

基于 UNet+Swin-Transformer 的西瓜叶片病害识别的研究

向宇杰, 向元平

(湖南农业大学信息与智能科学技术学院, 湖南 长沙 410128)

✉ 2014176995@qq.com; 13619652@qq.com



摘要: 诊断与识别植物叶片的病虫害是农业生产中的一大难题。为了解决西瓜叶片病虫害的诊断与识别问题,方便瓜农评估西瓜叶片的健康状况,提出了一种先分割、后识别的西瓜叶片病害识别算法。该算法首先采用 UNet 模型对叶片进行分割,其次使用 Swin-Transformer 模型进行病虫害识别。通过在自建的西瓜叶片数据集上进行对比实验,文章所提算法识别准确率达到 92.9%,相比直接在原始图像上使用 Swin-Transformer 模型进行病虫害识别,准确率提高了 3.2%。实验结果表明,使用分割后的图像进行病虫害分类可以显著提高识别准确率。

关键词: UNet; Swin-Transformer; 语义分割; 病虫害识别

中图分类号: TP391 **文献标志码:** A

Research on Watermelon Leaf Disease Identification Based on UNet+Swin-Transformer

XIANG Yujie, XIANG Yuanping

(College of Information and Intelligence, Hunan Agricultural University, Changsha 410128, China)

✉ 2014176995@qq.com; 13619652@qq.com

Abstract: Diagnosis and identification of plant leaf diseases and pest are major challenges in agricultural production. In order to solve the problems of diagnosis and identification of watermelon leaf disease and pest, and to facilitate the evaluation of the health status of watermelon leaves by farmers, this paper proposes a watermelon leaf disease identification algorithm with segmentation first and recognition followed. Firstly, UNet model is used for leaf segmentation, and then Swin-Transformer model is used for pests and diseases identification. Through comparative experiments on a self-built watermelon leaf dataset, the proposed algorithm achieves a recognition accuracy of 92.9%, which is 3.2% higher than the one that uses Swin Transformer model directly on the original image for pest and disease identification. The experimental results show that using segmented images for pest and disease classification can significantly improve recognition accuracy.

Key words: UNet; Swin-Transformer; semantic segmentation; identification of pests and diseases

0 引言 (Introduction)

西瓜作为一种重要的经济作物,其叶片病虫害的防治一直是困扰瓜农的难题,及时发现并有效防治病虫害,有助于提高瓜农收入,降低损失。传统的人工观察和经验判断方法识别准确率和效率低,此外一部分瓜农受限于自身的学识和经验不足而无法辨认所有病害种类,仅凭数量有限的专家无法满足瓜农诊断病虫害的需求。因此,利用人工智能技术训练一个高识别率的神经网络模型,实现自动、准确和实时地识别西瓜叶片病虫害,对于提高农业生产效率和减轻瓜农负担具有重要意义。

目前,农业病虫害识别在国内已经成为一个被广泛研究的课题。已有的研究有深度迁移学习在花生叶部病害图像识别

中的应用、基于 Fisher 判别分析的玉米叶部病害图像识别、基于改进 VGG 卷积神经网络的棉花病害识别等,但少有将 Transformer 应用于西瓜叶片病虫害识别的研究^[1-3]。为了实现自动化识别西瓜叶片病虫害,本文提出了一种先使用 UNet 模型分割再使用 Swin-Transformer 模型识别的西瓜叶片病害识别算法,可以自动提取病虫害图像的数字特征,具有较高的识别准确度和良好的对新鲜样本的适应能力。

1 叶片分割 (Leaf segmentation)

UNet 完美的模型结构及其具备的各种先进技术,使其能够在图像分割任务中取得非常好的性能,成为自然图像分割、医学影像分割等领域的重要工具。

1.1 模型结构

UNet^[4]主要由编码器(Encoder)和解码器(Decoder)两个部分组成。在 Encoder 部分,采用由两个卷积核为 3×3 的卷积层和一个 2×2 的最大池化层组成的块提取图像的高级特征,并将输入图像中的信息压缩为一个较小的特征向量。在 Decoder 部分,使用由一个上采样层和两个卷积核为 3×3 的卷积层组成的块进行特征映射的上采样,将 Encoder 从图像中提取的特征向量还原为与原始图像大小相同的分割 Mask 图像。Encoder 和 Decoder 之间相邻的层次通过跳跃连接拼接在一起,使解码器可以利用编码器中较低级别的特征矩阵进行更加精细的分割,让解码器在训练时不局限于其中的高级特征。跳跃连接实现了特征图的“对齐”,从而在解码器中更好地使用编码器中的特征图进行分割,有助于准确地还原原始图像的分辨率并保留分割任务所需的细节信息。在模型的最后一层,使用一个卷积核为 1×1 的卷积层将特征向量映射出最终的分割 Mask 图像结果。UNet 模型结构示意图如图 1 所示。

此外,UNet 使用的 Batch Normalization 技术,可以提高模型的训练速度和性能。批归一化可以减少模型的内部协方差移动,使模型更加稳定,并且可以通过减少对权重的依赖提高模型的泛化能力。

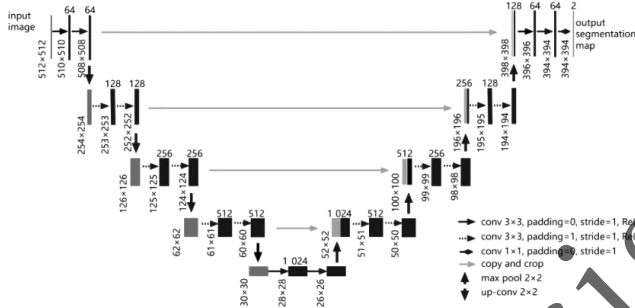


图 1 UNet 模型结构示意图

Fig. 1 UNet model structure diagram

1.2 模型训练

步骤 1:数据预处理。采用随机增加噪声和颜色变化等强度变换方法,以增强叶片图像。

步骤 2:设置损失函数。针对语义分割模型的训练,使用的 Loss 函数为 BCEWithLogitsLoss,实质上就是 Sigmoid 和 BCELoss 的结合。

$$I(x, y) = \frac{1}{N} \sum_{n=1}^N [y_n \cdot \log(\sigma(x_n)) + (1 - y_n) \cdot \log(1 - \sigma(x_n))] \quad (1)$$

其中: x 为模型的原始值, y 为对应的标签值, $\sigma(x)$ 为 Sigmoid 函数, $\log$ 为自然对数, N 为样本数。

步骤 3:设置评估指标。对于语义分割模型的训练,评估训练结果的指标为 PA (Pixel Accuracy) 和 $MIoU$ (平均交并比)。

$$PA = \frac{\sum_{i=1}^N I(y_i = \hat{y}_i)}{N} \quad (2)$$

其中: N 是图片中像素点的总个数, y_i 是第 i 个像素点的真实值, $\hat{y}_i$ 是第 i 个像素点的预测值, $I(\cdot)$ 是指示函数,当条件为真时, $I(\cdot)$ 等于 1,否则等于 0。 PA 为预测结果是真的 pixel

个数与总 pixel 个数的比值。

$$MIoU = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FP_i + FN_i} \quad (3)$$

其中: TP_i 是第 i 类的真阳性像素数, FP_i 是第 i 类的假阳性像素数, FN_i 是第 i 类的假阴性像素数。 $MIoU$ 为数据集的每一个类的交并比的均值。

步骤 4:设置其他参数训练模型。实验中设置 $batch_size$ 和 $num_workers$ 为 4,学习率($learning_rate$)为 0.01,迭代轮数($epoch$)为 100, num_class 为 2。每迭代一轮后,对测试集进行一次评估,并统计模型的效果。

2 叶片病虫害识别(Identification of leaf diseases and pests)

Swin-Transformer^[5]是一种基于 Transformer 的神经网络模型,专门针对大规模视觉数据集的处理而设计,可以用于图像分类等各种识别任务。与传统的 Transformer^[6]模型不同,Swin-Transformer 通过分层分块的方式将输入图像分成多个小块,并在不同的块之间建立跨块连接,从而在保留图像全局信息的同时,降低了计算和存储成本,提高了模型的效率和可扩展性;Swin-Transformer 在许多图像分类的公开数据集的基准测试中取得了优异的表现,并且在处理大规模图像数据集时,其效率和可扩展性也得到了显著提升。因此,近年来研究人员正在不断探索 Swin-Transformer 在计算机视觉领域的应用和优化。

2.1 模型结构

Swin Transformer 主要包括两个组成部分:基础网络和分层分块机制。基础网络采用传统的 Transformer 结构,包括多个 Transformer 编码器层和一个全连接层。分层分块机制则是在输入图像的特征图上进行的,通过将特征图分成多个小块,并在不同的块之间建立跨块连接,使得模型能够更好地处理大规模图像数据集。Swin-Transformer 模型结构示意图如图 2 所示。

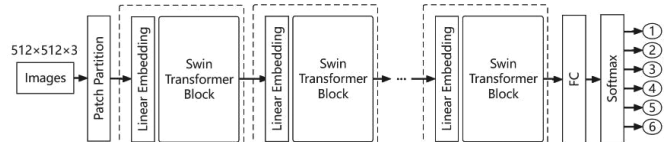


图 2 Swin-Transformer 模型结构示意图

Fig. 2 Swin-Transformer model structure

2.2 模型训练

步骤 1:设置损失函数。训练分类模型时使用交叉熵损失 (CrossEntropyLoss)。

$$CELoss(p, y) = - \sum_{i=1}^N y_i \cdot \log(p_i) \quad (4)$$

其中: p_i 是模型预测样本 i 呈现为各个类别的概率值, y_i 是样本 i 的真实标签, N 是样本总数。

步骤 2:设置评价指标。对于病虫害分类识别模型,主要的评价指标为识别准确率,即预测类别结果与实际类别结果一致的图像数量占测试集图像总数的比例。

步骤 3:设置其他参数训练模型。实验中设置 $batch_size$ 和 $num_workers$ 为 4,学习率($learning_rate$)为 0.01,迭代轮数($epoch$)为 100, num_class 为 2。每迭代一轮后,对测试集进行一次评估,并统计模型的效果。

步骤 4:调试 *patch_size* 参数的大小。将训练图片切割成合适大小的块,依次设置 *patch_size* 为 4、8、16、32,经过训练调试后发现 *patch_size* 为 16 时最合适。

步骤 5:调试 Swin-Block 模块的层数。依次测试了 12、24、32 层。由于西瓜叶片图像数据集的图片尺寸较大,图像中所含的信息量较多,所以需要更多的网络参数进行拟合。经过训练调试后发现最优网络层数为 32 层。

3 实验及结果分析(Experiment and result analysis)

3.1 数据集

本研究采用的数据集是研究人员在自然环境下拍摄的照片(图 3)。数据集中包含 3 928 张西瓜叶片图像,图像像素大小为 512×512。病虫害种类繁多,包括叶斑病、叶枯病、炭疽病、斑点病、蔓枯病、黑斑病、病毒病、红蜘蛛病、疫病、霜霉病等。但是,在数据集中,有部分病虫害种类的西瓜叶片数据量有限,无法满足训练分类模型的需求,实验时删除病虫害种类中数据量少于 50 张图片的数据。最终根据病虫害的分类,将数据集划分为正常、红蜘蛛病、炭疽病、叶枯病、叶斑病和黑斑病六类,共计 3 811 张图像。

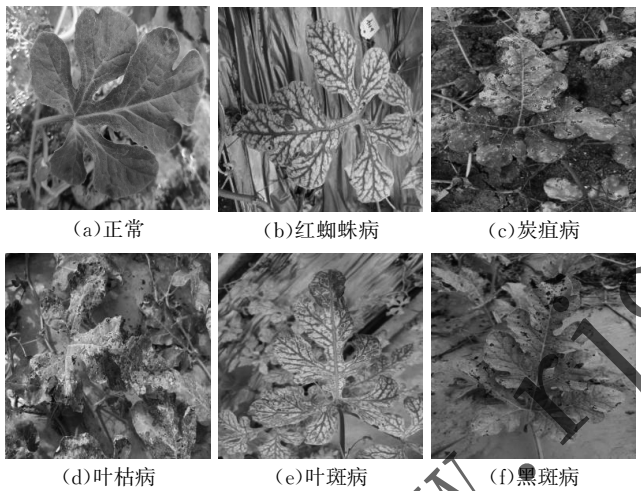


图 3 西瓜叶片图像

Fig. 3 Watermelon leaves

使用标注工具 LabelMe 对西瓜叶片进行了语义分割,从图像中框选主要待识别的叶片,从而形成相应的 Mask 图片集。对西瓜叶片图像的标注如图 4 所示。

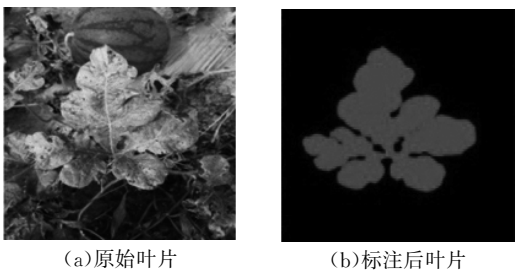


图 4 西瓜叶片图像的标注

Fig. 4 Annotation of watermelon leaf images

3.2 实验结果

实验运行环境为 Ubuntu16.04 Server, GPU 为 Tesla T4 15 GB, CUDA 版本为 10.2, Python 版本为 3.7, 实验基于 PyTorch 框架进行。

3.2.1 叶片分割

实验对 UNet、Trans-UNet^[6] 和 Swin-UNet^[7] 三个模型应用于西瓜叶片图像语义分割的效果进行了比较。表 1 列出了三个模型语义分割的最终评价指标。实验结果表明, UNet 模型的表现最佳, 显著优于 Trans-UNet 模型, 略胜于 Swin-UNet 模型。

表 1 三种模型对应的 PA 和 MIoU 指标

Tab.1 PA and MIoU indicators corresponding to the three models

模型	PA/%	MIoU/%
UNet	93.3	97.9
Trans-UNet	67.2	71.6
Swin-UNet	84.4	88.1

图 5 展示了 UNet 模型对西瓜叶片图像进行语义分割的效果。



图 5 UNet 模型效果图

Fig. 5 Diagram of UNet model effect

3.3.2 病虫害识别

实验对 LeNet5、ResNet18、ResNet50、ResNet101、ResNet152^[8]、Vision-Transformer^[9] 和 Swin-Transformer^[10] 七个模型在未分割西瓜叶片和分割后西瓜叶片在病虫害识别方面的表现进行了比较。表 2 列出了七个分类模型使用分割前后图像进行训练所对应的 Acc 指标。实验结果可以看出, 除 Vision Transformer 模型外, 使用分割后的西瓜叶片图像比使用原始西瓜叶片图像的效果更好。直接在原始图像上使用 Swin-Transformer 模型进行病虫害识别, 识别准确率为 89.7%, 本文提出的先使用 UNet 模型分割, 再使用 Swin-Transformer 模型进行病虫害识别, 识别准确率提高了 3.2%, 达到了 92.9%。

表 2 使用分割前后图像进行训练所对应的 Acc 指标

Tab.2 The corresponding Acc indicators trained on images before and after segmentation

模型	Acc_Train_By_Origin_Pic/%	Acc_Train_By_Seg_Pic/%
LeNet-5	45.4	59.7
ResNet-18	56.2	66.7
ResNet-50	61.6	70.5
ResNet-101	73.0	80.3
ResNet-152	82.5	85.3
Vision-Transformer	74.3	70.4
Swin-Transformer	89.7	92.9

(下转第 73 页)