

doi: 10.6046/zrzyg.2023169

引用格式: 潘俊杰, 慎利, 鄢薪, 等. 一种基于对抗学习的高分辨率遥感影像语义分割无监督域自适应方法[J]. 自然资源遥感, 2024, 36(4): 149-157. (Pan J J, Shen L, Yan X, et al. An adversarial learning-based unsupervised domain adaptation method for semantic segmentation of high-resolution remote sensing images[J]. Remote Sensing for Natural Resources, 2024, 36(4): 149-157.)

# 一种基于对抗学习的高分辨率遥感影像 语义分割无监督域自适应方法

潘俊杰, 慎利, 鄢薪, 聂欣, 董宽林

(西南交通大学地球科学与环境工程学院, 成都 610097)

**摘要:** 语义分割模型在高分辨率遥感影像中表现良好的关键是训练集和测试集之间域的高度一致。然而, 不同数据集之间存在域偏差, 包括地理位置、传感器成像方式和天气条件的差异, 导致在一个数据集上训练的模型在另一个数据集上预测时准确性会显著下降。域自适应是解决上述问题的有效策略, 该文从域自适应模型的角度, 基于对抗学习方法提出了一种用于高分辨率遥感图像语义分割任务的无监督域自适应框架。该框架对全局域对齐模块和局部域对齐模块分别融入熵值加权注意力和逐类别域特征聚合机制, 缓解源域和目标域之间的域偏差; 此外, 引入了对象上下文表征(object context representation, OCR)模块和空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块, 以充分利用影像中的空间级和对象级上下文信息, 并提出了OCR/ASPP双分类器组合策略, 以提高分割精度和准确性。实验结果表明, 该方法在公开的2个数据集中实现了优越的跨域分割性能, 并超过了同类型的其他方法。

**关键词:** 高分辨率遥感图像; 语义分割; 对抗学习; 无监督域自适应

**中图法分类号:** TP 751; P 237 **文献标志码:** A **文章编号:** 2097-034X(2024)04-0149-09

## 0 引言

随着遥感技术的进步, 获取大量高分辨率遥感影像(high-resolution remote sensing images, HRSI)变得更加便捷, 推动了地球观测技术的精细化发展<sup>[1-3]</sup>。语义分割模型的应用使人们能够高效地将遥感影像中的每个像素分配到预定义的语义类别, 例如建筑物、道路、植被等<sup>[4-6]</sup>。HRSI的语义分割技术在城市建设规划、智能交通管理、规模化农业生产和自然灾害监测等遥感分析领域扮演着至关重要的角色<sup>[7-10]</sup>, 因此引起了广泛关注和研究。

近年来, 深度卷积神经网络(deep convolutional neural network, DCNN)以其卓越的特征表示能力逐渐受到广泛认可, 同时也使得基于DCNN的语义分割算法成为研究热点, 如FCN<sup>[11]</sup>, UNet<sup>[12]</sup>, SegNet<sup>[13]</sup>, PSPNet<sup>[14]</sup>, Deeplabv2<sup>[15]</sup>等多个算法相继被

提出, 并不断刷新着语义分割算法精度的里程碑。然而, 这些方法在训练集和测试集之间的数据域高度一致时, 才能表现出优异的性能, 并且对数据域的变化十分敏感。当训练数据和测试数据存在域偏差时, 例如影像地理位置、传感器成像方法和获取时天气条件等因素的影响会导致影像数据之间出现颜色、纹理、空间分辨率和背景的差异, 这使得在一个数据集(源域)上训练的模型一旦应用于另一个数据集(目标域)进行预测时, 模型的准确性可能会显著下降。

为解决上述问题, 最常见的方法是在目标域上进行大量像素级标注并重新训练模型, 然而, 这种方法成本高且耗时费力, 限制了其在实践中的可行性。另外, 利用已经训练好的模型在目标域上进行微调, 或者在源域数据集上进行样本扩增也是一种选择, 但这些方法并未取得理想的效果。相应地, 无监督域自适应(unsupervised domain adaptation, UDA)方

收稿日期: 2023-06-09; 修订日期: 2023-08-16

基金项目: 国家重点研发计划项目“时空知识图谱服务平台与应用验证”(编号: 2022YFB3904205)、国家自然科学基金项目“基于弱监督深度学习的高分辨率遥感影像灾后损毁建筑物提取研究”(编号: 42071386)、“基于匀质化分解与解析式合成的栅格类别数据尺度效应建模”(编号: 41971330)和四川省科技厅基本科研业务费项目“耕地‘非粮化’调查监测成果的知识化服务研究”(编号: 2023JDKY0017-3)共同资助。

第一作者: 潘俊杰(1998-), 男, 硕士研究生, 主要从事摄影测量与遥感方向研究。Email: peter\_panjunjie@163.com。

通信作者: 慎利(1986-), 男, 博士, 副教授, 主要从事遥感影像智能解译、资源环境遥感研究等方面的教学与科研工作。Email: rssh-enli@outlook.com。



## 1.2 OCR/ASPP 双分类器组合

HRSI 具有出色的空间分辨率和广泛的覆盖范围,这使得目标对象的细节和全局上下文信息变得至关重要。对象上下文表征(object context representation, OCR)模块<sup>[28]</sup>通过引入注意力机制和上下文感知,利用对象之间的关联信息来提供更丰富的对象表示,从而提升分割模型对图像中对象间关系的理解能力。同时,空洞空间金字塔池化(atrous spatial pyramid pooling, ASPP)模块<sup>[15]</sup>通过并行空洞卷积和全局平均池化操作,捕获多尺度的上下文信息,扩展感受野并实现全局特征融合,从而增强分割模型对不同尺度对象的识别能力。以往一些研究者的工作也证明了 OCR 模块对目标对象上下文特征关系优秀的学习能力<sup>[29-30]</sup>,以及 ASPP 模块对于全局多尺度上下文信息的获取能力<sup>[15]</sup>。这 2 个模块对于提升高分遥感影像中复杂场景目标对象的准确识别和分割能力有着重要的意义。

为了实现高质量的分割结果,采用 Deeplabv2<sup>[15]</sup>作为基础分割网络。在分割网络  $G$  的具体实现中,如图 2 所示,在经过特征提取器处理得到的特征层之后插入一个 OCR 模块,随后进行分类,并将其作为主分类器。辅助分类器使用 Deeplabv2 中带有 ASPP 模块的分类器。带有 OCR 模块的主分类器通过对对象间关系建模,捕捉目标之间的空间布局 and 语义相互作用,从而增强目标的上下文表示能力;而带有 ASPP 模块的辅助分类器,通过多尺度感受野,能够捕捉不同尺度上的语义信息,从而适应遥感影像中的目标变化和尺度差异。将 2 个分类器相结合,实现了特征的多尺度感知和目标对象之间关系的建模。这样的组合能够充分利用 HRSI 的丰富信息,提高分割的准确性和鲁棒性,以获得更好的域适应性能。

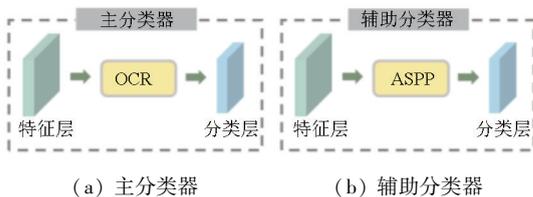


图 2 OCR/ASPP 双分类器组合

Fig.2 A combination of OCR/ASPP dual classifiers

## 1.3 EWG 模块

从整体的角度上来看,大多数 UDA 框架都是以相同的强度对齐目标域和源域的不同区域,但不同区域中的内容不同,导致其对齐难度也是不同的。并且,目标域中大部分预测不准确的区域通常是由于这些区域未能得到良好的对齐。

鉴于对齐难度与空间位置相关,并且辅助分类

器可以更好地利用目标域多尺度特征和全局上下文信息,可以利用辅助分类器的预测结果进行熵值处理。通过观察预测结果的熵值判断其偏差大小,高熵值表示预测结果的偏差较大,低熵值表示预测结果的偏差较小。因此,如图 3 所示,EWG 模块将高熵值区域作为对域对齐的重点关注对象,并给予这些区域更大的损失权重,迫使鉴别器在对抗训练中对齐目标域中高熵值区域的域特征分布,同时降低对低熵值区域的关注程度。

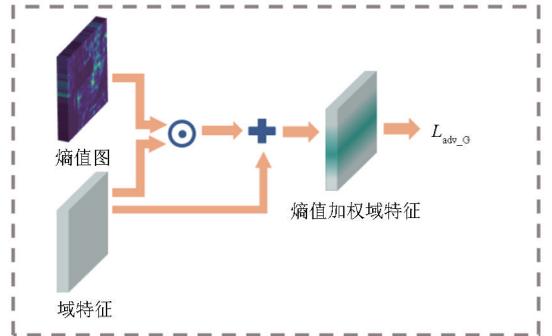


图 3 EWG 模块

Fig.3 EWG module

具体可表示为,用辅助分类器的预测结果  $P_1^{\text{aux}}$  计算香农熵来生成熵值图  $E_c$ ,再对其进行归一化处理得到  $E_1$ ,  $E_1 \in [0, 1]^n$ ,由归一化到  $[0, 1]$  范围的独立像素熵组成,即

$$[E_c]_c^n = - [P_1^{\text{aux}}]_c^n \times \log_2([P_1^{\text{aux}}]_c^n), \quad (1)$$

$$[E_1]_c^n = (\sum_{c=0}^C [E_c]_c^n) / \log_2(C), \quad (2)$$

式中:  $C$  为类别数;  $[\cdot]_c^n$  表示在维度为  $c$ 、空间位置为  $n$  处的值,  $n \in N, c \in C, N$  为图像像素数;  $[\cdot]_c^n$  为  $[\cdot]_c^n$  有且仅有一个维度时的情况,后式同理。

$E_1$  中空间位置处  $n$  的熵值越高,表示在  $n$  处对齐分布就越困难。所以本文将熵值图与注意力机制结合,来强制鉴别器模型关注难以对齐的区域,其可以公式化为:

$$[A_1]_c^n = - \sigma([E_1]_c^n \times [D_1]_c^n + [D_1]_c^n), \quad (3)$$

式中:  $\sigma$  为 sigmoid 函数; 目标域的域特征  $D_1$  为  $D_1^{\text{main}}$  与  $D_1^{\text{aux}}$  的和,  $D_1^{\text{main}}$  和  $D_1^{\text{aux}}$  分别为在目标域上主分类器和辅助分类器输出结果经过鉴别器处理后得到的域特征;  $[A_1]_c^n$  可以被视为加权运算,其最大程度地利用熵图来自适应地调整不同位置处的对应权重,最后利用 sigmoid 函数对求得的加权熵值进行归一化处理,以保持模型训练的稳定性。

基于熵值加权注意力机制的全局对齐损失  $L_{\text{adv}_G}$  可以表示为:

$$L_{\text{adv}_G} = - \sum_{n=0}^N \log_2([A_1]_c^n). \quad (4)$$

通过对域特征和熵值图进行加权运算,给予难以对齐的区域更大的损失权重,以便鉴别器更加专注地对齐这些区域。这种整体性的区域加权策略使得容易对齐和困难对齐的区域具有不同的权重,避免了使用相同权重对所有区域进行对齐的原始方法可能引发的负迁移问题的出现。

#### 1.4 CAL 模块

从类别的角度来看,目标域域特征的不同空间位置对应的类别属性可能不同,因此整体域对齐方式可能会导致类别的负迁移现象。在本文中,CAL先将不同空间位置的域特征与相应的类别属性进行匹配,再利用主分类器对类别对象间上下文语义特征敏感的特点,通过伪标签生成策略结合目标域的主分类器预测结果  $P_1^{\text{main}}$  完成对类别信息的获取。伪标签生成策略可表示为:

$$c = \operatorname{argmax} [P_1^{\text{main}}]_c^n, \quad (5)$$

$$[\text{pseudo}_1]_c^n = \begin{cases} c & ([P_1^{\text{main}}]_c^n \geq T_c) \\ 255 & ([P_1^{\text{main}}]_c^n < T_c) \end{cases}, \quad (6)$$

式中:  $T_c$  为概率阈值,如果空间位置  $n$  处像素的类别  $c$  的预测概率值大于等于  $T_c$ ,则将该像素处的标签值设为  $c$ ,表示存在一个可信的类别标签。如果小于  $T_c$ ,则将该像素处的标签值设为 255,表示不存在可信的标签值,  $c' \in C$ ,最终得到目标域影像的伪标签  $[\text{pseudo}_1]_c^n$ 。

图 4 为 CAL 模块示意图。

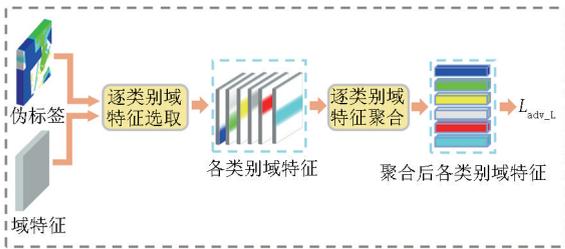


图 4 CAL 模块

Fig.4 CAL module

如图 4 所示,一旦获得了此处的伪标签,便可以通过伪标签为目标域的域特征  $D_1$  的不同区域分配相应的类别值,以便逐类别进行对齐。类别  $c$  的域特征选择策略可以用以下公式表示:

$$\text{mask}_c = ([\text{pseudo}_1]_c^n = c), \quad (7)$$

$$D_1^c = D_1 \times \text{mask}_c, \quad (8)$$

式中:  $D_1$  与式(3)中同值;  $\text{mask}_c$  为  $[\text{pseudo}_1]_c^n$  中值为类别  $c$  的空间位置掩模; 各类别域特征  $D_1^c$  表示在目标域中类别  $c$  对应的域特征。

在进行了不同类别的域特征区域选取后,利用平均池化对该类别的域特征进行聚合,域特征逐类

聚合策略可表示为:

$$A_D = \left( \sum_{m=0}^M [D_1^c]_c^m \right) / M, \quad (9)$$

式中:  $A_D$  为聚合后的各类别域特征;  $M$  为类别  $c$  对应的域特征的像素数量。

最后,基于逐类别域特征聚合的局部对齐损失  $L_{\text{adv}_L}$  可以被表示为:

$$L_{\text{adv}_L} = - \sum_{c=0}^C \log_2(A_D). \quad (10)$$

与直接整体对齐域特征的方法相比,本文提出的逐类别域特征聚合策略能够使域鉴别器知道每个类别的域特征区域,并对每个类别的域特征进行分别聚合和对齐,互不干扰。此外,对于那些不确定类别的区域,由于缺乏类别标签,不进行域特征聚合,即域特征分布不会调整,这 2 点都有效防止类别负迁移现象的出现。

#### 1.5 模型训练

本文提出的 UDA 语义分割框架主要使用 4 个损失函数进行训练,包括分割损失、自适应对抗损失、全局对齐损失和局部对齐损失。其中,本文使用多类交叉熵损失函数计算分割损失  $L_{\text{seg}}$ ,在带标签的源域数据上对分割网络  $G$  进行全监督训练,可表示为:

$$L_{\text{seg\_main}}(P_{s\_main}, Y_s) = S(P_{s\_main}, Y_s), \quad (11)$$

$$L_{\text{seg\_aux}}(P_{s\_aux}, Y_s) = S(P_{s\_aux}, Y_s), \quad (12)$$

$$L_{\text{seg}} = \lambda_{\text{seg\_main}} L_{\text{seg\_main}} + \lambda_{\text{seg\_aux}} L_{\text{seg\_aux}}, \quad (13)$$

式中:  $S()$  为多类交叉熵损失函数<sup>[24]</sup>;  $P_{s\_main}$  和  $P_{s\_aux}$  分别为主分类器和辅助分类器在源域上的预测结果;  $Y_s$  为标签值;  $\lambda_{\text{seg\_main}}$  和  $\lambda_{\text{seg\_aux}}$  为训练分割网络时候的损失平衡权重。

对于模型中的鉴别器,本文使用对抗训练和二类交叉熵损失来对其进行训练,其损失函数  $L_D$  可以表示为:

$$L_D^{\text{main}} = (1-r)d(D_s^{\text{main}}) + rd(D_t^{\text{main}}), \quad (14)$$

$$L_D^{\text{aux}} = (1-r)d(D_s^{\text{aux}}) + rd(D_t^{\text{aux}}), \quad (15)$$

$$L_D = L_D^{\text{main}} + L_D^{\text{aux}}, \quad (16)$$

式中:  $D_s^{\text{main}}$  和  $D_s^{\text{aux}}$  分别为在源域上主分类器和辅助分类器输出结果经过鉴别器处理后得到的域特征;  $d()$  为二类交叉熵损失函数<sup>[24]</sup>;  $r$  为对抗训练中的域类别,如果影像来自目标域,则  $r = 1$ ,如果影像来自源域,则  $r = 0$ 。

为了更好地对齐全域特征与目标域特征的分布,本文使用 EWG 模块和 CAL 模块来进行域对齐,

对抗性损失  $L_{adv}$  定义如下:

$$L_{adv} = \lambda_{adv\_G} L_{adv\_G} + \lambda_{adv\_L} L_{adv\_L}, \quad (17)$$

式中  $\lambda_{adv\_G}$  和  $\lambda_{adv\_L}$  为 2 个模块在对齐过程中的所占的损失权重比例。

## 2 实验设计

### 2.1 实验数据

本实验采用 ISPRS 在遥感图像分割和对象识别领域中常用的 2 个 HRSI 数据集 Vaihingen 和 Potsdam。其中 Vaihingen 数据集包含 33 幅不同大小的遥感影像,正射影像和数字地表模型(digital

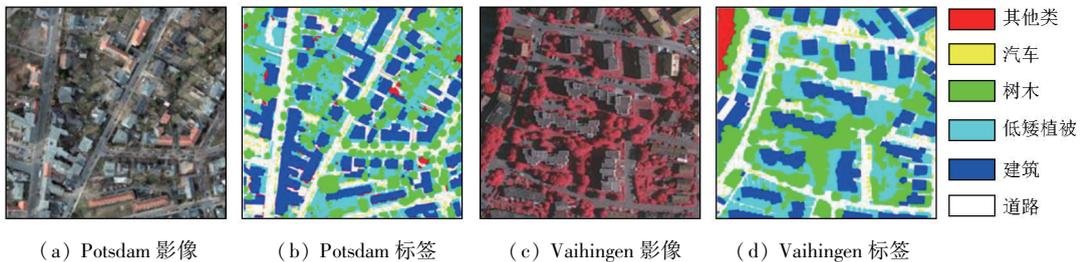


图5 源域和目标域数据集样本示例

Fig.5 Sample of source and target domain datasets

### 2.2 网络结构设置

本文提出的基于对抗学习的 UDA 框架是基于 Adaptseg<sup>[24]</sup>实现,对于分割网络  $G$ ,带有 OCR 模块的分类器作为主分类器,将带有 ASPP 模块的分类器作为辅助分类器。ASPP 的采样率固定为 [6, 12, 18, 24] 以获得不同的感受野。此外,还执行多输出策略<sup>[24]</sup>以提高语义分割网络的性能。

对于域鉴别器  $D$ ,采用了深度卷积生成对抗网络(deep convolutional generation adversarial network, DCGAN)<sup>[31]</sup>。在得到概率预测图后,将其输入到 DCGAN 的 4 个带有泄露修正线性单元(leaky ReLU)的卷积层中,并通过  $1 \times 1$  卷积层输出输入值的域类别(源域或目标域)。

### 2.3 实验参数设置

本文实验在 PyTorch(1.7.1)框架及单个 8 G 的 NVIDIA 3060TI GPU 下进行。参考 Adaptseg 中的设置,语义分割网络  $G$  使用随机梯度下降(stochastic gradient descent, SGD)优化器<sup>[32]</sup>,动量为 0.9,学习率为 0.002 5,权重衰减为 0.001。域鉴别器  $D$  使用 Adam 优化器<sup>[33]</sup>,学习率为 0.001,训练迭代次数设置为 150 000。

经过多次实验发现将  $T_c$  设为 0.75 时,模型能够达到最佳效果。对于加权因子  $\lambda_{seg\_main}$ ,  $\lambda_{seg\_aux}$ ,  $\lambda_{adv\_G}$  和  $\lambda_{adv\_L}$ ,本文采用了 Adaptseg<sup>[24]</sup>的设置,将它们固定为  $\lambda_{seg\_main} = 1$ ,  $\lambda_{seg\_aux} = 0.1$ 。同时,通过多

surface model, DSM) 的空间分辨率为 0.09 m, Potsdam 数据集包含 38 幅相同尺寸的遥感影像,正射影像和 DSM 的空间分辨率为 0.05 m。

如图 5 所示, Potsdam 数据集和 Vaihingen 数据集都具有较高的分辨率和丰富的地物细节。并且 Potsdam 数据集中的对象类别与 Vaihingen 类似,包括建筑物、道路、树木、汽车等。在本实验中将这 2 个数据集裁剪成 512 像素  $\times$  512 像素大小,横向和纵向重叠度均为 50%,并将裁剪后的 Potsdam 数据集作为源域, Vaihingen 数据集作为目标域。最后, 2 个数据集都随机划分 20% 的图像作为测试集,剩下的 80% 作为训练集。

次实验调整了比例,将  $\lambda_{adv\_G}$  设置为 0.03,  $\lambda_{adv\_L}$  设置为 0.02 以实现最佳效果。在测试阶段,本文仅使用分割网络  $G$  而不使用鉴别器  $D$ ,并使用多级输出策略<sup>[24]</sup>,在不引入额外参数的情况下,实现分割网络  $G$  对目标域影像分割性能的优化。

### 2.4 精度评价指标

为了公正评估分割网络  $G$  在每个类别上的性能,计算各类别的交并比(intersection over union, IOU)和平均交并比(mean intersection over union, mIOU)<sup>[11]</sup>来综合考虑模型对于各类别的性能并评估模型的整体性能。公式分别为:

$$IOU = TP / (TP + FP + FN), \quad (18)$$

$$mIOU = \sum_{i=0}^C \{ [TP_i / (TP_i + FP_i + FN_i)] \} / C, \quad (19)$$

式中:  $TP$  为真阳性;  $FP$  为假阳性;  $FN$  为假阴性;  $i$  为类别序号;  $C$  为类别总数。

### 2.5 对比方法及实验结果分析

本文实验中引入了几个具有代表性的基于对抗学习的 UDA 语义分割框架作为对比方法,包括 AdaptSegNet, CLAN, ADVENT 和 Metacorection。并使用 Potsdam 数据集作为源域, Vaihingen 数据集作为目标域。为了公平比较,都使用 ResNet50<sup>[34]</sup>作为基准模型 Deeplabv2 的特征提取器,并确保所有模型都经过充分训练。本文提出的 OA-GAL 和对比

方法的分割结果如表1所示,可以看出 OA-GAL 相比其他模型有更好的表现,并且没有明显的错误分割。同时,从第4列和第8列的图像可以明显观察

到相较于基线模型(AdaptSegNet),OA-GAL 更好地减少了负迁移的发生。

表1 Potsdam→Vaihingen 的 UDA 分割结果示例

Tab.1 Example of UDA segmentation results for Potsdam → Vaihingen

序号	目标域影像	标签影像	AdaptSegNet	CLAN	ADVENT	Metacorection	OA-GAL
1							
2							
3							
4							
5							

对分割效果进行定量评价,实验结果见表2,第一行是 Deeplabv2(仅使用源域数据进行训练的分割模型,用于预测目标域图像)。从第2—5行是与 UDA 相关的语义分割模型的对比方法,第6行是本文提出的 OA-GAL,加粗字体表示每列的最高值。从表2可以看出,相较于未进行域自适应的 Deeplabv2,OA-GAL 在各类别的 *IOU* 上都有显著提升,

*mIOU* 从 0.264 7 提升到 0.474 8。此外,相对于其他模型,OA-GAL 在树木、低矮植被、建筑、道路等类别的 *IOU* 明显优于其他模型,但在其他类别和汽车这 2 个小类别的 *IOU* 略低于个别模型。总体而言,本文提出的 OA-GAL 优于其他模型,并获得了更高的 *mIOU* 分数。

表2 Potsdam→Vaihingen 对比实验的结果评价

Tab.2 Evaluation of the results of the Potsdam → Vaihingen comparative experiment

模型	<i>mIOU</i>	<i>IOU</i>					
		其他类	汽车	树木	低矮植被	建筑	道路
Deeplabv2	0.264 7	0.066 3	0.074 5	0.186 5	0.244 3	0.488 6	0.527 9
AdaptSegNet	0.423 1	0.075 2	<b>0.263 4</b>	0.457 8	0.401 1	0.720 1	0.620 9
CLAN	0.410 1	0.084 7	0.164 0	0.544 1	0.274 1	0.773 0	0.619 9
ADVENT	0.434 8	<b>0.168 6</b>	0.221 8	0.510 7	0.316 9	0.768 2	0.622 4
Metacorection	0.440 4	0.102 8	0.249 5	0.517 1	0.400 1	0.744 8	0.628 1
OA-GAL	<b>0.474 8</b>	0.114 8	0.219 5	<b>0.573 2</b>	<b>0.435 5</b>	<b>0.818 2</b>	<b>0.687 4</b>

此外,如图6所示,本文还选取2个示例,可视化了框架中 CAL 模块的伪标签生成结果和 EWG 模块的熵值图生成结果。本文将预测概率图的阈值  $T_c$  设定为 0.75,即将黑色区域定义为不确定性区域。通过对比伪标签中的黑色区域和熵值图的部分,可以观察到这些不确定的黑色区域在熵值图中呈现为

高亮区域,表示难以对齐的区域。这也是本文思路的出发点,对于确定性区域采用基于逐类别域特征聚合的局部对齐策略,确保局部对齐过程互不干扰。而对于不确定性区域,即高熵值区域,本文采用基于熵值加权的全局对齐策略,更加关注这些不确定区域,以促进目标域中困难区域的对齐。

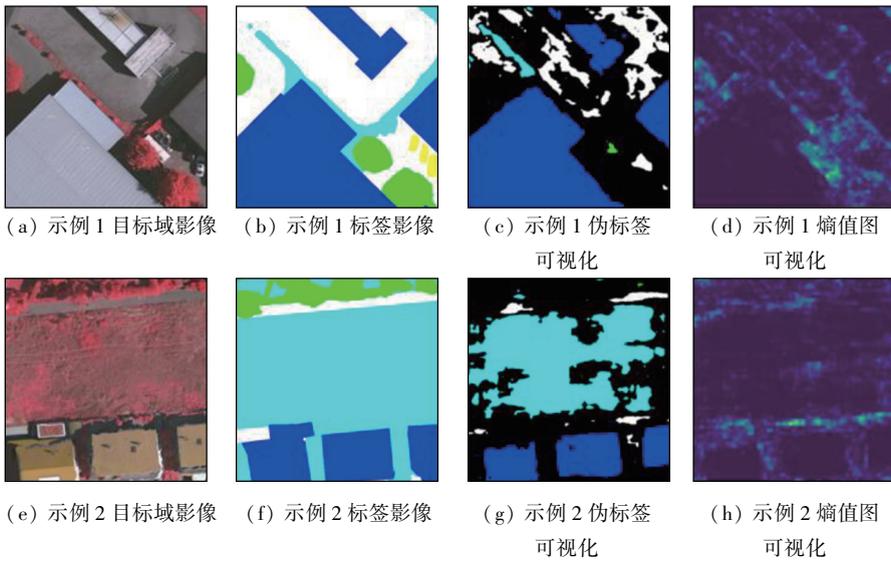


图6 伪标签和熵值图可视化示例

Fig.6 Example of visualization of pseudo labels and entropy regions

## 2.6 消融实验

为了验证所提出方法的有效性,本文在不同设置下评估了从 Potsdam 到 Vaihingen 的模型测试结果。基线模型表示 AdaptSegNet 模型(不采用 OCR/ASPP 双分类器组合、CAL 模块、EWG 模块),具体结果如表 3 所示。第 2—5 行展示了在基线基础上逐渐添加 OCR/ASPP 双分类器组合、CAL 模块和 EWG 模块后的结果。

表 3 消融实验

Tab.3 Ablation experiment

模型	OCR/ASPP	CAL	EWG	<i>mIOU</i>
基线模型				0.423 1
OCR/ASPP	✓			0.438 2
OCR/ASPP+CAL	✓	✓		0.451 9
OCR/ASPP+EWG	✓		✓	0.453 7
OA-GAL	✓	✓	✓	0.474 8

此外,从表 2 的第 1 行可以看出,不进行域自适应的分割模型(Deeplabv2)的 *mIOU* 仅为 0.264 7,而在本文实验中使用源域数据训练的分割模型 Deeplabv2 在源域测试时的 *mIOU* 为 0.722 3,这种显著的精度下降反映了源域和目标域之间存在巨大的域偏差。对比表 3 中的基线模型和表 2 中的 Deeplabv2 的预测结果,可以发现基线模型在一定程度上通过对齐输出空间中的分布,使得目标域和源域有所对齐,但效果一般。

当引入 OCR/ASPP 双分类器组合到模型后,*mIOU* 提高到 0.438 2。这表明 OCR/ASPP 双分类器组合可以分别从类别上下文感知和多尺度上下文信息的角度出发,去增强分类模型的理解能力和泛化能力,有助于更好地进行域对齐。

在分别将 CAL 和 EWG 插入到域鉴别器之后,

*mIOU* 进一步提高到 0.451 9 和 0.453 7。这证明了针对每个类别单独对齐域特征且互不干扰的方法以及根据不同区域给予不同关注度的域对齐方法的有效性。最后,与仅插入 CAL 或 EWG 相比,当同时添加 CAL 和 EWG 时,*mIOU* 提升到 0.474 8,均有所提升。这验证了 CAL 和 EWG 模块在域对齐过程中的互补性,也验证了对不同区域进行不同强度的域对齐策略的可行性。

## 3 结论

本文提出了一个基于对抗学习的无监督自适应框架 OA-GAL,用于 HRSI 的分割,并与其他同类型方法进行了对比实验,最后对实验结果和精度进行了对比分析,验证了 OA-GAL 性能的有效性与优越性,结论如下:

1) 本文在 OA-GAL 的分割网络  $G$  中提出了 OCR/ASPP 双分类器组合,实现了特征的多尺度感知和目标对象间的信息关联,提高了对 HRSI 分割的准确性和鲁棒性。

2) 在 OA-GAL 的鉴别器  $D$  中提出了 EWG 模块和 CAL 模块,其中 EWG 专注于解决目标域中难以实现对齐的区域,而 CAL 则负责实现各类别间相互独立的域对齐效果。两者的结合产生了更精细、全面的对齐结果,提高了鉴别器  $D$  的域自适应效果。

3) 与大多数同类型方法采用的整体性域对齐方式不同,OA-GAL 采用针对不同特征区域的差异性关注度进行域对齐。实验证明,OA-GAL 能够更有效地缓解不同域之间的域偏差问题,并在无标签的目标域上获得良好的分割结果。

通过分析,认识到 OA-GAL 对于类似于汽车、其他类等小类别地物的分割能力方面存在一些不足,因此在后续工作中会着重去优化小类别地物的分割。值得注意的是,OA-GAL 在建筑和道路的分割结果方面表现较为出色,后续会将该框架应用于建筑物或道路等单类域自适应分割问题进行进一步验证。

## 参考文献 (References):

- [1] 卢晓燕.面向高分辨率遥感影像大范围道路提取的深度学习研究方法研究[J].武汉大学学报(信息科学版),2023,48(5):821.  
Lu X Y. Deep learning method for large-scale road extraction from high resolution remote sensing imagery[J]. Geomatics and Information Science of Wuhan University, 2023, 48(5): 821.
- [2] 薛源,覃超,吴保生,等.基于多源国产高分辨率遥感影像的山区河流信息自动提取[J].清华大学学报(自然科学版),2023,63(1):134-145.  
Xue Y, Qin C, Wu B S, et al. Automatic extraction of mountain river information from multiple Chinese high-resolution remote sensing satellite images[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(1): 134-145.
- [3] 赵会芹,于博,陈方,等.基于高分辨率卫星遥感影像滑坡提取方法研究现状[J].遥感技术与应用,2023,38(1):108-115.  
Zhao H Q, Yu B, Chen F, et al. Research status of landslide extraction methods based on high-resolution satellite remote sensing images[J]. Remote Sensing Technology and Application, 2023, 38(1): 108-115.
- [4] 王丽梅,王延正.基于高分辨率遥感影像的建筑物提取[J].测绘通报,2023(6):180-183.  
Wang L M, Wang Y Z. Buildings extraction based on high-resolution remote sensing imagery[J]. Bulletin of Surveying and Mapping, 2023(6): 180-183.
- [5] 张新长,黄健锋,宁婷.高分辨率遥感影像耕地提取研究进展与展望[J].武汉大学学报(信息科学版),2023,48(10):1582-1590.  
Zhang X C, Huang J F, Ning T. Progress and prospect of cultivated land extraction from high-resolution remote sensing images[J]. Geomatics and Information Science of Wuhan University, 2023, 48(10): 1582-1590.
- [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 X C, Liu Z Y, Jiang Y, 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.
- [7] 杨军,于茜子.结合空洞卷积的FuseNet变体网络高分辨率遥感影像语义分割[J].武汉大学学报(信息科学版),2022,47(7):1071-1080.  
Yang J, Yu X Z. Semantic segmentation of high-resolution remote sensing images based on improved FuseNet combined with atrous convolution[J]. Geomatics and Information Science of Wuhan University, 2022, 47(7): 1071-1080.
- [8] Chen B, Xia M, Qian M, et al. MANet: A multi-level aggregation network for semantic segmentation of high-resolution remote sensing images[J]. International Journal of Remote Sensing, 2022, 43(15/16): 5874-5894.
- [9] Wang Y, Zeng X, Liao X, et al. B-FGC-net: A building extraction network from high resolution remote sensing imagery[J]. Remote Sensing, 2022, 14(2): 269.
- [10] Guo Y, Liu Y, Georgiou T K, et al. A review of semantic segmentation using deep neural networks[J]. International Journal of Multimedia Information Retrieval, 2018, 7(2): 87-93.
- [11] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE, 2017: 640-651.
- [12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [13] 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.
- [14] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 6230-6239.
- [15] 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, 2018, 40(4): 834-848.
- [16] Zhao S, Yue X, Zhang S, et al. A review of single-source deep unsupervised visual domain adaptation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 473-493.
- [17] Xu M, Wu M, Chen K, et al. The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data[J]. Remote Sensing, 2022, 14(17): 4380.
- [18] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2242-2251.
- [19] Yang Y, Soatto S. FDA: Fourier domain adaptation for semantic segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 4084-4094.
- [20] Hoffman J, Tzeng E, Park T, et al. CyCADA: Cycle-consistent adversarial domain adaptation[J/OL]. arXiv, 2017. <https://arxiv.org/abs/1711.03213.pdf>.
- [21] Ma H, Lin X, Wu Z, et al. Coarse-to-fine domain adaptive semantic segmentation with photometric alignment and category-center regularization[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 4050-4059.
- [22] Zou Y, Yu Z, Vijaya Kumar B V K, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training[C]//Computer Vision - ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III. ACM, 2018: 297-313.

- [23] Zheng Z, Yang Y. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation [J]. *International Journal of Computer Vision*, 2021, 129(4): 1106–1120.
- [24] Tsai Y H, Hung W C, Schuster S, et al. Learning to adapt structure output space for semantic segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 7472–7481.
- [25] Luo Y, Zheng L, Guan T, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 2502–2511.
- [26] Vu T H, Jain H, Bucher M, et al. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 2512–2521.
- [27] Guo X, Yang C, Li B, et al. MetaCorrection: Domain-aware meta loss correction for unsupervised domain adaptation in semantic segmentation [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 3926–3935.
- [28] Yuan Y, Chen X, Wang J. Object-contextual representations for semantic segmentation [J/OL]. arXiv, 2019. <https://arxiv.org/abs/1909.11065.pdf>.
- [29] Yuan Y, Chen X, Chen X, et al. Segmentation transformer: Object-contextual representations for semantic segmentation [J/OL]. arXiv, 2019(2021-04-30). <https://arxiv.org/abs/1909.11065v2>.
- [30] Huang S, Han W, Chen H, et al. Recognizing zucchinis intercropped with sunflowers in UAV visible images using an improved method based on OCRNet [J]. *Remote Sensing*, 2021, 13(14): 2706.
- [31] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [J/OL]. arXiv, 2015. <https://arxiv.org/abs/1511.06434.pdf>.
- [32] Bottou L. Large-scale machine learning with stochastic gradient descent [C]//Proceedings of COMPSTAT2010: 19th International Conference on Computational Statistics. Physica-Verlag HD, 2010: 177–186.
- [33] Kingma D P, Adam B J. Adam: A method for stochastic optimization [J/OL]. arXiv, 2014(2017-01-30). <https://arxiv.org/abs/1412.6980>.
- [34] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770–778.

## An adversarial learning-based unsupervised domain adaptation method for semantic segmentation of high-resolution remote sensing images

PAN Junjie, SHEN Li, YAN Xin, NIE Xin, DONG Kuanlin

(Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 610097, China)

**Abstract:** The key to the high performance of semantic segmentation models for high-resolution remote sensing images lies in the high domain consistency between the training and testing datasets. The domain discrepancies between different datasets, including differences in geographic locations, sensors' imaging patterns, and weather conditions, lead to significantly decreased accuracy when a model trained on one dataset is applied to another. Domain adaptation is an effective strategy to address the aforementioned issue. From the perspective of a domain adaptation model, this study developed an adversarial learning-based unsupervised domain adaptation framework for the semantic segmentation of high-resolution remote sensing images. This framework fused the entropy-weighted attention and class-wise domain feature aggregation mechanism into the global and local domain alignment modules, respectively, alleviating the domain discrepancies between the source and target. Additionally, the object context representation (OCR) and Atrous spatial pyramid pooling (ASPP) modules were incorporated to fully leverage spatial- and object-level contextual information in the images. Furthermore, the OCR and ASPP combination strategy was employed to improve segmentation accuracy and precision. The experimental results indicate that the proposed method allows for superior cross-domain segmentation on two publicly available datasets, outperforming other methods of the same type.

**Keywords:** high-resolution remote sensing images; semantic segmentation; adversarial learning; unsupervised domain adaptation