

嘉应学院

毕业论文（设计）

（2021 届）

题目：基于 Faster R-CNN 和无人机多光谱影像的
梅州柚树单株识别方法研究

申请学位_____理学学士_____

院 别_____地理科学与旅游学院_____

专 业_____地理信息科学_____

学生姓名_____赖为乾_____

学 号_____171080135_____

导师姓名_____熊永柱_____

2021 年 5 月 20 日

嘉应学院本科毕业论文



勤 俭 诚 信
立 己 树 人

毕业论文（设计）诚信承诺书

本人郑重承诺：

1.本人所呈交的毕业论文（设计），是在指导教师的指导下，严格按照学校有关规定完成的成果。

2.本人在毕业论文（设计）中引用他人的文献、数据、图件、资料等均加以注释和说明，毕业论文（设计）的结论等成果为本人独立完成。

3.本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4.如本人的毕业论文（设计）涉及抄袭或剽窃等行为，本人愿意承担由此而造成的一切后果及法律责任。

毕业论文（设计）作者：（签名）

指导老师（已阅）：（签名）

年 月 日

基于 Faster R-CNN 和无人机多光谱影像的梅州柚树单株识别方法研究

摘要

梅州柚树单株分布信息可以实现柚树的高效统计和管理。基于 Faster R-CNN 深度目标检测网络，提出了一种面向无人机多光谱遥感影像的柚树单株识别方法。通过遥感影像预处理技术，将可见光影像融合多光谱信息和植被指数先验信息，并采用 SEnet 模块优化 ResNet 网络，增加“Channel Attention”机制，使得增加遥感影像的波段时，模型精度也不会下降。以 2021 年 2-4 月梅州市梅县区石扇镇金柚产区的试验田的无人机多光谱影像为研究对象，在 2 月份的遥感影像上构建并训练模型，在 4 月份的遥感影像上测试模型的泛化能力。实验结果表明，融合多光谱信息和植被指数先验信息的遥感影像在改进后的 Faster R-CNN 模型的 $\text{IoU}=0.5$ 和 $\text{IoU}=0.5:0.05:0.95$ 的平均精度分别是 83.19%和 53.12%，比可见光遥感影像提高了 3.37、4.52 个百分点，获得了较好的分类性能。因此，本文方法能从野外柚树多光谱遥感影像中学习到表达力更强的语义特征，从而获得准确的柚树单株识别结果，为利用无人机多光谱遥感影像识别柚树提供了一种新的技术方法。

关键词：深度目标检测；柚树识别；无人机；多光谱遥感影像；Faster R-CNN

Research on Meizhou Pomelo Individual Plant Recognition Method Based on Faster R-CNN and UAV Multispectral Images

Abstract

The individual plant distribution information of Meizhou pomelo can realize efficient statistics and management of pomelo trees. Based on the Faster R-CNN deep target detection network, a pomelo individual plant identification method for UAV multi-spectral remote sensing images is proposed. Through the remote sensing image preprocessing technology, the visible light image is fused with multi-spectral information and vegetation index prior information, and the SEnet module is used to optimize the ResNet network, and the "Channel Attention" mechanism is added, so that the model accuracy will not decrease when the band of the remote sensing image is increased. Taking the UAV multi-spectral image of the experimental field in the "Golden Pomelo" production area of Shishan Town, Meixian District, Meizhou City from February to April 2021 as the research object, the model was constructed and trained on the remote sensing image in February, and the model was used in the remote sensing image in April. Test the generalization ability of the model. The experimental results show that the improved Faster R-CNN model of the remote sensing image fused with multi-spectral information and vegetation index prior information has an average accuracy of 83.19% and 53.12%, respectively, of IoU=0.5 and IoU=0.5:0.05:0.95. Visible light remote sensing images have increased by 3.37, 4.52 percentage points, and better classification performance has been obtained. Therefore, the method in this paper can learn more expressive semantic features from the multi-spectral remote sensing images of pomelo trees in the wild, so as to obtain accurate individual pomelo tree recognition results, which provides a way for the use of UAV multi-spectral remote sensing images to identify pomelo trees. A new technical method.

Keywords: Target Detection of Deep learning; Pomelo Detection; Drone; Multispectral Remote Sensing Image, Faster R-CNN

目录

摘要	I
Abstract	II
1 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
2 Faster R-CNN 架构及其改进方法.....	4
2.1 Faster R-CNN 历史与网络结构	4
2.1.1 Faster R-CNN 发展史.....	4
2.1.2 网络组织.....	4
2.2 区域生成网络（Region Proposal Network）	7
2.3 骨干网络（backbone）	7
2.3.1 ResNet（残差网络）	8
2.3.2 SEResNet（SEnet+ResNet）	8
2.4 改进后的 Faster R-CNN	10
3 研究区概况与数据集	11
3.1 研究区域概述.....	11
3.2 数据采集.....	12
3.3 数据预处理.....	14
3.4 数据增强.....	15
3.5 数据集构建.....	16
3.6 识别成果后处理.....	17
4 实验研究方案.....	19
4.1 实验方案.....	19
4.2 评价准则.....	20
4.3 模型训练.....	22
4.3.1 训练平台配置.....	22
4.3.2 模型超参数	23
5 实验结果分析.....	24
5.1 实验结果.....	24
5.1.1 改造模型对分类性能的影响	24
5.1.2 光谱特征对分类性能的影响	24
5.1.3 植被指数特征对分类性能的影响	25
5.2 测试结果.....	26

5.2.1 测试精度..... 26

5.2.2 测试集单株柚树分布图 27

6 结论与讨论.....28

6.1 结论.....28

6.2 讨论.....28

6.2.1 不足与改进 28

6.2.2 展望..... 28

附录.....30

附录 1 单波段影像归一化代码.....30

附录 2 随机水平翻转、随机亮度增强代码.....30

附录 3 判断相交图形重叠部分是否超过阈值代码.....31

附录 4 各个数据集训练结果（取最后 10 个 epoch）32

参考文献 35

致谢 36

1 绪论

1.1 研究背景及意义

梅州作为中国的金柚之乡，源先是从民国初年开始种植沙田柚，后经梅州本地的提纯复壮后，逐渐形成了具有梅州特色的“金柚”。梅州栽种金柚已有近百年的历史，金柚已成为梅州最具区域特色的支柱产业。截止 2020 年，梅县区种植金柚的面积为 25.5 万亩，产量约 55.6 万吨，是全国最大的金柚商品基地。2019 年梅县金柚鲜果产值 26 亿元，人均年收入 8170 元，占农业人口人均年收入的 40%，梅县金柚公共品牌评估价值高达 113.27 亿元，位居全省第一。

梅州市的柚树种植面积非常大，每个县均有种植。梅州市种植柚果有着多年经验，其中有水晶柚、蜜柚、沙田柚等多个品种。以梅州市行政县划分为基础，每个县均有自己地区的特有品种，如梅县区以种植沙田柚为主，而大埔区以种植蜜柚为主^[1]。梅州不仅柚果基地众多，梅州本地人也喜好在院子或自家田地里栽种柚树，因此梅州柚树有着局部集中但整体范围广的分布特点。柚树以柚果产业基地为主，以成行的片区形式栽种，此外还有个体户种植的柚树，以单株的形式零散分布。另外，金柚在种植过程中不仅容易受到潜叶蛾、疮痂病等病虫害还会受到干旱、缺素等不可控因素胁迫，对金柚进行长势监测能够掌握柚树的健康状况，做到精细化管理从而减少不必要的损耗，降低柚树的培育成本，间接提高柚树产值。

传统的柚树统计方法是基于农户调查和统计上报。这两种方法存在着周期较长、费用较高和数据不一致等问题。因此，探索一种高效地对柚树进行精准化的柚树单株识别技术，对柚树的长势监测、面积调查和产值估算的技术有着现实意义。

在 20 世纪 70 年代起，机器视觉技术开始被应用于农业中代替人工劳作，降低农业强度，提高工作效率^[2]。传统机器学习在遥感图像上的目标识别方法多采用 SVM（支持向量机）、最大似然法、随机森林、决策树、K 近邻和 KNN 等算法，这些算法只能学到人工设计的特征，如：纹理、颜色、形状等，应用在遥感影像上识别精度较低，无法满足现实需要。

近年来，随着人工智能领域信息挖掘技术的不断发展，无人机多光谱遥感数据有了新的机遇和挑战^[3]。通过深度学习等人工智能技术方法可以提取无人机多光谱遥感数据中更本质、抽象的可区分性特征，充分利用其丰富的光谱信息和精细的空间信息的数据优势、显著提高分类识别的精准度，在目标检测、场景分类、土地利用、覆盖制图等领域愈加广泛的应用^[3]。利用深度目标检测技术加高光谱高空间分辨率的遥感数据，可以有效提高柚树单株识别的准确率，对柚树进行时段性的病毒虫害监测和基于语义分割的面积调查。

本文采用的 Faster R-CNN 模型是深度卷积神经网络中比较成熟的目标检测算法。Faster R-CNN 是基于候选区域的 Two-stage 检测方法，该方法先提取到目标候选区域，然后再由神经网络做分类和回归，同时获得目标的边界和置信度，再通过 NMS（非极大值抑制）算法，返回置信度最高的预测框，实现单株柚树的识别。

1.2 国内外研究现状

随着精准农业的发展,利用遥感信息快速准确地分类农作物以获取其种植面积及空间分布信息的需求越来越大^[4]。利用遥感技术及新兴技术结合做农作物的长势监测、健康检测、产量估计和灾害评估等工作是对农业生产技术的管理和决策的基本前提。

目前,利用遥感技术检测作物的方式主要有卫星和无人机两种。高空卫星遥感覆盖面积大、探测周期短,是大范围农情信息检测的重要手段^[4]。相关应用主要包括提取地表植被指数、光谱、纹理和地形特征,以及融合雷达与光学影像等数据识别地物。另一方面,低空无人机遥感因具有灵活机动、时效性强和分辨率高等特点,已被广泛用于获取农田级区域的农情信息,特别是田地种植分布信息的提取。例如,杨蜀秦等基于深度语义分割的无人机多光谱遥感作物分类方法,利用改进后的 DeepLab V3+模型,有效地处理多光谱遥感图像,获得了比可见光分类作物更准确地分类结果^[4]。韩文霆等基于可见光遥感影像,结合支持向量机和决策树分类模型,进行土地利用和覆被分类研究^[5]。戴建国等从无人机遥感影像中提取纹理和低通滤波特征,采用支持向量机实现对农田作物等分类^[6]。WU 等利用农作物收获后和作物生长期的数字表面模型(DSM),采用基于对象的分类方法和最大似然法,结合遥感图像和作物高度信息识别作物类型^[7]。ESRI 公司在新版的软件中发布在 ArcPro 软件上基于深度学习目标检测算法在遥感图像上对棕榈树进行单株识别教程。

可以发现,对于无人机多光谱遥感的分类主要是基于人工选择特征,采用支持向量机等传统机器学习方法,分类结果具有一定局限性^[4]。在深度学习模型提出后,在农业分类上都取得了不错的效果。总体来说,对于野外植被的分类,基于深度学习的算法会比传统的基于人工选择特征的机器学习算法优势更明显。原因是野外植被容易受到光照、季节、太阳高度角、无人机成像精度等因素的影响,植被在遥感影像上具有多变性。因此,具有非线性思维的深度学习模型在遥感影像上的泛化性能比机器学习更好。

为了精准获取单株柚树的分布信息,本文提出一种基于无人机多光谱农田遥感影像的深度目标检测模型。首先用 Python 的 GDAL 库重写了数据 I/O 的方法,实现 TIFF 格式图像读取的端对端模型,单株柚树的预测框写入到 shapefile 矢量文件中,赋予每株柚树地理位置信息,后处理可以结合人工审核对预测框进行质量把控。由于融合了多光谱信息和植被指数先验信息的遥感影像波段数量会逐步增加,影像的波段数量的增加除了会带来更多的特征信息外,也会带来一定程度的噪声,这些噪声如果被模型当成特征学习,那么波段数量的增加反而会降低模型的分类性能。为了使模型能够正确获取融合了多光谱信息和植被先验信息的遥感影像特征,降低噪声的影响,本文采用的解决方法是用 SEnet 模块嵌入到 ResNet 网络的残差块(residual)中,从而给 ResNet 分类网络增加“Channel Attention”机制。SEnet 模块会给每一个通道进行加权运算,增强通道间的相关性以优化 Faster R-CNN 模型,克服其在多光谱植被遥感解译领域的限制和不足。为了验证模型的泛化能力,本文基于研究区域 2021 年 2 月至 4 月的遥感影像,取 2 月份影像制作训练集用来训练模型,4 月份影像制作测试

集，用来验证模型的泛化能力，3 月份影像作为补充数据。通过传统可见光遥感影像和融合多光谱信息和植被指数先验信息的遥感影像的对比，验证多光谱影像对于深度学习模型的分

类精准提升的有效性。

2 Faster R-CNN 架构及其改进方法

2.1 Faster R-CNN 历史与网络结构

2.1.1 Faster R-CNN 发展史

加州大学伯克利分校的 Ross Girshick^[8]及其同事在 2013 年发表了 R-CNN 方法的文章，将 CNN 方法引入目标检测领域，大大提高了目标检测效果，改变了目标检测领域的主要研究思路，紧随其后的系列文章：Fast R-CNN^[9]，Faster R-CNN^[10]显示了该方法的不断改进。这 3 种算法结合 VGG-16 网络结构应用到 PASCAL VOC 2012 标准数据集上的效果对比见表 1，Faster R-CNN 相比前面的成果，在图像检测精度与速度上都有了显著的提升。

表 1 R-CNN 系列算法的精度与速度对比表

Tab.1 Comparison table of accuracy and speed of R-CNN series algorithms

算法	mAP	fps
R C-NN	66.0	0.02
Fast R-CNN	66.9	0.45
Faster R-CNN	73.2	7.00

2.1.2 网络组织

基于 Faster R-CNN 的目标检测框架通过神经网络解决两个主要问题，一是 Faster R-CNN 通过区域生成网络生成可能包含目标的预测框；二是 Faster R-CNN 通过卷积神经层计算图像每个预测框的类别概率，并将类别概率最大值作为分类结果。

Faster R-CNN 的网络组织主要分为三个部分，头网络，区域生成网络（RPN）和分类网络。头网络使用预训练的分类网络，如 Resnet-50，VGG16，来提取图像特征从而得到特征图（Convolutional Feature Maps），头网络部分可以直接通过迁移学习的方法，获得一些训练的比较好的权重值，这样能达到更快收敛网络的效果。接在头网络后面的区域生成网络（RPN）和分类网络(backbone)主要用于学习由底层特征构成的更高级别的内容，实现解决具体方法的功能^[11]。

由头网络产生的卷积特征图进入由卷积层和全连接层构成的区域生成网络（RPN），区域生成网络将会根据特征图提取图像中可能包含目标前景的预测框，并在特征图中裁剪出这些预测框，这称为“Crop Pooling”，分类网络会对每个“Crop Pooling”中包含的对象进行分类，如图 2.1。

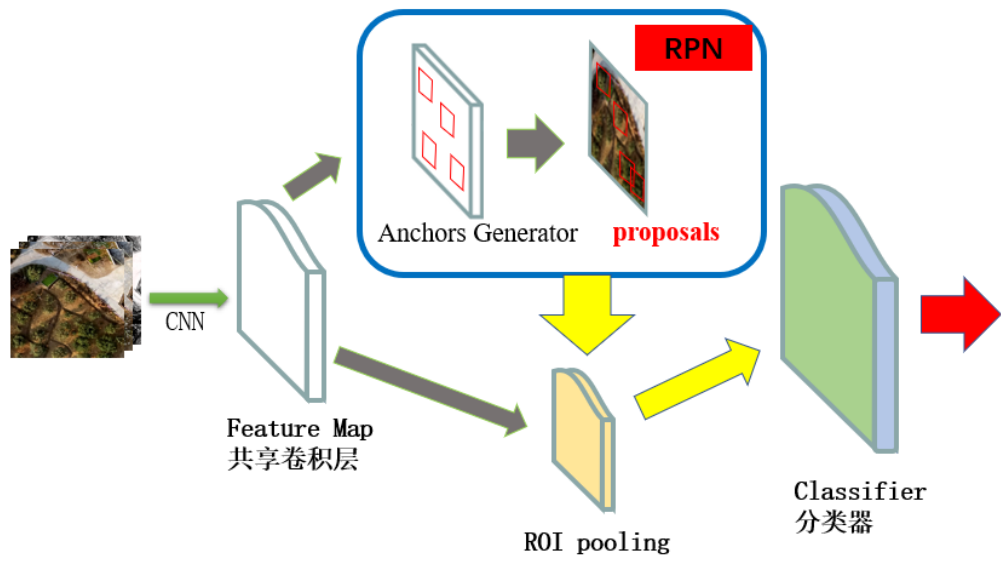


图 2.1 Faster R-CNN 架构图

Fig.2.1 Faster R-CNN architecture diagram

卷积神经网络一般由卷积层、池化层、非线性层、全连接层组成^[12]。

(1) 卷积层

卷积层的作用主要是提取图像的局部特征，通过多个卷积层的运算可以得到更深层次的特征图。在 Faster R-CNN 卷积神经网络中对所有的卷积都进行扩边处理（Pad=1，即填充一圈 0），使原图变为 $(M+2) * (N+2)$ 大小，再做 $3*3$ 卷积后输出 $M*N$ 。正是这种设置，导致 Conv layers 中的 conv 层不改变输入和输出图像矩阵大小^[11]。

(2) 池化层

池化层的作用是对输入的特征图进行压缩，虽然局部连接可以有效的降低参数数量，但其中包含了较多冗余信息。通过池化层，一方面为使特征图变小，简化网络计算复杂度；一方面进行特征压缩，提取主要特征。池化常采用平均值池化或者最大池化，分别对应滤波区域的最大值和平均值的处理, 如图 2.2。

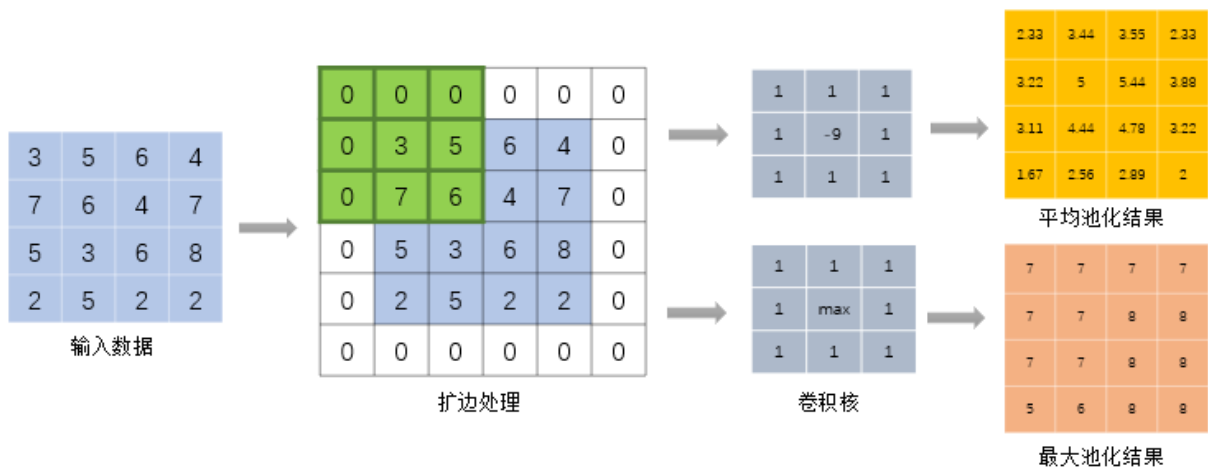


图 2.2 数据卷积过程图解

Fig.2.2 Diagram of data convolution process

(3) 非线性层

激活函数使深度神经网络具备分层的非线性映射学习能力。常见的激活函数主要有 sigmoid 函数、tanh 函数、ReLU 函数。ReLU 函数是目前深度学习中最常见的激活函数，表达式见式 2.1，对应图像见图 2.3。激活函数对输入信息进行非线性变换，变换后结果输入到下一层神经元当中。如果没有激活函数，那么神经元之间都是线性组合，输入值和输出值存在线性关系，这样会导致模型的泛化能力大幅降低。在神经元之前加入激活函数，在众多神经元的组合下，将线性函数变成非线性函数，可以使得特征表达逼近任何非线性函数，从而神经网络可以应用到众多的非线性模型中。

$$g(x) = \max\{0, x\} \quad (2.1)$$

式中， x 为激活函数前级输出单元的输出， $g(x)$ 表示只取正值。

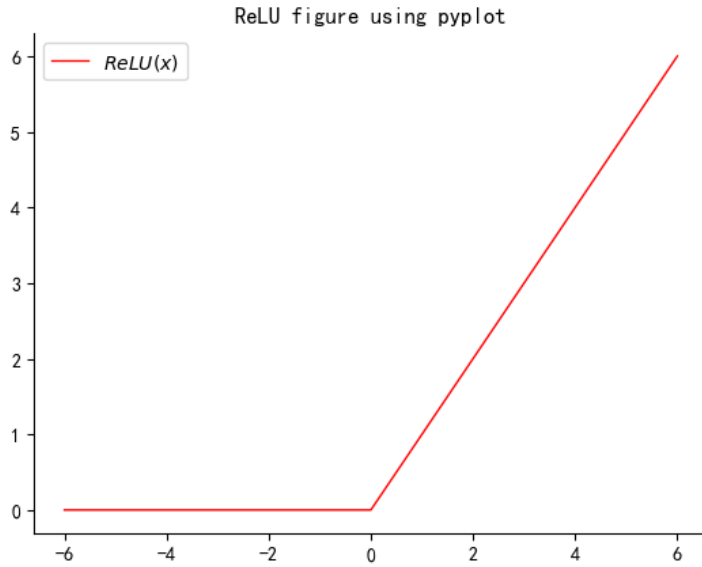


图 2.3 ReLU 函数图像

Fig.2.3 ReLU function

(4) 全连接层

全连接层在整个卷积神经网络中主要起到一个分类器的作用，但也能用作加权运算（SENet 模块中就这样使用），还能用于弱化参数，节省计算量。简而言之，卷积核取的是局部特征，全连接层的作用就是把局部特征重新通过权值矩阵组装成完整的图。常见的分类器有 Softmax、logistic、Boosting、Adaboost，其中 Softmax 分类器是卷积神经网络中常用的分类器，表达式见式 2.2。

$$s_i = \frac{e^{v_i}}{\sum_i^C e^{v_i}} \quad (2.2)$$

式中， v_i 是分类器前级输出单元的输出， i 表示类别索引，总的类别个数为 C 。 s_i 表示的是当前元素的指数与所有元素指数和的比值。

2.2 区域生成网络（Region Proposal Network）

区域生成网络（Region Proposal Network）第一次出现在世人眼中是在 Faster R-CNN 这个结构中，专门用来提取候选框，运行机制如图 2.4。在 R-CNN 和 Fast R-CNN 等物体检测架构中，主要是使用 Select Search 算法去生成预测框，平均一幅图像在 CPU 中耗时 2 s，耗时较长。作者提出区域生成网络，专门用来提取候选框，一方面 RPN 耗时更少，在 GPU 中一幅图的推理时间可达到 5 fps，几乎可以达到实时检测，另一方面 RPN 可以简单结合到 Fast R-CNN 成为一个整体。区域生成网络的引入，是真正意义上把物体检测整个流程整合到一个神经网络中，这个网络结构叫做 Faster R-CNN。简而言之，Faster R-CNN 就 RPN 结构 + Fast R-CNN 框架。

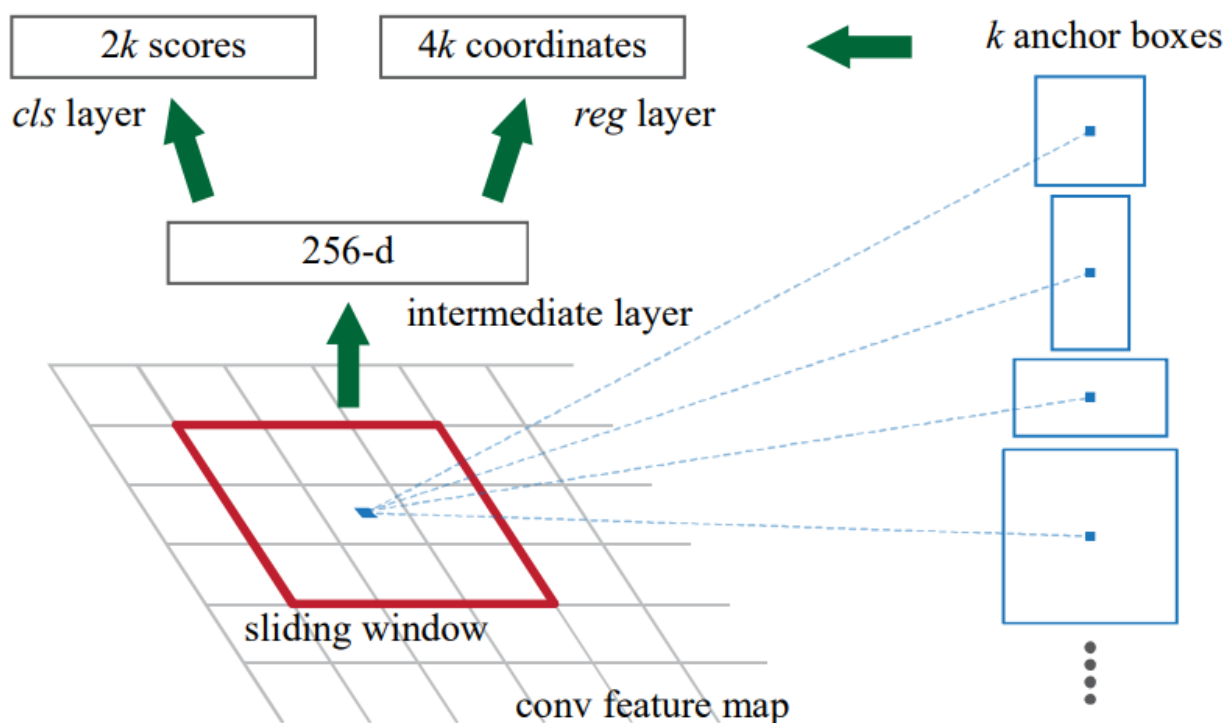


图 2.4 RPN 运作机制^[10]

Fig.2.4 RPN operation mechanism^[10]

2.3 骨干网络（backbone）

骨干网络的功能主要是对输入照片进行特征提取和对候选框进行分类。本文提出的目标检测模型中主要用到了三种骨干网络，分别是 ResNet-50 网络、SEResNet-50 网络和 SEResNet-101 网络。

2.3.1 ResNet（残差网络）

ResNet (Residual Neural Network)，又称残差网络，是由微软研究院的 Kaiming He^[13]等四名华人提出，通过使用 ResNet Unit 成功训练出 152 层的神经网络，并在 ILSVR2015 比赛中取得冠军。相比 VGG^[14]网络，ResNet 的网络层次更深，但是参数量却更少，效果突出。

传统的卷积网络或全连接网络在信息传递的时候会存在信息丢失、损耗等问题，同时还有导致梯度消失或者梯度爆炸的问题，导致网络层次越深，网络性能却倒退。ResNet 通过在卷积层之间增加 shortcut 或者 skip connections 结构，使得输入信息可以绕道传到输出，保护信息的完整性。如图 2.5。

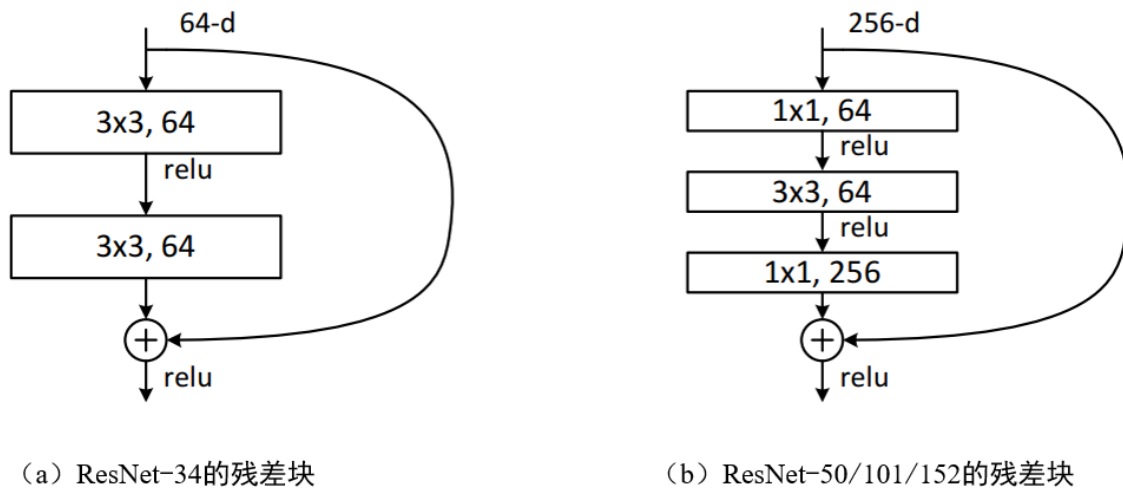


图 2.5 卷积残差结构^[13]

Fig.2.5 Convolutional residual structure^[13]

图 2.5a 的二层卷积残差结构用于网络层数不深的模型，层数一般在 50 层以内，超过 50 层的 ResNet 网络的残差块会使用图 2.5b 的三层残差结构，通过在 3*3 卷积层的上下方加入一个 1*1 的卷积核，能够有效降低模型的参数量。需要注意的是，使用三层的残差结构时，主分支的卷积核维度会发生变化，因此，shortcut 分支也要对 x 做一个维度变换，使其与主分支的维度一致才能够进行相加。实验证明，残差网络的确解决了退化的问题。

2.3.2 SEResNet（SEnet+ResNet）

SEnet^[15]是 2017 年国内汽车自动驾驶公司 Momenta 即魔门塔在 CVPR2017 上提出的，其团队提出的 SEnet 斩获了最后一届 ImageNet2017 竞赛 Image Classification 任务的冠军。SEnet 是一个轻量级的模型，能够嵌入到含有 skip-connections 的模块中。

SEnet 模块，通过显式地建模特征通道之间的相互依赖关系。即，采用一种全新的“特征重标定”策略，其本质就是在通道之间引入“Channel Attention”机制。SEnet 模块能够通

过学习的方式来自动获取到每个特征通道的重要程度，然后依照这个重要程度去提升有用的特征并抑制对当前任务影响较小的特征^[15]。SEnet 模块主要包括 Squeeze 和 Excitation 两个操作过程, 如图 2.6。

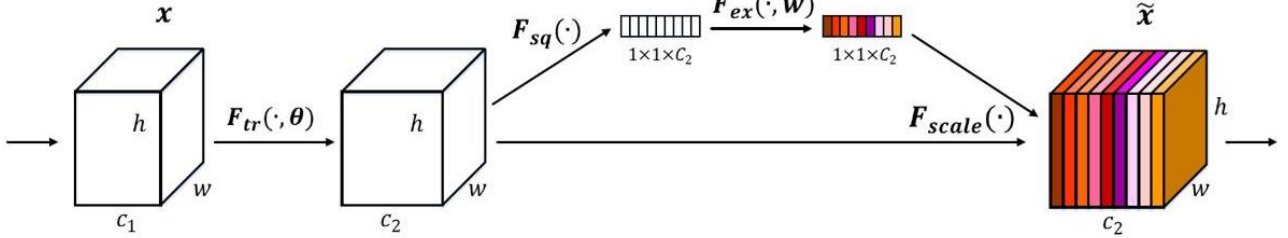


图 2.6 SENet 网络结构示意图^[15]

Fig.2.6 SENet network structure diagram^[15]

从图 2.7 可以看出，在主干网络中第一步采用全局池化（Global pooling）操作, 空间特征压缩成一系列的实数（Squeeze 操作），其后接的是一个全连接层 FC，在这个过程中一般会把输入维度降为 $1/r$ （ r 的值一般为 16），然后通过是一个 Relu 层和 FC 层，把维度恢复为输入大小（Excitation 操作），再通过一个 Sigmoid 层把权重归一化到 0~1 之间，最后通过 Scale 通道加权到先前的特征上，完成一次通道加权运算的过程。

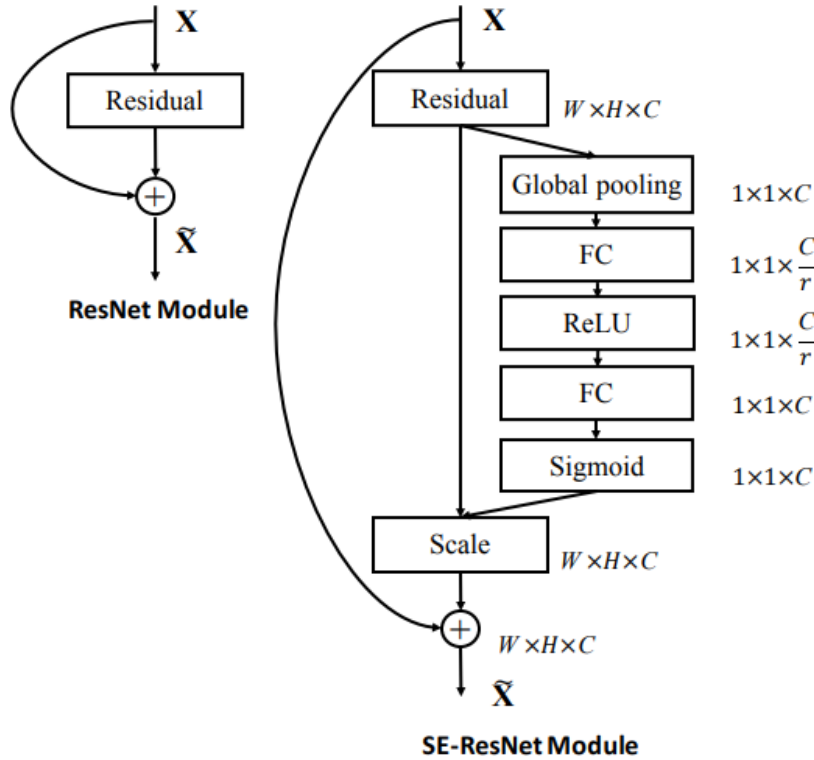


图 2.7 SENet 嵌入 Residual^[15]

Fig.2.7 SENet insert Residual^[15]

2.4 改进后的 Faster R-CNN

网上开源的深度目标检测模型直接应用于无人机多光谱影像会存在一些限制。例如，不能支持处理 TIFF 格式图像、当波段数量增多时模型容易学习到噪声导致模型分类性能下降、预测框不具有地理位置信息、识别成果绘制在影像上等问题，不利于实际生产需要。因此，针对这些不足，本文基于 Fater R-CNN 模型进行改进，从 TIFF 图像的输入、卷积层的修改、成果的输出等 3 个方面提出相应改进，构建适用于无人机多波段遥感影像的深度目标检测柚树单株识别模型。

模型的主要改进包括：针对无法读取 TIFF 图像的问题，本文使用 Python 的 GDAL 库重写数据读取函数，除了能够读取 TIFF 图像信息还能够访问 TIFF 图像的头文件获取图像的地理位置信息，为赋予预测框地理位置信息做数据准备；另外，针对融合光谱特征和植被指数特征导致信息冗余、噪声增多和模型不易收敛问题，运用 SENet 中提出的方法建模特征通道间的相关性，采用全局池化生成每个特征通道的权重，提取最重要的分类特征。最后，考虑到模型输出的预测框都是绘制在图像上，不方便进行地理分析和处理，本文使用 Python 的 GDAL 库改写了模型的预测框输出方法，将每个预测框根据地理位置信息、类别、置信度和交并比阈值（IoU）写入到 shapefile 矢量文件中，如图 2.8。

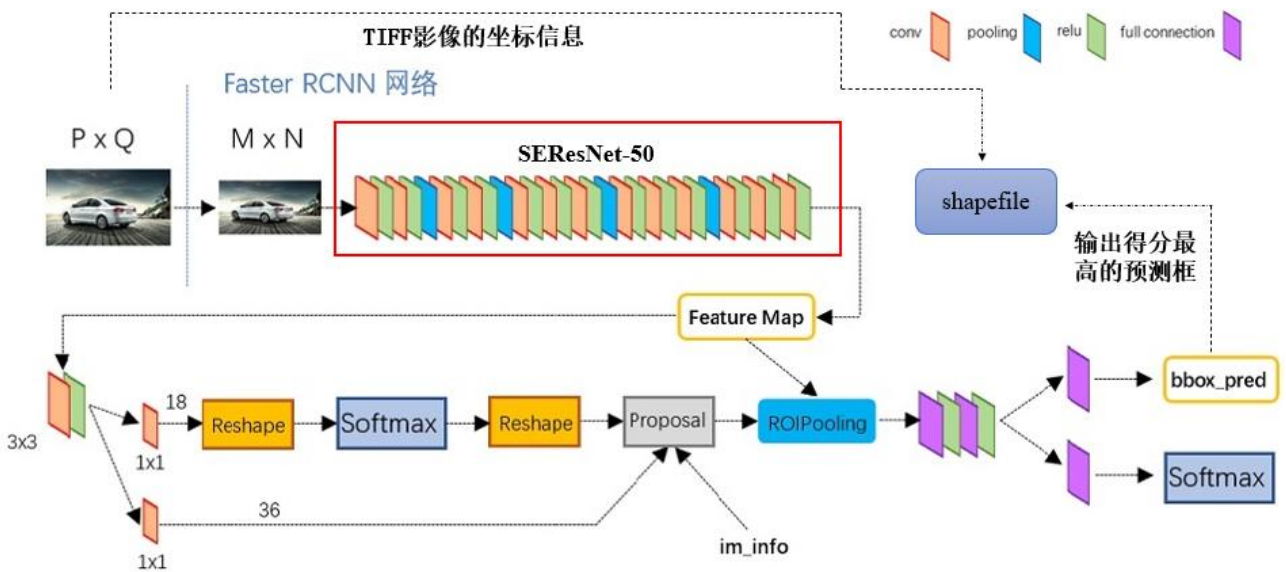


图 2.8 改进后 Faster R-CNN 模型架构，据^[15]改编

Fig.2.8 Improved Faster R-CNN model architecture, adapted from^[15]

3 研究区概况与数据集

3.1 研究区域概述

研究区位于梅州市梅县区石扇镇的金柚产区，如图 3.1，其中心点坐标为东经 $116^{\circ} 5' 40''$ ，北纬 $24^{\circ} 25' 39''$ 。该区南北长，东西宽，面积约 400 亩，海拔约 110 米，地处亚热带季风气候区，是南亚热带和中亚热带气候区的过渡地带，邻近南海，受太平洋和山地的特定地形影响，夏日长，冬日短，全年气温高，冷热悬殊、光照充足、气流闭塞、雨水丰盈而集中。梅州土壤多为红壤，PH 值处于 4.5~6.8 之间。梅州的土壤的温度、水分、PH 值均适宜梅州柚的生长发育。石扇镇作为金柚老产区，已经种植金柚 30 多年，种植经验丰富，试验区的柚树有专人管理，栽种的柚树树龄全面，南北成行，株距在 1~1.5 米，满足实验采集要求。

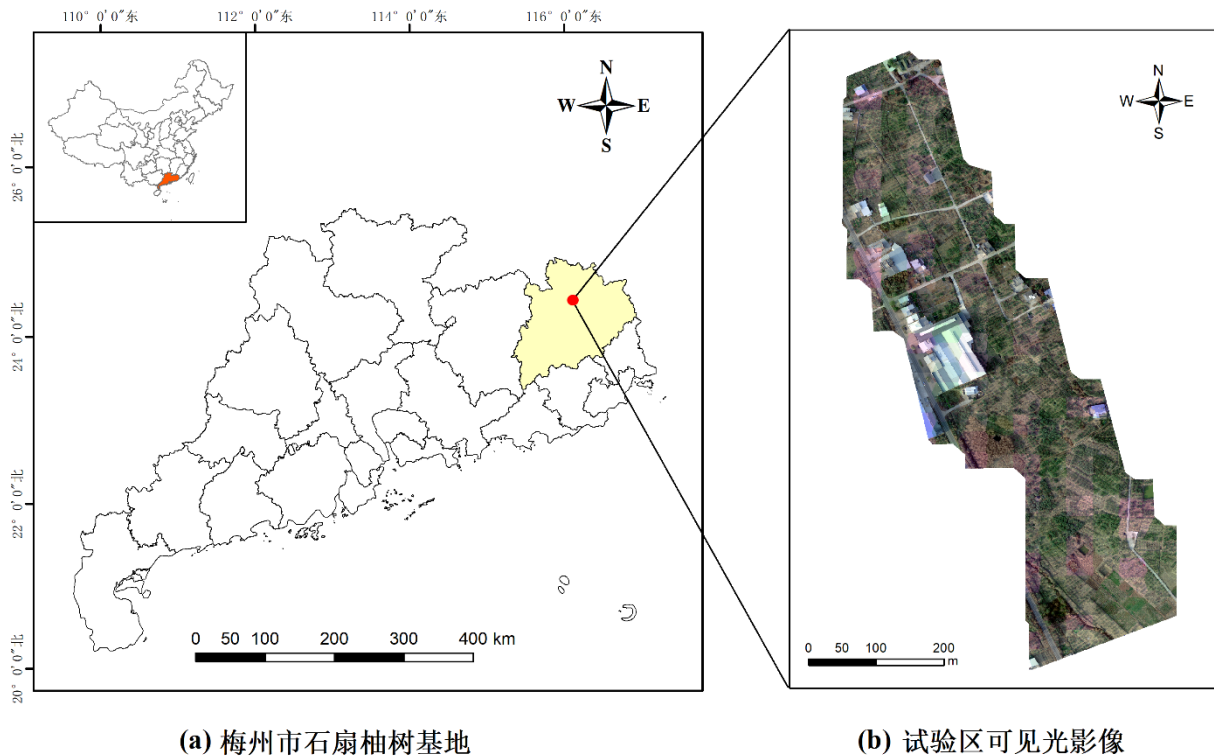


图 3.1 研究区域试验田示意图

Fig.3.1 Sketch map of the experimental field in the study area

3.2 数据采集



图 3.2 无人机及机载传感器

Fig.3.2 Drones and airborne sensors

实验采用深圳市大疆创新科技有限公司的大疆精灵 4 多光谱版旋翼无人机，如图 3.2，搭载的一体式多光谱成像系统集成成了 1 个可见光和 5 个多光谱相机（蓝、绿、红、红边及近红外）。所有相机均拥有 200 万像素解析力及配备全局快门，成像系统装置于三轴云平台上，成像清晰稳定，如表 2 所示。

表 2 大疆精灵 4 传感器参数

Tab.2 DJI Phantom 4 sensor parameters

参数	数值
影像传感器	6 个 1/2.9 英寸 CMOS；1 个用于可见光成像的彩色传感器； 5 个用于多光谱成像的单色传感器； 单个传感器：有效像素 208 万（总像素 212 万）
滤光片	蓝（B）：450 nm ± 16 nm； 绿（G）：560 nm ± 16 nm； 红（R）：650 nm ± 16 nm； 红边（RE）：730 nm ± 16 nm； 近红外（NIR）：840 nm ± 26 nm
镜头	FOV：62.7°； 焦距：5.74 mm（35 mm 格式等效：40 mm）； 无穷远固定焦距；光圈：f/2.2
彩色传感器 ISO 范围	200 – 800
单色传感器增益	1 – 8 倍
电子全局快门	1/100 – 1/20000 s（可见光成像）； 1/100 – 1/10000 s（多光谱成像）
照片最大分辨率	1600×1300（4:3.25）
照片格式	JPEG（可见光成像） + TIFF（多光谱成像）

本次实验所用影像数据于 2021 年的 2 月 3 日、3 月 12 日和 4 月 12 日对试验区采集共 3 期无人机多光谱遥感影像。多光谱遥感影像的波长范围为 450nm~856nm，包括波长为 450nm (B)、560nm (G)、650nm (R) 的 3 个可见光波段和波长为 730nm (RE)、840nm (NIR) 的两个近红外波段。在试验区内预定飞行轨迹点连续航拍，飞行高度为 110 米，航向重叠率 60%，旁向重叠率 60%，镜头垂直航拍，完整覆盖整个区域。为减少太阳高度角的影响，拍摄时间在中午 12~13 时。

实验应用大疆智图 (DJI Terra) 软件进行无人机遥感影像数据的拼接处理。主要过程为：将筛选后的单张影像图片及对应的 POS 数据导入 Terra 软件中，设置好模型清晰度及坐标系，如果有需要可以进行像控点校正。随后模型就会逐一进行点云提取和立体模型的建立，提取纹理，构造地物的特征形态；输出 1 个可见光和 5 个单波段的正射影像图 (DOM)。获得多光谱的正射影像数据后，结合实地调查和目视解译对拼接后的正射影像进行单株柚树标注，建立学习样本标签，之后进行裁剪处理并导出可见光和多光谱深度学习图像切片数据集 (通过 5 个可见光波段及可见光波段生成的植被指数波段的自由组合形成了 7 个数据集，如表 3)，如图 3.3 所示。

表 3 不同波段组合的切片数据集

Tab.3 Sliced data sets of different band combinations

数据集名称	波段选择	波段数量
RGB	R、G、B	3
R_RE_NIR	R、RE、NIR	3
RGB_RE	R、G、B、RE	4
RGB_NIR	R、G、B、NIR	4
ALL_5bands	R、G、B、RE、NIR	5
植被指数	DVI、EVI、NDVI、NDWI	4
ALL_9bands	R、G、B、RE、NIR、DVI、EVI、NDVI、NDWI	9

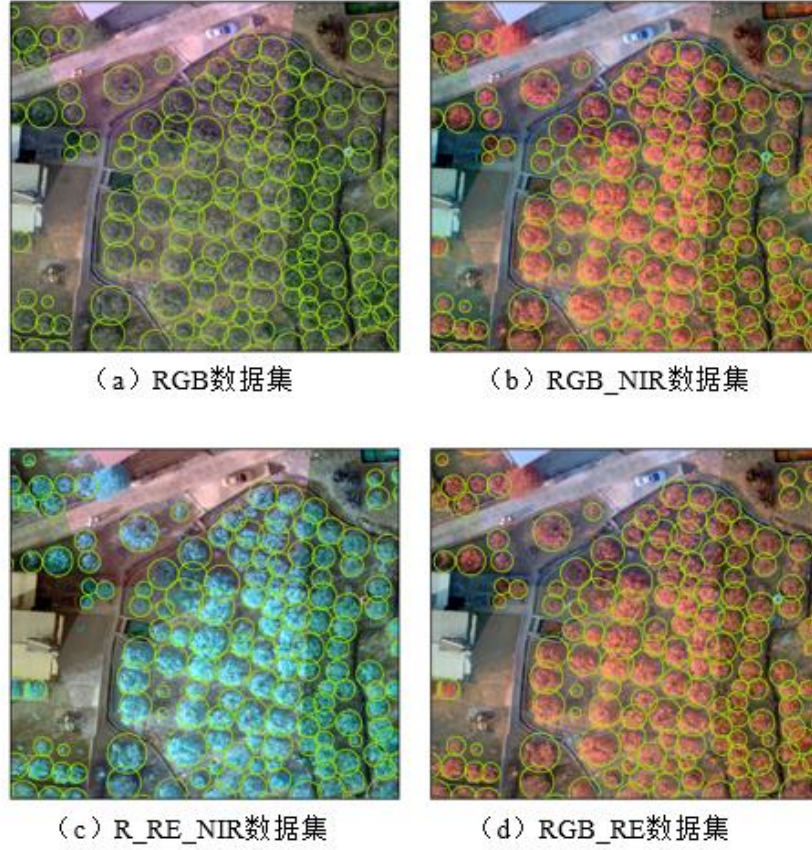


图 3.3 样本数据集示例

Fig.3.3 Sample data set example

3.3 数据预处理

在影像输入模型之前，需要对采集后进行影像裁剪、波段选择、归一化、波段合成、数据增强、数据集构建和数据集样本检查等一系列预处理操作后才能输入到网络中进行训练。本文选取 2021 年 2 月的遥感影像作为模型的训练集和验证集。利用 Python 的 `arcpy` 库编写脚本对单通道多光谱遥感影像的图像降维和归一化进行自动化处理，表达式见下式 3.1，代码实现见附录 1。归一化后，分别按照表 3 的波段组合，合成 7 份数据集。

当 $g(x) < 0$ 时，归一化使用式 3.1，

$$f(x) = \frac{255}{\max(g(x) - \min(g(x)))} \quad (3.1)$$

当 $g(x) \geq 0$ 时，归一化使用式 3.2，

$$f(x) = \frac{255}{\max(g(x))} \quad (3.2)$$

式中， $g(x)$ 表示原栅格的像素值， $f(x)$ 表示归一化后的像素值。

3.4 数据增强

深度学习图像数据集的数据增强主要有两种方式，一种是模型外的数据增强，另一种是在模型内的数据增强。模型外的数据增强可以利用 Python 的 `imgaug` 库对图像进行旋转、翻转、合成、高斯模糊、噪声增强、亮度增强等操作对原有图像进行变换，变换后图像与原数据集为进行合并，完成数据增强处理。模型内的数据增强指在模型内部图形进行标准化过程中对图像进行随机水平翻转、随机旋转、随机噪声增强、随机高斯模糊等操作，从而达到增强数据的效果。

本文结合柚树数据集的实际情况选择了较为简单的数据增强方式，模型外的增强是在裁剪数据集切片时设置 50% 的重叠率，增强四倍样本数据集的数据量；模型内的增强则是在标准化过程中添加随机水平翻转和随机亮度增强两个函数，如图 3.4 所示，代码实现见附录 2。



图 3.4 样本数据增强示例

Fig.3.4 Sample data enhancement example

3.5 数据集构建

本文采用 2021 年 2 月份影像作为深度目标检测模型的训练集和验证集，4 月份影像作为测试集，3 月份影像作为补充数据。训练集用来训练模型，验证集随训练集一起输入模型但不参与训练，用于调整模型的超参数和评估模型，测试集用以检验模型的泛化性能。根据不同的波段数量构建可见光遥感影像数据集和多光谱遥感影像数据集，并按照 3:1 将图像分为训练集和验证集，单幅图像的大小为 600*600，格式为 TIFF，训练集的图像数量为 360 幅，验证集的数量为 154 幅，共打了 4166 个柚树标签。部分样本数据集示意图如 3.5。

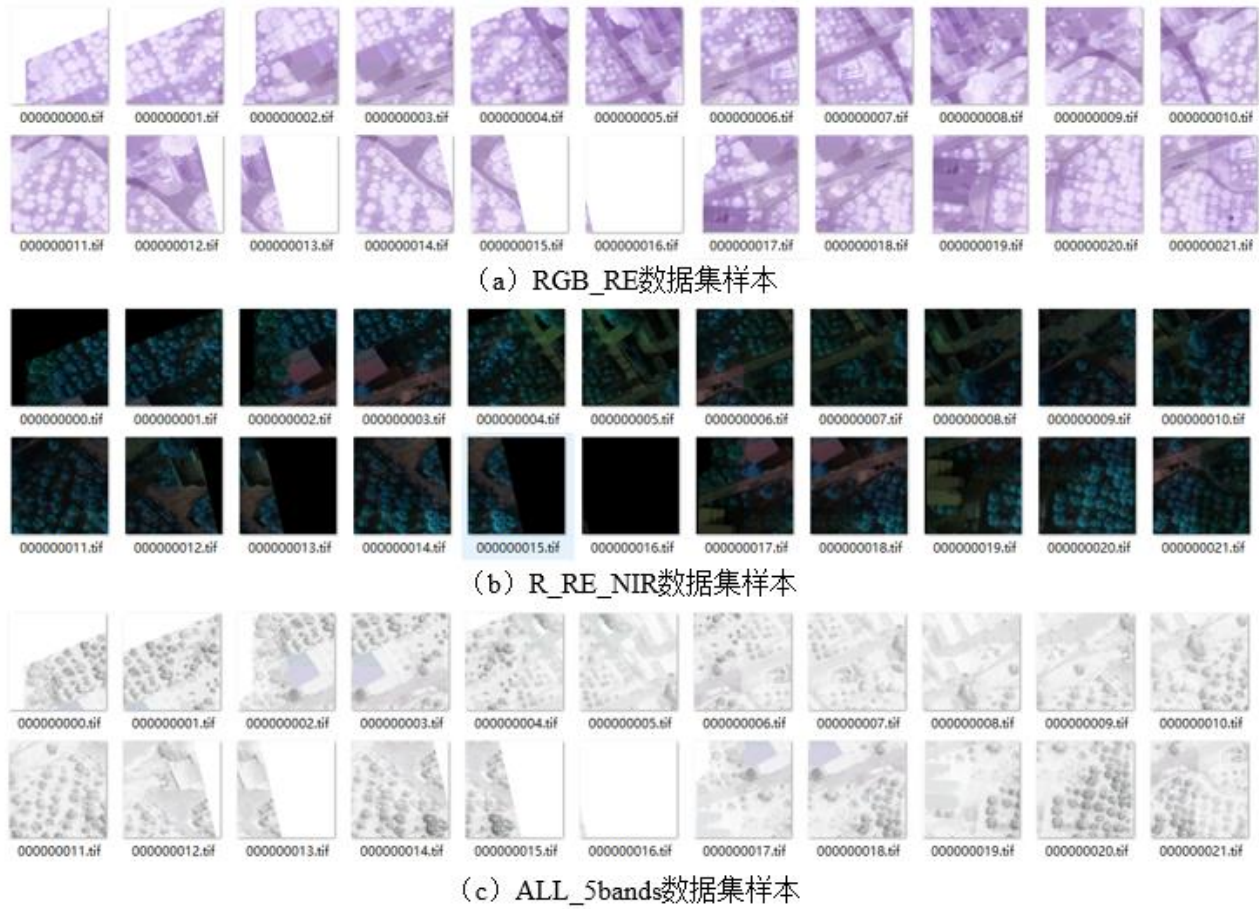


图 3.5 柚树数据集部分样本展示

Fig.3.5 Partial sample display of the pomelo data set

关于训练集和验证集的划分，常用的划分方法是随机从训练集中划分出一定比例的样本充当验证集，如图 3.6a。但由于本文的数据集是按照重叠率进行切片输出的，如果按照传统的方法进行随机划分验证集会导致同一个区域的影像既出现在训练集中，又出现在测试集中，导致出现数据泄露的情况。为了避免数据泄露，本文提出了按块划分验证集的方法，如图 3.6b。这种划分方法确保了大部分的验证集样本不会出现在训练集当中，保证训练结果的可靠性。

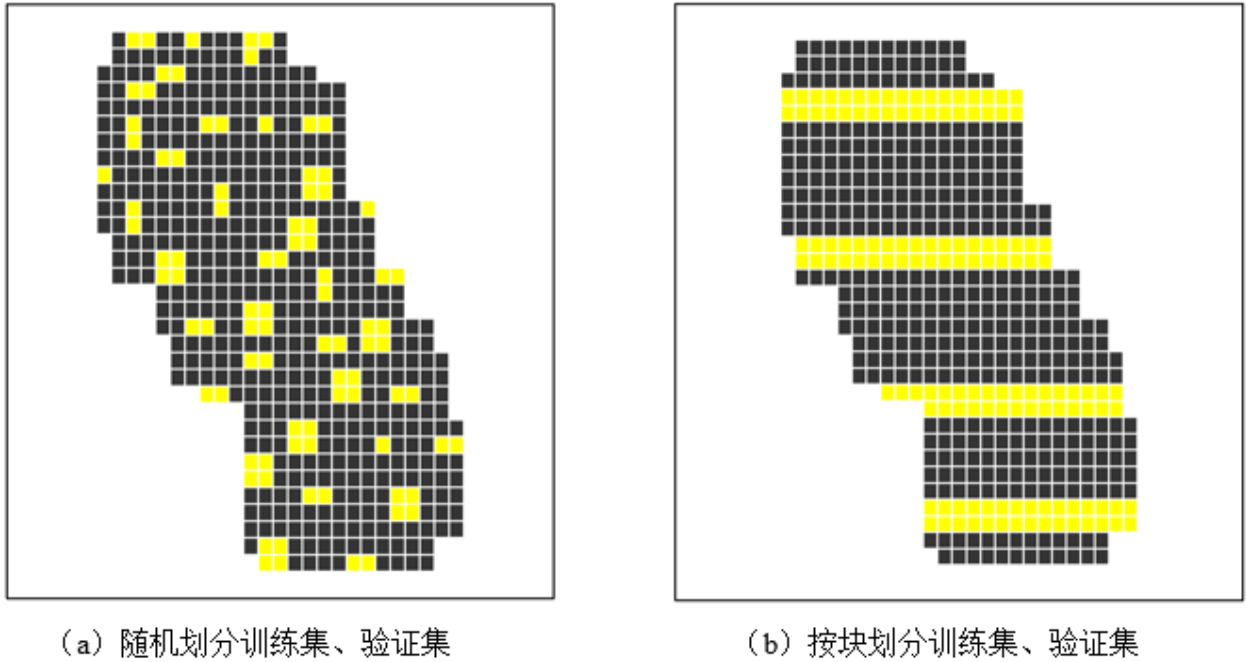


图 3.6 数据集划分方式示例

Fig.3.6 Example of how the data set is divided

3.6 识别成果后处理

遥感影像数据集的目标检测与常用数据集的目标检测不同，遥感数据集的目标检测要求识别结果含有地理位置信息，即具有空间位置连续性。但常用的目标检测针对的数据集并非遥感数据集，因此并不具有空间位置连续性，只需要准确识别目标即可。

本文的训练图像格式是 TIFF，不同于常用的 JPEG 图像，在 ArcMap 软件导出的 TIFF 格式影像切片具有地理位置信息，深度模型网络对图像生成的预测框坐标信息是相对于图像的分辨率而言的，即图像左上角为原点，横轴为图像的宽，纵轴为图像的高（本文图像的宽和高都为 600 像素值），利用 TIFF 格式图像的地理位置信息计算出图像上每一个预测框的地理位置信息，从而实现获取每一个预测框的地理位置信息。除此以外，切片影像在预测的过程中，位于图像的分割处的标签只有半个，即同一个目标被分成了多个标签。在识别柚树的过程中，为了获取到每一株柚树的完整预测框，本文选择把含有重叠度的影像让模型进行识别，这样会导致位于重叠区域的柚子树会被重复识别多次。为了解决这个问题，本文通过 Python 的 GDAL 库，编写脚本把预测框写入到 shapefile 矢量文件中，借助 Python 的 shapely 库判断两个预测框是否相交，如图 3.7 所示，代码看附录 3。通过设置交并比 (IoU) 的阈值比例（本文阈值为 0.8），删除重复识别的预测框，实现非极大值抑制 (NMS)，如图 3.8，从而使预测框在空间上具有唯一性和连续性（即一株柚树只有一个预测框），保证了预测框的数量精度，能够更加精准地统计单株柚树的数量。

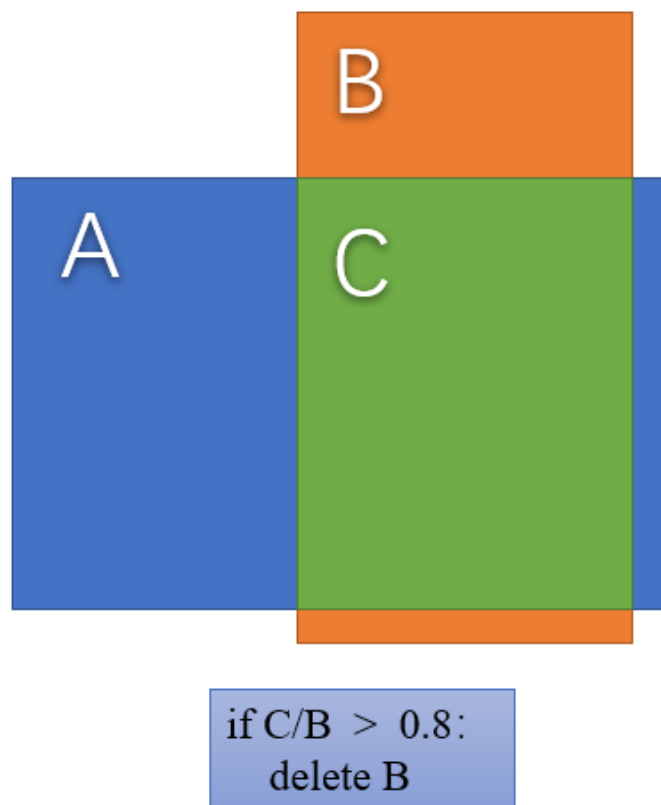


图 3.7 NMS 方法图解

Fig.3.7 Illustration of NMS method



图 3.8 NMS 处理前后对比

Fig.3.8 Comparison before and after NMS treatment

4 实验研究方案

4.1 实验方案

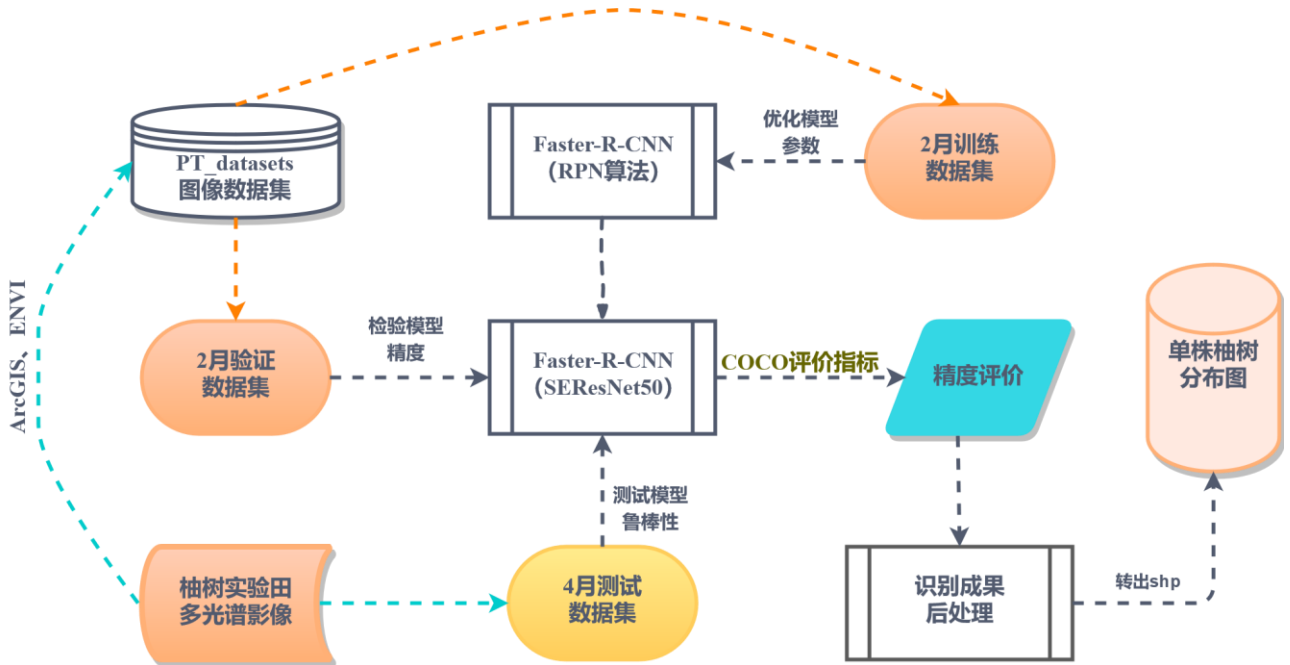


图 4.1 实验方案技术路线

Fig.4.1 Technical route of experimental scheme

本文的实验方案的技术路线如图 4.1，将试验田 2 月份影像制作训练集和验证集，4 月份影像制作测试集，训练集用于优化模型的参数，验证集用于评估模型精度和调整模型的超参数，测试集用于检验模型的泛化能力。所用的目标检测模型是 Faster R-CNN，是一种比较成熟的目标检测算法，骨干网络采用的是 SEResNet-50 替换掉原论文的 VGG16，评价指标使用 PSACAL VOC 数据集和 COCO 数据集的主要评价指标，分别是 $mAP/IoU=0.5$ 和 $mAP/IoU=0.5:0.05:0.95$ ，前者指的是预测的 bounding box 与人工标注的 Groundtruth box 的交并比达到 50% 的各类 AP 的平均值，后者指的是不同 IoU 阈值（从 0.5 到 0.95，步长 0.05）（0.5、0.55、0.6、0.65、0.7、0.75、0.8、0.85、0.9、0.95）上的各类 AP 的平均值。识别成果后处理使用的是 python 的 GDAL 库编写的自定义 NMS 方法，把预测成果写入到 shapefile 文件中，制作成单株柚树分布数据集。

为了验证提出算法的有效性，根据技术路线设置三个实验方案，分别是对比模型改造前后对模型分类精度的影响、对比光谱特征对模型分类精度的影响和对比植被指数先验知识对模型分类精度的影响。

（1）改造后的模型对模型分类精度的影响

通过改造前后模型在 RGB 数据集、RGB_NIR 数据集和 RGB_RE_NIR 数据集三份数据集上的表现，验证改造目标检测模型对识别无人机多光谱遥感影像的必要性。

(2) 不同光谱特征下模型分类精度的表现

通过控制变量法研究不同波段组合（RGB、R_RE_NIR、RGB_RE、RGB_NIR、ALL_5bands 等数据集）的多光谱遥感影像在模型上的表现

(3) 融合植被指数先验知识的影像对模型分类精度的影响

使用表 4 的遥感指数计算公式将多光谱影像运算后生成 NDVI、EVI、DVI、NDWI 四种植被指数信息，再通过波段合成，把四种植被指数信息合并成一份含有植被指数的数据集，另外再结合原来的 5 个波段，合成一份含有 9 个波段信息的融合多光谱信息和植被指数先验知识的数据集，探究二者对模型精度的影响。

表 4 遥感指数公式

Tab.4 Remote sensing index formula

指数	NDVI	EVI	DVI	NDWI
公式	$NDVI = \frac{NIR - R}{NIR + R}$	$EVI = \frac{2.5(NIR - R)}{NIR + 6R - 7.5B + 1}$	$DVI = NIR - R$	$NDWI = \frac{G - NIR}{G + NIR}$

注：NDVI 为归一化植被指数，EVI 为增强植被指数，DVI 为差值植被指数，NDWI 为归一化水指数，R 为红色通道，G 为绿色通道，B 为蓝色通道，NIR 为 840nm 波长的近红外通道。

4.2 评价准则

本文采取的评估模型的指标是 mAP/IoU=0.5 和 mAP/IoU=0.5:0.05:0.95，IoU=0.5 指当交并比等于或者大于 0.5 的预测框为正例预测框（表 5 中的 TP），mAP/IoU=0.5:0.05:0.95 是不同 IoU 阈值（从 0.5 到 0.95，步长 0.05）的正例预测框的平均值。前者是验证 PASCAL VOC 数据集分类精度的主要指标（如表 5 的 APIoU=0.50 指标），后者则是 COCO 数据集提出的评估模型分类精度的指标（如表 5 的 AP 指标），该指标更能反映出探测器的定位性能。

表 5 正负样本说明

Tab.4 Positive and negative sample description

GroundTruth（真实框）	Predict（预测情况）	
	Positive（正例）	Negative（反例）
True（真例）	TP	TN
False（假例）	FP	FN

注：TP（True Positives），意思是被正确分类的正样本；TN（True Negatives），意思是被正确分类的负样本；FP（False Positives），意思是没有被正确分类正样本；FN（False Negative），意思是没有被正确分类的负样本。

Precision 精准率（查准率）：表示正确识别物体的个数占总识别出的物体个数的比例，即预测为正例中预测正确的概率，表达式如 4.1。

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

Recall 召回率（查全率）：表示正确识别物体占测试集中物体总个数的比例，即所有正例中预测正确的概率，表达式如 4.2。

$$Recall = \frac{TP}{TP + FN} \quad (4.2)$$

Accuracy 准确率：正确分类（正例分为真例，负例分为假例）的样本数除以所有的样本数，准确率越高，正确率越高，分类器越好。表达式如 4.3。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.3)$$

交并比（Intersection-over-Union, IoU），目标检测中使用的一个概念，是产生的候选框（candidate box）与原标记框（ground truth box）的交叠率，即它们的交集与并集的比值。最理想情况是完全重叠，即比值为 1。IoU 的计算公式如 4.4，示意图如图 4.2。

$$IoU = \frac{area(C) \cap area(G)}{area(C) \cup area(G)} \quad (4.4)$$

式中，C 表示模型生成的候选框，G 表示原标记框。



图 4.2 IoU 计算示意图

Fig.4.2 IoU calculation diagram

表 5 COCO 物体检测器性能指标

Tab.5 COCO object detector performance index

类型	指标	IoU
Average Precision (AP)	AP	%AP at IoU=.5:0.05:.95 (primary challenge metric)
	AP _{IoU=0.50}	% AP at IoU=0.50 (PASCAL VOC metric)
	AP _{IoU=0.75}	% AP at IoU=0.75 (strict metric)
	AP _{small}	% AP for small objects: area < 322
AP Across Scales	AP _{medium}	% AP for medium objects: 322 < area < 962
	AP _{large}	% AP for large objects: area > 962
Average Recall (AR)	AR _{max=1}	% AR given 1 detection per image
	AR _{max=10}	% AR given 10 detections per image
	AR _{max=100}	% AR given 100 detections per image
	AR _{small}	% AR for small objects: area < 322
AR Across Scales	AR _{medium}	% AR for medium objects: 322 < area < 962
	AR _{large}	% AR for large objects: area > 962

在表 5 中，除非另有说明，否则 AP 和 AR 在多个交汇点（IoU）值上取平均值。具体来说，我们使用 10 个 IoU 阈值 0.50:0.05:0.95。这是对传统的一个突破，其中 AP 是在一个单一的 0.50 的 IoU 上计算的（这对应于我们的度量 AP_{IoU=0.5}）。超过均值的 IoUs 能让探测器更好定位（Averaging over IoUs rewards detectors with better localization.）。

在 COCO 中的 AP 是所有类别的平均值。传统上，这被称为“平均精确度”（mAP, mean average precision）。表 5 中没有区分 AP 和 mAP（同样是 AR 和 mAR），本文的柚树目标识别模型是一个二分类模型，因此 AP 和 mAP 的值是一样的。AR 是在每个图像中检测到固定数量的最大召回（recall），在类别和 IoU 上平均。AR 与提案评估（proposal evaluation）中使用的同名度量相关，但是按类别计算。所有度量标准允许每个图像（在所有类别中）最多 100 个最高得分检测进行计算。除了 IoU 计算（分别在框（box）或掩码（mask）上执行）之外，用边界框和分割掩码检测的评估度量在所有方面是相同的。

在 COCO 中，比大物体相比有更多的小物体。具体地说，大约 41% 的物体很小（面积 < 322），34% 是中等（322 < area < 962），24% 大（area > 962）。测量的面积（area）是分割掩码（segmentation mask）中的像素数量。本文结合实际情况，在标签的预处理时，删除了 area < 400 的 GroundTruth Boxes，因此并不含有 COCO 数据集定义小目标物体。

4.3 模型训练

4.3.1 训练环境说明

本文深度学习运行环境在国内的一个叫“Featurize”的租用云服务器平台，租用的服务器配置是：显卡型号：GeForce RTX 2080Ti，CPU 型号：Intel(R) Core(TM) i3-10100F CPU @ 3.60GHz，核数：8 核，显存：11GB，内存：31GB，磁盘空间：466GB，代码运行平台为 linux 系统的 Jupyter Notebook，如图 4.3。



图 4.3 租用服务器的配置参数

Fig.4.3 Configuration parameters of the leased server

4.3.2 模型超参数

本文采用控制变量法,对所有的数据集采用一样的超参数进行训练,由此探究光谱特征、植被指数特征对改进后的模型分类精度的影响。本文先通过用不同学习率策略去训练数据集,最后确定了以批次(epoch)为 300,批尺寸(batchsize)为 6,学习率(learning rate)为 0.015,学习率的步降为每 25 个批次乘 0.77 作为模型训练的超参数,最后获取相同超参数下不同数据集的分类结果。模型最终的分类结果取每个训练集的最后 10 个批次的平均值,由于模型训练到最后 10 个批次时已经收敛的比较好,验证集的损失(loss)也降到 0.03 以下,因此最后 10 个批次的平均值能反映出模型真实精度。

5 实验结果分析

5.1 实验结果

5.1.1 改造模型对分类性能的影响

通过表 7 的分类结果我们可以看出, 当通道数量不多时, 添加了“Channel Attention”机制的 SEResNet-50 网络与没有添加“Channel Attention”机制的 ResNet-50 网络的分类精度差别不大, 但当数据集增加到 5 个波段时, 没有“Channel Attention”机制的网络表现的不如只有 4 个波段的数据集。相反拥有“Channel Attention”机制的网络表现出了更优异的性能, 能够根据有效波段的增加逐步提高模型的分类精度, 原因是 SEnet 网络能够给各个波段进行加权运算, 建立通道间的相关性, 这样能有效屏蔽图像的噪音, 使模型学习到真正有效的特征。

表 7 改进前后网络上数据集的结果

Tab.7 The results of the data set on the network before and after the improvement

分类方法	通道选择	骨干网络	mAP/IoU=.5	mAP/IoU=.5:0.05:.95
Faster-R-CNN	R、G、B	ResNet-50	80.12	47.77
Faster-R-CNN	R、G、B、NIR	ResNet-50	82.99	50.36
Faster-R-CNN	R、G、B、RE、NIR	ResNet-50	81.54	50.20
改进后 Faster-R-CNN	R、G、B	SEResNet-50	80.29	47.14
改进后 Faster-R-CNN	R、G、B、NIR	SEResNet-50	82.40	49.61
改进后 Faster-R-CNN	R、G、B、RE、NIR	SEResNet-50	82.64	50.39

注: epoch:300; lr:0.015/25epoch*0.77。

5.1.2 光谱特征对分类性能的影响

相对于可见光影像, 5 波段的多光谱影像提供了更丰富的信息。除用于构建植被指数外, 近红外光谱信息对深度学习模型性能也会产生影响。在不考虑构建植被指数通道的前提下, 利用改进的模型测试不同近红外波段组合的遥感影像分类性能, 结果如表 8 所示, 其中, RE 为 730nm 波长的近红外通道, NIR 为 840nm 波长的近红外通道。通过 RGB_RE 数据集和 RGB_NIR 数据集的对比, 可以看到 RE 近红外波段比 NIR 近红外波段对于模型分类精度更明显。通过 RGB 数据集和 R_RE_NIR 数据集的分类结果对比, 可以发现, 同样只有 3 个波段, 但是后者的 mAP/IoU=0.5 和 mAP/IoU=0.5:0.05:0.95 指标分别比前者提高了 2.24%、4.21%。因此, 可认为模型在该数据集下的分类性能受到近红外波段的影响更大。

表 8 不同波段组合的分类结果

Tab.8 Classification results of experimental groups with different band combinations

数据集名称	通道选择	通道数量	骨干网络	mAP/IoU=.5	mAP/IoU=.5:0.05:.95
RGB	R、G、B	3	SEResNet-50	80.29	47.14
R_RE_NIR	R、RE、NIR	3	SEResNet-50	82.53	51.35
RGB_RE	R、G、B、RE	4	SEResNet-50	83.08	50.02
RGB_NIR	R、G、B、NIR	4	SEResNet-50	82.40	49.61
ALL_5bands	R、G、B、RE、NIR	5	SEResNet-50	82.64	50.39

注：epoch:300；lr:0.015/25epoch*0.77。

5.1.3 植被指数特征对分类性能的影响

通过表 3 的遥感指数计算公式构建 DVI, EVI, NDVI, NDWI 四个单波段影像，通过波段合成合并成一个 4 波段的遥感影像。植被指数数据集（训练结果如表 9）的训练结果要比表 7 的所有结果都略好，证明植被指数对模型的分类精度的提升是有效的。为了探究融合多光谱信息和植被指数先验信息的影像对比可见光影像对模型分类精度的影响，本文合并可见光波段、近红外波段和植被指数波段形成一个 9 波段的数据集，但该数据集在 SEResNet-50 网络的训练结果与 4 波段的植被指数数据集精度差别不大。考虑到可能是波段特征加深，但分类网络层次的深度不够，学习不到有效特征，因此使用更深层次的分类网络 SEResNet-101 去再次训练，训练结果的 mAP/IoU=0.5 和 mAP/IoU=0.5:0.05:0.95 指标分别提高了 0.2%、1.54%，再次验证了融合多光谱信息和植被指数先验信息的图像对模型精度的提升是有效的。

表 9 植被指数与多光谱组合的分类结果

Tab.9 Classification results of experimental groups with different band combinations

数据集名称	通道选择	通道数量	骨干网络	mAP/IoU=.5	mAP/IoU=.5:0.05:.95
植被指数	DVI、EVI、NDVI、NDWI	4	SEResNet-50	82.54	52.4
ALL_9bands	R、G、B、RE、NIR、DVI、EVI、NDVI、NDWI	9	SEResNet-50	82.99	51.58
ALL_9bands	R、G、B、RE、NIR、DVI、EVI、NDVI、NDWI	9	SEResNet-101	83.19	53.12

注：epoch:300；lr:0.015/25epoch*0.77。

5.2 测试结果

5.2.1 测试精度

本文使用 4 月份影像数据制作模型的测试集, 验证模型的泛化能力。使用表 7-9 训练出来的权重文件对 4 月份的测试集进行测试, 所得结果如表 10, 实验结果表明, 模型的泛化能力较差, 不足以用于生产。另用 3 月份影像补做了一份 RGB 测试集给模型测试, 3 月份影像的 RGB 测试集的 $mAP/IoU=0.5$ 和 $mAP/IoU=0.5:0.05:0.95$ 指标分别是 62.3 和 40.5, 模型在 3 月份的影像上泛化能力比 4 月份更好。预测结果输入到 shapefile 中, 通过 NMS 处理后如图 5.1b 所示, 人工绘制的标签如图 5.1a, 通过两者对比, 判断测试结果良好。

表 10 测试集分类结果 (4 月份影像)

Tab.10 The result of text set (April images)

数据集名称	通道选择	骨干网络	权重精度 $mAP/IoU=.5$	$mAP/IoU=.5$	$mAP/IoU=.5:0.05:.95$
RGB	R、G、B	SEResNet-101	81.1	40.1	20.5
RGB_RE	R、G、B、RE	SEResNet-50	83.3	20.5	12.5
RGB_NIR	R、G、B、NIR	SEResNet-50	83.3	22.6	10.8
ALL_5bands	R、G、B、RE、NIR	SEResNet-50	83.4	17.8	10.2
植被指数	DVI、EVI、NDVI、NDWI	SEResNet-50	83.2	32.4	18.8
ALL_9bands	R、G、B、RE、NIR、 DVI、EVI、NDVI、NDWI	SEResNet-101	83.9	33.5	20.9



(a) 3 月份影像标签数据



(b) 3 月份影像识别结果

图 5.1 模型预测结果部分样本

Fig.5.1 Partial sample of model prediction results

5.2.2 单株柚树分布图（3 月份影像测试集）

选取泛化能力最好的 3 月份 RGB 影像给模型进行预测，测试结果经过非极大值抑制后处理，识别结果制图如图 5.2，处理后预测框数量为 4751 个，3 月份影像的标签数据为 6356 个。3 月份 RGB 测试集的测试精度为 62.3%，说明正确预测框个数应为 3960，通过计算，判断处理后的预测框还存在约 790 个被误识别的预测框。

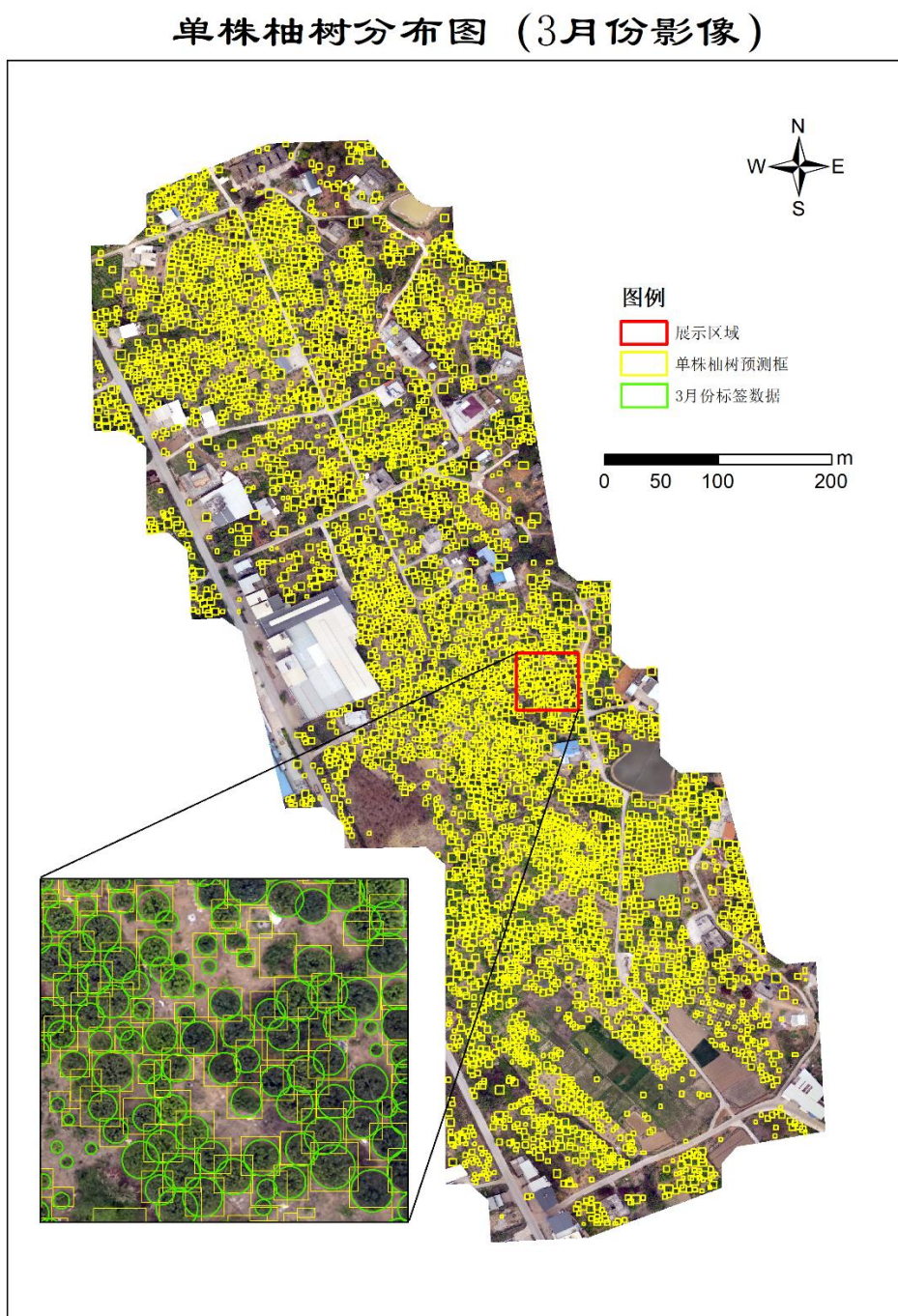


图 5.2 3 月份影像预测结果图示

Fig.5.2 Image prediction result icon for March

6 结论与讨论

6.1 结论

针对野外包含柚树的无人机多光谱遥感影像，将深度目标检测模型应用于野外柚树分类。利用卷积神经网络端到端的强大的特征学习能力，基于 Faster R-CNN 模型，通过数据预处理获取含有多光谱信息和植被指数先验信息的影像，并修改骨干网络，添加“Channel Attention”机制。为了控制变量对比可见光影像和多光谱影像对模型分类精度的影响，本文用 2 月份影像的 RGB 数据集，使用 SEResNet-101 补充做了一个实验（附录 4 含实验结果），实验结果的 $mAP/IoU=0.5$ 和 $mAP/IoU=0.5:0.05:0.95$ 指标分别是 79.82%、48.60%，与同样使用 SEResNet-101 网络的 ALL_9bands 的数据集的分类精度相差 3.37、4.52 个百分点，差距明显。实验结果表明在相同的模型参数下，融合了多光谱信息和植被指数先验信息的遥感影像对模型分类精度提升明显。

6.2 讨论

6.2.1 不足与改进

（1）模型细节没有做好，没有将植被指数通道的生成模块和识别成果后处理结合进模型里，因为 EVI 植被指数的公式计算比较复杂，应编写专门的植被指数生成模块，在模型中自动生成植被指数通道。

（2）识别成果后处理，本文使用自定义的 NMS 方法来删除无效的预测框，但实际上切片影像的交界处预测框情况复杂，单靠简单的重叠面积阈值来判断和删除预测框会导致出现正确的预测框被误删除，也无法删除所有的无效预测框。

（3）模型泛化能力不足，在表 10 中可以发现模型在波段数量越多的影像的泛化能力反而越差，这不符合逻辑，本文并没有针对这一实验结果进行深入研究。

6.2.2 展望

本文通过实验，已基本实现利用无人机多光谱遥感影像对柚树的单株识别，为实现柚树的精细化管理奠定了基础。但模型在测试集上的表现并不好，泛化能力较差。目前来说，要增强模型的泛化能力的主要途径还是通过增加更多的柚树样本量来实现。

除了柚树的单株识别之外，要实现对单株柚树的长势监测、面积调查及产值估算仍需要结合更多技术。柚树的长势监测需要时段性的获取单株柚树的高光谱波段信息，这需要搭载高光谱传感器才可以实现，在本文中的多光谱无人机不足以获取到所需要的波段信息，或是

通过结合高分 2 号卫星的高光谱数据，利用柚树成长过程中光谱信息的变化从而判断柚树的长势是否健康。面积调查及产值估算则使用深度学习的语义分割模型，该模型能够精准地识别单株柚树在影像上的轮廓，借此可以更精准地统计柚树面积，为之后的估产奠定数据基础。如今深度学习发展的愈发火热，深度学习技术早已应用到各行各业当中，将深度学习技术应用于地理信息的遥感影像分类将会是地理信息行业中的一个重要课题。

附录

附录 1 单波段影像归一化代码

```
import arcpy
import os
from arcpy.sa import *

def tif_tranf_0_255(tif_dir,save_dir):
    for tif in os.listdir(tif_dir):
        if tif.endswith(".tif"):
            tif_path = os.path.join(tif_dir, tif)
            arcpy.CalculateStatistics_management(tif_path, "", "", "", "")
            Max_result = arcpy.GetRasterProperties_management(tif_path, "MAXIMUM", 'Band_1')
            Min_result = arcpy.GetRasterProperties_management(tif_path, "MINIMUM", 'Band_1')
            Max = Max_result.getOutput(0)
            Min = Min_result.getOutput(0)
            if Min < 0:
                out1 = Plus(tif_path, Min)
                out2 = Times(out1, 255 / float(Max))
                outInt = Int(out2)
                output_save = os.path.join(save_dir, tif)
                outInt.save(output_save)
                print output_save
            else:
                out2 = Times(tif_path, 255 / float(Max))
                outInt = Int(out2)
                output_save = os.path.join(save_dir, tif)
                outInt.save(output_save)
                print output_save
```

附录 2 随机水平翻转、随机亮度增强代码

```
class RandomHorizontalFlip(object):
    # 随机水平翻转图像及 bboxes
    def __init__(self, prob=0.5):
```



```

self.prob = prob
def __call__(self, image, target):
    if random.random() < self.prob:
        height, width = image.shape[-2:]
        image = image.flip(-1)
        bbox = target["boxes"]
        bbox[:, [0, 2]] = width - bbox[:, [2, 0]]
        target["boxes"] = bbox
    return image, target

```

```

class Random_Brightness(object):
    # 随机提升图像亮度
    def __init__(self, prob=0.5):
        self.prob = prob
    def __call__(self, image, target):
        if random.random() < self.prob:
            image = image * 1.2
        return image, target

```

附录 3 判断相交图形重叠部分是否超过阈值代码

```

from shapely.geometry import Polygon
def intersects(a_box, b_box, thresh=0.65):
    # box 为一个包含图形坐标值的列表 (list)
    object_a = Polygon(a_box)
    object_b = Polygon(b_box)
    # 判断是否相交
    if object_a.intersects(object_b):
        diff = object_a.intersection(object_b).area
        # 判断是否为本身，若是跳过
        if a_box == b_box:
            return False
        else:
            # 判断重叠度是否超过预设阈值，超过则返回标记
            if diff / object_b.area > thresh:

```

```

        return True
    else:
        return False

else:
    return False

```

附录 4 各个数据集训练结果（取最后 10 个 epoch）

数据集	RGB			
骨干网络	ResNet-50		SEResNet-50	
评价精度	mAP/IoU=.5:0.05:.95	mAP/IoU=.5	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	0.4933	0.8088	0.4815	0.8036
epoch292	0.5041	0.8107	0.4814	0.8189
epoch293	0.4842	0.808	0.4716	0.8025
epoch294	0.438	0.7688	0.4923	0.8069
epoch295	0.4597	0.7938	0.4565	0.7893
epoch296	0.4605	0.7953	0.4674	0.8014
epoch297	0.495	0.8098	0.4892	0.8159
epoch298	0.4927	0.8108	0.4568	0.7983
epoch299	0.5045	0.8223	0.4596	0.8003
epoch300	0.4449	0.7836	0.458	0.7923
平均值	0.47769	0.80119	0.47143	0.80294

数据集	RGB			
骨干网络	ResNet-101		SEResNet-101	
评价精度	mAP/IoU=.5:0.05:.95	mAP/IoU=.5	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	0.4881	0.7949	0.506	0.8048
epoch292	0.5091	0.8116	0.459	0.7856
epoch293	0.4681	0.7998	0.4938	0.8019
epoch294	0.4923	0.8064	0.5041	0.8052
epoch295	0.4693	0.793	0.5036	0.8009
epoch296	0.4161	0.7556	0.4828	0.7984
epoch297	0.4914	0.8029	0.4961	0.8042
epoch298	0.5017	0.8044	0.4524	0.7835
epoch299	0.4911	0.8004	0.4618	0.7859
epoch300	0.4677	0.7992	0.5006	0.8111
平均值	0.47949	0.79682	0.48602	0.79815

数据集	RGBNIR			
骨干网络	ResNet-50		SEResNet-50	
评价精度	mAP/IoU=.5:0.05:.95	mAP/IoU=.5	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	0.5103	0.8303	0.4992	0.8254
epoch292	0.5046	0.8298	0.5015	0.8244

epoch293	0.5008	0.8284	0.5107	0.8283
epoch294	0.5078	0.8302	0.5054	0.8275
epoch295	0.5044	0.8286	0.4804	0.8204
epoch296	0.5057	0.8363	0.4983	0.8237
epoch297	0.4946	0.8277	0.493	0.8235
epoch298	0.4862	0.8264	0.4958	0.8187
epoch299	0.5136	0.8304	0.4958	0.8239
epoch300	0.5081	0.8305	0.4805	0.8243
平均值	0.50361	0.82986	0.49606	0.82401

数据集		RGBRENIR		
骨干网络	ResNet-50		SEResNet-50	
评价精度	mAP/IoU=.5:0.05:.95	mAP/IoU=.5	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	0.5075	0.8227	0.4915	0.8193
epoch292	0.5016	0.8135	0.5046	0.8268
epoch293	0.4935	0.8124	0.4949	0.8211
epoch294	0.5055	0.8231	0.5034	0.8277
epoch295	0.5134	0.8169	0.5149	0.8286
epoch296	0.4991	0.8137	0.5104	0.8274
epoch297	0.4875	0.8086	0.5075	0.8343
epoch298	0.4975	0.8134	0.5097	0.828
epoch299	0.5013	0.8141	0.5077	0.8283
epoch300	0.5135	0.8154	0.4945	0.8222
平均值	0.50204	0.81538	0.50391	0.82637

数据集		RRENIR		
骨干网络	/		SEResNet-50	
评价精度	/	/	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	/	/	0.5207	0.8229
epoch292	/	/	0.5084	0.8292
epoch293	/	/	0.5114	0.8263
epoch294	/	/	0.5137	0.828
epoch295	/	/	0.5133	0.8277
epoch296	/	/	0.5147	0.8223
epoch297	/	/	0.5123	0.8195
epoch298	/	/	0.5086	0.8261
epoch299	/	/	0.5144	0.8291
epoch300	/	/	0.5172	0.8222
平均值	/	/	0.51347	0.82533

数据集		RGBRE		
骨干网络	/		SEResNet-50	
评价精度	/	/	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	/	/	0.4821	0.8236

epoch292	/	/	0.4912	0.8324
epoch293	/	/	0.5155	0.8305
epoch294	/	/	0.4803	0.8242
epoch295	/	/	0.4979	0.8318
epoch296	/	/	0.5192	0.8387
epoch297	/	/	0.4991	0.8321
epoch298	/	/	0.4953	0.8325
epoch299	/	/	0.5033	0.8332
epoch300	/	/	0.518	0.8293
平均值	/	/	0.50019	0.83083

数据集	DVI_EVI_NDVI_NDWI			
骨干网络	SEResNet-50			
评价精度	/	/	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	/	/	0.5255	0.8245
epoch292	/	/	0.5207	0.8239
epoch293	/	/	0.5251	0.8322
epoch294	/	/	0.5241	0.8244
epoch295	/	/	0.5238	0.8254
epoch296	/	/	0.5236	0.825
epoch297	/	/	0.5245	0.8252
epoch298	/	/	0.5219	0.8248
epoch299	/	/	0.5206	0.8239
epoch300	/	/	0.5262	0.8242
平均值	/	/	0.5236	0.82535

数据集	ALL_9bands			
骨干网络	SEResNet-50		SEResNet-101	
评价精度	mAP/IoU=.5:0.05:.95	mAP/IoU=.5	mAP/IoU=.5:0.05:.95	mAP/IoU=.5
epoch291	0.5111	0.8287	0.5335	0.834
epoch292	0.5183	0.8295	0.5319	0.8345
epoch293	0.515	0.8291	0.5287	0.8267
epoch294	0.5119	0.8284	0.5338	0.8346
epoch295	0.5155	0.8356	0.5295	0.8257
epoch296	0.5186	0.828	0.5321	0.8347
epoch297	0.5157	0.8287	0.5307	0.8336
epoch298	0.516	0.8263	0.5276	0.826
epoch299	0.5153	0.8361	0.5318	0.8346
epoch300	0.5203	0.829	0.5325	0.8349
平均值	0.51577	0.82994	0.53121	0.83193

参考文献

- [1] 毛凤麟. 农产品协会主导的区域品牌发展模式研究[D].广州.华南农业大学, 2018.
- [2] 查志华, 周文静, 吴杰. 基于迁移学习 Faster R-CNN 模型田间红提葡萄果穗的识别[J/OL].石河子大学学报(自然科学版)
- [3] 刘康, 周壮, 李盛阳, 等.2020.天宫一号高光谱遥感场景分类数据集及应用. 遥感学报, 24(9): 1077-1087
- [4] 杨蜀秦, 宋志双, 尹瀚平, 等. 基于深度语义分割的无人机多光谱遥感作物分类方法. 农业机械学报
- [5] 韩文霆, 郭聪聪, 张立元, 等. 基于无人机遥感的灌区土地利用与覆被分类方法[J].农业机械学报, 2016,47(11):270-277.
- [6] 戴建国, 张国顺, 郭鹏, 等. 基于无人机遥感可见光影像的北疆主要农作物分类方法阴. 农业工程学报, 2018, 34(18): 122-129.
- [7] WU M, YANG C, SONG X, et al. Evaluation of orthomosaics and digital surface models derived from aerial imagery for crop type mapping[J]. Remote Sensing, 2017, 9(3): 239-243.
- [8] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[J]. IEEE Computer Society, 2013.
- [9] Girshick R. Fast R-CNN[J]. Computer Science, 2015.
- [10] Ren Shaoqing, He Kaiming, Girshick Ross, Sun Jian. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6).
- [11] 张德钧. 基于 Faster R-CNN 和多光谱图像的漂浮 HNS 检测与分类研究[D].杭州.浙江大学, 2019.
- [12] 祝思君. 基于深度学习的无人机遥感图像目标识别方法研究[D]. 北京. 北京建筑大学, 2018.
- [13] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [14] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [15] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).

致谢

感谢嘉应学院，感谢地理学院，感谢地理信息科学实验室，感谢大学期间遇到的每一个人，你们在我的人生中，留下了浓墨重彩的一笔，一同帮我描绘了在嘉大成长的故事。

感谢指导老师熊永柱老师的悉心教导，在我还对自己专业方向感到迷茫的时候，引导我参加了基于深度学习识别柚树的应用研究，让我度过了愉快又充实，同时又富有挑战性的大学时光。在这个过程中，我不断学习新的技术并享受着新技术给我带来的奇妙体验。

感谢实验室的陈志云老师、陈燕奎老师及实验室的小伙伴们。志云老师和燕奎老师做事负责、认真、严谨，对待专业的热情的优秀品质让我收益颇丰，让我不仅在专业上获益良多，也能领悟到生活的道理。在实验室团队四年里，感谢实验室优秀的师兄师姐，他们都是我在专业方向上的引路人。

在完成毕业论文的这段时光里，曾晓峰、李雁欢、吴泽欣和伍丽君同学给予我很大的帮助，不仅给我提出了许多宝贵的意见而且还帮我解决了许多论文上遇到的问题。特别感谢曾晓峰同学，由于改动深度学习框架需要有一定的开发知识，而我的开发是个半吊子，很多时候都要请教曾晓峰同学。

感谢地理信息科学的郑春燕、胡华科、张鹏飞、钟广锐、谢作轮、张侃、朱长柏、张豫等专业老师们，你们的辛勤工作与付出不仅为我们专业创造了良好的学习氛围，还为学生树立了很好的榜样。

感谢校重点领域专项（新一代信息技术）项目：基于无人机高分遥感与深度学习的梅州柚单株自动识别方法研究（编号：2020ZDZX3044）

感谢学院 2020 年广东省科技创新战略专项资金（“攀登计划”专项资金）项目：基于高分遥感和深度学习的梅州金柚单株识别、面积调查及健康评估（编号：pdjh2020b0552）