



重庆大学

## 研究生课程考核试卷

(适用于课程论文、提交报告)

科 目: 人工智能与知识工程 教 师: 叶俊勇

姓 名: 吕容飞 学 号: 201808021013

专 业: 仪器科学与技术 类 别: 学术

上课时间: 2019 年 4 月至 2019 年 6 月

考 生 成 绩:

课程综合成绩

阅卷评语: \_\_\_\_\_

\_\_\_\_\_

阅卷教师 (签名) \_\_\_\_\_

重庆大学研究生院制



# 目录

1. 引言.....	1
2. 技术方案.....	3
2.1 总体流程.....	3
2.2 数据集可视化及预处理.....	4
2.3 模型构建和网络训练.....	5
2.4 预测和保存结果.....	7
3. 实验.....	8
3.1 实施细节.....	8
3.2 结果与分析.....	9
4. 总结.....	10
5. 参考文献.....	11



# 2019 年县域农业大脑 AI 挑战赛

## 1. 引言

我国地区区域差异大、种植结构复杂、地块破碎严重，因此在使用对地观测卫星遥感数据获取大尺度农作物数据的基础上，无人机遥感测量技术作为空间信息技术的重要组成部分，既能作为星载遥感影像的重要补充，又能有效替代人工实地调查，凭借着降低地面人工调查强度和调查成本、快速获取实时高分辨数据的优势，成为农业统计调查工作中的一大创新点，同时也是精准农业的重要方向之一。本次大赛赛题聚焦无人机遥感影像分析技术和产能预测模型构建具有非常大的现实意义和应用价值。农作物的资产盘点与精准产量预测是实现农业精细化管理的核心环节。

当前，我国正处于传统农业向现代农业的加速转型期，伴随着农业的转型升级，政府宏观决策、社会各界对农业数据的需求不断增加，现有农业统计信息的时效性与质量，已不足以为市场各主体的有效决策提供科学依据。在农作物资产盘点方面，传统的人工实地调查的方式速度慢、劳动强度大，数据采集质量受主观因素影响大，统计数据有较大的滞后性，亟待探索研究更高效准确度更高的农业调查统计技术。在产量预测方面，及时准确地获取区域作物单产及其空间分布信息，对作物进行精准的产能预测，对于农业生产安全预警、农产品贸易流通，以及农业产业结构优化具有重要意义<sup>[1]</sup>。

本次大赛，主办方选择了具有独特的地理环境、气候条件以及人文特色的贵州省兴仁市作为研究区域，聚焦当地的特色优势产业和支柱产业——薏仁米产业，以薏仁米作物识别以及产量预测为比赛命题，要求选手开发算法模型，通过无人机航拍的地面影像，探索作物分类的精准算法，识别薏仁米、玉米、烤烟三大作物类型，提升作物识别的准确度，降低对人工实地勘察的依赖，提升农业资产盘点效率，并结合产量标注数据预测当年的薏仁米产量，提升农业精准管理能力。其空间分布信息，对作物进行精准的产能预测，对于农业生产安全预警、农产品贸易流通，以及农业产业结构优化具有重要意义。

官方提供的原始数据为 3 通道自然图像，标签为 8 为单通道图像，每个像素点值表示原始图片中对应位置所属类别，其中“烤烟”像素值 1，“玉米”像素值

2, “薏仁米” 像素值 3, 其余所有位置视为 “其他” 像素值 0。下图 1 为官方提供的两张训练集大图以及对应的标签 overlap 到原图上的结果, 分辨率大小分别为 47161x50141, 77470x46050, 其中空白部分为官方做的屏蔽处理。

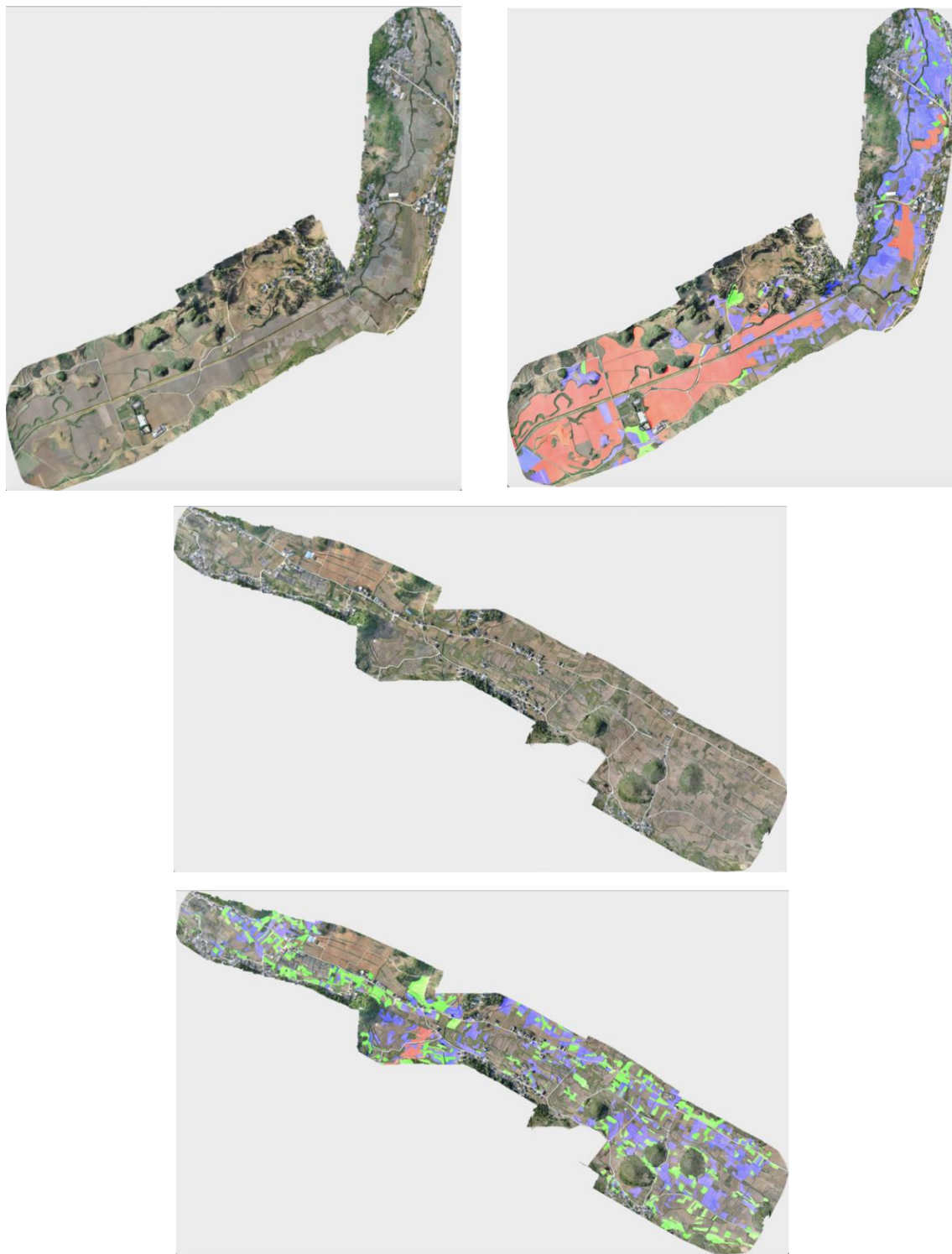


图 1 训练集概览

本次比赛采用平均交并比 (Mean Intersection over Union) 作为评价标准, 即求

出每一类的 IOU 取平均值。IOU 指的是，真实标签和预测结果的两块区域交集/并集。评估只考虑“烤烟”，“玉米”，“薏仁米”三种作物。针对每种作物所有的预测结果，统计每个类别的真实标签和预测结果，根据 Jaccard Index 计算 IOU，最后取平均。

具体的，针对 3 种作物每张图片首先计算 TP, FP, and FN；求和得到所有图片的 TP, FP, and FN；根据  $\text{Jaccard Index} = \text{TP}/(\text{TP}+\text{FP}+\text{FN})$ ，计算得到 IOU。最后对所有 3 个类别的 IOU 取平均，得到最后的 MIOU 作为评测结果。

## 2. 技术方案

### 2.1 总体流程

由于这次比赛给的遥感图像很大，分辨率大约有 50000x50000，这么大的图肯定不能直接投入网络训练。所以在简单的进行数据可视化之后，我们需要对图像进行切割。通常切割后的大小为 512x512，或者 1024x1024。为了使得图像内包含地物类别尽量多，我们选择切割为 1500x1500，并在之后进行了数据增强，弱缩放为 1024x1024，最后归一化。深度学习模型选择了在各大数据集表现优异的 U-Net<sup>[2]</sup>，之后进行训练，最后选取验证集表现最好的一次模型进行预测并提交结果。下图 2 为本次遥感图像语义分割比赛的总体流程图。

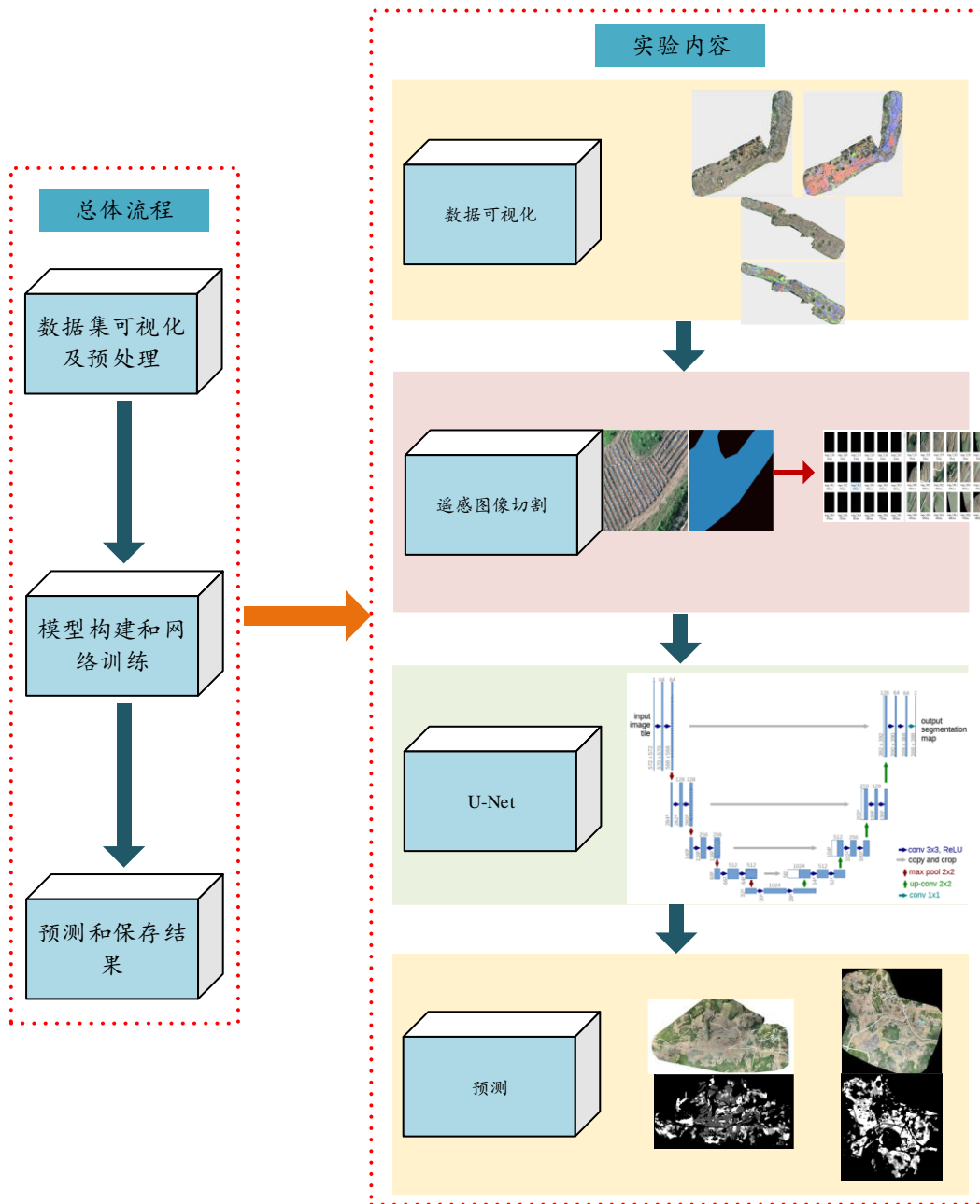


图 2 总体流程图

## 2.2 数据集可视化及预处理

图像是 4 个 channel，前三个是 RGB，第四个是 alpha 通道（透明）丢掉。通过观察图 1 可视化的结果，我们发现图像的区域很多是空白的，所以需要将这部分图像滤掉，不作处理。然后我们切割图像的时候需要有冗余，大小不能按照完全的 512, 1024 等切割，这样可以使得边缘部分能得到较好的分割效果。并且我们切割的尺度要略大于最后送入网络的图像尺度，因为数据在后期需要增强：弱缩放，旋转等。这里我们采用的切割尺度为 1500x1500，采用 1024x1024 作为训练的输入



图像大小。切图时我们采用了切割阈值 0.7，即切割遍历如屏蔽区域占比超过 0.7，我们就将其切割的图像丢弃，不作为训练图像。下图 3 为我们最终切割好的图像概览。由于切割好的标签图为单通道灰度图，其中“烤烟”像素值 1，“玉米”像素值 2，“薏仁米”像素值 3，其余所有位置视为“其他”像素值 0，故人眼无法分辨，导致看上去全为黑色，实际并不是，而是已经做好的原图像对应的 label。切割后总共有 2217 张图片，均保存为 bmp 格式。

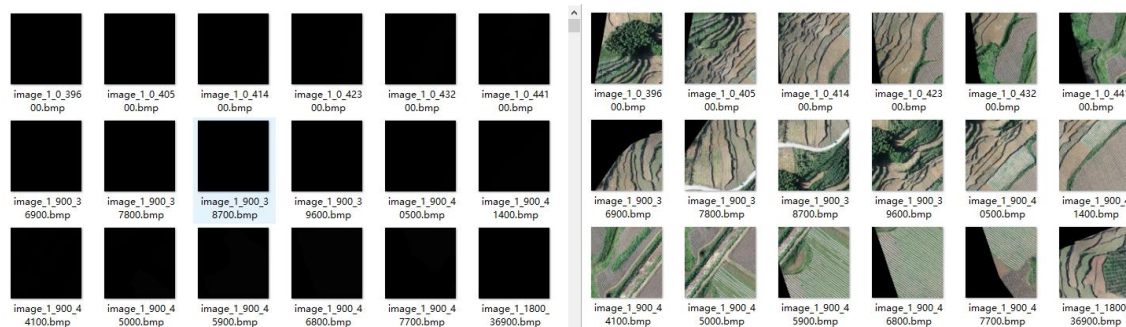


图 3 切割好的图像

之后我们进行了数据增强：原图和 label 图都需要旋转：90 度，180 度，270 度；沿 y 轴的镜像操作；模糊操作；原图做光照调整操作等<sup>[3]</sup>。这样我们的训练数据就得到了充分的扩充，使得我们训练的模型更加具有鲁棒性。

## 2.3 模型构建和网络训练

深度卷积神经网络通过组合低层特征形成更加抽象的高层表示、属性类别或特征，给出数据的分层特征表示，非线性操作的层级数多。浅层学习依靠人工经验抽取样本特征，网络模型学习后获得的是没有层次结构的单层特征；而深度卷积神经网络通过对原始信号进行逐层特征变换，将样本在原空间的特征表示变换到新的特征空间，自动地学习得到层次化的特征表示，从而更有利于分类或特征的可视化。用于分类的深度卷积神经网络通常包含卷积层、池化层、全连接层，其中卷积层与下采样层组合成卷积组，逐层提取特征，最终通过若干个全连接层完整分类识别。可以看出，这样的网络结构尽管能够取得非常高的识别准确率，却抛弃了图像的空间结构特征，能够知道图像中包含那些目标，却不能知道目标的位置、大小、形态等信息，即丢掉了图像语义<sup>[4]</sup>。

图像语义分割需要一种端到端的网络结构，以保证输出图像应与输入图像有相同的尺度大小。语义分割任务最初流行的深度学习方法是图像块分类（patch

classification)，即利用像素周围的图像块对每一个像素进行独立的分类。使用图像块分类的主要原因是分类网络通常是全连接层（full connected layer），且要求固定尺寸的图像。

2014 年，加州大学伯克利分校的 Long 等人提出全卷积网络（FCN）<sup>[5]</sup>，这使得卷积神经网络无需全连接层即可进行密集的像素预测，CNN 从而得到普及。使用这种方法可生成任意大小的图像分割图，且该方法比图像块分类法要快上许多。之后，语义分割领域几乎所有先进方法都采用了该模型。

除了全连接层，使用卷积神经网络进行语义分割存在的另一个大问题是池化层。池化层不仅扩大感受野、聚合语境从而造成了位置信息的丢失。但是，语义分割要求类别图完全贴合，因此需要保留位置信息。因此在 2015 年医学影像顶会 ISBI 有学者提出了 U-Net 网络<sup>[6]</sup>，如下图 4。

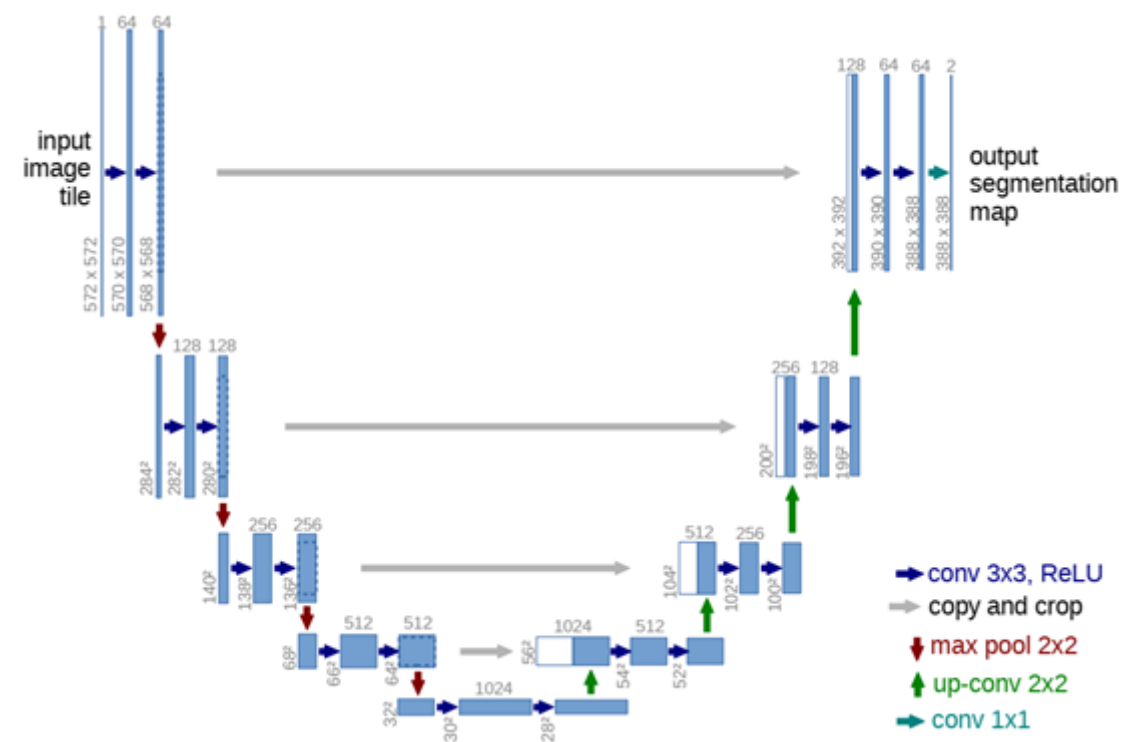


图 4 U-Net 网络结构图

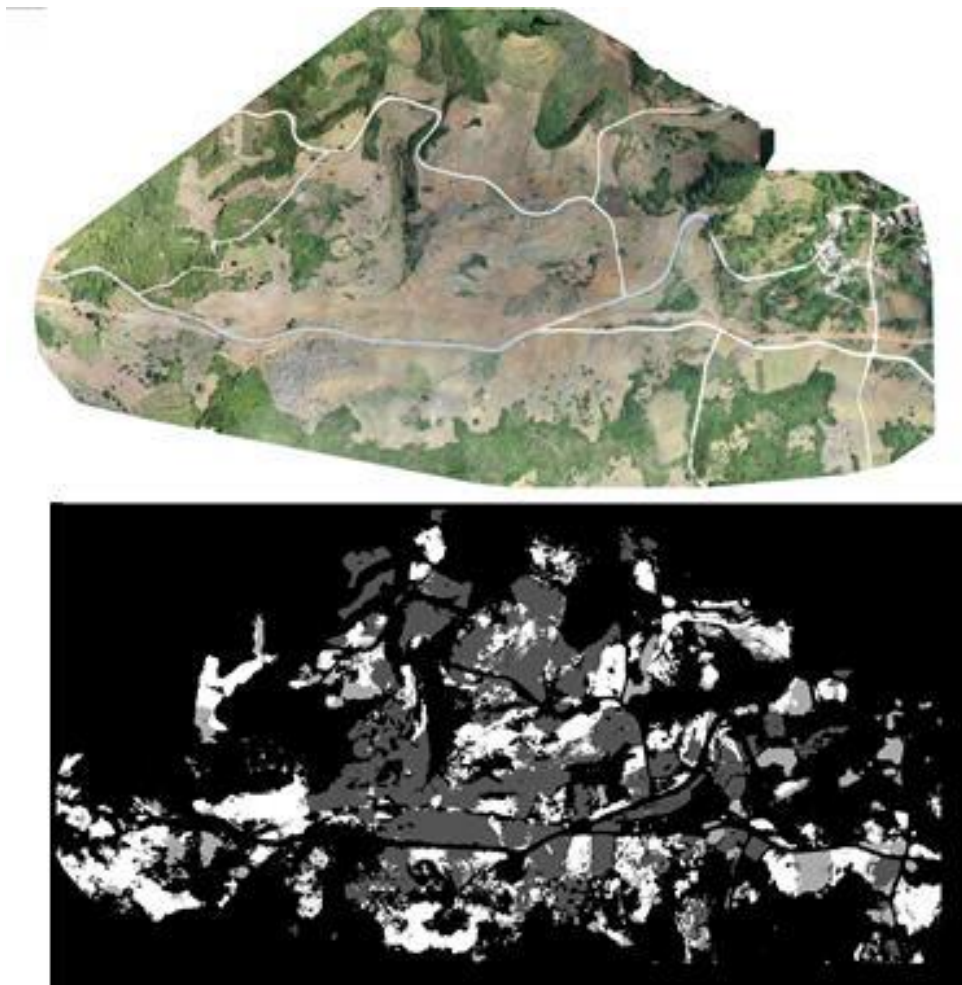
U-Net 建立在 FCN 的网络架构上，作者修改并扩大了这个网络框架，使其能够使用很少的训练图像就得到很精确的分割结果。U-Net 网络非常简单，前半部分作用是特征提取，后半部分是上采样。在一些文献中也把这样的结构叫做编码器-解码器结构。由于此网络整体结构类似于大写的英文字母 U，故得名 U-Net。U-Net 与其他常见的分割网络有一点非常不同的地方：U-Net 采用了完全不同的特征融合

方式：拼接，U-Net 采用将特征在 channel 维度拼接在一起，形成更厚的特征。而 FCN 融合时使用的对应点相加，并不形成更厚的特征。

pooling layer 实现了网络对图像特征的多尺度特征识别。之后上采样部分会融合特征提取部分的输出，这样做实际上是将多尺度特征融合在了一起，以最后一个上采样为例，它的特征既来自第一个卷积 block 的输出(同尺度特征)，也来自上采样的输出(大尺度特征)，这样的连接是贯穿整个网络的，你可以看到上图的网络中有四次融合过程，相对应的 FCN 网络只在最后一层进行融合。因此，概括地说，U-Net 融合了图像的高级语义特征以及低级局部特征，使得图像分割可以达到期望的效果。

## 2.4 预测和保存结果

在模型训练结束，结束后，保存的分割模型输入为  $4 \times 1024 \times 1024$ ，输出为  $1024 \times 1024$ 。故将整个图切块为很多张  $1024 \times 1024$  大小的图像，然后对每张图像预测，最后拼接在一起。如下图 5 为我们预测的结果。



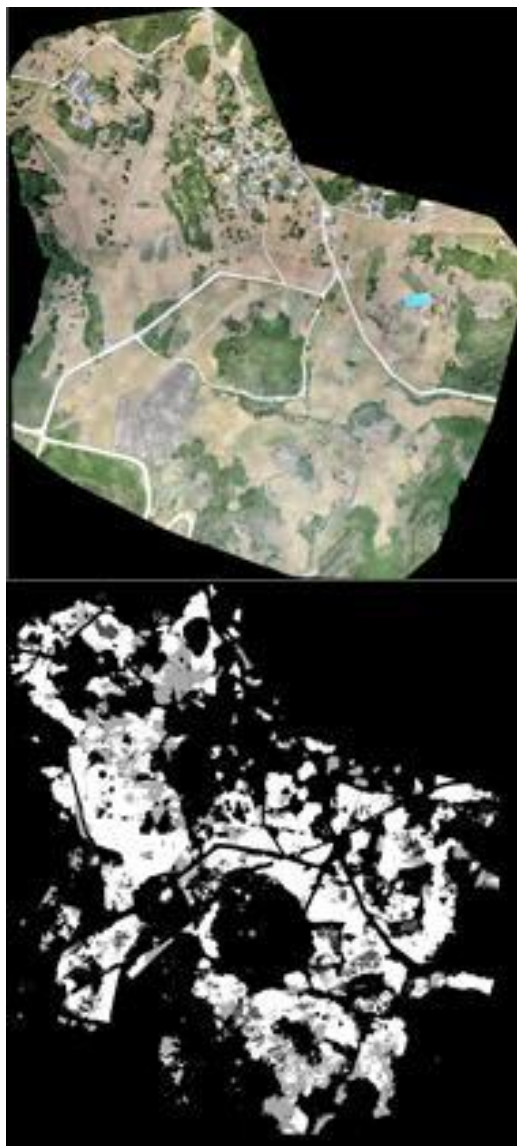


图 5 预测结果

### 3. 实验

#### 3.1 实施细节

实验采用的数据集为天池官方提供的两张遥感大图及对应的标签图像，其分辨率大小分别为 47161x50141，77470x46050，我们对图像进行了切割，切割大小为 1500x1500，最终切出 2217 张图片，随机分出其中 100 张用作验证集，验证训练模型的有效性，并且我们对训练集的 2117 张图片进行了数据增强。

我们基于 Pytorch 深度学习框架设计实验，运行环境为双 NVIDIA TitanXP GPU，32G CPU，评价标准以官网为准选择 MIoU，损失函数选用加权交叉熵损失函数。optimizer 选用 Adam，learning rate=1e-4，batch size 选择 8，epochs 为 100。我们每

隔 3 次保存一次模型。

### 3.2 结果与分析

经过迭代 100 次，绘出了训练曲线图，如下图 6 为训练损失曲线图，图 7 为训练精度曲线图。

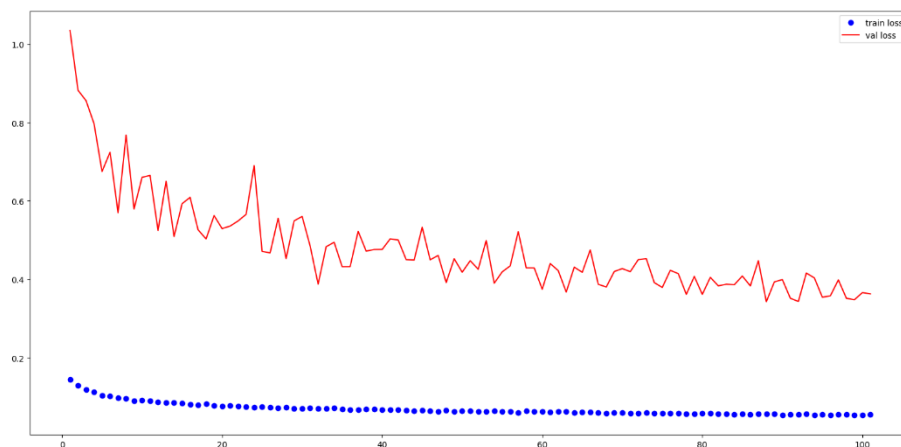


图 6 训练损失

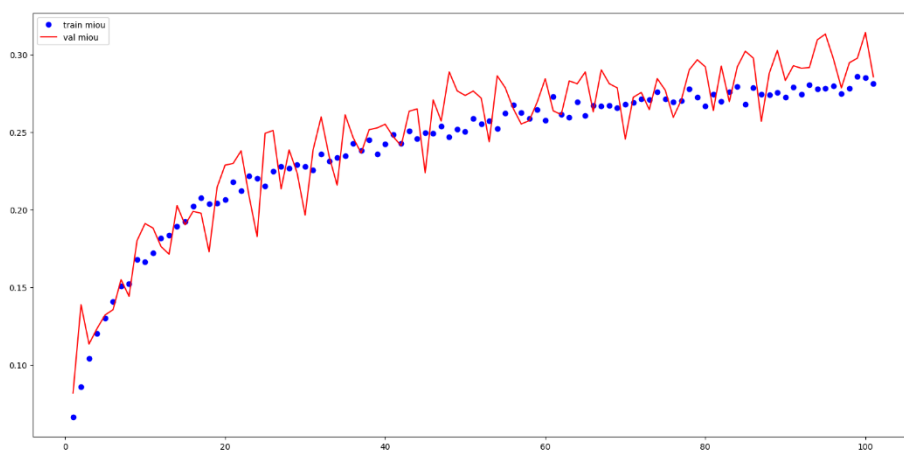


图 7 训练精度

通过观察训练曲线图，可以发现训练过程较为收敛，且稍微有点欠拟合，因为验证集的精度非常接近甚至超过了训练集的精度。因为每隔 3 次保存一次模型，所以我们使用了第 99 次模型进行测试集的预测，并进行提交，如下图 8，第一次提交得到的分数为 0.3028，在当时排名是 80/1520。



日期: 2019-07-09 14:57:00

↑ score: 0.3028

77↓ <sup>3</sup>	你看到我的小熊了吗	jialidun	0.305	2019-07-05
78↓ <sup>3</sup>	A-Force		0.305	2019-07-09
79↓ <sup>3</sup>	lengxia	华东师大	0.305	2019-07-03
80↓ <sup>3</sup>	CQU_IIPLab	重庆大学	0.303	2019-07-09

图 8 提交结果

这之后，我们针对欠拟合的问题对网络结构进行了调整，加深了一层下采样与上采样，即按照之前的规律，分别在编码-解码加了一层，使得结构可以提取更高级的特征。由于加深一层之后训练速度变得非常慢，由于时间关系，我们最终只迭代了 50 次，并选择第 48 次生成的模型进行预测和提交。最终结果如下图 9，分数达到了 0.3047，可以看到相比上次成绩上升了 0.19 个百分点。截止初赛结束最终我们的排名也是定格在了 99/1520，勉强挤进复赛，复赛即将与 7.29 发布数据集。

日期: 2019-07-19 11:49:00

↑ score: 0.3047

95↓ <sup>7</sup>	不行退学	中国科学院	0.307	2019-07-14
96↓ <sup>7</sup>	下了阵小雨	安徽工业大学	0.306	2019-07-18
97↓ <sup>6</sup>	iansn_sss	无	0.305	2019-07-18
98	心意合一	重庆大学	0.305	2019-07-19
99↓ <sup>7</sup>	CQU_IIPLab	重庆大学	0.305	2019-07-19
100↓ <sup>6</sup>	yodalucas	南京大学	0.304	2019-07-16

图 9 最终结果

## 4. 总结

遥感图像中的目标通常具有多尺度、特殊视角、多方向和背景复杂的特点，对语义分割提出了更高的要求。为了适应这种特殊性，基于经典的 U-Net 分割模型，我们加深了一层网络，使其适应遥感图像的语义分割任务，结果得到了提升；针对样本不平衡的问题，我们对交叉熵损失函数进行了加权，赋予不同类别不同权重，具体实施可以见代码部分。最终的 MIoU 达到了 0.3047，具有较为准确的语义分割

10

效果。

这是我第一次参加比赛，感谢叶老师提供这么一个契机，使得我的编程能力和实际应用能力得到了极大的加强。个人觉得最难的部分在于数据预处理部分，这部分的处理极大地锻炼了我们查找文献资料以及编程能力。最后希望后面的复赛可以取得一个比较好的成绩！

## 5. 参考文献

- [1] 杨红卫,童小华.中高分辨率遥感影像在农业中的应用现状[J].农业工程学报,2012,28(24):138-149.
- [2] Li R, Liu Wenjie, Yang Lei, et al. DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2017, PP(99):1-9.
- [3] 吴止锲,高永明,李磊,薛俊诗.类别非均衡遥感图像语义分割的全卷积网络方法[J].光学学报,2019,39(04):401-412.
- [4] 苏健民,杨岚心,景维鹏.基于 U-Net 的高分辨率遥感图像语义分割方法[J].计算机工程与应用,2019,55(07):207-213.
- [5] Long J , Shelhamer E , Darrell T . Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):640-651.
- [6] Ronneberger O , Fischer P , Brox T . U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. 2015.