

基于多尺度特征交互和边界优化的遥感图像语义分割

张晗 吴希文 刘勋

Semantic segmentation of remote sensing images based on multi-scale feature interaction and boundary optimization

ZHANG Han, NG Alex Hay-Man, LIU Xun

引用本文:

张晗, 吴希文, 刘勋. 基于多尺度特征交互和边界优化的遥感图像语义分割[J]. 应用光学, 2025, 46(3): 612–621. DOI: 10.5768/JAO202546.0302001

ZHANG Han, NG Alex Hay-Man, LIU Xun. Semantic segmentation of remote sensing images based on multi-scale feature interaction and boundary optimization[J]. Journal of Applied Optics, 2025, 46(3): 612–621. DOI: 10.5768/JAO202546.0302001

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0302001>

您可能感兴趣的其他文章

Articles you may be interested in

基于改进DeepLabv3+的无人车夜间红外图像语义分割

Semantic segmentation of nocturnal infrared images of unmanned vehicles based on improved DeepLabv3+

应用光学. 2020, 41(1): 180–185 <https://doi.org/10.5768/JAO202041.0106002>

基于改进RoI Transformer的遥感图像多尺度旋转目标检测

Multi-scale oriented object detection based on improved RoI Transformer in remote sensing images

应用光学. 2023, 44(5): 1010–1021 <https://doi.org/10.5768/JAO202344.0502001>

基于语义特征的遥感舰船关键部位检测网络

Detection network of critical parts for remote sensing ship based on semantic features

应用光学. 2023, 44(3): 595–604 <https://doi.org/10.5768/JAO202344.0303004>

基于掩码一致性机制的弱监督图像语义分割研究

Weakly supervised image semantic segmentation based on masked consistency mechanism

应用光学. 2024, 45(4): 741–750 <https://doi.org/10.5768/JAO202445.0402003>

基于深度学习的铸件X射线图像分割研究

Casting X-ray image segmentation based on deep learning

应用光学. 2021, 42(6): 1025–1033 <https://doi.org/10.5768/JAO202142.0602003>

基于改进BiSeNet的非结构化道路分割算法研究

Unstructured road segmentation algorithm based on improved BiSeNet

应用光学. 2023, 44(3): 556–564 <https://doi.org/10.5768/JAO202344.0302003>



关注微信公众号，获得更多资讯信息

文章编号: 1002-2082 (2025) 03-0612-10

引用格式: 张晗, 吴希文, 刘勋. 基于多尺度特征交互和边界优化的遥感图像语义分割 [J]. 应用光学, 2025, 46(3): 612-621.

ZHANG Han, NG Alex Hay-Man, LIU Xun. Semantic segmentation of remote sensing images based on multi-scale feature interaction and boundary optimization[J]. Journal of Applied Optics, 2025, 46(3): 612-621.



在线阅读

基于多尺度特征交互和边界优化的 遥感图像语义分割

张 晗¹, 吴希文², 刘 勋¹

(1. 广东工业大学 信息工程学院, 广东 广州 510006; 2. 广东工业大学 土木与交通工程学院, 广东 广州 510006)

摘 要: 对遥感图像进行语义分割是一项具有重要理论意义和实用价值的任务。遥感图像中含有丰富的地物信息, 边界处的像素信息也较难判定, 分割难度较大。基于 Mask2Former 架构, 提出了改进的多尺度特征交互架构 Mask2Former-MS 和边界优化架构 Mask2Former-BR。前者利用双线性插值法进行上采样和下采样, 来达到特征融合的效果, 并引入通道注意力机制减少冗余信息的影响; 后者通过空洞卷积金字塔池化模块 (atrous spatial pyramid pooling, ASPP) 进行特征提取, ASPP 的每个并行空洞卷积分支使用不同的空洞率捕捉不同尺度的特征信息, 利用 ReLU 层和 BN(batch normalization) 层进行激活和归一化处理抑制梯度消失, 使边界处的像素更加准确。实验结果显示, 在 GID(Gaofen image dataset) 数据集上, 通过对比 U-Net 网络和 Mask2Former 架构, 所改进的 Mask2Former-MS 架构和 Mask2Former-BR 架构的最优精确度和最优准确度分别为 88.82%、85.90% 和 89.56%、87.46%, 改进架构的分割效果更优。

关键词: 深度学习; 语义分割; 遥感图像; 多尺度特征; 边界优化

中图分类号: TN201; TP751

文献标志码: A

DOI: 10.5768/JAO202546.0302001

Semantic segmentation of remote sensing images based on multi-scale feature interaction and boundary optimization

ZHANG Han¹, NG Alex Hay-Man², LIU Xun¹

(1. School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China;

2. School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Semantic segmentation of remote sensing images is a task of great theoretical importance and practical values. Remote sensing images contain rich feature information, and the pixel information at the boundary is also more difficult to determine, making segmentation more difficult. Based on the structure of Mask2Former, an improved multi-scale feature interaction structure Mask2Former-MS and a boundary optimization structure Mask2Former-BR were proposed, the former utilized bilinear interpolation for up-sampling and down-sampling to achieve the effect of feature fusion, and a channel attention mechanism was introduced to reduce the effect of redundant information. The latter utilized the atrous spatial pyramid pooling

收稿日期: 2024-07-05; 修回日期: 2024-10-16

基金项目: 国家自然科学基金 (42274016); 广东省林业科学数据中心项目 (2021B1212100004); 粤港澳大湾区湿地系统水质改善、生态修复与多维调控项目 (2019ZT08L213)

作者简介: 张晗 (1999—), 男, 硕士研究生, 主要从事遥感图像语义分割算法研究。E-mail: ubtamqud540786@foxmail.com

通信作者: 吴希文 (1983—), 男, 教授, 博士生导师, 主要从事卫星遥感图像分析方法及应用研究。

E-mail: hayman.ng@foxmail.com

(ASPP) module for feature extraction, the different atrous rates was used to capture feature information at different scales by each parallel atrous convolution branch of ASPP, and the ReLU layer and batch normalization (BN) layer for activation and normalization were used to suppress gradient vanishing and made the pixels at the boundary more accurate. The experimental results show that by comparing the U-Net network and the Mask2Former structure on the Gaofen image dataset (GID), the optimal precision and optimal accuracy of the improved Mask2Former-MS structure and the Mask2Former-BR structure are 88.82%, 85.90% and 89.56%, 87.46%, respectively, and the segmentation effect of the improved structures is better.

Key words: deep learning; semantic segmentation; remote sensing image; multi-scale features; boundary optimization

引言

随着全球城市化进程的加速、自然资源的枯竭以及环境问题的日益严重,对土地利用覆盖的监测与管理变得越来越重要。遥感技术作为一种重要的对地观测方式,提供了获取地表信息的有效途径。在遥感技术中,土地利用语义分割是一项关键的研究任务,其目的是将遥感影像中的像素按其所代表的地物类别进行分割,从而实现对地表的精细化描述和定量分析。土地利用语义分割能够提供更精确、更全面的地表信息,可以准确识别出各种地表特征,为城市规划、资源管理和环境保护^[1]等提供了更详细的空间信息,有助于政府部门和相关机构更好地制定决策。土地利用语义分割在环境监测和生态保护方面也具有重要的意义^[2],及时监测和评估土地利用覆盖的变化,发现潜在的问题,并采取相应的措施加以防治,维护生态系统的相对稳定。综上所述,开展土地利用语义分割的研究具有重要的现实意义。

在上个世纪末形成了最初开创的传统机器学习语义分割算法,例如 HEARST M A 等人在 1998 年提出的支持向量机 (support vector machine, SVM) 算法^[3],其在统计学框架或 VC 理论 (Vapnik-Chervonenskis theory) 的基础上提出了监督学习分类器;柳秀山^[4]等人利用支持向量机的高分遥感影像对道路进行了提取;2001 年, BREIMAN L 在学术期刊上发表了关于随机森林^[5]的机器学习算法,其采用了集成学习的思想,将多个弱学习器(决策树)组合成一个强学习器(森林),能够提高算法的性能和泛化能力;梁锦涛^[6]等人基于融合指数和主成分分量的随机森林方法得到高精度的分类结果。虽然这些算法在遥感领域中取得了不错的效果,但因其具有提取的特征因子较为单一、语义分割的效率较低以及最终的结果主观性强的劣势,在面

对要素更多样化的高分辨率遥感影像时,效果逐渐不能满足实际需求。

在深度学习日益发展的今天,一些经典的语义分割算法逐渐涌现了出来。例如全卷积网络^[7] (fully convolutional networks, FCN),它通过对输入的图像进行像素级别的分类,第一次实现了图像的语义级分割;沈言善^[8]等人改进了一种全卷积神经网络,实现了对遥感地物信息的有效分割;SegNet^[9]是一种专为像素级图像分割任务设计的深度学习模型,其具有的 Encoder-Decoder 架构和池化索引传递机制,在遥感图像分割领域中达到了良好的效果;关世杰^[10]等人利用改进 SegNet 网络实现对遥感图像分割精度的明显提升;U-Net^[11]是一种著名的卷积神经网络架构,因其独特的“U”形结构而得名,该模型能够有效捕捉遥感图像中的上下文信息并实现精确定位,同时,U-Net 架构具有灵活的适用范围,在深度学习遥感图像分割领域占据着重要的地位。

2020 年, CHENG B 等人设计出一种在图像分割领域有更好通用性的架构 Mask2Former^[12],该框架提出一种全新的 Masked-Attention 机制,通过考虑输入特征图的上下文信息和目标掩码,从而有选择性地关注模型感兴趣的区域。Mask2Former 架构还利用多尺度特征交互策略,能够更加有效地融合不同尺度的信息;另外,在多个公开数据集上对 Mask2Former 架构进行了验证,表明其在图像分割任务中有着良好的表现。虽然 Mask2Former 架构在图像分割任务中取得了不错的效果,但在一些特定情况下存在缺点,例如分割后的精度不是特别理想、没有更好地利用图像的特征融合特点以及在像素点与像素点的边界处难以界定等,这些不足之处仍然需要做进一步的改进和优化。

本文基于 Mask2Former 提出了新的语义分割

改进框架 Mask2Former-MS 和 Mask2Former-BR。前者设计了一种改进的多尺度特征交互模块^[13],并将通道注意力机制^[14]引入到其中,能够筛选并且聚焦于对语义分割任务更重要的特征通道,减少对任务无关的冗余信息,提高模型的泛化能力和鲁棒性;后者在分割后的处理阶段设计了一个改进的边界优化模块^[15],可以更有效地区分相近或重叠的像素,从而提高整体分割的性能。最后利用 GID(Gaofen image dataset) 数据集^[16]将本文改进的架构与 Mask2Former 架构在评价指标中的交并比、精确率、准确率和 F 分数进行了对比,得到了比 Mask2Former 架构更优的分割结果,验证了其在遥感图像语义分割任务中有着更好的表现。

1 架构设计

Mask2Former 架构是图像分割任务中的一种通用方案,在语义级、实例级和全景级都取得了不错的效果,其拥有的 Transformer 解码器^[17]使用屏蔽注意力机制^[18]代替常规的交叉注意力机制^[19],使模型的训练收敛更快,同时使模型的性能得到提升;Mask2Former 架构还使用少量随机采样点计算掩码损失,节省了3倍的训练内存。与 Mask2Former 相比,改进后的 Mask2Former-MS 架构在主干特征提取后引入了改进后的多尺度特征交互模块,并插入了通道注意力机制,将特征送入到 Transformer 解码器中进行解码,得到分割结果;Mask2Former-BR 架构将解码器输出的特征经过逐级像素相乘操作,再结合特征交互处理后的部分特征,将其输入到改进的边界优化模块中,经其作用后得到精细的输出图像,其整体结构框架如图1所示。

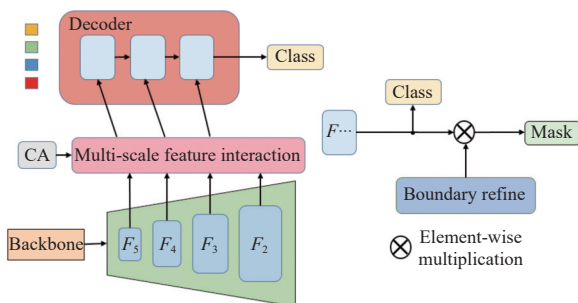


图1 整体结构框架图

Fig. 1 Framework diagram of overall structure

1.1 多尺度特征交互

在遥感图像语义分割任务中,多尺度特征交互是指利用不同尺度特征图之间的信息交互,从而

提高模型对各种尺度信息的利用能力。遥感图像通常具有不同的空间分辨率,目标类别往往受到周围环境的影响,其语义标签与其周围区域的关联性较强,通过改进的多尺度特征交互,可以更好地平衡图像的局部细节和全局一致性,确保分割结果能够准确识别到细微的目标特征。不同特征层级的特征表征信息有着一定的差异,浅层的低频特征空间信息较为丰富,但是其语义信息不够高级,特征数量较少,不利于模型的分割任务;而随着网络深度的增加,特征图的尺寸会逐渐变小,深层的高频特征空间信息就会较为欠缺,导致图像分辨率下降或者空间信息丢失。通过上采样和下采样的方式去融合不同层级的特征,从而达到特征互补的效果。多尺度特征交互的过程如式(1)所示:

$$F_{\text{fusion}} = \sum_{i=1}^N F_i \cdot \alpha_i \quad (1)$$

式中: F_{fusion} 为融合后的特征表示; F_i 为不同尺度的特征图; α_i 为对应尺度特征的权重参数。

上采样将低分辨率特征图扩大到与高分辨率特征图相同的尺寸,以便与其他尺度的特征进行交互融合;下采样将输入特征图缩小为较小尺寸,提取更加全面的重要特征信息,同时能够减少计算量和内存消耗。本文将上采样和下采样操作结合使用,如图2所示,均通过双线性插值法来实现。

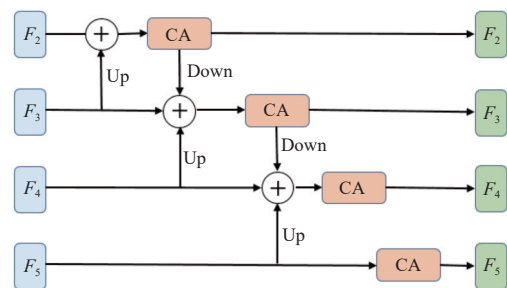


图2 多尺度特征交互结构

Fig. 2 Structure diagram of multi-scale feature interaction

双线性插值假设在一个像素点的4个邻近点之间的像素值是线性变化的,并根据这些邻近点的像素值来计算新像素点的像素值。给定一幅输入特征图 F_{in} 及其目标尺寸,对于每个目标尺寸的像素点,找到其最近的4个邻近像素点 (x_1, y_1) 、 (x_2, y_1) 、 (x_1, y_2) 和 (x_2, y_2) , 然后根据目标像素点与最近的邻近像素点之间的相对位置关系,利用双线性插值算法计算目标像素点的像素值,即:

$$F_{\text{out}}(x, y) = (1 - \alpha) \cdot (1 - \beta) \cdot F_{\text{in}}(x_1, y_1) + \alpha \cdot (1 - \beta) \cdot F_{\text{in}}(x_2, y_1) + (1 - \alpha) \cdot \beta \cdot F_{\text{in}}(x_1, y_2) + \alpha \cdot \beta \cdot F_{\text{in}}(x_2, y_2) \quad (2)$$

式中: α 和 β 为目标像素点与最近点的相对位置; $F_{\text{in}}(x_1, y_1)$ 、 $F_{\text{in}}(x_2, y_1)$ 、 $F_{\text{in}}(x_1, y_2)$ 和 $F_{\text{in}}(x_2, y_2)$ 为 4 个最近像素点的像素值。

由于图像特征通过上采样和下采样的方式进行了融合拼接, 导致其数量在一定程度上增长, 所以可能会出现冗余信息, 因此我们在多尺度特征交互模块中引入通道注意力机制。这一做法能够根据每个通道的重要性动态调整特征图的权重, 使得改进后的模型更加关注与之重要相关的特征通道, 帮助模型抑制无关或者关联很小的特征通道, 减少冗余信息的影响, 从而提高特征的表示能力。

首先, 输入的特征图 $F \in \mathbb{R}^{H \times W \times C}$ (H 和 W 分别为特征图的高度和宽度, C 为通道数), 经过一系列卷积操作得到一组特征图, 接着对每个通道 c 的特征进行全局池化, 得到每个通道对应的全局信息向量, 其表达形式如式 (3) 所示:

$$z_c = \text{AvgPool}(F[:, :, c]) \quad (3)$$

利用全局信息向量, 计算得到每个通道的权重, 此举用来调整特征图中每个通道的重要程度, 计算方法如式 (4) 所示:

$$\alpha_c = \sigma(W_z \cdot z_c) \quad (4)$$

式中: σ 为激活函数; W_z 为参数矩阵。把每个通道的权重应用到与之对应的特征图中, 得到加权特征图 F' , 特征通道加权的计算方法为

$$F'^{[:, :, c]} = \alpha_c \cdot F[:, :, c] \quad (5)$$

最后进行特征融合, 得到最终的特征表示 F_{end} , 其计算方法如式 (6) 所示:

$$F_{\text{end}} = \text{Cat}(F'^{[:, :, 1]}, F'^{[:, :, 2]}, \dots, F'^{[:, :, C]}) \quad (6)$$

式中: Cat 为通道的拼接操作。通道注意力机制通过学习每个通道的权重, 使网络模型能够自适应地调整特征图中通道的重要程度, 从而提高语义分割任务的性能和效果。

1.2 边界优化

在语义分割任务中, 边界处的像素点往往包含了重要的语义信息, 分割目标通常具有复杂的边界结构, 一般的语义分割模型在处理边界时往往存在像素点模糊、断裂等问题, 导致分割边界的精度较低。因此, 本文引入改进的边界优化模块能

够增强边界锐利度并减少边界错分, 从而提高分割边界的准确性和连贯性, 同时提升图像的视觉感知效果。

在边界优化模块中, 高级特征的特征数量较多, 有利于边界像素特征的区分, 这是因为网络架构通常通过多层堆叠来扩展感受野, 意味着在更加深层的网络中提取的特征能够覆盖更广泛的空间上下文信息, 从而能更全面地理解像素周围的语义信息。但是相对于其他特征而言, 高级特征的空间信息较为缺少, 所以在对边界处的像素进行预测时, 通过与其他层级的低级特征交互作用, 提供更加丰富的空间信息来提升边界的预测效果。如图 3 所示, 高级特征 F_5 通过空洞卷积金字塔池化模块 (atrous spatial pyramid pooling, ASPP)^[20] 进行特征提取, ASPP 模块一般由几个并行的空洞卷积分支构成, 每个分支使用不同的空洞率来捕捉不同尺度上的特征信息, 同时, ASPP 模块中还有一个全局池化分支^[21], 用于捕捉全局上下文信息。空洞卷积操作的数学表达形式为

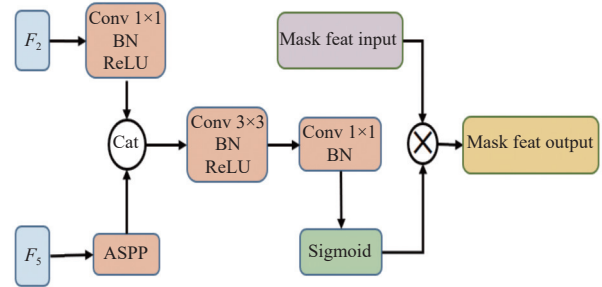


图 3 边界优化结构

Fig. 3 Structure diagram of boundary optimization

$$y_i = \sigma(W_i * F) \quad (7)$$

式中: σ 与 F 同上文一样, 分别为激活函数和输入特征图; W_i 为第 i 个卷积核; $*$ 为卷积操作。经过 ASPP 模块处理后的高级特征与低级特征拼接, 通过 ReLU 激活层^[22] 作用于提升模型的非线性表达能力与 BN(batch normalization) 层^[23] 对特征通道的归一化处理, 抑制梯度消失的现象, 计算方法如式 (8)、式 (9) 所示:

$$\text{ReLU}(x) = \max(0, x) \quad (8)$$

$$\text{BN}(x) = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \cdot \gamma + \beta \quad (9)$$

式中: μ 为特征均值; σ 为特征标准差; γ 和 β 为可学习的缩放和偏移参数; ε 为一个极小的常量, 防止分母为零, 使得每层的输入特征具有相对一致

的分布,利于网络模型的训练。在 ReLU 层,将所有不大于 0 的值设置为 0,而大于 0 的值保持原值,这样保留了正数部分的特征,抑制负数部分的特征;在 BN 层,对每个特征通道进行归一化操作,使特征分布接近于标准正态分布,经过 BN 层后,利用作用于输出层的 Sigmoid 函数^[24],将模型的特征输出转换为 0 至 1 范围内的概率值,有利于判断像素点分割的结果,即:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

式中: x 为输入的边界特征。最后将经过 Sigmoid 激活后的边界特征与解码器输出的掩码特征进行逐元素相乘,将边界特征与掩码特征进行有效的信息融合,可降低不相关的信息对边界预测的影响,使得边界处的像素更加准确。

2 实验方案

2.1 数据集

本文采用的数据集是由高分 2 号卫星 (GF-2) 提供的。高分 2 号卫星具有较高的空间分辨率,能够提供高质量的地球观测数据;同时,它还有多个波段的传感器,用来捕获不同波段的地表信息,包括可见光、红外线等。高分 2 号卫星具有广泛的

应用前景,成为了中国遥感卫星系统中的重要数据资源。

GID 数据集是由武汉大学童心仪等人利用高分 2 号卫星开发的一套土地覆盖分类并且具有较好注释的数据集。GID 数据集由两部分组成,一个大规模的土地覆盖分类集和一个精确的土地覆盖分类集,本文使用了大规模的土地覆盖分类集。大规模土地覆盖分类集包含了 150 张像素级别注释的 GF-2 遥感图像,光谱信息为 RGB 可见光波段,其中 R 代表红色, G 代表绿色, B 代表蓝色。数据集分为了 6 个类别的训练和验证图像,分别为建设用地、森林、耕地、草地、水域和其他类别,其语义信息的注释见表 1,实验使用的遥感图像训练集示例如图 4 所示。

表 1 语义信息注释

Table 1 Semantic information of labels		
Class	No.	Label color (R, G, B)
Built-up	0	(255, 0, 0)
Forest	1	(0, 255, 255)
Farmland	2	(0, 255, 0)
Meadow	3	(255, 255, 0)
Water	4	(0, 0, 255)
Others	5	(0, 0, 0)



图 4 遥感图像训练集示例

Fig. 4 Example of training set for remote sensing images

2.2 实验平台及参数设置

本实验的数据训练和预测都在版本号为 20.04.5 LTS 的 Ubuntu 操作系统上进行,建立在 Python3.8+Pytorch2.0+Cuda11.8 的虚拟环境上,并使用 PyCharm2020.1 编辑器进行编码,利用 ArcGIS-10.8 软件对遥感影像数据进行裁剪等数据处理操

作。在硬件设施方面,显卡使用的是 Nvidia GeForce RTX 3090,为实验顺利开展提供了有力的保障。

实验将数据集按照 7 : 3 的比例划分,即 22050 张图像作为训练集,9450 张图像作为验证集。学习率是控制模型参数在训练过程中的更新速度的一个超参数,根据以往经验设置学习率为 0.0005

时,模型收敛速度缓慢,且易陷入局部最优解;设置学习率为 0.005 时,训练过程中出现梯度爆炸。最终设置合适的学习率为 0.001,实验迭代次数为 40,批尺寸为 4,优化器为 Adam,在计算梯度的一阶矩估计和二阶矩估计时使用了指数加权移动平均。

2.3 评价指标

本实验使用常见的语义分割指标来对架构的性能进行全方位的评价,其使用到的评价指标有交并比(intersection over union, IoU)^[25]、精确率(Precision)、准确率(Accuracy)和 F 分数(F -Score)。

IoU 在深度学习的目标检测和语义分割中是十分重要的一项指标,它定量分析了不同图像相互重叠的程度,能够测量出真实值与预测值之间的相关性。IoU 的值越大,说明其相关性越好,通常情况下, IoU 的值大于 0.5^[26],就能说明网络结构具有较好的分割性能。IoU 的计算方法如式 (11) 所示:

$$I = \frac{\sum t_p}{\sum f_n + \sum t_p + \sum f_p} \quad (11)$$

式中: t_p 为预测结果中正确分类为该类别的像素点的数量; f_n 为预测结果中本属于该类但实际为其他类别的像素点的数量; f_p 为预测结果中属于其他类但实际分为了该类的像素点的数量。

精确率是深度学习中一种常见的评价指标。具体来说,它是真正类与真正类和假正类之和的比例,即反映了算法模型预测出的属于该类的像素点中实际属于该类的像素点个数所占的比例,

其表达式如式 (12) 所示:

$$P = \frac{\sum t_p}{\sum t_p + \sum f_p} \quad (12)$$

准确率是一种用于衡量模型能够成功识别正确类别样本能力的评价指标,它是指在所有的像素点中,成功预测为该类的像素点的占比,数学表达形式如式 (13) 所示:

$$A = \frac{\sum t_p + \sum t_n}{\sum t_p + \sum t_n + \sum f_p + \sum f_n} \quad (13)$$

在深度学习中, F -Score($F1$ -Score) 也叫 F 分数,是一种综合考虑了模型的精确率和召回率的评价指标,平衡了模型在正样本和负样本方面的表现,可以用来衡量模型在分割任务中的整体性能。一般来说,其值越高表示模型的性能越好, F 分数的计算方法如式 (14) 所示:

$$F = 2 \times \frac{P \cdot R}{P + R} \quad (14)$$

式中 R 为模型评价指标中的召回率。

3 实验结果与分析

本文利用 MMSegmentation 中的 Mask2Former 架构,与在此基础上引入改进的多尺度特征交互架构 Mask2Former-MS 和改进的优化边界架构 Mask2Former-BR,对 GID 数据集进行遥感图像语义分割,得到分割后的可视化影像,原始影像、原始影像对应的标签可视化图及各个架构得到的分割结果和细节展示分别如图 5 和图 6 所示。

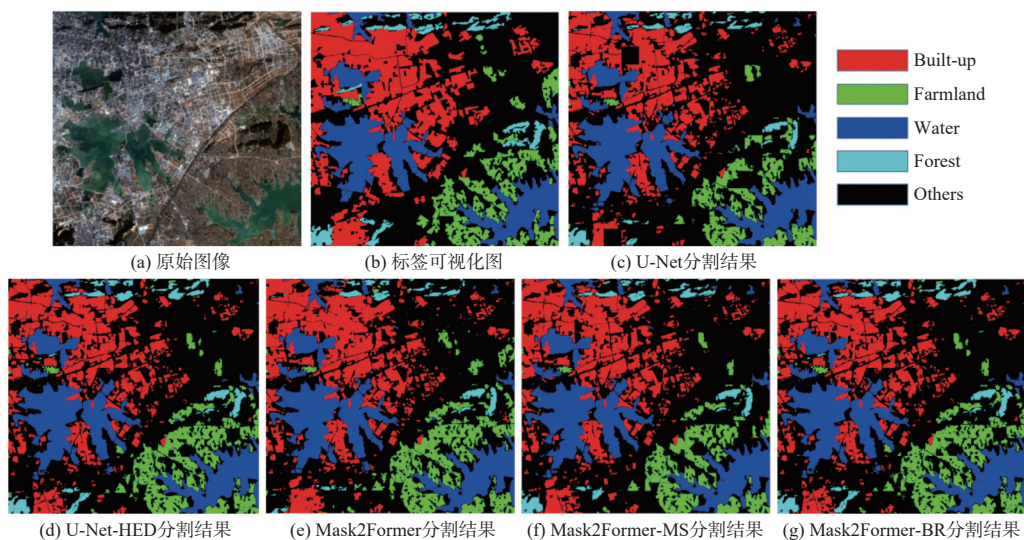


图 5 分割结果对比

Fig. 5 Comparison of segmentation results

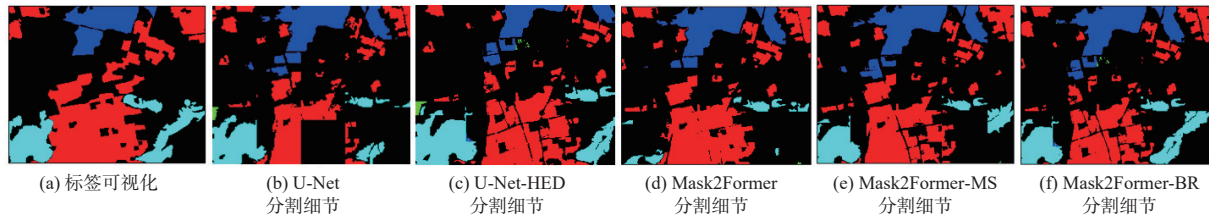


图 6 细节对比

Fig. 6 Comparison of details

由图 5 和图 6 可以看出,与 U-Net、U-Net-HED 和 Mask2Former 架构得到的分割结果进行比较,改进后的 Mask2Former-MS 架构与 Mask2Former-BR 架构得到的分割结果更接近于原始图像对应的标签可视化图像。其中,Mask2Former-BR 架构得到的分割结果视觉感知更加良好,是因为经过改进的边界优化模块处理得到的像素点类型更加准确,被错误分割的像素点数量减少,使得像素分割更加准确。

实验对 5 类土地利用类型的各项指标进行了记录和对比,对比结果如表 2 所示。由表中数据可以看到,在运用本文改进的 Mask2Former-MS 架构和 Mask2Former-BR 架构后,5 种地类的各项评价指标相较于 U-Net、U-Net-HED 和 Mask2Former 架构都得到了不同程度的提升。其中,在具体的地类方面,Mask2Former-MS 和 Mask2Former-BR 架构

在建筑类的 IoU 相对于 U-Net、U-Net-HED 和 Mask2Former 架构分别有 7.1%、7.5%、4.9%、5.4% 和 3.3%、3.8% 的提升;在水体类的准确度 Accuracy 分别有 9.8%、1.9%、8.6%、8.8% 和 11.0%、2.0% 的提高;在森林类的精确度 Precision 分别有 10.9%、1.9%、7.9%、10.8% 和 13.7%、4.8% 的提升;在草地类的 IoU 分别有 8.0%、2.6%、6.4%、6.6% 和 8.2%、2.8% 的提升;在耕地类的 F 分数分别有 9.2%、2.7%、6.8%、6.9% 和 9.3%、2.8% 的提升。以上数据指标的提升充分说明了引入通道注意力机制改进的多尺度特征交互模块能更准确地捕捉到一些类别的细节和特征,优化边界模块的改进使模型在得到分割结果时能够更准确地划分不同类别像素之间的边界,减少了类别之间的混淆,提高了遥感图像语义分割的效果。

表 2 各架构评价指标

Table 2 Evaluation indicators for each structure

Structure	Evaluation	Built-up	Water	Forest	Meadow	Farmland
U-Net	IoU	0.603 2	0.816 5	0.626 9	0.714 8	0.632 7
	Accuracy	0.667 4	0.835 8	0.805 2	0.771 5	0.706 3
	Precision	0.823 6	0.839 1	0.702 2	0.783 5	0.816 4
	F -Score	0.673 7	0.848 9	0.734 6	0.786 8	0.744 1
U-Net-HED	IoU	0.624 9	0.830 8	0.648 3	0.730 5	0.659 1
	Accuracy	0.690 6	0.857 6	0.822 3	0.783 2	0.718 8
	Precision	0.851 3	0.876 6	0.731 7	0.807 4	0.851 9
	F -Score	0.696 4	0.874 9	0.759 7	0.804 4	0.768 4
Mask2Former	IoU	0.641 1	0.898 3	0.685 5	0.769 4	0.693 9
	Accuracy	0.710 0	0.925 1	0.881 8	0.855 6	0.744 6
	Precision	0.938 0	0.968 2	0.792 0	0.882 0	0.903 1
	F -Score	0.778 7	0.947 1	0.794 3	0.856 2	0.809 3
Mask2Former-MS (Ours)	IoU	0.674 4	0.924 0	0.706 9	0.795 3	0.718 3
	Accuracy	0.741 5	0.944 4	0.903 9	0.897 9	0.879 0
	Precision	0.957 8	0.988 6	0.811 1	0.911 1	0.925 3
	F -Score	0.808 4	0.960 5	0.821 4	0.886 0	0.836 5
Mask2Former-BR(Ours)	IoU	0.678 7	0.933 0	0.713 6	0.796 7	0.720 2
	Accuracy	0.742 6	0.945 3	0.941 1	0.893 4	0.880 9
	Precision	0.963 5	0.982 9	0.839 9	0.924 5	0.925 5
	F -Score	0.811 4	0.959 9	0.816 4	0.874 3	0.837 3

在模型的训练过程中,每一次迭代的各项指标的平均值及损失曲线如图7和图8所示。

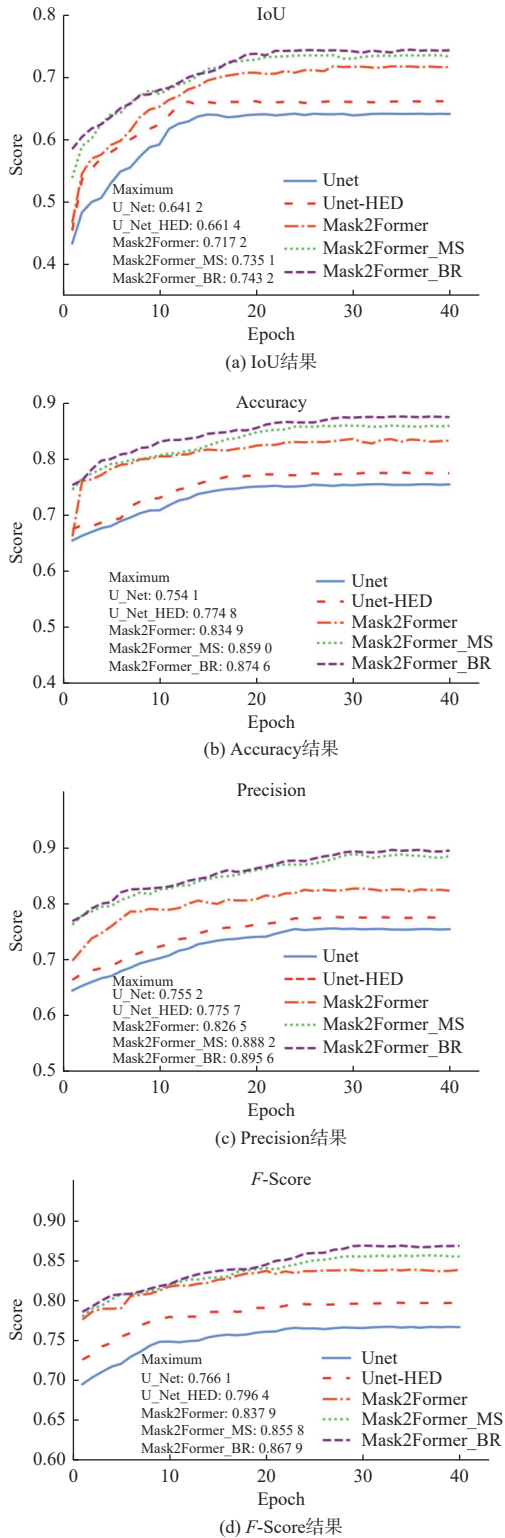


图7 实验结果

Fig. 7 Experimental results

由图7分析得到,5种架构的IoU和Accuracy指标基本呈现先上升后稳定的趋势,而运用U-

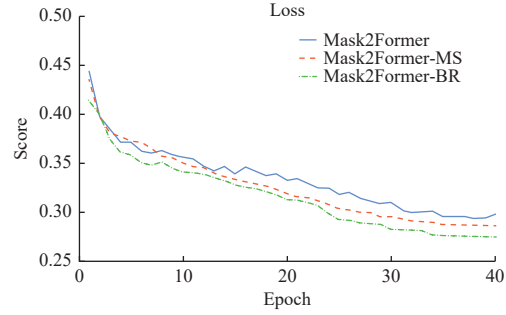


图8 损失曲线

Fig. 8 Loss curves

Net、UNet-HED 和 Mask2Former 架构进行语义分割任务时,模型的初始状态可能不太理想,在第1次迭代时没有学习到足够的特征,导致了其值出现较低的情况;而当插入通道注意力机制、改善多尺度特征交互模块,并运用 Mask2Former-MS 架构后,这种情况就大大改善了;在此基础上插入改进后的边界优化模块,使得 Mask2Former-BR 架构的评价指标变化趋势更为平滑,提高了架构的稳定性和可靠性。通过架构的指标对比分析,发现在 GID 数据集上进行语义分割时,Mask2Former-BR 和 Mask2Former-MS 架构的表现最好,且在两者比较时,Mask2Former-BR 架构稍占优势,Mask2Former 和 U-Net 架构次之。通过进行几种架构评价指标最优值的比较,发现其在 Mask2Former-MS 和 Mask2Former-BR 架构上都有不同程度的提高,提高程度在 U-Net 架构上最为明显,在 Mask2Former 原架构上次之。

在深度学习语义分割任务中,损失(loss)是用来衡量模型架构输出的预测结果与实际标签之间的距离或差异程度。由图8可以看到,通过对比 Mask2Former、Mask2Former-MS 和 Mask2Former-BR 架构的训练损失,改进后的架构的损失都有着一一定的降低。具体来说,Mask2Former-MS 架构减少了由于特征表示不准确而导致的 loss 波动。Mask2Former-BR 架构更稳定地划分不同类别之间的边界,使得损失函数的曲线更加平缓,进一步提高了 loss 的稳定性。

4 结论

本文基于 Mask2Former 的基础设计了一种引入通道注意力机制的改进版多尺度特征交互架构,利用双线性插值算法进行图像特征的上采样和下采样操作,以达到特征互补和融合的效果,从

而获得更准确、更丰富的特征表示,而通道注意力机制能够动态调整特征的权重,减少冗余信息的影响;另外基于 Mask2Former 设计了一种在分割后处理阶段的改进的边界优化架构,使在边界处不同类别的像素点的分割结果更加准确。实验结果表明:1) 与 Mask2Former 架构相比,Mask2Former-MS 和 Mask2Former-BR 架构在 GID 数据集上有着更良好的语义分割表现;2) 多尺度特征交互有助于对特征信息进行融合,引入通道注意力机制能够对特征进行筛选,调整特征的重要程度;3) 优化边界是分割后对边界处的像素进行更为精细分类的一种方法,利用 ReLU 激活函数和 BN 层能够增强模型的非线性表达能力及训练的稳定程度。

本文在遥感图像语义分割的任务中取得了一定的精度提升,但是其提升效果还有一定的上升空间。另外,实验利用 GID 数据集取得了较好的结果,但仅使用单一数据集可能会限制方法的通用性与适用性,后续工作考虑增加更多种类的数据集对方法进行验证说明,例如考虑使用卫星影像和光学影像的多源融合数据集^[27],从而进一步提升架构的鲁棒性。

参考文献:

- [1] 孙乔, 吴佳伟. 高精度测绘在城市规划中的作用[J]. 城市建设理论研究, 2024(11): 175-177.
SUN Qiao, WU Jiawei. The role of high-precision mapping in urban planning[J]. The Oretical Research in Urban Construction, 2024(11): 175-177.
- [2] RAJABI R, ZEHTABIAN A, SINGH K D, et al. Hyperspectral imaging in environmental monitoring and analysis[J]. Frontiers in Environmental Science, 2023, 19: 885-894.
- [3] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18-28.
- [4] 柳秀山, 彭炜锋, 蔡君. 支持向量机的高分辨率光学遥感影像道路提取研究[J]. 激光杂志, 2021, 42(4): 72-75.
LIU Xiushan, PENG Weifeng, CAI Jun. Research on road extraction from high-resolution optical remote sensing image based on support vector machine[J]. Laser Journal, 2021, 42(4): 72-75.
- [5] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5-32.
- [6] 梁锦涛, 陈超, 张自力, 等. 一种融合指数与主成分分量的随机森林遥感图像分类方法[J]. 自然资源遥感, 2023, 35(3): 35-42.
LIANG Jinchao, CHEN Chao, ZHANG Zili, et al. A random forest-based method integrating indices and principal components for classifying remote sensing images[J]. Remote Sensing for Nature Resources, 2023, 35(3): 35-42.
- [7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 3431-3440.
- [8] 沈言善, 王阿川. 基于深度学习的遥感图像地物分割方法[J]. 液晶与显示, 2021, 36(5): 733-740.
SHEN Yanshan, WANG Achuan. Remote sensing image feature segmentation method based on deep learning[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(5): 733-740.
- [9] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [10] 关世杰, 刘继豪, 姜月秋. 基于改进 SegNet 网络的遥感图像语义分割方法研究[J]. 兵器装备工程学报, 2023, 44(7): 237-242.
GUAN Shijie, LIU Jihao, JIANG Yueqiu. Research on semantic segmentation algorithm of remote sensing images basedon improved SegNet[J]. Journal of Ordnance Equipment Engineering, 2023, 44(7): 237-242.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Cham: Springer International Publishing, 2015: 234-241.
- [12] CHENG B, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 1290-1299.
- [13] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2881-2890.

- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
- [15] YUAN Y, XIE J, CHEN X, et al. Segfix: model-agnostic boundary refinement for segmentation[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XII 16. Cham: Springer International Publishing, 2020: 489-506.
- [16] TONG X Y, XIA G S, LU Q, et al. Land-cover classification with high-resolution remote sensing images using transferable deep models[J]. *Remote Sensing of Environment*, 2020, 237: 111322.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 13756489.
- [18] ABDULLA W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow[EB/OL]. [2024-07-01] https://github.com/matterport/Mask_RCNN.
- [19] LIU M, YIN H. Cross attention network for semantic segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP). New York: IEEE, 2019: 2434-2438.
- [20] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[EB/OL]. (2014-11-22)[2024-07-01]. <https://arxiv.org/abs/1412.7062?file=1412.7062>.
- [21] LIN M, CHEN Q, YAN S. Network in network[EB/OL]. (2013-12-16)[2024-07-01]. <https://arxiv.org/abs/1312.4400>.
- [22] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). [S. l.]: ACM, 2010: 807-814.
- [23] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. [S. l.]: PMLR, 2015: 448-456.
- [24] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834-848.
- [25] REZATOFIGHI H S, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 658-666.
- [26] 张哲晗, 方薇, 杜丽丽, 等. 基于编码-解码卷积神经网络的遥感图像语义分割[J]. *光学学报*, 2020, 40(3): 46-55. ZHANG Zhehan, FANG Wei, DU Lili, et al. Semantic segmentation of remote sensing image based on encoder-decoder convolutional neural network[J]. *Acta Optica Sinica*, 2020, 40(3): 46-55.
- [27] ZHAO J, ZHANG D, SHI B, et al. Multi-source collaborative enhanced for remote sensing images semantic segmentation[J]. *Neurocomputing*, 2022, 493: 76-90.