

一种基于改进 U-Net 的植物图像分割算法

赵艳杰,郭晓丽,刘洋,娜茜泰,李雅婷,张铭梓,秦文强

(内蒙古大学电子信息工程学院, 呼和浩特 010021)

摘要: 图像分割技术是图像分析的重要步骤。本文针对野外采集的植物图片存在多尺度及理想数据量不足等问题, 提出一种多尺度 Res-Att-Unet 模型以实现自主植物图像分割。该方法在 U-Net 的收缩网络中增加多尺度机制, 在输出端以相应的 Label 图进行深监督, 并在模型的扩张网络部分加入双通道注意力机制, 以抑制无关信息表达, 各层新增的残差块使得模型可以提取更加丰富的特征信息, 实验表明无论在花朵型图像还是叶片型图像中均取得不错的分割效果。

关键词: 图像分割; U-Net; 残差网络; 注意力机制

中图分类号: O422 文献标识码: A

A Plant Image Segmentation Algorithm Based on Improved U-Net

ZHAO Yan-jie, GUO Xiao-li, Liu Yang, NA Xi-tai,
ZHANG Ming-zi, QIN Wen-qiang

(College of Electronic Information Engineering, Inner Mongolia University of China, Hohhot, 010021, China)

Abstract: Image segmentation technology is an important step in image analysis. In this paper, a multi-scale Res-Att-Unet model is proposed to achieve autonomous plant image segmentation in order to solve the problems of insufficient multi-scale and ideal data of plant images collected in the field. Model in the contraction of U-Net network with multi-scale mechanism, increase the corresponding figure for deep Label supervision, and in the expansion of the network model part to join the dual channel attention mechanism, inhibition of irrelevant information expression, in addition the new residual blocks in each layer make the model can extract more abundant image information, the experiments show that good segmentation results are achieved in both flower-shaped images and leaf-shaped images.

Key words: image segmentation; U-Net; residual network; attentional mechanism

1 引言

图像分割是理解和分析图像的重要环节, 即基于图像的颜色、灰度、纹理等特征对图像的主体部分进行提取^[1], 目前被广泛的应用于目标识别、目标追踪、3D 重构等技术的基础处理过程, 本文主要

的分割对象为野外采集的花朵型植物和叶片型植物。T. Saitoh 等人^[2]通过选择局部代价与路径长度之比最小的路径来提取边界, 并开发了一种分割识别一体的界面, 实验对 600 幅图片进行测试, 获得了 97% 的花朵分割准确率, 但是当待分割主体中心处有清晰边缘时, 会出现分割不完整需要手动在边界输入分割点的情况。K Yang^[3]等人通过收集标注得到 2500 张复杂背景叶片图像, 分割和识别网络分别为 Mask-R-CNN 和 VGG16 网络的第一部分选择候选目标区域, 接着将目标候选区域输入网络的第

基金项目: 国家自然科学基金项目 (62071257,61941304)
通信作者: 刘洋 (1981-), 男 (满族), 内蒙古通辽市人, 博士, 教授, yangliuimu@163.com。

二部分产生叶片分割图和识别率，最终的误分率（Misclassification Error, ME）为 1.15%，但其实验数据量不够充足且模型训练和验证的时间相对较长。刘永波等人^[4]使用两组 U-Net 进行并行计算，实现玉米的叶部病斑分割，其叶片分割的 MIOU 值达到 96.33%，但其只考虑了对叶片的分割，很多病症存在于植物果实等部位，不具有普适性。邓颖等人^[5]针对花量密集、多遮挡场景，证明当以 ResNeXt50 为主卷积网络，Mask-R-CNN 获得较高准确率和较低误差率且模型运行时间短，该方法创造性的实现了密集、大量场景中花朵数量的统计但其准确率有待进一步提升。钟伟镇等人^[6]为对植物后续的生长状态识别和监控，提出一种基于 Mask-RCNN 网络的分割和识别复杂背景多叶片的情况，其中将 ResNet101 和 FPN 作为系统的主干架构用于图片的特征提取，兴趣区域（ROI）分类器将最终的植物进行分类，另外系统还设置了边框回归器来对目标区域进行位置校正，另一网络分支为 FCN 网络^[7,8]来实现对植物图像的分割，最终将识别和分割的网络相结合生成一幅与原始图片大小相同的输出，但是其样本图片拍摄时对背景要求较高，模型效果容易受拍摄角度、光照等的影响。Hu J 等人^[9]提出一种在多通道关系上搭建的注意力模型^[10-13]，在分类任务中取得了更好的识别率，实验表明注意力机制的应用可以有效的提高网络整体性能。仝真等人^[14]结合 ResNet34^[15]和 U-Net 网络特点形成 ResNet-U-Net 模型对立木图像进行初步分割，后使用动态阈值分割前景和背景，最后验证该方法误分率降低了 3.5%，运用评分奖惩思想，在立木分割中获得较高准确率和较强的鲁棒性，但是当光照出现过亮过暗的情况时，其图像分割误差较大。

在对植物分割过程中其花朵和叶片都是主要研究部位，针对目前网络对于两种植物部位分割缺乏普适性，且植物生长环境中图片数据取样存在多尺度、分辨率等问题，本文提出多尺度 Res-Att-UNet 模型，实验表明该方法具有较强的泛化性，无论在花朵型数据集还是叶片型植物数据集中都表现出良好的收敛性、较高的精度等特点。

2 基于改进 U-Net 植物图像分割模型

实验中整个项目流程如图 1 所示，本文的工作内容主要为第三部分即将无人机回传的植物图片进行分割操作。分割作为后期识别的基础，分割的好

坏直接决定了植物识别准确率的高低。



图 1 项目实施过程

由于原始 U-Net 网络在植物图像分割过程中存在特征提取不够全面、缺乏对多尺度植物图像包容性等问题，并不适合实际项目需求，故基于 U-Net 模型做出如下三方面改进：

（1）残差模块：针对 U-Net 网络层数少提取到的植物图像特征信息并不充分，网络中增加残差模块在提升网络性能的同时避免了模型加深导致的梯度消失问题。

（2）多尺度机制：无人机在野外拍摄植物图片时，由于拍摄角度等原因会采集到不同尺度的图像，在网络的收缩部分引入多尺度输入机制。输入 U-Net 网络中的不同尺度数据与本身对应通道信息融合可以提取到更加丰富的上下文信息，无论大尺度还是小尺度植物图片均能取得较好的分割效果。

（3）双通道注意力机制：为了使网络对植物的主体部分实现更精准的捕捉，在模型的扩张部分增加双通道注意力机制，可以有效提高整个网络的分割准确度。改进后的 U-Net 模型结构如图 2 所示。

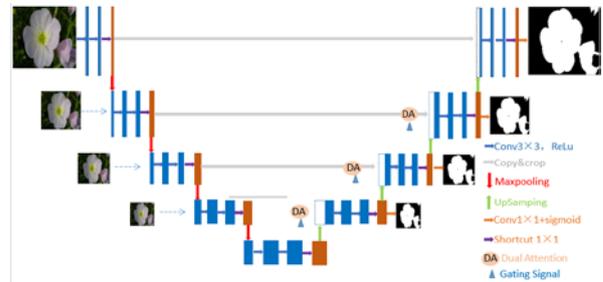


图 2 改进后的 U-Net 分割模型

2.1 残差结构模块

由于原始 U-Net 网络层数较少，对于植物图像的特征信息提取不够全面准确，造成分割后的图像粗糙且精度低，于是本文提出在原有的 U-Net 网络中加入残差块，其独特的 shortcut 连接方式，相当于在网络中加入了一条跳过权重层的新路径，即在最终的输出上新增一项没有经过权重衰减的梯度，可以融合植物图像浅层和深层特征信息。目前残差块有两层或者三层如图 3 所示，由于该场景下两层和三层残差块得到的分割准确率相差十分微小，考虑减少训练时间复杂度，本文改进的模型中采用两层残差块。引入残差块解决了当网络加深到某一程

度，由于梯度消失导致模型的分割效果出现不升反降的情况。

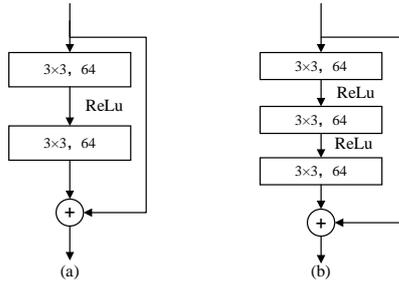


图 3 ResNet 的残差学习块^[16]

在原始 U-Net 模型中加入残差块，使得网络模型在相对简单的情况下^[17]，降低模型训练参数和训练时长且可以在本层的末端引入了初始端的植物特征信息，增强了各层之间的特征融合，并在一定程度上解决了网络层加深后导致分割模型性能退化的问题，但是野外拍摄植物图片时由于距离角度等原因得到的多尺度图片同样会影响分割精度，接下来将利用多尺度目标分割模块对该问题做出解决。

2.2 多尺度目标分割

在实际项目中，无人机对植物的拍摄会得到不同尺度的植物图像，这为植物图像分割带来了一定难度。从 SPPNET^[18]中将多尺度特征融合的思想引用至本文的模型中。为了得到不同尺度的输入图片，首先将已有的植物图片通过连续下采样之后得到分辨率逐渐降低的一系列图像构成图像金字塔，其次将得到的多尺度图像输入构建好的网络提取不同尺度下的特征信息，最后使用相应尺寸下的 Label 图对植物分割结果进行深监督。

如图 2 所示，模型的左侧为多输入的下采样部分，通过对原始图片进行缩小，得到相当于原始图片大小 1/2、1/4、1/8 的三组数据，原始尺寸图像输入网络之后经过两个 3×3 的卷积层以及池化层得到 1/2 原始图像大小的特征图，此时输入该层的多尺度图像经过两层卷积以及批量归一化 (BN)，再经过 ReLu 激活函数进行非线性映射与上一层得到的特征图进行特征拼接，新得到的拼接特征再进行两次卷积、批量归一化、激活函数、最大池化，以此类推。在模型的上采样部分，对 Sigmoid 函数输出的分割预测图，使用与该分割图相对应尺度监督标签样本进行深监督，以此加快网络收敛速度。

多尺度输入机制解决了多尺度图像对后期分割带来的困难。融合了金字塔结构的 U-Net 网络把提

取到的多尺度信息与本身网络对应通道信息进行融合从而实现不同感受野的上下文信息融合以及植物图像的表层轮廓信息和深层纹理信息的融合，无论是大尺度目标分割还是小尺度目标分割均可以获得更完备的语义信息。但是由于 U-Net 独特的结构特征在对编码和解码特征进行融合的同时，一些对分割无用的信息同样被传递到解码层造成分割精度降低，针对这样的问题提出双通道注意力机制来抑制分割过程的无关信息表达。

2.3 双通道注意力机制

注意力模型最初是由人类的注意力机制启发而来。仿照人类这一机制衍生的注意力模型首先被应用于机器翻译中，后期被广泛的用于计算机、语音、自然语言处理等领域^[19]。

由于传统的 U-Net 网络为融合更加丰富的信息，在模型的扩张网络部分会接收到模型收缩网络中的浅层信息，在接收到有用特征信息的同时会接收到对最终分割目标无用的特征信息，此时注意力机制的引入会让网络学习我们更为关注的区域，增强模型的分割性能。为此本文融合了两种注意力机制用于提高网络信息处理的效率和准确性。

(1) 通道注意力机制

通道注意力机制(Channel Attention, CA)^[20]概念同样来源于人类的注意力机制，在神经网络中，每一张输入图片可以表示为 (W, H, C) 形式，其中 W 和 H 分别为输入图像的长和宽， C 为图像的通道数。例如在植物图像分割网络中，输入的初始图像 $(256, 256, 3)$ 通过不同卷积核来提取图像中的特征信息，在经过 64 的卷积核后变为 $(256, 256, 64)$ 新的特征图，64 个特征图中有的特征图上占据的关键信息较多，有的对于最终的分割任务并没有多大作用，此时就需要对各个特征图进行权值分配来调节对最终结果影响程度的比例。

在 U-Net 网络中不同的特征通道对于最终的表现各不相同，同样的在响应的过程中很可能有一些重要的信息被削弱，在上采样的自下而上的三层输出端各加入通道注意力，使得特征通道对最终的分割表现有高响应性。用 x 表示网络模型某一层的特征图， $x = [x_1, x_2, \dots, x_c]$ 中的 x_i 表示第 i 个特征图，然后需要对 c 个特征图做平均合并得到该通道的特征向量，之后通过两个全连接层来得到各通道之间的关联关系，最终使用 Sigmoid 函数将特征向量映射到 $[0, 1]$ 范围之内。

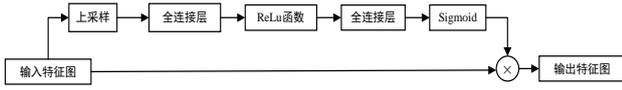


图 4 通道注意力模块

本文的通道注意力模块的设计如图 4 所示：

$$CA = f(x, W) = \sigma(fc_2(\delta(fc_1(x, W_1)), W_2)) \quad (1)$$

上述公式中， W 表示该通道注意力模块中所有的参数， σ 表示 Sigmoid 函数， fc 表示全连接层， δ 为 ReLu 函数，得到每层特征图的 CA 系数后，对各层特征图进行加权得到最终输出特征值：

$$\tilde{f}_h = CA \square x_h \quad (2)$$

通道注意力机制的本质是给不同特征赋以不同的权重，从而学习对植物分割任务更有用的特征通道。

(2) 空间注意力机制

在图像处理过程中，低级特征往往存在很多图像的细节信息，分割任务中我们更希望得到目标区域和背景之间的边界轮廓，对其余区域的细节纹理信息不进行过多关注，即对图像所有的空间信息不能进行同等考虑。如果将通道注意力机制比作让网络去看什么，那空间注意力机制(Spatial Attention, SA)^[21-22]则是让网络去看哪里，具体模型如图 5 所示，模型中为了扩大感受野和获取更加丰富的全局特征，使用两个相邻的卷积核，最后使用 Sigmoid 函数将最终特征映射到[0,1]之间。

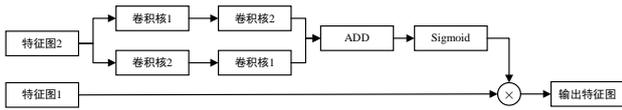


图 5 空间注意力机制模块

图 5 中的特征图 2 为经过通道注意力机制得到的高维特征图，特征图 1 为当前通道特征图，现将涉及到的公式列举如下：

$$C_1 = conv_2(conv_1(\tilde{x}, W_1^1), W_1^2) \quad (3)$$

$$C_2 = conv_1(conv_2(\tilde{x}, W_2^1), W_2^2) \quad (4)$$

$$SA = \sigma(C_1 + C_2) \quad (5)$$

最终的输出结果为：

$$\tilde{f}_l = SA \square x_l \quad (6)$$

其中的 W 为空间注意力模块的参数， $conv_1$ 、 $conv_2$ 为卷积层， \tilde{x} 表示经过通道注意力机制得到的高维特征图， σ 为 Sigmoid 函数， C_1 、 C_2 为实现扩大感受野及捕获高维特征空间信息得到的两分支卷

积输出值， x_l 为当前通道的低维特征图。在进行图像分割时，注意力机制的引用可以更好的关注任务相关区域，提高模型的分割准确率。

融合后的注意力模块见图 6 所示，其中高维输出特征图经过通道注意力机制后进行上采样，再通过空间注意力模块与经过跳跃结构后低维特征 1 相乘，相乘之后的结果与经过上采样后高维特征融合，得到处理之后的低维特征 2，通道注意力机制和空间注意力机制的融合可以使模型在学习过程中抑制无用特征信息表达。

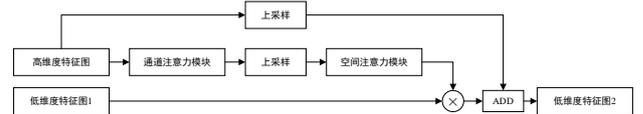


图 6 双注意力通道模块

3 仿真实验

3.1 实验环境及数据集

为确保实验数据的有效性和准确性，本文所有实验均采用同一计算机进行实验完成，计算机的相关硬件配置及所用软件环境情况如表 1 所示：

表 1 实验中计算机的硬件配置和软件环境

名称	配置说明
操作系统	Windows10
CPU	Intel (R) Core(TM)i5-10300H CPU@2.5GHZ
GPU	NVIDIA GeForce GTX 1660 Ti/内存:16GB
开发语言	Python3.6
CUDA	CUDA 11.0
Cudnn	cuda 11.2
深度学习框架	Tensorflow GPU 2.2.0
Anaconda	Anaconda3

实验中所用到的花朵型植物数据集为 Oxford 102flowers 见图 7 所示，由于该数据集最初用于植物的识别领域，故本文对该数据集中的 8189 张图片进行手动分割，分割软件采用图像标注软件 Labelme，叶片型植物数据集采用 cvppp 数据集如图 8 所示，由于该数据集中只有 810 张已经分割好的图片，为提高模型的泛化能力，对已有的图片进行图像角度变换、翻转、尺度变换、对比度变换、添加噪声等^[23]数据增强操作，将植物叶片数据集扩充

为 8870 张。



图 7 花朵部分数据集



图 8 叶片部分数据集

3.2 评价指标

为了确保模型的有效性，本文采用多指标来对分割结果进行评估。其中 FN 表示被模型错误的分为负例，实际是正例；FP 为被模型错误的分为正例，实际是负例；TP 为被模型正确的分为正例，实际为正例；TN 为被模型正确的分为负例，实际为负例。

(1) 像素准确率

像素准确率 (Pixel accuracy, PA) 是一种比较简单的评价指标，它表示的是像素被正确分类的个数与总像素的比值，即图像中像素被正确分类的百分比，像素准确率的表达式为：

$$Pixel\ accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

(2) 均交并比

均交并比 (Mean Intersection over Union, MIOU) 是对 IOU 指标优化的结果，IOU 表示真实 Label 值与预测值之间的交集同真实值与预测值之间并集的比值，而 MIOU 是分别计算每个类的 IOU，然后对所有类的 IOU 求取均值，MIOU 被广泛的用于各种分割场景中。MIOU 的表达式为：

$$MIOU = \frac{TP}{FP + FN + TP} \quad (8)$$

(3) F1 score (Dice score)

F1 score 又称为 Dice score，是统计学用来衡量二分类任务模型精确度的指标，其数学表达式为。

$$DSC = \frac{2 \frac{TP}{TP + FN} \frac{TP}{TP + FP}}{\frac{TP}{TP + FN} + \frac{TP}{TP + FP}} = \frac{2TP}{2TP + FN + FP} \quad (9)$$

3.3 实验结果对比

实验过程分为两个部分，分别为花朵型植物数据集以及叶片型植物数据集在改进后的 U-Net 和 FCN 模型的对比情况，实验中使用的图片大小均为 $256 \times 256 \times 3$ ，在每层卷积操作之后都使用批量归一化和 ReLu 函数，在模型训练过程中使用 Adam 优化器，初始学习率设置为 $2e-4$ ，batch_size 设置为 2，迭代次数为 100 次，参数设置见表 2 所示。

表 2 训练过程部分超参数

单元通道数	64,128,256,512,1024
batch_size	2
损失函数	Cross Entropy Loss
优化算法	Adam
初始学习率	0.0002
迭代次数	100

(1) 花朵型数据集各模型分割精度对比

图 9 为花朵型数据集在 FCN-8s、原始 U-Net、Res-Unet、多尺度 Res-Unet、多尺度 Res-Att-Unet 模型训练过程中 Loss 值的变化情况对比。

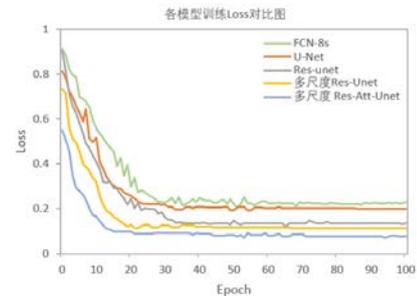


图 9 网络训练过程中各模型 Loss 值

可以看到在 100 次迭代范围内五条曲线的 Loss 值均经过持续下降趋于收敛，证明模型对于该场景的植物图像分割有效，其中多尺度 Res-Att-Unet 模

型在经过短暂的训练之后达到收敛，模型的训练时间短且模型训练过程 Loss 曲线稳定下降，模型稳定性及用时均优于其他模型。

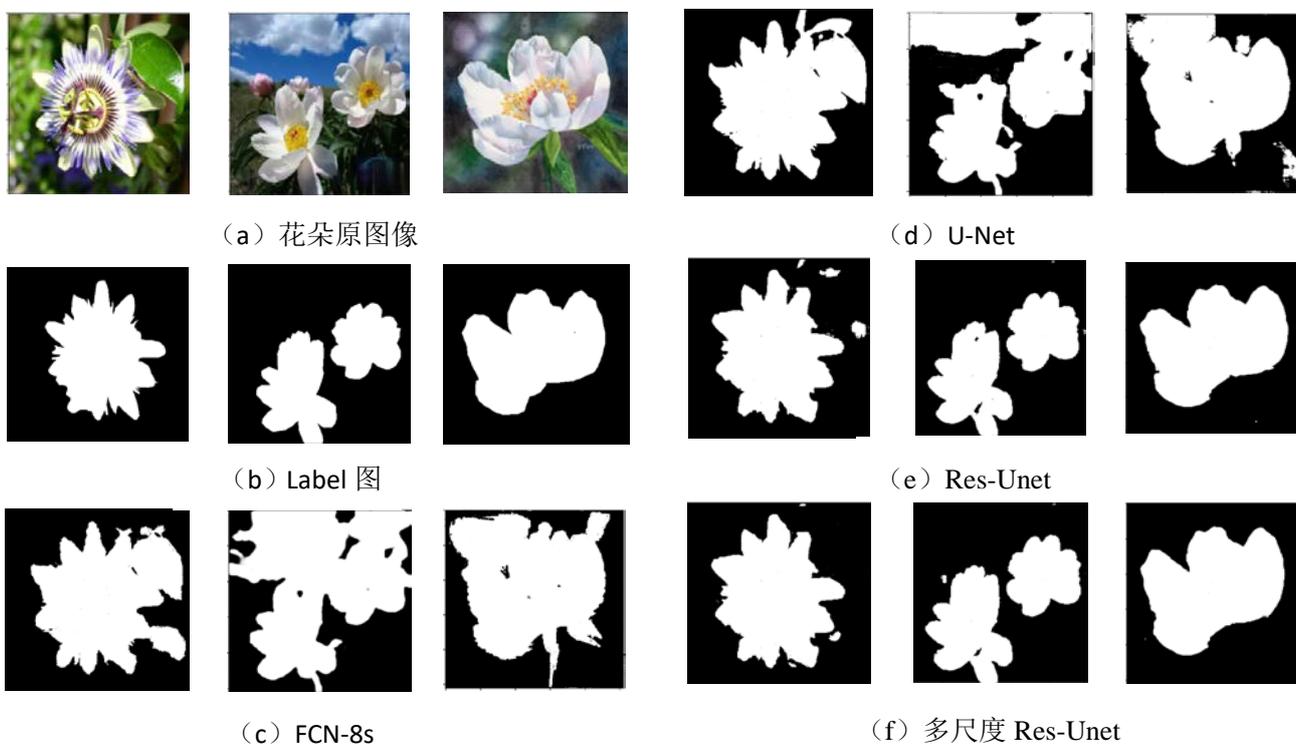
表 3 为各训练模型最终在三种精度评价指标上的表现情况，由表可知 U-Net 的整体表现要优于 FCN-8s 模型，这是由于 U-Net 独特的结构融合了植物图像的浅层轮廓信息和深层纹理信息，改进后的多尺度 Res-Att-Unet 模型像素准确率、MIOU、F1 score 指标较原始 U-Net 网络分别提高 0.1274、0.1209、0.1402。表明改进后的模型在主体为花朵的植物数据集中取得较 FCN-8s、原始 U-Net 模型更优

的分割准确率。

现将测试图片输入已经训练好的模型中观测最终分割效果，各模型分割对比情况如图 10 所示，其中 (a) 为原始图像 (b) 原始图像对应图片的 Label 图 (c) 为 FCN-8s 模型分割图 (d) 为 U-Net 模型分割图像 (e) 为加入残差模块 Res-Unet 模型的分割效果图 (f) 为加入残差模块及多输入机制的多尺度 Res-Unet 模型的分割图 (g) 为加入残差模块、多尺度输入机制及双通道注意力机制的多尺度 Res-Att-UNet 模型的分割效果。

表 3 花朵型数据集在不同模型上的精度表现情况

模型	像素准确率	均交并比	F1 score
Model	Pixel accuracy	Mean IOU	Dice score
FCN-8s	0.7732	0.7124	0.7486
U-Net	0.7959	0.7119	0.7647
Res-Unet	0.8654	0.7412	0.8497
多尺度 Res-Unet	0.8864	0.7491	0.8633
多尺度 Res-Att-Unet	0.9233	0.8328	0.9049





(g) 多尺度 Res-Att-Unet

图 10 各模型测试情况

观测图 10，可以看到虽然 FCN-8s 模型融合了底层的特征信息，但是其对图像细节区域的分割仍十分粗糙，U-Net 网络融合了各自通道的浅层信息，其分割效果虽优于 FCN-8s 模型，但由第一幅图像可以看出 U-Net 进行分割的过程中仍将叶子误分为花朵，在模型中加入残差结构和多尺度输入机制后该现象有明显好转但是部分细节信息表现仍欠佳，在引入双通道注意力机制后基本实现花朵主体部分的分割。第二张对于图像主体和背景颜色差异非常小的植物图片，原始 U-Net 模型和 FCN-8s 模型均将与分割主体颜色很相近的白云误分为花朵，优化后的模型该误分现象被有效避免。

(2) 叶片型数据集各模型分割精度对比

项目后续的植物识别操作主要是对植物的主体部分进行特征识别，由于有的植物并没有花朵结构，所以对叶片型植物进行分割同样十分必要。现就叶片型植物在各模型上进行分割对比验证。图 11 为 FCN-8s 和多尺度 Res-Att-Unet 模型在训练过程中损失函数的变化情况。可以看到两条曲线在开始时损失值快速下降，当下降到一定程度曲线趋于平缓，证明模型在训练过程中均达到收敛，但是多尺度 Res-Att-Unet 模型收敛时间和最终的损失值均优

于 FCN-8s 模型。

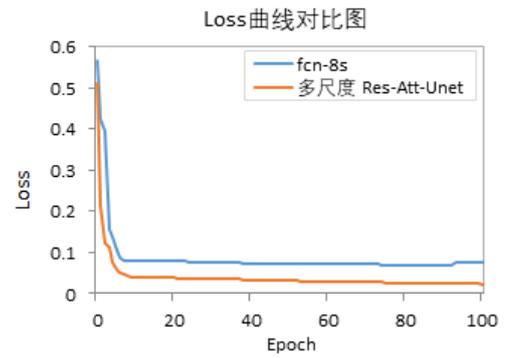


图 11 FCN-8s 和改进 U-Net 训练过程 Loss 值

表 4 为 FCN 经过三种不同上采样得到的模型与已改进的 U-Net 模型进行分割效果的对比情况。可以看到 FCN-32s 模型是直接将 conv7 层的特征图进行 32 倍上采样，其准确率分别为 0.8719、0.7119、0.8986。对于 FCN-16s 的模型，将 conv7 进行一次卷积操作，得到的特征图的大小为 conv7 的 2 倍大小，此时 2 倍大小的 conv7 与 pool4 的特征图大小相同，将 $2 \times \text{conv7}$ 与 pool4 进行拼接，并进行 16 倍的上采样后得到的分割准确率分别为 0.9233、0.7284、0.9123。对于 FCN-8s 模型先将 conv7 进行 4 倍的上采样，pool4 经过 2 倍的上采样，再将 pool3、 $2 \times \text{pool4}$ 、 $4 \times \text{conv7}$ 进行拼接，最后增加一个 8 倍的上采样层得到和原始图片相同大小的图像。此时的准确率分别为 0.9354、0.7389、0.9225。当在模型中增加多尺度残差双注意力机制后准确率则达到了 0.9462、0.8232、0.9365。相比表现效果最佳的 FCN-8s 模型准确率分别提高 0.0108、0.0843、0.014。

表 4 叶片型植物数据集在不同模型上的精度表现情况

模型	像素准确率	均交并比	F1 score
Model	Pixel accuracy	Mean IOU	Dice score
FCN-32s	0.8719	0.7119	0.8986
FCN-16s	0.9233	0.7284	0.9123
FCN-8s	0.9354	0.7389	0.9225
多尺度 Res-Att-Unet	0.9462	0.8232	0.9365

现将叶片测试图片分别输入训练好的

FCN-8s、FCN-16s、FCN-32s 和多尺度 Res-Att-Unet

模型得到图 12。因为 FCN 模型在进行像素分类的过程中没有充分考虑像素与像素之间的关系，其分割边缘存在明显锯齿状。采用多尺度 Res-Att-Unet 模型来进行叶片图像的分割，得到的叶片分割图像与 Label 图像非常接近，这是因为多尺度 Res-Att-Unet 模型在模型构建过程中融合了多层信息，且多尺度机制可以学习更加丰富的尺度信息，双通道注意力机制使模型更加专注图像分割主体，最终得到的分割图更加细腻精确。实验证明改进后的多尺度 Res-Att-Unet 模型在叶片型植物分割中同样取得不错的效果，达到预期分割要求。

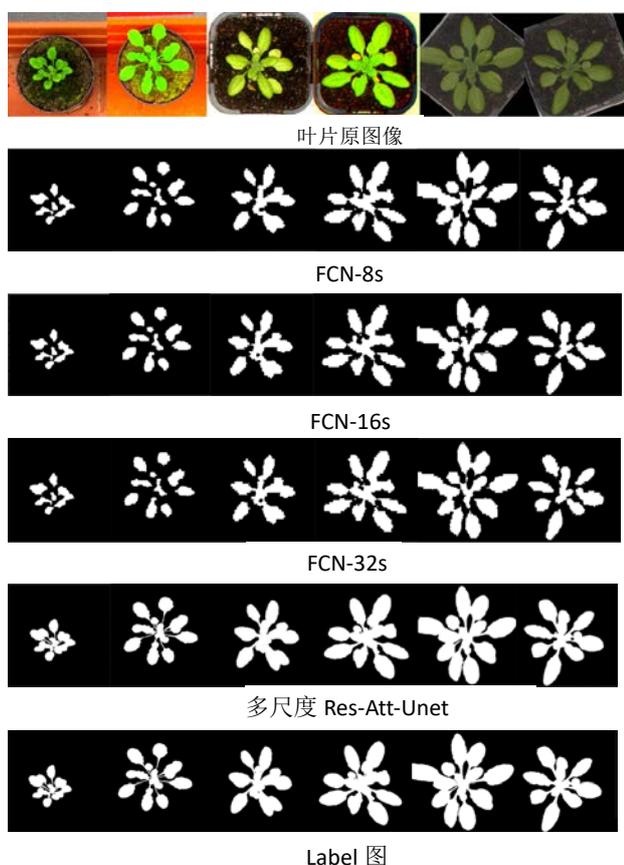


图 12 改进后的 U-Net 与 FCN 三种模型的测试效果

4 总结及未来工作

本文从实际项目需求背景出发，考虑到无人机采集到的野外植物图片存在多尺度问题，以及目前的网络对于花朵植物和叶片型植物缺乏普适性，对比验证了 U-Net 网络和 FCN 网络在两部分数据集上的表现，对表现较优的 U-Net 模型做出如下改进，针对其网络层数少导致分割精度低的问题，在网络中加入残差块；其次野外植物拍摄的多尺度问题通过带监督的多尺度模块得以解

决；最后由于 U-Net 网络自身的不足在融合底层有利信息的同时也融合到无关特征的问题，在网络解码部分加入双通道注意力机制。最终改进的多尺度 Res-Att-Unet 模型无论对于花朵型植物还是叶片型植物都具有不错的分割效果，证明了该分割算法在本项目背景下具有一定的可行性和有效性。

接下来的工作进一步探究将项目中无人机的植物图片自主采集、双目标定算法确定植物的具体位置、本文的植物图片分割、后续植物种类的识别这四个部分更好的结合起来，实现全自动植物种类识别，另外目前对于卷积神经网络的训练往往都需要大量的数据集进行深监督，但是在很多植物生长的环境中获取大量样本本身不是一件容易的事，在以后将针对小样本数据进行神经网络模型训练研究。

参考文献

- [1] 武鹏. 图像特征提取与分割算法在苹果图像中的应用[D]. 重庆:重庆大学, 2010.
- [2] Saitoh T, Aoki K, Kaneko T. Automatic recognition of blooming flowers[J]. *Ice Transactions on Information & Systems Pt*, 2004, 88(7): 27-30.
- [3] Yang K, Zhong W, Li F. Leaf Segmentation and Classification with a Complicated Background Using Deep Learning[J]. *Agronomy*, 2020, 10(11): 1721.
- [4] 刘永波,胡亮,曹艳等. 基于 U-Net 的玉米叶部病斑分割算法[J]. *中国农学通报*, 2021,37(5): 88-95.
- [5] 邓颖, 吴华瑞, 朱华吉. 基于实例分割的柑橘花朵识别及花量统计[J]. *农业工程*, 2020, 036(007): 200-207.
- [6] 钟伟镇, 刘鑫磊, 杨坤龙, 等. 基于 Mask-RCNN 的复杂背景下多目标叶片的分割和识别[J]. *浙江农业学报*, 2020, 32(11): 2059-2066.
- [7] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4): 640-651.
- [8] 周昌雄. 基于活动轮廓模型的图像分割方法研究[D]. 南京:南京航空航天大学, 2005.
- [9] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[C]. *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018, 7132-7141.
- [10] Xu K, Ba J, Kiros R, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[J]. *Computer Science*, 2015: 2048-2057.

-
- [11] Cho Kyunghyun, Courville Aaron, Bengio Yoshua. Describing Mul-timedia Content Using Attention-Based Encoder-Decoder Networks [J]. IEEE Transactions on Multimedia, 2015, 17(11): 1875-1886.
- [12] Raffel Colin, Ellis Daniel P W. Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems[EB/OE]. <https://arxiv.org/abs/1512.08756v4>, 2016-09.
- [13] Fe I W , Jiang M , Chen Q , etc. Residual Attention Network for Image Classification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 3156-3164.
- [14] 徐爱俊, 仝真. 一种基于改进 ResNet-UNet 网络的立木图像分割方法 [J]. 中南林业科技大学学报, 2021,41(01), 132-139
- [15] Tianqi Wang, Dong Eui Chang. Improved Reinforcement Learning through Imitation Learning Pretraining Towards Image-based Autonomous Driving[C]. 2019 19th International Conference on Control, Automation and Systems (ICCAS). 2019, 1306-1310.
- [16] 张效荣. 基于卷积神经网络的深度学习算法与应用研究[D]. 西安:西安电子科技大学, 2015.
- [17] Shangbo Mao, Deepu Rajan, Liang Tien Chia. Deep Residual Pooling Network for Texture Recognition[J]. Pattern Recognition, 2021.
- [18] He K, Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-16.
- [19] 杜亚伟. 图像高层语义理解的关键问题研究[D]. 成都: 电子科技大学, 2019.
- [20] Chen L, Zhang H, Xiao J, et al. SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [21] Tootell R B, Hadjikhani N, Hall E K, et al. The Retinotopy of Visual Spatial Attention[J]. Neuron, 1998, 21(6): 1409-1422.
- [22] Haochen Mei, Wenhui Lei, Ran Gu, et al. Automatic Segmentation of Gross Target Volume of Nasopharynx Cancer using Ensemble of Multiscale Deep Neural Networks with Spatial Attention[J]. Neurocomputing, 2021, 438: 211-222.
- [23] 王琢, 汪雅婷, 宋文龙, 等. 基于深度学习的叶片图像分割算法[J]. 森林工程, 2019,035(001): 42-46.