

# 基于改进DeepLabV3+模型的遥感图像语义分割

蔡翔宇

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2023年2月23日; 录用日期: 2023年3月24日; 发布日期: 2023年3月31日

## 摘要

我国自乡村振兴战略施行以来, 不断促进乡村耕地的改良扩张, 通过遥感图像对其面积的检测提取对于保障我国农村产业发展起到至关重要的作用。针对本文采用的数据集, 普通DeepLabV3+模型缺少对条形结构区域以及离散结构区域特征的关注, 并且缺少自适应的特征层融合过程, 针对该问题, 本文提出一种以MobileNetV2为骨干网络的轻量级改进DeepLabV3+模型。首先, 通过MobileNetV2提取遥感图像特征; 其次, 将空间金字塔模块(ASPP)中的全局平均池化层替换为条形池化层, 便于提取条形区域以及离散区域的特征信息; 最后, 为了获得多尺度特征信息, 加入自适应特征融合模块, 对不同特征层的特征进行融合, 提高网络的特征表达能力。实验结果表明在不同类型的数据集中, 改进DeepLabV3+模型的性能相比普通DeepLabV3+模型以及现有主流网络模型得到了一定提升, 一定程度上改善了DeepLabV3+的不足, 提高了遥感图像语义分割能力。

## 关键词

DeepLabV3+, 遥感图像, 语义分割, 条形池化层, 自适应特征融合模块

# Semantic Segmentation of Remote Sensing Image Based on Improved DeepLabV3+ Model

Xiangyu Cai

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Feb. 23<sup>rd</sup>, 2023; accepted: Mar. 24<sup>th</sup>, 2023; published: Mar. 31<sup>st</sup>, 2023

## Abstract

Since the implementation of rural revitalization strategy, constantly promoting the improvement and expansion of rural farmland, the detection and extraction of its area through remote sensing image plays a vital role in guaranteeing the development of rural industry. Aiming at the data set adopted in this paper, the ordinary DeepLabV3+ model lacks attention to the features of strip structure region and discrete structure region, and lacks adaptive feature layer fusion process. To solve this problem, this paper proposes a lightweight improved DeepLabV3+ model with MobileNetV2 as the backbone network. Firstly, remote sensing image features were extracted by MobileNetV2. Secondly, the global average pooling layer is replaced by the strip pooling layer in the spatial pyramid module (ASPP), which is convenient to extract the feature information of the strip region and discrete region. Finally, in order to obtain multi-scale feature information, an adaptive feature fusion module is added to fuse features of different feature layers and improve the feature representation ability of the network. Experimental results show that in different types of data sets, the performance of the improved DeepLabV3+ model is improved to some extent compared with the ordinary DeepLabV3+ model and the existing mainstream network models, which to some extent improves the deficiency of DeepLabV3+ and improves the semantic segmentation capability of remote sensing images.

## Keywords

DeepLabV3+, Remote Sensing Image, Semantic Segmentation, Strip Pool Layer, Adaptive Feature Fusion Module

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

图像语义分割作为人工智能的热点之一,在农业、医学、交通等领域发挥着至关重要的作用。在我国,实施乡村振兴战略之后,通过 Landsat8 卫星、Modis 卫星或无人机提供的遥感图像或鸟瞰图对水稻种植面积提取、农业耕地识别等图像语义分割技术为普惠金融、农业保险以及农村产业发展提供了有力的支撑。

对于图像语义分割,2012 年 AlexNet [1]的提出为深度学习的发展奠定了快速发展的基石;2013 年深度卷积网络(DCNN) [2]通过提取网络更深的语义信息,与传统的机器学习方法相比具有更强大的特征提取能力;2015 年 UNet [3]和 SetNet [4]等基于编码 - 解码结构的网络结构发挥了全卷积神经网络(FCN) [5]的优势,其中 UNet 对不同通道级通过下采样与上采样进行特征融合,使得模型的特征表达能力更加突出;2014 年开始提出 DeepLab 系列,DeepLabV1 [6]通过去掉池化层保证了特征的平移不变性,并加入空洞卷积扩大感受野避免获得有限的特征信息,DeepLabV2 [7]提出空洞空间金字塔池化层(ASPP)以获得多尺度的特征信息,DeepLabV3 [8]加深了网络结构,在空间金字塔池化层中融合了位置信息,并在 DeepLabV3+ [9]则采用了编码 - 解码结构,通过对深层特征上采样与浅层特征进行融合,增加了图像的语义分割能力,但同时也存在一定的不足,如:边缘信息不明确、分割大目标存在孔洞等。

随着深度学习的发展,注意力机制成为提高模型精确度的必要手段之一,一些学者尝试将注意力机

制融入到网络模型结构当中去。2018 年提出的残差注意力机制[10]通过对不同通道赋予不同的权重,抑制非必要的通道,从而更加突出所需要的特征,这对于多波段遥感图像语义分割起到了极大的作用;同时 2018 年提出的 CBAM (convolutional block attention module) [11]同时关注了通道与空间两个因素,提出了通道注意力机制与空间主语里机制,丰富了特征的提取;此外 2021 年提出的坐标注意力机制[12]缓解了全局平均池化对于特征提取不足的影响,将位置信息嵌入到通道注意力中,对不同空间方面的特征进行提取。

此外,深度学习的发展一定程度上促进着适用于不同场景的语义分割模型的发展,特别是在我国乡村振兴战略支持下,针对于农作物而提出的诸如水稻面积提取、小麦倒伏识别、农作物病态情形抓捕等问题提出的深度学习模型。2019 年 Pena [13]将 SegNet 用于农作物语义分割当中,与传统的机器学习方法相比,精确度得到了大幅度的提升;2021 年宁纪锋,倪静[14]等人通过将注意力机制加入到 DeepLabV3+模型当中,精度高于原有的 DeepLabV3+模型,且对于无人机拍摄的多光谱遥感图像农田识别精度高于可见光;2021 年,邓泓,杨滢婷[15]通过调整空间金字塔池化层中空洞卷积的膨胀率,从而提高了水田的边缘分割精度;2022 年任鸿杰,刘萍[16]等人通过改进空间金字塔池化层,加入了处理正负样本不平衡的损失函数,优化了对于农田模型边缘的识别能力;2022 年董荣胜,马雨琪[17]通过融合不同尺度农作物类别的特征信息,将浅层与深层的特征结合起来,提高特征的表达力。

从上述等研究发现,现有的针对于农作物语义分割的模型存在一定的不足,表现为:1) 没有针对实际情况而提出的模型,如现有农作物种植区域存在着条形分布、离散分布等情形,采用普通的语义分割模型不具有针对性;2) 直接采用传统特征融合方式,未能抑制不重要的特征层的特征表达,从而造成信息污染和干扰;3) 未能缓解由于样本不平衡问题带来的影响。

针对以上问题,本文针对普通 DeepLabV3+模型进行改进并进行模型评估,提出了以下改进思路:1) 使用全局平均池化不可避免地合并来自无关区域的污染信息,因此替换原有空间金字塔中的全局平均池化层为条形池化层;2) 加入自适应特征融合模块,在特征融合模块中加入多尺度通道注意力机制,赋予不同深度特征层相应权重,并提高对小尺度目标的关注度,降低特征融合过程中的噪声干扰,提升模型的鲁棒性;3) 利用类别平衡函数 Focal Loss [18]与 Dice Loss [19]缓解了样本不平衡的问题。

## 2. 网络模型

### 2.1. DeepLabV3+模型

DeepLabV3+模型在 DeepLabV3 模型的基础上改进为编码-解码结构。首先,针对于编码器部分,该模型通过骨干网络进行特征提取并输出浅层特征,再通过空间金字塔池化层增加卷积核的感受野,获取多尺度信息;其次,针对于解码器部分,该部分对编码器部分输出的下采样结果进行上采样,并与编码器部分输入的浅层特征进行结合;最终,通过卷积操作输出预测图像,具体网络结构如图 1 所示。

改进过后的 DeepLabV3+模型通过增加空间金字塔池化获取多尺度信息,增加卷积核的感受野的同时缓解了因为下采样中池化层降低图片维度而导致的信息损失,最后为了减少下采样过程因为网络加深而导致的信息损失,DeepLabV3+模型采用了编码-解码的方式融合浅层特征与深层特征达到特征融合的目的,其中对深层特征上采样的方式采用线性插值方法,通过特征融合使得特征表达能力增强。

### 2.2. 条形池化层

条形池化层整体采用残差网络结构,通过水平面与垂直面的一维卷积对输入图像做一维池化,最后再进行融合,并与原有特征图进行点乘。如图 2 所示,首先条形池化层通过  $H \times 1$  和  $1 \times W$  的两个一维卷积对输入图像进行一维卷积,如公式(1)、公式(2)所示:

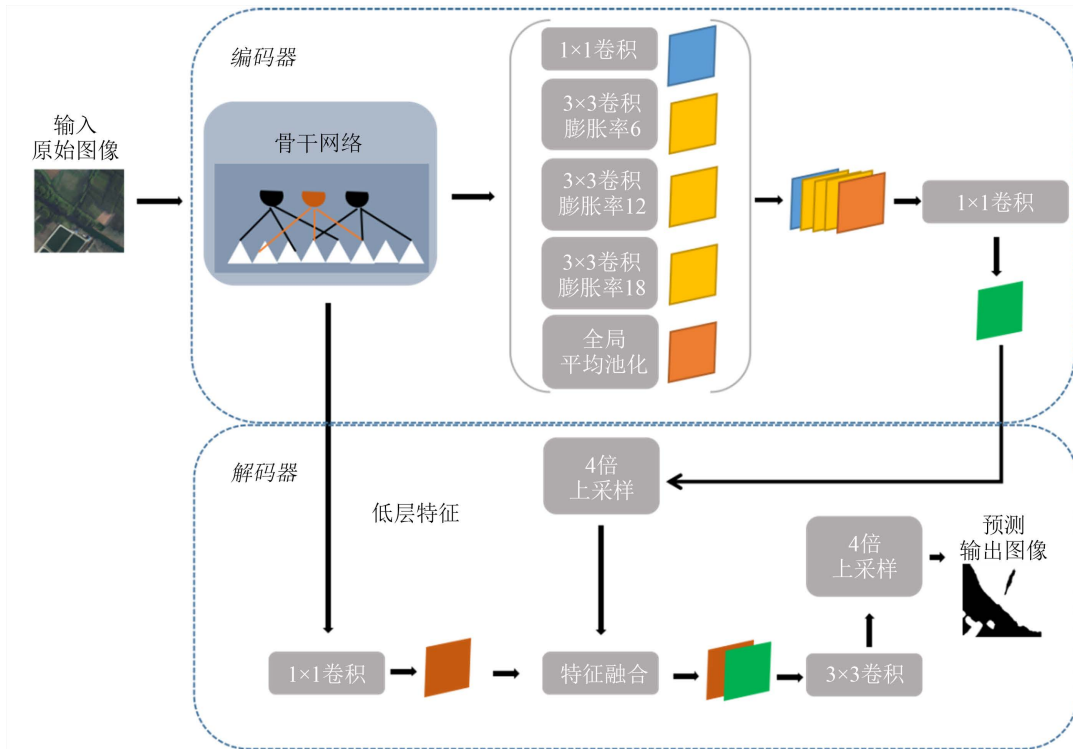


Figure 1. DeepLabV3+ Model  
图 1. DeepLabV3+模型

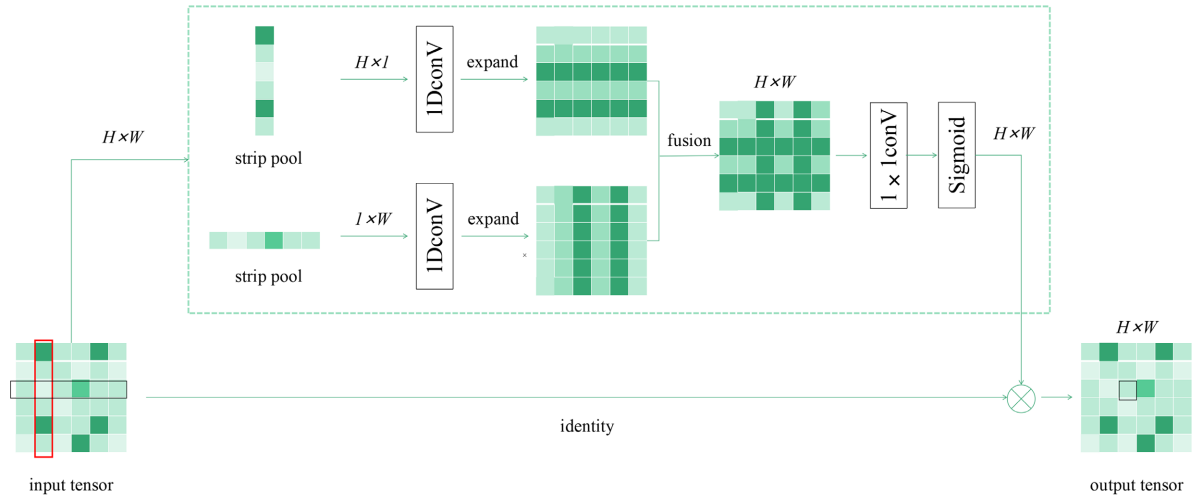


Figure 2. Strip pool layer model  
图 2. 条形池化层模型

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \tag{2}$$

其中  $W$  为输入图像的宽， $H$  为输入图像的高， $c$  为输入图像的通道数， $z_c^h(h)$  为在水平方向做一维池化

的输出结果， $z_c^w(w)$  为在垂直方向做一维池化的输出结果。

其次，通过上采样还原为与原图相同维度的特征图，再将两个方向池化过的特征图进行融合，最后通过 Sigmoid 激活函数输出每个像素点的权重并与原图进行点乘形成新的特征图，如公式(3)、公式(4)所示：

$$z = \sigma(\text{expand}(z_c^h(h)) + \text{expand}(z_c^w(w))) \tag{3}$$

$$\tilde{z} = \sigma(F(z)) \otimes z \tag{4}$$

其中， $\text{expand}$  为上采样操作， $z$  为上采样之后两个特征图融合的结果， $\sigma$  为 Sigmoid 激活函数， $F$  为  $1 \times 1$  卷积， $\otimes$  为点乘操作， $\tilde{z}$  为最终输出特征图。

改进过后的空间金字塔模块(SP\_ASPP+)如图 3 所示。

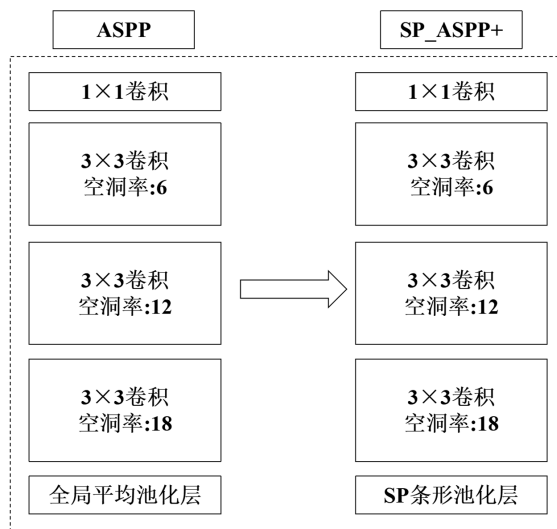


Figure 3. SP\_ASPP+ Module

图 3. SP\_ASPP+模块

条形池化层对于具有长条形结构或离散分布的目标对象具有良好的特征提取作用，特别是在提取农作物的面积时，采样点的农作物种植地呈现如图 4 所示的长条型结构以及离散分布状态，因此采用如全局最大池化或者全局平均池化不能很好地提取特征，其将不可避免地合并来自无关区域的污染信息。

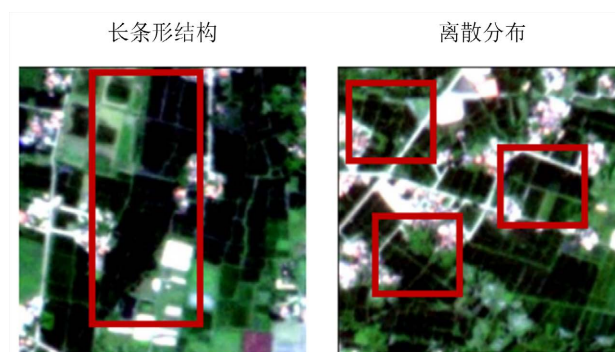


Figure 4. Part of the special samples

图 4. 部分特殊样本展示

### 2.3. 自适应特征融合模块

#### 2.3.1. 局部通道注意力机制

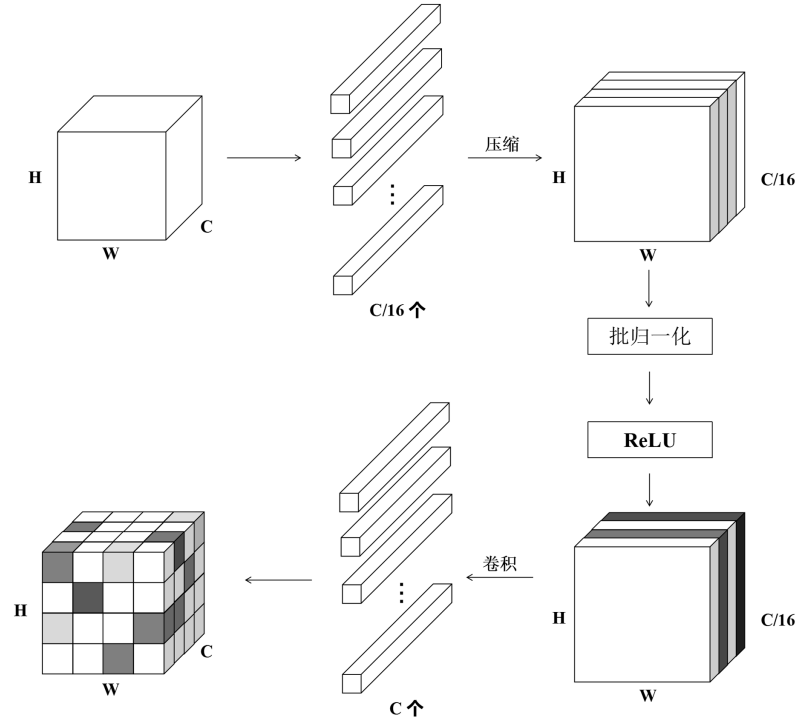


Figure 5. Local channel attention  
图 5. 局部通道注意力机制

局部通道注意力机制如图 5 所示，首先，通过全局平均池化将  $H \times W \times C$  的特征图压缩为  $1 \times 1 \times C$  的特征图。具体过程如公式(5)所示：

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), Z \in R^C \quad (5)$$

其中， $H$ 、 $W$  和  $C$  分别表示为输入图像的高、宽和通道数； $F_{sq}$  为全局池化操作； $u_c$  为输入的特征图。

其次，通过卷积层压缩通道数，获得通道之间的关系，再通过批归一化和 ReLU 激活函数规范权重分布，得到每个通道的权重  $s$ 。具体过程如公式(6)所示：

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 ReLU(BN(W_1 z))) \quad (6)$$

其中  $W_1 \in R^{r \times C}$ ， $W_2 \in R^{C \times r}$ ， $BN$  为批归一化操作， $F_{ex}$  为激励操作； $z$  为挤压操作得到的池化特征图； $W_1$  和  $W_2$  分别为第一层卷积层和第二层卷积层的权重； $r$  为降维系数。最后将权重与原始特征相乘，得到新的特征。具体过程如公式(7)所示：

$$\tilde{x}_c = F_{scale}(u_c, s_c) \quad (7)$$

其中  $F_{scale}$  为点乘操作。

局部通道注意力使得神经网络在训练过程中可以赋予不同重要程度的通道不同的权重，自适应选择通道的有效信息，排除某些通道上的噪声信息。全局通道注意力结构以残差注意力机制的结构为基础，



通过增加批归一化和 ReLU 激活用于调整权重分布，残差结构增强了模型对强噪声和高冗余数据的处理能力。

### 2.3.2. 全局通道注意力机制

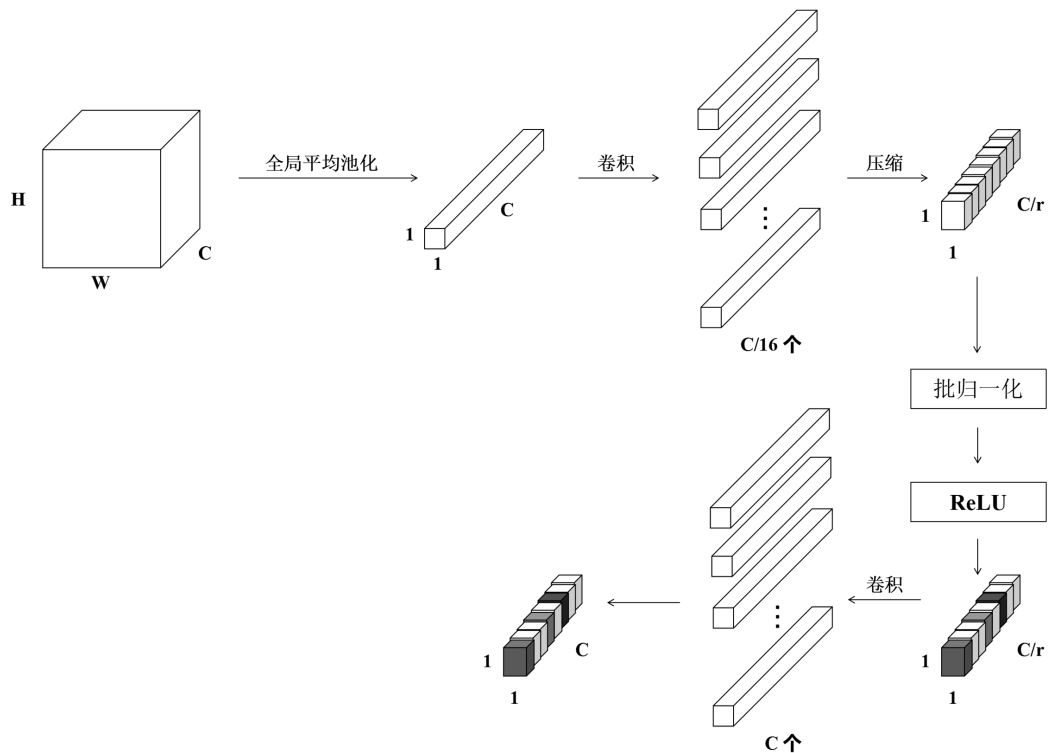


Figure 6. Global channel attention  
图 6. 全局通道注意力机制

全局通道注意力与局部通道注意力相似如图 6 所示，输入同样为  $H \times W \times C$  的特征图，输出为维相同维度的通道权重矩阵。局部通道注意力机制通过全局平均池化简化了位置信息，将位置信息压缩到一个像素点之上，使得神经网络更好地关注于通道之间的关系，在对小尺寸目标的特征提取上具有较好的鲁棒性，但由于将水平方向和垂直方向上的信息压缩到一个像素点会导致重要的位置信息损失，因此全局通道注意力模块舍弃了全局平均池化，关注全局信息，直接采用  $1 \times 1$  卷积将输入特征图的通道数降为原来的  $1/16$ ，通过卷积建模各通道间的相关性，并将特征图通道数恢复为  $C$ 。

### 2.3.3. 多尺度通道注意力机制

多尺度通道注意力结合全局通道信息和局部通道信息，如图 7 所示。首先局部通道注意力机制更好地关注了通道之间的关系并且提高了神经网络对小尺度目标的关注度；其次，全局通道注意力机制舍弃全局平均池化，更加关注位置上的特征信息。通过结合全局通道信息和局部通道信息，降低特征融合过程中的噪声干扰，提升图像语义分割算法的鲁棒性。具体过程如公式(8)所示：

$$\tilde{Z} = Z \otimes (\sigma(A(Z) + B(Z))) \quad (8)$$

其中， $Z$  为输入的特征图， $A$  为局部通道注意力， $B$  为全局通道注意力， $\sigma$  为 Sigmoid 激活函数， $\otimes$  为对应元素点乘， $\tilde{Z}$  为最后输出特征图。

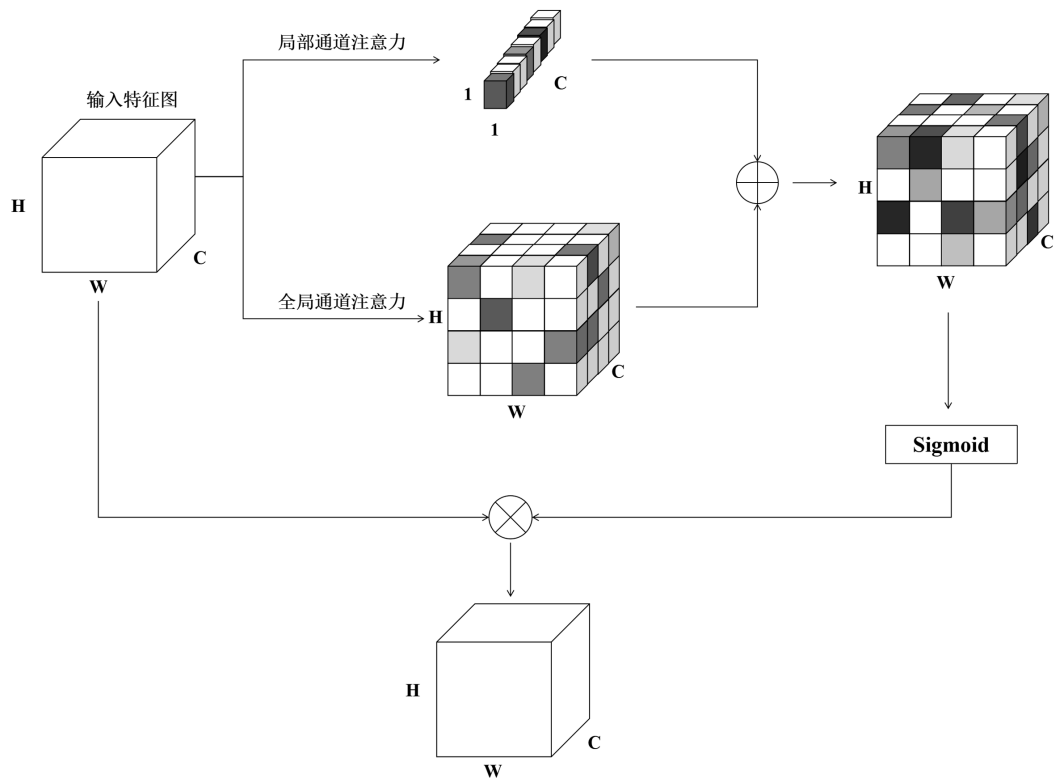


Figure 7. Multi-scale channel attention  
图 7. 多尺度通道注意力机制

### 2.3.4. 自适应特征融合模块

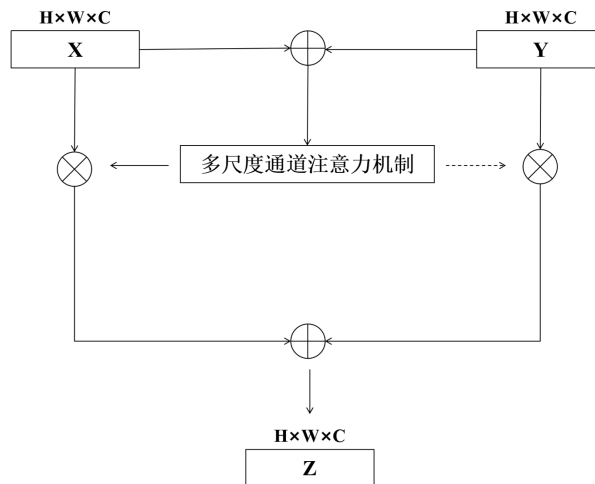


Figure 8. Adaptive feature fusion module  
图 8. 自适应特征融合模块

自适应特征融合模块如图 8 所示。首先，将不同尺度的特征图对应元素进行相加，并通过多尺度通道注意力机制获得不同通道的权重；其次，对不同尺度的特征图的通道赋予权重；最后，将赋予权重后的不同尺度的特征图对应元素相加，以达到特征融合的目的。具体过程如公式(9)所示：



$$\tilde{Z} = G(X+Y) \otimes X + (1-G(X+Y)) \otimes Y \quad (9)$$

其中,  $X$  和  $Y$  为维度相同的两个不同特征层的特征图,  $G$  为多尺度通道注意力,  $\otimes$  为对应元素点乘,  $\tilde{Z}$  为最后输出特征图。

自适应特征融合模块等同于对不同尺度的特征图进行加权平均以达到选择较重要的特征层的作用, 均衡了输入特征所携带信息所占比重。

## 2.4. 改进 DeepLabV3+模型

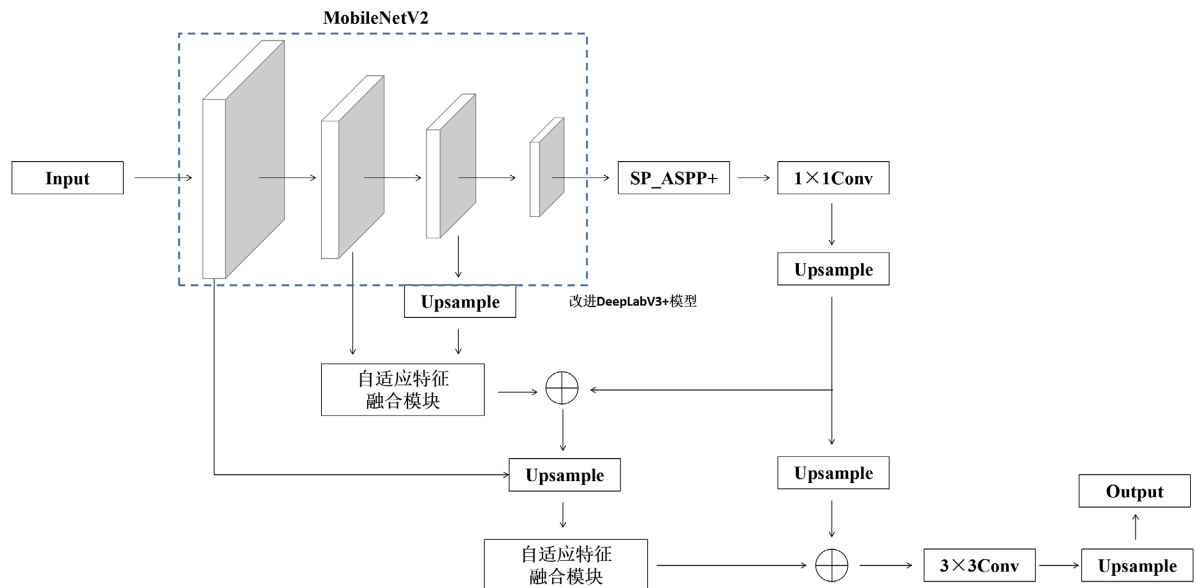


Figure 9. Improved DeepLabV3+ model

图 9. 改进 DeepLabV3+模型

改进 DeepLabV3+模型如图9所示。首先, 通过特征提取器 MobileNetV2 骨干网络提取特征; 其次, 通过 SP\_ASPP+模块利用不同感受野的卷积核获得多尺度目标信息; 最后, 通过两次自适应特征融合模块对不同尺度的特征信息进行融合。

改进 DeepLabV3+模型首先改善了由于全局平均池化带来的冗余信息的干扰, 针对样本本身出现的条形结构以及离散结构提出 SP 条形池化层, 增加了特征信息的提取能力; 其次, 由于低层特征分辨率更高, 包含更多位置、细节信息, 而高层特征具有更强的语义信息, 但是分辨率很低, 对细节的感知能力较差, 因此通过自适应特征融合模块可以自适应地选择对提升语义分割精度有用的特征层信息, 缓解了普通特征融合模块带来的由于尺度和语义信息不一致带来的信息差错。

## 2.5. 损失函数

为了缓解类别不平衡问题, 引入 Focal Loss 与 Dice Loss 替代交叉熵损失函数。本文考虑改进的 Focal Loss, 其同时考虑了“正负样本平衡”和“难以分辨样本”的问题, 根据不同类别样本的数量设置权重, 并给予更难辨别的样本更大的权重。具体过程如公式(10)所示:

$$Loss_{Focal} = -\alpha(1-p)^\gamma \log p \quad (10)$$

$\alpha$  为样本类别权重, 给不同类别的样本损失函数给予权重;  $p$  为样本预测置信度;  $\gamma$  减少预测置信度

较高样本损失函数的比重,使得模型更关注置信度较低样本损失函数的优化。由于 Dice Loss 训练时不稳定,因此与改进的 Focal Loss 结合使用.具体过程如公式(11)所示:

$$Loss = Loss_{Focal} + Loss_{Dice} \quad (11)$$

### 3. 实验

#### 3.1. 数据集

为了验证改进 DeepLabV3+模型的有效性,本文通过实地考察采集,选取成都市青白江区以及大邑县中两个试验地的数据,并通过人工标注区分水稻以及非水稻区域并制作标签,水稻区域标注为 1,非水稻区域标注为 0,对这两类样本进行语义分割实验。

通过裁剪、翻转、旋转以及加入噪点对数据进行数据增强分别产生 1152 张尺寸大小为  $256 \times 256$  的三波段遥感图像,其中训练集 736 张,验证集 185 张,测试集 231 张。遥感图像格式为 jpg 格式,标签格式为 png 格式,实验中不更改遥感图像和标签的图像格式,如图 10,图 11 所示。



Figure 10. Display of datas  
图 10. 数据集展示

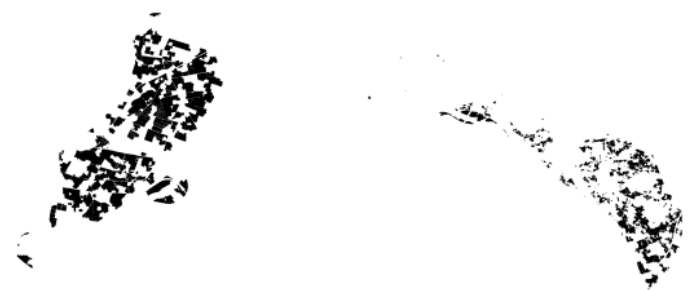


Figure 11. Display of Datas' Labels  
图 11. 数据集标签展示

#### 3.2. 实验参数与环境

为了保证实验的公平性,本文通过控制变量法保证每个实验中的超参数相同,5 个实验均将初始学习率设置为 0.005,学习率下降方式选择余弦下降方式,优化器选择 Adam 优化器,动量设置为 0.9,weight decay 设置为 0.0001,下采样倍数设置为 8,迭代次数设置为 200 次,骨干网络使用 MobileNetV2 的预训练权重进行训练.实验在 Linux 系统中执行,实验平台为 Python3.7,深度学习框架为 Pytorch1.10.0,CUDA 版本为 11.6,显卡型号为 NVIDIA Tesla A40。

### 3.3. 评价指标

为了评估模型性能, 本文在选择交并比(mIoU)、召回率(Recall)、训练参数 2 种指标来评价模型的精度, 以上 2 种指标标准值越高则精度越高。

### 3.4. 实验结果

为了验证改进 DeepLabV3+模型有效性, 设计了 4 个实验, 分别为对照组实验、在 DeepLabV3+模型中加入 SP\_ASPP+模块、在 DeepLabV3+模型中加入自适应特征融合模块、在 DeepLabV3+模型中加入 SP\_ASPP+模块和自适应特征融合模块, 并对不同模型的一个 Batch 的训练参数进行统计, 实验结果如表 1、表 2、表 3 所示。

**Table 1.** Ablation results (IoU)

**表 1.** 消融实验结果(IoU)

实验编号	Methods	青白江区试验地			大邑县试验地		
		总体	水稻	非水稻	总体	水稻	非水稻
1	对照组	81.77	68.32	95.22	76.05	56.3	95.81
2	SP_ASPP+	84.82	73.52	96.13	76.58	57.32	95.84
3	自适应特征融合模块+ASPP	87.6	78.28	96.91	77.81	59.46	<b>96.16</b>
4	自适应特征融合模块+SP_ASPP+	<b>88.02</b>	<b>79.07</b>	<b>96.96</b>	<b>78.17</b>	<b>60.23</b>	96.12

**Table 2.** Ablation results (Recall)

**表 2.** 消融实验结果(Recall)

实验编号	Methods	青白江区试验地			大邑县试验地		
		总体	水稻	非水稻	总体	水稻	非水稻
1	对照组	90.16	83.06	97.26	85.08	72.32	97.83
2	SP_ASPP+	92.00	86.19	97.81	86.16	74.63	97.7
3	自适应特征融合模块+ASPP	93.64	89.01	<b>98.26</b>	86.63	75.29	<b>97.97</b>
4	自适应特征融合模块+SP_ASPP+	<b>94.72</b>	<b>91.42</b>	98.02	<b>88.05</b>	<b>78.4</b>	97.7

**Table 3.** Ablation results (Total Params)

**表 3.** 消融实验结果(Total Params)

实验编号	Methods	ToalParams
1	对照组	5.813266M
2	SP_ASPP+	5.895442M
3	自适应特征融合模块+ASPP	4.54153M
4	自适应特征融合模块+SP_ASPP+	4.623706M

由表 1、表 2 发现, 首先, 在 DeepLabV3+模型中加入 SP\_ASPP+模块后, 在 2 种数据集的消融实验中平均 IoU 指标值与平均召回率指标值均得到了提升; 其次, 在 DeepLabV3+模型中加入自适应特征融合模块后, 在 2 种数据集的消融实验中平均 IoU 指标值与平均召回率指标值均得到了更大的提升, 其中, 青白江区试验地的平均 IoU 指标值与平均召回率指标值分别提升了 5.83%和 3.52%; 最后, 在 DeepLabV3+模型同时加入 SP\_ASPP+模块与自适应特征融合模块后, 效果比单独加入两类模块, 在两个试验地上的

平均 IoU 指标值与平均召回率指标值分别提升了 6.25%、4.56%和 2.12%、2.97%。这两类实验更好验证了改进 DeepLabV3+模型的有效性。同时由表 3 发现,改进 DeepLabV3+模型相对减少了训练参数,在数量级上保证了轻量型网络的特征,同时提升了相对精度,具有良好的性能。

此外,为了说明改进 DeepLabV3+模型的性能,与当下主流网络(FCN8S、SegNet 以及 UNet++)进行对比,对比实验结果如表 4、表 5 以及表 6 所示。通过对比实验说明,改进 DeepLabV3+模型相比当下主流网络而言,在保证轻量型网络的基础上同时提升了语义分割的精度,具有较好的应用效果。

**Table 4.** Comparative experimental results (IoU)

**表 4.** 对比实验结果(IoU)

Methods	青白江区试验地			大邑县试验地			参数
	总体	水稻	非水稻	总体	水稻	非水稻	
FCN8s	83.93	72.01	95.86	63.07	36.57	89.58	23.953442M
SegNet	81.32	67.56	95.07	63.37	37.54	89.2	29.444162M
UNet++	84.24	72.39	96.09	65.1	37.2	93.0	83.32233M
自适应特征融合模块+SP_ASPP+	<b>88.02</b>	<b>79.07</b>	<b>96.96</b>	<b>78.17</b>	<b>60.23</b>	<b>96.12</b>	<b>4.623706M</b>

**Table 5.** Comparative experimental results (Recall)

**表 5.** 对比实验结果(Recall)

Methods	青白江区试验地			大邑县试验地			参数
	总体	水稻	非水稻	总体	水稻	非水稻	
FCN8s	91.36	85.02	97.69	76.94	60.66	93.21	23.953442M
SegNet	89.76	82.31	97.22	78.82	65.25	92.39	29.444162M
UNet++	84.24	72.39	96.09	70.55	42.61	<b>98.5</b>	83.32233M
自适应特征融合模块+SP_ASPP+	<b>94.72</b>	<b>91.42</b>	<b>98.02</b>	<b>88.05</b>	<b>78.4</b>	97.7	<b>4.623706M</b>

**Table 6.** Comparative experimental results (Toal Params)

**表 6.** 对比实验结果(Total Params)

Methods	ToalParams
FCN8s	23.953442M
SegNet	29.444162M
UNet++	83.32233M
自适应特征融合模块+SP_ASPP+	<b>4.623706M</b>

首先,改进 DeepLabV3+模型通过 SP\_ASPP+模块中的 SP 条形池化层缓解了由于全局平均池化带来的位置信息损失,针对数据集中的“条型结构”以及“离散结构”提出了较好的解决方案,将方形卷积核拆分为两个不同方向的一维卷积核,并通过残差网络结构实现长期依赖关系,达到提高特征提取的能力;其次,自适应特征融合模块通过全局和局部通道注意力机制缓解尺度变化和小对象引起的问题,弥补不同尺度间的特征不一致,实现特征融合。最后的预测图像如图 12 所示,在原始图像 1 中,水稻区域呈现离散分布,对照组出现了预测水稻区域粘连和多余的情景,通过加入 SP\_ASPP+模块缓解了由于全局平均池化带来的信息干扰,自适应特征融合模块使得特征提取更加完善,而通过两类模块融合达到了比较好的预测效果,缓解了区域粘连和多余的问题;在原始图像 2 中,方框 1 中为条型结构,方块 2 中

为离散结构，对照组未能将条形结构区域预测出来，而 SP\_ASPP+模块虽然缓解了部分条形结构区域消失的问题，但由于特征信息提取不完全，未能将水稻区域完全预测，自适应特征融合模块增加了特征信息，但同时增加了部分冗余信息，而通过两类模块融合使得特征信息有较好的表达。

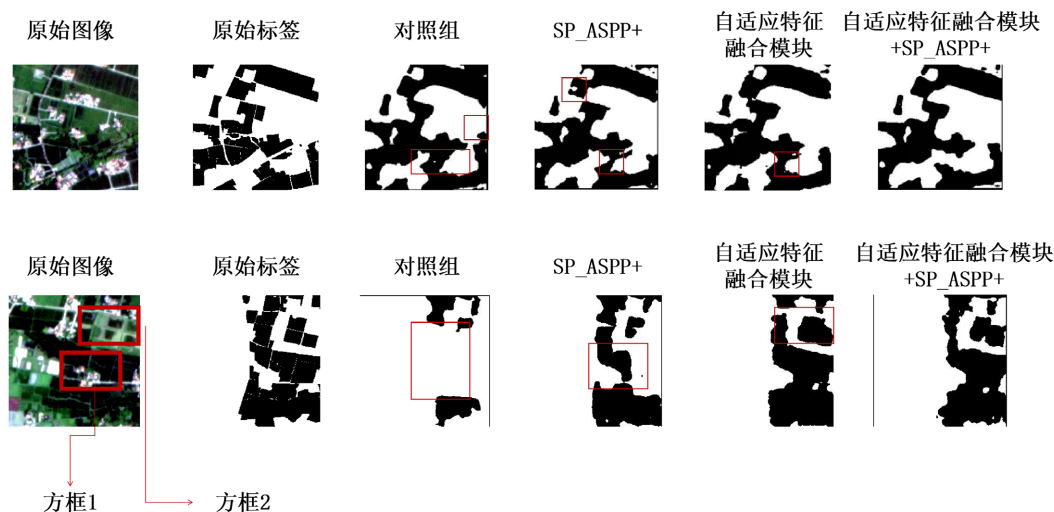


Figure 12. Prediction images

图 12. 预测图像

#### 4. 结束语

本文提出改进 DeepLabV3+模型，并利用 2 种试验地数据进行训练和测试，得到的主要结论为：

首先，针对样本出现的条形结构区域和离散结构区域，加入 SP\_ASPP+模块缓解了由于全局平均池化带来的位置信息损失，减少了冗余信息带来的干扰；其次，通过自适应特征融合模块自适应选择特征层的信息进行融合，抑制不重要的特征层的信息表达，克服输入特征之间语义和规模不一致的问题。

由于骨干网络为轻型 MobileNetV2 网络，因此若要得到更高的精确度，可以替换如 Xception 网络、ResNet 网络作为骨干网络，在实验条件允许的情况下，后期可以通过替换骨干网络以达到提升网络性能的目的。

#### 参考文献

- [1] Krizhevsky, A., Sutskever, I., Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012*, Lake Tahoe, 3-6 December 2012, 2.
- [2] Sun, Y., Wang, X., Tang, X. (2013) Deep Convolutional Network Cascade for Facial Point Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 23-28 June 2013, 3476-3483. <https://doi.org/10.1109/CVPR.2013.446>
- [3] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer, Cham.
- [4] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [5] Shelhamer, E., Long, J. and Darrell, T. (2014) Fully Convolutional Networks for Semantic Segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 3431-3440.
- [6] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K.P. and Yuille, A.L. (2014) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**,

- 834-848.
- [7] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K.P. and Yuille, A.L. (2016) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
  - [8] Chen, L., Papandreou, G., Schroff, F. and Adam, H. (2017) Rethinking Atrous Convolution for Semantic Image Segmentation.
  - [9] Chen, L., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Computer Vision—ECCV 2018 15th European Conference*, Munich, 8-14 September 2018, 833-851. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)
  - [10] Hu, J., Shen, L. and Sun, G. (2017) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
  - [11] Woo, S., Park, J., Lee, J. and Kweon, I. (2018) CBAM: Convolutional Block Attention Module. *European Conference on Computer Vision*, Munich, 8-14 September 2018, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
  - [12] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13708-13717. <https://doi.org/10.1109/CVPR46437.2021.01350>
  - [13] Pena, J., Tan, Y.M. and Boonpook, W. (2019) Semantic Segmentation Based Remote Sensing Data Fusion on Crops Detection. *Journal of Computer and Communications*, **7**, 53-64. <https://doi.org/10.4236/jcc.2019.77006>
  - [14] 宁纪锋, 倪静, 何宜家, 李龙飞, 赵志新, 张智韬. 基于卷积注意力的无人机多光谱遥感影像地膜农田识别[J]. *农业机械学报*, 2021, 52(9): 213-220.
  - [15] 邓泓, 杨滢婷, 刘兆朋, 刘木华, 陈雄飞, 刘鑫. 基于深度学习的无人机水田图像语义分割方法[J]. *中国农机化学报*, 2021, 42(10): 165-172.
  - [16] 任鸿杰, 刘萍, 岱超, 史俊才. 改进 DeepLabV3+网络的遥感影像农作物分割方法[J]. *计算机工程与应用*, 2022, 58(11): 215-223.
  - [17] 董荣胜, 马雨琪, 刘意, 李凤英. 加强类别关系的农作物遥感图像语义分割[J]. *中国图象图形学报*, 2022, 27(11): 3382-3394.
  - [18] Lin, T., Goyal, P., Girshick, R.B., He, K. and Dollár, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/ICCV.2017.324>
  - [19] Milletari, F., Navab, N. and Ahmadi, S. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *4th International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3DV.2016.79>