彭莹琼,饶宇翔,廖牧鑫,等. 交叉注意力机制引导的无监督域自适应图像分类模型构建及其在细粒度实蝇识别中的应用[J]. 江苏农业学报,2026,42(4):756-762.
doi:10.3969/j.issn.1000-4440.2026.04.012

交叉注意力机制引导的无监督域自适应图像分类模型构建及其在细粒度实蝇识别中的应用

彭莹琼[1,2], 饶宇翔[1,2], 廖牧鑫[3], 钟文博[1]
(1.江西农业大学软件学院,江西 南昌 330045; 2.江西省高等学校农业信息技术重点实验室,江西 南昌 330000; 3.江西农业大学计算机与信息工程学院,江西 南昌 330045)

摘要: 现有害虫图像分类模型在跨域识别时常出现性能下降。为此,本研究提出一种基于焦点区域的交叉注意力机制引导的无监督域自适应图像分类模型 FC-DroNet。该模型首先在特征提取中引入掩膜处理,并利用级联交叉注意力模块融合水平、垂直及全局空间特征,以增强对细粒度局部特征的捕捉能力。同时,通过引入一致性约束机制抑制源域过拟合,从而提升跨域泛化性能。构建包含瓜实蝇(*Bactrocera cucurbitae*)、具条实蝇(*Bactrocera scutellata*)、南瓜实蝇(*Bactrocera tau*)和橘小实蝇[*Bactrocera dorsalis* (Hendel)]4 类图像的数据集 FD4Set,对 FC-DroNet 模型性能进行验证。结果表明,FC-DroNet 模型在测试集上的精确率达 99.41%,*F*1 达 99.22%,均高于 ResNet50 模型、AlexNet 模型、VGG-16 模型、LeNet-5 模型、ConvNext 模型、MobileVit 模型。本研究结果为田间害虫智能识别提供了技术支持。

关键词: 实蝇; 无监督域自适应; 注意力机制; 图像分类
中图分类号: Q969.456.8 **文献标识码:** A **文章编号:** 1000-4440(2026)04-0756-07

Construction of an unsupervised domain adaptive image classification model guided by cross-attention mechanism and its application in fine-grained fruit fly recognition

PENG Yingqiong[1,2], RAO Yuxiang[1,2], LIAO Muxin[3], ZHONG Wenbo[1]
(1.*School of Software*, *Jiangxi Agricultural University*, *Nanchang* 330045, *China*; 2.*Key Laboratory of Agricultural Information Technology of Colleges and Universities in Jiangxi Province*, *Nanchang* 330000, *China*; 3.*School of Computer Science and Engineering*, *Jiangxi Agricultural University*, *Nanchang* 330045, *China*)

Abstract: Existing pest image classification models often suffer from performance degradation in cross-domain recognition. To address this issue, this study proposed an unsupervised domain adaptive image classification model based on focal regions, named FC-DroNet. First, the model introduced mask processing in feature extraction and utilized a cascaded cross-attention module to fuse horizontal, vertical, and global spatial features, thereby enhancing the ability to capture fine-grained local features. Meanwhile, a consistency constraint mechanism was introduced to suppress overfitting on the source domain, thus improving cross-domain generalization performance. A dataset, FD4Set, containing four types of images, namely *Bactrocera cucurbitae*, *Bactrocera scutellata*, *Bactrocera tau*, and *Bactrocera dorsalis* (Hendel), was constructed to verify the performance of the FC-DroNet model. The results showed that the precision of the FC-DroNet model on the test set reached 99.41% and the *F*1 reached 99.22%, both higher than those of the ResNet50

收稿日期:2025-10-15
基金项目:国家自然科学基金项目(62262028);江西省自然科学基金面上项目(20242BAB25082)
作者简介:彭莹琼(1979-),女,江西萍乡人,硕士,教授,硕士生导师,主要从事农业信息化、图像处理研究。(E-mail)jneyq@jxau.edu.cn
通讯作者:钟文博,(E-mail)jneyq_pyq@jxau.edu.cn

model, AlexNet model, VGG-16 model, LeNet-5 model, ConvNext model, and MobileVit model. The findings of this study provide technical support for the intelligent identification of field pests.

Key words: *Bactrocera*; unsupervised domain adaptation; attention mechanism; image classification

图像分类是计算机视觉的基础任务,旨在通过分析图像内容将其准确归类。其流程包括图像预处理、特征提取与分类器设计。预处理通过尺寸调整、去除噪声等操作提升数据质量[1];特征提取将图像信息转化为更具代表性的低维特征(如颜色、纹理)[2];分类器则学习特征与类别标签之间的映射关系[3]。根据分类粒度,任务可分为粗粒度和细粒度图像分类。细粒度分类侧重于区分外观相似的类别(如不同种类的害虫),这对模型的局部辨别能力提出了更高要求[4]。害虫图像通常背景复杂、类间差异细微,需要通过局部特征(如触角、翅膀纹理等)进行分析。现有方法常通过注意力机制、对比学习等方式增强模型对细节的捕捉能力。

自注意力机制在自然语言处理中取得显著成效[5],随后被引入视觉任务,基于自注意力机制的Transformer在图像分类[6]和目标检测[7]中得到广泛应用。在图像分类中,自注意力机制通过将图像划分为块并展平为序列建模,有效捕获二维图像特征[8-9]。在语义分割等需要全局上下文的视觉任务中,PSPNet通过多尺度金字塔池化融合层级特征[10-11],MDC-SpecNet模型采用扩张卷积扩大感受野以捕获上下文信息[12]。自注意力机制通过预测注意力图实现全局加权聚合,如非局部网络[13]可实现全图信息感知,但计算量与存储量较大。为此,Huang等[14]提出使用连续稀疏注意力替代单层密集注意力,在降低模型复杂度的同时保持全局依赖建模能力。然而,上述方法多假设数据独立分布,难以应对跨域场景下的域偏移问题。加之实蝇图像存在背景复杂、边缘模糊等特点,传统注意力机制在跨域应用时分类精度受限。为解决上述问题,本研究拟提出一种基于焦点区域的无监督域自适应实蝇图像分类模型,该模型以ResNet-50为骨干,结合U-Net提取实蝇目标掩膜并与全局特征融合,通过交叉注意力机制建模像素级全局依赖,同时引入跨域一致性约束增强特征泛化能力,以期为农业虫害监测提供技术支持。

1 材料与方法

1.1 数据集构建

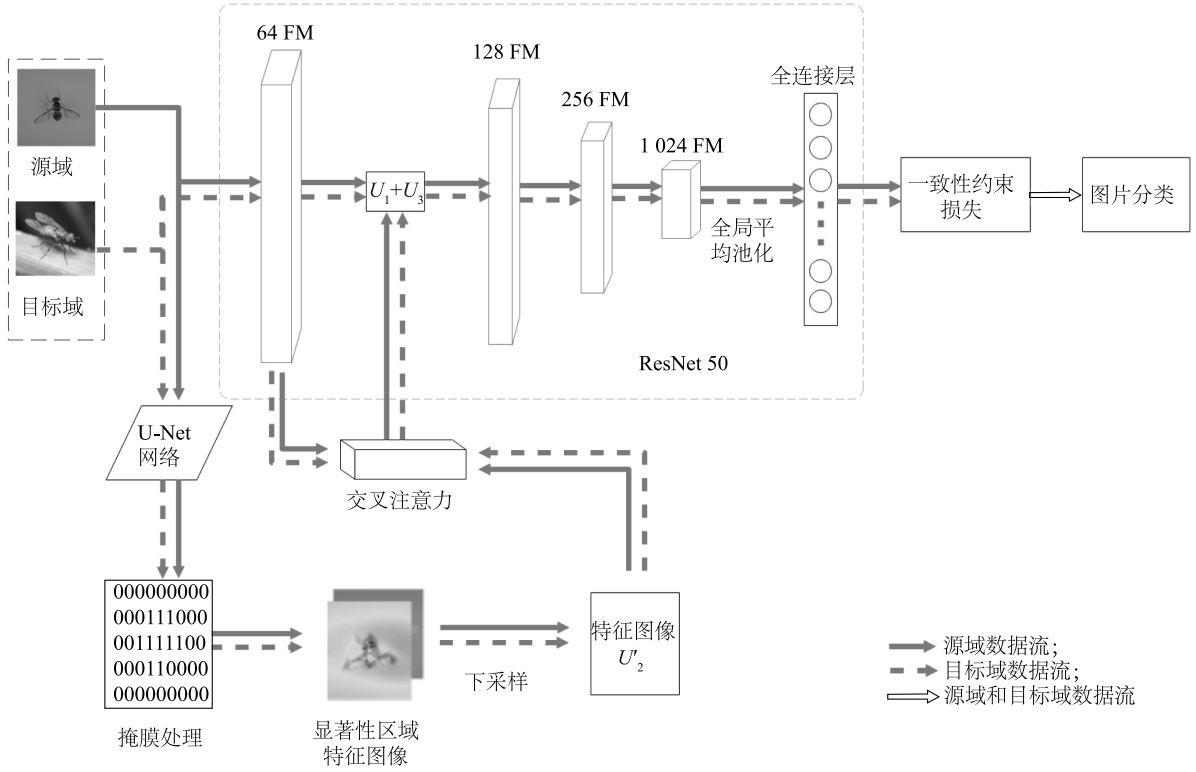
本研究构建了一个面向果蝇害虫的专用数据集——FD4Set。该数据集图像于2024年7月至8月在中国江西省拍摄。FD4Set数据集包含瓜实蝇(*Bactrocera cucurbitae*)、具条实蝇(*Bactrocera scutellata*)、南瓜实蝇(*Bactrocera tau*)和橘小实蝇[*Bactrocera dorsalis* (Hendel))4类图像。数据集共包含3 600张图像,被划分为3个子集:源域数据集(2 000张背景简单的果蝇图像)、训练目标域数据集(1 200张以叶片为背景的果蝇图像)以及测试目标域数据集(400张经过旋转与亮度调整背景的果蝇图像)。各子集均包含全部4种果蝇物种。为增强数据多样性并降低偏差,拍摄过程中涵盖不同角度与多种背景。数据集中所有图像均经过人工筛选,剔除模糊及低质量的图像,随后统一调整为1 024×1 024像素并进行归一化处理,以适配网络输入要求。物种标注由昆虫学专家审核,确保标签准确性。

此外,为验证模型的鲁棒性与泛化能力,本研究还构建了一套OPset数据集,该数据集基于公开的果园害虫数据集,并进行了调整。数据集中包含4类果园害虫图像:星天牛、日本金龟子、绿叶蝉、梨小食心虫。在源域数据集中,半数图像的背景被统一涂黑以简化学习特征,其余图像则通过旋转、添加噪声及调整亮度等方式进行增强处理,以提升模型泛化能力。经过上述处理的图像被划分为两部分,分别作为训练目标域数据集与测试目标域数据集。

1.2 FC-DroNet模型构建

本研究提出一种面向焦点区域的无监督域自适应实蝇图像分类网络(FC-DroNet)模型。该网络以ResNet50为基础架构,整体结构如图1所示。首先,利用预训练的U-Net模型对输入图像进行掩膜处理,将掩膜图像与原始图像相乘,从而提取仅包含实蝇的目标区域。随后,网络引入交叉注意力机制,以促进局部特征的传递,从而获取实蝇图像在水平与垂直方向上的信息。该机制将初始特征图像传递至后续模块,进一步提取互补性上下文表达,最终在像素层面建立全局依赖关系。经注意力加权后的特征图像与原始特征图像相乘,生成新的特征图像,使其更关注图像中复杂背景、细微边缘结构及低对比度特征。此外,通过对新特征图像施加一致性约束,限

制模型对源域数据的过拟合,从而提升在目标域上的表现,增强模型泛化能力。



64FM、128 FM、256 FM、1024 FM:对应通道数的特征图;U_1:ResNet50 提取的全局特征图像;U'_2:对掩膜图像进行下采样与最大池化得到的区域掩膜特征;U_3:经交叉注意力机制输出的特征图像。

图 1 FC-DroNet 模型结构

Fig.1 Architecture of FC-DroNet model

1.2.1 可重复使用的交叉注意力模块 引入可重复使用的交叉注意力模块,用于以轻量级计算在局部特征中建立全局图像的依赖关系[14]。如图 2 所示,交叉注意力模块使每个像素都能与图像全局信息建立依赖关系。对像素特征的分析聚焦于空间维度的上下文信息传递,而不涉及通道维度间的相互差异,因此仅使用 1×1 卷积核的卷积层,其作用可等效为对不同空间位置的像素特征进行直接的线性组合与连接操作。特征图像 H 上的一个目标像素点 n [坐标(x,y)]获取特征图像H'上另一个源像素点 m [坐标(x',y')]的信息,这一过程可通过两步计算实现。首先定义从源位置(x',y')到目标位置权重 $A_{i,x,y}$的映射函数:$A_{i,x,y}=f(A,x,y,x',y')$,其中 A 为全局关联权重张量。当位置 U 与 u 位于不同行和不同列时,信息传递过程可表示为:

$$H'_u \leftarrow [f(A,m_x,n_y,n_x,n_y)\cdot f(A',m_x,m_y,m_x,n_y)+f(A,n_x,m_y,n_x,n_y)\cdot f(A',m_x,m_y,n_x,m_y)]\cdot H_\theta \quad (1)$$

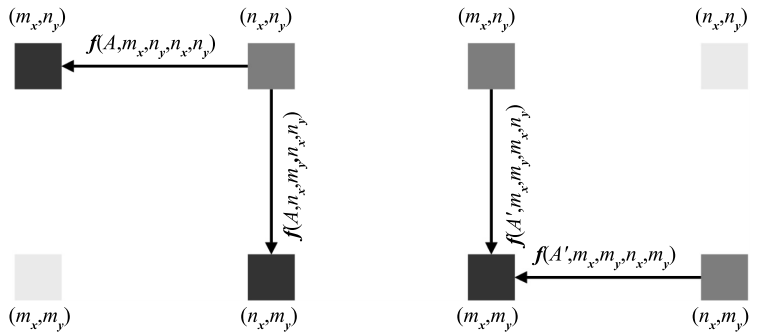
1.2.2 掩膜处理与特征融合的改进 由于实蝇个体较小,因此在注意力机制前增加掩膜预处理,以增强模型对局部细节的捕捉能力。具体而言,首先通过卷积网络提取初始特征图 $U_1 \in R^{H\times W\times C}$。为突出显著区域,使用 U-Net 生成掩膜图 U_2。进一步将掩膜信息与特征图像融合后输入注意力模块,并利用原始特征对注意力输出进行补强,以保持信息完整性。U_2计算公式如下:

$$U_2=f_{\text{U-Net}}(X) \quad (2)$$

式中,$f_{\text{U-Net}}$表示 U-Net 网络的前向传播过程,输出的U_2保持与输入图像具有相同的空间尺寸 $H\times W$,即 $U_2 \in R^{H\times W}$。

由于特征提取网络 ResNet50 通常在较低分辨率的特征图像上进行计算,因此对掩膜图像进行下采样处理,通过最大池化操作得到压缩后的信息图像 $U'_2 \in R^{H'\times W'}$($H'<H$,$W'<W$),计算公式如下:

$$U'_{2(i',j')}=\text{Max}_{(i,j)\in R(i',j')}U_{2(i,j)} \quad (3)$$



m、n:像素点;x、y、x'、y':坐标;$f(A,m_x,n_y,n_x,n_y)$、$f(A,n_x,m_y,n_x,n_y)$、$f(A',m_x,m_y,m_x,n_y)$、$f(A',m_x,m_y,n_x,m_y)$分别为信息传递的过程。

图 2 建模全局图像依赖的图像编辑方法

Fig.2 Image editing method for modeling global image dependencies

式中,$U_{2(i,j)}$为原始掩膜图像U_2在空间位置(i,j)处的像素值;Max 为最大池化操作。

将两个网络提取的特征图像进行融合,并通过交叉注意力机制进一步增强特征表示。在交叉注意力模块中,查询向量。$(Q)=\boldsymbol{W}_{\boldsymbol{Q}}U_1$,键向量$(K)=\boldsymbol{W}_{\boldsymbol{K}}U_1$,值向量$(V)=\boldsymbol{W}_{\boldsymbol{V}}U_1$。$\boldsymbol{W}_{\boldsymbol{Q}}$、$\boldsymbol{W}_{\boldsymbol{K}}$、$\boldsymbol{W}_{\boldsymbol{V}}$为可学习的卷积权重矩阵。自注意力机制的计算过程如下:

$$A=\mathrm{softmax}\left(\frac{Q\cdot K\cdot V}{\sqrt{d_k}}\cdot U'_2\right) \tag{4}$$

softmax 为归一化函数,U'_2为对掩膜图像进行下采样与最大池化得到的区域掩膜特征,使得模型能够聚焦掩膜所标注的显著区域,从而增强注意力机制的目标导向性。

注意力加权输出采用残差连接方式,使得最终得到的注意力特征图像U_3同时包含交叉注意力信息与原始特征信息,其计算公式如下:

$$U_3=A+U_1 \tag{5}$$

式中,A 表示交叉注意力信息,U_1表示原始特征信息。

尽管交叉注意力机制能够强化特定区域的特征表达,但也可能过度抑制部分非显著区域的信息。因此,将注意力模块输出的U_3与未经注意力加权处理的原始特征信息U_1结合,以保留完整的全局特征,避免重要信息损失。最终融合为新特征图像 U:

$$U=\lambda U_1+(1-\lambda)U_3 \tag{6}$$

式中,λ 是超参数,用于调节原始特征与注意力增强特征在融合时的贡献比例。

1.3 评价指标

采用精确率和 $F1$ 对模型性能进行了全面评估。精确率(*Accurancy*)和 $F1$ 计算公式如下:

$$Accurancy=\frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Precision=\frac{TP}{TP+FP} \tag{8}$$

$$Recall=\frac{TP}{TP+FN} \tag{9}$$

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall}=\frac{2TP}{2TP+FP+FN} \tag{10}$$

式中,TP 表示模型正确预测为正类的样本数,FP 表示模型错误预测为正类的样本数,FN 表示模型错误预测为负类的样本数,TN 表示模型正确预测为负类的样本数。

2 结果与分析

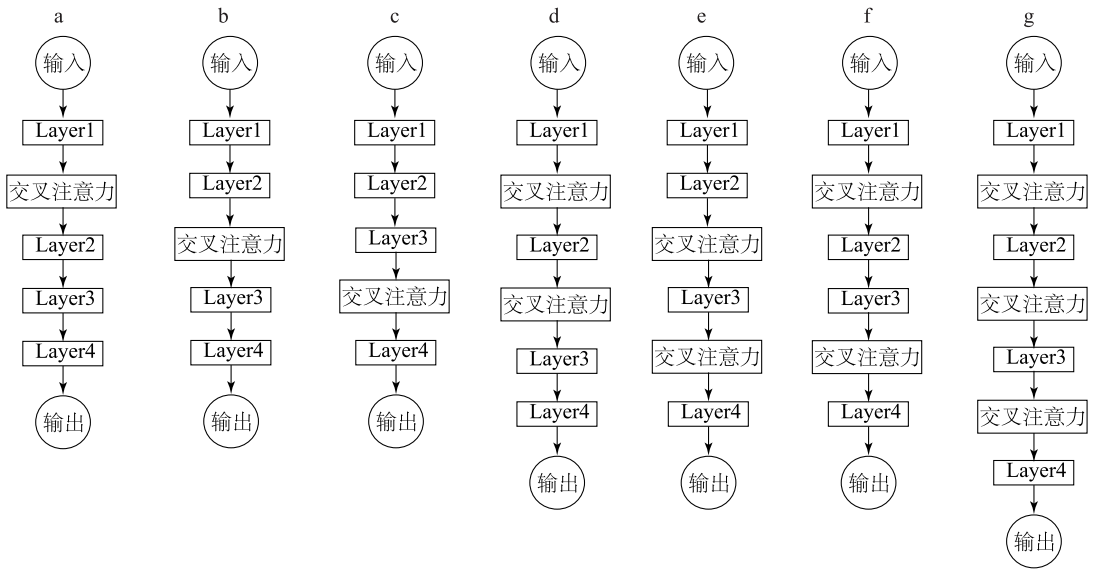
2.1 注意力机制插入位置对模型性能的影响

如图 3 所示,将交叉注意力机制插入 ResNet50 网络不同的残差块后。如表 1 所示,在 FD4set 数据集上,将交叉注意力机制插入 layer1 残差块后。模型表现出更优的性能。这主要是因为交叉注意力机制在高分辨率特征图(如 layer1 残差块获得的特征图)中能够充分利用丰富的空间细节,在大范围内捕获细粒度信息,有利于细粒度目标的定位;而在低分辨率特征图(如 layer3 残差块获得的特征图)中,由于空间信息有限,交叉信息机制难以有效挖掘细粒度目标的判别性特征,因此不适合作为细粒度识别的主要空间特征来源[15-16]。

2.2 一致性约束方法对模型性能的影响

如表 2 所示,当参数为 1.0 ,L1 正则化约束的模

型取得最佳效果。这是因为该参数下更多不重要的特征被压缩至零,降低了噪声干扰,增强了模型的泛化能力,进而在测试集上表现更优。因此将参数固定为 1.0,对比 L1 正则化[17]、余弦距离[18]和 Jensen-Shannon 散度[19]3 种一致性约束方法。如表 3 所示,L1 正则化约束的模型表现最佳,这主要得益于其更强的特征稀疏化能力,能够更有效地去除冗余特征。



Layer1、Layer2、Layer3、Layer4:ResNet50 中 4 个位置不同的残差块。

图 3 注意力机制插入位置

Fig.3 Insertion position of the attention mechanism

表 1 注意力机制插入不同位置的模型的性能

Table 1 The performance of models with attention mechanism inserted at different positions

插入位置	精确率(%)	*F*1(%)
Layer1 后	91.71	91.67
Layer2 后	72.79	72.58
Layer3 后	71.77	71.53
Layer1 后+Layer2 后	34.91	33.72
Layer2 后+Layer3 后	35.93	34.34
Layer1 后+Layer3 后	39.50	38.18
Layer1 后+Layer2 后、Layer3 后	29.97	29.17

Layer1、Layer2、Layer3 为 ResNet50 中不同位置的残差块。

表 2 不同参数设置的 L1 正则化约束的模型的性能

Table 2 The performance of models constrained by L1 regularization with different parameter settings

参数量	轮次	学习率	精确率(%)	*F*1(%)
1.000	20	0.001	99.17	99.08
0.100	20	0.001	99.41	98.69
0.010	20	0.001	98.47	97.57
0.001	20	0.001	98.23	95.88

表 3 不同一致性约束方法约束的模型的性能

Table 3 The performance of models constrained by different consistency constraint methods

一致性约束方法	轮次	学习率	精确率(%)	*F*1(%)
L1 正则化	20	0.001	99.17	99.08
余弦距离	20	0.001	99.06	98.79
Jensen-Shannon 散度	20	0.001	99.41	98.90

2.3 不同模型性能对比

将交叉注意力机制插入 ResNet50 网络的 layer1 残差块后,并采用参数为 1.0 的 L1 正则化进行一致性约束,得到 FC-DroNet 模型。如图 4 所示,在 FD4set 数据集上,FC-DroNet 模型的损失函数值与准确率均表现出稳定且良好的收敛趋势。

在 FD4set 数据集上,将 FC-DroNet 模型与 ResNet50 模型[20]、AlexNet 模型[21]、VGG-16 模型[22]、LeNet-5 模型[23]以及基于注意力机制的 ConvNeXt 模型[24]、MobileVit 模型[25]、EfficientNetV2 模型[26]进行对比。如表 4 所示,FC-DroNet 模型精确率为 99.41%,*F*1 分数为 99.22%,优于其他模型。

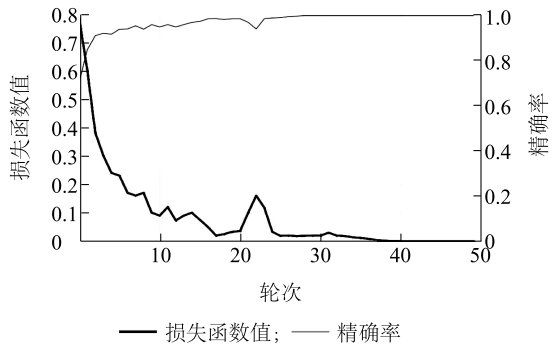


图 4 本研究构建的模型的损失函数值和精确率曲线

Fig. 4 Loss function and precision curves of the model constructed in this study

表 4 FD4set 数据集上不同模型的性能

Table 4 The performance of different models on the FD4set dataset

模型	精确率(%)	*F*1(%)
ResNet50	76.77	75.93
AlexNet	78.23	77.78
VGG-16	58.92	60.38
LeNet-5	69.52	72.34
ConvNext	72.13	71.64
MobileVit	99.19	99.19
FC-DroNet(本研究构建的模型)	99.41	99.22

如表 5 所示,在果园虫害数据集上,FC-DroNet 模型相精确率和 *F*1 同样优于其他模型。

表 5 果园虫害数据集上不同模型的性能

Table 5 The performance of different models on the orchard pest dataset

模型	精确率(%)	*F*1(%)
ResNet50	84.66	70.76
AlexNet	85.80	85.17
VGG-16	54.46	50.82
LeNet-5	79.07	74.91
ConvNext	71.39	54.87
MobileVit	98.35	96.17
FC-DroNet(本研究构建的模型)	99.53	99.35

利用 Grad-CAM[27] 生成最后一个特征提取层的输出图像。如图 5 所示,与原始 ResNet50 模型对比,FC-DroNet 模型能更准确地聚焦于关键区域,更关注细节特征。

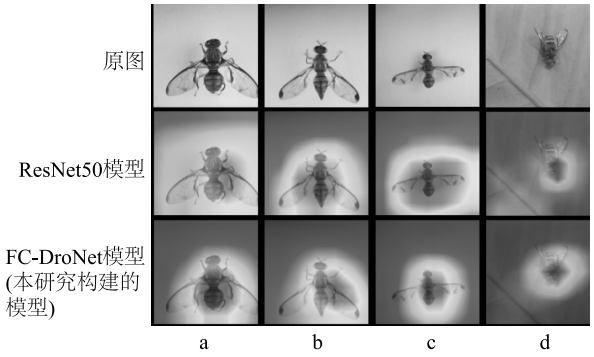


图 5 FC-DroNet 模型和 ResNet50 模型性能对比

Fig.5 Performance comparison between the FC-DroNet model and the ResNet50 model

2.4 不同模块对 FC-DroNet 性能的影响

利用消融试验对交叉注意力机制与基于 U-Net 的下采样模块的作用进行了独立评估。如表 6 所示,引入交叉注意力机制后,模型精确率和 *F*1 提升,表明该机制能更有效地识别小尺寸目标;引入 U-s 模块后,模型精确率和 *F*1 提升。同时引入交叉注意力机制和 U-s 模块,模型整体性能提升显著。

表 6 不同模块对 FC-DroNet 模型性能的影响

Table 6 The impact of different modules on the performance of FC-DroNet model

模型	精确率(%)	*F*1(%)
ResNet50	76.77	75.93
ResNet50+CAA	91.71	91.67
ResNet50+U-s	80.12	82.69
ResNet50+CAA+U-s	99.66	99.63

CAA:交叉注意力机制;U-s:基于 U-Net 的下采样模块。

3 结 论

本研究提出了一种基于焦点区域的无监督域自适应实蝇图像分类网络(FC-DroNet)模型。该模型首先通过预分类模块对图像进行掩膜与下采样处理,生成信息图像并与卷积特征图像融合,接着引入交叉注意力机制,引导局部特征传递,捕获图像在水平与垂直方向上的信息。此外,通过对特征图像施加一致性约束,抑制模型对源域数据的过拟合,增强其在目标域上的泛化能力。为验证模型效果,构建了包含瓜实蝇、具条实蝇、南瓜实蝇和橘小实蝇 4 类

实蝇图像的 FD4Set 数据集。试验结果表明,FC-DroNet 模型在测试集上精确率达 99.41%,*F*1 达 99.22%,均优于 ResNet50 模型、AlexNet 模型、VGG-16 模型、LeNet-5 模型、ConvNext 模型、MobileViT 模型。本研究结果为田间害虫智能识别提供了理论依据和技术支持。

参考文献:

[1] 胡 婷,孙晓海,宋海龙,等. 基于层次标注和自适应预处理的多源农业病害图像数据集构建[J]. 吉林大学学报(理学版),2025,63(3):815-821.

[2] 魏超宇,韩 文,庞 程,等. 基于多尺度特征融合和密集连接网络的疏果期黄花梨植株图像分割[J]. 江苏农业学报,2021,37(4):990-997.

[3] 孙 进,张 洋,王 宁,等. 融合机器视觉和 CAN 总线的玉米种粒分类器设计与试验[J]. 中国农机化学报,2020,41(8):81-89,120.

[4] 郝月华,吕卫东,张幽迪,等. 基于多分类自适应聚焦损失与 B-CNN 的棉田昆虫细粒度图像分类研究[J]. 现代电子技术,2025,48(5):43-48.

[5] ZHOU M, DUAN N, LIU S J, et al. Progress in neural NLP: modeling, learning, and reasoning[J]. Engineering, 2020, 6(3): 275-290.

[6] 蒋东山,刘金洋,张浩淼,等. 基于 CNN 和 Transformer 的绿豆干旱胁迫识别模型[J]. 江苏农业学报,2025,41(1):87-100.

[7] 唐秀英,孙中清,杨琳琳,等. 基于改进 YOLO v8n 轻量化的番茄叶霉病发病程度分级检测[J]. 江苏农业学报,2025,41(10):1985-1996.

[8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//ECVA. European Conference on Computer Vision. Cham: Springer International Publishing, 2020.

[9] 温世雄,智 敏. 视觉 Transformer 在细粒度图像分类中的应用综述[J]. 计算机工程与应用,2025,61(23):24-37.

[10] ZHAO H H, SHI J P, QI X J, et al. Pyramid scene parsing network [C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2017.

[11] ZHANG H, DANA K, SHI J P, et al. Context encoding for semantic segmentation[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2018.

[12] 李 强,陈 蓓,张 芳. 基于多尺度扩张卷积神经网络的近红外光谱定量分析模型研究[J]. 分析化学,2025,53(3):451-463.

[13] CHENG J P, DONG L, LAPATA M. Long shortterm memory-networks for machine reading[EB/OL]. (2016-09-20)[2025-10-15]. https://arxiv.org/pdf/1601.06733.

[14] HUANG Z, WANG X, HUANG L, et al. Ccnet: criss-cross attention for semantic segmentation[C]//IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE Computer Society, 2019.

[15] YING X, ZHANG Y L, WEI X, et al. MSDAN: multi-scale self-attention unsupervised domain adaptation network for thyroid ultrasound images[C]//IEEE. 2020 IEEE International Conference on Bioinformatics and Biomedicine. Piscataway, NJ: IEEE, 2020.

[16] ZHANG Z L, ZHANG X Y, PENG C, et al. Exfuse: enhancing feature fusion for semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham, Switzerland: Springer, 2018.

[17] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996, 58(1): 267-288.

[18] CHOWDHURY G G. Introduction to modern information retrieval [M]. London: Facet Publishing, 2010.

[19] LIN J. Divergence measures based on the Shannon entropy[J]. IEEE Transactions on Information Theory, 1991, 37(1): 145-151.

[20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2016.

[21] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2025-10-08]. https://doi.org/10.48550/arXiv.1409.1556LE.

[23] CUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 2002, 86(11): 2278-2324.

[24] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2022.

[25] MEHTA S, RASTEGARI M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer[J]. (2021-10-05)[2025-10-15]. https://doi.org/10.48550/arXiv.2110.02178.

[26] TAN M X, LE Q V. Efficientnetv2: smaller models and faster training[EB/OL]. (2021-04-01)[2025-10-15]. https://doi.org/10.48550/arXiv.2104.00298.

[27] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization [C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Computer Society, 2017.

(责任编辑:成纾寒)