

基于改进 AdvSemiSeg 的半监督遥感影像作物制图方法

翟雪东¹ 韩文霆^{1,2} 马伟童³ 崔欣¹ 李广⁴ 黄沈锦⁵

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100; 2. 西北农林科技大学中国旱区节水农业研究院, 陕西杨凌 712100;
3. 西北农林科技大学水利与建筑工程学院, 陕西杨凌 712100; 4. 中国科学院重庆绿色智能技术研究院, 重庆 400714;
5. 哈尔滨工业大学计算学部, 哈尔滨 150001)

摘要: 作物精准遥感制图对于农业资源调查与管理具有重要意义。深度学习为实现精准高效作物制图提供了技术支持。为了缓解深度学习对标记样本的依赖,本文提出了一种改进 AdvSemiSeg 的半监督遥感影像作物制图方法。所提方法引入 STMF-DeepLabv3+ 作为对抗学习中的生成网络,通过 Swin Transformer (ST) 和多尺度特征融合 (Multi-scale fusion, MF) 模块提高生成网络特征编码能力和语义表达能力,改善遥感影像作物分割效果;此外,在判别网络中引入通道注意力 (Efficient channel attention, ECA) 模块,对不同通道特征图的表征信息进行自适应学习,增强判别网络对不同通道特征的感知能力。模型训练过程中,判别网络为生成网络提供高质量的伪标签和对抗损失,有效提高生成网络的泛化能力。采用所提方法与几种先进的半监督语义分割方法对内蒙古河套灌区遥感影像种植信息进行提取,本文方法性能最优。

关键词: 遥感; 作物制图; 半监督学习; 生成对抗网络; 多尺度特征融合; 通道注意力

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-1298(2024)08-0196-09

OSID:



Semi-supervised Network for Remote Sensing Crop Mapping Based on Improved AdvSemiSeg

ZHAI Xuedong¹ HAN Wenting^{1,2} MA Weitong³ CUI Xin¹ LI Guang⁴ HUANG Shenjin⁵

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Institute of Water-saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling, Shaanxi 712100, China
3. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
4. Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China
5. Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Crop precision remote sensing mapping holds significant importance for agricultural resource surveys and management. Deep learning provides technical support for achieving accurate and efficient crop mapping. To alleviate the dependency of deep learning on labeled samples, an improved semi-supervised remote sensing crop mapping method was proposed based on AdvSemiSeg. The proposed method introduced STMF-DeepLabv3+ as the generator in the adversarial learning framework, enhancing the feature encoding and semantic expression capabilities of the generator through Swin Transformer (ST) and multi-scale fusion (MF) modules, thus improving the segmentation performance of remote sensing crop images. Additionally, the efficient channel attention (ECA) module was introduced after each convolutional layer of the discriminator to adaptively learn the representation information of different channel feature maps, enhancing the discriminator's perception of different channel features. During the training process, the discriminator provided high-quality pseudo-labels and adversarial losses to the generator, effectively improving the generalization ability of the generator. Compared with several advanced semi-supervised semantic segmentation methods, the proposed method achieved optimal performance in extracting planting information from remote sensing images in the Hetao Irrigation District

收稿日期: 2023-11-23 修回日期: 2024-02-25

基金项目: 国家重点研发计划项目(2022YFD1900802)、国家自然科学基金项目(51979233)和中央高校基本科研业务费专项资金项目(2452023078)

作者简介: 翟雪东(1997—),女,硕士生,主要从事遥感监测研究,E-mail: zhaixd0000@nwafu.edu.cn

通信作者: 韩文霆(1972—),男,研究员,博士生导师,主要从事农情信息智能感知及精准灌溉技术研究,E-mail: hanwt2000@126.com

of Inner Mongolia.

Key words: remote sensing; crop mapping; semi-supervised learning; generative adversarial network; multi-scale features fusion; channel attention

0 引言

作物制图是农业遥感和地理信息科学领域的一项重要应用^[1-3], 卫星遥感和深度学习算法的快速发展, 为实现大范围作物制图提供了重要技术支撑^[4]。目前利用深度学习算法对遥感影像作物制图的相关研究主要采用全监督训练范式^[5-6], 此类方法需要大量标记训练数据^[7-8], 而人工标注训练数据成本高, 在实际遥感应用场景中很难满足这一要求。相对于全监督学习, 半监督学习方法根据标记数据, 挖掘未标记数据的有效信息, 可达到降低人工成本和提升模型性能的目的^[9-11]。

现有半监督学习方法大多采用自训练^[12]、一致性正则化^[13-14]、对抗训练^[15]等策略, 其中, 通过对抗训练利用未标记数据是半监督学习的研究热点。对抗训练起源于生成对抗网络 (Generative adversarial network, GAN)^[16-17], GAN 分为生成网络和判别网络, 生成网络负责生成高质量的虚假图像以欺骗判别网络, 判别网络则需要判断出真实图像和虚假图像。训练过程中, 生成网络和判别网络相互博弈, 逐渐提高模型整体性能。对抗训练思想在半监督语义分割方法设计上具有重要的启发性, 例如 AdvSemiSeg^[18]采用不同分割网络作为生成网络和判别网络, 生成网络接收有标签和无标签的训练样本, 输出预测概率图; 判别网络接收有标签样本的预测概率图和真实标签图, 输出图像像素来自真实标签的置信度。在 AdvSemiSeg 模型参数优化过程中, 生成网络与判别网络对抗训练, 一方面鼓励生成网络可以输出接近真实标签图的预测概率分布以欺骗判别网络; 另一方面不断提高判别网络的鉴别能力, 以期从判别网络输出的置信图中选择高置信度像素赋予伪标签, 为生成网络提供无标签样本的监督损失, 挖掘无标签样本的图像信息。

AdvSemiSeg 在自然图像半监督语义分割任务中取得了显著效果, 许多研究将该方法的设计思路应用于遥感领域。耿艳磊等^[19]将 UNet 和 FCN 分别作为生成网络和判别网络, 利用生成网络对遥感影像进行地物 (建筑、地表、低矮植被、树木、汽车) 分割, 证明了对抗训练策略对该遥感任务的有效性。WANG 等^[15]提出了一种多尺度对抗学习方法用于改善遥感影像小目标地物识别效果, 该方法的生成网络接收 3 种不同像素尺寸 (1、0.75、0.5 倍) 的输

入图像, 学习遥感影像地物的多尺度特征。此外, CHEN 等^[20]设计了一种应用于遥感影像道路分割任务的 SemiRoadExNet 方法, 与 AdvSemiSeg 不同的是, 该方法的生成网络输出类别概率图和熵图, 其判别网络也分为类别概率判别网络和熵判别网络, 通过增加生成网络与熵判别网络的对抗训练过程, 提高道路分割精度。逯焕宇等^[21]提出了一种 SWTUNet 方法用于遥感影像水体分割, 该方法将 Swin Transformer (ST) 引入生成网络用于改善水体提取效果。这些研究表明, 诸如 AdvSemiSeg 的对抗训练策略在遥感影像各种分割应用中具有广泛适用性。

为了缓解深度学习遥感影像作物制图对标记样本的数量要求, 本文基于对抗训练策略提出一种改进 AdvSemiSeg 的半监督语义分割方法。对于生成网络, 设计一种 STMF-DeepLabv3+ 结构, 将 Swin Transformer 作为生成网络的骨干网络, 建模遥感图像中的长距离依赖关系, 以提高生成网络的特征编码能力。为进一步利用骨干网络不同阶段特征图的语义信息, 设计一个自下而上的逐阶段多尺度特征融合 (Multi-scale fusion, MF) 模块, 以增强生成网络语义表达能力。对于判别网络, 之前大多研究主要采用 4 个卷积层的 FCN 实现。在对抗训练过程中, 判别网络接收生成网络预测的类别概率图或训练样本的真实标签图, 仅有单通道信息的真实标签图经过 4 个卷积层的通道扩增和压缩处理后, 可能存在一定的信息损失, 最终影响判别网络对输入来源的正确判断。为此, 将通道注意力机制引入判别网络, 以提高判别网络对关键通道特征的感知能力, 避免对抗训练时产生的信息损失问题。通过优化 AdvSemiSeg 的生成网络和判别网络, 以期提升遥感影像作物制图精度。

1 研究区与数据源

1.1 研究区概况

选取内蒙古河套灌区内 5 个旗县 (磴口县、杭锦后旗、临河区、五原县、乌拉特前旗乌梁素海以西) 为研究区域。如图 1 所示, 河套灌区 (40.1°~41.4°N, 106.1°~109.4°E) 东西长 270 km, 总面积超过 1.18 × 10⁶ hm², 现有耕地 7.337 × 10⁵ hm², 主要种植玉米、西葫芦、向日葵、小麦等^[22]。该区复种指数相对较小, 大多数耕地一年一作, 作物生长季节为 4 月至 9 月 (图 2)。

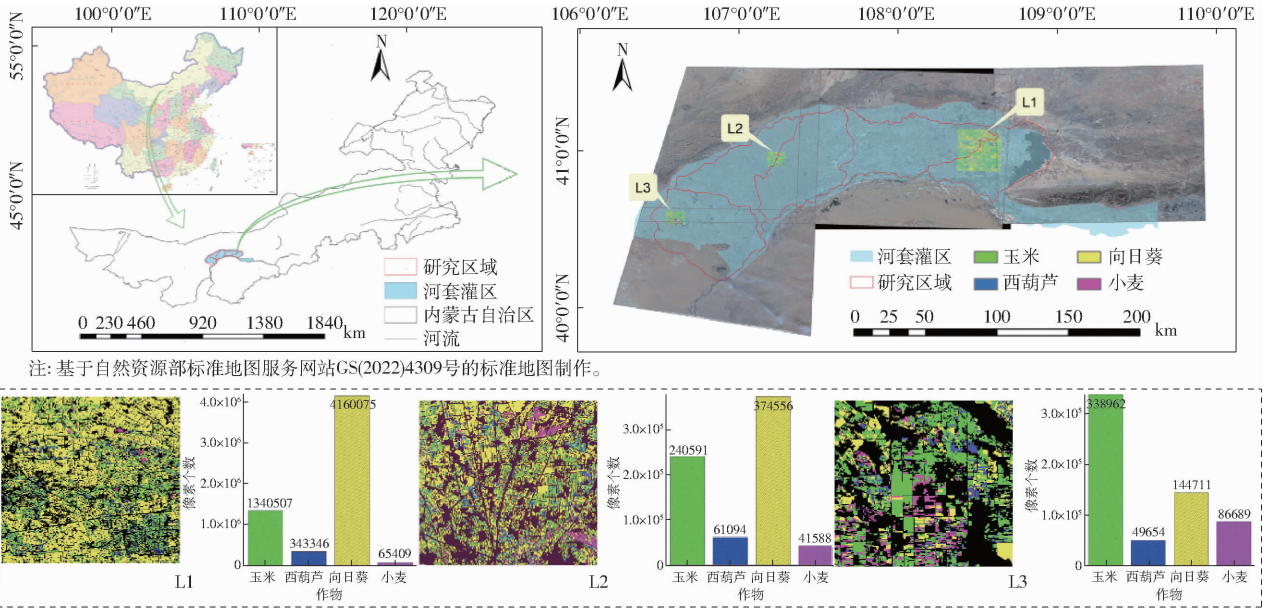


图 1 研究区域概况

Fig. 1 Overview of study area

作物	4月			5月			6月			7月			8月			9月		
	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下	上	中	下
玉米			播种		出苗			拔节期			抽穗			乳熟期			成熟期	
向日葵						播种			幼苗期			现蕾		开花			成熟	
西葫芦				播种		出苗			幼苗期		开花		结瓜		晾晒			
小麦	播种	出苗	分蘖		拔节			抽穗		乳熟期		成熟						

图 2 河套灌区主要作物物候期

Fig. 2 Phenological calendar of main crops in Hetao Irrigation District

1.2 Sentinel-2 影像数据集

本研究使用 Sentinel-2 L1C 数据,根据玉米、西葫芦、向日葵和小麦作物的物候期数据,在美国地质调查局地球探测器数据门户网站下载 2022 年 4—9 月包含研究区域的 Sentinel-2 数据(每月一景,时间点分别为 4 月 21 日、5 月 26 日、6 月 30 日、7 月 25 日、8 月 24 日和 9 月 23 日)。此后对下载影像进行大气校正(Sen2Cor)、重采样(SNAP)、波段组合(ENVI 5.6)、镶嵌(ENVI 5.6)、图像配准(ENVI 5.6)和研究区裁切(Argis 10.8)预处理操作,所有影像投影为 UTM/WGS84。预处理过程中移除对地表作物监测作用小的卷云波段,共获取 72 波段的多时相 Sentinel-2 遥感影像数据。依据研究团队 2022 年河套灌区野外实地调研结果,在 Argis 10.8 上标注图 1 中 L1、L2、L3 区域的作物,包含背景、玉米、西葫芦、向日葵和小麦 5 种类别。在制作 Sentinel-2 遥感数据集时,按尺寸 256 像素 × 256 像素和重叠度 75% 对 L1、L2 区域的遥感影像进行裁剪,获得 3 581 幅图像。本研究将 L1、L2、L3 区域对应的图像分别作为训练集(3 190 幅)、验证集(391 幅)和测试集,同时在训练集中选取 25% 的图像作为标记样本,剩余 75% 作为未标记样本。

2 研究方法

本文提出一种改进 AdvSemiSeg 的半监督语义分割方法对遥感影像进行作物制图,该方法采用 AdvSemiSeg 作为基础框架,包括优化的生成网络和判别网络。

2.1 AdvSemiSeg 概述

AdvSemiSeg 框架如图 3 所示,主要由生成网络 $G(\cdot)$ 和判别网络 $D(\cdot)$ 构成。生成网络接收输入图像 x ,包括标记图像 x^l 和未标记图像 x^u ,输出类别概率 $G(x)$;判别网络接收生成网络的预测结果 $G(x)$ 或真实标签 y ,输出像素来自真实标签 y 的置信度。训练过程中,对于判别网络 $G(x^l)$,使用标记图像 x^l 和真实标签 y ,通过判别网络输出结果 $D(G(x^l))$ 与

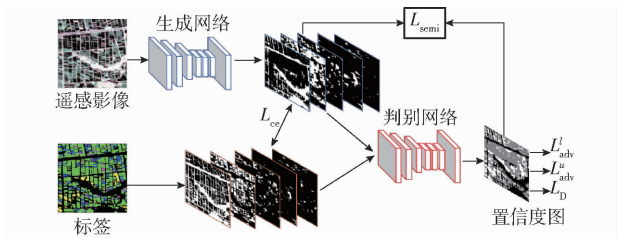


图 3 AdvSemiSeg 总体框架

Fig. 3 Overall framework of AdvSemiSeg

$D(y)$ 的二元交叉熵损失 L_D , 确保判别网络对输入来源做出正确判断。对于生成网络 $G(\cdot)$, 使用标记图像 x^l 、未标记图像 x^u 、真实标签 y , 通过多任务损失 L_G 优化网络训练, L_G 计算式为

$$L_G = L_{ce} + \lambda_{adv}^l L_{adv}^l + \lambda_{adv}^u L_{adv}^u + \lambda_{semi} L_{semi} \quad (1)$$

式中 L_{ce} ——标记图像交叉熵损失

L_{adv}^l ——标记图像对抗损失

L_{adv}^u ——未标记图像对抗损失

L_{semi} ——未标记图像半监督损失

$\lambda_{adv}^l, \lambda_{adv}^u, \lambda_{semi}$ ——权重因子

L_{ce} 用于约束分割结果 $G(x^l)$ 与真实标签 y 的一致性。 L_{adv}^l 和 L_{adv}^u 期望生成网络输出的概率分布接近真实标签, 使判别网络误认为样本数据的预测结果来自真实标签。 L_{semi} 旨在挖掘无标记图像高置信度像素的有效信息。为了防止模型受到初始不准确预测结果的影响, 先利用标记图像训练生成网络, 然后随机采样标记图像和未标记图像进行对抗训练, 联合更新生成网络和判别网络的模型参数。AdvSemiSeg 的对抗训练策略可以有效利用无标签样本, 提高模型泛化能力。

2.2 生成网络

生成网络用于分割遥感影像作物类型。以 DeepLabv3 +^[23] 作为生成网络的基线网络, DeepLabv3 + 是计算机视觉领域一种经典的语义分割网络, 该网络采用编码器-解码器结构。编码器包括骨干网络和空洞卷积特征金字塔 (Atrous spatial pyramid pooling, ASPP) 模块。骨干网络用于从原始图像中提取特征, 生成低级语义特征图; 而 ASPP 模块通过一系列操作 (包括 1×1 卷积, 扩张率为 6、12、18 的 3×3 空洞卷积和全局平均池化) 对低级语义特征图进行多尺度特征提取, 生成高级语义特征图。解码器则负责将低级和高级语义特征图进行通道融合, 输出分割结果。

为了提升基线网络特征编码能力, 将 Swin

Transformer^[24] 作为骨干网络。此外, 引入 MF 模块进一步提升基线网络的语义表达能力。为了区分于基线网络, 将这一改进后的生成网络命名为 STMF-DeepLabv3 +, 其结构如图 4 所示。Swin Transformer 由 4 个阶段 (Stage) 组成, 每个阶段通过窗口多头自注意力和移动窗口多头自注意力机制建立特征图的远程依赖关系, 优化特征表示。对于给定尺寸为 $H \times W$ 的输入图像, 经过 4 个阶段处理后, 分别得到尺寸为 $H/4 \times W/4$ 、 $H/8 \times W/8$ 、 $H/16 \times W/16$ 和 $H/32 \times W/32$ 的特征图 S_1 、 S_2 、 S_3 和 S_4 。这些特征图反映了遥感影像在不同层次上的语义表达, 即浅层特征图主要表达局部细节和纹理等相关信息, 而深层特征图则表达了更抽象的语义概念。

为了利用不同语义表达的特征图, MF 模块采用自下而上的逐阶段融合策略。具体而言, 先将 S_3 和 S_4 融合为 S_{34} , 再将 S_{34} 与 S_2 融合为 S_{234} , 最后将 S_{234} 与 S_1 融合为 S_{1234} 。这种自下而上的融合策略可以最大程度地保留浅层特征图的细节信息, 提高遥感影像作物田块的边缘分割效果, 而逐阶段的融合策略则有助于避免不同阶段特征图直接融合时可能存在的语义鸿沟问题。在 MF 模块处理过程中, 为了使不同阶段特征图的像素尺寸保持一致, 融合前对较低分辨率的深层特征图执行 2 倍上采样操作, 对齐较高分辨率的浅层特征图, 然后采用通道叠加方式进行融合。此外, 为了提高训练效率, 采用 Conv-BN-ReLU 层对融合前后的特征图进行通道数量减半的卷积处理, 减小 MF 模块计算量。借鉴 DeepLabv3 + 方法, 将 MF 模块输出特征图 S_{1234} 定义为低级语义特征图, 随后利用 ASPP 模块对其进行多尺度特征提取, 生成高级语义特征图。低级与高级语义特征图的融合过程以及解码器的处理过程与 DeepLabv3 + 保持一致。尽管 MF 模块和 ASPP 模块都执行了多尺度操作, 但它们的侧重有所不同。MF 模块侧重于融合不同语义表达的多尺度特征图; 而

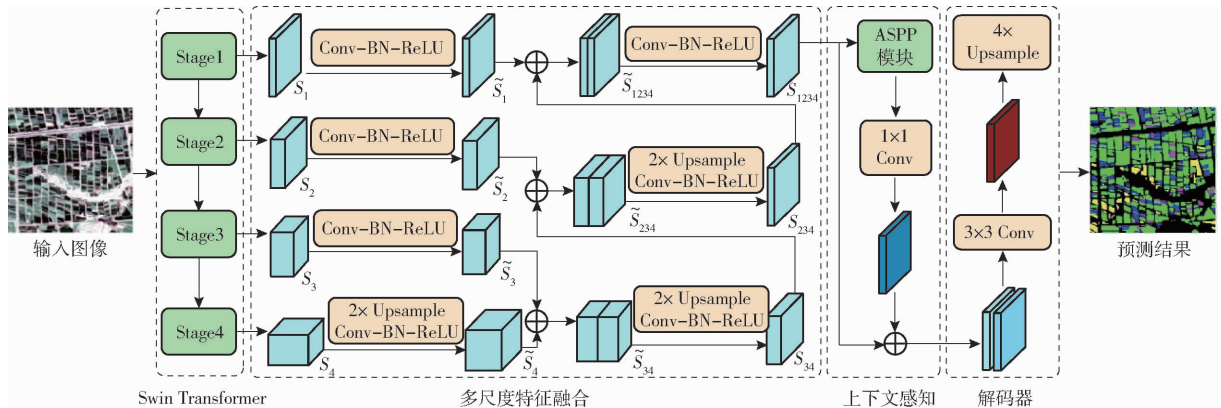


图 4 生成网络

Fig. 4 Network structure of generator

ASPP 模块侧重于在特征图中提取多尺度信息,用于改善网络的上下文感知能力。

2.3 判别网络

判别网络用于判断输入样本是否来自真实标签。本研究采用与文献[18,25]相同的全卷积神经网络作为判别网络,该网络输出与原始图像同等尺寸的置信度图。为了进一步提升判别网络性能,引入通道注意力机制对不同通道特征图的表征信息进行自适应学习,通过选择性地加权不同通道特征,优化判别网络对关键信息的感知能力。

判别网络结构如图 5 所示,主要由 4 个 Conv-ReLU-ECA 层和 1 个 Conv-Upsampling 层构成。卷积层的核大小为 4×4,步长为 2,通道数量分别为 64、128、256、512、1。在前 4 个卷积层后面均加入 ReLU 函数作为激活函数提升判别网络的非线性表达能力。随后采用通道注意力(Efficient channel attention,ECA)网络^[26]对 ReLU 函数输出的不同通道特征进行自适应学习。ECA 网络的前向计算过程如下:首先利用核大小为 $k \times k$ 的一维卷积对经过全局平均池化(Global average pooling,GAP)后的特

征图进行卷积处理;然后利用 Sigmoid(σ)函数将一维卷积的输出结果映射到 0 到 1 之间,得到每个通道的注意力分数;最后根据注意力分数对每个通道特征图进行加权处理。利用一维卷积计算注意力分数主要有 2 点优势:①没有对通道特征降维,避免了降维操作引起的信息损失。②通过计算每个通道及其 k 个相邻通道的特征来捕获局部跨通道交互信息,这种处理方式更加轻量与高效。一维卷积的核尺寸(k)根据输入的特征通道数量自适应调整,计算公式为

$$k = \left\lceil \frac{\text{lb}C + 1}{2} \right\rceil_{\text{odd}}$$

(2)

式中 $\lceil \cdot \rceil_{\text{odd}}$ ——取最近邻奇数操作

C ——通道数量

在判别网络中,最后一个卷积层融合所有通道特征图的信息,输出通道数为 1 的预测结果;上采样层将预测结果重新缩放至输入图像尺寸;Sigmoid 函数将预测结果映射至 0 到 1 之间,表示输入来自真实标签的概率。本研究采用优化的判别网络提高对输入样本来源的判断能力。

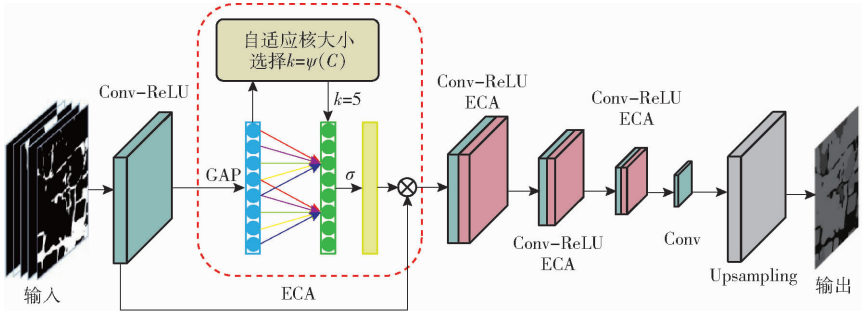


图 5 判别网络
Fig. 5 Network structure of discriminator

3 实验

3.1 实验设置

使用 PaddlePaddle 框架(2.0.4 版本)实现所提方法,在一块 GPU-NVIDIA GeForce RTX 3090 显卡上训练模型,批处理大小为 16,训练周期共 100 轮,第 50 轮后联合优化生成网络和判别网络的模型参数。训练生成网络和判别网络时,使用动量参数为 0.9、衰减系数为 4.0×10^{-5} 的随机梯度下降算法和初始学习率为 10^{-4} 的多项式衰减策略优化模型训练。对于各损失项权重因子,将 λ_{adv}^l 、 λ_{adv}^u 、 λ_{semi} 设置为 0.01、0.001、0.1。生成伪标签阈值设置为 0.2。

为了评价所提方法性能,选择一些先进的半监督语义分割算法进行比较,主要包括 ClassMix^[13]、CutMix^[14]、AdvSemiSeg^[18]、ST++^[27]、FixMatch^[28]、

UniMatch^[29]、CCT^[30]方法,其中 AdvSemiSeg 方法在原文使用 DeepLabv2 作为生成网络。为了与 STMF-DeepLabv3+ 公平比较,本研究同时采用 DeepLabv3+ 作为 AdvSemiSeg 的生成网络,该方法标记为 AdvSemiSeg*。

对于每类作物的分割结果,采用交并比(Intersection of union,IoU)进行评估;对于遥感影像的作物制图结果,采用 mIoU、F1 值和总体分类精度(Overall accuracy,OA)进行评估。

3.2 实验结果

3.2.1 不同方法分割精度比较

各方法性能评估精度如表 1 所示,在玉米、西葫芦、小麦分割结果中,所提方法表现最佳,其 IoU 分别达到 73.78%、64.53%、63.32%;在背景和向日葵的分割结果中,CutMix 方法表现最佳,所提方法达到次优水平,其 IoU 分别为 78.3%、60.43%。综

合考虑所有作物的分割性能,所提方法 mIoU、F1 值、OA 均实现了最优,依次为 68.07%、80.81%、84.82%,相比于次优的 CutMix 分别提高 1.57、1.24、0.07 个百分点。此外,对比采用不同生成网

络的 AdvSemiSeg 方法,所提方法通过优化生成网络和判别网络,大幅提高了各种作物分割精度。这些结果表明,本文方法在 Sentinel-2 遥感影像作物制图上具有明显优势。

表 1 不同方法分割精度
Tab.1 Accuracy of different methods

模型	IoU					mIoU	F1 值	OA
	背景	玉米	西葫芦	向日葵	小麦			
AdvSemiSeg	58.83	56.49	41.17	39.91	50.30	49.34	65.72	69.41
AdvSemiSeg*	65.52	63.86	52.27	41.38	55.17	55.64	71.08	74.56
ST++	60.67	64.88	47.70	37.73	40.59	50.31	66.27	71.87
FixMatch	72.60	67.65	56.63	50.58	50.38	59.57	74.26	80.22
UniMatch	73.51	69.40	54.48	56.87	51.68	61.19	75.57	81.20
ClassMix	73.34	66.60	51.07	56.00	43.68	58.14	72.96	80.09
CutMix	78.43	73.11	58.25	66.59	56.13	66.50	79.57	84.75
CCT	69.79	62.21	44.63	41.32	42.92	52.17	67.83	76.62
本文方法	78.30	73.78	64.53	60.43	63.32	68.07	80.81	84.82

3.2.2 不同方法识别效果比较

图 6 展示了各方法对不同作物的识别效果。这 4 种作物中,所提方法对玉米的识别效果最佳。玉米在遥感影像中的冠层纹理均匀,田块边缘分界线清晰,识别难度最低,其它方法对玉米的识别效果亦是如此。其次,所提方法对小麦和西葫芦的识别效果优于其它方法。这 2 种作物在研究区域内的种植比例很小,有标记训练样本的数量也相对较少。所提方法通过优化判别网络,充分利用未标记训练集中的有效信息,提高识别精度。在向日葵的识别中,CutMix 方法表现最好,所提方法次之。这可能因为向日葵抗逆性较强,可以在不同盐碱化程度

的土壤上种植,而不同土壤在卫星遥感影像中的反射率差异很大,受种植环境影响,向日葵易与其它类别混淆。CutMix 方法在训练过程中通过混合训练样本,将向日葵置于不同背景下,提高了模型的抗干扰能力。此外,从各作物田块边缘的分割效果来看,所提方法具有明显优势,这可能归因于多尺度特征融合模块能够最大化保留浅层特征图的局部细节信息,使田块分割完整。综合比较所有方法对各作物识别效果,本文方法优于其它对比方法。

3.2.3 消融实验

为了探究 STMF-DeepLabv3+ 和在判别网络中

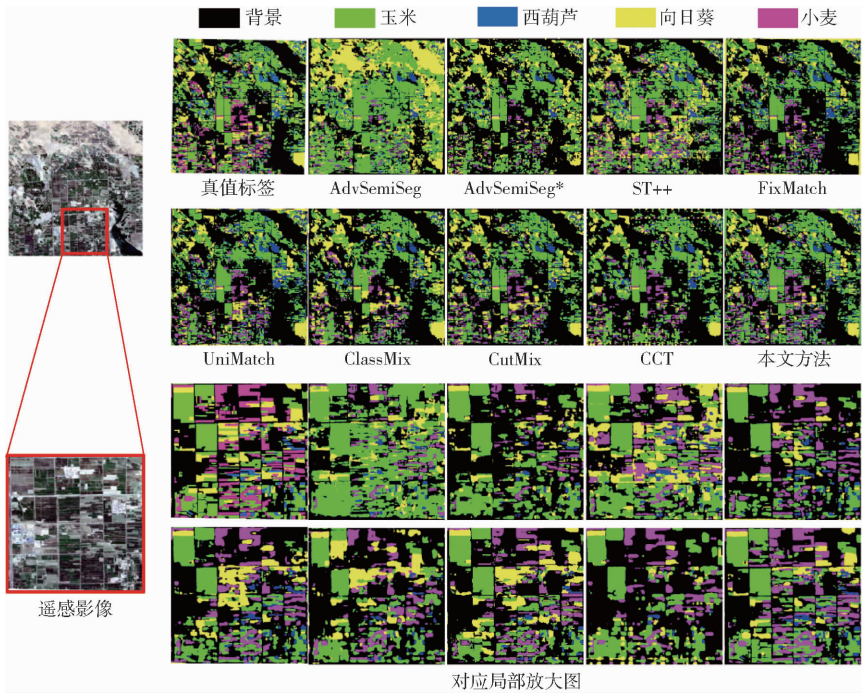


图 6 不同方法识别效果
Fig.6 Segmentation results of different methods

引入 ECA 模块对模型性能的影响,将 AdvSemiSeg^{*} 作为基线方法,在基线方法中分别引入 ST、MF 模块、ECA 模块。实验结果如表 2 所示,引入上述模块可以在不同程度上提高模型对遥感影像的制图精度。与采用 ResNet101 作为骨干网络的基线方法相比,采用 ST 进行遥感影像特征提取更具有优势,其 mIoU 提高 5.67 个百分点;MF 模块通过融合 ST 各阶段特征图,增强了网络的语义表达能力,其 mIoU 提高 3.24 个百分点。以上结果表明了 STMF - DeepLabv3 + 网络的先进性。此外,引入 ECA 模块可以增强判别网络对关键通道特征的感知能力,相比于基线方法,其 mIoU 提高 7.25 个百分点。综合引入所有模块能够实现最佳性能,实验结果表明了本研究从优化生成网络和判别网络两方面对 AdvSemiSeg 改进的有效性。

表 2 消融实验结果

Tab.2 Ablation results			%		
生成网络		判别网络	mIoU	F1 值	OA
ST	MF	ECA			
×	×	×	55.64	71.08	74.56
√	×	×	61.31	75.16	80.37
√	√	×	64.55	77.03	82.24
×	×	√	62.89	76.64	81.96
√	√	√	68.07	80.81	84.82

注:√表示添加此模块,×表示不添加此模块。

3.2.4 作物种植信息提取

将本文方法应用于河套灌区 2022 年 Sentinel - 2 遥感影像,提取各作物种植区域及面积。如图 7 所示,该灌区主要种植玉米和向日葵,这两种作物在图 7 中分布较均匀,其像元数量分别为 19 956 028、24 045 944,约 1.996×10^5 、 2.405×10^5 hm²;而西葫

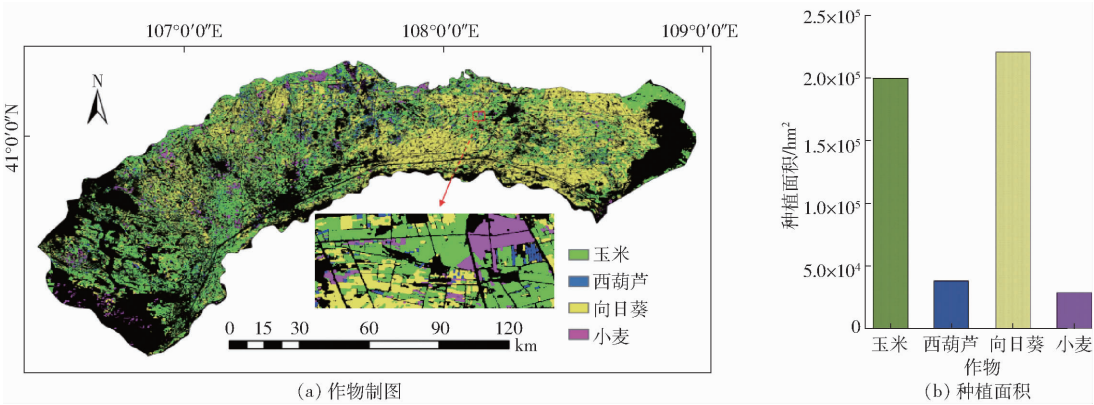


图 7 河套灌区作物种植信息
Fig. 7 Planting information in Hetao Irrigation District

芦及小麦种植面积相对较小,从遥感影像中提取的像元数量分别为 3 784 059、2 915 076,约 3.78×10^4 、 2.92×10^4 hm²,其中小麦主要种植在河套灌区的中西部区域。

4 讨论

4.1 样本尺寸对模型性能的影响

制作遥感数据集通常需要将原始影像按照一定尺寸进行裁剪,以便适应网络训练的需求。不同裁剪尺寸可能对模型性能产生不同程度影响。为了探

究该参数对 Sentinel - 2 遥感影像作物制图的影响,将裁剪尺寸分别设置为 64 像素×64 像素、128 像素×128 像素、256 像素×256 像素和 512 像素×512 像素,获得不同样本尺寸的遥感数据集,使用这些数据集训练所提方法。实验结果如图 8 所示,当样本图像尺寸为 64 像素×64 像素或 128 像素×128 像素时,模型性能基本类似。当样本尺寸从 128 像素增加到 256 像素时,模型精度明显提升,这可能因为网络能够从增大的样本图像中学习到更丰富的上下文信息。然而当样本尺寸从 256 像素增加到 512 像素

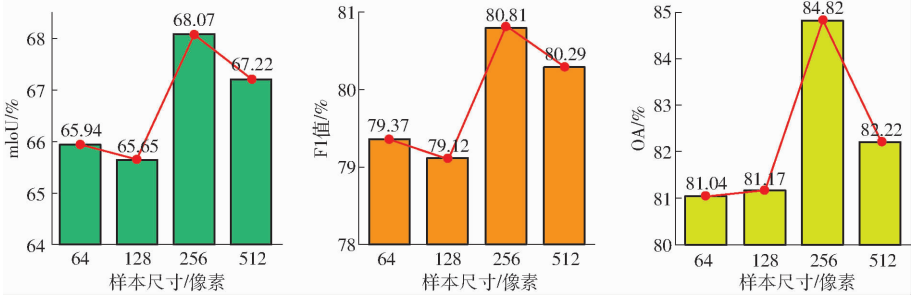


图 8 不同样本尺寸对模型性能的影响
Fig. 8 Model performance with different cropping sizes

时,模型精度轻微下降,这可能由于尺寸增大导致了数据集中有效样本数量减少,增加了模型在训练过程中出现过拟合的风险。实验结果表明,在一定范围内增加样本尺寸可以提高 Sentinel-2 遥感影像作物制图精度,其中 256 像素的样本尺寸表现最佳。

4.2 局限性与展望

基于 AdvSemiSeg 方法,对生成网络和判别网络分别进行了优化。然而在实际过程中,采用固定阈值方式筛选无标记数据高置信度像素,缺乏灵活性,在处理不同数据集时,需要进行多次调参实验才能实现较优的语义分割性能;此外,在整个训练阶段,仅使用单个阈值也不是一种最合理的方式。今后研究中,可以根据网络性能动态调整伪标签阈值,类似于 ZHANG 等^[31]提出的在线课程学习方式。其次,无标记数据低置信度的像素仍具有一些潜在和有价值的信息,这些像素可能是地物的边界组成像素^[20]。本文方法缺乏对这些像素的合理利用,如何设计一种安全的训练策略或网络架构,利用低置信度的像素信息仍是一个值得探索的研究点。综合上述,提升半监督遥感作物分类方法的性能将不仅通过增强生成网络或者判别网络实现,也可以进一步优化高置信度像素的筛选机制和提高低置信度像素

的利用率,采用分级策略对这 2 种不同置信度的像素进行训练,合理利用未标记数据的全部信息,提高遥感影像作物分割效果。

5 结束语

提出了一种改进 AdvSemiSeg 的半监督语义分割方法,用于 Sentinel-2 遥感影像作物制图任务。所提方法采用 STMF-DeepLabv3+ 作为生成网络预测作物类别,以及带有通道注意力的全卷积网络作为判别网络预测输入样本来自真实标签的置信度。在生成网络中,引入 Swin Transformer 和多尺度特征融合模块增强生成网络的特征编码能力和语义表达能力,改善遥感影像分割效果。在判别网络中,引入 ECA 模块对不同通道特征图的表征信息进行自适应学习,增强判别网络对关键通道特征的感知能力,避免真实标签在输入判别网络后出现信息损失的问题。整个训练过程中,判别网络为生成网络提供未标记数据的对抗损失和高质量伪标签,提高生成网络的泛化性能。在河套灌区作物种植信息提取上进行了相应实验,与 AdvSemiSeg、ST++、FixMatch、UniMatch、ClassMix、CutMix、CCT 半监督语义分割方法相比,所提方法性能明显提升。

参 考 文 献

[1] 陈水森,柳钦火,陈良富,等. 粮食作物播种面积遥感监测研究进展[J]. 农业工程学报, 2005, 21(6): 166-171.
CHEN Shuisen, LIU Qinhua, CHEN Liangfu, et al. Review of research advances in remote sensing monitoring of grain crop area[J]. Transactions of the CSAE, 2005, 21(6): 166-171. (in Chinese)

[2] 刘哲,刘帝佑,朱德海,等. 作物遥感精细识别与自动制图研究进展与展望[J]. 农业机械学报, 2018, 49(12): 1-12.
LIU Zhe, LIU Diyou, ZHU Dehai, et al. Review on crop type fine identification and automatic mapping using remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(12): 1-12. (in Chinese)

[3] KYERE I, ASTOR T, GRAB R, et al. Agricultural crop discrimination in a heterogeneous low-mountain range region based on multi-temporal and multi-sensor satellite data[J]. Computers and Electronics in Agriculture, 2020, 179: 105864.

[4] CASTELO-CABAY M, PIEDRA-FERNANDEZ J A, AYALA R. Deep learning for land use and land cover classification from the Ecuadorian Paramo[J]. International Journal of Digital Earth, 2022, 15(1): 1001-1017.

[5] 赵龙才,李粉玲,常庆瑞. 农作物遥感识别与单产估算研究综述[J]. 农业机械学报, 2023, 54(2): 1-19.
ZHAO Longcai, LI Fenling, CHANG Qingrui. Review on crop type identification and yield forecasting using remote sensing [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 1-19. (in Chinese)

[6] 董秀春,刘忠友,蒋怡,等. 基于 WorldView-2 影像和语义分割模型的小麦分类提取[J]. 遥感技术与应用, 2022, 37(3): 564-570.
DONG Xiuchun, LIU Zhongyou, JIANG Yi, et al. Winter wheat extraction of WorldView-2 image based on semantic segmentation method[J]. Remote Sensing Technology and Application, 2022, 37(3): 564-570. (in Chinese)

[7] COSTA H, FOODY G M, BOYD D S. Supervised methods of image segmentation accuracy assessment in land cover mapping [J]. Remote Sensing of Environment, 2018, 205: 338-351.

[8] WANG Zhimin, WANG Jiasheng, YANG Kun, et al. Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+ [J]. Computers and Geosciences, 2022, 158: 104969.

[9] REDDY Y, VISWANATH P, REDDY B E. Semi-supervised learning: a brief review[J]. International Journal of Engineering and Technology, 2018, 7(1.8): 81-85.

[10] CHEN Jifa, SUN Buyang, WANG Lizhe, et al. Semi-supervised semantic segmentation framework with pseudo supervisions for land-use/land-cover mapping in coastal areas[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: 102881.

[11] 陈理,韩毅,杨广,等. 基于改进半监督模型的空间异质性农田特征提取研究[J]. 农业机械学报, 2023, 54(12): 173-185.

- CHEN Li, HAN Yi, YANG Guang, et al. Spatially heterogeneous cropland characteristic extraction based on improved semi-supervised models[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 173 – 185. (in Chinese)
- [12] ZHANG Mingmei, XUE Yongan, ZHAN Yuanyuan, et al. Semi-supervised semantic segmentation-based remote sensing identification method for winter wheat planting area extraction[J]. Agronomy, 2023, 13(12): 2868.
- [13] OLSSON V, TRANHEDEN W, PINTO J, et al. ClassMix: segmentation-based data augmentation for semi-supervised learning [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 1369 – 1378.
- [14] YUN S, HAN D, OH S J, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6023 – 6032.
- [15] WANG Jiaqi, LIU Bing, ZHOU Yong, et al. Semisupervised multiscale generative adversarial network for semantic segmentation of remote sensing image[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1 – 5.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, 2014: 2672 – 2680.
- [17] HAN Wei, FENG Ruyi, WANG Lizhe, et al. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 23 – 43.
- [18] HUNG W C, TSAI Y H, LIOU Y T, et al. Adversarial learning for semi-supervised semantic segmentation [J]. arXiv Preprint, arXiv:1802.07934, 2018.
- [19] 耿艳磊, 邹嵘嵘, 何帅帅. 基于半监督生成对抗网络的遥感影像地物语义分割[J]. 测绘与空间地理信息, 2020, 43(4): 36 – 39.
- GENG Yanlei, ZOU Zhengrong, HE Shuaishuai. Semantic segmentation of objects in remote sensing images based on semi-supervised generative adversarial nets [J]. Geomatics and Spatial Information Technology, 2020, 43(4): 36 – 39. (in Chinese)
- [20] CHEN Hao, LI Zhenghong, WU Jiangjiang, et al. SemiRoadExNet: a semi-supervised network for road extraction from remote sensing imagery via adversarial learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 198: 169 – 183.
- [21] 逯焕宇, 张永宏, 马光义, 等. 基于半监督对抗学习的遥感图像水体提取[J]. 计算机工程, 2024, 50(7): 251 – 263.
- LU Huanyu, ZHANG Yonghong, MA Guangyi, et al. Semi-supervised adversarial learning based water body extraction from remote sensing images[J]. Computer Engineering, 2024, 50(7): 251 – 263. (in Chinese)
- [22] LI Guang, CUI Jiawei, HAN Wenting, et al. Crop type mapping using time-series Sentinel – 2 imagery and U-Net in early growth periods in the Hetao irrigation district in China[J]. Computers and Electronics in Agriculture, 2022, 203: 107478.
- [23] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European Conference on Computer Vision, 2018: 801 – 818.
- [24] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012 – 10022.
- [25] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv Preprint, arXiv:1511.06434, 2015.
- [26] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA – Net: efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 – 11542.
- [27] YANG Lihe, ZHUO Wei, QI Lei, et al. ST + +: make self-training work better for semi-supervised semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4268 – 4277.
- [28] SOHN K, BERTHELOT D, CARLINI N, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[J]. Advances in Neural Information Processing Systems, 2020, 33: 596 – 608.
- [29] YANG Lihe, QI Lei, FENG Litong, et al. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7236 – 7246.
- [30] OUALI Y, HUDELLOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12674 – 12684.
- [31] ZHANG Bowen, WANG Yidong, HOU Wenxin, et al. FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling[J]. Advances in Neural Information Processing Systems, 2021, 34: 18408 – 18419.