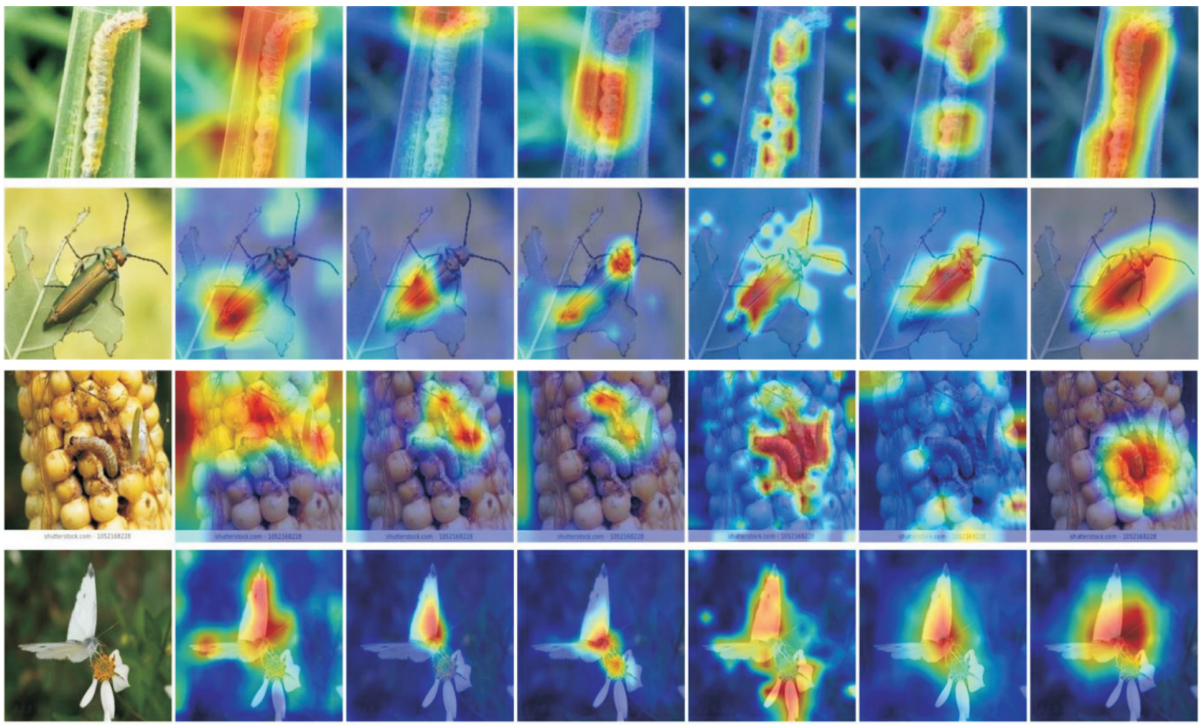
表 4　IP102 数据集上的消融实验结果

网络	HMSP	AF-Mamba	AWF	L_{EQ}	准确率/%	F1 分数/%
单分支	√	×	×	×	64.51	59.53
	×	√	×	×	66.94	65.13
	√	×	×	√	65.73	61.17
	×	√	×	√	67.45	65.76
双分支	√	√	×	×	68.57	67.28
	√	√	√	×	70.32	69.56
	√	√	×	√	69.43	69.17
CFFDS-Net	√	√	√	√	71.19	71.08

由表 4 可知，在单分支网络的消融实验中，与交叉熵损失函数相比，在使用 L_{EQ} 后，HMSP 模块的准确率和 F1 分数分别提高了 1.22 个百分点和 1.64 个百分点，AF-Mamba 模块的准确率和 F1 分数分别提高 0.51 个百分点和 0.63 个百分点，说明 L_{EQ} 能够有效缓解害虫图像的类别不平衡问题，有效地提升了网络对稀有类别的识别精度。在双分支网络的消融实验中，首先，“HMSP＋AF-Mamba＋特征相加”虽然不如完整的 CFFDS-Net 有效，但相比于单分支网络仍有一定的性能提升，这说明 HMSP 和 AF-Mamba 模块在特征提取和表示学习方面具有互补性。其次，相比于“HMSP＋AF-Mamba＋特征相加”，“HMSP＋AF-Mamba＋AWF”网络的准确率和 F1 分数分别提高了 1.75 个百分点和 2.28 个百分点，由此可见 AWF 模块能够充分利用局部特征和全局特征之间的互补优势，增强了网络对特征的判别能力。最后，“HMSP＋AF-Mamba＋AWF＋L_{EQ}”的准确率达到 71.19%，F1 分数达到 71.08%，完整网络 CFFDS-Net 取得了最优性能，说明各模块之间具有良好的互补性。

3.6　特征可视化分析

为了直观展现 CFFDS-Net 的有效性并直观展示经 AWF 模块融合所得互补特征图谱对害虫的感知能力，采用 Grad-CAM[33] 技术对 IP102 数据集中的部分害虫图像进行了可视化分析，如图 6 所示。由图 6 可以看出，与其他网络相比，CFFDS-Net 在害虫目标区域定位的精度和关注度方面表现出显著优势。具体而言，CFFDS-Net 能够精确聚焦目标区域并关注害虫细节，以第二行中的绿芫菁为例，Restnet-50、EfficientNet 等网络易陷入局部关注，但 CFFDS-Net 的类激活图能够完全覆盖害虫的关键部位，并且关注到害虫的触角、前肢等细节信息。其次，CFFDS-Net 在能够有效抑制背景干扰。例如在第三行含有复杂背景的玉米害虫图像中，CFFDS-Net 的激活区域紧密贴合害虫本体，而其他模型的激活区域则较为分散，部分包含了无关背景信息。此外，CFFDS-Net 在不同害虫类型上展现出高度一致的目标关注区域，这种一致性反映了模型在不同害虫类别和环境条件下的卓越泛化能力，确保了其在多样化农业场景中的准确及稳定识别。综上所述，CFFDS-Net 在定位精度、背景抑制和目标细节关注方面均表现出显著优势，验证了其在害虫识别任务中的优越性能。



(a) 原图像 (b) Restnet-50 (c) EfficientNet (d) Mobilent-V3 (e) ViT (f) SwinViT (g) CFFDS-Net

图 6 部分害虫在不同网络上的可视化结果

4 结语

针对真实农业场景中害虫识别任务面临的类间细节差异小、背景干扰严重及类别分布不平衡的挑战，设计了互补特征融合双流网络 CFFDS-Net 实现对害虫图像的准确识别。利用 HMSP 模块和 AF-Mamba 模块捕获局部细节特征和全局语义特征，通过 AWF 模块动态融合高相关性特征，增强网络的语义表征能力并抑制背景噪声干扰。构建均衡损失函数并自适应调整各类别的权重，增强网络对少数类别的关注，有效缓解了多数类主导训练过程的问题，从而提升了网络对类别分布不均衡数据集的分类性能。实验结果表明，CFFDS-Net 能够有效识别害虫图像，所设计的模块均对提高网络的识别精度和泛化能力起到了积极作用。未来工作将进一步探索更轻量化的网络设计，以实现在移动设备上的高效部署，为智能化害虫监测与精准农业管理提供强有力的技术支持。

参 考 文 献

[1] FU X Q, MA Q Y, YANG F F, et al. Crop pest image recognition based on the improved ViT method[J]. Information Processing in Agriculture, 2024, 11(2): 249-259.

[2] ODOUNFA M G F, GBEMAVO C D S J, TAHI S P G, et al. Deep learning methods for enhanced stress and pest management in market garden crops: A comprehensive analysis[J]. Smart Agricultural Technology, 2024, 9: 100521.

[3] VENKATASAICHANDRAKANTH P, IYAPPARAJA M. A survey on pest detection and classification in field crops using artificial intelligence techniques[J]. International Journal of Intelligent Robotics and Applications, 2024, 8(3): 709-734.

[4] 李睿，王莹，王恒. 基于树莓派的室内植物病虫害识别系统设计[J]. 电子设计工程，2022，30(8)：114-118.
LI R, WANG Y, WANG H. Design of house plants diseases and insect pests recognition system based on Raspberry Pi [J]. Electronic Design Engineering, 2022, 30(8): 114-118. (in Chinese)

[5] 胡海峰，倪宗煜，赵海涛，等. 无人机场景下基于 Transformer 的轻量化行人重识别[J]. 南京邮电大学学报(自然科学版)，2024，44(3)：48-62.
HU H F, NI Z Y, ZHAO H T, et al. Transformer based lightweight person re-identification in unmanned aerial vehicle images[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2024, 44(3): 48-62. (in Chinese)

[6] GUO B Y, WANG J J, GUO M H, et al. Overview of

pest detection and recognition algorithms[J]. Electronics, 2024, 13(15): 3008.

[7] 刘志，翟瑞芳，彭万伟，等. 融合注意力机制的 Cascade R-CNN 田间害虫检测方法[J]. 华中农业大学学报，2023，42(3)：133-142.
LIU Z, ZHAI R F, PENG W W, et al. Field pest detection method based on improved Cascade R-CNN by incorporating attention mechanism [J]. Journal of Huazhong Agricultural University, 2023, 42(3): 133-142. (in Chinese)

[8] LIU W J, WU G Q, REN F J. Deep multibranch fusion residual network for insect pest recognition[J]. IEEE Transactions on Cognitive and Developmental Systems, 2021, 13(3): 705-716.

[9] NANDHINI C, BRINDHA M. Visual regenerative fusion network for pest recognition[J]. Neural Computing and Applications, 2024, 36(6): 2867-2882.

[10] HECHEN Z, HUANG W, YIN L, et al. Dilated-windows-based vision transformer with efficient-suppressive-self-attention for insect pests classification [J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107228.

[11] FANG M W, TAN Z P, TANG Y, et al. Pest-conformer: A hybrid CNN-transformer architecture for large-scale multi-class crop pest recognition[J]. Expert Systems with Applications, 2024, 255: 124833.

[12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//9th International Conference on Learning Representations (ICLR). Austria: ICLR, 2021:15-35.

[13] PACAL I. Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model[J]. Expert Systems with Applications, 2024, 238: 122099.

[14] 李雨晴，陈燕红，李永可，等. 作物害虫图像智能识别方法[J]. 新疆农业科学，2023，60(12)：2973-2981.
LI Y Q, CHEN Y H, LI Y K, et al. Image intelligent recognition method of crop pests[J]. Xinjiang Agricultural Sciences, 2023, 60(12): 2973-2981. (in Chinese)

[15] XIA W S, HAN D Z, LI D, et al. An ensemble learning integration of multiple CNN with improved vision transformer models for pest classification[J]. Annals of Applied Biology, 2023, 182(2): 144-158.

[16] AYAN E, ERBAY H, VARCıN F. Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks[J]. Computers and Electronics in Agriculture, 2020, 179: 105809.

[17] LIU Y G, YU J Z, HAN Y H. Understanding the effective receptive field in semantic image segmentation [J]. Multimedia Tools and Applications, 2018, 77(17): 22159-22171.

[18] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces[EB/OL]. [2024-12-08]. https://doi.org/10.48550/arXiv.2111.00396.

[19] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces [EB/OL]. [2024-12-08]. https://doi.org/10.48550/arXiv.2312.00752.

[20] LIU Y, TIAN Y J, ZHAO Y Z, et al. VMamba: Visual state space model[EB/OL]. [2024-12-08]. https://doi.org/10.48550/arXiv.2401.10166.

[21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//31st International Conference on Neural Information Processing Systems (NIPS'17). Long Beach: Curran Associates, 2017: 6000-6010.

[22] XIONG Y W, LI Z Q, CHEN Y T, et al. Efficient deformable ConvNets: Rethinking dynamic and sparse operator for vision applications[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 5652-5661.

[23] CHEN B, ZOU X C, ZHANG Y, et al. LEFormer: A hybrid CNN-transformer architecture for accurate lake extraction from remote sensing imagery[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing. Seoul: IEEE, 2024: 5710-5714.

[24] WU X P, ZHAN C, LAI Y K, et al. IP102: A large-scale benchmark dataset for insect pest recognition [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 8779-8788.

[25] XIE C J, WANG R J, ZHANG J, et al. Multi-level learning features for automatic classification of field crop pests[J]. Computers and Electronics in Agriculture, 2018, 152: 233-241.

[26] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1314-1324.

[27] WANG C, ZHANG J R, HE J, et al. A two-stream network with complementary feature fusion for pest image classification[J]. Engineering Applications of Artificial Intelligence, 2023, 124: 106563.

[28] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966-11976.

[29] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 12021-12031.

[30] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 9992-10002.

[31] SU Z B, LUO J Q, WANG Y, et al. Comparative study of ensemble models of deep convolutional neural networks for crop pests classification[J]. Multimedia Tools and Applications, 2023, 82(19): 29567-29586.

[32] ZHU L H, LIAO B C, ZHANG Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model[EB/OL]. [2024-12-08]. https://doi.org/10.48550/arXiv.2401.09417.

[33] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

[作者简介]

李大湘(1974—),男,湖南麻阳人,博士,西安邮电大学副教授,主要研究方向为遥感图像分类、作物图像病虫害识别与机器学习。E-mail: www_ldx@163.com

孙家宁(2000—),女,陕西大荔人,西安邮电大学硕士研究生,主要研究方向为植物病虫害识别。E-mail: jianingsun@stu.xupt.edu.cn

刘颖(1972—),女,陕西户县人,博士,西安邮电大学教授,主要研究方向为图像处理与模式识别。E-mail: ly_yolanda@sina.com

[责任编辑:祝 剑]