

分类号 _____

密 级 _____

U D C _____

编 号 10486

武汉大学
硕 士 学 位 论 文

基于生成对抗网络的跨域遥感
图像语义分割

研 究 生 姓 名 : 史特

学 号 : 2020202130028

指导教师姓名、职称 : 李彦胜 副教授

专 业 名 称 : 模式识别与智能系统

研 究 方 向 : 计算机视觉

二〇二三年五月

Master Dissertation of Wuhan University

Cross-Domain Remote Sensing Image
Semantic Segmentation Based on Generative
Adversarial Networks

By

Te Shi

Supervised by

Associate Prof. Yansheng Li

School of Remote Sensing and Information Engineering

Wuhan University

May, 2023

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已发表或撰写的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：史物

2023 年 5 月 24 日

学位论文使用授权书

本论文作者完全了解学校关于保存、使用学位论文的管理办法及规定，即学校有权保留并向国家有关部门或机构递交论文的复印件和电子版，允许论文被查阅和借阅，接受社会监督。本人授权武汉大学可以将本学位论文的部分或全部内容编入学校有关数据库进行信息服务，也可以采用影印、缩印或扫描等复制手段保存或汇编本学位论文。

本论文提交当年/一年/两年/三年以后，同意发布。

若不选填则视为一年以后同意发布。

注：保密学位论文，在解密后适用于本授权书。

作者签名: 史特

导师签名: 李彦阳

2023 年 5 月 24 日

武汉大学研究生学位论文作者信息

论文题目	基于生成对抗网络的跨域遥感图像语义分割				
姓 名	史特	学号	2020202130028	答辩日期	2023 年 5 月 24 日
论文级别	博士 <input type="checkbox"/> 硕士 <input checked="" type="checkbox"/>				
院/系/所	遥感信息工程学院	专业	模式识别与智能系统		
联系电话	17643230599	E_mail	te.shi@whu.edu.cn		
通信地址(邮编): 湖北省武汉市珞喻路 129 号武汉大学遥感信息工程学院					
备注:					

本论文创新点

本文研究的关注点主要是跨域遥感图像语义分割问题，以生成对抗网络为主线，提出了一系列解决方案。本文主要创新点有以下几方面：

(1) 提出基于对偶生成对抗网络的域适应遥感图像语义分割方法。首先，该方法利用对偶生成对抗网络将源域图像迁移至目标域图像的风格，用以减少源域与目标域的像素级分布差异。随后，提出了带有多重弱监督约束的目标优化损失函数，进一步减小域偏移问题的影响。具体包括弱监督迁移不变约束、弱监督伪标签约束以及弱监督旋转一致性约束。为了充分发挥各项约束的优势，该方法采取了一种动态优化策略，在训练过程中动态调整各项约束的权值，避免深度语义分割网络发生退化。本方法在跨地理位置和跨成像模式两个遥感图像跨域语义分割任务上相比同期代表性算法取得了更优的结果。

(2) 提出基于语义保持生成对抗网络的域适应遥感图像语义分割方法。现有的基于生成对抗网络的图像迁移模型未充分考虑语义信息，致使迁移后的图像往往会产生偏差，导致迁移效果有限。针对上述问题，该方法通过在生成对抗网络模型中引入表征不变约束和语义保持约束，以此实现源域图像到目标域风格的无偏迁移。此外，考虑到源域和目标域数据的类别分布存在差异，提出了类别分布对齐语义分割模块，实现类别分布对齐。本方法在跨地理位置、跨成像模式以及跨景观布局三个典型的跨域遥感图像语义分割任务都取得了最优的性能。

摘要

随着全球高分对地观测系统的建设与完善，遥感数据的获取能力得到了极大提升，其数据量呈现指数级增长，这为遥感图像智能化解译提供了数据支撑。作为遥感影像智能解译的基础任务之一，遥感图像语义分割旨在对图像中的每一个像素点进行分类，为其赋予一个预定义的语义类别标签，进而刻画出目标物体的轮廓信息，其结果被广泛应用于地表专题制图、智能交通、环境监测、城市规划等方面。随着人工智能技术的蓬勃发展，深度学习方法已经逐渐成为遥感图像语义分割的主流方法，并实现了对传统图像分割方法的全面超越。然而，深度语义分割网络的优越表现很大程度上依赖于带有标签的海量目标数据，使得这些方法所需的时间和金钱成本十分高昂。为减轻标注数据工作的压力，一种常用的方法是使用已有的标注数据集（源域）进行训练，在未标注的数据集（目标域）上进行测试与推理。但不同遥感图像数据集之间往往有着较大的差异，存在严重的“同物异谱”和“同谱异物”现象，而传统的深度学习模型要求源域与目标域数据分布相似，这势必将导致在源域图像场景下训练的语义分割模型难以泛化到目标域图像场景。针对这一难题，本文对遥感图像跨域语义分割问题展开了研究，提出了一系列的创新性解决方案，并通过实验验证了方案的有效性。论文的主要工作和贡献如下：

(1) 提出基于对偶生成对抗网络的域适应遥感图像语义分割方法。本方法首先训练对偶生成对抗网络，将源域图像映射至目标域的风格，缩小源域图像与目标域图像的像素级分布差异。为了更好的将在标签丰富的源域数据集训练的网络迁移应用至无标签的目标域数据集，创新性地引入了联合迁移后图像语义分割损失、目标域图像伪标签损失以及目标域图像旋转不变一致性损失的多重弱监督约束。同时，为了充分发挥多重约束的作用、避免网络退化，该方法采取了一种动态优化策略，在训练过程中动态地调整目标约束的各项权重。

(2) 提出基于语义保持生成对抗网络的域适应遥感图像语义分割方法。主流方法通常直接利用已有的生成对抗网络模型，进行源域图像到目标域图像的迁移工作。然而，由于现有的基于生成对抗网络的图像迁移模型未充分考虑语义信息，致使迁移后的图像往往会产生偏差，导致迁移效果有限。为解决该问题，本方法通过在生成对抗网络模型中引入表征不变约束和语义保持约束，来实现从源

域到目标域的语义保持图像迁移。此外，考虑到源域和目标域数据的类别分布存在差异，提出了类别分布对齐语义分割模块。该模块包括两个层面：在模型输入层面，引入类别混合操作—ClassMix 操作，通过随机采样的方式从迁移后的图像中随机选择一定的类别，将相应的图像块粘贴到目标域图像上构成混合图像，接着将迁移后的图像与混合图像一起输入语义分割网络进行训练；在模型输出层面，提出了边界增强损失，用来细化语义分割网络在目标边界的性能。

跨域遥感图像语义分割是一个具有挑战性的新兴课题，在遥感大数据时代具有很强的实用价值。本文以生成对抗网络图像迁移为主线，开展了一系列相关研究。首先，利用已有的对偶生成对抗网络模型进行源域到目标域的图像迁移，缩小了源域和目标域图像像素级分布的差异。其次，充分挖掘目标域数据的弱监督信息，如伪标签损失、旋转不变性损失，进一步提升语义分割模型在目标域的性能。随后，针对已有的生成对抗网络在图像迁移过程中存在偏差的问题，提出了基于语义保持的生成对抗网络方法。最后，考虑到源域与目标域数据类别分布差异的问题，进一步地提出了类别分布对齐语义分割模块，在输入和输出层面进行了进一步对齐。大量的实验结果表明，本文提出的方法在经典的跨域遥感图像语义分割任务中都取得了最优的结果，证明了本文所提出的方法的有效性和通用性，为跨域遥感图像语义分割任务提供了新的范式。

关键词：领域自适应语义分割；生成对抗网络；多重弱监约束；语义保持；类别分布对齐

ABSTRACT

With the construction and improvement of global high-resolution earth observation systems, obtaining remote sensing images has become much easier. Thus, the amount of data has grown exponentially. This provides data support for the intelligent interpretation of remote sensing images. As one of the fundamental tasks of intelligent interpretation of remote sensing images, semantic segmentation aims to classify each pixel in one image and assign a pre-defined semantic category label to describe the contour information of the target object. The semantic segmentation results are widely used in thematic mapping, intelligent transportation, environmental monitoring, urban planning, and other fields. With the booming development of artificial intelligence technology, deep learning methods have gradually become the mainstream method of remote sensing image semantic segmentation, and have achieved comprehensive surpassing of traditional image segmentation methods. However, the superior performance of deep semantic segmentation networks highly depends on massive labeled target data, which will cost a lot of time and money. A common method to reduce the dependence on annotated data is to use existing labeled datasets (source domain) for training and testing or inference on the unannotated dataset (target domain). However, due to significant differences between different remote sensing image datasets, there are serious phenomena of the same object with different spectra and different objects with the same spectra. Traditional deep learning models require similar data distribution between source and target domains, which will inevitably lead to difficulty in generalizing the semantic segmentation model trained in the source domain image scene to the target domain image scene. In response to this problem, this paper investigates the problem of cross-domain semantic segmentation of remote sensing images, proposes a series of innovative solutions, and verifies the effectiveness of the proposed solutions through experiments. The main work and contributions of the paper are as follows:

(1) A domain adaptative semantic segmentation method based dual generative adversarial network is proposed. First, a generative adversarial network (GAN) is trained to map source domain images to the style of the target domain, reducing the pixel-level distribution differences between source and target domain images. In order to better transfer the network trained on a source domain dataset with rich labels to an

unlabeled target domain dataset, multiple weakly-supervised constraints of joint translated image semantic segmentation loss, target domain image pseudo-label loss, and target domain image rotation consistency loss are innovatively introduced. To fully utilize the role of multiple constraints and avoid network degradation, a dynamic optimization strategy is proposed to dynamically adjust the weight of the constraint terms in the objective function during training.

(2) A domain adaptive semantic segmentation method based on semantic-preserved generative adversarial network is proposed. The mainstream methods directly use existing generative adversarial network models for source-to-target domain image translation. However, the existing generative adversarial network-based image transfer models do not fully consider semantic information, which often leads to deviation in the translated images and limited transfer performance. To address this issue, representation-invariant constraints and semantic-preserved constraints are introduced into the generative adversarial network model to achieve semantic-preserved source-to-target image translation. In addition, considering the difference in category distribution between the source and target domain datasets, a class distribution alignment semantic segmentation module is proposed. The class distribution alignment semantic segmentation module includes two levels: At the input level of the model, the class mixing operation (ClassMix) is introduced to randomly select certain classes from the translated image by random sampling strategy, and paste the corresponding image patches onto the target domain image to form a mixed image. The translated images with the mixed images are fed into the semantic segmentation network for training. At the model output level, boundary enhancement constraint is proposed to refine the performance of the target object boundary.

Cross-domain remote sensing image segmentation is a challenging emerging topic with strong practical value in the era of remote sensing big data. In this paper, a series of related researches are carried out with generative adversarial network image translation as the main line. Firstly, existing GAN model is used for image translation from the source domain to the target domain, reducing the difference in pixel-level distribution between the two domains. Second, unsupervised information of the target domain data, such as pseudo-label loss and rotation invariant loss, is fully exploited to further improve the performance of semantic segmentation models in the target domain. Subsequently, in response to problem of the bias in the translated images by existing

GAN, a semantic-preserved GAN is proposed. Finally, the issue of differences in class distribution between the source and target domain data is also addressed by introducing a class distribution alignment semantic segmentation module, which aligns the distributions at the input and output levels. Numerous experimental results demonstrate that the proposed methods achieve optimal results in classic cross-domain remote sensing image semantic segmentation tasks, proving the effectiveness and generality of the proposed method. This provides a new paradigm for cross-domain remote sensing image semantic segmentation tasks.

Key words: domain adaptative semantic segmentation; generative adversarial networks; multiple weakly-supervised constraints; semantic preservation; class distribution alignment

目 录

摘 要.....	I
ABSTRACT.....	III
图索引.....	VIII
表索引.....	X
第 1 章 绪论	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.2.1 传统全监督语义分割技术.....	3
1.2.2 领域自适应语义分割技术.....	7
1.3 主要工作及贡献.....	9
1.4 论文组织结构.....	10
第 2 章 相关研究理论基础	12
2.1 领域自适应方法.....	12
2.1.1 基于模型的领域自适应.....	12
2.1.2 基于样本的领域自适应.....	13
2.1.3 基于特征的领域自适应.....	13
2.2 生成对抗网络方法.....	15
2.2.1 生成对抗网络.....	15
2.2.2 循环一致生成对抗网络.....	17
2.3 本章小结.....	18
第 3 章 基于对偶生成对抗网络的域适应遥感图像语义分割	19
3.1 概述.....	19
3.2 研究方法.....	20
3.2.1 无监督风格迁移.....	21
3.2.2 多重弱监督约束的目标函数.....	22
3.2.3 优化策略.....	24
3.3 实验设计.....	25
3.3.1 实验数据集.....	25

3.3.2 实验设置与评价指标.....	26
3.4 实验结果与分析.....	27
3.4.1 方法对比.....	27
3.4.2 消融实验.....	32
3.4.3 超参数分析.....	32
3.5 本章小结.....	34
第 4 章 基于语义保持生成对抗网络的域适应遥感图像语义分割	35
4.1 概述.....	35
4.2 研究方法.....	38
4.2.1 语义保持的生成对抗网络.....	38
4.2.2 类别分布对齐语义分割模块.....	40
4.3 实验设计	42
4.3.1 任务设定.....	42
4.3.2 网络结构.....	43
4.3.3 评价指标.....	43
4.4 实验结果与分析.....	44
4.4.1 SPGAN 模型的优越性验证	44
4.4.2 方法对比.....	47
4.4.3 模型分析.....	52
4.5 本章小结.....	55
第 5 章 总结与展望	57
5.1 研究总结.....	57
5.2 未来展望.....	58
参考文献	59
攻读硕士学位期间发表的科研成果	66
攻读硕士学位期间参与的课题情况	67
致 谢.....	68

图索引

图 1-1 遥感图像语义分割应用场景	1
图 1-2 不同域间遥感图像示例图	2
图 1-3 全监督语义分割与跨域领域自适应语义分割区别	3
图 1-4 全卷积神经网络	4
图 1-5 基于编码器-解码器的语义分割网络 SegNet 模型	4
图 1-6 不同膨胀率的空洞卷积	5
图 1-7 ASPP 模块	5
图 1-8 Deeplab v3 plus 模型结构	6
图 1-9 SegFormer 模型结构	7
图 1-10 CyCADA 算法示意图	7
图 1-11 FCN wild 算法示意图	8
图 1-12 CBST 算法示意图	9
图 1-13 论文组织结构	11
图 2-1 传统机器学习与领域自适应学习异同	12
图 2-2 生成对抗网络结构图	15
图 2-3 GAN 训练过程示意图	17
图 2-4 CycleGAN 图像迁移示例	17
图 2-5 CycleGAN 结构示意图	18
图 3-1 多重弱监督约束下的遥感图像跨域语义分割算法	21
图 3-2 利用 DualGAN 进行图像迁移示意图	22
图 3-3 跨域遥感图像语义分割任务设定	26
图 3-4 经 DualGAN 迁移后的图像	28
图 3-5 跨地理位置 Potsdam IR-R-G → Vaihingen IR-R-G 语义分割结果可视化	29
图 3-6 跨成像模式 Potsdam R-G-B → Vaihingen IR-R-G 语义分割结果可视化	31
图 3-7 Potsdam IR-R-G → Vaihingen IR-R-G 不同伪标签阈值 τ_{pse} 下语义分割结果可视化	33
图 3-8 Potsdam IR-R-G → Vaihingen IR-R-G 不同伪标签阈值 τ_{pse} 下语义分割结	

果可视化.....	33
图 4-1 图像迁移中的偏差现象	35
图 4-2 三种典型的跨域遥感图像语义分割任务	37
图 4-3 提出的 SPGAN-DA 框架	38
图 4-4 ClassMix 操作示意图	41
图 4-5 Potsdam IR-R-G 数据集图像迁移结果	45
图 4-6 Potsdam R-G-B 数据集图像迁移结果.....	46
图 4-7 Urban R-G-B 与 Rural R-G-B 数据集图像迁移结果.....	47
图 4-8 Potsdam IR-R-G → Vaihingen IR-R-G 语义分割结果可视化	48
图 4-9 Potsdam R-G-B → Vaihingen IR-R-G 语义分割结果可视化.....	49
图 4-10 Rural R-G-B → Urban R-G-B 语义分割结果可视化	52
图 4-11 t-SNE 特征分布可视化	55

表索引

表 3-1 跨地理位置 Potsdam IR-R-G → Vaihingen IR-R-G 跨域语义分割结果	28
表 3-2 跨成像模式 Potsdam R-G-B → Vaihingen IR-R-G 跨域语义分割结果.....	30
表 3-3 Potsdam IR-R-G → Vaihingen IR-R-G 消融实验	32
表 3-4 Potsdam R-G-B → Vaihingen IR-R-G 消融实验.....	32
表 3-5 Potsdam IR-R-G → Vaihingen IR-R-G 伪标签阈值 τ_{pse} 参数分析.....	33
表 3-6 Potsdam R-G-B → Vaihingen IR-R-G 伪标签阈值 τ_{pse} 参数分析	33
表 4-1 Potadam IR-R-G/R-G-B 迁移后的图像 FID 值（值越小图像质量越高） .	45
表 4-2 Urban R-G-B 与 Rural R-G-B 迁移后的图像 FID 值(值越小图像质量越高)	46
表 4-3 跨地理位置下不同域适应方法的 mIoU 对比结果	48
表 4-4 跨成像模式下不同域适应方法的 mIoU 对比结果	50
表 4-5 跨景观布局下不同域适应方法的 mIoU 对比结果	51
表 4-6 SPGAN-DA 算法消融实验	53
表 4-7 Potsdam IR-R-G/R-G-B → Vaihingen IR-R-G 阈值参数 λb 分析	54
表 4-8 Urban R-G-B ↔ Rural R-G-B 阈值参数 λb 分析	54

第1章 绪论

1.1 研究背景与意义

遥感图像是一种重要的信息载体，是人类认知世界的重要信息来源。随着遥感科学技术与相关产业的飞速发展，大量搭载功能各异传感器的卫星平台被发射升空，使得图像类型多种多样、图像分辨率提高、图像更新速度变快。遥感图像的数据量呈现井喷式增长，我们俨然迈入遥感数据时代，这为遥感能力化解译提供了强有力的数据支撑^[1-4]。利用遥感图像解译的结果，人们能够快速深入地了解全球各地的复杂情况。这对于城市建设规划、自然要素资源监测、灾害防控管理、生态环境保护、经济发展预测等领域都有着重要且深远的意义^[5-8]。在这些领域中，遥感图像解译技术为研究和实践提供了极其有价值的信息来源和工具。进一步探索和应用遥感图像解译技术，将有助于促进相关领域的发展和创新。

作为遥感影像智能解译的基础任务之一，遥感图像语义分割旨在对图像中的每一个像素点进行分类，为其赋予一个预定义的语义类别标签，进而刻画出目标物体的轮廓信息。随着人工智能技术的蓬勃发展，深度学习方法已经逐渐成为遥感图像语义分割的主流方法，并实现了对传统图像分割方法的全面超越。如图 1-1 所示，基于深度学习方法的遥感图像语义分割可以广泛用于开展地物分类、云区检测、建筑物检测、道路提取等相关工作^[9-14]。

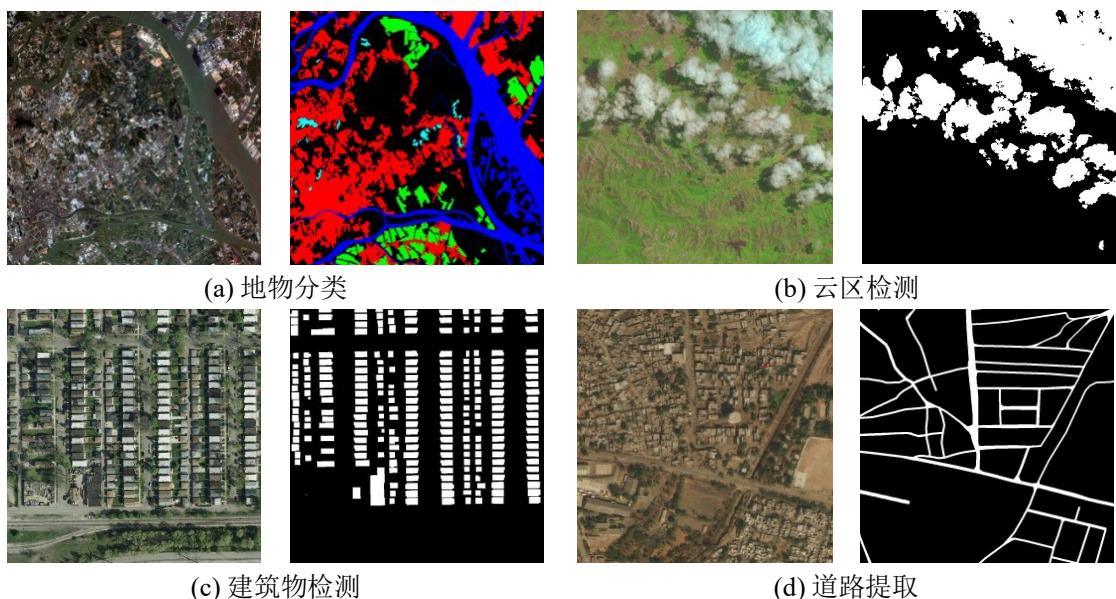


图 1-1 遥感图像语义分割应用场景

总体而言，基于深度学习的语义分割方法已经取得了十分不错的表现，但对于遥感图像语义分割仍旧面临如下两方面的问题：

1. 需要大量精准标注数据。深度语义分割模型的优越表现依赖于大量的人工精准标注数据作为训练集。语义分割是像素级别的分类任务，故需要人工精准标注到每个像素。然而数据的标注成本十分高昂，需要消耗大量的时间、人力和物力。如 Cityscapes^[15]数据集中所描述的，标注一张尺寸为 1024×2048 的图像至少需要 90 分钟。相比之下，遥感图像的幅面更大，地物种类更加复杂、密集，甚至需要一定的领域专家知识，标注的速度难以跟上图像更新的速度。

2. 泛化性不足。深度语义分割模型通常基于一个假设，即数据之间是独立同分布的。但在现实应用中很难满足这一条件，测试场景往往非可控，当训练数据和测试数据分布存在差异，即通常所说的域偏移（domain shift）现象，那么模型在测试集上的表现将急剧下降。在遥感领域中，因传感器类型、地理位置、地物分布等差异的影响，不同数据集间的遥感影像差异很大，呈现出十分明显的“同谱异物、同物异谱”现象，如图 1-2 所示。

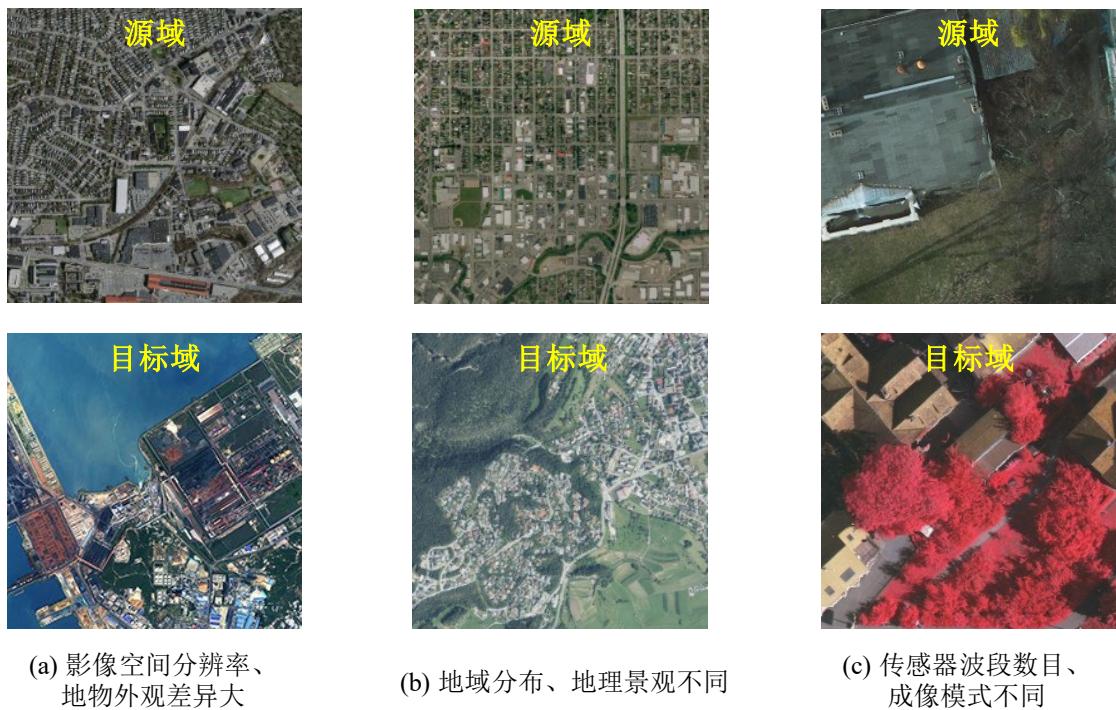


图 1-2 不同域间遥感图像示例图

为了解决上述问题，领域自适应（domain adaptation）这一技术被提出。领域自适应技术旨在让模型在源域（训练集）中学习的知识可以在目标域（测试集）中得到应用，并呈现出较好的性能。通常源域为一个或多个，其包含丰富的精准标签数据，目标域中标签数据十分稀少或无标签数据。领域自适应的核心思路就是将源域和目标域的数据分布对齐，从而使深度学习模型能够完成对目标域数据的正确分类。基于此核心思路，涌现出了许多优秀的研究成果。从基于 MMD 距离损失函数的训练方法，到基于特征空间或输出空间对抗学习的方法；从直接迁

移源域图像风格至目标域，到约束模型在输出空间低熵等。研究学者所提出的杰出的工作为领域自适应提供了广阔视野，推动了领域自适应技术的发展。然而，针对遥感图像的跨域语义分割问题还有很大的研究空间，如何高效地对齐源域和目标域的数据分布，进而提升跨域遥感图像语义分割模型的割精度是研究的难点和亟需解决的关键问题。

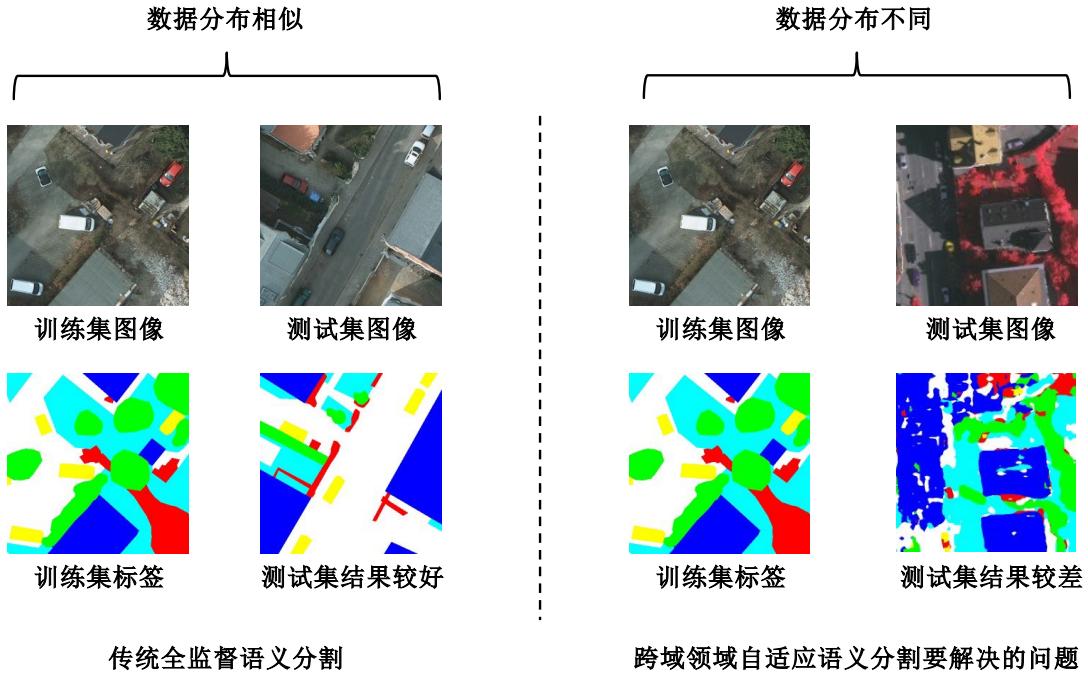


图 1-3 全监督语义分割与跨域领域自适应语义分割区别

1.2 国内外研究现状

语义分割是图像理解任务中的一个重要环节，可以被视为传统图像分割算法上的进一步深入拓展，其对场景的理解更为细致，提取的信息量更加丰富。语义分割旨在对图像中的每一个像素进行分类，为其分配一个预定义的语义类别标签（如属于建筑、道路、植被或汽车等），进而刻画出目标物体的轮廓信息。目前基于深度学习的语义分割方法取得了显著成效，本节首先针对基于深度学习传统全监督语义分割相关代表性方法进行概述。随后，针对基于深度学习的领域自适应语义分割相关代表性方法进行阐述。

1.2.1 传统全监督语义分割技术

全卷积神经网络语义分割模型：全卷积神经网络（Fully Convolutional Networks, FCN）^[16]于 2015 年被 Long 等人提出，这是深度学习技术用于语义分割的开山之作。传统的卷积神经网络模型（如 AlexNet^[16]、VGG^[16]等）在经过一系列卷积、池化操作后会使用全连接层，经过 Softmax 产生类别的概率信息用于

获得分类结果。而在 FCN 结构中将全连接层替换为卷积层，这样就可以获得一张二维特征图，后面续接 Softmax 层用来得到对每个像素点的预测信息，从而获得语义分割结果。此外，通过引入跳级（skip）结构，也就是把倒数几层的输出与最后的输出进行加和融合，获得更加稠密精细的预测结果。与传统的分割算法相比，FCN 无需人为手工设计特征，极大地简化处理流程，实现了基于端到端的卷积神经网络的图像语义分割，并在测试集上取得了优秀的性能，该项工作被认为是深度学习语义分割方法的一个里程碑。

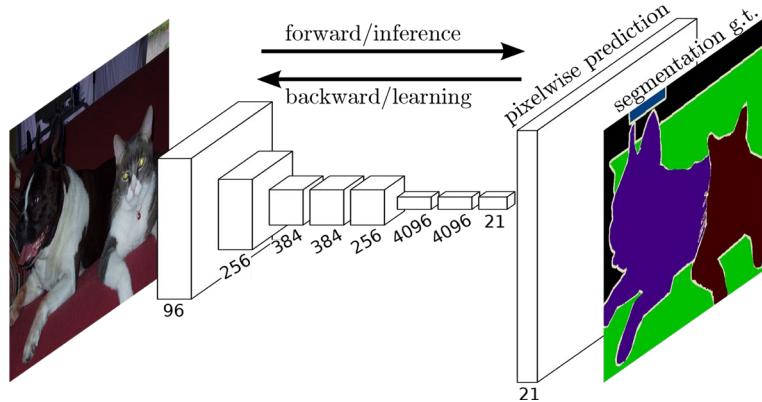


图 1-4 全卷积神经网络^[16]

基于编码器—解码器架构的语义分割模型：由于全卷积网络中涉及融合底层池化特征，这会导致物体的分割边缘不够清晰。此外全卷积网络没有充分考虑像素之间关系，导致分割结果缺乏空间结构一致性。针对上述问题，基于“编码器—解码器”这一结构的语义分割模型被创提出。其中最具有代表性的为 Unet^[17] 和 SegNet^[18]。在 SegNet 网络模型中“编码器”利用卷积层和池化层不断提取图像的抽象信息，具体表现为图像在宽和高维度不断减小，通道维度不断增加。“解码器”则是通过卷积和上采样操作将特征转换为最终的语义分割预测图。如图 1-5 所示的即为 SegNet 的网络结构。具体而言，该模型使用 VGG16 的前 13 层卷积网络作为“编码器”，“解码器”采用了与“编码器”完全对称的镜像结构。此外，编码器在进行最大池化操作时，会记录最大池化的索引。解码器根据池化的索引值进行非线性采样，实现从低分辨率的编码特征图恢复出细粒度的语义分割预测。

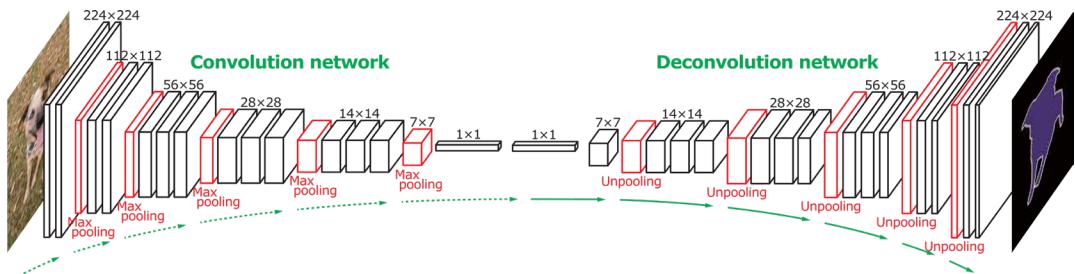


图 1-5 基于编码器-解码器的语义分割网络 SegNet 模型^[18]

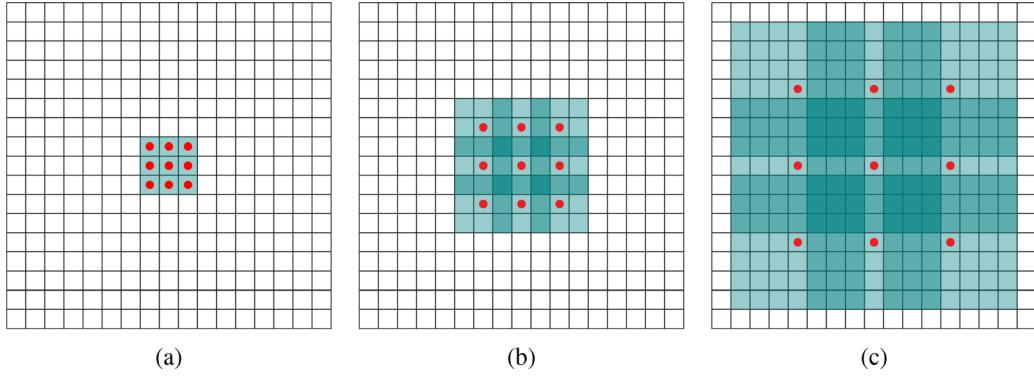


图 1-6 不同膨胀率的空洞卷积

基于空洞卷积的语义分割模型:对于语义分割任务，增大卷积网络在每个像素上感受野，可以获取更大范围的上下文信息，有助于对于每个像素点的精准分类。图 1-6 展示了不同膨胀率下的空洞卷积。Google 团队提出了基于空洞卷积的 DeepLab 语义分割网络，切实提升了语义分割算法的精度。DeepLab v1^[19]是基于卷积神经网络和概率图模型的语义分割模型。具体地，DeepLab v1 以 VGG 作为骨架网络，引入空洞卷积增大感受野，以此来减小池化和下采样操作所导致的细节丢失。通过条件随机场（Conditional Random Field, CRF）对分割图进行后处理，进一步提升语义分割模型对细节的处理能力。DeepLab v2^[20]在 DeepLab v1 的基础上进行了进一步改善。首先，DeepLab v2 将骨架网络进行了更改，由 VGG 网络换成了特征提取能力更强的 ResNet^[21]网络。此外，引入了空间空洞金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）结构，实现了对不同尺度的特征感知融合。在 DeepLab v3^[22]中 CRF 被移除。ASPP 模块被进一步实施了优化，采用了不同的采样率的空洞卷积，并引入 BatchNorm 层。Google 团队随后又发布了 DeepLab v3 plus^[23]模型，该模型以 Xception^[24]作为骨干网络，使用了编码—解码架构，并引入了带孔可分离卷积。DeepLab v3 plus 在 Pascal VOC 2012^[25]上实现了 89% 的 mIoU，成为了当时最领先的语义分割方法。

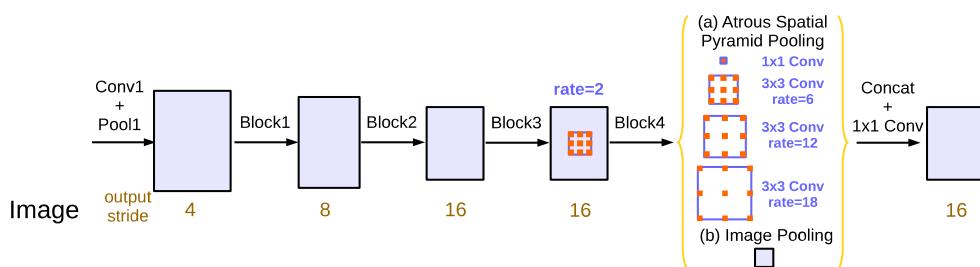
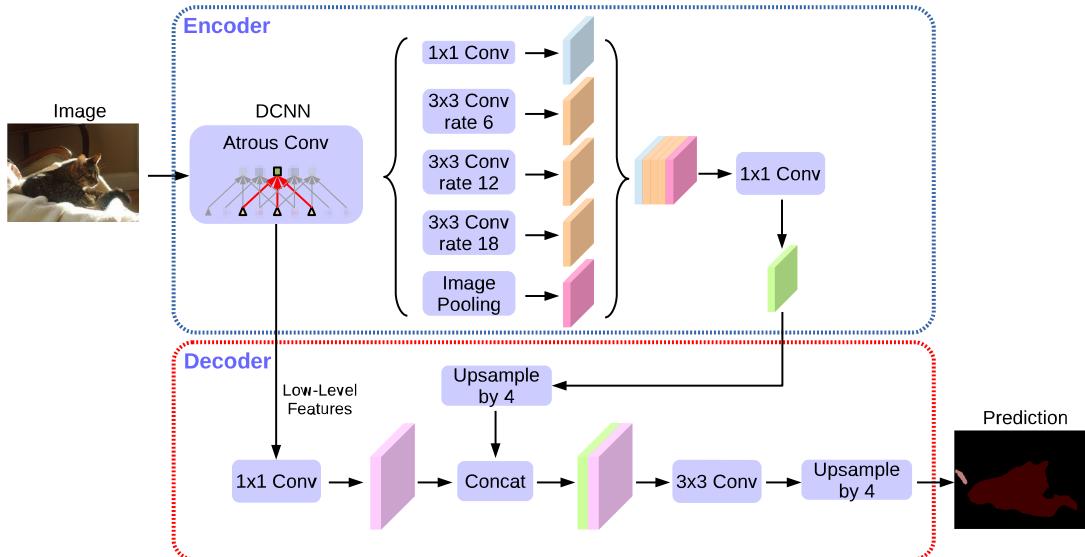
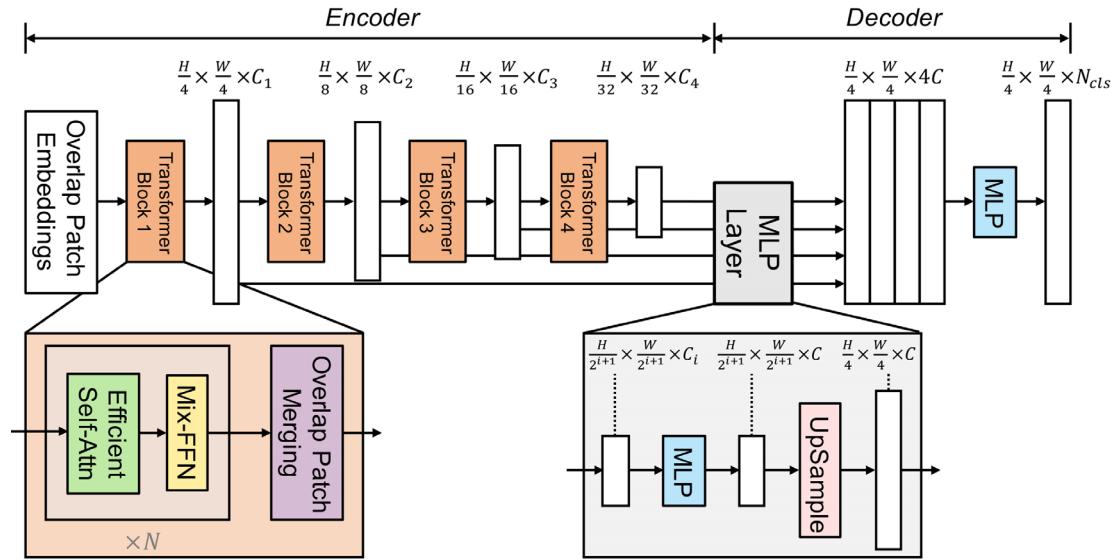


图 1-7 ASPP 模块

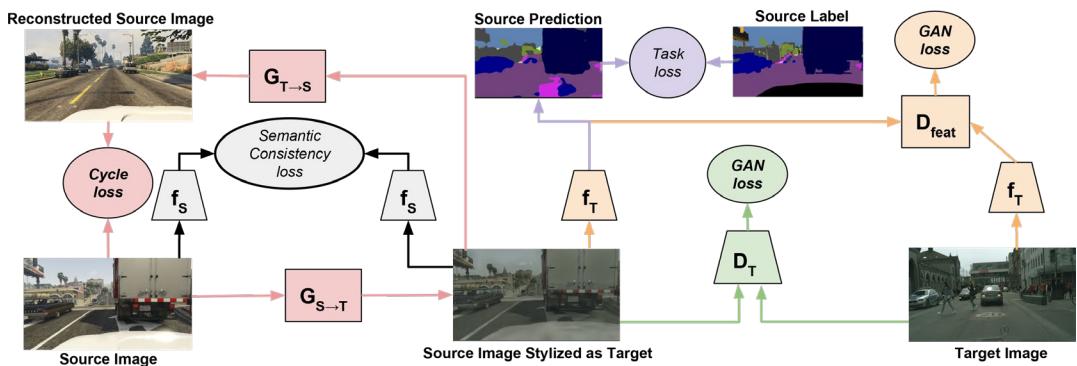
图 1-8 DeepLab v3 plus 模型结构^[23]

基于 Transformer 的语义分割模型：随着 Transformer^[26] 模型架构在自然语言处理领域的大放异彩，计算机视觉领域的研究学者针对该模型架构开展了大量研究，涌现出了许多基于 Transformer 架构的语义分割模型。得益于 Transformer 自注意力机制所建立的长效远距离依赖关系，这些语义分割模型取得了显著的成效。Segmenter^[27]基于 Transformer 架构，构建了“编码器-解码器”形式的语义分割模型。与基于卷积神经网络语义分割模型存在两点重要不同：一是，在图像输入 Segmenter 模型前，需要将图像裁剪成不同的 patch，并将其展平为一维向量序列。由于在这个过程中会导致空间维度信息的缺失，为此需要进行位置编码操作。二是 Segmenter 的编码器可以对图片的全局上下文信息进行建模。Segmenter 提供了两种解码器架构：线性解码器和基于 Transforer 的掩码解码器。其中，线性解码器通过对 patch 进行线性映、变形、上采样以及 Softmax 操作，得到语义分割结果。基于 Transforer 的解码器则添加了一组可学习的类嵌入向量到解码器的输入，实现了更优的语义分割效果。虽然 ViT^[28]架构在视觉领域取得了一定的成功，但其仍旧存在一些缺点，如参数量大、ViT 的柱状结构设计对语义分割的不友好等。为此 SegFormer^[29]算法被提出，针对 ViT 的上述劣势进行了优化。SegFormer 主要进行了以下几方面的改进：(1) 设计了层次化的 Transformer 编码器，其结构简单，只有 6 个线性层，因此参数量和计算量都比较小。此外，在进行 Patch 嵌入时将其设计成有重叠的，这样可以保证局部连续性；(2) 彻底去掉了位置编码，取而代之的是 Mix FFN 结构，即让模型自己去学习位置信息。SegFormer 不仅在经典语义分割数据集（如 ADE20K^[30], Cityscapes^[15]等）上取得了 SOTA 的精度，而且在 Cityscapes-C^[31]（对测试图像加各种噪声）上大幅度超过之前的方法（如 DeepLab v3+），反映出其强大的性能与良好的鲁棒性。

图 1-9 SegFormer 模型结构^[29]

1.2.2 领域自适应语义分割技术

基于图像迁移的方法: 受益于生成对抗网络 (Generative Adversarial Network, GAN) ^[32] 在图像迁移方向取得了重大研究进展 (例如, CycleGAN^[33]等模型), 许多基于 GAN 的领域自适应语义分割方法被提出。这些方法的核心思路是通过对源域图像进行迁移, 使其在视觉上与目标域图像相似, 这有助于在训练语义分割模型之前减少域偏移的影响。Hoffman 等人首先提出了 CyCADA^[34], 他们使用 CycleGAN 生成目标风格化的图像, 实现了特征级和像素级对齐。DCAN^[35]在生成器和分割网络中都探索了通道级特征对齐。Choi 等人^[36]提出了一种基于 GAN 的自集成数据增强方法, 通过迁移源图像样式来实现域对齐。CPN^[37]和 FDA^[38]通过简单的傅里叶变换及其反变换来迁移源图像的风格。CRIN^[39]利用 CycleGAN 生成内容不变图像, 并进一步引入辅助分类器模块, 重点关注像素级的域偏移, 进一步提高域自适应性能。

图 1-10 CyCADA 算法示意图^[34]

基于对抗学习的方法: 基于对抗学习的无监督领域自适应方法在语义分割方

面得到了广泛的研究。对抗训练通过引入判别器作为语义分割网络的附加模块，用于最小化源域和目标域的特征分布差异。具体地，判别器从语义分割网络中获取特征图，并尝试区分输入的域。语义分割网络被训练产生领域不变特征，使判别器无法分辨特征图所属的域。与此同时，语义分割网络在源域数据进行监督学习，又可以使提取到的特征具有类别可分性。经过这样的训练，语义分割网络在目标域上就能够产生良好的分割结果。Hoffman 等人提出的 FCN wild^[40]算法是第一个应用对抗性学习来对齐语义分割网络在全球范围内两个不同领域之间提取的特征图。Tsai 等人发现对于语义分割任务，直接在输出空间进行分布对齐更加有效。在输出空间的领域自适应能够实现预测和特征的联合优化，为此提出了 AdaptSegNet^[41]算法。此外，ADVENT^[42] 通过引入熵对抗损失，来强制目标域的预测具有高预测置信度（低熵）来实现领域自适应，为输出空间对抗学习提供了一种新的思路。

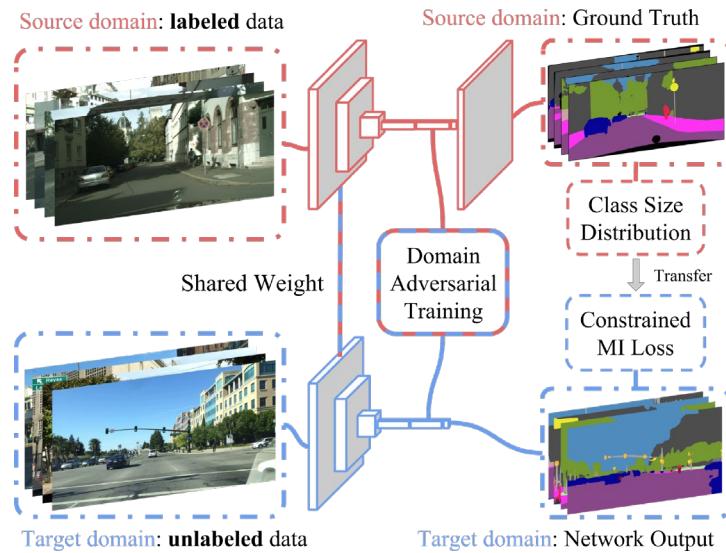


图 1-11 FCN wild 算法示意图^[40]

基于自学习的方法：自学习在无监督或半监督学习领域中得到了广泛的研究。其关键思想是利用集成模型或先前模型的高置信度预测作为未标记数据的伪标签，迫使模型以隐式方式学习域不变特征。CBST^[43]算法使用了一种迭代的自训练的策略，通过动态调整类别权重，从而在目标数据上生成更准确的伪标签，并使用这些伪标签对模型进行再训练。Xu 等人^[44]提出 SEANet 方法，利用自集成注意网络提取注意感知特征进行领域自适应。Wang 等人^[45]提出了一种多原型聚类方法，增强了目标域的类内紧凑性和类间分离性，便于构建任务特定决策边界。Yuan 等人^[46]提出了一个同时考虑类别同质性和多样性的多层次无监督领域适应框架。DACS^[47]通过跨域混合采样策略，将源域的图像、标签以及目标域的图像、伪标签进行混合，随后利用源域样本和混合样本进行训练，实现领域自适应语义

分割。

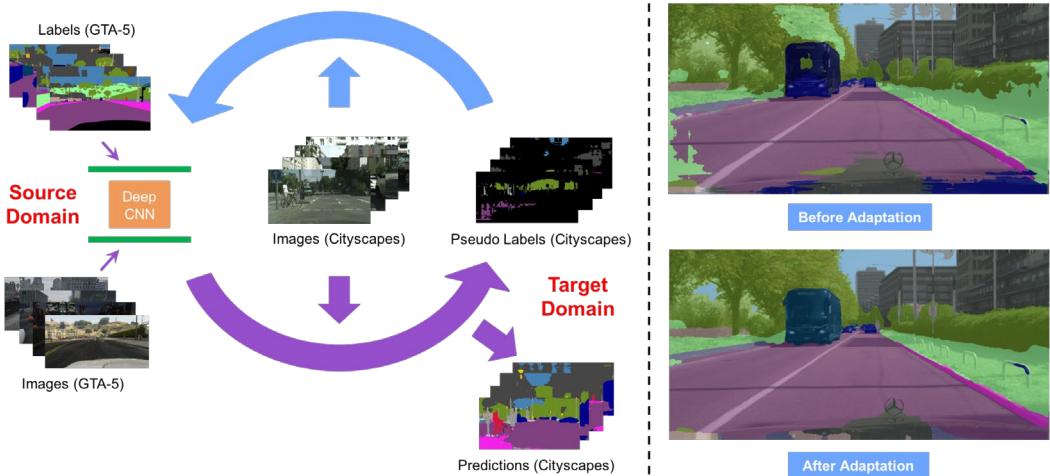


图 1-12 CBST 算法示意图^[43]

1.3 主要工作及贡献

遥感影像语义分割是遥感图像理解工作中的重要一环，其应用前景广泛。随着深度学习技术的再次兴起，基于深度学习的遥感图像语义分割算法已经成为了主流，并取得了优越的性能表现。但如何缓解对标注工作的压力，如何利用已有数据集训练模型，泛化到目标域数据集成为了现今的研究热点。本文开展的工作即为基于生成对抗网络的遥感图像跨域语义分割问题。本文以生成对抗网络图像迁移为主线，开展了关于跨域遥感图像语义分割的研究工作，提出了两种高效的跨域语义分割方法：一是基于多重弱监督约束的域适应遥感图像语义分割方法；二是基于语义保持生成对抗网络的域适应遥感图像语义分割方法。下面分别对二者进行介绍。

(1) 基于生成对抗网络的域适应遥感图像语义分割方法。

针对源域训练的语义分割模型难以泛化到目标域场景，提出了一种基于多重弱监督约束的遥感图像跨域语义分割方法。该方法可以在有标注的源域数据集上训练，应用在目标域无标注的数据集上测试，一定程度克服了域偏移问题的影响，在目标域数据集仍旧能保持较好的性能。具体地，该方法提出了一种带有多重弱监督约束的损失函数，由弱监督迁移不变约束、弱监督伪标签约束和弱监督旋转一致性约束构成。首先，利用生成对抗网络将源域图像迁移至目标域风格，减少像素级的分布差异。然后，利用迁移后的图像和其原始标签训练语义分割网络，构建弱监督迁移不变损失；通过设置阈值自适应地生成目标域图像的伪标签，构建弱监督伪标签约束损失；基于遥感图像成像角度任意这一特性，刻画了弱监督旋转一致性约束损失。为了充分发挥多个约束的优点，该方法提出了一种动态优化策略，在训练过程中动态调整目标函数的约束权值，有助于缓解深度语义分割

网络的退化。本方法在跨地理位置和跨成像模式两个遥感图像跨域语义分割任务上相比同期代表性算法取得了更优的结果。

(2) 基于语义保持生成对抗网络的域适应遥感图像语义分割方法。

针对现有的基于生成对抗网络的图像迁移模型未充分考虑语义信息，致使迁移后的图像往往会产生偏差，导致迁移效果有限这一问题，提出了一种语义保持的生成对抗网络用于域适应遥感图像语义分割。该方法通过在生成对抗网络模型中引入表征不变约束和语义保持约束，实现从源域到目标域的语义保持图像风格迁移。此外，考虑到源域和目标域数据的类别分布存在差异，提出了类别分布对齐语义分割模块。在模型输入层面，引入类别混合操作—ClassMix 操作，通过随机采样的方式从迁移后的图像上随机选择一定的类别，将相应的图像块粘贴到目标域图像上，构成混合图像。将迁移后的图像与混合图像一同送入语义分割网络进行训练。在模型输出层面，提出了边界增强损失，来细化语义分割网络在目标边界的性能。所提出的方法，在跨地理位置、跨成像模式以及跨景观布局三个典型的跨域遥感图像语义分割任务都取得了最优的性能，充分证明了算法的优越性与普适性。

1.4 论文组织结构

本文共分为五个章节，论文组织结构如图 1-13 所示，每章的具体安排如下：

第一章：绪论。本章首先论述了遥感图像语义分割的概念、应用前景，描述了遥感大数据时代下深度语义分割模型面临的困难与挑战。随后，进一步阐述了跨域遥感图像语义分割的研究背景与意义；介绍了与本文研究密切相关的语义分割和领域自适应语义分割的研究现状。最后，对本文的研究内容与目标进行了阐述，以及介绍了论文的主要研究内容和结构安排。

第二章：相关理论研究基础。首先，本章介绍了领域自适应的相关概念以及三种典型的方法，具体包括基于模型的领域自适应、基于样本的领域自适应以及基于特征的领域自适应算法。随后，介绍了生成对抗网络的理论基础，以及与本文密切相关的基于生成对抗网络的图像迁移模型。

第三章：基于对偶生成对抗网络的域适应遥感图像语义分割。本章进一步介绍遥感图像跨域语义分割的研究背景、研究动机，并提出了一种基于多重弱监督约束的遥感图像跨域语义分割方法。该方法由弱监督迁移不变约束、弱监督伪标签约束和弱监督旋转一致性约束构成。首先，利用对偶生成对抗网络将源域图像迁移至目标域风格，减少像素级的分布差异。然后，利用迁移后的图像和其原始标签训练语义分割网络，构建弱监督迁移不变损失；通过设置阈值自适应地生成目标域图像的伪标签，构建弱监督伪标签约束损失；基于遥感图像成像角度任意

这一特性，刻画了弱监督旋转一致性约束损失。此外，该方法提出了一种动态优化策略，在训练过程中动态地调整更新损失约束的权重，充分地发挥了上述约束的优势。

第四章：基于语义保持生成对抗网络的域适应遥感图像语义分割。本章针对已有生成对抗网络存在的不足以及在遥感领域中跨域语义分割任务中面临的难题，进行了更全面的探索。首先，本章提出了基于语义保持的生成对抗网络，实现源域到目标域的无偏图像迁移。此外，进一步考虑到源域数据集的类别分布和目标域数据集的类别分布存在差异，提出了类别分布对齐语义分割模块，进一步缩小了不同数据集之间的景观布局差距。在模型输入层面，引入类别混合操作—ClassMix 操作，通过随机采样的方式从迁移后的图像上随机选择一定的类别，将相应的图像块粘贴到目标域图像上，构成混合图像。将迁移后的图像与混合图像一起送入语义分割网络进行训练。在模型输出层面，提出了边界增强来细化目标边界的性能。利用这两个方面协同训练出领域自适应语义分割模型。

第五章：总结与展望。对本文所展示的两个方法进行了全面总结，反思了方法存在的不足之处，并对今后的研究方向和内容进行了展望。

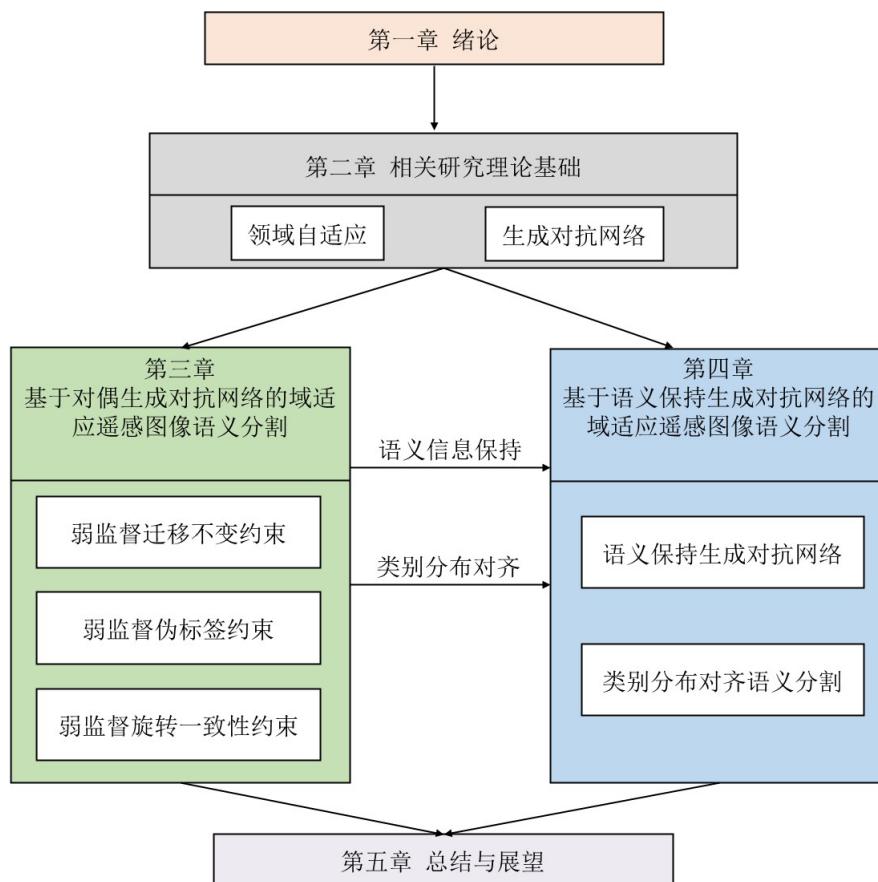


图 1-13 论文组织结构

第2章 相关研究理论基础

2.1 领域自适应方法

经典的机器学习算法通常会基于“独立同分布”这一假设，即训练数据与测试数据服从相似的概率分布。但在现实的许多场景中，这一假设很难得到满足。这将使得在训练集上表现优越的模型在实际应用过程中，往往无法将训练场景中学到的知识运用于测试数据中，致使结果无法达到预期。若测试场景与训练场景存在域偏移（domain shift），即二者的概率分布不一致，那么训练场景的模型应用在测试场景时性能会较差，领域自适应（domain adaptation）技术由此应运而生。

领域自适应技术是迁移学习中的一个细分研究方向，通常源域（source domain）的样本含有丰富的标签信息，目标域（target domain）的样本含有少量标签信息或者不含有标签信息。领域自适应的目的是利用源域样本实现领域不变知识的跨域迁移和复用，提升模型的泛化能力。图 2-1 形象地表示了传统机器学习算法和领域自适应算法的区别。下面给出域适应的形式化定义。给定一个有标签的源域 $D_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ 和一个无标签的目标域 $D_t = \{x_j^t\}_{j=1}^{n_t}$ 。设二者共享相同的特征空间和标签空间，即 $\mathcal{X}_s = \mathcal{X}_t$, $\mathcal{Y}_s = \mathcal{Y}_t$ 。此外，二者的条件概率分布也相同 $P(y^s | x^s) = P(y^t | x^t)$ 。但边缘概率分布不同，即 $P(x^s) \neq P(x^t)$ 。领域自适应算法的目标就是利用有标签的源域样本 D_s 学习一个神经网络模型 $f: x \mapsto y$ ，来实现对目标域样本 D_t 预测其标签 $y^t \in \mathcal{Y}_t$ 。通过在不同的阶段采取领域自适应的相关措施，可以将常见的方法大致分为三类：（1）基于模型的领域自适应；（2）基于样本的领域自适应；（3）基于特征的领域自适应。

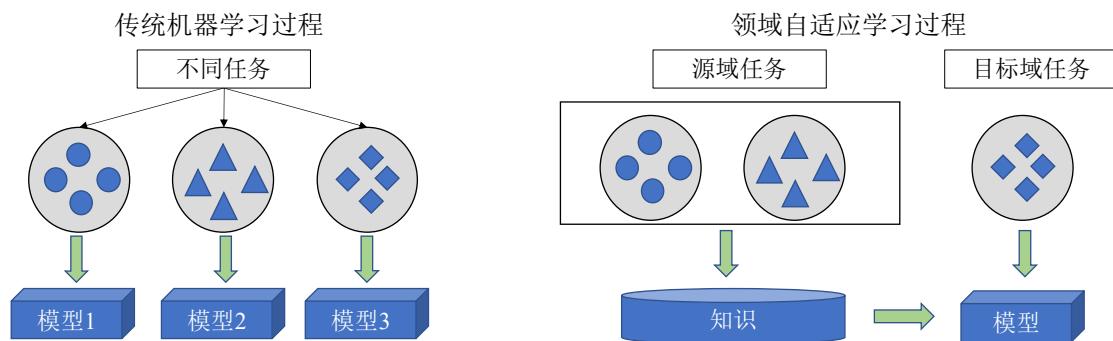


图 2-1 传统机器学习与领域自适应学习异同

2.1.1 基于模型的领域自适应

基于模型迁移的方法是指从模型参数中找出源域和目标域可以共享的部分，以此来实现域适应。例如现有一个训练过的深度分类网络，当该分类深度分类网

络要应用到一个新的图像分类任务中，往往不会从零开始训练，通常会冻结住全连接层前的网络部分，只需要新任务的数据集上微调（fine-tuning）全连接层的参数即可实现对新任务的图像分类。Schwaighofer 等人^[48]提出用通过共享模型的参数 w_0 实现不同域之间的连接，倾向于直接对模型参数进行迁移；Yosinski 等人^[49]率先对神经网络的可迁移性进行了探索，得出结论认为神经网络的浅层学习的是通用特征，迁移效果会比较好；神经网络深层学习到的特征是与任务强相关的，因此迁移性较差。许夙晖^[50]等人提出了基于极限机器学习的域适应参数迁移算法，主要思想是将目标域分类器的参数向源域参数空间进行投影，使两个分类器的参数分布尽可能相似。余欢欢等人^[51]提出模型参数自适应迁移的方法（Adaptive Transfer for Model Parameter, ATMP），以私有和共有的方式对源域的模型参数进行字学习，随后将在源域学到的模型参数字典作为目标域模型的参数字典，再对字典的系数进行稀疏化约束，从而实现模型参数由源域到目标域的自适应选择迁移。

2.1.2 基于样本的领域自适应

源域和目标域的差异主要是二者的数据不相似，因此样本层面的适应是指重新加权源域的有标注样本，进而实现源域数据项目标域数据的对齐。这种方法的核心思路是对于每个样本计算出其属于源域或目标域的似然比，随后用这一似然比对样本进行加权。最常用的加权方法是通过计算源域与目标域数据分布的最大均值差异（Maximum Mean Discrepancy, MMD），从而实现对样本的加权。Huang 等人提出核均值匹配（Kernel Mean Matching, KMM）^[52]对概率分布进行估计，从而拉近源域和目标域的概率分布。此外，TrAdaBoost^[53]方法也是样本加权法的典型代表。该方法将 AdaBoost 的思想应用到迁移过程中。在对目标分类器进行训练的过程中，对有利于目标任务的样本提升其权重。反之，对不利于目标任务的样本则要对其赋予更低的权重，然后在更新权重后的样本上继续训练目标分类器，如此迭代进行。Zhang 等人提出重要性加权对抗网络（Importance weighted adversarial net, IWAN）^[54]。IWAN 由特征提取器 F_s 和 F_t 、领域鉴别器 D 和 D_0 以及一个类别预测器 C 组成。首先，对 F_s 和 C 进行预训练；在随后的训练过程中固定 F_s ，并利用 F_s 的参数初始化 F_t 。接下来利用领域鉴别器 D 的输出对源域样本加权。最后，对目标域特征提取器 F_t 与另一个领域鉴别器 D_0 进行对抗训练。此外，Cao 等人^[55]也构建了类似的用于部分迁移学习的选择性对抗网络。

2.1.3 基于特征的领域自适应

特征层面的领域自适应方法的主要思想是减少源域和目标域在特征空间的

差异。最大均值差异（Max Mean Discrepancy, MMD）是衡量源域和目标域差异最常用的指标。对于给定的源域 D_s 和目标域 D_t , MMD 定义如下：

$$\begin{aligned} d_{\mathcal{H}}^2(\mathcal{D}^s, \mathcal{D}^t) = M_{\text{MMD}}(X^s, X^t) = \\ \left\| \frac{1}{n^s} \sum_{i=1}^{n^s} \phi(x_i^s) - \frac{1}{n^t} \sum_{j=1}^{n^t} \phi(x_j^t) \right\|_{\mathcal{H}}^2 \end{aligned} \quad (2-1)$$

其中 ϕ 是核函数，用于将原始数据映射到一个再生核希尔伯特空间（Reproducing kernel Hilbert space, RKHS）。

MMD 通过计算原始数据在再生核希尔伯特空间中的均值来衡量源域和目标域数据分布的差异。在浅层领域自适应方法中，Pan 等人提出迁移成分分析算法（Transfer Component Analysis, TCA）^[56]。TCA 算法通过 MMD 来衡量源域和目标域的特征差异，学习一个从经验核特征空间到低维空间的线性映射。随后可以在这个低维空间中产生新的表示，利用传统的机器学习算法训练模型，在目标域进行推理与测试。在深度网络领域自适应中，2015 年 Long 等人提出深度适应网络（Deep Adaptation Network, DAN）^[57]。该方法以条件概率不变为基准假设，通过向 AlexNet 网络最后 3 个全连接层添加关于不同域间分布差异的度量，从而拓展为多核最大均值差异（Multi-kernel MMD, MK-MMD）除 MMD 外，还有最大均值匹配距离（Kernel Mean Matching, KMM）^[52]、关联对齐距离（Correlation alignment, CORAL）^[58,59]等领域距离度量方式，也涌现出了许多基于此类距离度量的算法。

此外，通过直接对源域图像迁移，进行源域与目标域图像的像素分布对齐也可视为特征层面的领域自适应。这种方法通常与图像风格迁移技术紧密相关，即通过将源域图像的风格迁移至目标域图像风格，从而缩小域偏移问题的影响。Hoffmann 等人率先提出 CyCADA^[34]方法，该方法利用 CycleGAN 将源域图像的风格迁移至目标域，利用迁移后的图像和其原始标签进行训练语义分割网络，并同时考虑到了特征级和像素级的域差异。Tasar 等人提出 ColorMapGAN^[60]，其生成器十分简洁，没有任何卷积层和池化层，仅需要学习两个矩阵，通过执行一次矩阵元素点乘和一次矩阵加法操作，即可将源域数据映射到目标域的色彩风格。

除上述研究外，也涌现出了大量基于对抗学习的领域自适应方法。对抗学习的思想源自于生成对抗网络（Generative Adversarial Network, GAN）。通常该类方法会有一个特征提取器 F ，分类器 C ，以及一个域鉴别器 D 。其对应参数分别用 θ_F 、 θ_C 和 θ_D 表示。域判别器 D 通过训练来区分源域和目标域的特征，而特征提取器 F 通过训练学习具有领域不变性的特征来“骗过”判别器 D。特征提取器 F 和分类器 C 在源域数据进行监督学习，又可以使提取到的特征具有类别可分性。上述的原理可以概括为如下的目标函数：

$$\begin{cases} \min_{\theta_C, \theta_F} \mathcal{L}_C(\mathcal{D}_S, \mathcal{Y}_S; \theta_C, \theta_F) - \lambda \mathcal{L}_D(\mathcal{D}_S, \mathcal{D}_T, \theta_D; \theta_F) \\ \min_{\theta_D} \mathcal{L}_D(\mathcal{D}_S, \mathcal{D}_T, \theta_F; \theta_D) \end{cases} \quad (2-2)$$

Ganin 等人基于上述框架率先提出域对抗网络 (DANN)^[61], 融入了 GAN 的对抗思想, 并取得了重大突破。随后基于深度对抗网络的领域自适应算法逐渐成为了主流。

2.2 生成对抗网络方法

2.2.1 生成对抗网络

受博弈论中零和博弈的启发, 生成对抗网络 (Generative Adversarial Networks, GAN)^[32]于 2014 年被 Ian Goodfellow 等人提出。GAN 的模型结构如图 2-2 所示, 包括一个生成器 (Generator, G) 和一个判别器 (Discriminator, D)。

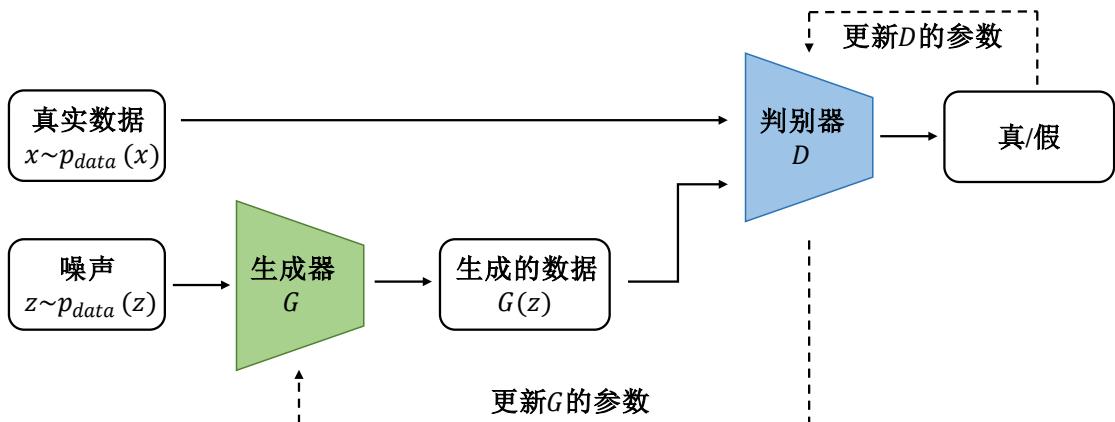


图 2-2 生成对抗网络结构图

生成器 G 的输入为服从标准正态分布 $N \sim (0,1)$ 的随机噪声向量 z , 输出得到生成的图像样本 $G(z)$ 。在生成器训练过程中, 其目标是“骗过”判别器 D , 使判别器 D 无法判断数据来自真实样本还是生成样本。判别器 D 在训练过程中用于判别数据是来自真实样本 x 还是生成样本 $G(z)$, 最终的输出为数据是真实样本的概率, 起到一个二元分类器的作用。判别器会为真实图像给予标签 1, 对生成的图像给予标签 0。而生成器试图让判别器将生成的图像“误判”为 1。通过这种方式, 生成器与判别器交替不断地进行对抗训练, 二者在相互对抗的过程中不断优化, 优化后的生成器 G 所产生的图像也变得越来越完美, 不断地逼近真实的图像数据。训练收敛后整个模型达到纳什平衡的状态, 即生成器 G 生成的样本分布和训练数据的样本分布一样, 判别器 D 对其输入的数据给出的都是 0.5 的分类概率。设 P_{data} 表示真实图像 x 的概率分布, P_g 表示生成图像 $G(z)$ 的概率分布, P_z 表示随机噪声向量 z 的数据分布 $N \sim (0,1)$ 。生成对抗网络的优化目标如公式 (2-3) 所示。

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (2-3)$$

其中 $\mathbb{E}_{x \sim p_{data}(x)}$ 表示判别器对真实数据的评分期望值, $\mathbb{E}_{z \sim p_z(z)}$ 表示生成数据的评分期望值。

在训练过程中, 通过优化生成器 G 的参数来最小化损失函数 $V(G, D)$, 而通过优化生成器 D 的参数来最大化损失函数 $V(G, D)$ 。生成器 G 和判别器 D 的优化目标相反, 构成一个博弈的过程。对生成器 G 和判别器 D 进行交替训练, 直到最终二者达到纳什平衡的状态。训练的过程如算法所示。

算法 1: 生成对抗网络的训练算法

```

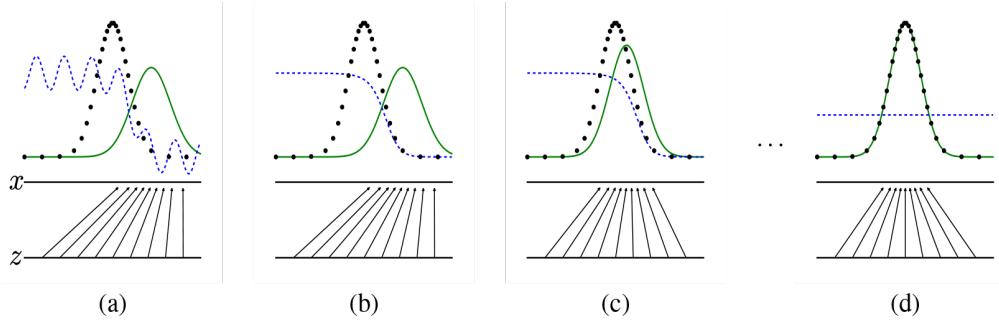
1  初始化生成器 G 和判别器 D;
2  while 模型还没有收敛 do
3      for k 次迭代 do
4          由  $P_z \sim N(0,1)$  随机采样  $m$  个样本  $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ ;
5          由  $P_{data}$  中随机采样  $m$  个样本  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ ;
6          使用梯度上升法更新判别器  $D$ :
7              
$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log (1 - D(G(z^{(i)})))]$$

8      end for
9      由  $P_z \sim N(0,1)$  随机采样  $m$  个样本  $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ ;
10     使用梯度下降法更新生成器 G:
11         
$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)})))$$

10  end

```

图 2-3 给出了生成对抗网络模型训练优化过程的示意图, 其中真实数据的分布用黑色的虚线表示, 生成数据的分布用绿色的实线表示, 判别器 D 的判别函数用蓝色的虚线表示。在图 2-3 (a) 中判别器 D 未得到充分训练, 是一个部分精准判别器, 对生成数据和真实数据的区分能力有限; 当训练到图 2-3 (b) 的过程, 判别器 D 得到了一定的训练, 能较好地对生成数据和真实数据进行区分, 判别器 D 会指导生成器 G 更新梯度从而产生高质量的图像; 在图 2-3 (c) 中随着不断进行交替迭代训练, 绿色实线一步一步地向着黑色虚线移动, 即生成的图像会不断地逼近真实图像。最后的训练过程收敛至图 2-3 (d), 将得到 $p_g = p_{data}$, 即判别器 D 难以区分真真实和生成分布, $D(x) = D(G(z)) = 0.5$ 。

图 2-3 GAN 训练过程示意图^[32]

2.2.2 循环一致生成对抗网络

随着生成对抗网络的问世，它涉猎的领域也愈发的宽泛，而在图像风格迁移的领域，它的表现更加卓越，因为生成对抗网络可以生成更加清晰、线条更加细致真实的样本图像，并在图像风格迁移上有着出色的效果。图像风格迁移可以视为一种生成任务，输入是一种高维数据，输出是另一种高维数据。根据是否需要配对的数据，基于生成对抗网络的图像风格迁移方法可以分为两类。一类是有监督的学习方式，代表性的就是基于条件生成对抗网络（Conditional GAN）^[62]的图像风格迁移，这种方法是在数据源中找到配对的原始图像和目标图像。目标图像即相当于原始图像的标签。另一类是无监督的学习方式，代表性的就是循环一致性生成对抗网络（CycleGAN）^[33]，该方式的数据源中的数据之间没有映射关系，即为无监督学习。在实际的应用场景中，很难获取完全匹配的图像数据，比如照片到莫奈油画的转换、马到斑马的转换等。因此在多数场景下，无法使用 cGAN 这类监督学习方法。针对上述问题，2017 年 CycleGAN、DualGAN、DiscoGAN 这些以无监督方式进行的，不需要配对数据训练的模型被提出。图 2-4 直观的展示了 CycleGAN 的图像迁移示例。下面将以 CycleGAN 为例，对基于 GAN 的图像迁移模型进行介绍。

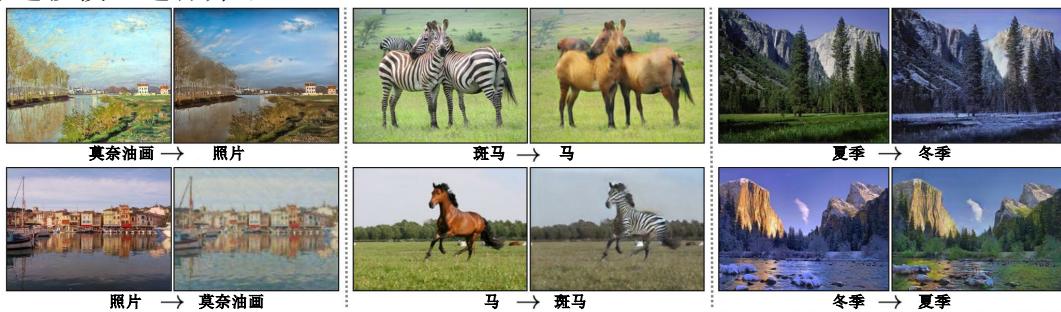


图 2-4 CycleGAN 图像迁移示例

CycleGAN 的结构如图 2-5 所示，包含两个生成器 G 和 F ，两个判别器 D_X 和 D_Y 。CycleGAN 需要学习两个方向的映射， $G: X \mapsto Y$ 和 $F: Y \mapsto X$ 。判别器用于区分域 x 的图像和生成后的图像 $F(y)$ 。 D 用于区分域 Y 的图像和迁移后的图像 $G(x)$ 。

损失函数包括对抗损失和一致性损失。其中 $X \mapsto Y$ 方向 G 和 D_Y 的对抗损失为

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))]$$

对于 $Y \mapsto X$ 方向的对抗损失与其类似。

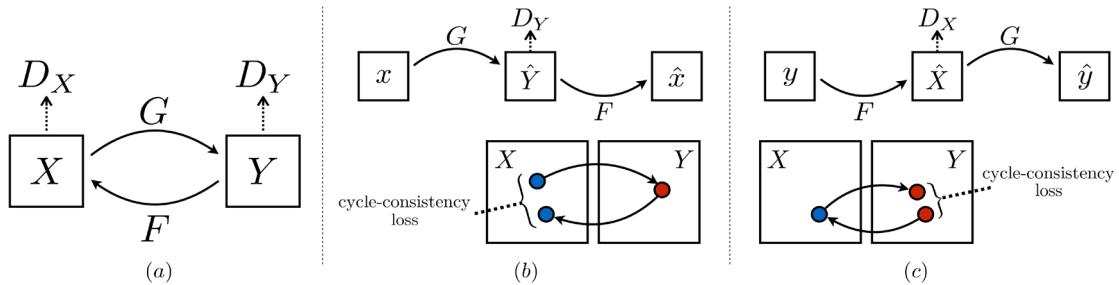


图 2-5 CycleGAN 结构示意图

另一项损失为循环一致性损失，该损失确保了非配对数据的有效图像迁移过程，避免模型崩溃。如图所示，对于域 X 中的图像 x ，首先经过转换得到 $G(x)$ ，随后又经过一次转换得到 $F(G(x))$ 。要求经两次转换后的图像与原始图像保持一致，这个过程可表示为 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ 。因此循环一致性可以表示为公式(2-4) 所示。

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1] \quad (2-4)$$

因此总的损失项可表示为：

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(G, D_X, Y, X) + \mathcal{L}_{cyc}(G, F) \quad (2-5)$$

2.3 本章小结

本章主要对领域自适应基础理论和生成对抗网络理论及其应用进行了基本的介绍。首先，阐述了领域自适应方法与传统机器学习算法的区别。详细介绍了三种典型的领域自适应算法，包括基于模型的领域自适应算法、基于样本的领域自适应算法以及基于特征的领域自适应算法。其中基于特征的领域自适应方法被研究的最为深入广泛。本文后续的研究均属于该方式下的领域自适应算法，进而对齐源域和目标域的特征，使语义分割模型在目标域数据有良好的性能。随后，本章对生成对抗网络的理论基础进行了介绍。讲述了 GAN 在图像迁移方向的应用。本文的后续研究也将围绕 GAN 的图像迁移展开，并通过分析已有的 GAN 模型在遥感图像迁移过程中的不足，充分考虑到遥感图像数据特性提出新的 GAN 模型用于遥感图像迁移。

第3章 基于对偶生成对抗网络的域适应遥感图像语义分割

3.1 概述

随着遥感、计算机科学和通信工程等多个领域的快速发展，遥感图像的规模呈爆炸式增长，为大规模地表监测提供了可能。因此，我们进入了遥感大数据时代^[4, 63, 64]。在这个遥感大数据时代，遥感图像自动判读对于有效挖掘遥感大数据的价值具有重要意义。由于其在城市规划^[65, 66]、作物评估^[67, 68]、环境监测^[69]和智能交通^[70]等方面的广泛应用，近年来遥感图像语义分割引起了越来越多的研究兴趣。具体来说，遥感图像语义分割旨在为遥感图像中的每个像素分配一种土地覆盖类型（如建筑物、树木）。近年来，卷积神经网络（Convolutional Neural Networks, CNN）在遥感图像检索^[71]、遥感图像目标检测^[9, 72]和遥感图像场景分类^[73, 74]得到了广泛应用。深度语义分割网络作为 CNN 的一种扩展，最早由计算机视觉领域的学者提出^[75]，并被进一步引入解决遥感图像语义分割问题^[76-79]。当遥感图像配有精准的人工标注时，深度语义分割网络可以端到端方式进行有效训练，且性能明显优于传统方法（如手工制作的基于特征的方法）^[80]。然而，深度语义分割网络的出色性能在很大程度上依赖于强有力的监督训练（即具备大量标注训练数据）和相似的数据分布（即测试数据与训练数据具有相似的外观特征）。

如 Cityscapes^[15]数据集中所描述，一幅自然城市景观图像的像素级标注平均需要近 90 分钟。与自然图像相比，遥感图像通常呈现出更复杂的结构。由于遥感图像的类间混淆，标注过程往往需要大量的领域专家知识。总体而言，遥感图像复杂的结构和类间混淆使得像素级的遥感图像标注更加耗时和昂贵。在遥感大数据时代，遥感图像的收集变得越来越容易，但如何为遥感图像构建像素级的标签成为了实际的挑战。在此基础上，遥感图像语义分割还需要进一步研究如何减少标注数据的监督依赖。一种可能的解决方案是用源域标记的遥感图像训练深度语义分割网络，然后利用训练后的深度语义分割网络为目标域的遥感图像预测。但不同遥感图像数据集之间往往有着较大的差异，在源域训练的深度语义分割网络应用在目标域时往往性能严重退化。

与计算机视觉领域的跨领域语义分割相比，跨领域遥感图像语义分割会面临一些特有的挑战。由于遥感图像采集条件的多样性，包括成像传感器、不同的地理空间区域、地面采样距离和任意的拍摄角度^[81, 82]，遥感图像往往呈现出成像模式多样、物体尺度多、色彩饱和度多样等鲜明特征。在现实中，这些遥感特征往往交织在一起，极大地扩大了遥感图像的跨域变化。因此，直接将已有方法应用

在跨域遥感图像语义分割场景中效果不一定理想。

基于上述考虑,本章对跨域遥感图像语义分割进行研究。为了最大程度地减少源域和目标域数据偏移对深度语义分割网络学习的不利影响,本章提出了一种带基于多重弱监督约束的跨域语义分割网络方法(Multiple Weakly-supervised Constraints for Cross-domain remote sensing image Semantic Segmentation,记为MWCSS)。该方法有多个弱监督约束构成,其中包括弱监督迁移不变约束(weakly-supervised transfer invariant constraint, WTIC)、弱监督伪标签约束(weakly-supervised pseudo-label constraint, WPLC)和弱监督旋转一致性约束(weakly-supervised rotation consistency constraint, WRCC)。为了充分发挥上述约束的优势,MWCSS采用了一种动态优化策略,在训练过程中动态调整目标函数的约束权值,可以防止深度语义分割网络退化。考虑到跨域遥感图像语义分割任务的特殊性,本章构建了两种实验场景:(1)跨地理位置变化;(2)跨成像模式变化。广泛的实验表明,所提方法MWCSS在这两种实验设置下的性能可以明显优于当前最先进的方法。

3.2 研究方法

为了便于阐明方法,令 S 表示源数据集, T 表示目标数据集。 $S = \{(I_1^S, L_1^S), (I_2^S, L_2^S), \dots, (I_N^S, L_N^S)\}$ 表示包含了 N 张遥感图像和其对应的标签。 $I_i^S \in \mathbb{R}^{H \times W \times C}$ 表示第 i 张影像, $L_i^S \in \{0,1\}^{H \times W \times C}$ 表示相应的标签。其中 H 和 W 分别代表图像的高和宽, C 代表语义分割的类别数。使用 $T = \{I_1^T, I_2^T, \dots, I_M^T\}$ 表示目标域 M 张无标注的图像。此外,图3-1给出了提出的框架的概述。如图3-1所示,本章提出了一种新的带有多重弱监督约束的目标函数用于深度语义分割网络学习优化,其中多个弱监督约束包括WTIC,WPLC和WRCC。更具体地,首先利用DualGAN将源域图像迁移至目标域风格,减少像素级的分布差异,利用迁移后的图像和其原始标签训练语义分割网络,以实现WTIC。通过设置阈值对置信度进行筛选,自适应地选择目标域图像中具有伪标签的锚点进行约束,构建WPLC。基于遥感图像成像角度任意这一特性,刻画了WRCC。为了平衡这些约束,提出了一种动态优化策略。通过在训练过程中,动态调整各项约束的权重,可以有效地避免深度语义分割网络陷入退化情况。在这些弱监督约束下训练后,深度语义分割网络能够在目标数据集上很好地进行预测。

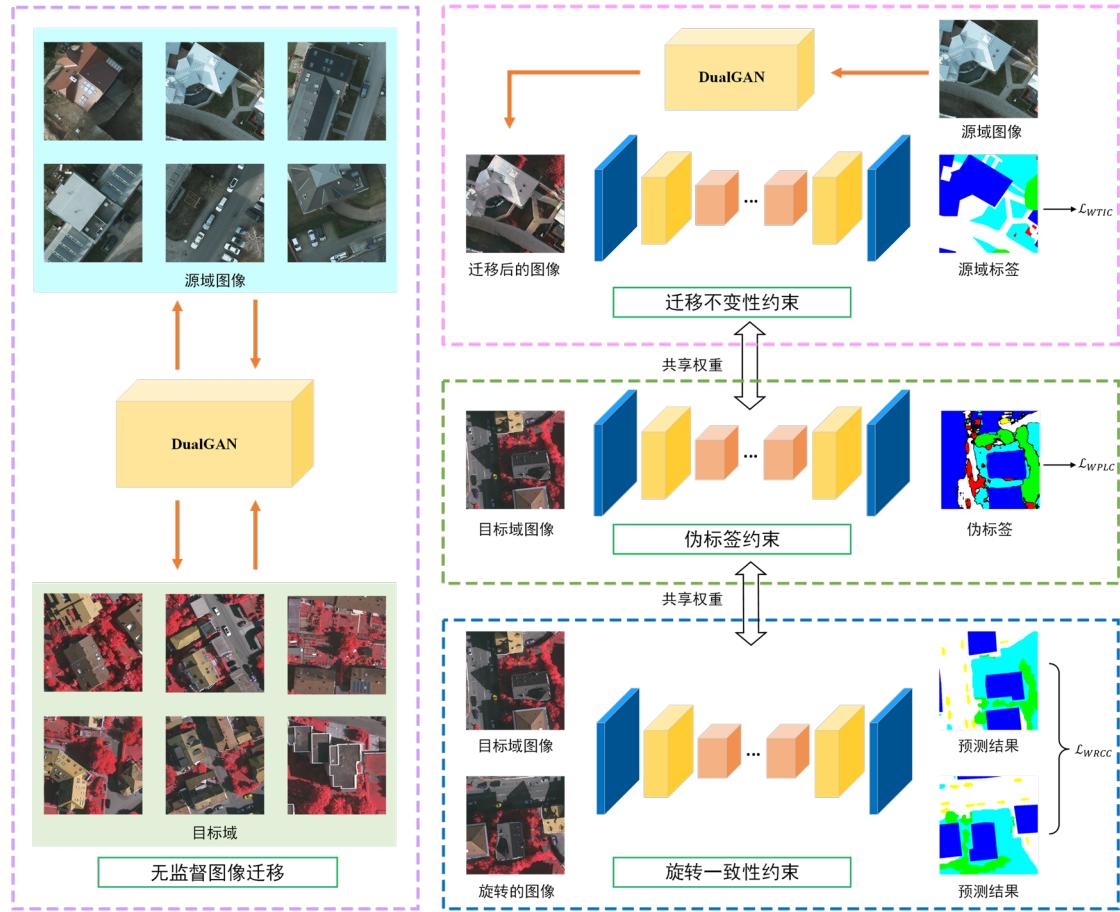


图 3-1 多重弱监督约束下的遥感图像跨域语义分割算法

3.2.1 无监督风格迁移

对于遥感图像跨领域语义分割任务，源领域有大量带标签的数据，而来自目标领域的图像没有任何标签。显然，如何充分挖掘不同领域遥感图像数据的不变语义特征，高效地进行跨领域遥感图像语义分割仍然是一个开放问题。通过挖掘源域和目标域之间的不变语义特征，无监督风格迁移可以将源图像映射到目标域的风格。然后，利用迁移后的具有原始标签的图像训练深度语义分割网络。因此，这种策略自然降低了域偏移的影响。在无监督图像迁移方向，CycleGAN 通过提出循环一致性，来实现非配对图像域间的迁移，取得不错的效果。在同时期，还有 DiscoGAN^[83]和 DualGAN^[84]（对偶生成对抗网络）以类似的思想被提出，来进行无监督下的图像迁移，同样取得了不错的效果。这种无监督的图像到图像的迁移方法，在设计风格变换任务中表现良好，特别是涉及颜色和纹理变化^[33]。与 CycleGAN 和 DiscoGAN 相比，DualGAN 采用了 Wasserstein GAN(WGAN)^[85]使用的一种基于 Wasserstein Distance 有效损失函数，使得优化过程相对更加彻底，有利于生成更接近目标域视觉效果的高质量图像。此外，DualGAN 的网络结构相对简单，容易训练与收敛。综合考虑，这里使用 DualGAN 进行图像迁移。

工作。

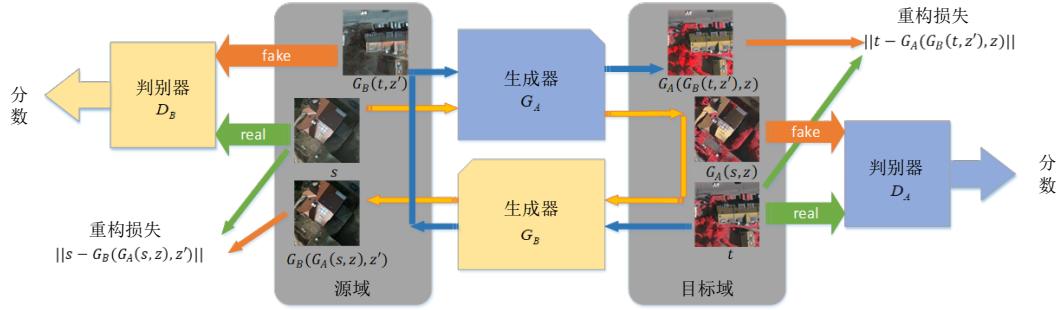


图 3-2 利用 DualGAN 进行图像迁移示意图

DualGAN 的工作流程如图 3-2 所示。对于源域图像 $I^S \in S$ 通过生成器 G_A 将其迁移为目标域风格。然后，利用 D_A 来衡量迁移后的图像 $G_A(I^S, z)$ 接近真实目标域图像的程度。其中 z 为随机噪声，用于进行数据增强。然后通过生成器 G_B 将 $G_A(I^S, z)$ 转换回源域的风格样式，输出为 $G_B(G_A(I^S, z), z')$ ，其为源域图像 $I^S \in S$ 的重构。同样地， $I^T \in T$ 转换到源域风格，记为 $G_B(I^T, z')$ 。然后重构为 $G_A(G_B(I^T, z'), z)$ 。判别器 D_A 以 I^T 为正样本， $G_A(s, z)$ 为负样本进行训练，这意味着判别器给目标域 T 中真实样本 I^T 一个较高的分数，给生成的样本 $G_A(I^S, z)$ 低分。 D_B 也是按照这样方式进行训练的。生成器 G_A 和 G_B 被优化为模拟“假”输出，以混淆对应的鉴别器 D_A 和 D_B ，并最小化重构损失 $\|I^S - G_A(G_B(I^S, z'), z)\|$ 和 $\|I^T - G_B(G_A(I^T, z'), z')\|$ 。生成器与判别器按照上述的方式进行对抗训练。

使用训练好的 DualGAN 模型，将源数据集 S 迁移至目标域的风格。其中，对于图像 $I^S \in S$ ，得到 $I^G = G_A(I^S, z)$ 。将迁移后的图像与原始标签相结合，得到一个与目标域风格相似的数据集，记为 $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \dots, (I_N^G, L_N^S)\}$ 。

3.2.2 多重弱监督约束的目标函数

根据前面的描述，首先使用 DualGAN 将源域图像迁移为目标域的样式。输出为一个新的数据集，记为 $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \dots, (I_N^G, L_N^S)\}$ ，用于训练过程中进行 WTIC。为了进一步挖掘目标域数据的可学习信息，在提出的方法中引入了 WPLC 和 WRCC，其中 WPLC 通过挖掘目标域图像的伪标签信息来实现，WRCC 描述了来自目标域的图像的广义旋转一致性特性。总的来说，所有三个约束都是弱监督的，并且图像标签来自源域在 WTIC 中得到充分的考虑。总损失函数如公式 (3-1) 所示。

$$\mathcal{L}_{Total} = \left(1 - \frac{1}{2}(\alpha + \beta)\right) \mathcal{L}_{WTIC} + \alpha \mathcal{L}_{WPLC} + \beta \mathcal{L}_{WRCC} \quad (3-1)$$

式中 \mathcal{L}_{WTIC} 、 \mathcal{L}_{WPLC} 和 \mathcal{L}_{WRCC} 分别表示弱监督转移不变约束、弱监督伪标签约束和弱监督旋转一致性约束。 α 和 β 是两个重要的超参数，分别表示弱监督伪标签约束

和弱监督旋转一致性约束的权值。在训练的初始阶段，网络模型不够稳定。因此，如果 α 和 β 过高，即使是对标记数据的训练也会受到干扰，并且网络很容易陷入退化的状况，无法获得有意义的语义分割结果。然而，如果 α 和 β 太小，就无法从未标记的数据中获益。考虑到所有这些因素，采用了 Gaussian ramp-up^[86]曲线来动态调整不同约束的贡献。更准确地说， α 和 β 从零开始，沿着曲线 $\exp[-5(1-t)^2]$ 上升，其中 t 首先等于零，然后在每次迭代中线性前进，最终增加到1。因此，第一项 \mathcal{L}_{WTIC} 的权重被设置为 $(1 - \frac{1}{2}(\alpha + \beta))$ ，随着训练过程的进行从1减少到0，其中 $1/2$ 是一个归一化常数，目的是避免权重变成一个没有意义的负值。

(1) 弱监督迁移不变约束

训练 DualGAN 后，源域数据集中的图像将被迁移为接近于目标域风格的图像。迁移后的图像与其原始的源域标签构成一个新的数据集，标记为 $S' = \{(I_1^G, L_1^S), (I_2^G, L_2^S), \dots, (I_N^G, L_N^S)\}$ ，这有利于最大限度地减少不同域之间数据移位的影响。使用迁移后的数据集 S' 来训练一个深度语义分割网络，交叉熵损失函数如公式 (3-2) 所示。

$$\begin{aligned} \mathcal{L}_{WTIC}(S; \theta) \\ = & -\sum_{i=1}^N \ell_{CE}(P_i^G, L_i) \\ = & -\sum_{i=1}^N \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C \left(L_i(h, w, c) \cdot \log(P_i^G(h, w, c)) \right) \end{aligned} \quad (3-2)$$

其中 ℓ_{CE} 表示交叉熵损失函数， $P_i^G = f_\theta(I_i^G)$ 表示概率图， $f_\theta(\cdot)$ 表示一个参数为 θ 语义分割模型。 H 和 W 分别代表图像的高度和宽度， C 为土地覆盖类别的数目。

(2) 弱监督伪标签约束

由于迁移后的图像不能与真实目标域图像完全一致，因此迁移后的图像在该算法框架中主要起到导引作用。Lee 等人^[87]在半监督学习中提出了伪标签学习，并通过大量的实验和分析证明了其简洁有效。因此，考虑在准备阶段使用 S' 上训练的深度语义分割网络为目标域图像生成高置信度的伪标签，以提高后续动态优化阶段深度语义分割网络在目标域的性能。这里使用带有 c 类的 $w \times h$ 图像来说明如何衡量伪标签的置信度。当图像通过最终的 $softmax$ 函数时，将得到一个大小为 $w \times h \times c$ 的矩阵，每个像素都有一个大小为 $1 \times c$ 的向量，向量 top1 值的索引表示类别。计算 top1 和 top2 值之间的差值。如果差值大于阈值 τ_{pse} ，则认为模型对该位置的预测是可靠的，将用作伪标签。否则认为该位置的预测不可信，将其忽略，不参与损失的计算。伪标签的生成过程由可由公式 (3-3) 刻画。经过这一步，就可以得到目标域图像的伪标签。通过将这个过程应用到 T 中的每个图像，即可得到数据集 $T' = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \dots, (I_M^T, L_M^E)\}$ 。

$$\hat{y}_{ij} = \begin{cases} \arg \max_c p_{ij}(c), & \text{top1}(p_{ij}) - \text{top2}(p_{ij}) > \tau_{pse} \\ \text{ignore}, & \text{otherwise} \end{cases} \quad (3-3)$$

在动态优化步骤中，来自目标域的图像及其伪标签参与训练网络，这部分的交叉熵损失函数公式（3-4）所示：

$$\begin{aligned} & \mathcal{L}_{WPLC}(T'; \theta) \\ &= -\sum_{i=1}^M l_{CE}(P_i^T, L_i^E) \\ &= -\sum_{i=1}^M \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C \left(L_i^E(h, w, c) \cdot \log(P_i^T(h, w, c)) \right) \end{aligned} \quad (3-4)$$

其中， l_{CE} 表示交叉熵损失函数， $P_i^T = f_\theta(I_i^T)$ 为概率图， $f_\theta(\cdot)$ 表示参数为的 θ 语义分割网络。 H 和 W 分别代表图像的高度和宽度， C 代表土地覆盖类别数。

（3）弱监督旋转一致性约束

为了挖掘目标域未标记的数据，弱监督旋转一致性约束被引入到算法框架中。本文所做的是像素级分类（即语义分割），这与按场景分类略有不同。只有在变换完全可逆的情况下，才能计算出一致性损失。因此，对目标域无标注图像 $I_i^T \in T$ 采用旋转变换。具体地，旋转变换 φ （随机旋转 90°、180°、270°）作用于图像 I_i^T ，得到 $\tilde{I}_i^T = \varphi(I_i^T)$ 。 I_i^T 和 \tilde{I}_i^T 同时被送入深度语义分割网络，得到两个输出 $P_i^T = f_\theta(I_i^T)$ ， $\tilde{P}_i^T = f_\theta(\tilde{I}_i^T)$ 。不同于场景分类任务，为了计算两个输出的像素级一致性，必须执行逆变换，将每个像素旋转回到原始位置。令随机旋转的逆变换表示为 φ^{-1} 。因此，可以得到逆变换后的输出 $\bar{P}_i^T = \varphi^{-1}(\tilde{P}_i^T)$ ，就可以计算出弱监督的一致性损失。一致性损失项经常使用均方误差（Mean Square Error, MSE），它能促进在不同的随机旋转变换下输出的像素级一致性。损失函数可描述为公式（3-5）所示的形式。

$$\begin{aligned} & \mathcal{L}_{WRCC}(T; \theta) \\ &= \sum_{i=1}^M \ell_{MSE}(P_i^T, \bar{P}_i^T) \\ &= \frac{1}{H \times W} \sum_{i=1}^M \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C \left(\|P_i^T(h, w, c) - \bar{P}_i^T(h, w, c)\|_2 \right) \end{aligned} \quad (3-5)$$

其中 H 、 W 分别表示图像的高度和宽度， C 表示土地覆盖类别数。

3.2.3 优化策略

在联合动态优化之前，需要进行一些准备工作。该方法首先训练 DualGAN 网络，然后将源领域的图像迁移到目标域的风格图像，得到迁移后的数据集 S' ，实现源领域图像和目标领域图像之间像素级分布的对齐。该步骤在一定程度上消除了领域偏移的影响，为深度语义分割网络提供初始训练数据。下一步是让 $\alpha = 0, \beta = 0$ ，并用迁移后的数据集 S' 来训练网络。这有助于模型学习目标数据集的模式，拥有更好的泛化能力。训练后，使用深度语义分割网络为目标域数据集生成伪标签，目标域图像与生成的伪标签构成一个新的数据集，记为 $T' = \{(I_1^T, L_1^E), (I_2^T, L_2^E), \dots, (I_M^T, L_M^E)\}$ 。

在上述准备工作的基础上，最后一步是以端到端的方式，从零开始通过

WTIC、WPLC 以及 WRCC 的约束共同训练深度语义分割网络。优化过程是动态的，可以有效地平衡这些多约束，避免深度语义分割网络陷入退化的情况。通过引入这些约束，深度语义分割网络可以获得更高的性能，特别是对象的边界会更加清晰。最后，语义分割网络适用于目标数据集上的工作。算法 2 总结了深度语义分割网络的整个训练优化算法。

算法 2：动态优化策略

输入： $S = \{(I_1^S, L_1^S), (I_2^S, L_2^S), \dots, (I_N^S, L_N^S)\}; T = \{I_1^T, I_2^T, \dots, I_M^T\}$

输出： 深度语义分割网络参数

1 准备阶段：

2 使用 S 和 T 中的遥感图像训练 DualGAN；
 3 使用训练后的 DualGAN 将 S 迁移成 S' ；
 4 令公式 (3-1) 中 $\alpha = 0, \beta = 0$ ，使用迁移后的数据集 S' 来训练深语义分割网络；
 利用该深度语义分割网络为目标域图像预测伪标签，记为 T' ；

5 动态优化阶段：

6 **for** $epoch = 1:epochs$ **do**
 7 计算基于当前 $epoch$ 的动态权重 α 和 β ；
 8 基于当前的 α 和 β ，根据公式 (3-1) 训练优化深度语义分割网络；
 9 **end for**
 10 **end**

3.3 实验设计

3.3.1 实验数据集

为了充分验证跨域遥感图像语义分割的有效性，采用 ISPRS 2D^[88]语义分割基准数据集中的 Potsdam 和 Vaihingen 数据集进行实验。Potsdam 数据集提供了 3 种不同成像模式的遥感图像，具体为 IR-R-G、R-G-B 以及 R-G-B-IR，这里使用前两种。Vaihingen 只提供了 IR-R-G 这一成像模式下的遥感图像。Potsdam 数据集包含 38 张超高分辨率的正射影像，图像大小为 6000×6000 。Vaihingen 数据集包括 33 张超高分辨率的正射影像，图像大小约为 2000×2000 。Potsdam 与 Vaihingen 包含了相同的 6 个类别，其他、不透水层、汽车、树木、低矮植被和建筑。

具体来说，提供了两种跨领域的实验设置：(1) 地理位置的变化，如图 3-3 (a) 所示。Potsdam IR-R-G 数据集作为源域，Vaihingen IR-R-G 数据集作为目标

域。为了提高计算效率，本文将 Potsdam IR-R-G 数据集及其对应的标签裁剪为 512×512 大小，水平和垂直步幅均为 512 像素，共获得近 4000 幅图像。所有这些图像都参与了 WTIC 模块。对于 Vaihingen IR-R-G 数据集，将图像裁剪为 512×512 大小，水平和垂直步幅均为 256 像素，获得了近 1700 幅图像。所有的图像在训练阶段被用于 WPLC 和 WRCC 模块。此外，在 1700 张图像中，500 张图像（由编号为 2、5、7、8、13、20、22、24 的原始图像裁剪用于验证），以选择最优模型，近 1200 张图像（由剩余图像裁剪）用于测试，以评估算法的性能。

(2) 成像模式的变化，如图 3-3 (b) 所示。Potsdam R-G-B 数据集作为源域，Vaihingen IR-R-G 数据集作为目标域。其余的设置与 (1) 中类似。

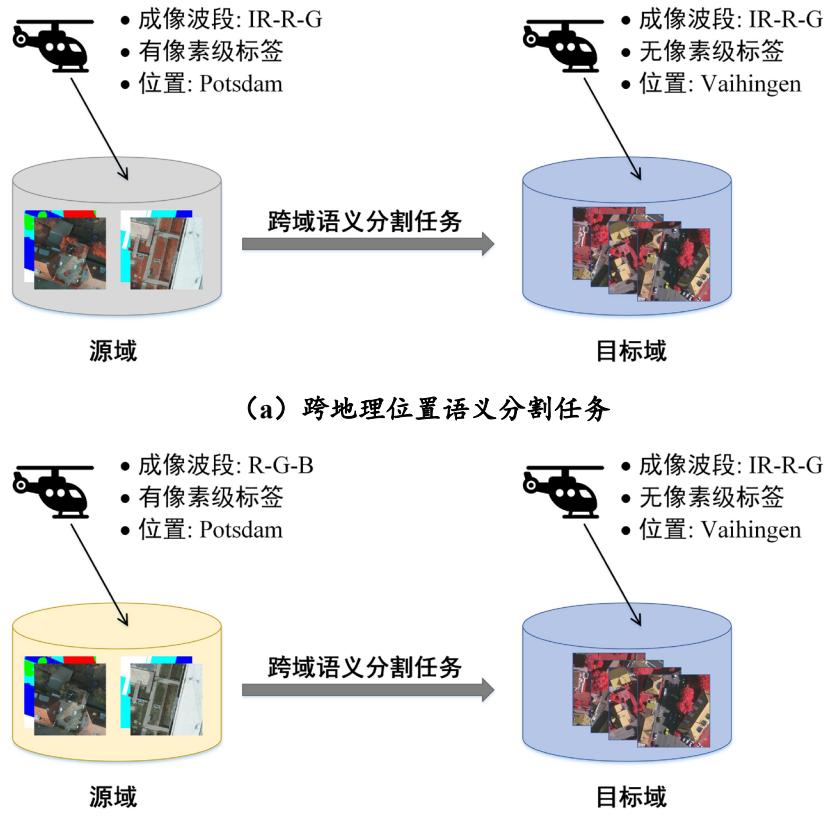


图 3-3 跨域遥感图像语义分割任务设定

3.3.2 实验设置与评价指标

这项工作由 Pytorch 实现，并在单个具有 24 GB RAM 的 Nvidia TITAN RTX GPU 上进行训练。对于 DualGAN 网络。使用 RMSProp 作为优化器，初始学习率设置为 $lr=0.00005$ ，衰减率 $decay=0.9$ ，训练 50 epochs。对于语义分割网络使用 DeepLab v2^[20]作为语义分割模型并以 ResNet-101^[21]作为 backbone。使用 SGD 作为优化器进行训练，学习速率 5×10^{-4} ，权重衰减 5×10^{-4} ，以及指数为 0.9 的多项式衰减，训练 10 epochs。

在评价指标方面，采用了交并比（IoU）来评价每个类别的语义分割结果的准确性。随后，计算所有类别的平均交并比（mIoU）来评价语义分割结果的整体精度。IoU 的计算方法如公式（3-6）所示。

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3-6)$$

其中 A 是真实标签的像素集合， B 是预测像素的集合。 \cap 和 \cup 分别表示交集和并集。 $|\cdot|$ 表示集合中的像素数目。

3.4 实验结果与分析

3.4.1 方法对比

在本小节中，与现有的跨域语义分割算法在两个跨域实验设置下进行了对比实验：（1）跨地理位置；（2）跨成像模式。为了公平比较，所有的算法都通过 Vaihingen IR-R-G 的验证集选择最佳模型，并在 Vaihingen IR-R-G 的测试集上进行评估。

（1）跨地理位置

为了证实提出的方法对主要由区域变化引起的域偏移的有效性，使用 Potsdam IR-R-G 数据集作为源域，Vaihingen IR-R-G 数据集作为目标域。经过 DualGAN 迁移后的图像，如图 3-4（a）所示。可以看出，在经过迁移后，确实减小了源域与目标域的视觉差异。跨域语义分割的可视化结果，如图 3-5 所示。其中 Source Only 是一个基线模型，即只利用源域进行训练，未采取任何领域自适应措施，直接在目标域上直接测试得到的结果。可以发现 Source Only 模型存在着严重的域偏移问题。它的语义分割结果通常表现为嘈杂的噪声以及错误的上下文。经过不同领域自适应算法的调整，域偏移问题在很大程度上得到了缓解。与其他方法相比，本章提出的方法产生了更好的分割结果。通过图 3-5 可以看出对于其他方法，很容易将低矮植被与树木进行混淆，产生错误的语义分割结果。而本章提出的方法在学习域不变特征的同时也保证了类别的可分性，相较于其他方法能更好地区分低矮植被与树木。此外，在一些小目标类别的物体，如“汽车”上 MUCSS 也有着更好的表现。

跨域语义分割的定量结果列在表 3-1 中，其中 Source Only 的 mIoU 值为 31.04%，表现出明显的域偏移问题。在经过不同的领域适应方法处理后，语义分割性能均得到了一定程度的提高。但可以发现有的方法在某些类别上发生了“负迁移”的现象。例如，Benjdir's、SEANet 以及 MRNet 在“其他”这个类别上的 IoU 值小于 Source Only 方法，说明出现了负迁移。本章提出的模型实现了 45.38% 的 mIoU，相比基线提高了近 14%。与其他方法相比，本章提出的模型仍然具有

更高的性能，在各个类别上也都具有良好的表现，这表明所提出的方法更有利于消除域偏移带来的负面影响。

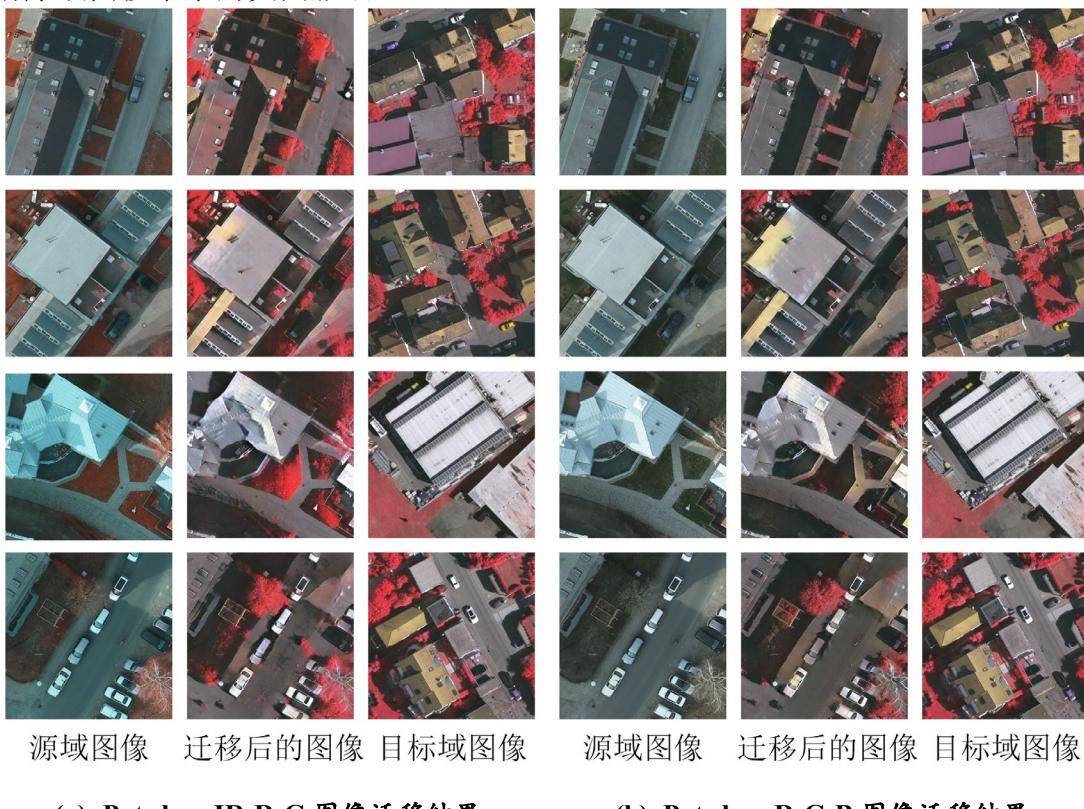


图 3-4 经 DualGAN 迁移后的图像

表 3-1 跨地理位置 Potsdam IR-R-G → Vaihingen IR-R-G 跨域语义分割结果

方法	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
Source Only	5.71	35.84	20.27	54.95	17.88	51.59	31.04
Benjdir's ^[89]	2.12	39.88	8.20	26.56	26.53	40.97	24.04
SEANet ^[44]	0.80	42.24	24.41	53.26	35.54	58.21	35.75
AdaptSegNet ^[41]	4.60	54.39	6.40	52.65	28.98	63.14	35.02
CsDA ^[90]	9.85	46.22	31.14	52.04	31.11	52.39	37.12
MRNet ^[91]	3.81	55.02	34.37	54.79	26.46	76.39	41.80
CLAN ^[92]	9.89	58.23	37.25	59.10	36.74	59.10	43.38
MWCSS	29.66	49.41	34.34	57.66	38.87	62.30	45.38

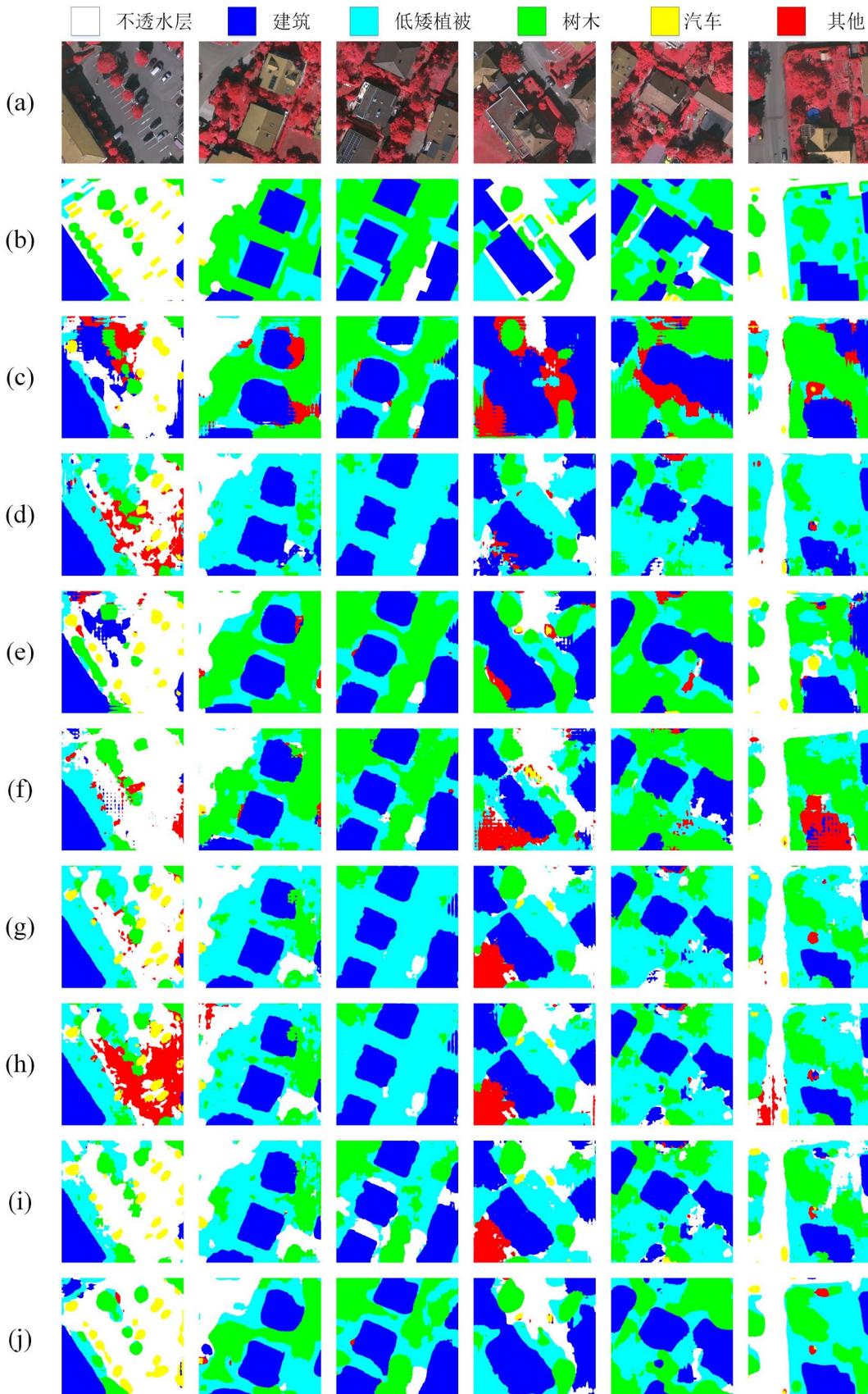


图 3-5 跨地理位置 Potsdam IR-R-G → Vaihingen IR-R-G 语义分割结果可视化

说明: (a) 目标域图像 (b) 目标域标签 (c) Source Only (d) Benjdir's (e)
SEANet (f) AdaptSegNet (g) CsDA (h) MRNet (i) CLAN (j) 本章方法

(2) 跨成像模式

在该实验设定条件下, Potsdam R-G-B 数据集作为源域, Vaihingen IR-R-G 数据集作为目标域, 以评估所提出的方法对成像模式变化引起的域偏移的有效性。经过 DualGAN 迁移后的图像, 如图 3-4 (b) 所示。可以观察出, 在经过迁移后, 确实减小了源域与目标域的视觉差异。语义分割可视化结果见图 3-6。同样, Source Only 方法严重受到域偏移问题的影响。经过各种域适应语义分割算法调整, 每个类别的分割性能都得到了明显的改善。与其他方法相比, 本章所提出的方法可以保留更多的边缘细节, 这从定性的角度证明了所提方法的优越性。

表 3-2 描述了本章方法和其他竞争方法的定量结果。Source Only 方法的性能非常差, mIoU 为 24.01%。对于“低矮植被”这一类别, mIoU 下降的程度很大。通过观察两个数据集的原始数据, 发现由于这两个数据集使用的成像模式不同, 导致视觉效果差异很大, “低矮植被”的颜色在 Potsdam R-G-B 数据集中呈现出绿色的视觉效果, 而在 Vaihingen IR-R-G 数据集中呈现为红色。这些差异直接导致了 Source Only 方法性能的急剧下降。本章的方法的性能优于其他方法, mIoU 值可以达到 39.93。上述实验结果证明了本章方法的优越性和有效性。

综上所述, 本章提出的方法在处理主要由地理位置和成像模式变化引起的两个领域的域偏移方面具有良好的性能。提出的方法显示了较好的鲁棒性和领域泛化性。

表 3-2 跨成像模式 Potsdam R-G-B → Vaihingen IR-R-G 跨域语义分割结果

方法	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
Source Only	1.76	26.86	16.70	44.48	12.56	41.67	24.01
Benjdir's ^[89]	4.48	31.78	21.70	41.76	23.67	52.36	29.31
SEANet ^[44]	0.80	42.24	24.41	53.26	35.54	58.21	35.75
AdaptSegNet ^[41]	1.08	50.05	14.18	56.45	20.73	62.61	34.18
CsDA ^[90]	0.55	44.82	23.81	52.04	20.74	53.39	32.56
MRNet ^[91]	0.81	54.11	29.39	54.99	16.16	75.39	38.47
CLAN ^[92]	0.84	57.30	17.28	59.10	24.94	59.19	36.44
MWCSS	3.94	46.19	40.31	55.82	27.85	65.44	39.93

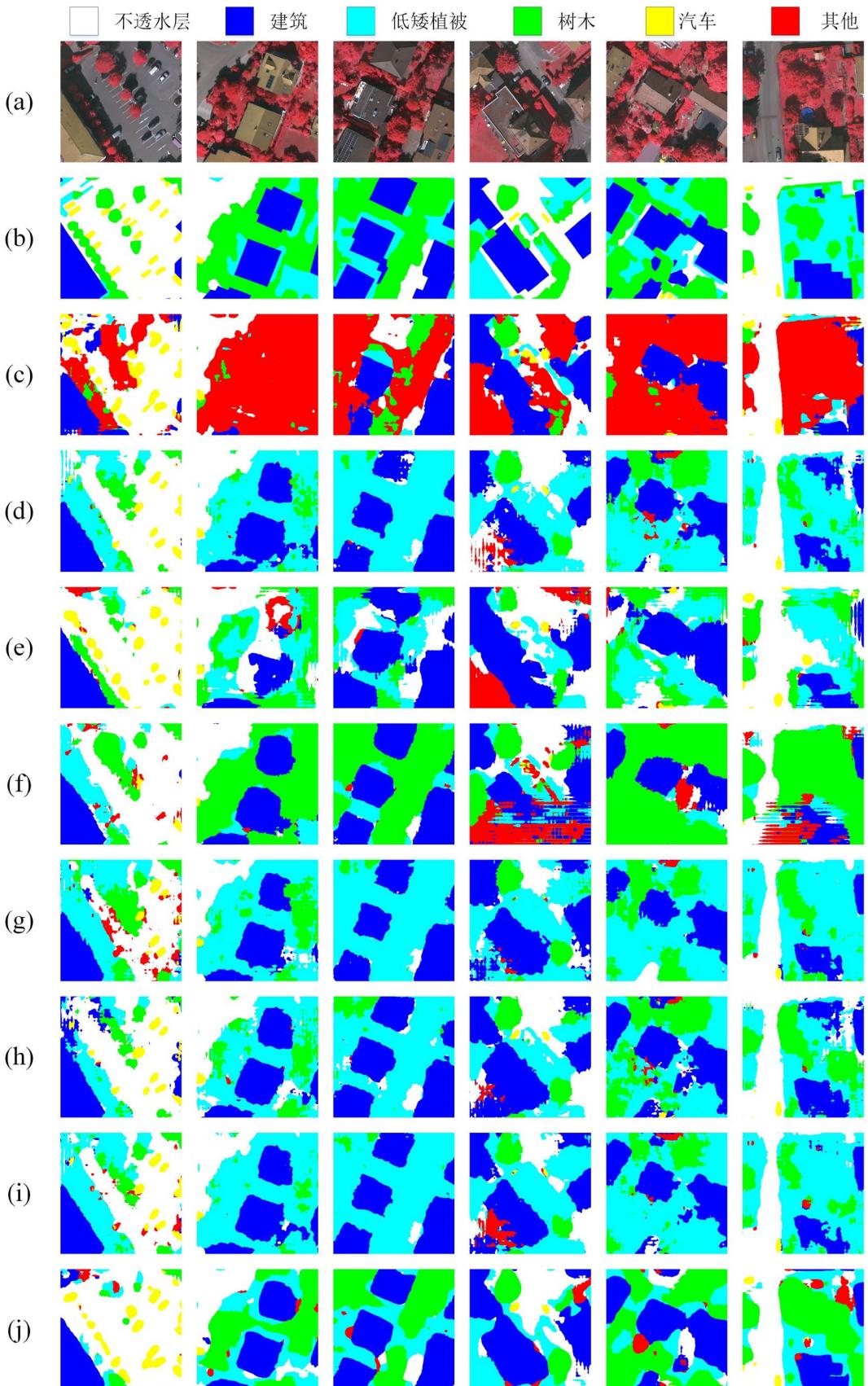


图 3-6 跨成像模式 Potsdam R-G-B → Vaihingen IR-R-G 语义分割结果可视化

说明: (a) 目标域图像 (b) 目标域标签 (c) Source Only (d) Benjdir's (e) SEANet (f) AdaptSegNet (g) CsDA (h) MRNet (i) CLAN (j) 本章方法

3.4.2 消融实验

该部分验证了 WPLC 和 WRCC 模块的有效性。具体地，四种类型的设置 ($\alpha = 0, \beta \neq 0$; $\alpha \neq 0, \beta = 0$; $\alpha = 0, \beta = 0$; $\alpha \neq 0, \beta \neq 0$)。例如， $\alpha = 0, \beta \neq 0$ 表示弱监督迁移不变约束和弱监督旋转一致性约束的结合，即仅使用 WTIC 和 WRCC 模块训练深度语义分割网络。其他三个设置与此类似。实验结果如表 3-3 和表 3-4 所示，WPLC 和 WRCC 模块有助于提高深度语义分割网络的性能。特别是当 WTIC、WPLC 和 WRCC 这三个约束条件协同工作时，它们可以相互受益，获得良好的性能。

表 3-3 Potsdam IR-R-G → Vaihingen IR-R-G 消融实验

损失项	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
$\alpha = 0, \beta \neq 0$	0.63	43.82	39.50	57.14	33.00	61.41	39.25
$\alpha \neq 0, \beta = 0$	0.80	42.24	24.41	53.26	35.54	58.21	35.75
$\alpha = 0, \beta = 0$	10.46	40.30	30.95	56.95	36.00	59.16	38.97
$\alpha \neq 0, \beta \neq 0$	34.82	42.34	28.65	58.45	43.15	59.75	44.53

表 3-4 Potsdam R-G-B → Vaihingen IR-R-G 消融实验

损失项	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
$\alpha = 0, \beta \neq 0$	0.70	40.32	39.19	50.73	31.11	60.93	37.16
$\alpha \neq 0, \beta = 0$	4.71	39.67	31.79	52.73	25.75	60.28	35.82
$\alpha = 0, \beta = 0$	0.69	35.73	36.45	50.86	27.36	58.83	34.99
$\alpha \neq 0, \beta \neq 0$	0.62	42.83	40.70	57.15	31.24	62.63	39.19

3.4.3 超参数分析

在 WPLC 约束中，阈值 τ_{pse} 用于指导伪标签的生成过程。为了分析该框架对 τ_{pse} 的敏感性，分别对 Potsdam IR-R-G 到 Vaihingen IR-R-G 和 Potsdam R-G-B 到 Vaihingen IR-R-G 的跨域语义分割任务在不同 τ_{pse} 下的表现进行了评估。

对于伪标签阈值 τ_{pse} ，一般来说，高阈值 τ_{pse} 不能产生良好的性能，因为它会忽略图像中的更多区域。因此，参与 WPLC 过程的像素较少，导致性能较差。为了获得最佳的阈值 τ_{pse} ，在验证集上进行了一个完整的实验过程。两种任务的不同阈值 τ_{pse} 的定量结果见表 3-5 和表 3-6，语义分割可视化结果见图 3-7 和图 3-8。对于从 Potsdam IR-R-G 到 Vaihingen IR-R-G 的跨域语义分割任务，在阈值

τ_{pse} 设置为 0.4 的条件下，在验证集上的 mIoU 为 44.53%，性能最佳。对于另一项任务，可以得出在阈值 τ_{pse} 值为 0.5 下，在验证集上的 mIoU 为 39.19%。

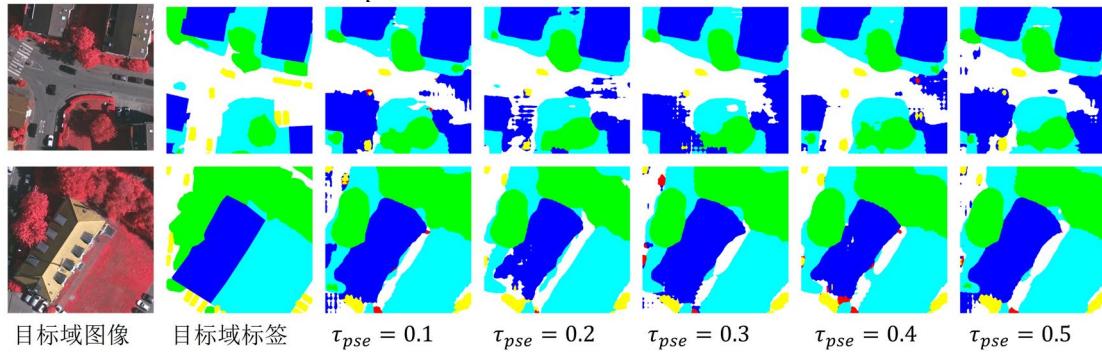


图 3-7 Potsdam IR-R-G → Vaihingen IR-R-G 不同伪标签阈值 τ_{pse} 下语义分割结果可视化

表 3-5 Potsdam IR-R-G → Vaihingen IR-R-G 伪标签阈值 τ_{pse} 参数分析

τ_{pse}	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
0.1	19.30	45.80	33.51	53.85	39.39	60.78	42.11
0.2	21.03	41.44	33.78	58.06	42.38	59.75	42.74
0.3	21.70	49.07	33.04	54.69	43.08	63.95	44.26
0.4	34.82	42.34	28.65	58.45	43.15	59.75	44.53
0.5	37.50	37.59	32.54	56.84	40.62	57.66	43.79

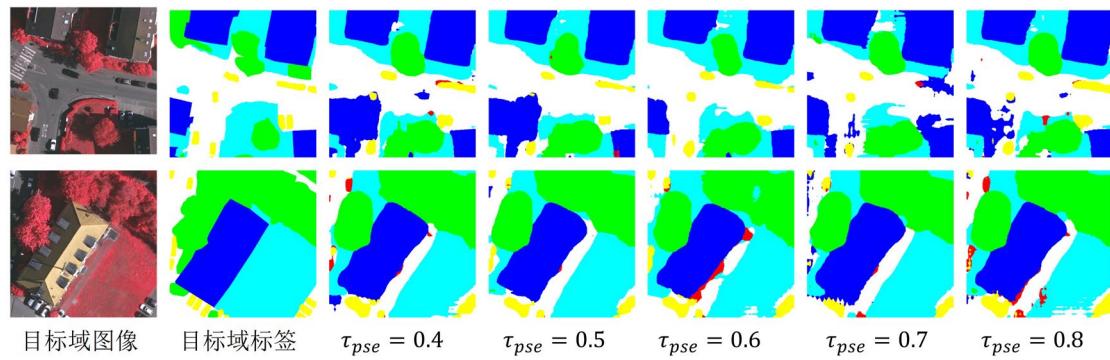


图 3-8 Potsdam IR-R-G → Vaihingen IR-R-G 不同伪标签阈值 τ_{pse} 下语义分割结果可视化

表 3-6 Potsdam R-G-B → Vaihingen IR-R-G 伪标签阈值 τ_{pse} 参数分析

τ_{pse}	IoU						mIoU
	其他	不透水层	汽车	树木	低矮植被	建筑	
0.4	0.75	38.45	39.86	53.69	29.36	61.00	37.19
0.5	0.62	42.83	40.70	57.15	31.24	62.63	39.20
0.6	0.63	39.44	39.05	52.00	31.10	31.94	32.36
0.7	0.52	45.66	36.34	57.70	30.57	60.59	38.56
0.8	0.70	39.51	41.15	55.03	27.84	62.02	37.71

3.5 本章小结

由于深度语义分割网络出色的特征提取能力，其已被广泛应用于遥感图像语义分割中，并取得了巨大的成功。然而，深度语义分割网络的优越性高度依赖于大量的标记训练数据，并且要求测试数据和训练数据的分布是相同的。在实际应用中，很难满足这些条件。因此，如何充分挖掘不同领域的遥感图像数据的不变语义特征，高效地进行跨领域的遥感图像语义分割，引起了人们的关注。在这一强烈需求的驱动下，本章提出了一种具有多个弱监督约束的新型目标函数，用于学习跨域遥感图像语义分割的深度语义分割网络，其中多个弱监督约束包括 WTIC、WPLC 和 WRCC。与基于领域适应性来学习不同领域之间的不变特征的方法不同，所提出的方法直接学习将图像从源头映射到目标，并保持生成的图像内容与原始图像相似。更具体地说，DualGAN 被用于在源域和目标域之间进行无监督的风格迁移，以进行 WTIC。为了平衡这些多重约束，优化过程是动态进行的，这可以有效地避免深度语义分割网络陷入退化的情况。在多种约束条件下进行训练后，深度语义分割网络可以在目标数据集上表现良好。为了验证本章提出的方法，使用了两个跨领域的实验设置：（1）地理位置变化和（2）成像模式变化。在上述两个典型的跨域设置下进行的大量实验表明，本章提出方法的表现明显优于其他代表性算法。

第4章 基于语义保持生成对抗网络的域适应遥感图像语义分割

4.1 概述

在上一章中通过 DualGAN 将源域图像映射到目标域的风格，在一定程度上减弱了域偏移问题的影响。虽然这在一定程度上缩小了源域和目标域之间的视觉差距，但如何追求语义不变的内容仍然是一个挑战。众所周知，理想状态下的图像迁移是保持语义内容不变，并使风格与目标域高度相似。然而，在迁移后的图像中经常出现不平滑、不协调的区域，如图 4-1 所示的黄色框所示区域。本文将迁移后的图像中的这种现象定义为偏差。基于 GAN 的图像迁移由于没有充分考虑源域的语义标签，不可避免地会在迁移后的图像中产生偏差。在特征层面，偏差意味着生成器无法彻底地将同一类的特征聚在一起，即在迁移过程中，有些类可能会与其他类混淆。在图像层面，偏差表现为迁移后图像中的不平滑和不协调区域，这将导致迁移后图像的内容与原始语义标签不一致。毫无疑问，这种现象会影响接下来的语义分割训练过程。之前的一些工作^[34,39]尝试使用源预训练分割模型来计算源图像预测和迁移图像预测的语义一致性损失，从而达到语义保持的源到目标迁移的目的。但这种方式存在一定的不足：一方面，这种约束高度依赖预训练的模型，因此不能以端到端方式实现；另一方面，一旦预先训练的模型是固定的，并且没有来自分割模型的反馈，导致有限的性能增益。

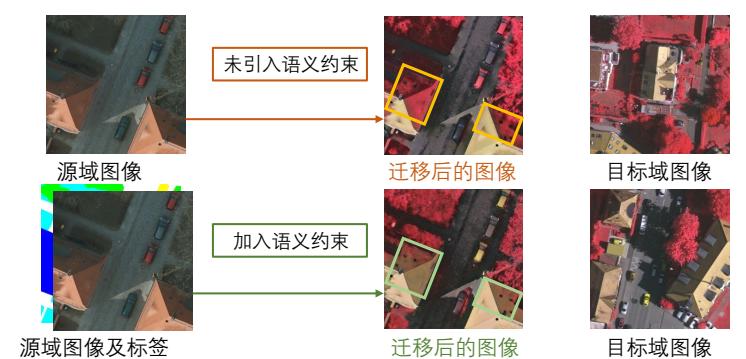


图 4-1 图像迁移中的偏差现象

说明：通过对有无语义约束的图像迁移的实例分析，揭示了语义约束存在的必要性。其中，源域图像为 R-G-B 成像模式，而目标域图像为 IR-R-G 成像模式

对于遥感图像跨域语义分割任务而言，域偏移的主要挑战可以概括为：(1) 跨地理位置：在此情况下，源图像与目标图像处于相同的成像模式，如图 4-2 (a) 所示，但由于亮度、时间差异和地理位置等原因，光谱信息存在或多或少的差异。

例如对于 Potsdam IR-R-G 其各个光谱均值分别约为 97, 92, 86。而 Vaihingen IR-R-G 各个光谱均值约为 120, 81, 80, 存在较明显差异。从给出的图像样本示意图也能直观地感受到两个数据集的外观差异。这将导致不同的光谱分布，从而导致域偏移问题。(2)跨成像模式：在这种情况下，源和目标成像模式是多样化的，如图 4-2 (b) 所示。例如，源图像以 R-G-B 成像模式被拍摄，而目标域图像的成像模式为 IR-R-G 甚至更为不同。在外观方面，源域和目标域的图像风格可以有很大差异。在分布上，源和目标的谱统计完全不同。这将导致经源域学习的语义分割模型应用到目标域时失效。(3)跨景观布局：这通常发生在城乡数据集域适应之间，如图 4-2 (c) 所示。特别是对于城市和农村地区，土地覆盖在景观布局和对象风格上的表现完全不同。例如，城市的建筑和道路占比明显高于农村。而农村的林地和农田占比又明显高于城市。此外，城市地区的建筑排列整齐，形状各异，而农村地区的建筑则杂乱无章，形状更简单。城市的道路是宽阔的。相比之下，乡村的道路是狭窄的。城乡景观布局不一致，进一步增加了模型泛化的难度。

基于上述域偏移问题的分析和 GAN 的局限性，提出了一种新的语义保持生成对抗网络（Semantic-Preserved Generative Adversarial Network, SPGAN），可以进行无偏（即语义保持）的源到目标图像迁移，并进一步实现光谱信息对齐甚至成像模式。这是通过在基于 GAN 的图像迁移模型中引入表征不变约束和语义保持约束来实现的。其次，数据集之间的景观布局存在一定的差异，尤其是在城市和农村的领域偏移问题中。因此，在提出的方法中引入半监督学习中的类别混合操作—ClassMix 操作^[93]，实现类别分布对齐。具体地，在模型输入层面使用 ClassMix 操作，通过随机采样的方式从迁移后的图像上随机选择一定的类别，将相应的图像块粘贴到目标域图像上，构成混合图像。将迁移后的图像与混合图像一同送入语义分割网络进行训练。在模型输出层面，提出了边界增强损失，来细化语义分割网络在目标边界的性能。上述模块耦合在一起，形成一个统一的框架，称为 SPGAN-DA (SPGAN-Domain adaptation)。本章的主要贡献总结如下：

(1) 提出了一种新型的 SPGAN，它从源域到目标域进行无偏迁移（即视觉内容不变迁移），以对齐光谱信息或成像模式。这是通过在 GAN 框架中引入表示不变和语义保留约束来实现的，并以端到端方式进行优化。

(2) 提出了一种新的类分布对齐语义分割模块，进一步缩小了不同数据集之间的景观布局差距。在模型输入级别，首先通过 ClassMix 操作将迁移后的图像中的部分物体粘贴到目标域图像上，形成混合图像。将迁移后的图像与混合图像一同送入语义分割网络进行训练。在模型输出层面，提出了边界增强来细化物体边界的分割效果。通过这两个层面协同训练领域自适应语义分割模型。

(3) 所提出的 SPGAN-DA 在经典的遥感跨域语义分割基准上始终表现良好。大量的实验结果证明了提出的 SPGAN-DA 框架的有效性和通用性。

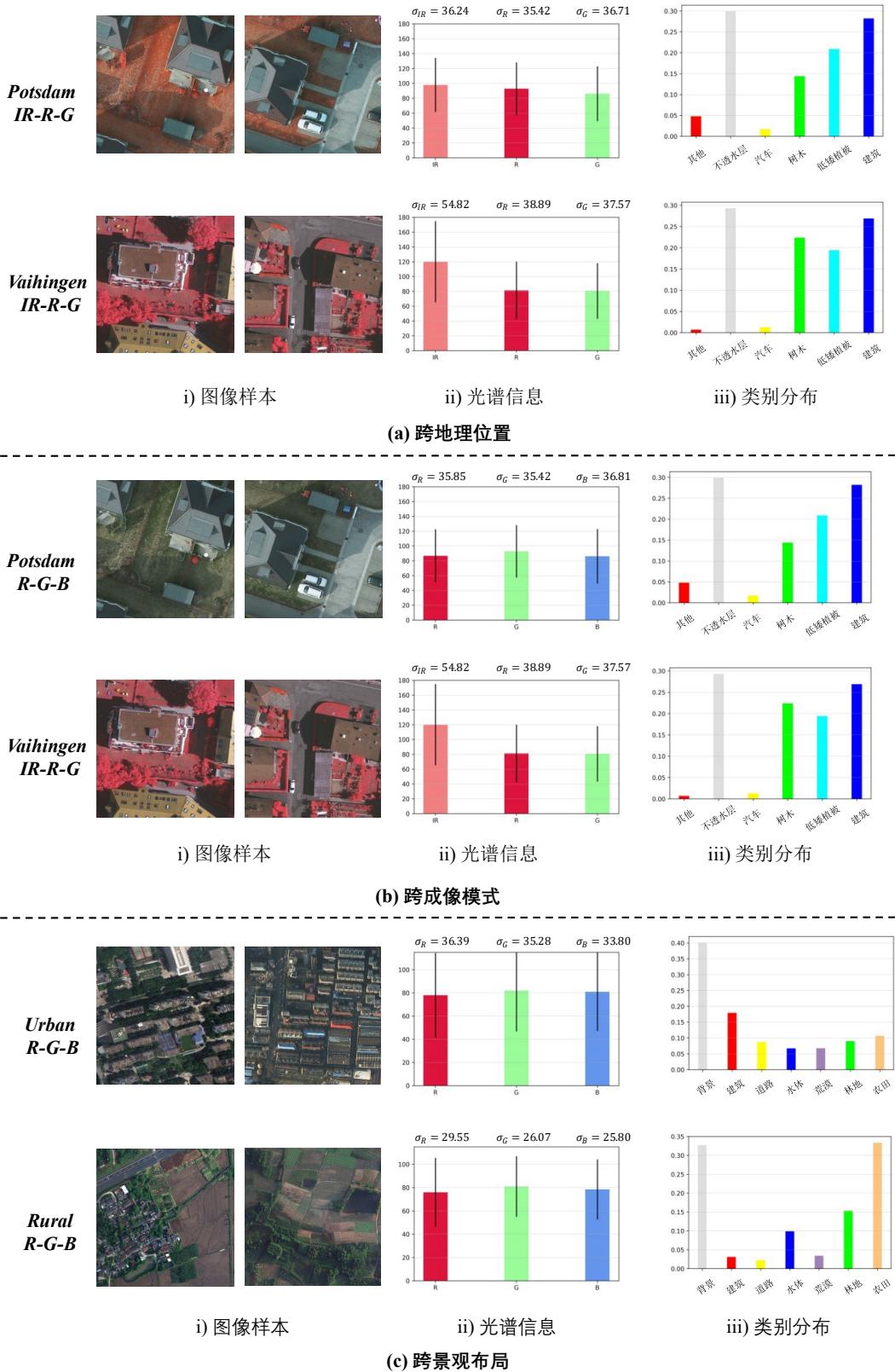


图 4-2 三种典型的跨域遥感图像语义分割任务

说明: (a) 跨地理位置 (b) 跨成像模式 (c) 跨景观布局

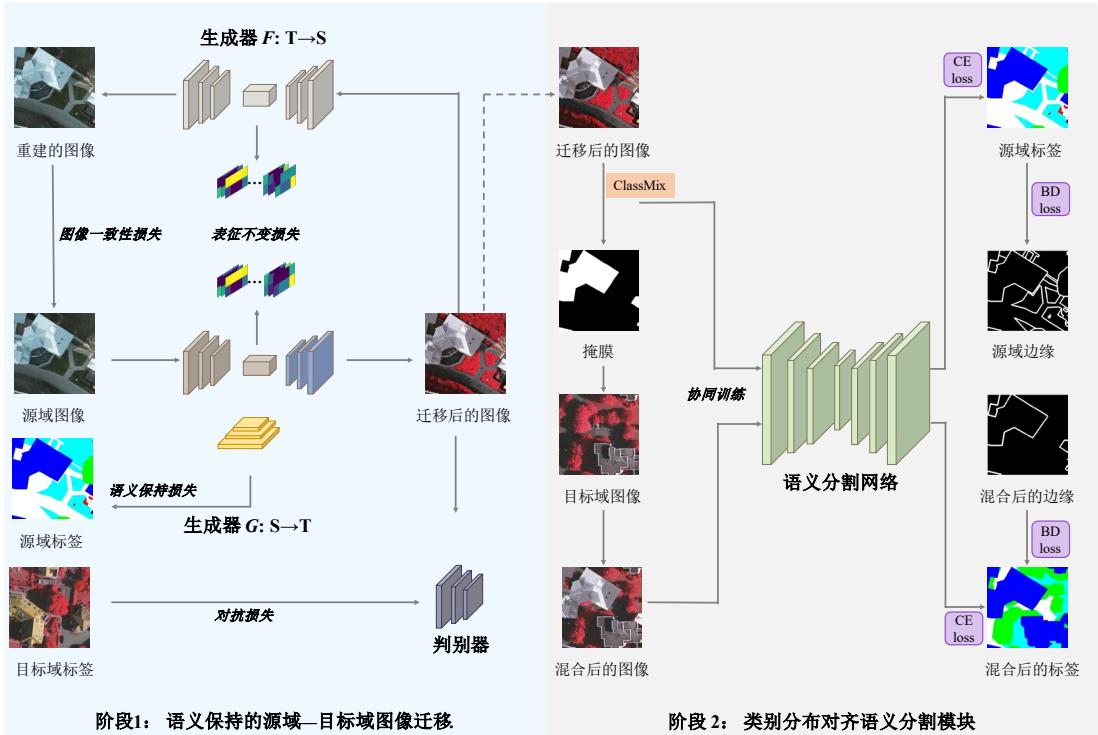


图 4-3 提出的 SPGAN-DA 框架

说明：首先，通过提出的 SPGAN 进行语义保持的源域到目标图像迁移。随后，将迁移后的图像和混合后的图像一起输入语义分割网络进行协同训练

4.2 研究方法

在源域中，有 N_s 图像和对应的像素级标签，记为 $I_s = \{X_s, Y_s\}$ 。图像样本 $x_s^i \in \mathbb{R}^{N \times H \times W}$ ，标签样本 $y_s^i \in \{0,1\}^{C \times H \times W}$ ， H 和 W 分别为高和宽， N 表示图像样本的通道数， C 表示语义分割的类别数目， $i = 1 \dots N_s$ 。对于目标域，只有 N_t 个未标记的图像可用，记为 $I_T = \{X_T\}$ 。目标是训练一个分割模型来为 I_T 预测准确的对应标签。本章提出的框架 SPGAN-DA 如图 4-3 所示。可以看出，框架由两个阶段组成。首先，通过使用提出的 SPGAN 进行图像迁移，将源域图像映射到目标域的风格。其次，利用迁移后的图像和目标域图像协同训练类别对齐语义分割模块。

4.2.1 语义保持的生成对抗网络

阶段一首先执行语义保持的源到目标图像迁移，以减少源和目标域之间的像素级差异，这是由提出的 SPGAN 完成的。该方法的目的是将源域图像映射为目标域图像，减小像素级层面的分布差异。如图 4-3 所示，SPGAN 有两个方向映射的生成器分别为 $G: S \rightarrow T$ 和 $F: T \rightarrow S$ ，两个对抗性鉴别器 D_S 和 D_T 。值得注意的是，提出的 G 不同于传统的 G ，它由三个部分组成， $G = \{G_{enc}, G_{dec}, G_{seg}\}$ ， F 与以往的生成器一样 $F = \{F_{enc}, F_{dec}\}$ 。其中， $G_{img} = G_{enc} \circ G_{dec}$ 适用于产生目标域风

格的图像，并“骗过”判别器 D_T 。 $G_{sem} = G_{enc} \circ G_{seg}$ 用于以有监督的方式进行语义分割，从而达到语义信息保持的目的。下面将按照各个损失模块进行介绍。

对抗损失（Adversarial Loss, AL）表示为公式（4-1）所示

$$\begin{aligned}\mathcal{L}_{GAN}(G, F, D_S, D_T, I_S, I_T) \\ = E_{x_t \sim X_T} [\log D_T(x_t)] \\ + E_{x_s \sim X_S} [\log (1 - D_T(G_{img}(x_s)))] \\ + E_{x_s \sim X_S} [\log D_S(x_s)] \\ + E_{x_t \sim X_T} [\log (1 - D_S(F(x_t)))]\end{aligned}\quad (4-1)$$

图像的一致性损失（Image Consistency Loss, ICL）：为了实现非配对图像间的迁移，这里同样采用了图像一致性约束^[33, 83, 84]。要求将一个源样本从源映射到目标，再映射回源，重构原始样本。通过计算重构样本与原始样本的差异实现图像一致性约束。

$$\begin{aligned}\mathcal{L}_{img}(G, F, I_S, I_T) \\ = E_{x_s \sim X_S} [\|F(G_{img}(x_s)) - x_s\|_1] \\ + E_{x_t \sim X_T} [\|G_{img}(F(x_t)) - x_t\|_1]\end{aligned}\quad (4-2)$$

其中 $\|\cdot\|_1$ 表示 ℓ_1 范数。

现有的基于GAN方法的“源域—目标域”图像迁移方法大多只考虑上述两种损失，从而产生偏差。这一现象可以用 $\lim_{\theta_G} |C(x_s) - C(G_{img}(x_s))| = b$ 进行形式化描述，其中 θ_G 表示生成器网络的参数， $C(\cdot)$ 表示抽象内容提取函数， G_{img} 表示源到目标生成器。

表征不变损失（Representation-Invariant Loss, RIL）：之前的工作只考虑了图像级的一致性，这里进一步考虑了表征不变信息，因为表征层包含更多高频和抽象的信息。通过观察发现，当两个相反方向的生成器网络的中间表征层被强制服从相似的分布。最终，生成器生成的图像在分布上更接近目标，并且细节将得到很好的保留。形式上，要使 $G_{enc}(x_s) \approx F_{enc}(G_{img}(x_s))$ 和 $F_{enc}(x_t) \approx G_{enc}(F(x_t))$ 同时满足。因此，表征不变损失可以表示为公式（4-3）所示。

$$\begin{aligned}\mathcal{L}_{rep}(G, F, I_S, I_T) \\ = E_{x_s \sim X_S} [\|G_{enc}(x_s) - F_{enc}(G_{img}(x_s))\|_1] \\ + E_{x_t \sim X_T} [\|F_{enc}(x_t) - G_{enc}(F(x_t))\|_1]\end{aligned}\quad (4-3)$$

语义保持损失（Semantic-Preserved Loss, SPL）：最后，旨在生成目标域风格化的图像，其中语义内容得到完好保留。一旦迁移后的图像发生语义内容的改变，那么和其原始标签将发生像素级的错位，这将影响后续的语义分割性能。已有学

者^[34,39]尝试使用源预训练的分割模型，并固定其权重，计算源图像预测和迁移后图像预测的语义一致性损失，从而达到语义保持的“源域—目标域”迁移的目的。然而，这种约束高度依赖于预训练的模型，因此不能以端到端方式实现，从而导致性能提升有限。相比之下，本方法在源域不需要任何预先训练的模型，并且语义约束可以在训练过程中以一种优雅而有效的方式实现。回顾前文的描述，所提出的生成器 G 由三部分组成， $G = \{G_{enc}, G_{dec}, G_{seg}\}$ ，其中 $G_{img} = G_{enc} \circ G_{dec}$ 用于进行“源域—目标域”方向的图像迁移。 $G_{sem} = G_{enc} \circ G_{seg}$ 则以有监督的方式促进生成器 G 在图像迁移的过程中保持语义信息，这被视为一个辅助任务。由于辅任务与主任务存在高度的语义信息耦合，这已被证明有利于图像迁移这个主任务^[94-96]。通过这种语义约束，网络可以从中受益，并将保留图像中对象的语义信息而不失真。将语义保持损失定义为交叉语义损失：

$$\mathcal{L}_{sem}(G, I_S) = E_{(x_s, y_s) \sim (X_S, Y_S)}[\ell(G_{sem}(x_s), y_s)] \quad (4-4)$$

其中 $\ell(\cdot)$ 表示常用的交叉熵损失函数， y_s 表示源域的标签， $G_{sem}(x_s)$ 表示预测的概率值。

根据上述损失项，SPGAN的整体损失函数为：

$$\begin{aligned} \mathcal{L}_{SPGAN}(G, F, D_S, D_T, I_S, I_T) \\ = \mathcal{L}_{GAN}(G, F, D_S, D_T, I_S, I_T) \\ + \lambda_1 \mathcal{L}_{img}(G, F, I_S, I_T) \\ + \lambda_2 \mathcal{L}_{rep}(G, F, I_S, I_T) \\ + \lambda_3 \mathcal{L}_{sem}(G, I_S) \end{aligned} \quad (4-5)$$

其中 λ_1 的值通常设置在[10, 20]这一区间范围内， λ_2 通常被设置为1，而 λ_3 的数值随着训练时间的增加而逐渐增大。

通过在基于GAN的迁移模型中引入表示不变和语义保持约束，可以将偏差现象最小化，即 $\lim_{\theta_G \rightarrow \theta^*} b = 0$ ，其中 θ_G 表示生成器的网络参数， θ^* 表示生成器网络的最佳参数。

经过良好的SPGAN训练后，利用SPGAN进行语义保留的源域—目标域图像迁移，迁移后的图像与其原始源标签构成一个新的数据集，记为 $I_S' = \{X_S', Y_S\}$ 。

4.2.2 类别分布对齐语义分割模块

为了进一步缩小迁移图像与目标图像之间的差距，提出了多级细化语义分割模块（Class Distribution Alignment, CDA）。具体来说，多级细化语义分割由两个方面组成。一方面，通过在模型输入阶段引入ClassMix操作来减小类别分布差异；另一方面，提出了边界增强损失，用以细化物体边界的分割效果。

首先，对于迁移后的数据集 $I_S' = \{X_S', Y_S\}$ ，会直接参与语义分割网络的训练，

该部分损失如公式 (4-6) 所示:

$$\mathcal{L}_{src_seg} = \mathcal{L}_{ce}(X_s'^{(h,w)}, Y_s^{(h,w)}) \quad (4-6)$$

下面将介绍 ClassMix 的过程。形式上，给定一个有标记的迁移后的图像 x'_s 和目标域未标记的目标图像 x_t ，使用将 M_{mask} 表示为中随机选中类别的位置掩膜。其中 $M_{mask}^{(h,w)} = 1$ 表示位于第 h 行，第 w 列的像素点属于所选类。反之， $M_{mask}^{(h,w)} = 0$ 则表示位于第 h 行，第 w 列的像素点不属于所选类。混合图像可以形成如下：

$$x_m = M_{mask} \odot x'_s + (1 - M_{mask}) \odot x_t \quad (4-7)$$

其中， \odot 表示点积操作。

为获取混合图像的标签 y_m ，使用 Mean Teacher 模型^[97]为目标图像分配伪标签。将目标未标记图像 x_t 输入到教师语义分割网络中，得到伪标签 \hat{y} ，再通过同样的 ClassMix 操作得到混合图像 x_m 的标签。

$$y_m = M_{mask} \odot y_s + (1 - M_{mask}) \odot \hat{y} \quad (4-8)$$

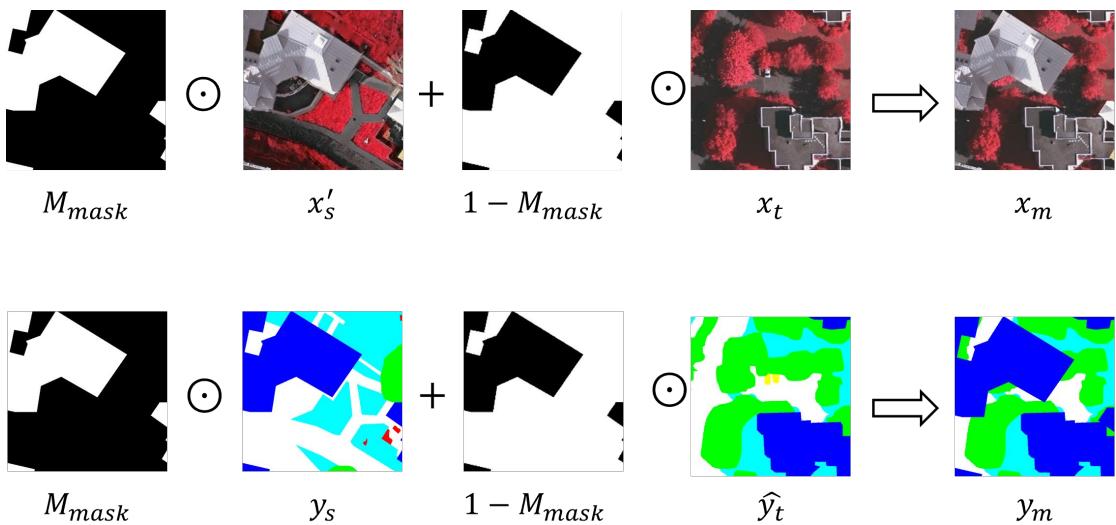


图 4-4 ClassMix 操作示意图

通过上述操作，可以获得混合图像数据集 $I_M = \{X_M, Y_M\}$ ，这部分数据会协同迁移后的数据集 $I_{\mathcal{S}}'$ ，一起参与语义分割网络的训练，该部分损失如公式 (4-9) 所示：

$$\mathcal{L}_{mix_seg} = \mathcal{L}_{ce}(X_M^{(h,w)}, Y_M^{(h,w)}) \quad (4-9)$$

边界是语义分割过程中的一个重要因素，然而现有方法往往只关注整体性能，而忽略了对象边界的重要性。因此，进一步提出了边界增强约束来实现这一目标。首先计算一个距离映射来描述每个像素到其最近的 ClassMix 掩膜边界的距离。

令 \mathcal{X}_b 表示为精确位于 ClassMix 掩膜边界的所有像素。对于混合图像 x_m 中，位于 (h, w) 这一坐标的像素点，记为 $x_m^{(h,w)}$ ，其距离映射值可计算为：

$$M_d^{(h,w)} = \min_{x \in \chi_b} d(x_m^{(h,w)}, x) \quad (4-10)$$

其中 $d(\cdot, \cdot)$ 表示欧氏距离。

$$M_{mix_b} = \left(1 - \frac{M_d}{\max(M_d)}\right) \odot \mathbb{1}[M_d < \lambda_d] \quad (4-11)$$

其中 $\mathbb{1}[\cdot]$ 是示性函数，也就是说主要关注距离小于 λ_d 的像素点，即更加关注 ClassMix 边界上的像素点， λ_d 根据经验设置为 4。

因此，可以通过公式 (4-12) 计算混合图像的边界增强损失：

$$\mathcal{L}_{mix_b} = \frac{1}{HW} \sum_{h,w}^{H,W} M_{mix_b} \odot \mathcal{L}_{ce}(X_m^{(h,w)}, Y_m^{(h,w)}) \quad (4-12)$$

对于源域迁移后的图像，很容易从其标签中获得边界掩码 M_{src_b} 。值得注意的是，源域图像中位于边界的像素点数量要明显多于混合图像。因此，这里考虑一个比例因子 γ ，用来平衡源域图像和混合图像对边界增强的贡献。源域迁移后的图像边界损失可按照公式 (4-13) 计算：

$$\mathcal{L}_{src_b} = \frac{\gamma}{HW} \sum_{h,w}^{H,W} M_{src_b} \odot \mathcal{L}_{ce}(X_s^{(h,w)}, Y_s^{(h,w)}) \quad (4-13)$$

其中 $\gamma = |\{m_{i,j} = 1 \mid m \in M_{mix_b}\}| / |\{n_{i,j} = 1 \mid n \in M_{src_b}\}|$ 表示用于平衡源边界和混合边界像素数的缩放因子。

基于以上考虑，类别分布对齐语义分割模块的总体目标可以写成公式 (4-14) 所示：

$$\mathcal{L}_{total} = \mathcal{L}_{src_seg} + \mathcal{L}_{mix_seg} + \lambda_b \cdot (\mathcal{L}_{src_b} + \mathcal{L}_{mix_b}) \quad (4-14)$$

其中 λ_b 是用于控制边界增强模块权值的超参数。

4.3 实验设计

4.3.1 任务设定

ISPRS 2D: 影像来自于 ISPRS WG III/4^[88] 的一个 2D 语义标注比赛，其中包括 Potsdam 和 Vaihingen 两个子集。Potsdam 数据集包含 3 种不同的成像模式：3 通道(IR-R-G)，3 通道(R-G-B)，4 通道(R-G-B-IR)。本文使用前两种。Vaihingen 数据集只包含一种成像模式：3 通道(IR-R-G)。Potsdam 与 Vaihingen 包含了相同的 6 个类别，其他、不透水层、汽车、树木、低矮植被和建筑。具体的数据划分与裁剪方式，采用了与上一章相同的策略。

LoveDA: LoveDA^[98] 数据集包含 5987 张 1024×1024 大小的 R-G-B 遥感图像，空间分辨率为 0.3m。具体包括 7 个类别，即建筑、道路、水、荒芜、森林、农业和背景。数据集包括 Urban 和 Rural 两个场景，收集自中国三个城市（南京、

常州和武汉)。

具体来说，提供了三种跨域遥感图像语义分割实验设置：(1) 跨地理位置，即 Potsdam IR-R-G 数据集作为源域，Vaihingen IR-R-G 数据集作为目标域。(2) 跨成像模式。Potsdam R-G-B 数据集是作为源域，Vaihingen IR-R-G 数据集作为目标域。(3) 跨景观布局。LoveDA 中的 Urban R-G-B 作为源域，Rural R-G-B 作为目标域。此外，也在相反的方向进行了尝试。

4.3.2 网络结构

对于生成器，使用了相等的 8 个卷积层（卷积核大小为 4×4 ，步长均为 2，输出通道分别为 $\{64, 128, 256, 512, 512, 512, 512, 512\}$ ）和相应的 8 个反卷积层。此外，还配置了镜像卷积层和卷积层之间的跳跃连接，使其构成一个 U 型网络。判别器有 5 个卷积层（卷积核大小为 4×4 ，步长分别为 $\{2, 2, 2, 1, 1\}$ ，输出通道分别为 $\{64, 128, 256, 512, 1\}$ ）。采用 RMSProp 优化器对模型进行优化，其中学习率为 5×10^{-5} ，权值衰减为 0.1。模型总共训练了 45 个 epochs，batch size 大小为 2。 λ_1 和 λ_2 也将分别设置为 1 和 20。通过经验观察到，为了提高语义保留损失的有效性，权值的动态递增是必要的。因此， λ_3 的值遵循公式 $\lambda_3 = 5 \cdot e^{-5(1-x)^2}$ 。其中 $x \in [0, 1]$ 表示当前 epoch 与整个 epoch 的比值。

为了对比的公平性，在 Potsdam IR-R-G/R-G-B \rightarrow Vaihingen IR-R-G 的任务设定下采用了 DeepLab v2 作为语义分割模型，并以 ResNet-101 作为骨干网络。使用 SGD 作为优化器进行训练，学习速率为 5×10^{-4} ，权重衰减为 5×10^{-4} ，以及指数为 0.9 的多项式衰减。此外还应用颜色抖动和高斯模糊进行数据增强。模型总共进行了 30000 次迭代训练，批处理大小为 2。对于 Urban R-G-B \rightarrow Rural R-G-B，采用了 DeepLab v2 分割模型，使用 ResNet-50 作为骨干网络。使用 SGD 优化器进行训练，学习速率 5×10^{-4} ，权重衰减 5×10^{-4} ，以及指数为 0.9 的多项式衰减。此外还应用颜色抖动和高斯模糊进行数据增强。模型总共训练了 30000 次迭代。以上所有的实验都是由 PyTorch 实现的，并在单个 Nvidia GeForce RTX 3090 上进行训练。

4.3.3 评价指标

(1) 图像生成质量

使用的 Frechet Inception Distance (FID)^[99] 指标评价生成图像的质量。FID 是通过在深度网络空间中估计真实图像和生成图像的分布，并计算它们之间的散度，来进行评价的。因此，FID 值越小，说明生成的图像质量越高。直观地说，如果生成的图像足够真实，则应该在特征空间中具有与真实图像相似的统计信息。

(2) 语义分割精度

这里同样针对每个类别来计算交并比 (IoU)，并最终计算所有类别的平均交并比 (mIoU) 来衡量语义分割的精度。

4.4 实验结果与分析

4.4.1 SPGAN 模型的优越性验证

上述三种典型任务在源域和目标域之间存在或多或少的光谱差异甚至成像波段差异。根据所提出的方法，GAN 将首先用于对齐光谱或波段。事实上，所有基于 GAN 的图像迁移方法 (CycleGAN, DualGAN, DiscoGAN 等) 都可以对齐光谱或成像模式的差异。然而这类方法缺少对语义信息的考虑，导致迁移后的图像经常出现偏差 (图像中表现为不平滑和不协调的区域)。因此，本部分将首先验证所提出的 SPGAN 的优点。

首先从 Potsdam IR-R-G 到 Vaihingen IR-R-G 数据集进行实验。根据图 4-2 (a)，可以得出 Potsdam IR-R-G 与 Vaihingen IR-R-G 之间的光谱统计存在差异。图像迁移结果如图 4-5 所示。可以发现，在比较方法中，都存在不同程度的偏移现象，呈现出不和谐甚至失真的区域。例如，在迁移后的图像中，部分建筑与低矮植被混合在一起。出现这种现象的原因是 GAN 注重全局统计信息的对齐，而缺乏对底层语义信息的关注。然而，由于引入了语义约束，SPGAN 的迁移结果更加真实流畅。进一步，使用 FID 对迁移后图像的质量进行评估，如表 4-1 所示。SPGAN 的 FID 评分为 107.22，明显小于其他方法，说明迁移后图像的分布更接近目标域图像的分布。

然后，在 Potsdam R-G-B 到 Vaihingen IR-R-G 的设置下进行实验，不同的成像模式更具挑战性。迁移后的图像如图 4-6 所示。同样，使用 FID 来评估迁移后图像的质量，如表 4-1 所示。SPGAN 的 FID 评分为 84.24，明显小于其他方法，说明迁移后图像的分布更接近目标域图像。

最后，进行了 Urban R-G-B 与 Rural R-G-B 间的图像迁移实验。在这种情况下，源域和目标域之间的光谱有微小的差异。然而，GAN 中类别分布的巨大差异和对语义信息的不重视仍然导致迁移图像的偏差。将会出现“无中生有”的现象。例如，在 Urban R-G-B 到 Rural R-G-B 的迁移中，植被出现在建筑之上。而在 Rural R-G-B 向 Urban R-G-B 迁移的过程中，植被上出现了零散的建筑斑块，如图 4-7 最后一行所示。相反，所提出的 SPGAN 生成语义一致的目标风格图像，看起来是无噪声的。如表 4-2 所示，SPGAN 的 FID 评分在 Urban R-G-B 到 Rural R-G-B 的实验设置下为 78.32，由 Rural R-G-B 到 Urban R-G-B 为 89.75，明显小

于其他方法。总体而言，提出的 SPGAN 在不同场景下都能很好地输出语义一致且无噪声的目标风格化的图像，证明了其有效性。

表 4-1 Potadam IR-R-G/R-G-B 迁移后的图像 FID 值（值越小图像质量越高）

任务	方法	FID
Potsdam IR-R-G ↓ Vaihingen IR-R-G	CycleGAN	135.08
	DualGAN	133.17
	DiscoGAN	135.16
	SPGAN	107.22
Potsdam R-G-B ↓ Vaihingen IR-R-G	CycleGAN	120.71
	DualGAN	118.75
	DiscoGAN	125.86
	SPGAN	84.24

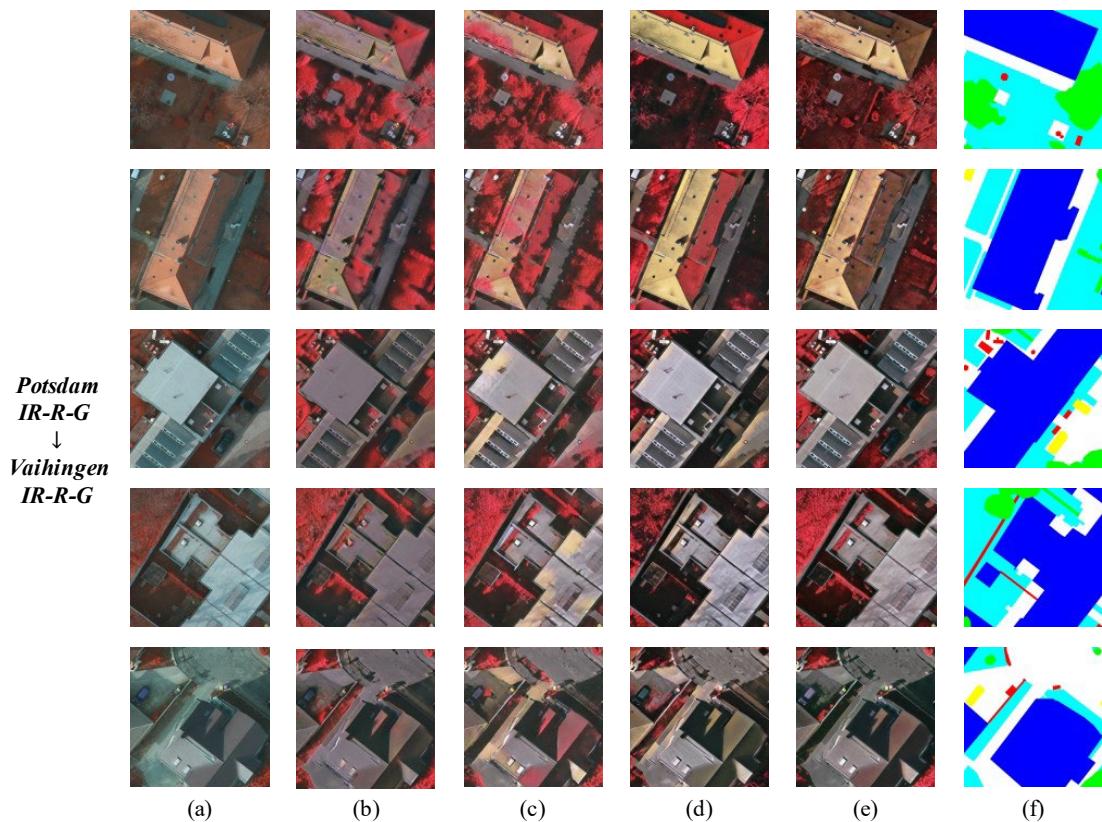


图 4-5 Potsdam IR-R-G 数据集图像迁移结果

说明：(a) 源域图像 (b) 经 CycleGAN 迁移后的图像 (c) 经 DualGAN 迁移后的图像
(d) 经 DiscoGAN 迁移后的图像 (e) 本章提出的 SPGAN 迁移后的图像 (f) 源域标签

表 4-2 Urban R-G-B 与 Rural R-G-B 迁移后的图像 FID 值（值越小图像质量越高）

任务	方法	FID
Urban R-G-B ↓ Rural R-G-B	CycleGAN	83.34
	DualGAN	81.55
	DiscoGAN	82.28
Rural R-G-B ↓ Urban R-G-B	SPGAN	78.32
	CycleGAN	96.87
	SPGAN	89.75

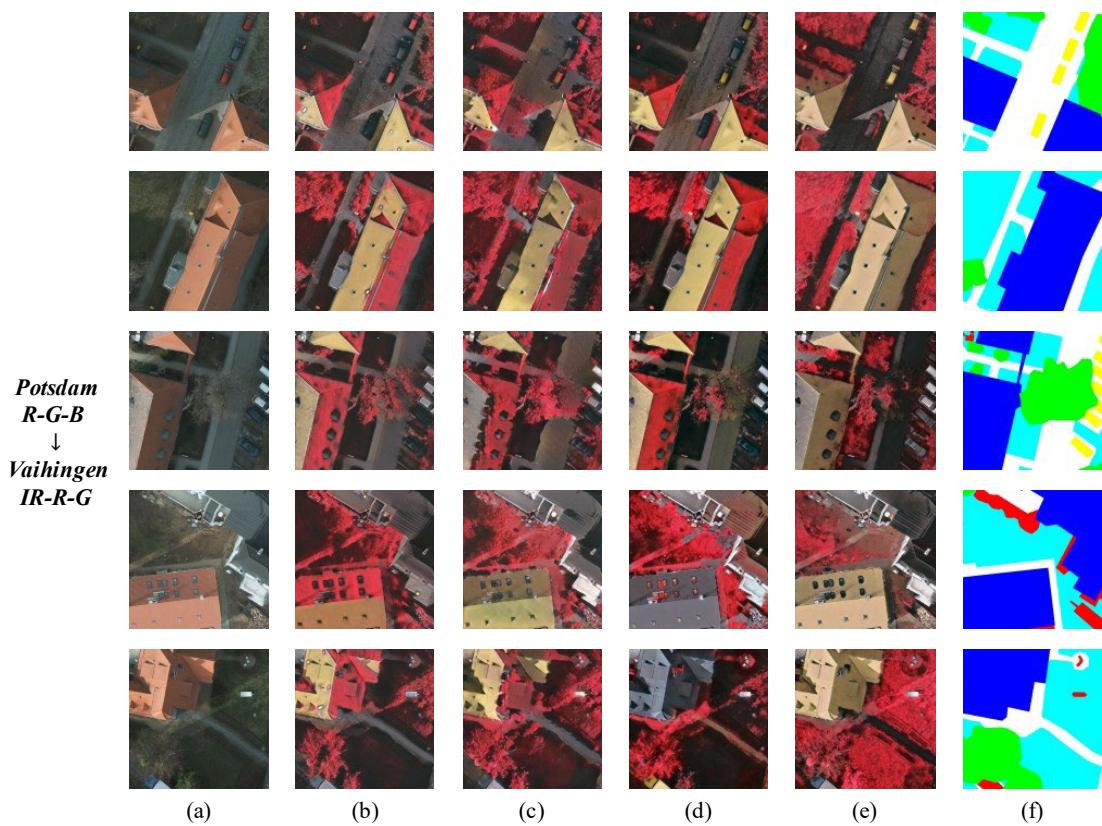


图 4-6 Potsdam R-G-B 数据集图像迁移结果

说明：(a) 源域图像 (b) 经 CycleGAN 迁移后的图像 (c) 经 DualGAN 迁移后的图像
(d) 经 DiscoGAN 迁移后的图像 (e) 本章提出的 SPGAN 迁移后的图像 (f) 源域标签

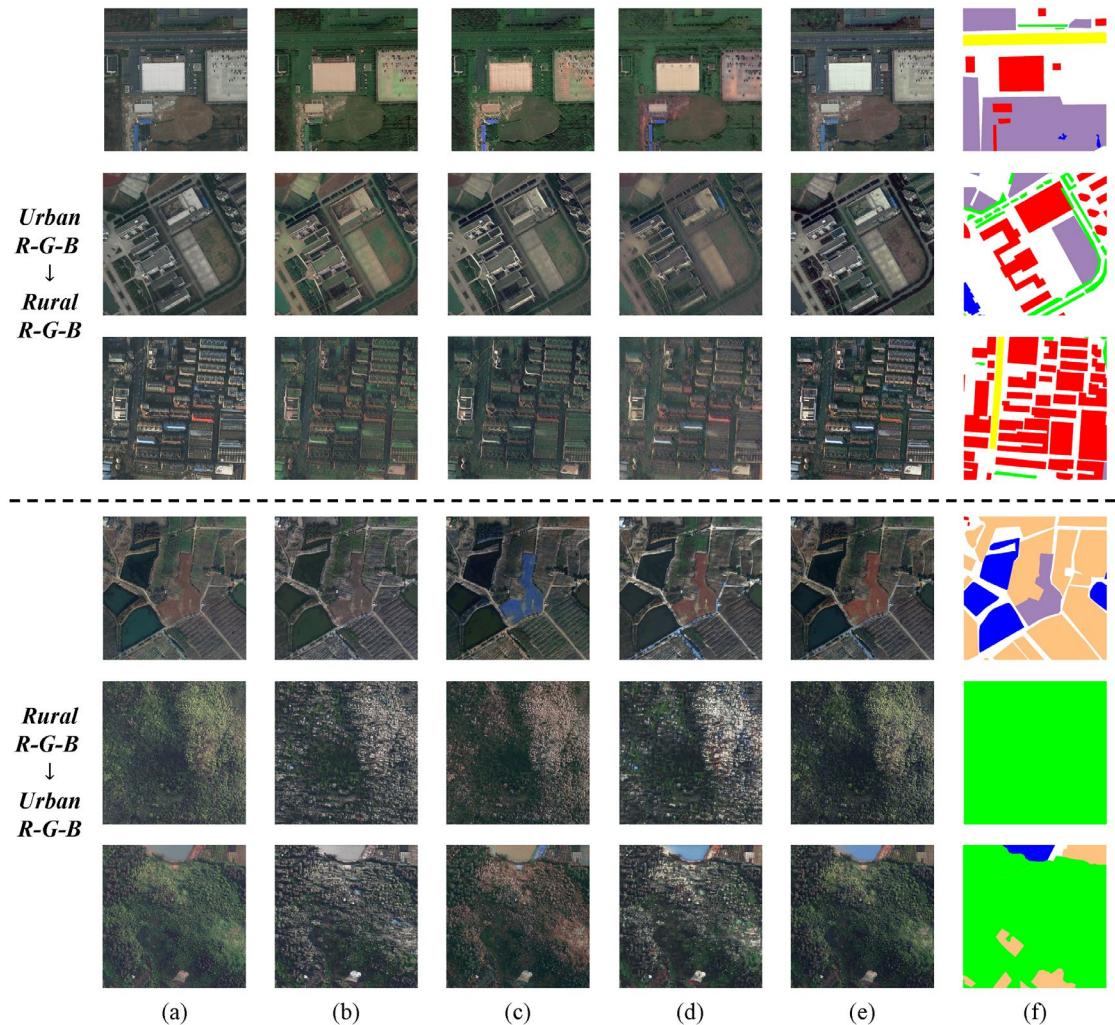


图 4-7 Urban R-G-B 与 Rural R-G-B 数据集图像迁移结果

说明: (a) 源域图像 (b) 经 CycleGAN 迁移后的图像 (c) 经 DualGAN 迁移后的图像
 (d) 经 DiscoGAN 迁移后的图像 (e) 本章提出的 SPGAN 迁移后的图像 (f) 源域标签

4.4.2 方法对比

(1) 跨地理位置

对于跨域语义分割部分,表 4-3 给出了所提出的方法与先进方法比较的实验结果。由于两个数据集之间存在严重的域偏移, Source Only 性能较差, 严重受到了域偏移问题的影响。总体来说, 领域自适应方法显著提高了总体指标和单个类别的性能, 展示了它们将知识从标记的源领域迁移到未标记的目标领域的能力。最佳模型 SPGAN-DA mIoU 达到了 53.31%, 从而相比 Source Only 提高了 22.27%。与其他竞争方法相比, SPGNA-DA 模型仍然具有更高的性能, 并在一些具有挑战性的类别中做出足够强的预测, 例如“汽车”、“树木”和“建筑”。这充分显示了 SPGAN-DA 方法的优越性。图 4-8 展示了基线和提出的方法 SPGAN-DA 在 Potsdam IR-R-G 到 Vaihingen IR-R-G 上获得的一些定性语义分割示例。由于

Potsdam IR-R-G 和 Vaihingen IR-R-G 之间存在严重的域偏移问题，Soure Only 预测通常出现噪声分割或错误上下文，丢失目标边界和结构信息。经过 SPGAN-DA 领域自适应后，结果有了很大的改善，并保留了更多的结构信息和细节，特别是在边缘。

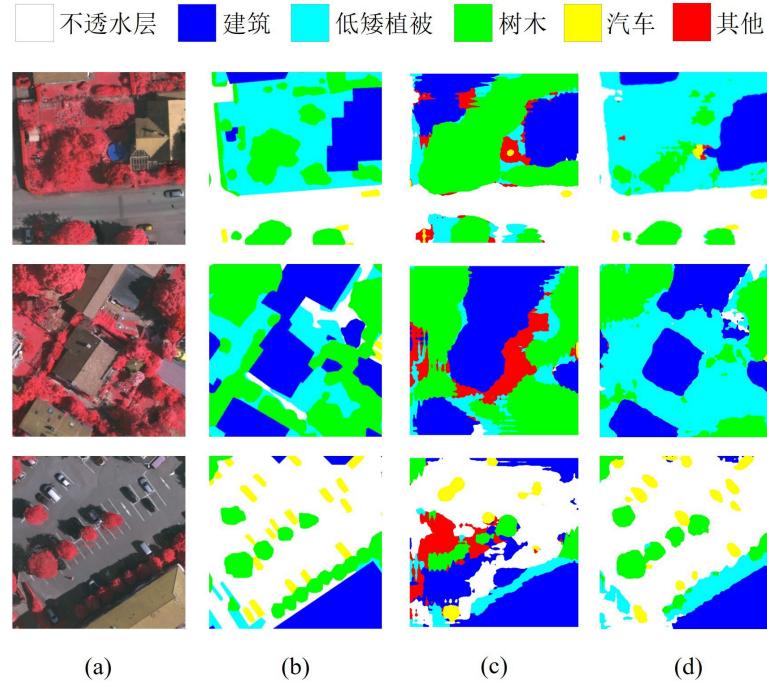


图 4-8 Potsdam IR-R-G → Vaihingen IR-R-G 语义分割结果可视化

说明：(a) 目标域图像 (b) 目标域标签 (c) Sourc Only (d) SPGAN-DA

表 4-3 跨地理位置下不同域适应方法的 mIoU 对比结果

任务	方法	IoU							mIoU
		其他	不透水层	汽车	树木	低矮植被	建筑		
Potsdam IR-R-G ↓ Vaihingen IR-R-G	Source Only	5.71	35.84	20.27	54.95	17.88	51.59	31.04	
	Benjdir's ^[89]	2.12	39.88	8.20	26.56	26.53	40.97	24.04	
	MWCSS	29.66	49.41	34.34	57.66	38.87	62.30	45.38	
	AdaptSegNet ^[41]	4.60	54.39	6.40	52.65	28.98	63.14	35.02	
	CsDA ^[90]	9.85	46.22	31.14	52.04	31.11	52.39	37.12	
	AdvEnt ^[42]	10.18	57.03	35.28	59.02	33.65	67.41	43.76	
	CLAN ^[92]	9.89	58.23	37.25	59.10	36.74	59.10	43.38	
	LSR-EGA ^[100]	11.25	59.61	38.56	56.55	38.73	71.04	45.95	
	MRNet ^[91]	3.81	55.02	34.37	54.79	26.46	76.39	41.80	
	SPGAN-DA	11.94	65.33	48.25	66.04	45.20	83.11	53.31	

(2) 跨成像模式

为了进一步验证方法的泛化能力，在跨成像模式设定下，即 Potsdam R-G-B 到 Vaihingen IR-R-G 数据集的实验。表 4-4 报告了本文方法与先进方法在 Potsdam R-G-B 和 Vaihingen IR-R-G 数据集上的语义分割结果对比。由于两个数据集之间存在严重的域偏移，Source Only 方法的性能非常差。一般来说，领域自适应方法显著提高了整体指标和单个类别的性能，显现出了它们将领域知识从标记的源领域迁移到未标记的目标领域的能力。本章提出的 SPGAN-DA 的 mIoU 值达到了 54.41%，相比 Source Only 高了 30.40%。与其他竞争方法相比，SPGAN-DA 在“低矮植被”、“其他”和“汽车”等具有挑战性的类别中仍然具有更高的性能和足够强的预测能力，这证明了 SPGAN-DA 既能充分学到领域不变特征，又能保持对类别的鉴别能力。主要原因是，与自然图像相比，由于成像波段不同，两种数据集之间存在巨大差异，其他方法可能会比 SPGAN-DA 方法逊色。对于本章提出的方法，它首先通过 SPGAN 对齐风格（即在此任务中的成像波段），其次使用 ClassMix 策略，以便模型同时看到源域对象和目标域对象。因此，即使在处理如此大的领域偏移时，它也足以对目标领域产生高置信度预测。语义分割的可视化结果如图 4-9 所示。与前面的实验任务设置类似，Source Only 的预测通常表现为有噪声的分割或错误的上下文。本章提出的 SPGAN-DA 结果与真实标签相似。

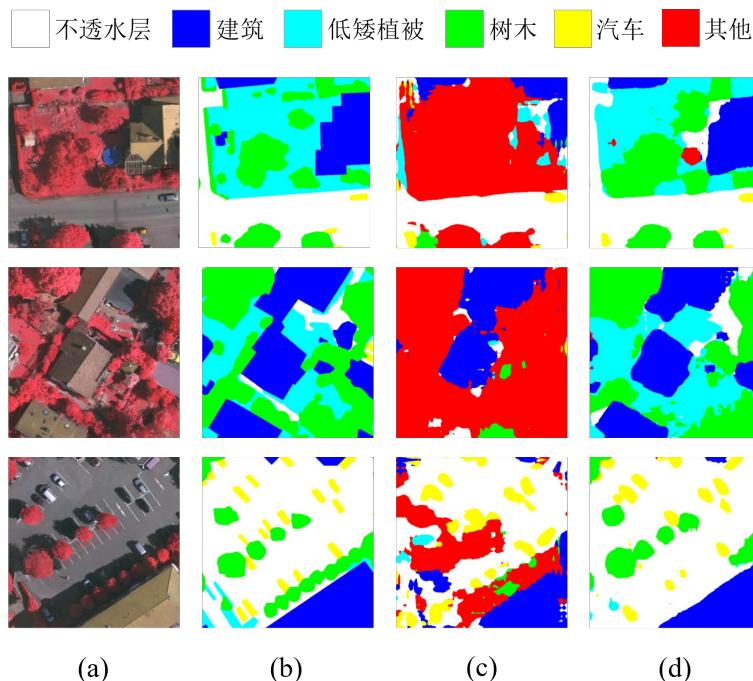


图 4-9 Potsdam R-G-B → Vaihingen IR-R-G 语义分割结果可视化

说明：(a) 目标域图像 (b) 目标域标签 (c) Sourc Only (d) SPGAN-DA

表 4-4 跨成像模式下不同域适应方法的 mIoU 对比结果

任务	方法	IoU						mIoU
		其他	不透水层	汽车	树木	低矮植被	建筑	
Potsdam R-G-B ↓ Vaihingen IR-R-G	Source Only	1.76	26.86	16.70	44.48	12.56	41.67	24.01
	Benjdir's ^[89]	4.48	31.78	21.70	41.76	23.67	52.36	29.31
	MWCSS	3.94	46.19	40.31	55.82	27.85	65.44	39.93
	AdaptSegNet ^[41]	1.08	50.05	14.18	56.45	20.73	62.61	34.18
	CsDA ^[90]	0.55	44.82	23.81	52.04	20.74	53.39	32.56
	AdvEnt ^[42]	0.73	55.43	28.28	59.02	20.73	68.49	38.78
	CLAN ^[92]	0.84	57.30	17.28	59.10	24.94	59.19	36.44
	LSR-EGA ^[100]	5.84	57.11	23.65	56.55	28.73	70.24	40.35
	MRNet ^[91]	0.81	54.11	29.39	54.99	16.16	75.39	38.47
	SPGAN-DA	23.45	67.88	49.79	55.26	47.27	82.83	54.41

(3) 跨景观布局

为了评估本章提出的方法在更多情况下的有效性，在 Urban R-G-B 和 Rural R-G-B 的双向设置中进行了实验。在这种情况下，语义分割模型存在较小的光谱或成像模式差异，但存在较大的类别分布差异。

Urban R-G-B → Rural R-G-B. 这组实验的结果如表 4-5 所示。从 Source Only 的结果可以看出，在源域训练的语义分割网络应用到目标域时，由于域偏移问题而失去了有效性。这证实了该任务的复杂性，这是由于源和源之间的强且不一致的类别分布。以城市景观为主，有建筑和高速公路，但很少有自然景观。从表 4-5 中可以看出，基于对抗的方法（如 TransNorm, FADA）可能会受到负迁移的影响，最终的效果甚至差于 Source Only。基于自学习训练的方法在目标图像上生成伪标签，随着目标样本的加入，在训练过程中一定程度上消除了类别分布的差异的不利影响。对于所提出的 SPGNA-DA 算法，其在目标域 Rural 数据集的 mIoU 可以达到 42.07%，显著优于其他代表性方法。此外，对于源域 Urban 数据集中占比不高的类别（如水体），SPGNA-DA 算法在对目标域数据集进行识别时，仍能保持较好的分类效果。

Rural R-G-B → Urban R-G-B. 在这种情况下，源域 Rural 数据集以大规模的自然物体和少量的人造样本为主。具体表现为“农田”、“林地”、“水体”这些类别占比相对较高，“建筑”、“道路”这些类别占比较低。这组实验的定量结果如表 4-5 所示。在 mIoU 方面，SPGAN-DA 达到 48.56%，超过 Source Only 模型

17.29%。Rural R-G-B 到 Urban R-G-B 实验的语义分割可视化结果如图 4-10 所示。由于源域 Rural 数据集中建筑物占比较少，一些基于对抗的方法（如 TransNorm, FADA）可能会受到负迁移的影响，难以识别目标域 Urban 图像中的建筑物。相比之下，基于自学习的方法的识别效果会好一些。这是因为该类方法会生成目标图像的伪标签，使目标域数据参与训练，一定程度上缓解类别分布差异的影响。由于 SPGAN-DA 综合考虑了光谱波段和类别分布差异，能成功地识别出建筑物和道路，并且语义分割结果是最接近真是标签的。此外，得益于边界增强损失的约束，对于图 4-10 (a) 右上角“弧形”的建筑，SPGAN-DA 方法也能对其很好地识别。

表 4-5 跨景观布局下不同域适应方法的 mIoU 对比结果

任务	方法	IoU							mIoU	
		背景	建筑	道路	水体	荒漠	林地	农田		
Urban R-G-B	Source Only	24.16	37.02	32.56	49.42	14.00	29.34	35.65	31.74	
	DDC ^[101]	25.61	44.27	31.28	44.78	13.74	33.83	25.98	31.36	
	AdaptSegNet ^[41]	26.89	40.53	30.65	50.09	16.97	32.51	28.25	32.27	
	FADA ^[102]	24.39	32.97	25.61	47.59	15.34	34.35	20.29	28.65	
	CLAN ^[92]	22.93	44.78	25.99	46.81	10.54	37.21	24.45	30.39	
	↓ TransNorm ^[103]	19.39	36.30	22.04	36.68	14.00	40.62	3.30	24.62	
	Rural R-G-B	PyCDA ^[104]	12.36	38.11	20.45	57.16	18.32	36.71	41.90	32.14
	CBST ^[43]	25.06	44.02	23.79	50.48	8.33	39.16	49.65	34.36	
	IAST ^[105]	29.97	49.48	28.29	64.49	2.13	33.36	61.37	38.44	
Urban R-G-B	UDA-CL ^[106]	28.55	49.69	35.74	53.52	4.96	31.36	52.26	36.58	
	SPGAN-DA	55.06	50.71	33.80	65.01	9.07	25.43	55.44	42.07	
Rural R-G-B	Source Only	43.30	25.63	12.70	76.22	12.52	23.34	25.14	31.27	
	DDC ^[101]	43.60	15.37	11.98	79.07	14.13	33.08	23.47	31.53	
	AdaptSegNet ^[41]	42.35	23.73	15.61	81.95	13.62	28.70	22.05	32.68	
	FADA ^[102]	43.89	12.62	12.76	80.37	12.70	32.76	24.79	31.41	
	CLAN ^[92]	43.41	25.42	13.75	79.25	13.71	30.44	25.80	33.11	
	↓ TransNorm ^[103]	38.37	5.04	3.75	80.83	14.19	33.99	17.91	27.73	
	Rural R-G-B	PyCDA ^[104]	38.04	35.86	45.51	74.87	7.71	40.39	11.39	36.25
	CBST ^[43]	48.37	46.10	35.79	80.05	19.18	29.69	30.05	41.32	
	IAST ^[105]	48.57	31.51	28.73	86.01	20.29	31.77	36.50	40.48	
Urban R-G-B	UDA-CL ^[106]	48.15	37.44	45.05	84.29	16.68	26.66	34.12	41.77	
	SPGAN-DA	35.45	50.31	51.02	68.42	42.68	40.17	51.90	48.56	

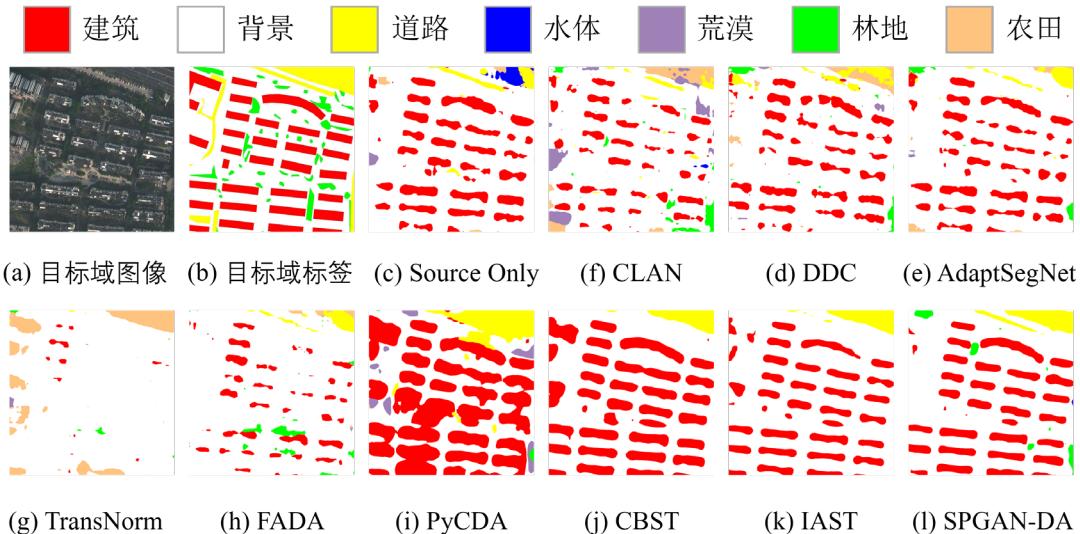


图 4-10 Rural R-G-B → Urban R-G-B 语义分割结果可视化

4.4.3 模型分析

(1) 消融实验

在本节中，进行了消融实验，以验证在 Potsdam IR-R-G/R-G-B → Vaihingen IR-R-G 和 Urban R-G-B ↔ Rural R-G-B 设定下各个模块起到的作用。

SO: 以 Source Only 为基线，即在源域训练，目标域测试，未采取任何领域自适应措施。Potsdam IR-R-G → Vaihingen IR-R-G 的 mIoU 为 31.0%。Potsdam R-G-B → Vaihingen IR-R-G 的 mIoU 为 24.0%。此外，在 Urban R-G-B ↔ Rural R-G-B 设置下，mIoU 分别为 31.7% 和 31.3%。通过这组实验可以说明仅在源域训练未经任何领域自适应的模型，直接应用在目标域性能十分有限，尤其像 Potsdam R-G-B → Vaihingen IR-R-G 这种域偏移十分大的情况。

AL+ICL: “AL+ICL”是传统 GAN 图像迁移模型中使用的对抗损失和图像一致性损失。AL+ICL 在 Potsdam IR-R-G/ R-G-B → Vaihingen IR-R-G 下的 mIoU 相比于 SO 分别提高了 7.9% 和 12.9%。这表明在这两种任务设定下，光谱或成像模式的差异是跨域语义分割的主要影响因素。然而，AL+ICL 在 Urban R-G-B ↔ Rural R-G-B 设定下，并没有像前两种情况那样显著提升。这表明，在这种情况下，模型受光谱差异的影响较小，而主要受到类别分布差异的影响。

RIL: “RIL”表示提出的表征不变损失，对应于公式 (4-3)。

SPL: SPL 表示语义保持损失，对应于公式 (4-4)。由于 SPL 的加入，mIoU 相比 SO 提高了 14.4%，分别为 20.2%、4.4% 和 6.9%。结果表明，该算法能够在不同场景下输出语义一致、无噪声的目标样式图像。

CDA: “CDA”表示类别分布对齐模块。CDA 通过 ClassMix 操作缓解类别分布的差异，边界增强损失细化语义分割物体对象的边界，有助于提高模型性能。

尽管 Potsdam IR-R-G/R-G-B \rightarrow Vaihingen IR-R-G 中类别分布差异不大，但 CDA 通过对齐两个领域的建筑等方面的差异，提升了模型的泛化能力。具体来说，Vaihingen 的建筑更为集中，相比之下 Potsdam 的建筑更为分散，因此 Potsdam 的道路也相对更宽阔。对于 Urban R-G-B \leftrightarrow Rural R-G-B，由于人口密度高，城市地区总是包含更多的人造物体，如建筑物和道路。相比之下，农村地区有更多的农业用地。也就是说，城市和农村的类别分布有很大的差异。CDA 直接对齐城市和农村之间的类别分布，显著提高了目标域的分割性能。

总体而言，在本章所提出的 SPGAN-DA 框架下，领域差距逐步缩小，每个组件/阶段都有助于提升跨域遥感图像语义分割的性能。

表 4-6 SPGAN-DA 算法消融实验

	SO	AL+ICL	RIL	SPL	CDA	mIoU	提升
	✓					31.0	
Potsdam IR-R-G		✓				38.9	+7.9
↓		✓	✓			43.2	+12.2
Vaihingen IR-R-G		✓	✓	✓		45.4	+14.4
	✓	✓	✓	✓	✓	53.3	+22.3
	✓					24.0	
Potsdam R-G-B		✓				36.9	+12.9
↓		✓	✓			41.5	+17.5
Vaihingen IR-R-G		✓	✓	✓		44.2	+20.2
	✓	✓	✓	✓	✓	54.4	+30.4
	✓					31.7	
Urban R-G-B		✓				33.5	+1.8
↓		✓	✓			34.1	+2.4
Rural R-G-B		✓	✓	✓		36.1	+4.4
	✓	✓	✓	✓	✓	42.0	+10.3
	✓					31.3	
Rural R-G-B		✓				34.2	+2.9
↓		✓	✓			35.9	+4.6
Urban R-G-B		✓	✓	✓		38.2	+6.9
	✓	✓	✓	✓	✓	48.6	+17.3

(2) 超参数分析

在协同自适应边界增强分割模块中， λ_b 是一个重要的超参数，它指导语义分割网络对物体边界的关注程度。为了分析 λ_b 的敏感性，本部分进行了相关实验。

直观来说,对于边界权值 λ_b ,如果 λ_b 的值过小,那么边界增强的作用就非常有限,带来的性能提升也很小。反之,如果过高,则会影响网络的正常优化,甚至导致性能下降。因此找出一个合适的 λ_b 值是至关重要的。上述任务权重不同的结果如表 4-7 所示。对于 Potsdam IR-R-G 到 Vaihingen IR-R-G 的跨域语义分割任务,在超参数 $\lambda_b = 3$ 的情况下,该方法在验证集上实现了 53.51% 的 mIoU。对于 Potsdam R-G-B 到 Vaihingen IR-R-G 的跨域语义分割任务,在超参数 $\lambda_b = 3$ 的情况下,在验证集上得到最优的效果, mIoU 为 55.14%。最后,对于 Urban R-G-B↔Rural R-G-B 如表 4-8 所示,在超参数 $\lambda_b = 2$ 的情况下达到最优。

表 4-7 Potsdam IR-R-G/R-G-B → Vaihingen IR-R-G 阈值参数 λ_b 分析

任务	值	IoU							mIoU
		其他	不透 水层	汽车	树木	低矮 植被	建筑		
Potsdam IR-R-G	$\lambda_b = 1$	19.29	63.67	44.38	62.61	44.43	81.78	52.69	
	\downarrow	15.53	63.25	48.21	64.17	46.35	83.27	53.46	
Vaihingen IR-R-G	$\lambda_b = 3$	13.66	63.72	48.98	66.64	44.25	83.84	53.51	
	$\lambda_b = 4$	11.25	58.83	51.89	60.12	43.71	83.83	51.60	
Potsdam R-G-B	$\lambda_b = 1$	12.64	67.99	51.72	61.01	48.38	84.99	54.45	
	\downarrow	29.06	65.83	50.41	56.24	46.97	82.36	55.14	
Vaihingen IR-R-G	$\lambda_b = 3$	14.06	66.93	51.49	62.36	49.83	85.03	54.94	
	$\lambda_b = 4$	8.90	66.49	51.18	60.31	46.93	83.25	52.84	

表 4-8 Urban R-G-B ↔ Rural R-G-B 阈值参数 λ_b 分析

任务	值	IoU							mIoU
		背景	建筑	道路	水体	荒漠	林地	农田	
Urban R-G-B	$\lambda_b = 1$	57.08	58.36	39.05	63.50	2.31	26.19	61.18	43.95
	\downarrow	59.32	52.14	36.80	66.33	10.92	29.59	57.14	44.60
Rural R-G-B	$\lambda_b = 3$	58.70	43.16	40.91	59.91	14.68	35.35	54.12	43.83
	$\lambda_b = 4$	61.05	53.87	38.07	61.17	11.66	13.87	55.79	42.21
Rural R-G-B	$\lambda_b = 1$	39.96	49.64	51.18	70.17	44.01	40.63	53.27	49.83
	\downarrow	37.61	53.31	53.02	72.09	45.00	40.17	54.90	50.87
Urban R-G-B	$\lambda_b = 3$	39.36	52.57	53.27	72.20	44.39	38.89	51.87	50.36
	$\lambda_b = 4$	36.71	53.33	52.01	55.78	44.14	49.95	54.26	49.45

(3) 特征分布可视化

为了更直观地说明所提出的模块的有效性,图 4-11 显示了 Source Only 和

SPGAN-DA 方法在二维空间的特征分布图。其中，二维空间的特征分布图是由通过 t-SNE^[107]算法获得。对于 Potsdam IR-R-G/R-G-B 到 Vaihingen IR-R-G，不同类别对应的特征是混淆在一起的，如图 4-11 (a) 和 (b) 所示。特别是不透水层、低矮植被、建筑和树木的特征高度重叠在一起，这无疑会导致分割结果较差。Urban R-G-B 和 Rural R-G-B 的实验也面临同样的问题，不同的类别特征混淆，导致语义分割结果较差。通过域自适应，SPGAN-DA 得到了比 Source Only 更准确的分割结果，并减弱了不同类别之间的特征混淆，如图 4-11 (c) 和 (d) 所示。该方法进一步提高了引入类特征的紧凑性和类间可分性。特别是，它增加了容易混淆的类别对之间的特征分布距离。

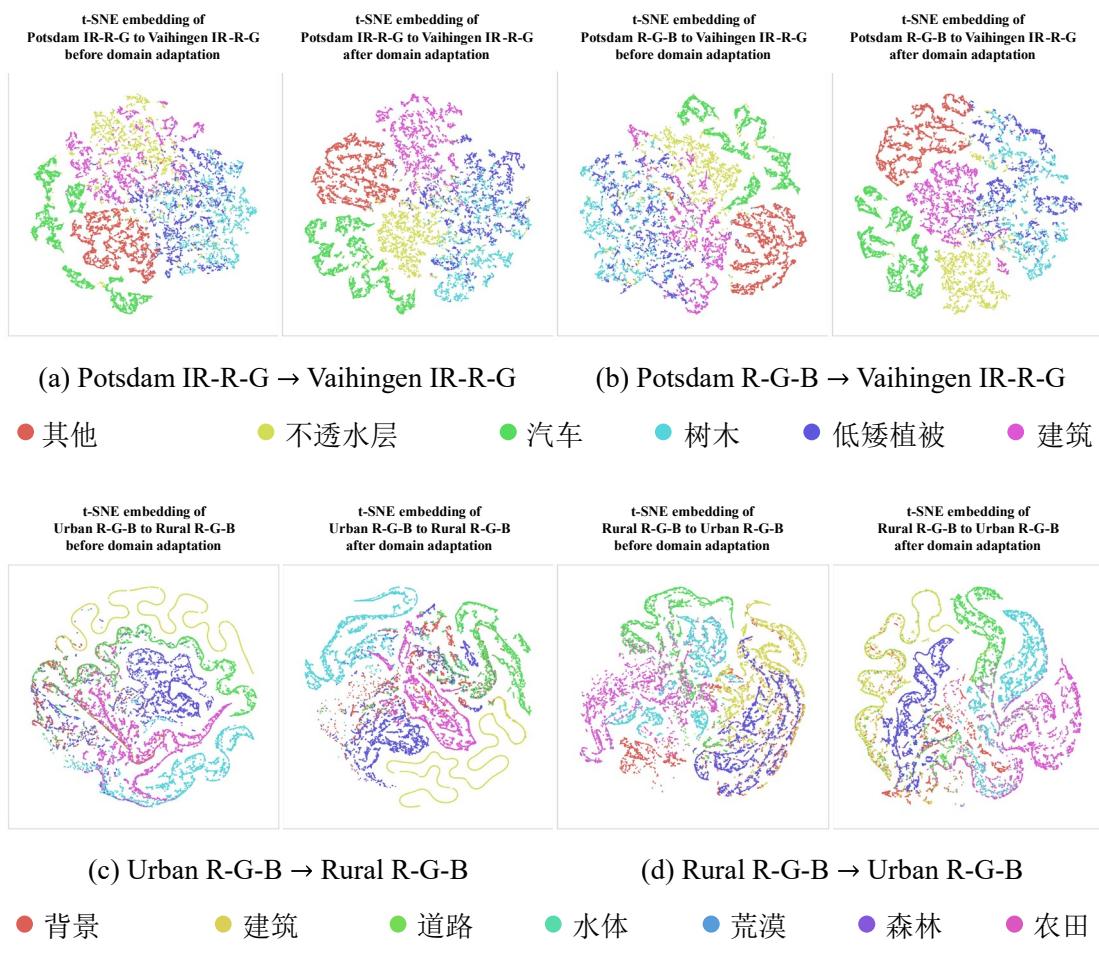


图 4-11 t-SNE 特征分布可视化

4.5 本章小结

在本章的研究中，提出了一个新颖的框架，称为 SPGAN-DA，以缓解域偏移问题对跨域遥感图像语义分割的影响。首先，提出的 SPGAN 将源图像迁移成目标域的风格，并保持语义信息不变，这将最大限度地减少光谱或成像模式的差异而不产生偏差。此外，提出了一个类分布对齐模块，利用模型输入和输出方面的

迁移后的类目标图像和目标域图像，协同训练一个域自适应语义分割模型。这为处理遥感图像跨域语义分割任务提供了一个新范式。在遥感基准数据集上的实验证明了提出的方法的有效性和通用性，与其他最先进的方法相比，该方法取得了有竞争力的结果表现。

第 5 章 总结与展望

5.1 研究总结

随着遥感大数据时代的到来，遥感数据的获取能力得到了极大提升，这为遥感智能化解译提供了数据支撑。遥感图像语义分割作为遥感影像智能解译的基础任务之一，其结果被广泛应用于地表专题制图、智能交通、环境监测、城市规划等方面。伴随着深度学习技术的日渐成熟，基于深度学习方法性能已远超传统方法。但深度学习的卓越表现依赖于大量的精准标注数据。为减轻标注数据工作的压力，一种常用的方法是使用已有的标注数据集（源域）进行训练，在未标注的数据集（目标域）上进行测试与推理。由于不同遥感图像数据集之间存在较大的差异，存在严重的“同物异谱”和“同谱异物”现象。而传统的深度学习模型训练数据与测试数据独立同分布，这势必将导致在源域图像场景下训练的语义分割模型难以泛化到目标域图像场景。无监督领域自适应跨域语义分割方法为这一思路提供了可能，可以将源域学习到的模型泛化到目标域并保持较好的性能。本文以生成对抗网络图像迁移为主线，开展了一系列关于跨域遥感图像语义分割的研究工作：

(1) 提出基于对偶生成对抗网络的域适应遥感图像语义分割方法 MWCSS。该方法可以在有标注的源域数据集上训练，应用在目标域无标注的数据集上测试，一定程度克服了域偏移问题的影响，在目标域数据集仍旧能保持较好的性能。其中多个弱监督约束由 WTIC、WPLC 和 WRCC 组成。为了充分发挥多个约束的优点，提出了一种动态优化策略，在训练过程中动态调整目标函数的约束权值，可以避免深度语义分割网络的退化。本方法在跨地理位置和跨成像模式的遥感图像跨域语义分割任务上取得了优越的结果。

(2) 提出基于语义保持生成对抗网络的域适应遥感图像语义分割方法 SPGAN-DA。由于现有的基于 GAN 的图迁移模型未充分考虑语义信息，致使迁移后的图像往往会产生偏差，导致迁移效果有限。因此提出了 SPGAN，通过在 GAN 中引入表征不变约束和语义保持约束，来实现从源域到目标域的语义保持迁移。此外，考虑到源域和目标域数据的类别分布存在差异，提出了类别分布对齐语义分割模块 CDA。类分布对齐语义分割包括两个层面：在模型输入层，通过使用 ClassMix 操作，以随机采样的方式在迁移后的图像上随机选择一定的类别，将相应的图像块粘贴到目标域图像上，构成混合图像。将迁移后的图像与混合图像一同送入语义分割网络进行训练。在模型输出层面，提出了边界增强损失来细化物体边界的分割性能。所提出的 SPGAN-DA 方法，在三个典型的跨域

遥感图像语义分割任务都取得了最优的性能，充分证明了算法的优越性与普适性。

5.2 未来展望

本文研究了基于生成对抗网络的遥感图像跨域语义分割问题，取得了一定进展。针对其中一些仍有提升空间的部分进行了展望，提出了未来研究中值得探索的方向，主要包括以下几个方面：

(1) 端到端一体化学习

在未来的工作中，将考虑在闭环的框架下共同优化生成网络的图像迁移模型和语义分割模型。一方面，这种方式的训练处理等流程将更加简洁易用，省时省力；另一方面，这种方式可以让生成对抗网络与语义分割模型发生交互，二者相互反馈。具体表现为，语义分割模型可独立判断图像迁移到何种程度最优，不再需要人为手工分析判断。此外，生成对抗网络可以与语义分割模型进行交互，二者可以互相反馈，获取更多的领域不变知识，切实提升在目标域的语义分割效果。

(2) 多源领域自适应方法

多源领域自适应能够更充分地利用已有数据集，实现对地面的全方位观测，在智能调度、安防监控、灾后重建等方面都有着广泛的应用前景。但由于地理位置、光照条件、传感器类型、地物类别分布等差异的存在，多源领域自适应语义分割将会面临更严峻的挑战。如何在已有的多个源域学习到领域不变知识，并迁移泛化到目标域，是摆在眼前的一个难题，值得在未来进行深入研究探索。

(3) 耦合遥感成像机理的图像迁移

目前针对遥感图像的迁移是直接利用生成对抗网络直接对齐光谱/波段，并未充分考虑到遥感图像的成像机理。通过耦合物理机理与深度学习，发展学习与机理双驱动的反演框架有着广阔的前景。具体地，可以通过反演出地表反射率，找出这一不变量，通过耦合遥感成像机理来指导遥感图像间的迁移过程。这样不仅能够处理复杂的物理特性，改善理论和优化系统，还可以弥补深度学习的可解释性，进一步提升模型的效率和精度，是未来的一个重要研究方向。

参考文献

- [1] 李德仁, 张良培, 夏桂松. 遥感大数据自动分析与数据挖掘[J]. 测绘学报, 2014, 43(12): 1211.
- [2] 张兵. 遥感大数据时代与智能信息提取[J]. 武汉大学学报(信息科学版), 2018, 43(12): 1861-1871.
- [3] 付琨, 孙显, 仇晓兰, 等. 遥感大数据条件下多星一体化处理与分析[J]. 遥感学报, 2021, 25(3): 691-707.
- [4] Chi M, Plaza A, Benediktsson J A, et al. Big data for remote sensing: Challenges and opportunities[J]. Proceedings of the IEEE, 2016, 104(11): 2207-2219.
- [5] 范一大, 吴玮, 王薇, 等. 中国灾害遥感研究进展[J]. 遥感学报, 2016, 20(5): 1170-1184.
- [6] 李德仁. 遥感用于自然灾害监测预警大有作为[J]. 科技导报, 2007, 25(0706): 1.
- [7] Cheng G, Guo L, Zhao T, et al. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA[J]. International Journal of Remote Sensing, 2013, 34(1): 45-59.
- [8] 张翰超. 基于遥感监测的中国典型城市时空格局演变及可持续性评价研究[D]. 武汉大学, 2019
- [9] Li Y, Chen W, Zhang Y, et al. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning[J]. Remote Sensing of Environment, 2020, 250: 112045.
- [10] Yan J, Ji S, Wei Y. A combination of convolutional and graph neural networks for regularized road surface extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
- [11] 陈凯强, 高鑫, 闫梦龙, 等. 基于编解码网络的航空影像像素级建筑物提取[J]. 遥感学报, 2020, 24(9): 1134-1142.
- [12] Tong X, Xia G, Lu Q, et al. Land-cover classification with high-resolution remote sensing images using transferable deep models[J]. Remote Sensing of Environment, 2020, 237: 111322.
- [13] 李彦胜, 张永军. 耦合知识图谱和深度学习的新一代遥感影像解译范式[J]. 武汉大学学报(信息科学版), 2022, 47(8): 1176-1190.
- [14] 李鑫伟, 李彦胜, 张永军. 弱监督深度语义分割网络的多源遥感影像水体检测[J]. 中国图象图形学报, 2021, 26(12): 3015-3026.
- [15] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 3213-3223.
- [16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 3431-3440.
- [17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image

- segmentation[C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, 234-241.
- [18] 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.
- [19] Chen L, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[C]. International Conference on Learning Representations (ICLR), 2015, 1-14.
- [20] Chen L, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [21] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 770-778.
- [22] Chen L, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv:1706.05587, 2017: 1-14.
- [23] Chen L, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. European Conference on Computer Vision (ECCV), 2018, 801-818.
- [24] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 1800-1807.
- [25] Everingham M, Van Gool L, Williams C K, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88: 303-338.
- [26] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2017, 5998-6008.
- [27] Strudel R, Garcia Pinel R, Laptev I, et al. Segmenter: Transformer for Semantic Segmentation[C]. International Conference on Computer Vision (ICCV), 2021, 7242-7252.
- [28] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]. International Conference on Learning Representations (ICLR), 2020, 1-22.
- [29] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2021, 12077-12090.
- [30] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 633-641.
- [31] Hendrycks D, Dietterich T G. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations[C]. International Conference on Learning Representations (ICLR), 2019, 1-16.
- [32] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets[C]. Annual

- Conference on Neural Information Processing Systems (NeurIPS), 2014, 2672-2680.
- [33] Zhu J, Park T, Isola P, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks[C]. International Conference on Computer Vision (ICCV), 2017, 2223-2232.
- [34] Hoffman J, Tzeng E, Park T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]. International Conference on Machine Learning (ICML), 2018, 1989-1998.
- [35] Wu Z, Han X, Lin Y, et al. Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation[C]. European Conference on Computer Vision (ECCV), 2018, 518-534.
- [36] Choi J, Kim T, Kim C. Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation[C]. International Conference on Computer Vision (ICCV), 2019, 6830-6840.
- [37] Yang Y, Lao D, Sundaramoorthi G, et al. Phase consistent ecological domain adaptation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 9011-9020.
- [38] Yang Y, Soatto S. Fda: Fourier domain adaptation for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020, 4085-4095.
- [39] Gao L, Zhang L, Zhang Q. Addressing domain gap via content invariant representation for semantic segmentation[C]. AAAI Conference on Artificial Intelligence (AAAI), 2021, 7528-7536.
- [40] Hoffman J, Wang D, Yu F, et al. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation[J]. arXiv:1612.02649, 2016.
- [41] Tsai Y, Hung W, Schulter S, et al. Learning to adapt structured output space for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 7472-7481.
- [42] Vu T, Jain H, Bucher M, et al. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 2517-2526.
- [43] Zou Y, Yu Z, Kumar B, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training[C]. European Conference on Computer Vision (ECCV), 2018, 289-305.
- [44] Xu Y, Du B, Zhang L, et al. Self-ensembling attention networks: Addressing domain shift for semantic segmentation[C]. AAAI Conference on Artificial Intelligence (AAAI), 2019, 5581-5588.
- [45] Wang S, Zhao D, Zhang C, et al. Cluster Alignment With Target Knowledge Mining for Unsupervised Domain Adaptation Semantic Segmentation[J]. IEEE Transactions on Image Processing, 2022, 31: 7403-7418.
- [46] Yuan B, Zhao D, Shao S, et al. Birds of a feather flock together: Category-divergence guidance for domain adaptive segmentation[J]. IEEE Transactions on Image Processing, 2022, 31: 2878-2892.

- [47] Tranheden W, Olsson V, Pinto J, et al. Dacs: Domain adaptation via cross-domain mixed sampling[C]. IEEE Winter Conference on Applications of Computer Vision (WACV), 2021, 1379-1389.
- [48] Schwaighofer A, Tresp V, Yu K. Learning Gaussian process kernels via hierarchical Bayes[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2004, 1209-1216.
- [49] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks?[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2014, 3320-3328.
- [50] 许夙晖, 慕晓冬, 柴栋, 等. 基于极限学习机参数迁移的域适应算法[J]. 自动化学报, 2018, 44(2): 311-317.
- [51] 余欢欢, 魏文戈. 模型参数自适应迁移的多源域适应[J]. 计算技术与自动化, 2020,(4): 87-90.
- [52] Huang J, Gretton A, Borgwardt K, et al. Correcting sample selection bias by unlabeled data[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2006, 601-608.
- [53] Dai W, Yang Q, Xue G, et al. Boosting for transfer learning[C]. International Conference on Machine Learning (ICML), 2007, 193-200.
- [54] Zhang J, Ding Z, Li W, et al. Importance weighted adversarial nets for partial domain adaptation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 8156-8164.
- [55] Cao Z, Long M, Wang J, et al. Partial transfer learning with selective adversarial networks[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 2724-2732.
- [56] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2010, 22(2): 199-210.
- [57] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks[C]. International Conference on Machine Learning (ICML), 2015, 97-105.
- [58] Sun B, Feng J, Saenko K. Return of frustratingly easy domain adaptation[C]. AAAI Conference on Artificial Intelligence (AAAI), 2016, 2058-2065.
- [59] Sun B, Saenko K. Deep coral: Correlation alignment for deep domain adaptation[C]. European Conference on Computer Vision (ECCV), 2016, 443-450.
- [60] Tasar O, Happy S L, Tarabalka Y, et al. ColorMapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7178-7193.
- [61] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C]. International Conference on Machine Learning (ICML), 2015, 1180-1189.
- [62] Isola P, Zhu J, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 1125-1134.
- [63] Ma Y, Wu H, Wang L, et al. Remote sensing big data computing: Challenges and opportunities[J]. Future Generation Computer Systems, 2015, 51: 47-60.

- [64] Li S, Dragicevic S, Castro F A, et al. Geospatial big data handling theory and methods: A review and research challenges[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 115: 119-133.
- [65] Shi H, Chen L, Bi F, et al. Accurate urban area detection in remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(9): 1948-1952.
- [66] Kampffmeyer M, Salberg A, Jenssen R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, 1-9.
- [67] Kussul N, Lavreniuk M, Skakun S, et al. Deep learning classification of land cover and crop types using remote sensing data[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5): 778-782.
- [68] Ozdarici-Ok A, Ok A O, Schindler K. Mapping of agricultural crops from single high-resolution multispectral images—Data-driven smoothing vs. parcel-based smoothing[J]. Remote Sensing, 2015, 7(5): 5611-5638.
- [69] Yu B, Yang L, Chen F. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(9): 3252-3261.
- [70] Zhang Y, Lu Y, Zhang D, et al. Risksens: A multi-view learning approach to identifying risky traffic locations in intelligent transportation systems using social and remote sensing[C]. 2018 IEEE International Conference on Big Data (Big Data), 2018, 1544-1553.
- [71] Li Y, Zhang Y, Huang X, et al. Large-Scale Remote Sensing Image Retrieval by Deep Hashing Neural Networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(2): 950-965.
- [72] Li Y, Zhang Y, Huang X, et al. Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 146: 182-196.
- [73] Huang X, Zhu Z, Li Y, et al. Tea garden detection from high-resolution imagery using a scene-based framework[J]. Photogrammetric Engineering & Remote Sensing, 2018, 84(11): 723-731.
- [74] Li Y, Zhang Y, Zhu Z. Error-Tolerant Deep Learning for Remote Sensing Image Scene Classification[J]. IEEE Transactions on Cybernetics, 2020, 51(4): 1756-1768.
- [75] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 3431-3440.
- [76] Hu W, Huang Y, Wei L, et al. Deep convolutional neural networks for hyperspectral image classification[J]. Journal of Sensors, 2015: 1-12.
- [77] Yue J, Zhao W, Mao S, et al. Spectral--spatial classification of hyperspectral images using deep convolutional neural networks[J]. Remote Sensing Letters, 2015, 6(6): 468-477.
- [78] Makantasis K, Karantzalos K, Doulamis A, et al. Deep supervised learning for hyperspectral

- data classification through convolutional neural networks[C]. International Geoscience and Remote Sensing Symposium (IGARSS), 2015, 4959-4962.
- [79] Mi L, Chen Z. Superpixel-enhanced deep neural forest for remote sensing image semantic segmentation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 140-152.
- [80] Zhang L, Zhang L, Du B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art[J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2): 22-40.
- [81] Bruzzone L, Carlin L. A multilevel context-based system for classification of very high spatial resolution images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(9): 2587-2600.
- [82] Tuia D, Persello C, Bruzzone L. Domain adaptation for the classification of remote sensing data: An overview of recent advances[J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2): 41-57.
- [83] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]. International Conference on Machine Learning (ICML), 2017, 1857-1865.
- [84] Yi Z, Zhang H, Tan P, et al. Dualgan: Unsupervised dual learning for image-to-image translation[C]. International Conference on Computer Vision (ICCV), 2017, 2849-2857.
- [85] Martin A, Soumith C, Leon B. Wasserstein generative adversarial networks[C]. International Conference on Machine Learning (ICML), 2017, 214-223.
- [86] Laine S, Aila T. Temporal ensembling for semi-supervised learning[C]. International Conference on Learning Representations (ICLR), 2016, 1-13.
- [87] Lee D. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]. Workshop on challenges in representation learning, ICML, 2013, 896.
- [88] ISPRS Test Project on Urban Classification, 3D Building Reconstruction and Semantic Labeling - 2D Semantic Labeling Contest[EB/OL].
<https://www.isprs.org/education/benchmarks/UrbanSemLab/default.aspx>.
- [89] Benjdira B, Bazi Y, Koubaa A, et al. Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images[J]. Remote Sensing, 2019, 11(11): 1369.
- [90] Fang B, Kou R, Pan L, et al. Category-sensitive domain adaptation for land cover mapping in aerial scenes[J]. Remote Sensing, 2019, 11(22): 2631.
- [91] Zheng Z, Yang Y. Unsupervised Scene Adaptation with Memory Regularization in vivo[C]. International Joint Conference on Artificial Intelligence (IJCAI), 2020, 1076-1082.
- [92] Luo Y, Zheng L, Guan T, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 2507-2516.
- [93] Olsson V, Tranheden W, Pinto J, et al. Classmix: Segmentation-based data augmentation for semi-supervised learning[C]. IEEE Winter Conference on Applications of Computer Vision

- (WACV), 2021, 1369-1378.
- [94] Lee K, Ros G, Li J, et al. SPIGAN: Privileged Adversarial Learning from Simulation[C]. International Conference on Learning Representations (ICLR), 2019, 1-14.
- [95] Xu D, Ouyang W, Wang X, et al. Pad-net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 675-684.
- [96] Zamir A R, Sax A, Shen W, et al. Taskonomy: Disentangling task transfer learning[C]. IEEE Conference on Computer Cision and Pattern Recognition (CVPR), 2018, 3712-3722.
- [97] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2017, 1195-1204.
- [98] Wang J, Zheng Z, Ma A, et al. LoveDA: A remote sensing land-cover dataset for domain adaptive semantic segmentation[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2021, 1-16.
- [99] Heusel M, Ramsauer H, Unterthiner T, et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2017, 6626-6637.
- [100] Zheng A, Wang M, Li C, et al. Entropy Guided Adversarial Domain Adaptation for Aerial Image Semantic Segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.
- [101] Tzeng E, Hoffman J, Zhang N, et al. Deep domain confusion: Maximizing for domain invariance[J]. arXiv:1412.3474, 2014: 1-9.
- [102] Wang H, Shen T, Zhang W, et al. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation[C]. European Conference on Computer Vision (ECCV), 2020, 642-659.
- [103] Wang X, Jin Y, Long M, et al. Transferable normalization: Towards improving transferability of deep neural networks[C]. Annual Conference on Neural Information Processing Systems (NeurIPS), 2019, 1951-1961.
- [104] Lian Q, Lv F, Duan L, et al. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach[C]. International Conference on Computer Vision (ICCV), 2019, 6758-6767.
- [105] Mei K, Zhu C, Zou J, et al. Instance adaptive self-training for unsupervised domain adaptation[C]. ECCV, 2020, 415-430.
- [106] Rana L, Ji C, Zhang S, et al. An Unsupervised Domain Adaption Framework for Aerial Image Semantic Segmentation Based on Curriculum Learning[C]. International Conference on Image, Vision and Computing (ICIVC), 2022, 354-359.
- [107] Van der Maaten L, Hinton G. Visualizing data using t-SNE.[J]. Journal of Machine Learning Research, 2008, 9(11): 2579-2605.

攻读硕士学位期间发表的科研成果

- [1] Li Y, **Shi T**, Zhang Y, et al. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175: 20-33.
(中科院一区 Top, IF=11.774, ESI 高被引论文。导师一作, 学生二作)
- [2] Li Y, **Shi T**, Zhang Y, et al. SPGAN-DA: Semantic-preserved generative adversarial network for domain adaptive remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023. (Major Revision)
- [3] **Shi T**, Li Y, Zhang Y. Rotation consistency-preserved generative adversarial networks for cross-domain aerial image semantic segmentation[C]. IEEE International Geoscience and Remote Sensing Symposium, 2021: 8668-8671. (学生论文竞赛 Final lists)
- [4] Li Y, **Shi T**, Chen W, et al. Unsupervised style transfer via dualgan for cross-domain aerial image classification[C]. IEEE International Geoscience and Remote Sensing Symposium, 2020: 1385-1388. (导师一作, 学生二作)

攻读硕士学位期间参与的课题情况

1. 阿里巴巴达摩院（横向协作项目），海量遥感影像场景下半监督和无监督图像识别技术的研究。
2. 国家自然科学基金委员会，国家自然科学基金面上项目，41971284，深度知识学习驱动的多源遥感影像场景分类及时空广义模型迁移，2020.01 至 2023.12，61 万元。

致 谢

行文至此，落笔为终。十余载漫漫求学路，即将画上句号。一个篇章结束，另一个篇章待启。回望珞珈山下的点点滴滴，心中思绪万千、感慨良多。

生逢盛世，与有荣焉。我要感谢伟大的党和祖国，有幸生于华夏，让我们可以不经战乱不缺衣食，可以拥有平等接受教育、改变命运的机会。感谢武汉大学遥感信息工程学院对我的培养，山水一程，三生有幸。

涓涓师恩，铭记于心。我要由衷的感谢我的导师李彦胜副教授。从论文选题、开题到定稿，老师无数次悉心指导，让我的论文得以顺利完成。在读研期间，何其有幸遇见李老师，还记得初见时，老师幽默随和的言语、亲切爽朗的笑容，让我有了莫大的安全感与归属感。每每与老师交流时，总能畅所欲言，亦师亦友。每每遇到困难、迷茫无所时，老师都以鼓励为主，总能让我重拾希望，继续努力。可惜我离老师的期望还有很大的差距，在此深表歉意！恩师乐观从容、严谨谦虚的生活和学术态度，都深深影响和鼓舞着我，将使我受益终身。在未来的日子里，祝愿老师能万事胜意、处处惊喜！感谢求学过程中遇到的每一位老师，愿你们万事顺遂，桃李芬芳。

愿岁并谢，与长友兮。感谢单炜琦和王开诚，在我参加夏令营时全程陪同我、鼓励我，如果没有你们的帮助，我也许无法来到武大深造。今年也将是我和王开诚认识的第十个年头，能够一起在武大读书、一起毕业，真是泰裤辣！感谢陈锦荣，在逆境中鼓励我，在困难时陪伴我，我们一起探索了许多新鲜事物，有着许多美妙的体验，希望我们继续谱写美好的诗篇，一起度过漫长岁月！感谢朱忠涛，你是我的挚友之一，你无私地帮助我、支持我，我都铭记于心！希望未来我们有机会多见面，一起旅游、一起猎奇。感谢宋金洁，我们每天都会互相分享生活，一起讨论问题、互相鼓励，希望我们友谊长存，继续保持这种分享欲。感谢柳力元，在我遇到困难或疑惑时，总喜欢找你倾诉，你总能帮分析情况，给予我帮助或给出合理的建议，愿你开心快乐。感谢曾畅、陈寅秋、齐志军，与你们抢体育馆、打羽毛球的日子真的很酣畅淋漓，期间球技大涨，逐渐掌握了高、吊、杀、搓、推、勾、扑、放等技能，希望未来的日子我们仍能有机会切磋球技。感谢赵迪，我们总是会一起远程观看羽毛球比赛，为国羽运动员的失意而沮丧，为国羽运动员的突破而激动，希望我们有机会可以现场观看羽毛球比赛。感谢在北京蔚来实习期间我的 mentor 闵捷对我的帮助，带我熟悉业务、教会我很多技能、给予我就业方面的建议与规划等，希望你工作顺利，有更多的时间做自己想做的事。感谢肖敏珍对我的帮助，在我刚去公司实习时带我了解工作情况、熟悉业务、陪我 debug。感谢实验室的伙伴们，与你们共同学习、共同进步的日子充实而又美

好！感谢已经毕业的孔德宇师兄、周煜晗师姐对我的帮助，你们在学习和生活上帮助了我很多，与你们相处的日子很开心快乐。感谢陈蔚师兄、王琳琳师姐帮我修改完善论文，提出了许多宝贵的建议。感谢魏凡漪师妹和党博师弟，你们总是与我分享新鲜的事物、畅谈人生。感谢大家的“好朋友”——中国国家羽毛球队女子单打运动员陈雨菲，你坚忍不拔不放弃的精神一直激励着我。你说“关关难过关过关”，我想说“前路漫漫亦灿灿”！希望你能够快乐羽毛球，在世锦赛夺金实现大满贯、在巴黎奥运会夺金实现卫冕，加油我的女战士！很幸运自己一路走来，在每个阶段都遇到了那么多美好、真温暖、合拍的人，他们伴我成长给我力量，见证彼此的青葱岁月，我大多数的快乐都来源于他们，或者说因为他们，我大多数都是快乐的，也正是他们，让我变得自信且阳光、快乐且向上，让这平淡如水的日子熠熠生辉，让我觉得这个世界始终充满美好。不管我们今后会继续陪伴，还是各奔东西，我都要感谢他们，温暖着我的每一刻时光。

春晖寸草，山高海深。感谢父母二十余载的培养和教育，从小到大，是你们的支持和培养，让我有直面未来的勇气和坚定的目标。是父亲的严格要求和耐心鼓励、母亲的细致关心和暖心照顾，让我能在风雨中不断前进、不断突破自己。你们用朴素又真挚的爱陪伴我走过人生的每一阶段，让我在健康、快乐、幸福的家庭环境中得以成长，让我站在你们的肩膀上看到了更大的世界、拥有了更广阔的天空。二十多年的养育之恩，无以回报，唯有不断努力，才能成为你们的骄傲！未来我一定会成为你们的避风港！

道阻且长，行则将至。我还要感谢始终坚持、始终向上、始终热烈、始终勇敢的自己。一路走来，煎熬过、崩溃过、迷茫过，但都从未放弃，跌跌撞撞的努力前进着，正是那些奋笔疾书、刻苦拼搏的日夜，无数自我治愈、自我坚持的时刻，使之成长为现在乐观、从容、自信、爱笑的自己。愿自己多年以后归来，仍是珞珈一少年。

凡是过往，皆为序章。始终相信“关关难过关过关，前路漫漫亦灿灿”，我将带着“自强、弘毅、求是、拓新”的珞珈精神，奔赴下一片山海！再见了，武汉大学！再见了，我美好的学生时代！希望我和我的武汉大学越来越好！

最后，衷心感谢论文评审专家、答辩专家在百忙之中抽出时间评阅论文和参加答辩！

武汉大学学位论文使用授权协议书

(一式两份，一份论文作者保存，一份留学校存档)

本学位论文作者愿意遵守武汉大学关于保存、使用学位论文的管理办法及规定，即：学校有权保存学位论文的电子版和印刷本，并提供文献检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在以教学与科研服务为前提下，学校可以向本校认证用户公布部分及全部内容。

- 1、 在本论文提交当年，同意向本校认证用户以及中国高等教育文献保障系统（CALIS）之高校学位论文系统提供查询及前十六页浏览服务。
- 2、 在本论文提交~~□当年/□一年/□两年/□三年以后~~，允许本校认证用户在线浏览全文，学校可以为存在馆际合作关系的图书馆提供全文文献传递服务。

论文作者（签名）： 史哲

学 号： 2020102130028

学 院： 遥感信息工程学院

日期：2023年5月24日