Vol. 45 No. 12 Dec. 2022

# 结合超分辨率和域适应的遥感图像语义分割方法

## 梁 敏 汪西莉

(陕西师范大学计算机科学学院 西安 710119)

摘 要 基于卷积神经网络的图像语义分割方法依赖于有真实标签的监督学习,但不能很好地推广到来源不同的无标签数据集中,目前无监督域适应给出了解决无标签目标域数据和有标签源域数据特征分布不一致的思路.但是在遥感图像处理领域,当源域和目标域的空间分辨率不相同时,无监督域适应方法应用效果不佳.本文结合图像超分辨率设计了一个新的端到端语义分割深度网络——结合超分辨率和域适应的语义分割模型(Semantic Segmentation Model Combining Super Resolution and Domain Adaption, SSM-SRDA),它可以缩小低分辨率源域和高分辨率目标域遥感图像的空间分辨率差异和特征分布差异,并完成无监督超分辨率和域适应语义分割任务. SSM-SRDA 模型包括三个部分:结合超分辨率的语义分割网络,将低分辨率的源域图像通过超分辨率模块生成带有目标域风格的高分辨率图像,缩小空间分辨率差异的同时帮助特征提取网络学习域不变特征,并通过特征相似性 FA-Loss 模块增强语义分割深层网络高分辨率特征的细节结构信息;像素级域判别器,用来缩小超分辨后源域和目标域间的像素级特征差异;输出空间域判别器,在 SoftMax 输出空间中缩小特征差异.通过语义分割网络和两个判别器的对抗性学习,SSM-SRDA 在分割模块的输入和输出级对齐源域与目标域的特征分布,通过域适应迁移应用于不同源的目标域数据分割,得到能够应用于实际且推广性更强的模型.实验在四组遥感图像数据集上证明SSM-SRDA 优于现有的域适应语义分割方法,在平均交并比上分别提高了 0.7%、1.7%、2.2%和 3.3%.

关键词 域适应;遥感图像;语义分割;超分辨率,对抗学习中图法分类号 TP391 **DOI**号 10.11897/SP. J. 1016, 2022.02619

## Semantic Segmentation Model for Remote Sensing Images Combining Super Resolution and Domain Adaption

LIANG Min WANG Xi-L

(School of Computer Science, Shaanxi Normal University, Xi'an (710119)

Abstract The semantic segmentation method based on convolutional neural network rely on supervised learning with ground truth, but it cannot be well generalized to unlabeled datasets with different sources. Unsupervised domain adaptation can solve the problem of inconsistent feature distribution between unlabeled target domain data and labeled source domain data. This is due to that remote sensing images are often come from different sources, they are variable in their spatial resolution and are influenced by different imaging regions, imaging conditions and imaging times. Even images from the same region may have large differences in spectral features. The generalization of the semantic segmentation model relies on the reduction of these inter-domain differences mentioned above. Therefore, unsupervised domain adaptation methods for remote sensing image should not only reduce the differences in features between domains, but also address the problem of different spatial resolutions. This paper designs a new end-to-end semantic segmentation deep network combined with image super resolution—Semantic Segmentation Model Combining Super Resolution and Domain Adaption, which can reduce the spatial resolution difference

and feature distribution difference between the low spatial resolution source domain and high spatial resolution target domain remote sensing images, and accomplish the super-resolution task for the source domain and the domain adaptation semantic segmentation task for the target domain. The SSM-SRDA model consists of three parts one is Semantic Segmentation Network with Super Resolution (SSNSR), the second is Pixel-level Domain Discriminator (PDD), and the third is Output-space Domain Discriminator (ODD). SSNSR consists of a semantic segmentation network and a super-resolution network, which share the same feature extraction network. The superresolution network generates a high-resolution synthetic image with target image style from a low-resolution source domain image, which can eliminate spatial resolution differences and style differences to help the feature extraction module learn the same features between the source and target domains. The Feature Affinity-Loss module enhances the learning of the semantic segmentation deep network by the feature maps with more detailed structural information obtained by superresolution. The pixel-level domain discriminator is used to reduce the pixel-level feature differences between the high-resolution target domain and the synthetic image of the source domain. High-resolution source domain synthetic images with target domain style are generated by generative adversarial learning with the super-resolution network and participate in the training of the model as additional training data. The output-space domain discriminator reduces the feature differences between source and target domain images in the output space of the semantic segmentation network. Through the adversarial learning of the semantic segmentation network and the two discriminators, SSM-SRDA aligns the feature distribution of the source domain and the target domain at the input and output stages of the segmentation network, and can be applied to the target domain datas of more different sources through domain adaptation. It is a practical and more popular model. Experiments show that SSM-SRDA is superior to the existing domain adaptive semantic segmentation methods on four sets of remote sensing image data sets, and the intersection ratio is improved by 0.7%, 1.7%, 2.2% and 3.3%, respectively.

**Keywords** domain adaptation; remote sensing image; semantic segmentation; super resolution; adversarial learning

## 1 引 言

遥感图像语义分割的目的是为图像的每个像素都预测一个类别标签,在很多实际问题中都得到了广泛应用,如城市规划[1]、土地利用调查<sup>[2]</sup>和环境监测<sup>[3]</sup>等.基于深度卷积神经网络(Convolutional Neural Networks, CNNs)的不同分割模型在语义分割任务中取得了显著效果<sup>[4]</sup>,它们都需要大量有标签的数据进行监督训练.对于遥感图像语义分割,在实际应用中常常面临的问题是:没有标签或缺少标签,这就需要耗费大量时间与人力进行标注,否则难以通过监督学习获得较好的结果.无监督域适应(Unsupervised Domain Adaptation, UDA)是处理标签不足问题的有效方法之一,它通过各种方法缩小域间差异,将基于公开的有标签的源域数据中学习

得到的语义分割模型迁移到没有标签的目标域中. 遥感图像不同域的数据往往来源不同、成像地面区域不同,存在光谱、分辨率等方面的差异.用于语义分割的 UDA 大都采用基于对抗的域适应学习方法,当源域数据与目标域数据特征分布之间的差异无法被域判别器区分时,就能在网络学习中对齐源域和目标域的特征分布<sup>[5]</sup>. Zhang 等人<sup>[6]</sup>将对抗损失应用于分割网络中,主要捕获图像外观信息的浅层;Li 等人<sup>[7]</sup>通过对抗训练将源域图像转化为目标域风格的图像,以最小化像素级域差异;Tsai 等人<sup>[8]</sup>认为特征空间中的域适应不是语义分割任务的最佳选择,而输出的分割结果虽然处于低维空间,但包含丰富的结构化场景布局信息,建议在输出空间对齐特征分布以缩小空间结构上的差异.

针对普通光学图像研究提出的 UDA 方法直接应用于遥感图像往往效果不佳,这是因为获取遥感

图像的传感器通常具有显著差异,不同源数据特征、 分辨率会有不同,此外,不同的获取条件会使同一场 景的图像在视觉上也可能呈现较大的色彩差异,如 果只考虑具有相同空间分辨率遥感图像的 UDA<sup>[9]</sup>, 那么可用的数据将被严重压缩,模型推广性受限.因 此,遥感图像的无监督域适应不仅要缩小源域和目 标域之间的特征差异,而且要解决不同空间分辨率 的问题. 现有的研究少有明确考虑空间分辨率差异 问题的遥感图像 UDA 方法,大多数 UDA 方法通过 简单的插值[10-11]来处理分辨率问题,当源域和目标 域的分辨率差异不严重时,一些研究者[12]忽略了分 辨率问题. 例如 Zhang 等人[10]使用插值来解决源域 和目标域的分辨率差异,然后通过在特征空间中对 齐源域数据和目标数据的特征分布来消除域差异; Tang 等人[11]使用双线性插值提高低分辨率源域图 像的分辨率,并使用了双重域判别器;Zhang 等人[12] 忽略了遥感图像域间空间分辨率差异,提出了一种 分阶段域适应模型,在第一阶段设计了一个特征金 字塔融合模块,在第二阶段提出了一种对抗性自监 督训练方法. 现有的文献采用简单的插值,难以恢复 图像细节信息,如果两个域图像存在风格差异,简单 的插值也无法解决,而这些问题的解决将有利于更 好地实现模型迁移,提升模型的推广性能.此外,目 前还没有检索到有文献来研究集成域适应和超分辨 完成语义分割的端到端的网络模型.

为此,本文针对遥感影像语义分割中不同域间存在空间分辨率差异和特征差异的问题,提出了一个结合超分辨率和域适应的端到端的语义分割模型 SSM-SRDA,同时完成遥感图像的无监督超分辨率和语义分割任务,并通过域适应提高模型从低分辨率源域遥感数据到高分辨率目标域遥感数据迁移应用的性能.图1示意了本文的研究问题设置:根据有标签的源域低分辨率遥感图像和无标签的目标域高分辨率遥感图像,结合超分辨和域适应在源域学习得到一个语义分割模型来预测目标域图像的标签,完成分割任务.

考虑遥感图像不同域间存在的图像特征、分辨率、风格方面的差异,为了提升语义分割模型的推广性能,我们进行了研究,提出了 SSM-SRDA 模型,研究思路受启发于深度学习领域的三类研究:

(1) 深度域适应. 随着深度学习技术的崛起,深度域适应已成为域适应研究领域的主流[13]. 根据深度神经网络在域适应中作用的不同,深度域适应方法可分为四种:基于域分布差异的方法,通过最小化

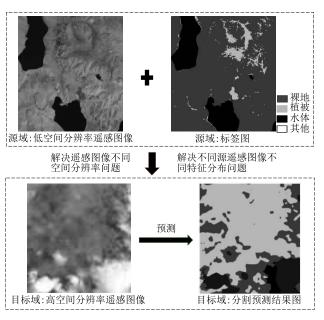


图 1 SSM-SRDA 模型研究问题设置示意图

两个域的分布差异对齐特征分布;基于对抗的方法,通过对抗学习来对齐分布;基于重构的方法,通过自编码器提取特征的同时抑制具有可迁移性信息的丢失;基于样本生成的方法,通过合成符合目标域分布的伪样本协助训练.本文采用对抗性域适应学习方法,其训练过程是特征提取器与域判别器之间的博弈过程,域判别器通过学习来区分源域特征与目标域特征,而特征提取器通过学习具有域不变性的特征表示来混淆域判别器,训练完成后的网络可以提取出具有域不变性的特征表示.

(2)图像超分辨率. 高空间分辨率的遥感图像 在各种实际应用中受到了广泛关注. 为了提高图像 的分辨率,已经提出了很多基于低-高分辨率图像对 的监督训练方法,然而,获取成对的低-高空间分辨 率遥感图像并不容易,在这种情况下,由于生成对抗 网络 GAN(Generative Adversarial Networks)[14] 的 无监督训练特性,基于 GAN 的无监督图像超分辨 率方法受到了更多关注. Wang 等人[15] 提出一种无 需低-高分辨率图像对的遥感图像无监督学习框架, 将内容相似的高分辨率参考图像转化为先验域迁移 图像,然后进一步从参考图像中获得纹理和结构信 息. Yang 等人[16]引入了一种更精确的方法:从参考 图像中获得相关纹理信息,并将其从参考图像迁移 学习到真实低分辨率图像中. 本文研究的源域与目 标域具有相似的内容,为将有分割标签的低空间分 辨率源域图像超分辨为目标域的高空间分辨率大 小,本文将目标域视为参考域,通过对抗学习生成具 有目标域风格的源域高分辨率图像.

(3)多任务(即图像超分辨率和分割)联合处理. 在遥感领域,包含更多细节信息的高分辨率图像对于语义分割任务非常重要[17].最近的研究表明,图像超分辨率和语义分割可以相互促进. Lei 等人[18]提出在分割网络中结合图像超分辨率,以获得超分辨率和分割任务的改进. Wang 等人[19]提出了一种双重超分辨率学习架构,将超分辨率的思想集成到现有的语义分割结构中,促使语义分割流的高分辨率特征被来自超分辨流的特征信息增强,保证低分辨率输入也能有高性能的高分辨率输出. Zhang 等人[20]在遥感图像数据集上验证了在分割网络中结合图像超分辨率,能够获得超分辨率和分割任务的改进. 但以上方法都是将两个任务联合处理的思想应用到语义分割监督学习任务中,而本文将这种思想拓展到遥感图像无监督域适应语义分割任务中.

综上,本文的主要贡献概括如下:(1)设计了一个用于遥感图像域适应的无监督超分辨率模块,缩小低空间分辨率源域和高空间分辨率目标域的空间分辨率和风格差异,促进后续域适应语义分割任务;(2)受超分辨率和语义分割相互促进的启发,在结合超分辨率的语义分割网络 SSNSR(Semantic Segmentation Network with Super Resolution)中设置了FA-Loss(Feature Affinity-Loss)模块,充分利用高分辨率图像的细节信息,提高语义分割模块的性能;(3)提出了一个端到端的结合超分辨率和域适应的语义分割模型 SSM-SRDA,其计算量远小于在域适应语义分割模型 SSM-SRDA,其计算量远小于在域适应语义分割任务前通过 GAN 生成具有目标域风格的源域参考数据的两阶段方法<sup>[7,9,40]</sup>.

实际遥感应用中,往往有一些数据有标签(即源域数据),但目标分割任务数据没有标签(即目标域数据),且这两类数据不同源,存在分辨率、特征等各方面的差异,我们基于超分辨和域适应思想设计的模型 SSM-SRDA 为解决不同于源域的目标域图像的分割任务提供了一种可行有效的方案,并通过实验进行了验证.

## 2 相关工作

#### 2.1 语义分割

语义分割是计算机视觉中最具挑战性的任务之一,它能够为输入的图像预测像素级语义标签,目前最新的语义分割方法主要基于深度神经网络. Long等人<sup>[21]</sup>提出可以将用于图像分类的卷积神经网络(如 AlexNet<sup>[22]</sup>、VGG<sup>[23]</sup>或 ResNet<sup>[24]</sup>)转换为用于

语义分割的全卷积网络,之后出现了多种方法来改进全卷积网络,如利用上下文信息[25]和多尺度特征融合[26].这些深度神经网络的训练依赖大量的有像素级标签的图像,为降低标签标注繁重的成本,近年来提出弱监督和半监督方法[27]完成语义分割任务.然而在大多数实际应用中,通过弱标注训练得到的模型不能很好地应用于无标签的数据集中,因此有必要通过域适应方法缩小数据集间的特征差异,使模型具有更好的推广性.

### 2.2 图像超分辨率

单幅图像超分辨率<sup>[28]</sup>旨在从相应的低分辨率 图像中恢复得到高分辨率图像. 传统的非 CNN 超 分辨率重建方法主要是基于插值的方法,但这类方 法恢复的高分辨率图像往往缺少细节信息,也不考 虑降低不同域图像存在的风格差异. Dong 等人<sup>[29]</sup> 提出的 SRCNN 首先将 CNN 引入单幅图像超分辨 率. FSRCNN<sup>[30]</sup>使用反卷积操作替代 SRCNN 模型 中的插值. ESPCN<sup>[31]</sup>的核心是亚像素卷积层,它通 过通道重排生成高分辨率图像,相比 FSRCNN 模 型中的反卷积层能学习到更多的细节特征. 本文在 无监督超分辨率模块使用亚像素卷积层,以获得 具有更多细节信息的高分辨率特征来提升后续分割 任务的效果. 此外,以降低感知损失<sup>[31]</sup>为目标,以 SRGAN<sup>[33]</sup>为代表的一系列方法通过使用对抗性训 练生成有更高视觉质量的超分辨率图像.

最近,不同于以往需要使用成对低-高分辨率图像来训练的超分辨率方法,基于 GAN 的超分辨率方法可以完成没有高分辨率图像标签的大型图像数据集超分辨率任务. Yuan 等人[34] 基于 CycleGAN提出一种具有多个 Cycle-in-Cycle 网络结构的无监督超分辨率算法. Jiang 等人[35] 提出一个基于 GAN的边缘增强网络(EEGAN),用于重建具有清晰边缘的高分辨率遥感图像. DASR[36] 将参考域高分辨率的图像生成具有真实图像风格的低分辨率合成图像,通过参考域的高-低分辨率图像对训练网络. 受DASR 的启发,本文的方法通过采样目标域的高分辨率图像,组成目标域低-高分辨率图像对,在目标域图像对训练超分辨率模块的同时,域判别器帮助源域生成高分辨率图像,学习到目标域的细节纹理特征.

#### 2.3 深度域适应

深度域适应领域针对源域向目标域迁移应用的问题出现了很多方法,其主要解决思想是对齐源域和目标域图像的特征分布. Ganin 等人[37]提出域对

抗神经网络 DANN 来对齐特征分布,此后出现了许 多具有不同损失函数或分类器的变体方法[38-39]. 针 对语义分割任务的域适应方法侧重于采用对抗式学 习的方法或使用生成网络将源域图像转换为类似目 标域的图像. Hoffman 等人[40]提出了一种基于全卷 积网络的无监督域自适应语义分割方法,对源域和 目标域特征进行判别训练,使两个域具有相似的特 征分布. CyCADA[41]使用 CycleGAN[42]通过对抗性 学习将源域图像生成具有目标域风格的额外训练数 据,缩小像素级特征差异. Tasar 等人[9] 为缩小训练 和测试数据光谱波段的差异,生成与训练图像语 义上完全相同但其光谱分布与测试图像分布相似 的伪训练图像.考虑到源域和目标域图像超分辨率 后在外观上具有明显差异,故本文除了对两个域的 SoftMax 输出采用判别训练,对源域和目标域的超 分辨率输出也采用了判别训练.

目前,遥感图像深度域适应语义分割领域少有解决不同域间差异的方法,特别是针对空间分辨率差异的方法.Wu等人[11]通过上采样或下采样的方式,使源域和目标域具有相同的分辨率后,再训练域适应语义分割模型,有效提高了对无标签目标域的分割性能,但该方法中也只使用了简单的插值法在输入级提高源域的分辨率.我们认为有必要结合超

分辨率和对抗性域适应方法解决遥感图像域适应领域面临的空间分辨率差异和特征差异问题.

## 3 本文方法

## 3.1 模型结构

本文提出的结合超分辨率和域适应的语义分割 模型 SSM-SRDA 由三部分组成:一是结合超分辨 率的语义分割网络(Semantic Segmentation Network with Super Resolution, SSNSR);二是像素级 判别器(Pixel-level Domain Discriminator, PDD); 三是输出空间级判别器(Output-space Domain Discriminator, ODD). 模型结构如图 2 所示,模型的输 入分别为源域和目标域,源域为较低分辨率的有标签 数据,目标域为较高分辨率的无标签数据. SSNSR 将 超分辨率模块结合到分割模块中,二者共享同一个 特征提取网络 ASPPnet<sup>[43]</sup>. 训练阶段,源域图像和 下采样目标域图像被送到 SSNSR 中,超分辨率模 块获得来自 ASPPnet 的特征图后,将有低高分辨 率图像对的目标域视为无监督超分辨率的参考域, 通过与像素级判别器 PDD 的生成对抗训练,对低分 辨率的源域进行无监督超分辨率学习,得到具有高 分辨率目标域图像细节纹理特征的高分辨率源域图

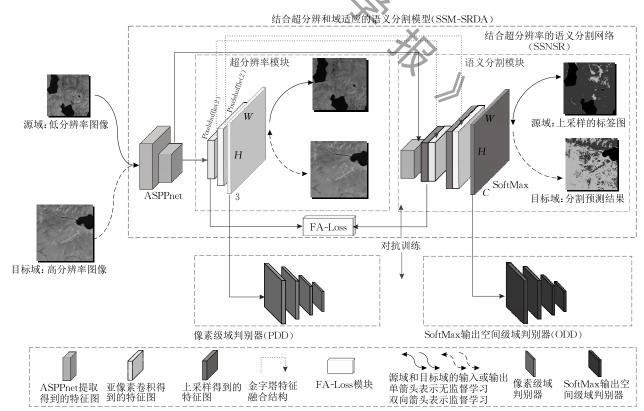
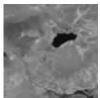


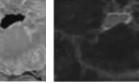
图 2 结合超分辨率和域适应的语义分割模型(SSM-SRDA)网络结构图

像,即对于源域图像完成了超分辨和与目标域图像 的风格差异处理. 分割模块得到来自 ASPPnet 的特 征图后,在每一次卷积操作前都与来自超分辨率模 块的特征图进行融合,对有分割标签的源域进行监 督学习,对目标域进行无监督学习,通过与输出空间 级判别器 ODD 的对抗训练,缩小两个域预测分割 结果的特征差异.其中,PDD的输入是超分辨率模 块输出的高分辨率图像,ODD 的输入是分割模块 SoftMax 层输出的预测分割结果. 测试阶段,目标域 测试图像被送入训练好的 SSNSR 中,得到预测分 割结果.

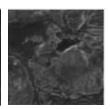
本文的目标是学习一个在无标签高分辨率目标 域图像上也应用良好的分割模型,如图2所示,为了 使超分辨率模块恢复的特征给分割模块解码器通过 上采样得到的特征图带来更多细节结构信息,提升 语义分割效果,本文在 SSNSR 的设计中使用了两 种策略:(1)为了将超分辨率模块提取得到的特征 迁移到分割模块,在两个模块之间建立了金字塔特 征融合结构[44],将超分辨率模块每一层得到的特征 图迁移融合到分割模块的对应层;(2)在超分辨率 和分割模块之间引入特征相似性 FA-Loss 模块,语 义分割模块在深层特征保留的细节结构信息较少、 语义信息较多,其深层特征的细节结构信息将被具 有特征相似性的超分辨率模块高分辨率特征进一步 增强. 图 3 为将 FA-Loss 模块的输入可视化后得到 的语义分割流与超分辨率流特征图可视化,通过比 较图 3(b)和(c),可以发现来自超分辨率流的图 3 (c)包含了更完整的结构信息,虽然这些结构还没有 明确表示出它们的类别,但是可以通过像素与像素 或者区域与区域之间的关系传递的语义信息进行有 效分类.



(a) 输入图像







(c) 超分辨率流特征 可视化

图 3 输入 FA-Loss 模块的语义分割流和超分辨率流 特征图可视化

在方法介绍中定义如下数学符号表示:将低分 辨率遥感数据集的源域 S 提供的低分辨率图像表 示为 $I_s$ 、其像素级标签为 $A_s$ ;将高分辨率遥感数据 集的目标域 T 提供的高分辨率图像表示为  $I_{T}$ . S 和 T 共享相同的标签空间 $\mathbb{F}^c$ ,其中 C 表示类别的数 量. 根据上面的定义,本文的目的是缩小S和T之 间的域差异,包括空间分辨率差异和特征差异,并通 过给定的  $I_s$ 、 $A_s$ 和  $I_T$ 学习一个语义分割模型来预测 目标域 T 的分割结果. 下面首先阐述结合超分辨的 语义分割网络 SSNSR,然后详细介绍 FA-Loss 模块 和基于对抗的像素级和输出空间级域适应方法.

#### 3.2 结合超分辨的语义分割网络 SSNSR

对于遥感图像,不同源数据的空间分辨率及场 景、获取条件等差异都是非常重要的域差异,严重影 响分割模型的迁移应用能力,消除空间分辨率及数 据特征方面的差异对于遥感图像跨域进行语义分割 具有重要意义. 很多语义分割任务中的解码器是简 单的上采样层,不能给低分辨率的输入带来有用的 附加信息,而非域适应领域的一些研究[18-20]表明超 分辨率和语义分割可以相互促进. 基于以上研究,本 文提出了一个结合超分辨的语义分割网络 SSNSR, 来减小源域和目标域之间空间分辨率和特征级别的 差距,其输入为源域图像和下采样为相同分辨率大 小的目标域图像. 超分辨率模块和语义分割模块的 输出均与高分辨率目标域图像大小相同,且输入与 输出图像分辨率的倍数差异和两个域的空间分辨率 的倍数差异相同. 下面分 SSNSR\_1 和 SSNSR\_2 两 个部分来介绍该模块. SSNSR\_1 由超分辨率模块和 分割模块通过金字塔特征融合结构连接组成,二者 共享同一个特征提取网络 ASPPnet<sup>[43]</sup>, SSNSR\_2 在 SSNSR\_1 的基础上增加特征相似性模块 FA-Loss. SSM\_SRDA 训练阶段使用 SSNSR\_2,测试阶段使 用 SSNSR\_1.

SSNSR\_1 的网络结构如图 4 所示. SSNSR\_1 网络使用残差 ASPPnet 作为共享的特征提取器. 超 分辨率模块由两层 PixelShuffle 层和一层上采样层 组成. 普通上采样插值或者反卷积方法只是增加了图 像的大小,并不能增强图像的高频细节信息,甚至会 带来额外误差,因此引入高效的亚像素卷积 Pixel-Shuffle 方法[31]. PixelShuffle 将低分辨率的特征图 通过卷积和通道重排恢复为高分辨率的特征图,可 以实现不同整数放大尺度的高分辨率图像重构. 第 一层 PixelShuffle 的输入为 ASPPnet 提取得到的特 征图,其大小为源域图像的1/2.由于本文实验中的数 据空间分辨率差异在2到4倍之间,因此两层 Pixel-Shuffle 层的放大因子均为 2,每经过一层亚像素卷 积层,特征图大小变为原来的2倍,通道数减为原来

(2)

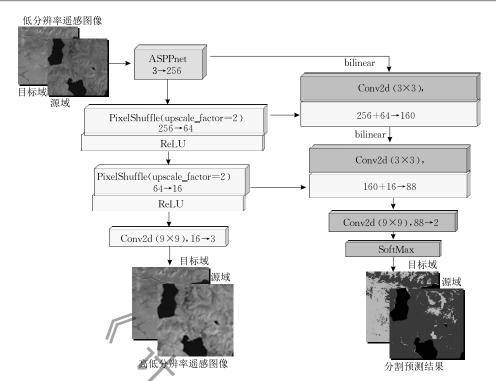


图 4 SSNSR\_1 网络结构(左为超分辨率模块,右为分割模块)

的 1/4,之后再接一层上采样层将特征图恢复到高分辨率目标图像大小.每个 PixelShuffle 层后接一个 ReLU 激活函数,超分辨率模块特征图通道数变化为 256,64,16,3.分割模块包括三层卷积层,为保证与超分辨率模块对应层的特征图大小相同,在每一个卷积层前都通过双线性插值的方法将特征图恢复到与超分辨率模块对应层的输出相同的大小,最后再通过一个 SoftMax 层输出预测的分割结果,其特征图通道数变化为 320,176,88,2.

SSNSR\_1的损失  $L_{SSNSR_1}$  由源域的分割损失  $L_{seg}$ 、源域的超分辨损失  $L_{idS}$ 和目标域的超分辨损失  $L_{idT}$ 构成. 其中,源域的分割损失  $L_{seg}$  由两部分组成:一是源域图像的分割预测结果和其上采样后的标签图计算的交叉熵损失;二是下采样源域超分辨后的高分辨率图像再送入分割网络得到的分割预测结果和其上采样后的标签图计算的交叉熵损失. 目标域的超分辨损失  $L_{idT}$  由图像超分辨率领域最广泛使用的均方误差损失函数计算. 由于源域图像没有对应的高分辨率图像标签,故本文采用感知损失  $L_{per}$  [32] 和定点损失  $L_{fp}$  [45] 加权求和来计算  $L_{idS}$ ,感知损失的目的是计算超分辨率后的图像和上采样的图像在样式与语义上的差异,定点损失的目的是比较下采样超分辨后的图像和原图提取得到的特征差异. 具体的公式表示如下:

$$L_{\text{SSNSR\_1}} = \alpha L_{\text{seg}} + \beta (L_{idT} + L_{idS})$$

$$L_{\text{seg}} = L_{\text{cel}} (S(I_S), \uparrow A_S) + L_{\text{cel}} (S( \downarrow R(I_S)), \uparrow A_S)$$

$$(1)$$

$$L_{idT} = L_{mse}(R(\downarrow I_T), I_T)$$
 (3)

$$L_{idS} = L_{per}(R(I_S), \uparrow I_S) + 0.5 \times L_{fp}$$
 (4)

$$L_{fp} = L_{L1}(E(\langle R(I_S)), E(I_S))$$
 (5)

其中,S,R 和 E 分别表示分割模块、超分辨率模块和共享的特征提取器, $\uparrow$  和  $\downarrow$  分别表示上采样和下采样操作.  $L_{cel}$  表示二维交叉熵损失, $L_{mse}$  表示均值方差损失, $L_{Li}$  表示 L1 标准损失.  $\alpha$  和  $\beta$  表示语义分割损失和超分辨率损失的加权因子.

#### 3.3 特征相似性 FA-Loss 模块

相较于语义分割模块,超分辨率模块提取得到的特征包含更完整的结构信息,因此引入特征相似性来帮助增强语义分割解码器特征图的结构信息.如式(6)所示, $L_{fa}$ 损失的目的是学习超分辨率与分割模块深层特征相似性矩阵之间的距离,若距离值小,其在 $L_{fa}$ 损失中的占比也小,证明超分辨率模块和分割模块提取得到特征的结构信息是相似的;若距离值大,其在 $L_{fa}$ 损失中的占比也大,促进分割模块学习超分辨率提取得到特征间的结构信息.相似性矩阵描述成对像素之间的相似关系,如式(7)所示.其中 $W' \times H'$ 表示空间维度, $C_{ij}$ 表示第i个和第j个像素之间的关系, $C_{ij}^{s}$ 表示语义分割模块特征图的相似性矩阵, $C_{ij}^{s}$ 表示超分辨率模块特征图的相似

性矩阵. 对于  $W' \times H' \times C'$  的特征图 F,需要计算每两个像素之间的相似关系,即关系图包含  $W'H' \times W'H'$  对连接,关系图中的每个值是其不同通道的相似关系之和. 为减少计算量,在超分辨率模块第一层和语义分割解码器第一层的特征图上计算每对像素的相似度. 实际操作中,在计算  $L_{fa}$  损失之前,对输入的特征图进行了特征维度的变换,由  $1 \times 1$  卷积层、BatchNorm 和 ReLU 层组成.

$$L_{fa} = \frac{1}{W^{\prime 2} H^{\prime 2}} \sum_{i=1}^{W^{\prime} H^{\prime}} \sum_{j=1}^{W^{\prime} H^{\prime}} \| C_{ij}^{S} - C_{ij}^{R} \|_{1}$$
 (6)

$$C_{ij} = \left(\frac{F_i}{\|F_i\|_2}\right)^{\mathsf{T}} \cdot \left(\frac{F_j}{\|F_j\|_2}\right) \tag{7}$$

将特征相似性 FA-Loss 模块加入 SSNSR 之后的损失函数为

$$L_{\text{SSNSR}_2} = \alpha L_{\text{seg}} + \beta (L_{idT} + L_{idS}) + \gamma L_{fa}$$
 (8)  $\gamma$  表示  $L_{fa}$  损失的加权因子.

#### 3.4 像素级域判别器

传感器、地理位置、成像条件和其他因素带来的 影响造成了遥感图像的固有差异,本文提出的 SSNSR 消除了源域和目标域之间的分辨率差异,没有减少 其他外观上的域差异(如纹理、颜色等),所以构建 PDD 域判别器进行对抗性域适应学习来减小域间 的风格差异. PDD 模块的目的是将源域图像向高分 辨率目标域图像的风格靠拢,因此选择目标域数据 集的高分辨率图像和源域超分辨率后得到的高分辨 率图像作为 PDD 的输入, PDD 判别器判断每个像素 来自源域或目标域,通过 SSNSR 和判别器的迭代训 练,当判别器无法分清该像素的域类时,表示源域和 目标域数据的特征已经对齐. 传统 GAN 判别器的设 计是输出一个对生成器生成的整幅图的评价值(真或 假),实验中应用了更能关注到图像细节的 Patch-GAN 作为 PDD 判别器, PatchGAN 使用卷积将输入 映射为 N×N 矩阵,矩阵中每个点都代表输入图像中 的一块区域,更有助于高分辨率图像体现更多局部特 征. PDD 损失  $L_{PDD}$ 定义如下:

$$I_{S}^{R} = R(I_{S}) \in \mathbb{F}^{H \times W \times 3}$$
 (9)

$$L_{\text{PDD}} = \mathbb{E}_{I_{S} \sim \rho_{\text{data}}(I_{S})} \left[ \log(1 - D_{\text{pdd}}(I_{S}^{R})) \right] + \mathbb{E}_{I_{T} \sim \rho_{\text{data}}(I_{T})} \left[ \log(D_{\text{pdd}}(I_{T})) \right]$$
(10)

其中,H 和 W 表示高分辨率目标域图像的高度和宽度, $p_{\text{data}}()$ 表示数据的分布, $\mathbb{E}()$ 表示分布函数的期望值, $D_{\text{pdd}}$ 是 PDD 模块, $I_T$ 为目标域高分辨率的图像,而 $I_5^{\mu}$ 为源域经过超分辨率模块后得到的高分辨率图像.

## 3.5 输出空间级域判别器

在遥感应用中,无论是来自源域还是目标域的图

像,其语义分割的输出图在空间和布局上都具有很强的相似性,如长方形的道路区域可能覆盖了汽车、行人的部分;建筑物周围经常生长着绿色植物等等. 仅在像素空间中进行对抗性域适应不足以使源域图像和目标域图像的分割预测结果  $P_s$  和  $P_{\tau}$  的特征分布更加接近,因此对分割模型 SoftMax 层预测的类别输出进行对抗性域适应学习来减小域间的特征差异. ODD 判别器由 5 个卷积核为  $4\times4$ 、步长为 2 的卷积层组成,其通道数分别为 64、128、256、512、1. ODD 损失  $L_{\text{ODD}}$  定义如下:

$$P_{S} = S(I_{S}) \in \mathbb{F}^{H \times W \times C} \tag{11}$$

$$P_T = S(\mathbf{1}_T) \in \mathbb{F}^{H \times W \times C} \tag{12}$$

$$L_{\text{ODD}} = -\sum_{h \cdot w} (1 - z) \log(D_{\text{odd}}(P_T)) + z \log(D_{\text{odd}}(P_S))$$
(13)

其中  $D_{\text{odd}}$ 是 ODD 模块, H 和 W 表示高分辨率目标域图像的高度和宽度, C 是类别数. 首先将源域和目标域图像经过分割模型 SoftMax 层之后的预测结果  $P_s$ 和  $P_T$ 作为 ODD 的输入, 然后通过 ODD 来判断输入样本的域类(源域或目标域), 如果样本来自目标域,则 z=0; 如果样本来自源域,则 z=1.

## 3.6 SSM-SRDA模型的目标函数

类似于标准的生成对抗网络,将 SSNSR\_2 作为生成器,PDD 和 ODD 作为判别器,对抗训练的最终目的为优化公式(14),最小化生成器 SSNSR\_2 的分割损失和超分辨损失,同时最大化目标域在判别器 D(包括 PDD 和 ODD)中被判别为源域的概率.

$$\min_{\text{SSNSR}_2} \max_{\text{D}} L_{\text{SSNSR}_2} + L_{\text{PDD}} + L_{\text{ODD}} \tag{14}$$

SSM-SRDA 模型的参数的优化是生成器和判别器交替迭代的优化过程. 首先,固定生成器的参数  $\theta_{SSNSR_2}$ 不变,最大化 PDD 和 ODD 损失函数来优化判别器的参数  $\theta_D$ ,然后,固定判别器的参数  $\theta_D$ ,并最小化 SSNSR\_2 损失而得到  $\theta_{SSNSR_2}$ .通过交替迭代训练,最终找到双方的损失平衡点,也就是生成数据拟合真实数据时,参数  $\theta_D$ 和  $\theta_{SSNSR_2}$ 都不再变化,SSM-SRDA 模型达到收敛状态,训练结束.

算法1描述了SSM-SRDA模型的训练、测试过程. 算法1. SSM-SRDA模型的训练、测试过程. 训练过程:

输入:源域数据集图像  $I_S$ ,目标域数据集图像  $I_T$ 和源域数据集标签  $A_S$ 

输出:具有目标域风格的高分辨率源域图像  $I_s^R$ ,目标域图像的分割预测标签  $A_T$ 

1. 通过 SSNSR\_2 的超分辨率和分割模块分别得到高 分辨率图像和 SoftMax 分类输出

$$I_S^R$$
,  $I_T^R = R(I_S, \bigvee I_T) \in \mathbb{R}^{H \times W \times 3}$   
 $P_S$ ,  $P_T = S(I_S, \bigvee I_T) \in \mathbb{R}^{H \times W \times C}$ 

2. 得到目标域的预测标签图

$$A_T = \max(P_T) \in \mathbb{R}^{H \times W \times 1}$$

3. 通过 PDD 域判别器区分超分辨图像的域类  $I_{\text{fake}}\,,I_{\text{true}}\!=\!D_{\text{pdd}}(I_{S}^{R}\,,I_{T})\!\in\!\mathbb{R}^{H\times W\times 1}$ 

- 4. 通过 ODD 域判别器区分 SoftMax 输出的域类  $P_{\rm true}\,,P_{\rm fake}=D_{\rm odd}(P_S\,,P_T)\in\mathbb{R}^{H\times W\times 1}$
- SSNSR\_2 与 PDD 和 ODD 的对抗性训练 如式(14)所示,通过 SSNSR\_2 与 PDD 和 ODD 的交 替训练进行 θ<sub>0</sub> 和 θ<sub>SSNSR\_2</sub> 的更新.

重复上述过程直到模型收敛.

测试过程:

输入:目标域数据集图像  $I_T$ 

输出:目标域图像的分割预测标签  $A_T$ 

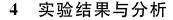
1. 通过 SSNSR\_1 的超分辨率和分割模块分别得到高分辨率图像和 SoftMax 分类输出

$$I_T^R = R(\downarrow I_T) \in \mathbb{R}^{H \times W \times 3}$$

$$P_T = S(\bigvee I_T) \in \mathbb{R}^{H \times W \times C}$$

2. 得到目标域的预测标签图

$$A_T = \max(P_T) \in \mathbb{R}^{H \times W \times 1}$$



#### 4.1 数据集

#### 4.1.1 二类语义分割数据集

使用Massachusetts Buildings Dataset,Inria Aerial Image Labeling Dataset 以及 WHU Building Dataset 三个建筑物数据集进行 Mass-Inria 和 Mass-WHU 两组实验.实验选择同一个有标签的源域数据集和不同源的无标签目标域数据集,以验证 SSM-SRDA 模型的泛化能力,较强的推广能力使其可以迁移应用到多个不同分布的目标域的分割任务,适应实际应用的需求.

Massachusetts Buildings Dataset 由 151 幅空间分辨率为 1 m 的波士顿地区航空图像组成,提供了两个语义类别:建筑物类和背景类(非建筑物). 将该数据集视为源域. 在 Mass-Inria 实验中,由于 Mass 与 Inria 数据集的空间分辨率差异大概为 3. 33 倍,所以将 Mass 建筑物训练集图像和标签裁剪为 114×114 像素,然后将标签图插值为 380×380 大小;在 Mass-WHU 实验中,由于 Mass 与 WHU 数据集的空间分辨率差异为 4 倍,所以将 Mass 建筑物训练集图像和标签裁剪为 128×128 像素,然后将标签图插值为 512×512 大小.

Inria Aerial Image Labeling Dataset 由 360 幅空间分辨率为 0.3 m 的航空图像组成,同样也提供了两个语义类别. 位置标号为 1 到 5 的图像组成测试集,6 到 36 组成训练集,将该数据集的训练集视为目标域,并在该数据集的测试集上验证了模型的分割精度.将 Inria 建筑物数据集图像和标签无重叠裁剪为 380×380 的 3875 张训练集图像和 1600 张测试集图像.

WHU Building Dataset 原始空间分辨率 $0.075\,\mathrm{m}$ ,实验中使用 Ji 等人 $[^{46}]$ 提供的下采样后空间分辨率为  $0.25\,\mathrm{m}$ 、大小为  $512\times512$  的  $8189\,\mathrm{张图片}$ . 将该数据集的训练集视为目标域,其中  $5773\,\mathrm{张用于训练}$ ,2416 张用于测试.

## 4.1.2 多类语义分割数据集

### (1) Vaih-Pots

使用 ISPRS 提供的 Vaihingen 和 Potsdam 两个城市分类数据集进行多类域适应语义分割实验. 每个数据集都有 6 个标签类别:不透水面、建筑物、低矮植被、树木、汽车、背景.

ISPRS Vaihingen 2D Semantic Labeling Challenge 数据集包括 33 张大小不同的遥感图像,空间分辨率为 9 cm,覆盖德国的 Vaihingen 市. 将该数据集中由近红外、红色和绿色 3 个波段组成的图像作为源域. ISPRS Potsdam 2D Semantic Labeling Challenge dataset 数据集由 38 张空间分辨率为 5 cm 的遥感图像组成,覆盖了德国 Potsdam 市. 实验中取标号从 2-10 到 5-12 的 19 幅图像作为训练集,其他 19 幅图像作为测试集. 实验中,由于源域与目标域的空间分辨率差异约为 2 倍,因此将 Vaih 数据集图像和标签裁剪为  $180 \times 180$  像素大小,将 Pots 数据集裁剪为  $360 \times 360$  像素大小.

#### (2) 青藏高原数据集 TPD\_S-TPD\_T

青藏高原数据集(Tibet Plateau Dataset, TPD) 选取了青藏高原地区 1990 年和 2020 年不同传感器 拍摄的不同空间分辨率遥感图像.

源域数据集 TPD\_S 使用 1990 年由对地观测卫星 Landsat 5 拍摄的青藏高原东部植被覆盖显著区域的两张 8192×8192 像素大小的遥感图像,其空间分辨率为 90 m. 为有利于对地表植被的识别,训练使用由短波红外波段、近红外波段、红波段组成的 RGB图像.此数据集先通过人工少量标注,再送入 Envi 使用监督分类得到精细的标签数据,共分 4 类,分别为植被、裸地、水体和其他.

目标域 TPD\_T 使用 2020 年由对地观测卫星

Landsat8 拍摄的青藏高原地区一张 5083×6454 像 素大小遥感图像,其空间分辨率为 30 m. 图像制作 和标签标注的设置与源域相同,实验中,由于源域与 目标域的空间分辨率差异为3倍,因此将源域训练 集图像和标签裁剪为 86×86 像素大小的 4536 张图 像,然后将标签图插值为 256×256 大小. 将目标域 裁剪得到 400 张 256×256 的训练集图像,以及 100 张测试集图像.

#### 4.2 评价指标和实验设置

#### 4.2.1 评价指标

#### (1) 交并比 IoU

在图像语义分割领域,采用交并比 IoU(Intersection-over-Union)作为主要评价指标,其定义如下:

$$IoU(P_m, P_{gt}) = \left| \frac{P_m \cap P_{gt}}{P_m \bigcup P_{gt}} \right| \tag{15}$$

其中, $P_m$ 为预测分割值, $P_{gt}$ 为真实值. mIoU(MeanIntersection-over-Union)通过计算多个类交并比的 平均值来评估模型性能.

(2) F1-score

$$F1 = \frac{precision \cdot recall}{precision + recall}$$

F1 分数(F1-score),被定义为精确率 precision

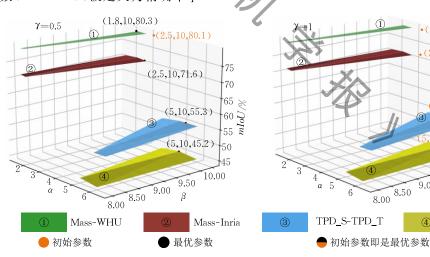


图 5 不同权重组合损失函数的 SSM-SRDA 在四组数据集上的 mIoU

图 5 表示的实验结果可看出, SSM-SRDA 使用 初始的参数取值在不同的数据集上能得到较好的结 果. 在不同数据集中,对参数进行微调,可能会有更 好的mIoU,但mIoU 的增长不超过 0.5%,说明本文 的方法 SSM-SRDA 对参数设置不敏感, 如在 Mass-WHU 实验中,固定  $\gamma$  为 0.5 和  $\beta$  为 10, $\alpha$  取 1.8 时的 mIoU 仅比α取 2.5 时多 0.2%.

#### 4.2.4 对比方法设置

实验采用 Resnet-101<sup>[24]</sup>结构的语义分割网络

和召回率 recall 的调和平均数. 在多分类任务中, 分别计算各个类的 precision 和 recall,得到各自的 F1-score 值,然后取平均值得到 Macro-F1.

#### 4.2.2 实验设置

本文在 24 GB 内存的 NVIDIA GeForce RTX™ 3090 GPU 以及深度学习框架 PyTorch 上实现并训 练了提出的 SSM-SRDA 模型. 使用 Adam 优化器, 动量为 0.9. 在实际训练过程中,每次实验都将全部 样本训练 100 次,前 50 次以 2×10<sup>4</sup>的学习率对模型 进行训练,后50次的学习率逐次递减直至为0.

## 4.2.3 损失函数权重值设置

式(8)中 $\alpha$ 、 $\beta$ 和 $\gamma$ 分别表示语义分割模块、超分 辨率模块和 FA-Loss 模块损失值的权重,为保证模 型的稳定性,使各个模块损失值的取值范围基本一 致,在二类分割数据集实验中初始设置  $\alpha$  为 2.5、 $\beta$ 为 10、γ 为 0.5,在多类分割数据集实验中初始设置  $\alpha$  为 5、 $\beta$  为 10、 $\gamma$  为 0.5. 手动对  $\alpha$ 、 $\beta$  和  $\gamma$  进行微调 后,SSM-SRDA 在四组数据集上的 mIoU 如图 5 所 示,左图中固定 $\gamma$ 为0.5,右图中固定 $\gamma$ 为 $1,\alpha$ 和 $\beta$ 在 两个图中的取值相同,在图中标出了在各个数据集 上初始参数值和获得最高 mIoU 值的最优参数值.

·(2.5,10,79.9)

75

70

50

10.00

Vaih-Pots

作为基线模型,为 NoAdapt. 不同于 SSNSR\_1 使用 超分辨率模块提高了源域图像分辨率,NoAdapt 不对 源域和目标域的空间分辨率差异作任何处理,仅仅用 源域数据集的训练集进行模型训练,然后用目标域测 试集进行测试. SSNSR 1 使用残差 ASPP 网络[41] 作 为特征提取器,其参数量小于 Resnet-101<sup>[24]</sup>的 1/2. 本文设置了5组消融实验: NoAdapt 和 SSNSR\_1,验 证超分辨率模块的作用;SSNSR\_1 和 SSNSR\_2 及 SSNSR\_1+PDD+ODD 和 SSM-SRDA(SSNSR\_2

9.00

+PDD+ODD)两组实验,验证了 FA-Loss 模块的作用;SSNSR\_1和SSNSR\_1+PDD,验证了 PDD 的作用;SSNSR\_1和SSNSR\_1+ODD,验证了 ODD 的作用.

现有研究中解决遥感图像不同域间空间分辨率 差异和特征差异的工作较少,我们选择以下三个具 有可比性的、开源的无监督对抗性域适应方法进行 比较:AdaptSegnet<sup>[8]</sup>、FCAN<sup>[6]</sup>和 SRDA<sup>[11]</sup>.不同于 AdaptSegnet 在语义分割模型的输出空间中采用域 判别器,本文的 SSM-SRDA 在像素级和输出空间采 用了双重判别器, FCAN 是一个两阶段的网络, 先使用 CycleGAN 将源域图像转化为目标域风格的图像,然 后将对抗性域判别器应用于分割网络的较浅层来捕 获图像的外观差异,而本文将超分辨率模块集成到 分割模块中获得具有目标域风格的源域高分辨率 图像,两个模块共享同一个特征提取网络,计算量 远小于两阶段的 FCAN 方法,并对高分辨率图像和 分割模块的输出使用双重对抗性域判别器. SRDA 使用双线性插值的方法来提高源域的分辨率。本文 使用了更强调恢复图像细节信息且更高效的亚像素 卷积方法,并通过 FA-Loss 模块使超分辨率模块提 取到的细节信息提高语义分割的效果. 为和本文的 模型有可比性,对忽略了图像分辨率问题的 Adapt-Segnet 和 FCAN 方法使用上/下采样的方式对齐源 域和目标域空间分辨率后再进行实验,并且在实验 中省略了 SRDA 的预训练过程.

## 4.3 消融实验结果及分析

## 4. 3. 1 Mass-Inria

Mass-Inria 数据集上的 SSM-SRDA 的逐步消融实验结果如表 1 所示,粗体值表示每列的最高精度.表中的测试时间为目标域测试集中每张图的测试时间.在无域适应的方法中,相对于忽略分辨率差异的 NoAdapt 方法,增加超分辨率模块的 SSNSR\_1 方法显著提高了建筑物类的 IoU(从32.7%提高到 37.8%,提高了 5.1%),不仅说明了在分割模块中结合图像超分辨率的有效性,也证明了 SSNSR\_1 方法通过消除分辨率差异获得的增益要大于使用上采样源域标签引入的误差.在超分辨率模块和语义分割模块间增加 FA-Loss 模块后,SSNSR\_2 的建筑物类 IoU 较 SSNSR\_1 提高了0.9%,mIoU提高了0.4%,说明超分辨率得到的细节信息可以提高对目标对象语义分割的效果.在SSNSR\_1的基础上分别加入 PDD 和 ODD 判别器,

根据 SSNSR\_1+PDD 和 SSNSR\_1+ODD 的实验结果发现,其 mIoU 较 SSNSR\_1 分别提升了 4.7% 和 2.4%,表明在分割类别较少的数据集中,学习域不变性特征方面 PDD 的作用比 ODD 更大. 在 SSNSR\_1 上集成 PDD 和 ODD 来减小特征差异的实验获得了更好的分割效果,较 SSNSR\_1 的 mIoU 提高了 8.1%,比只集成 PDD 或 ODD 在 mIoU 指标上提升超过 3%,F1-score 指标上提升超过 7%.增加 FA-Loss 模块后的 SSM-SRDA 取得了最好的mIoU 值(71.6%)及 F1-score 值(69.2%).

表 1 Mass-Inria 数据集上 SSM-SRDA 的逐步消融实验

方法	建筑物	背景 /%	mIoU / $%$	$_{\rm score/\%}^{F1\text{-}}$	训练 时间/h	测试 时间/s
NoAdapt	32.7	83.2	58.0	49.1	18	0.15
SSNSR_1	37.8	88.4	63.1	54.8	22	0.13
SSNSR_2	38.7	88.2	63.5	56.0	26	0.13
SSNSR_1+ PDD	45.4	90.2	67.8	61.1	29	0.16
SSNSR_1+ ODD	42.7	88.3	65.5	59.3	28	0.15
SSNSR_1+ PDD+ODD	52.1	90.3	71.2	68.5	35	0.18
SSM-SRDA	52.6	90. 5	71.6	69. 2	39	0.18

## 4. 3. 2 Mass-WHU

Mass-WHU 数据集上 SSM-SRDA 的逐步消融 实验结果如表 2 所示,表中的测试时间为测试集中 每张图的测试时间. 无域适应的方法中,使用超分辨 率 PixelShuffle 方法消除分辨率差异的 SSNSR\_1 显著提高了mIoU(从 69.1%提高到 73.7%,提高了 4.6%),说明了在分割模块中结合图像超分辨率的有 效性. 由于 SSNSR 1 中的特征提取网络为残差 AS-PP 网络<sup>[40]</sup>,其参数量远小于 Resnet-101<sup>[23]</sup> 网络,因 此即使 SSNSR\_1 中增加了超分辨率模块, SSNSR\_ 1与NoAdapt的测试时间基本相同(0.15s).同样的, 如表 1 所示,在源域和目标域空间分辨率差异更小的 Mass-Inria 数据集上, SSNSR\_1 的测试时间(0.13s) 小于 NoAdapt 的测试时间(0.15s). 根据 SSNSR\_1+ PDD 和 SSNSR\_1+ODD 与 SSNSR\_1 的实验结果对 比可以发现,加入 PDD 判别器后建筑物类的 IoU 提 升了 4.9%,加入 ODD 判别器后建筑物类的 IoU 提 升了 1.8%,在 SSNSR\_1 上集成 PDD 和 ODD 后,对 建筑物类有更好的分割效果(64.1%). 增加 FA-Loss 模块的 SSNSR\_2 与 SSM-SRDA(SSNSR\_2+PDD +ODD) 比 SSNSR\_1 与 SSNSR\_1+PDD+ODD 的 mIoU 分别提高了 0.8%、1.0%,提升效果明显,说 明超分辨率模块提取的特征信息对语义分割有促进 作用.

表 2 Mass-WHU 数据集上 SSM-SRDA 的逐步消融实验

方法	建筑 物/%	背景 /%	$^{mIoU}_{/\%}$	F1- score/%	训练 时间/h	测试 时间/s
NoAdapt	46.8	91.3	69.1	63.7	18	0.15
SSNSR_1	54.2	92.7	73.7	70.3	18	0.15
SSNSR_2	55.7	93.2	74.5	71.8	23	0.15
SSNSR_1+ PDD	59.1	94.0	76.6	74.6	19	0.20
SSNSR_1+ ODD	56.0	94.2	75. 1	72.1	19	0.20
SSNSR_1+ PDD+ODD	64.1	94.4	79.3	78.3	24	0.21
SSM-SRDA	66.0	96.7	80.3	80. 2	27	0.21

图 6显示了 Mass-WHU 数据集上逐步消融实验

的部分分割结果图.第一列是WHU测试数据集中的图像,第二列为真实标签,其余各列分别是SSNSR\_1、SSNSR\_1+PDD、SSNSR\_1+ODD、SSNSR\_1+PDD+ODD以及SSM-SRDA(SSNSR\_2+PDD+ODD)的预测分割结果图.可以观察到,加入PDD和ODD后,一些错误分割得到了有效的消除.增加FA-Loss模块后的SSM-SRDA分割结果的结构性和细节性有明显提升.此外,可以明显观察到在逐步消融实验中分割结果图变得越来越精确,小型建筑物的结构越来越清晰.

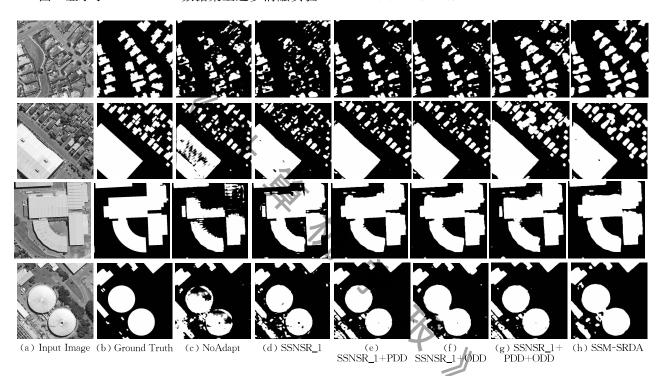


图 6 WHU 测试集分割预测结果图

### 4.3.3 Vaih-Pots

在 Vaih-Pots 数据集上的消融实验如表 3 所示. 比较表 3 中 SSNSR\_1+PDD(mIoU 为 36.6%)与 SSNSR\_1+ODD(mIoU 为 39.2%)的实验结果可知,在分割类别较多的数据集中,ODD 判别器在学习域不变性特征方面起到了更好的作用,其 mIoU 较没有域判别器的 SSNSR\_1 方法增加了 8.4%. SSNSR\_2较SSNSR\_1起到了更好的效果,mIoU增加了 3.1%,说明连接超分辨率模块和语义分割的FA-Loss 模块起到了增强细节结构特征的效果. SSM-SRDA 由 SSNSR\_2 与 PDD、ODD 两个判别器组成,取得了最好的实验结果,mIoU 为 45.2%,Macro-F1 为 61.1%. 图 7 给出了 SSM-SRDA 在 Vaih-Pots 数据

集上的消融实验结果,从视觉效果可以看出,分割结果变得越来越精细,特别是对建筑物有更好的提取效果.

表 3 Vaih-Pots 数据集上 SSM-SRDA 的逐步消融实验

方法	不透水 面/%	建筑 物/%	低矮植 被/%					
NoAdapt	46.7	47.8	17.7	28.0	24.4	10.0	29.1	43.3
SSNSR_1	38.4	39.6	21.1	27.3	37.6	20.5	30.8	46.4
SSNSR_2	49.0	43.4	40.4	20.1	35.9	14.7	33.9	49.3
SSNSR_1+ PDD	53.4	44.6	23. 2	31.3	44.5	22. 4	36.6	52.5
SSNSR_1+ ODD	56.9	40.7	44. 7	16.9	59.0	17. 1	39.2	54. 1
SSNSR_1+ PDD+ODD	58.5	49.9	44.6	30.2	59. 2	22.0	44.1	59.8
SSM-SRDA	58.8	55.3	41.5	39. 1	54.7	21.7	45. 2	61.1

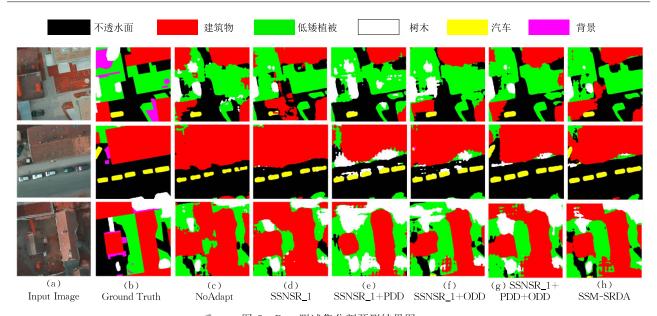


图 7 Pots 测试集分割预测结果图

### 4.3.4 青藏高原数据 TPD\_S-TPD\_T

青藏高原数据集 TPD\_S-TPD\_T 上的逐步消融实验结果如表 4 所示,各列的粗体值分别表示所有消融实验在四个语义类别上得到的最高 lot 值和 m IoU 值. 消除空间分辨率差异的 SSNSR\_1 较 No-Adapt 的 m IoU提高了 3.1%. 根据 SSNSR\_1+PDD、SSNSR\_1+ODD 和 SSNSR\_1 的实验结果对比可以发现,像素级域判别器 PDD 在高分辨率图像上减少外观差异对域适应任务的帮助更明显,较 SSNSR\_1 的 m IoU 提高了 8.2%,对四个类别的 IoU 分别提高了 3.0%、14.9%、10.1%和 5.0%. 在 SSNSR\_1 上集成 PDD和 ODD后获得了更好的 m IoU(53.2%). 增加 FA-Loss 模块的 SSNSR\_2 比 SSNSR\_1 的 m IoU 提升了 2.8%,其特征图细节结构信息的增强作用在不易区分的裸地和植被类别上效果更加突出,分别提高了 6.6%和 4.0%.

表 4 TPD\_S-TPD\_T 数据集上 SSM-SRDA 的逐步消融实验

方法	水体 /%	裸地 /%	植被/%	其他 /%	mIoU /%	Macro- F1/%
NoAdapt	15.9	23.1	41.5	70.7	37.8	42.6
SSNSR_1	17.8	33.7	45.5	66.5	40.9	52.4
SSNSR_2	19.1	40.3	49.5	65.9	43.7	55.7
SSNSR_1+ PDD	20.8	48.6	55.6	71.5	49.1	64.5
SSNSR_1+ ODD	22. 8	44.0	43.2	63.4	43.3	62.4
SSNSR_1+ PDD+ODD	22.6	52.8	56.7	80.5	53.2	66.8
SSM-SRDA	20.2	55. 2	63.4	82. 3	55.3	68. 1

本文在 Mass-Inria、Mass-WHU、Vaih-Pots 和TPD\_S-TPD\_T 四组数据集上进行了逐步消融实验,四组实验结果均验证了 SSM-SRDA 中超分辨率模块、PDD、ODD 和 FA-Loss 模块的有效性.

## 4.4 对比实验结果及分析

#### 4.4.1 Mass-Inria

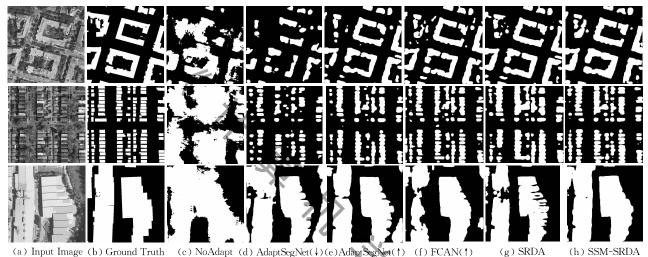
Mass-Inria 数据集上对比方法 AdaptSegnet<sup>[8]</sup>、FCAN<sup>[6]</sup>和 SRDA<sup>[12]</sup>与本文方法的比较结果如表 5 所示. 从结果可以看出, SSM-SRDA 的 *mIoU* 和 F1-score 的值(71.6%和 69.2%)均高于 AdaptSegnet (69.5%和 65.8%)、FCAN(69.4%和 65.2%)和 SRDA(70.9%和 67.8%). SSM-SRDA 的训练时间小于两阶段的 FCAN 和使用 ESPCN 提高源域分辨率的 AdaptSegnet,表明 SSM-SRDA 中超分辨率模块和分割模块共享同一个特征提取网络的设置大大

表 5 Mass-Inria 数据集上对比方法与本文方法的结果比较

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	111 141 3X J/	1 <del>*</del> ^	1 16 71 12	1742	/J /A H J = F	11/10/12
方法	建筑 物/%	背景 /%	mIoU / %	F1- score/%	训练 时间/h	测试 时间/s
NoAdapt	32.7	83.2	58.0	49.1	18	0.15
AdapSegNet (↓Inria)	34.5	89.4	61.9	51.2	30	0.17
AdapSegNet (↑Mass)	47.8	89.8	68.8	64.9	31	0.17
AdapSegNet (ESPCN)	49.3	89.7	69.5	65.8	41	0.17
FCAN (↓Inria)	39.8	89.4	64.6	56.8	46	0.15
FCAN (∱Mass)	48. 1	90. 6	69.4	65. 2	48	0.15
SRDA	51.2	90.4	70.9	67.8	49	0.24
SSM-SRDA	52. 6	90.5	71.6	69. 2	39	0.18

减少了模型计算量. SSM-SRDA 的测试时间小于用渐 进上采样提高源域图像分辨率的 SRDA,表明 pixshuffle 方法具有计算量较少的优势, 为了探究源域 和目标域的分辨率差异对域适应结果的影响,实验 对 AdaptSegnet 和 FCAN 设置两种输入情况:一是 下采样目标域图像到源域分辨率大小;二是上采样 源域图像到目标域分辨率大小; ↑ 和 ↓ 分别表示上 采样和下采样操作. 从结果可以发现上采样源域图 像的输入设置获得了更好的结果,这说明消除分辨 率差异获得的增益大于上采样插值引入的误差,相 比之下,本文提出的方法 SSM-SRDA 不需要考虑输 入的设置,对比同样采用两个判别器,使用双线性插值 方法恢复源域分辨率的 SRDA(mIoU 70.9%),可以 发现使用亚像素卷积方法的 SSNSR 1+PDD+ODD (mIoU 71.2%)获得了更好的结果.

图 8 显示了本文和对比方法的部分分割结果. 第一列中的图像选自 Inria 测试数据集,第二列为其 真实标签,其余各列为相应对比实验的预测结果. 可以观察到域适应方法的测试结果均优于没有消 除分辨率差异的无域适应的结果,而本文提出的 SSM-SRDA 方法的分割结果在结构与细节上均优 于 AdaptSegnet、FCAN、SRDA.



(g) SRDA

(h) SSM-SRDA

图 8 Inria 测试集分割预测结果

### 4.4.2 Mass-WHU

Mass-WHU 数据集上对比方法 AdaptSegnet [8]、 FCAN<sup>[6]</sup>和 SRDA<sup>[12]</sup>与本文方法的比较结果如表 6 所示,粗体值表示每列最高值.从结果可以看出, 本文方法 SSM-SRDA 在 Mass-WHU 数据集上取 得了最好的结果: mIoU 为 80.3%. SSM-SRDA 的 mIoU 高于 AdaptSegnet(76.0%)、FCAN(77.2%)

表 6 Mass-WHU 数据集上对比方法与本文方法的结果比较

方法	建筑 物/%	背景 /%	$^{mIoU}_{/\%}$	F1-score/ $%$	训练 时间/h	测试 时间/s
NoAdapt	46.8	91.3	69.1	63.7	18	0.13
AdapSegNet (↓Inria)	50.8	92.8	72.4	66.9	21	0.15
AdapSegNet (↑Mass)	58.0	94.0	76.0	73. 3	22	0.18
FCAN (↓Inria)	53.9	93.9	73.9	70. 1	27	0.14
FCAN (↑Mass)	59.3	95.1	77.2	74.7	28	0.14
SRDA	61.3	95.9	78.6	77.0	34	0.28
SSM-SRDA	66.0	96.7	80.3	80. 2	27	0.21

和 SRDA(78.6%). 由于Mass-WHU数据集空间分 辨率差异较大,因此 SSM-SRDA 的训练时长与两 阶段的 FCAN 相近,但仍小于用插值提高源域图像 分辨率的 SRDA. 结合表 5 和表 6 可以发现,将 SSM-SRDA 从同一个有标签的源域数据集 Mass 迁移应 用到不同源的目标域数据集(Inria和WHU)上均得 到了较好的效果,较最新的方法 SRDA, mIoU 分别 提高了0.7%和1.7%,证明本文提出的 SSM-SRDA 在实际应用中有泛化能力.

#### 4.4.3 Vaih-Pots

表 7 中对 AdapSegNet 进行的两种不同输入情况 的实验结果中,下采样目标域较上采样源域的 mIoU 高 5.6%, 而在 FCAN 的两组实验中, 上采样源域较 下采样目标域的 m IoU 高 1.0%,说明在源域和目标 域分辨率差异较小的情况下,很难确定采取哪种输入 设置. 然而,本文的 SSM-SRDA 方法不需要考虑输 入设置,与之前最好的方法 SRDA 对比,SSM-SRDA 获得了更好的性能(mIoU) 45.2%).

表 7 Vaih-Pots 数据集上对比方法与本文的比较结果

方法	不透 水面 /%	建筑物/%	低矮 植被 /%	树木	汽车 /%		mIoU /%	Macro- F1/%
NoAdapt	46.7	47.8	17.7	28.0	24.4	10.0	29.1	43.3
AdapSegNet (↓TPD_T)	48.4	41.0	41.7	20.7	36.6	22.3	35. 1	51.1
AdapSegNet (↑TPD_S)	54.8	44.5	42.2	34.0	46.9	21.5	40.7	57.0
FCAN (↓TPD_T)	56.8	40.2	44. 5	31.5	57.8	21. 1	42.0	57.9
FCAN (↑TPD_S)	54.7	44.5	42.2	34.0	46.9	21.5	41.0	57.0
SRDA	58.0	52.6	27.0	34.6	62. 2	23. 3	43.0	58.5
SSM-SRDA	58.8	55.3	41.5	39. 1	54.7	21.7	45. 2	61.1

#### 4.4.4 青藏高原数据 TPD\_S-TPD\_T

青藏高原TPD\_S-TPD\_T数据集上最新方法与本文方法的比较结果如表8所示,包括AdaptSegnet、FCAN和SRDA,粗体值表示每对最优结果.从表8可以看出,本文提出的方法SSM-SRDA在TPD\_S-

表 8 TPD\_S-TPD\_T 数据集上最新方法与本文的比较结果

方法	水体 /%	裸地 /%	植被/%	其他 /%	mIoU /%	Macro- F1/%
NoAdapt	15.9	23. 1	41.5	70.7	37.8	42. 6
AdapSegNet (↓TPD_T)	18.3	41.6	47.8	73. 2	45.2	57. 4
AdapSegNet (↑TPD_S)	21.6	52.9	50.3	71.4	49.1	63. 9
FCAN (↓TPD_T)	21.1	42.9	43.9	62.7	42.6	55.8
FCAN (†TPD_S)	22.7	44.8	50.9	67.1	46.4	64.4
SRDA	20.1	58. 9	57.4	71.6	52.0	65.7
SSM-SRDA	20.2	55.2	63. 4	82. 3	55.3	68. 1

TPD\_T 数据集上取得了最高的 mIoU(55.3%). SSM-SRDA 中利用 pixshuffle 方法提高分辨率的设置起到了提升性能的作用,其 mIoU 高于上采样目标域的 AdaptSegnet(49.1%)、FCAN(46.4%)和 SRDA (52.0%).

图 9 列出了三组在 TPD T 测试集上不同方法 的分割结果图,从视觉效果上可以发现 SSM-SRDA 分割结果最为精准. 图 10 给出了 SSM-SRDA 在青 藏高原目标域 TPD\_T 测试集上所有图像分割结果 的混淆矩阵,其中行代表每一类别的真实标签,列代 表预测得到的类别结果. 本文的方法 SSM-SRDA 对水体类分割精度较低,图 10 中的混淆矩阵表明, 大量水体类被分到了裸地类和植被类. 如图 11 所 示,观察源域和目标域数据集图像可以发现,青藏高 原数据集中有许多有云覆盖的区域,云附近有黑色 阴影区域,在标签中将黑色阴影划分为真实的裸地 或植被类别,而云层阴影和小的湖泊具有高度的相 似性,很难区分.目标域图像含大量小面积的湖泊 类,如图 9 展示的目标域 TPD\_T 测试图像第三行 所示,SSM SRDA 将小的湖泊类错误划分为了裸地 或植被类.

本文提出的 SSM-SRDA 在 Mass-Inria、Mass-WHU、Vaih-Pots 和 TPD\_S-TPD\_T 四组数据集上得到的 *mIoU* 或 F1-score 均优于具有可比性的 AdaptSegnet、FCAN 和 SRDA 方法,表明了在遥感图像域适应语义分割中对齐源域和目标域空间分辨率后缩小特征差异的方法的优势。

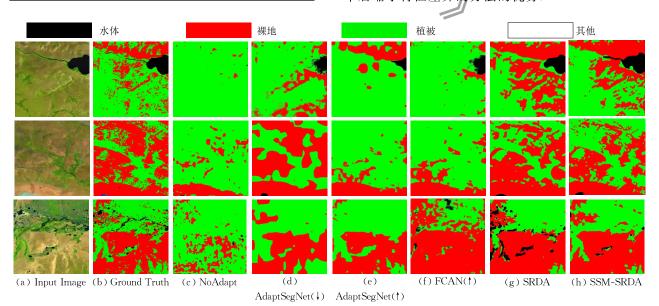


图 9 青藏高原目标域 TPD\_T 测试集分割预测结果图



图 10 SSM-SRDA 在青藏高原目标域 TPD\_T 测试集上 像素分类结果的混淆矩阵

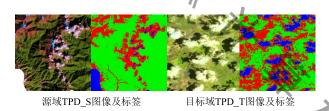


图 11 青藏高原源域 TPD\_S 及目标域 TPD\_T 多云 区域图像及标签

## 5 结 论

本文提出的是一个新的端到端无监督域适应方 法——结合超分辨率和域适应的语义分割模型 SSM-SRDA,解决遥感图像语义分割域适应领域中 源域和目标域图像存在空间分辨率差异的问题,并 对齐了两个域在视觉外观和 SoftMax 输出上的特 征差异. 首先建立了一个用于超分辨率和语义分割 的多任务模型 SSNSR,它不仅缩小了源域和目标域 之间的空间分辨率差异,生成了带有目标域风格的 高分辨率源域图像,还通过特征相似性 FA-Loss 模 块增加语义分割模块深层特征的细节结构信息. 其 次通过像素级域判别器 PDD 和输出空间级域判别 器 ODD 的对抗学习来引导 SSNSR 模型学习域不变 特征,有效提高了 SSM-SRDA 在目标域上迁移应用 的性能. 为了证明 SSM-SRDA 的有效性,在构建的 四对不同空间分辨率的数据集(Mass-Inria、Mass-WHU、Vaih-Pots和TPD\_S-TPD\_T)上进行实验, 均取得了最好的结果,表明当源域和目标域遥感图 像具有空间分辨率和特征差异时,SSM-SRDA效果 更好且没有额外的计算成本.未来研究将 SSM-SR-DA 扩展到更多不同需求的实际应用中,如高空间 分辨率源域和低空间分辨率目标域的应用场景、源域和目标域有较大光谱分布差异的应用场景等.

### 参考文献

- [1] Zheng C, Zhang Y, Wang L. Semantic segmentation of remote sensing imagery using an object-based Markov random field model with auxiliary label fields. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 3015-3028
- [2] Pan B, Shi Z, Xu X, et al. CoinNet: Copy initialization network for multispectral imagery semantic segmentation. IEEE Geoscience and Remote Sensing Letters, 2019, 16(5): 816-820
- [3] Mou L, Hua Y, Zhu X X. Relation matters: Relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(11): 7557-7569
- [4] Alam M, Wang J F, Guangpei C, et al. Convolutional neural network for the semantic segmentation of remote sensing images. Mobile Networks and Applications, 2021, 26(1): 200-215
- [5] Wu Z, Wang X, Gonzalez J E, et al. ACE: Adapting to changing environments for semantic segmentation//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Republic of Korea, 2019: 2121-2130
- [6] Zhang Y. Qiu Z. Yao T, et al. Fully convolutional adaptation networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018; 6810-6818
- [7] Li Y, Yuan L, Vasconcelos N. Bidirectional learning for domain adaptation of semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019; 6936-6945
- [8] Tsai Y H, Hung W C, Schulter S, et al. Learning to adapt structured output space for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7472-7481
- [9] Tasar O, Happy S L, Tarabalka Y, Alliez P. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58 (10): 7178-7193
- Zhang Z, Doi K, Iwasaki A, Xu G. Unsupervised domain adaptation of high-resolution aerial images via correlation alignment and self-training. IEEE Geoscience and Remote Sensing Letters, 2020, 18(4): 746-750
- [11] Wu J, Tang Z, Xu C, et al. Super-resolution domain adaptation

- networks for semantic segmentation via pixel and output level aligning. Frontiers in Earth Science, 2022, 10 (2022): 974325
- [12] Zhang L, Lan M, Zhang J, et al. Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60(2021): 1-13
- [13] Fan Cang-Ning, Liu Peng, Xiao Ting, et al. A review of deep domain adaptation: General situation and complex situation. Acta Automatica Sinica, 2021, 47(3): 515-548(in Chinese)
  (范苍宁, 刘鹏, 肖婷等. 深度域适应综述: 一般情况与复杂情况. 自动化学报, 2021, 47(3): 515-548)
- [14] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Communications of the ACM, 2020, 63(11): 139-144
- [15] Wang J, Shao Z, Liu T, et al. Unsupervised remote sensing super-resolution via migration image prior. arXiv preprint arXiv: 2105.03579, 2021
- [16] Yang X, Wang S, Han J, et al. RSAMSR: A deep neural network based on residual self-encoding and attention mechanism for image super-resolution. Optik, 2021, 245 (2021): 167736
- [17] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017; 4700-4708
- [18] Lei S, Shi Z, Wu X, et al. Simultaneous super-resolution and segmentation for remote sensing images//Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan, 2019: 3121-3124
- [19] Wang L, Li D, Zhu Y, et al. Dual super-resolution learning for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 3774-3783
- [20] Zhang Q, Yang G, Zhang G. Collaborative network for super-resolution and semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60(2021): 1-12
- [21] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440
- [22] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84-90
- [23] Liu Z, Li X, Luo P, et al. Semantic image segmentation via deep parsing network//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1377-1385
- [24] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778

- [25] Yuan Y, Huang L, Guo J, et al. OCNet: Object context for semantic segmentation. International Journal of Computer Vision, 2021, 129(8): 2375-2398
- [26] Wang W, Wang S, Li Y, et al. Adaptive multi-scale dual attention network for semantic segmentation. Neurocomputing, 2021, 460(2021): 39-49
- [27] Pathak D, Krahenbuhl P, Darrell T. Constrained convolutional neural networks for weakly supervised segmentation// Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015; 1796-1804
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. Munich, Germany, 2015: 234-241
- [29] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307
- [30] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 391-407
- [31] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 1874-1883
- [32] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for realtime style transfer and super-resolution//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 694-711
- [33] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017; 4681-4690
- [34] Yuan Y, Liu S, Zhang J, et al. Unsupervised image superresolution using cycle-in-cycle generative adversarial networks //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA, 2018, 701-710
- [35] Jiang K, Wang Z, Yi P, et al. Edge-enhanced GAN for remote sensing image superresolution. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5799-5812
- [36] Wei Y, Gu S, Li Y, et al. Unsupervised real-world image super resolution via domain-distance aware training//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, 2021; 13385-13394
- [37] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030
- [38] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks//Proceedings of the International Conference on Machine Learning. Lille, France, 2015: 97-105

- [39] Long M, Zhu H, Wang J, et al. Unsupervised domain adaptation with residual transfer networks. arXiv preprint arXiv:1602.04433, 2016
- [40] Hoffman J, Wang D, Yu F, et al. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv: 1612.02649, 2016
- [41] Hoffman J, Tzeng E, Park T, et al. CyCADA: Cycleconsistent adversarial domain adaptation//Proceedings of the International Conference on Machine Learning. Stockholm, Sweden, 2018; 1989-1998
- [42] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2223-2232
- [43] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks

- for object detection//Proceedings of the IEEE Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 2117-2125
- [44] Wang L, Wang Y, Liang Z, et al. Learning parallax attention for stereo image super-resolution//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 12250-12259
- [45] Kotovenko D, Sanakoyeu A, Ma P, et al. A content transformation block for image style transfer//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 10032-10041
- [46] Ji S, Wei S, Lu M. Fully convolutional networks for multi-source building extraction from an open aerial and satellite imagery data set. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(1): 574-586



**LIANG Min**, M. S. candidate. Her current research interests include machine learning and image processing.

WANG Xi-Li, Ph. D., professor. Her current research interests include pattern recognition, machine learning and image processing.

#### Background

Unsupervised semantic segmentation methods based on deep learning are often ineffective in the field of remote sensing images, because the sensors used to acquire remote sensing images usually have significant differences, resulting in significant differences in the feature distribution of remote sensing images in source domain and target domain, and different source data have different spatial resolutions. Therefore, unsupervised domain adaptation of remote sensing images should not only narrow the feature gap between source domain and target domain, but also solve the problem of different spatial resolution. In the existing research, few domain adaptation semantic segmentation methods explicitly consider the spatial resolution difference. Most of them deal with the resolution problem of remote sensing images through simple interpolation. When the resolution gap

between source domain and target domain is not serious, some researchers ignore the resolution problem.

This paper explicitly considers the difference in spatial resolution and feature distribution between source domain and target domain of remote sensing images, and proposes a semantic segmentation model that combines super resolution and domain adaptation—SSM-SRDA, which can implement super resolution and Semantic segmentation simultaneously. In the design of SSM-SRDA, we use domain adaptation methods to improve the performance of the model from low resolution remote sensing data migration and application to high resolution remote sensing data.

This study was supported by the Second Tibetan Plateau Scientific Expedition and Research and Research Program (No. 2019QZKK0405).