

分类号 \_\_\_\_\_

密级 \_\_\_\_\_

U D C \_\_\_\_\_

编号 \_\_\_\_\_

武汉大学

硕士学位论文

基于深度学习的高分辨率遥感影像道路  
提取研究

研究生姓名：魏瑶

学号：2018202130072

指导教师姓名、职称：季顺平 教授

专业名称：摄影测量与遥感

研究方向：深度学习与遥感影像地物提取

二〇二一年五月

Thesis for Master Degree

RESEARCH ON ROAD EXTRACTION  
FROM HIGH-RESOLUTION REMOTE  
SENSING IMAGE WITH DEEP LEARNING

By  
Wei Yao

Supervised by  
Ji Shunping

**May, 2021**

School of Remote Sensing and Information Engineering,  
Wuhan University, Wuhan 430079, China

## 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经发表或撰写过的研究成果。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）：

年 月 日

## 摘要

准确的道路地图在诸多应用中十分重要。大数据时代，遥感技术的发展为道路提取带来了机遇与挑战。由于道路本身及背景环境的复杂性，传统基于经验设计特征的提取方法在精度和效率方面难以满足实际需要。近年来，以卷积神经网络为代表的深度学习在遥感领域的应用中展现出强大的潜力，为道路提取提供了新的机遇。本文追踪国内外遥感与计算机视觉在深度学习领域的最新研究成果，关注基于深度学习的高分辨率遥感影像道路提取，包括用于道路语义分割和拓扑追踪的全监督学习方法和用于道路路面提取的弱监督学习方法。

首先，考虑到道路提取研究领域主流的深度学习方法尚有不足，尤其是在道路提取的拓扑连通性与完整性上有待改善。本文提出一种结合语义分割和拓扑追踪的道路提取算法，基于多个不同的卷积神经网络实现道路路面和道路中心线的同时提取。该算法采用多阶段框架，包括道路提升分割、道路多起点追踪、以及二者的融合，旨在提高道路提取的拓扑连通性和完整性。在道路分割方面，该算法比仅利用语义分割的先进算法高 7% 的连通性；在道路中心线提取方面，该算法比仅利用拓扑追踪的先进算法高 40% 的完整性。

其次，考虑到现有大多数道路提取算法都采用全监督学习，依赖于大量人工标注数据，这与自动化提取道路的目标仍有差距。由于道路路面数据的标注工作费时费力，而类似道路中心线的线状数据相比之下更易获取。本文借鉴弱监督学习在计算机视觉领域的进展，提出一种基于弱监督深度学习的道路路面提取算法。该算法利用道路属性与超像素图割，由线状弱标签生成道路分割代理标签，随后利用分割代理标签和辅助的边缘先验信息来训练用于路面分割的双分支网络。该算法与经典的线状弱监督学习算法相比 IoU 提高 20%，还超出了最新的相关研究工作 4%，它能够从稀疏的线状弱标签中学习，而不再需要密集标注的道路路面标签，是自动化道路提取研究的重要一步。

本文提出的创新性方法有望促进道路提取智能化与自动化水平的提高，具有很好的理论研究和实际应用前景。

**关键词：**遥感影像道路提取；深度学习；语义分割；拓扑追踪；弱监督学习

## Abstract

Accurate road maps are of great importance in a wide range of applications. In the era of big data, the development of remote sensing technology has brought opportunities and challenges to road extraction. Due to the diversity of the road and background, traditional hand-craft features-based extraction methods are impractical in terms of accuracy and efficiency. In recent years, deep learning represented by convolutional neural networks has shown great potential in remote sensing applications, which offer a promising avenue for road extraction. In this article, we track the latest research of deep learning applied in remote sensing and computer vision at home and abroad; and we focus on extracting road from high-resolution remote sensing images based on deep learning, including fully supervised learning-based road semantic segmentation and topologic tracing as well as weakly supervised learning-based road surface extraction.

First, considering the limitations of the mainstream road extraction methods, especially on topological connectivity and completeness, in this article, we propose a road extraction algorithm that combines the advantages of semantic segmentation and topologic tracing solutions to achieve simultaneously road surface and road centerline extraction. This algorithm contains three main steps: road boosting segmentation, road multiple starting points tracing, and the fusion process, which aims to improve the topological connectivity and completeness of road extraction. Regarding road segmentation, the proposed algorithm exceeds the state-of-the-art segmentation methods by 7% for road connectivity. With respect to road centerline extraction, the algorithm outperforms the state-of-the-art tracing methods by 40% for road completeness.

Second, considering most existing road extraction methods employed fully supervised learning and relied on large amounts of manually annotated data. As the annotation of road surface data is time-consuming and labor-intensive, the scribbles (such as road centerlines) are more accessible. Inspired by the latest advances of weakly supervised learning in computer vision fields, in this article, we propose a weakly

supervised deep learning algorithm for road surface extraction. By utilizing road attributes and super-pixel segmentation, the road segmentation proposal masks are generated from scribbles. Then, a two-branch neural network is trained with proposal masks and auxiliary edge information for road surface segmentation. This algorithm outperforms the classic scribble-supervised segmentation method by 20% for the intersection-over-union (IoU) indicator and also exceeds recent related research work by 4%. It can be an important step in automatic road extraction research as it learns from sparse scribbles without the need of densely labeled road surface annotations.

These innovative algorithms proposed in this article are expected to promote intelligence and automation of road extraction, and would provide good prospects for theoretical research and practical applications.

**Key words:** Road extraction from remote sensing image; Deep learning; Semantic segmentation; Topologic tracing; Weakly-supervised learning

# 目录

摘要.....	I
Abstract .....	II
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.....	2
1.3 研究内容.....	6
1.4 论文章节安排.....	7
第 2 章 深度学习基础理论与技术.....	9
2.1 卷积神经网络.....	9
2.2 全卷积神经网络.....	10
2.3 集成学习.....	12
2.4 弱监督学习.....	13
2.5 本章小结.....	14
第 3 章 道路提取实验准备.....	15
3.1 数据.....	15
3.2 性能评估.....	21
3.3 本章小结.....	24
第 4 章 结合语义分割和拓扑追踪的道路提取算法.....	25
4.1 引言.....	25
4.2 基于全卷积神经网络的道路提升分割.....	27
4.3 基于卷积神经网络的道路多起点追踪.....	34
4.4 道路分割和追踪的融合.....	39
4.5 实验分析与讨论.....	44
4.6 本章小结.....	47
第 5 章 基于弱监督深度学习的道路路面提取算法.....	48
5.1 引言.....	48
5.2 基于缓冲区推理和超像素图割的道路标签传播.....	50
5.3 用于弱监督道路分割的边缘感知双分支网络.....	54
5.4 实验分析与讨论.....	58
5.5 本章小结.....	62

第 6 章 总结与展望.....	63
6.1 全文总结.....	63
6.2 研究展望.....	64
参考文献.....	66
攻读硕士期间主要成果.....	71
致谢.....	73

武汉大学



# 第 1 章 绪论

## 1.1 研究背景与意义

遥感技术作为一种大范围、远距离感知的对地观测技术，通过获取研究区域的影像实现对地理信息的快速收集，结合相关遥感数据处理技术对这些信息进行定性或定量的分析，为经济社会发展和国家重大战略实施提供有力的数据和技术支持<sup>[1][2]</sup>。近年来，随着遥感技术的不断推进，遥感数据获取能力明显提升，主要表现在影像来源更丰富，时效性更强，分辨率更高。高分辨率遥感影像的数据量激增，使得处理分析这些影像数据的需求也大大增加。然而，相对于强大的遥感数据获取能力，当前遥感影像处理理论和方法还存在种种制约，遥感影像的智能处理和服务能力仍比较滞后<sup>[3]</sup>。在地物要素提取方面，迫切需要能够自动化实现遥感影像地物提取的可靠方法。

道路是具有复杂拓扑关系的典型地物要素，在城市规划、应急响应等方面发挥着关键作用，而且道路分布还是遥感影像场景理解研究中的重要因素，能够为建筑物、河流、植被及其他地物的识别提供先验知识<sup>[4]</sup>。但是时至今日，世界上仍有数百万公里甚至更多的道路没有录入到地图中，尤其在乡村偏僻的地区有大量的道路未进行数字化，即使在城市中也有许多道路不能及时更新。随着基于位置的移动服务以及自动驾驶技术的推广应用，更新及时的高精度道路地图显得愈加重要。然而，由于遥感影像中道路自身和背景环境的多样性、复杂性，从遥感影像中提取道路网一直是长期存在的挑战。遥感影像分辨率的不断提高，随之更多的道路目标与更丰富的细节特征为道路提取提供了更为有利的条件，与此同时也带来了诸多干扰因素。例如，受光照影响，影像中存在较多的行道树、高大建筑物、车辆的遮挡和阴影，使得针对中低分辨率遥感影像的提取方法不适用于高分辨率遥感影像的道路提取。另外，人工经验设计特征的传统道路提取方法对于场景相对简单的遥感影像能够取得良好的提取效果，却很难应对同物异谱、同谱异物等复杂场景带来的混淆，无法生成高质量的道路专题图，不能直接满足遥感影像产品大规模生产与应用的要求。

随着遥感与人工智能和机器学习的交叉融合，深度学习被引入遥感影像地物

提取研究中并取得了可喜进展<sup>[5]</sup>。人工智能是计算机科学的一个分支，研究如何让计算机模拟人类的思维和行为去感知环境。从二十一世纪初开始，在大数据驱动和高性能计算设备支持下，人工智能受到了广泛重视，特别是机器学习中的深度学习<sup>[6]</sup>已然成为研究热点，在计算机视觉、自然语言处理等任务中表现出色。其中，卷积神经网络（Convolutional Neural Network, CNN）能够自动学习将原始输入映射到指定标签或连续向量的表示形式，其表示学习的能力逐渐取代了传统的经验设计特征的方式，成为图像分类<sup>[7]</sup>、目标检测<sup>[8]</sup>等图像处理领域的主流模型，为遥感影像道路提取提供了新途径。

借鉴深度学习在图像处理方面的进展，研究人员围绕应用于道路提取的深度学习方法，改进和提出了颇多表现良好的模型和算法，可大致归纳为语义分割方法和拓扑追踪方法。道路分割的特点是逐像素预测类别，而道路追踪的特点是以道路拓扑结构为主体迭代式搜索道路网。前期研究发现，现有算法往往只关注其中任一类方法，即仅使用语义分割或仅使用拓扑追踪，导致其中任一类方法的上限在一定程度上制约了性能的大幅提升。因此，亟待研究结合语义分割和拓扑追踪的道路提取算法，通过两类方法的相互合作实现优势互补。

尽管到目前为止，深度学习已成功应用于道路提取任务并取得了较为丰硕的研究成果，但针对低标注成本的遥感影像道路自动化提取研究仍是富有挑战性的课题。大部分现有道路提取算法均为数据驱动的全监督学习，而道路路面标签数据的标注工作费时费力，大大限制了全监督学习方法在道路路面提取方面的泛化能力。仅有的基于弱监督学习的道路提取算法设计过于简单，导致难以与真实道路路面语义对齐。因此，亟待研究借助公开地图平台或少量标签数据实现道路路面提取的更为鲁棒的新方法。

基于上述背景，本课题将在深度学习下进行高分辨率遥感影像道路提取的相关研究，以期开发在精度和自动化程度等方面更为成熟有效的道路提取算法。

## 1.2 国内外研究现状

自二十世纪七十年代以来，国内外对遥感影像道路提取的研究一直在进行。道路提取是一个与数学、测绘、计算机视觉等多学科紧密关联的研究领域，由于各学科的迅速发展，道路提取方面的研究在二十世纪九十年代之后进入了高速发展阶段。

传统上, 遥感影像道路提取的工作集中在经验设计特征, 通常利用道路的光谱、形状、长度、纹理等特征进行道路提取<sup>[9][10]</sup>, 由于这些指标可能会随着季节、传感器质量等因素的变化而发生明显变化, 人工参与设计的特征常常不能完整、恰当地概括道路本身的信息, 导致道路提取精度很难有更大突破。基于初始聚类的提取方法<sup>[11][12]</sup>以同质区域而非像素为处理单元进行道路提取, 提高了效率, 但在停车场、裸地等复杂场景中会产生难以避免的混淆像素。结合先验知识的提取方法<sup>[13][14]</sup>利用多源数据融合进行道路提取, 但由于受实际应用中辅助数据匮乏的限制, 这类方法也并不是易再现的通用方法。

近年来, 研究人员开始利用深度学习探究遥感影像道路提取, 借助 GPU 并行计算能力, 提出了诸多基于 CNN 网络的模型及算法, 精度远超传统算法的结果。CNN 利用多个卷积层逐级提取特征, 能够从影像中自动学习复杂的非线性特征表示, 突破了传统方法的局限。经典 CNN 如 AlexNet<sup>[7]</sup>、VGGNet<sup>[15]</sup>、GoogleNet<sup>[16]</sup>、ResNet<sup>[17]</sup>等架构使用卷积层后的全连接层将特征图映射为固定长度的特征向量, 在图像分类任务上效果颇佳。自 2015 年起, 一种称为全卷积神经网络 (Fully Convolutional Network, FCN)<sup>[18]</sup>的特殊 CNN 架构得到发展, 进一步扩展了低维特征与高维特征融合的思想。FCN 采用卷积层取代了普通 CNN 中的全连接层, 并对 CNN 最后的特征图进行上采样操作生成与输入图像尺寸一致的输出, 从而为图像中所有像素都赋予类别。这种编码器 (下采样)-解码器 (上采样) 结构的 FCN 包括 DeconvNet<sup>[19]</sup>、U-Net<sup>[20]</sup>、DeepLab<sup>[21]</sup>等变种, 已成功应用于图像语义分割任务。

目前, 利用以 CNN 为代表的深度学习方法从遥感影像中提取道路的研究包括两个子任务, 分别是道路分割<sup>[4][22][23][25][27]</sup>和道路中心线提取<sup>[29][30][31][32][33]</sup>。前者通常被视为二分类问题, 即道路为前景, 其他非道路为背景, 利用 FCN 网络对影像进行逐像素分类, 预测每个像素属于道路的概率, 生成二值的道路分割图。后者旨在以中心线形式记录道路网的位置和结构, 由于直接从遥感影像中提取道路中心线难度较大, 通常将道路分割作为中间过程, 而更新的方法能够直接基于影像得到结构化的中心线。

现有道路分割方法在 CNN (特别是 FCN) 的基础上, 改进和设计了若干新的网络模型。文献<sup>[22]</sup>结合 U-Net 与 ResNet 用于道路提取, 简化深层网络训练的同时利用跳跃连接实现模型内部更好地传输信息。文献<sup>[23]</sup>将空洞卷积空间金字

塔池化 (Atrous Spatial Pyramid Pooling, ASPP)<sup>[24]</sup>与编码-解码结构网络相结合,更精细地提取道路特征。文献[25]在 U-Net 基础上设计了一个循环卷积神经网络单元,以探索有利于道路提取的空间上下文信息。基于 LinkNet<sup>[26]</sup>和空洞卷积 (Dilated convolution),文献[27]开发了一个名为 D-LinkNet 的编码器-解码器网络,借助更大的感受野并融合低级、高级语义特征从卫星影像中检测道路,赢得了 2018 年国际计算机视觉与模式识别会议中 DeepGlobe<sup>[28]</sup>道路提取挑战赛 (CVPR DeepGlobe Road Extraction Challenge) 的第一名。这些道路分割方法能够检测到影像中的大部分道路,但由于逐像素分类的方式忽略了道路的拓扑特性使得道路连接性较差。

现有道路中心线提取方法<sup>[29][30]</sup>一般包括两个步骤:先检测道路,然后提取中心线。有研究将这两个步骤合并为端到端的神经网络学习过程,Cheng 等<sup>[31]</sup>提出串行模式的 CNN 用于同时道路检测和中心线提取,Lu 等<sup>[32]</sup>在并行模式下提出多尺度特征融合的道路检测网络和中心线提取网络。文献[31]和[32]共同的缺点是,对于前阶段道路分割结果的依赖性,使得检测道路和提取中心线两个步骤都是像素级分割,而未考虑道路任何拓扑或结构相关信息。与上述借助道路分割的中心线提取方法不同,文献[33]开创性提出了一种名为 RoadTracer 的拓扑追踪方法,能够直接基于影像生成结构化的道路中心线。该方法从一个已知点出发,通过基于 CNN 的决策函数引导局部搜索窗口在道路上移动,迭代式搜索并逐段构建道路中心线。RoadTracer 虽保留了道路的拓扑连接性,但由于搜索容易受到桥梁等影响而中断使得道路完整性较差。

上述道路提取的深度学习方法多为全监督学习,训练神经网络时需要大量标签数据来优化参数,但创建例如 ImageNet<sup>[34]</sup>、COCO<sup>[35]</sup>这样大型样本数据库的人工标注成本较高。随着自发地理信息 (Volunteered Geographic Information, VGI) 平台的迅速发展,成千上万的用户能够在线创建和编辑地理数据,开源了愈来愈多的道路样本数据。作为最常见 VGI 之一,公开地图 OpenStreetMap (OSM)<sup>[36]</sup>提供了全球范围内大部分道路中心线的坐标位置,可被用作道路中心线提取任务中的标签。然而,OSM 一般不提供含有路宽信息的道路路面数据,无法为道路分割 (尤其是精确的路面分割) 任务提供公开可用的标签。考虑到目前测绘员的标注工作仍为先目视判读再利用计算机量测和描绘,这是一个相当耗时耗力的繁琐过程,所以实际应用中常出现道路路面标签不足或不准确的情况,大大限制了

全监督学习方法在道路路面提取方面的泛化能力。

为了解决全监督学习方法对逐像素标注的路面标签的依赖问题,弱监督学习技术为遥感影像道路路面提取提供了广阔的前景。与使用逐像素标注的全监督学习不同,弱监督学习是从稀疏或不完整的弱标签中学习提取有价值的特征,能够有效降低标注成本。针对基于弱监督学习的语义分割,常见的弱标签包括点状(click)<sup>[37]</sup>、线状(scribble)<sup>[38]</sup>、边界框(bounding box)<sup>[39]</sup>、图像级别(image-level)<sup>[40]</sup>。经典的弱监督学习往往采用交替优化的训练方案,所谓的“交替优化”即,首先由弱标签提供的稀疏信号生成“建议标签”,其次利用建议标签通过标准损失函数(例如交叉熵)监督 CNN 网络的训练,接着 CNN 的预测结果参与建议标签生成过程以改善建议标签的质量,如此在建议标签生成和网络训练两个步骤之间交替迭代,力求经过几轮迭代后提升 CNN 的性能。Lin 等<sup>[38]</sup>就是采用这种交替优化方案,提出了一种基于 scribble 弱监督深度学习的经典语义分割算法,命名为 ScribbleSup。尽管交替优化的方式提高了建议标签的质量,但由于该算法在建议标签生成时认为每个超像素内部的像素对应标签一致,限制了建议标签的准确度,而且标准损失函数很难区分建议标签中的错误标注,导致 CNN 在不同类别的目标边缘处不可避免地产生混淆像素。此外,交替优化的方式计算量相对较大。

目前,弱监督语义分割方面的研究开始关注端到端的方式而非交替优化,主要采用包括基于图论和基于边界的正则化操作,由稀疏的弱标签扩展更丰富的语义信息。Tang 等<sup>[41][42]</sup>引入了只在标记像素上计算的部分交叉熵损失,还引入了基于图割(graph cut)、密集 CRF(dense CRFs)等正则化技术的损失。Obukhov 等<sup>[43]</sup>提出了针对未标记像素的门控 CRF 损失,并协同针对标记像素的部分交叉熵损失监督分割网络。Kolesnikov 和 Lampert<sup>[44]</sup>应用一种边界约束规则来恢复弱监督分割的细节信息。Wang 等<sup>[45]</sup>设计了一个网络架构将边界直接嵌入分割网络中,该架构包括预测细化网络(Prediction Refinement Network, PRN)和边界回归网络(Boundary Regression Network, BRN),其中 BRN 用来指导 PRN 更好地定位边界。通过引入辅助的边缘检测网络,Zhang 等<sup>[46]</sup>基于 scribble 设计了一种仅关注图像显著目标所在区域的结构感知门控损失。

受启发于弱监督学习技术在语义分割研究方面的应用,特别是基于 scribble 弱监督深度学习方法的发展,类似 OSM 中心线的 scribble 可作为道路路面提取

的弱标签。但由于遥感影像的复杂性远超自然图像，将弱监督学习技术直接应用到遥感影像道路提取的研究极少。有几项早期研究<sup>[47][48][49]</sup>尝试借助 OSM 道路中心线数据进行航空影像道路提取，但都只限于将其视为简单的宽度估计问题。由于 OSM 道路数据中不仅包含中心线的坐标位置，同时还包含道路类型、车速等信息，Kaiser 等<sup>[49]</sup>训练了一个用于从航空影像提取建筑物和道路的 CNN，其中，道路标签由 OSM 中每个道路类别（例如高速公路、主干路、支路等）的平均道路宽度来确定。近期，Wu 等<sup>[50]</sup>提出一种道路提取的弱监督深度学习方法，基于 OSM 中心线生成代理的道路标签，利用代理标签监督由 FCN 改进的道路分割网络。显然，这些现有方法的共同缺点是经验性的简单策略仅能处理特定数据，在大规模复杂数据上应用时将存在局限性，难以得到与真实道路语义对齐的路面预测结果。

综上所述，从高分辨率遥感影像中提取道路的研究目前尚未完善，特别是在精度和自动化程度上。一方面，国内外主流的道路提取方法可归纳为道路语义分割和拓扑追踪，两类方法中的先进算法在道路识别精度上均需要进一步提高，尤其是道路拓扑连通性差、道路完整性差这两个关键问题有待解决。另一方面，利用低标注成本实现道路路面准确提取的研究至关重要，是“自动化”提取道路的含义所在，在实际生产中的应用价值极高，但是目前这方面的研究工作仍处于起步阶段，现有算法并未充分发挥神经网络从稀疏信号中挖掘目标潜在信息的能力。

### 1.3 研究内容

本文利用深度学习对高分辨率遥感影像进行道路提取，包括基于全监督学习的结合语义分割和拓扑追踪的道路提取、基于弱监督深度学习的道路路面提取。在不同数据源、不同国家和地区的道路提取数据集上实验，通过消融分析证明本文提出方法的有效性，并与现有先进方法对比验证所提出方法的先进性。

针对现有道路提取的精度和自动化程度有限等问题，本文提出一系列创新性方法，具体研究包括以下几个方面：

- 1) 为了解决已有道路提取方法中存在的道路结果不完整、拓扑连通性差等问题，提出一种结合语义分割和拓扑追踪的道路提取算法，基于多个不同的 CNN，分别在语义分割和拓扑追踪两类方法的先进算法上改进，并使二者相互合作、互为补充，与仅利用语义分割或拓扑追踪的方法对比，有效提高了道路提取的完整

性和拓扑连通性；

2) 为了解决全监督学习方法对逐像素标注的路面标签依赖问题, 提出了一种基于 FCN 弱监督学习的遥感影像道路路面提取算法, 仅在易获取的类似中心线的 scribble 弱标签监督下即可实现道路路面分割, 与基于全监督学习的道路提取对比, 该算法不需要路面标签, 大大降低了标注成本, 具有较高的应用价值。其性能远超利用 scribble 弱监督深度学习的经典算法 ScribbleSup, 也优于最近的几项相关研究工作, 在复杂场景下表现更为鲁棒, 有效提高了道路提取的自动化水平。

## 1.4 论文章节安排

本文将分为以下六章进行介绍:

**第一章 绪论。**本章阐述了论文选题的研究背景及研究意义, 介绍了遥感影像道路提取研究的国内外研究现状和发展趋势, 主要包括基于全监督深度学习的道路分割和拓扑追踪方法、基于弱监督深度学习的道路路面提取方法, 概述了本文的研究目标和主要研究内容, 最后介绍了本文的结构安排及组织情况。

**第二章 深度学习基础理论与技术。**本章是论文的理论基础部分, 以道路提取目的为驱使, 简要介绍了 CNN 的特点和一般结构, 并对 FCN、U-Net 和 D-LinkNet 网络模型进行了较为详细的介绍; 此外还概述了集成学习、弱监督学习技术, 为后续介绍提供铺垫。

**第三章 道路提取实验准备。**本章对论文涉及的实验数据、性能评估进行介绍, 其中实验数据方面具体包括本文使用的道路提取数据集介绍、数据预处理操作和数据增强策略说明; 性能评估方面具体包括实验环境配置、精度评价指标。

**第四章 结合语义分割和拓扑追踪的道路提取算法。**本章是论文的核心内容之一。首先探究基于 FCN 的道路提升分割方法, 其次探究基于 CNN 的道路多起点追踪方法, 最后针对道路语义分割和拓扑追踪两类方法提取的道路特征的融合进行探究。在不同道路数据集上进行实验验证本章所提出方法的有效性, 并与现有领先方法对比验证本章所提出方法的先进性。

**第五章 基于弱监督深度学习的道路路面提取算法。**本章是论文的核心内容之一。首先探究了用于从稀疏的弱标签生成分割代理标签的道路标签传播方案, 其次在全卷积神经网络基础上设计了用于道路分割的边缘感知双分支网络, 协同

不同的损失函数对网络进行训练。在不同道路数据集上进行实验验证本章所提出方法的有效性，并与现有领先方法对比验证本章所提出方法的先进性。

第六章 **总结与展望**。本章是对论文研究工作的总结，并展望了后续研究内容的潜在方向。

武汉大学



## 第 2 章 深度学习基础理论与技术

本文的理论和方法基础是深度学习中的卷积神经网络,涉及全卷积神经网络,以及激活函数、后向传播算法、损失函数等深度学习的理论和方法,还涉及集成学习、弱监督学习等相关方法技术。

### 2.1 卷积神经网络

卷积神经网络是由人工神经网络(Artificial Neural Network, ANN)不断发展而来的。受生物神经元的启发,人工神经网络模拟人脑中利用连接的神经元传递生物电信号的工作形式,通过构建连接的人工神经元,以期拟合目标任务中输入与输出之间的复杂关系。如图 2-1 (a) 所示,典型的人工神经网络包括输入层、输出层和它们中间的隐藏层。每一层的所有神经元均与上一层的所有神经元相连接,每处连接赋予相应的权重  $w$  代表输入值与该神经元的关联程度,权重值越大则关联程度越高。权重  $w$  加上偏置  $b$ ,再经过激活函数,就得到了该神经元的激活值。当然,隐藏层可以不止一层。如图 2-1 (b) 所示,层数越多,网络越深,结构越复杂。于是,利用大量的神经元和多层结构,深度神经网络(即深度学习<sup>[6]</sup>)能够建立输入样本空间与输出标签空间之间一系列的线性变换和非线性激活。

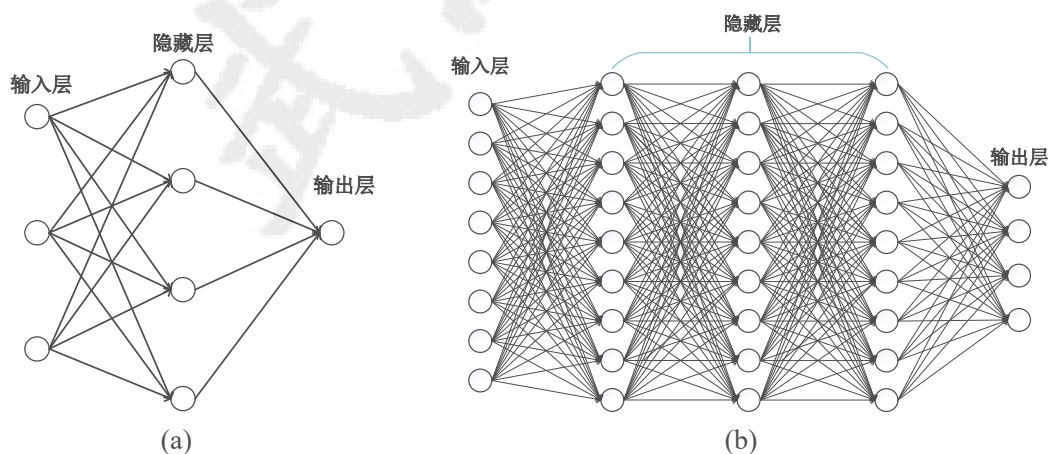


图 2-1 人工神经网络的一般结构

前馈神经网络是一种典型的深度学习模型。定义一个映射  $y=f(x, w)$ , 其中  $x$  代表输入样本,  $y$  代表输出标签。那么神经网络的目的是在  $x$  和  $y$  已知前提下, 通过学习参数  $w$ , 得到某个最优的近似函数  $f^*$ 。作为一种具有稀疏连接、权重共享等特点的前馈神经网络, CNN 使用卷积运算来代替常规的矩阵乘法运算。在处理图像方面, CNN 与全连接的一般人工神经网络相比存在参数量少、平移不

变性的优势。1) 参数量少: 不同于全连接神经网络中每个神经元与相邻网络层所有神经元都连接, CNN 采用稀疏连接的方式, 即一个神经元仅与输入层的局部区域的神经元相连接, 也就是“局部感受野”的概念。2) 平移不变性: 全连接神经网络需要将图像矩阵拉成一维张量, 而 CNN 的卷积核在图像上沿着水平和垂直两个方向滑动, 且一个卷积核对应相同的权重和偏置, 也就是“共享参数”的概念, 能够在提取特征的同时保留空间结构信息。

典型的 CNN 利用多个卷积层由低级到高级逐层提取特征, 越深层提取的特征越抽象。在卷积层之后, 通常使用最大值池化 (max pooling) 或平均值池化 (average pooling) 的池化层来降低特征维数, 一般步长为  $2 \times 2$ 。全连接层将卷积提取的特征图进行汇总并经过 Softmax 激活函数的非线性激活, 然后映射为固定长度的特征向量作为输出。

CNN 最早由 Lecun 等<sup>[51]</sup>于 1989 年首次提出并成功应用于手写字符识别任务, 但因为当时计算机硬件水平有限等问题, 其在很长时间内并未得到有效发展, 而真正为人熟知是在 2012 年 ImageNet 图像分类挑战赛上, Hinton 等<sup>[7]</sup>提出的 AlexNet 利用 CNN 网络和 GPU 高效的并行计算能力一举赢得最好成绩, 超越了人工经验设计特征的传统方法 10% 的精度。后来的几年内, VGGNet<sup>[15]</sup>、GoogleNet<sup>[16]</sup>、ResNet<sup>[17]</sup>等新的 CNN 架构层出不穷, 不断刷新在图像处理领域应用的性能。

## 2.2 全卷积神经网络

根据 2.1 节的介绍, 普通 CNN 网络是对输入图像进行全局反馈, 换言之, 普通 CNN 适合于图像级别分类或回归任务, 但针对语义分割这种图像像素级别分类的密集预测任务, 直接应用普通 CNN 架构并不合适, 原因有二。第一, 普通 CNN 以像素为中心在图像上开辟局部窗口作为输入, 输出中心像素的类别, 为了得到图像中每个像素的类别, 则需要逐像素的操作方式, 这对内存的占用较高, 会出现数据冗余问题且在边缘处产生混淆像素; 第二, 普通 CNN 的池化层通过下采样特征图的分辨率以期提取高级特征, 但代价是细节信息损失, 所以仅使用最后较为抽象的高级特征不利于精细分割结果的获得。

考虑到 CNN 在图像语义分割中的上述局限性, Long 等<sup>[18]</sup>于 2015 年提出全卷积神经网络 (FCN) 这种专门面向语义分割任务的特殊 CNN 架构。在普通 CNN

的基础上，FCN 采用卷积层代替全连接层，并对 CNN 最后的特征图进行上采样操作得到与输入尺寸一致的输出，其中还有跳跃连接（skip connection）使上采样阶段的特征图与前期下采样阶段的特征图相加（addition）。与普通 CNN 不同，FCN 及其变种网络模型（如 DeconvNet<sup>[19]</sup>、U-Net<sup>[20]</sup>、DeepLab<sup>[21]</sup>等）的这种编码器-解码器结构允许其可接受任意尺寸的输入。图 2-2 展示了 FCN 的网络结构，其中编码部分利用池化层进行下采样，解码部分利用反卷积层<sup>[19]</sup>对编码部分生成的特征图进行上采样，使其恢复到与输入图像相同尺寸的大小，得到图像每个像素的分类结果。文献[18]指出，相比直接将 pool5 结果进行 32 倍上采样的 FCN-32s，利用相加操作融合并逐层上采样的 FCN-8s 能够得到更精细的分割结果，这是由于高倍率的上采样损失了图像中细节信息的表达。

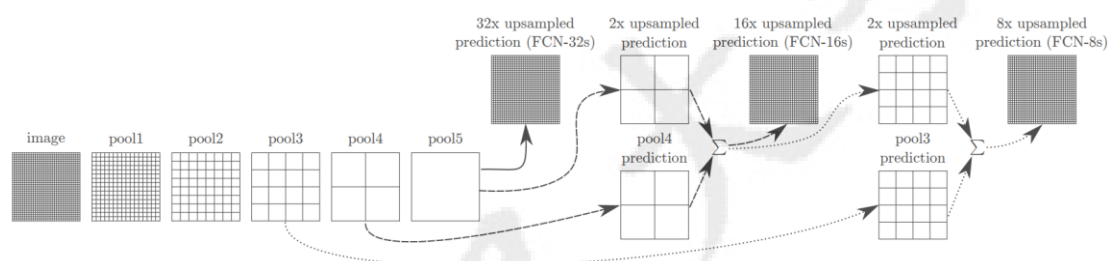


图 2-2 FCN 网络架构（引自[18]）

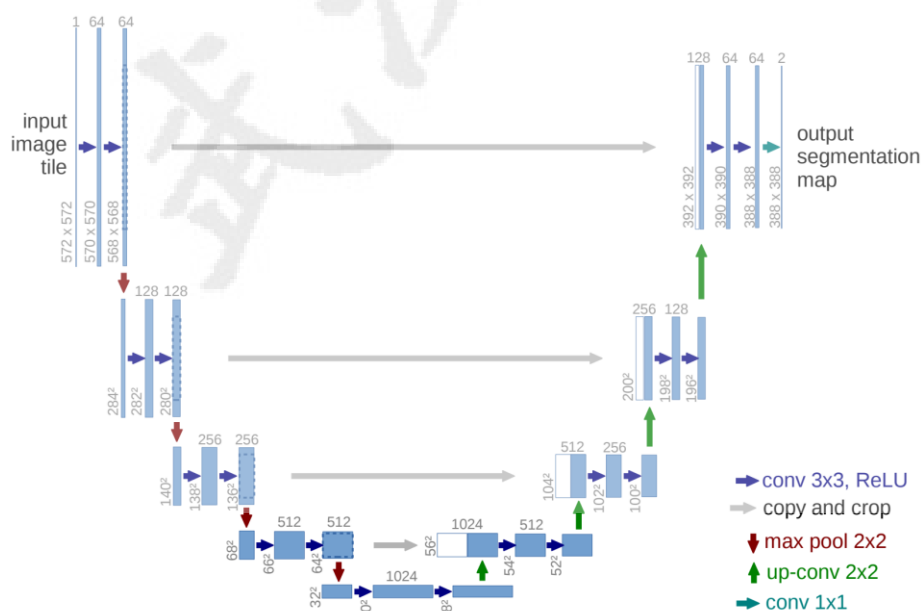


图 2-3 U-Net 网络结构（引自[20]）

在 FCN 的基础上，U-Net<sup>[20]</sup>被提出并应用于解决生物医学图像分割任务中，进一步扩展了低维特征与高维特征的融合，是经典的 FCN 变种网络架构。如图 2-3 所示，U-Net 网络呈左右对称的 U 形，延续了 FCN 编码器-解码器结构。输

入图像首先经过由多个堆叠的卷积层和池化层构成的编码部分，得到分辨率较低的特征图，然后经过由一系列反卷积层构成的解码部分，逐层上采样特征图。U-Net 将编码器和解码器对应层生成的特征图通过跳跃连接进行级联（concatenation），这样融合后既包含有低级特征的细节信息又包含有高级特征的抽象信息，能够在充分挖掘图像信息的同时又保留其原始信息。

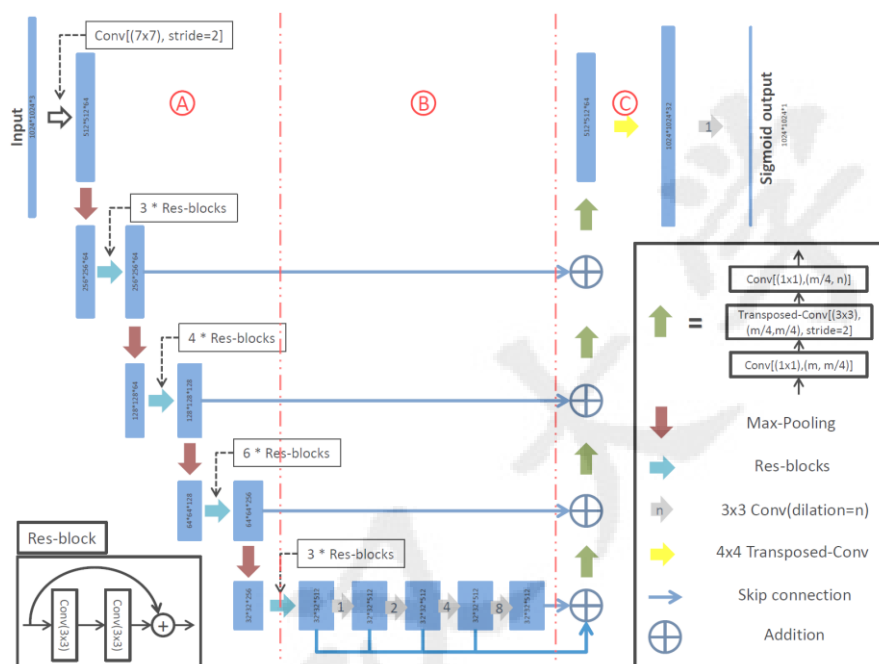


图 2-4 D-LinkNet 网络结构（引自[27]）

立足于道路分割任务，在此介绍赢得了 DeepGlobe<sup>[28]</sup>道路提取挑战赛的一个 FCN 架构，名为 D-LinkNet，其网络结构如图 2-4。D-LinkNet 同样延续了 FCN 编码器-解码器结构，采用已在 ImageNet<sup>[34]</sup>上预训练好的 ResNet<sup>[17]</sup>作为编码器（图 2-4 A），然后在道路数据集上微调，以加快训练过程。解码器（图 2-4 C）使用多个转置卷积层<sup>[52]</sup>进行上采样，以恢复特征图的分辨率。针对编码器和解码器之间的部分（图 2-4 B），结合并行和串行两种方式引入了具有不同空洞率的空洞卷积，以扩大感受野从而捕获图像的多尺度信息。

### 2.3 集成学习

集成学习<sup>[53]</sup>就是构建并组合多个机器学习器来完成学习任务，其关键在于：  
 1) 不同的学习器之间需要差异性，例如采用不同的算法及模型作为学习器，或者以不同的数据作为训练集等；  
 2) 根据目标任务确定个体学习器之间的集成方式。所谓“三个臭皮匠顶个诸葛亮”，集成学习技术的优势是利用多个学习器可

提高单个学习器的泛化性能，通过多种表示形式实现更接近目标的输出，降低陷入由算法特性或样本分布等因素导致的局部最优学习的风险。

目前集成学习大致有串行式集成和并行式集成两类，分别以 Boosting 与 Bagging 为代表。Boosting 采用序列化的模式，其工作原理是：从初始训练集上先训练一个学习器，根据它的预测性能调整训练样本分布，使下一个学习器在训练时更多地关注于先前学习器预测不准确的训练样本；利用新的样本分布再训练下一个学习器；重复上述过程，直至迭代轮数（即学习器数目）达到预定值。Bagging 的工作原理是：在给定数据集上并行训练多个学习器，且基于有放回随机采样以降低方差。上述简述的 Boosting 和 Bagging 的区别集中在学习器的训练阶段，那么，多个学习器训练完成后需要做的是，采用一定的策略将多个学习器的输出结果加权结合，常见的结合策略有平均法、投票法等。

## 2.4 弱监督学习

弱监督学习通常指利用较弱的监督来训练模型学习目标特征的研究，此处“较弱”是相对全监督学习而言的。全监督学习是最常见的一种监督学习方式，需要与输入样本空间完全对应的输出标签空间；而弱监督学习的输出标签空间并不一定与输入样本空间完全对应，换言之，弱监督学习是从不完整的弱标签中学习提取有价值的特征。针对语义分割任务，全监督学习方法需要预先逐像素标注好的密集标签，一般需要费时费力的手动注释工作。在基于弱监督学习的语义分割研究方面，常见的弱标签包括点状（click）<sup>[37]</sup>、线状（scribble）<sup>[38]</sup>、边界框（bounding box）<sup>[39]</sup>、图像级别（image-level）<sup>[40]</sup>。考虑到本文研究目标是典型的条带状地物，在此主要介绍现有基于线状 scribble 的弱监督分割方法。

Lin 等<sup>[38]</sup>采用计算量相对较大的交替优化方案，提出了一种基于 scribble 弱监督深度学习的经典语义分割算法，命名为 ScribbleSup。首先由弱标签提供的稀疏信号生成建议标签，其次利用建议标签通过标准损失函数监督 FCN 网络的训练，接着 FCN 的预测结果参与建议标签生成过程以改善建议标签的质量，如此在建议标签生成和网络训练两个步骤之间交替迭代，力求经过几轮迭代后提升 FCN 的性能。尽管交替优化的方式提高了建议标签的质量，但由于该算法在建议标签生成时认为每个超像素内部的像素对应标签一致，限制了建议标签的准确度，而且标准损失函数很难区分建议标签中的错误标注，导致 FCN 在不同类别的目

标边缘处不可避免地产生混淆像素。

最近,弱监督分割方面的研究开始关注端到端的方式而非交替优化,主要采用包括基于图论<sup>[41][42][43]</sup>和基于边界<sup>[44]</sup>的正则化操作,旨在丰富弱标签的语义信息。Tang 等<sup>[41][42]</sup>引入了只在标记像素上计算的部分交叉熵损失,还引入了基于图割 (graph cut)、密集 CRF (dense CRFs) 等正则化技术的损失。Obukhov 等<sup>[43]</sup>提出了针对未标记像素的门控 CRF 损失,并协同针对标记像素的部分交叉熵损失监督语义分割网络。Kolesnikov 和 Lampert<sup>[44]</sup>应用一种边界约束规则来恢复细节信息。近期,研究人员开始探索将边界直接嵌入分割网络中。Wang 等<sup>[45]</sup>设计了一个包括预测细化网络 PRN 和边界回归网络 BRN 的架构,利用 BRN 网络辅助和指导 PRN 更好地定位边界。Zhang 等<sup>[46]</sup>在借助边缘信息的同时还设计了一种仅关注图像显着目标所在区域的结构感知门控损失。

## 2.5 本章小结

在本章主要介绍了卷积神经网络和全卷积神经网络的特点和结构,并对 FCN、U-Net 和 D-LinkNet 网络模型进行了较为详细的介绍;此外概述了集成学习、弱监督学习技术,为后续研究内容的介绍提供铺垫。

## 第 3 章 道路提取实验准备

### 3.1 数据

#### 3.1.1 道路提取数据集及预处理

目前国际上开源的高分辨率遥感影像道路提取数据集有 Massachusetts<sup>[54]</sup>、Cities<sup>[33]</sup>、Cheng<sup>[31]</sup>、DeepGlobe<sup>[28]</sup>数据集，另外本文还整理并应用了 Shaoxing、Wuhan 两个道路数据集。这些数据集中场景多样，涵盖了全球多个国家的城市、郊区和农村地区，包含色彩、形状差异明显的不同道路样本。将它们在数据源、分辨率（ground sample distance, GSD）、覆盖面积、标签形式四个方面进行比较，如表 3-1 所示。

表 3-1 道路提取数据集的总体比较

数据集	数据源	GSD (m)	面积 (km <sup>2</sup> )	标签形式
Massachusetts <sup>[54]</sup>	航空	1.0	2600	OSM 等宽膨胀
Cities <sup>[33]</sup>	卫星	0.6	900	OSM 等宽膨胀
Cheng <sup>[31]</sup>	航空	1.2	132	分割&中心线
DeepGlobe <sup>[28]</sup>	卫星	0.5	1632	分割
Shaoxing	航空	0.6	228	分割
Wuhan	卫星	0.5	200	分割

接下来针对每个数据集详细地介绍其数据源、分辨率、覆盖面积等信息，分析它们各自的特点，并说明包括划分规则在内的数据预处理操作。

**Massachusetts 数据集:** 包含美国马萨诸塞州的 1171 张 1 米分辨率的航空影像，每张影像大小约为 1500×1500 像素。该数据集覆盖面积超过 2600 平方公里。标签是通过栅格化 OpenStreetMap 道路中心线生成的 7 像素宽的二值图，且不进行平滑处理。图 3-1 显示了该数据集中的两个代表性区域。经过检查发现，该数据集的特点：1) 在规模方面，多张影像可以拼接为范围更大的一幅影像（至少 4096×4096 像素），适合应用于道路拓扑追踪算法；2) 分割标签是由 OSM 等宽膨胀的，无法保证与真实道路路面的语义对齐，不适合应用于较为准确的道路路面提取；3) 存在部分错误数据，比如影像空白、标签缺失或偏移导致无法与影像对应等。在第 4 章中使用 Massachusetts 数据集验证本文提出的结合语义分割与拓扑追踪的道路提取算法。首先人工排除部分错误数据，将多张可以拼接的影像先拼接为一幅范围更大的影像，然后无缝裁剪为大小适宜（如 1024×1024 像

素) 的样本块, 其中 192 张作为训练集, 其余 64 张作为测试集。

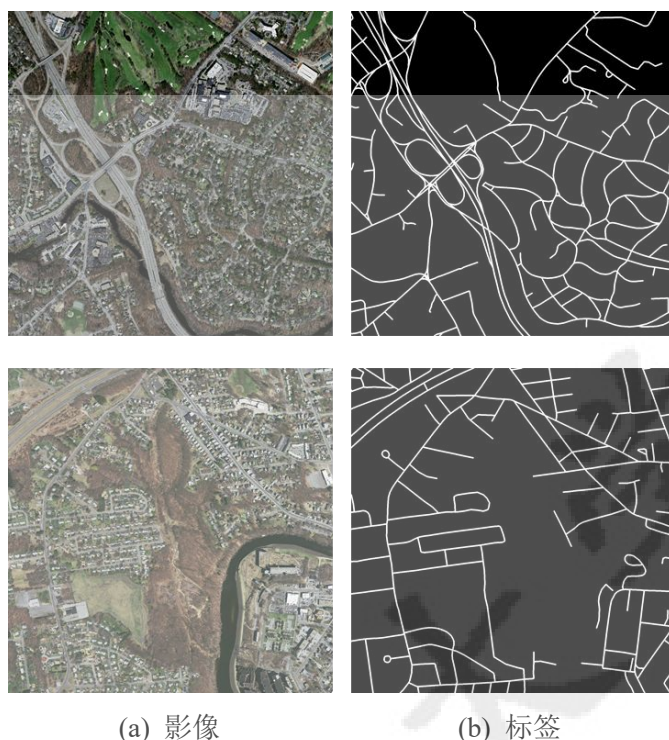


图 3-1 Massachusetts 数据集示例 (1500×1500 像素)

**Cities 数据集:** 以 0.6 米分辨率从 Google Earth<sup>[55]</sup>收集全球 37 个城市的卫星影像, 覆盖每个城市至少 24 平方公里的市中心区域, 这些城市主要来自北美洲、欧洲、亚洲。标签是通过栅格化 OpenStreetMap 道路中心线生成的 8 像素宽的二值图, 且不进行平滑处理。图 3-2 显示了该数据集中八个代表性城市的影像和标签。经过检查发现, 该数据集的特点: 1) 在规模方面, 每个城市对应一幅大范围影像 (至少 8192×8192 像素), 适合应用于道路拓扑追踪算法; 2) 分割标签是由 OSM 等宽膨胀的, 无法保证与真实道路路面的语义对齐, 不适合应用于较为准确的道路路面提取。在第 4 章中使用 Cities 数据集验证本文提出的结合语义分割与拓扑追踪的道路提取算法。采用与文献[33]一致的划分方案, 使用其中 25 个城市 (亚特兰大、奥斯汀、巴尔的摩、哥伦布、达拉斯、华盛顿、休斯顿、印第安纳波利斯、伦敦、路易维尔、迈阿密、密尔沃基、明尼阿波利斯、纳什维尔、奥兰多、费城、菲尼克斯、波兰特、圣安东尼奥、圣何塞、西雅图、旧金山、圣路易斯、坦帕、拉斯维加斯) 作为训练集, 另外 12 个城市 (阿姆斯特丹、芝加哥、丹佛、洛杉矶、蒙特利尔、巴黎、匹兹堡、盐湖城、圣迭戈、东京、多伦多、温哥华) 作为测试集。综合考虑计算机性能, 将遥感影像与对应标签无缝裁剪为大小适宜的样本块 (如 1024×1024 像素), 得到 1600 张训练集和 768 张测试集。



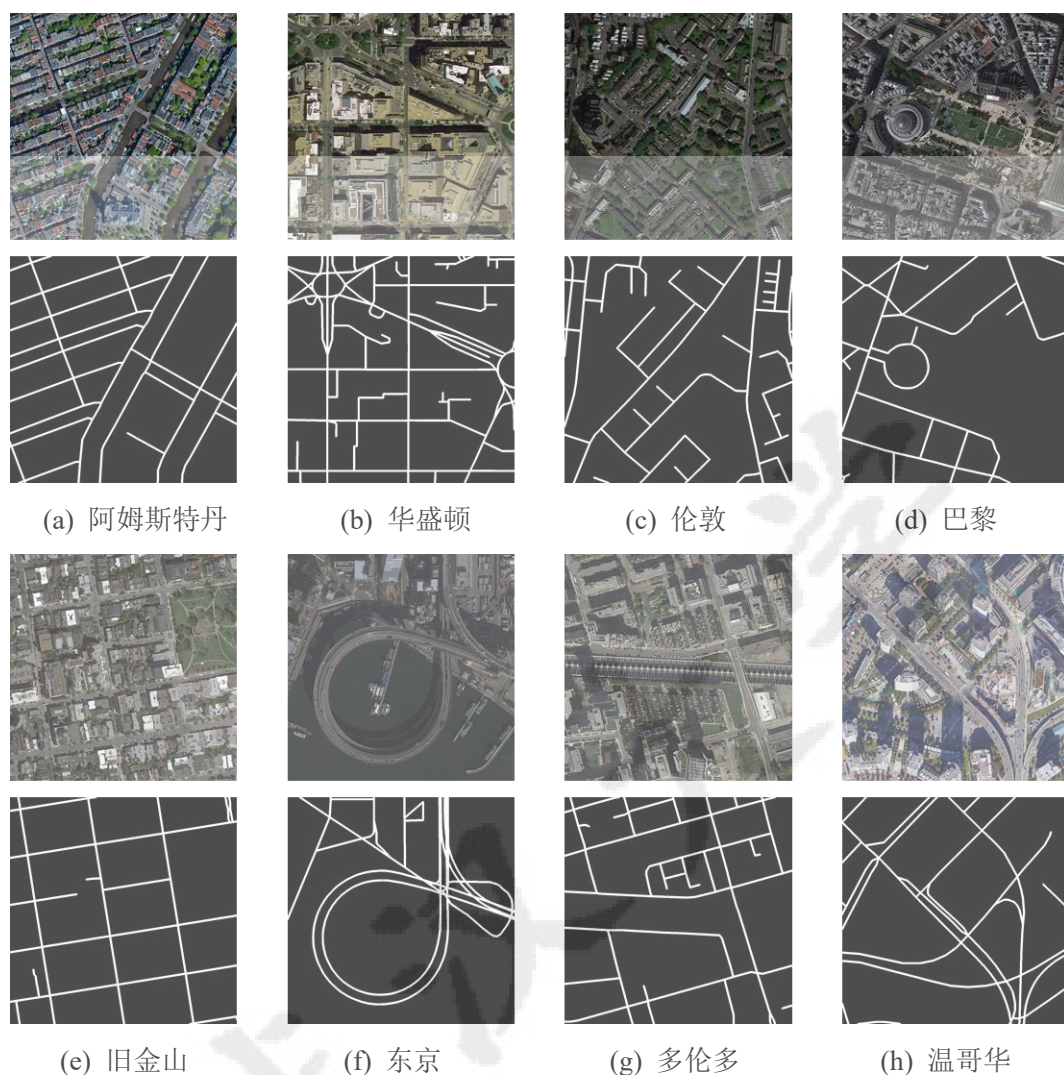


图 3-2 Cities 数据集示例 (1024×1024 像素)

**Cheng 数据集:** 以 1.2 米分辨率从 Google Earth 收集 224 张航空影像，每张影像大小不等但至少为 600×600 像素。标签是人工标注的，包括接近真实道路路面的道路分割标签以及单像素宽的道路中心线标签，均为二值图且不进行平滑处理。图 3-3 显示了该数据集中的两个代表性区域。经过检查发现，该数据集的特点：1) 在规模方面，单幅影像覆盖范围有限且无法拼接为范围更大的影像，不适合应用于道路拓扑追踪算法；2) 分割标签基本可以实现与真实道路路面的语义对齐，适合应用于较为准确的道路路面提取；3) 既有道路分割标签又有道路中心线标签。在第 5 章中使用 Cheng 数据集验证本文提出的基于弱监督深度学习的道路路面提取算法。遵循文献[31]的划分规则，并将遥感影像与对应的标签无缝裁剪为大小适宜的样本块（如 512×512 像素），得到 300 张训练集和 49 张测试集。

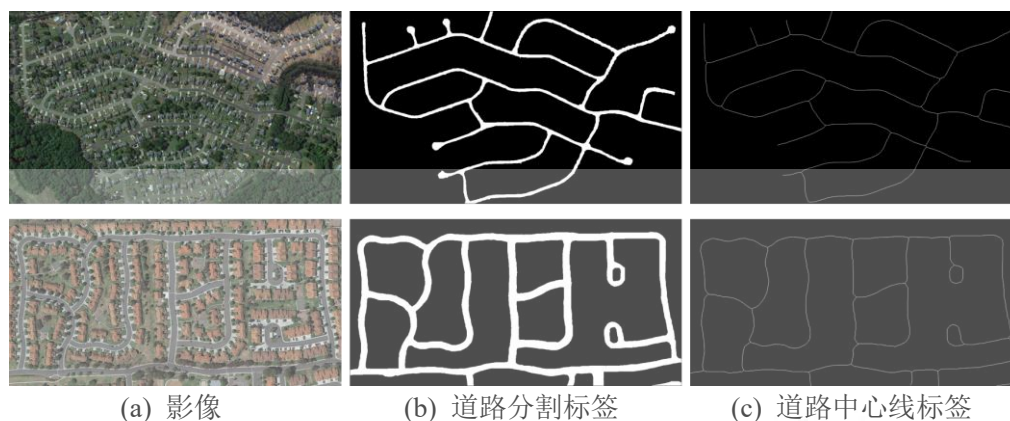


图 3-3 Cheng 数据集示例

**DeepGlobe 数据集：** CVPR DeepGlobe Road Extraction Challenge 道路提取挑战赛<sup>[28]</sup>的数据集，包含来自泰国、印度尼西亚、印度的 8570 张卫星影像，覆盖面积 2220 平方公里，每张影像尺寸为  $1024 \times 1024$  像素，分辨率为 0.5 米。该数据集涵盖了城市和乡村地区的多种道路类型。标签是接近真实道路的道路分割标签，而不是由中心线简单膨胀得到。图 3-4 显示了该数据集中的两个代表性区域。需要说明的是，由于只有原定为训练集的 6226 张影像对应的标签是开源的，所以本文仅使用这些有标签的数据，覆盖面积约 1632 平方公里。经过检查发现，该数据集的特点：1) 在规模方面，单幅影像覆盖范围有限且无法拼接为范围更大的影像，不适合应用于道路拓扑追踪算法；2) 分割标签基本可以实现与真实道路路面的语义对齐，适合应用于较为准确的道路路面提取。在第 5 章中使用 DeepGlobe 数据集验证本文提出的基于弱监督深度学习的道路路面提取算法。综合考虑计算机性能，将遥感影像与对应的标签无缝裁剪为大小适宜的样本块（如  $512 \times 512$  像素），15000 张作为训练集，其余 5333 张作为测试集。

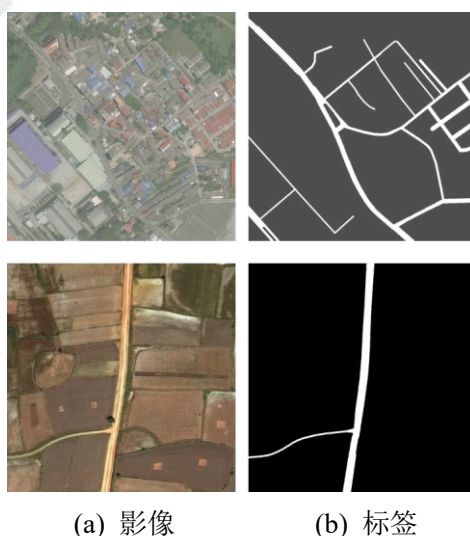
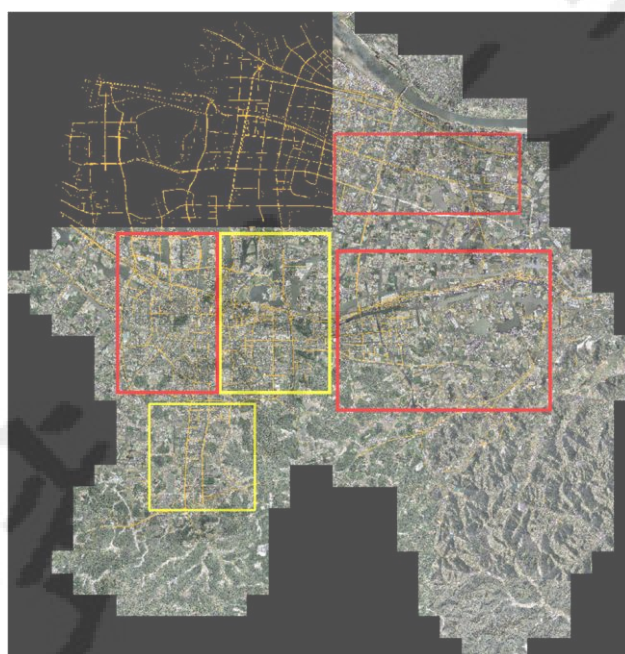
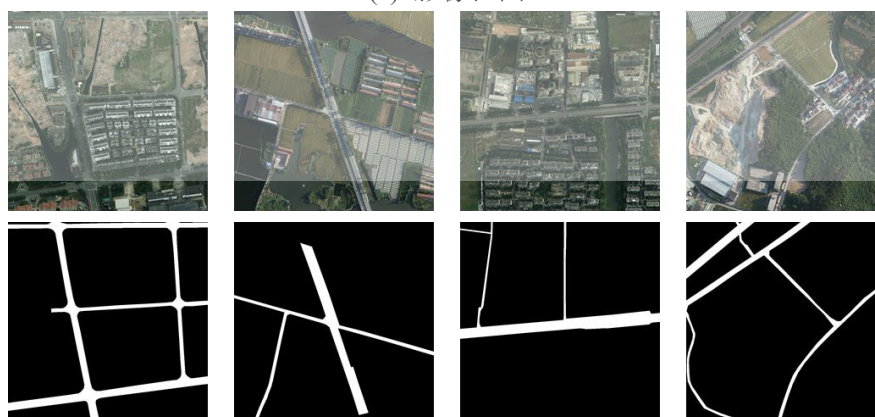


图 3-4 DeepGlobe 数据集示例（ $1024 \times 1024$  像素）

**Shaoxing 数据集:** 收集了关于浙江省绍兴市的 0.6 米分辨率的航空影像，覆盖面积约 228 平方公里，影像范围如图 3-5 (a) 所示。其中，红色框对应的区域用来训练，黄色框对应的区域用来测试。各区域对应一幅大范围影像（至少  $8192 \times 8192$  像素），适合应用于道路拓扑追踪算法。绍兴是我国具有江南水乡特色的城市，湖泊众多，河道密布，图 3-5 (b) 显示了该数据集中四个代表性区域的影像和标签。为了应用于道路提取研究，利用 ArcGIS 软件采用人工方式对道路矢量进行检查，然后通过栅格化生成道路分割标签。在第 4 章中使用 Shaoxing 数据集验证本文提出的结合语义分割和拓扑追踪的道路提取算法。综合考虑计算机性能，将遥感影像与对应的标签无缝裁剪为大小适宜的样本块（如  $1024 \times 1024$  像素），得到 372 张训练集和 160 张测试集。



(a) 影像范围



(b) 示例（ $1024 \times 1024$  像素）

图 3-5 Shaoxing 数据集

**Wuhan 数据集：**收集了关于湖北省武汉市的 0.5 米分辨率的卫星影像，由高景卫星采集，两幅 20000×20000 像素的影像分别位于武汉的江汉区和洪山区，其大致范围如图 3-6 (a) 所示，总覆盖面积 200 平方公里。图 3-6 (b) 显示了该数据集中四个代表性区域的影像和标签。为了应用于道路提取研究，利用 ArcGIS 软件采用人工方式对道路矢量进行检查，然后通过栅格化生成道路分割标签。在第 5 章中使用 Wuhan 数据集验证本文提出的基于弱监督深度学习的道路路面提取算法。综合考虑计算机性能，将遥感影像与对应的标签无缝裁剪为大小适宜的样本块（如 512×512 像素），以 3:1 划分训练数据和测试数据，得到 1944 张训练集和 648 张测试集。



(a) 影像范围



(b) 示例（512×512 像素）

图 3-6 Wuhan 数据集

### 3.1.2 数据增强

深度学习模型训练依赖大量的数据作为支撑，随着模型参数数量的增加，需要的数据量也越来越多。源数据集与目标数据集的差异较大时，用源数据集训练的模型在当前数据集上表现不俗，但可能在目标数据集上泛化能力降低。而制作和整理道路提取数据集的成本比较高昂，数据集有限的问题在实际生产应用中是常见的。因此，如何选择适当的策略来充分发挥有限数据集的能力，也是基于深度学习进行遥感影像道路提取的重要研究内容。

为了加强深度学习的泛化能力，本文采用数据增强（data augmentation）策略。第一种是在训练网络模型的前端，利用有限的模拟数据来模拟出更丰富的输入数据集，可以有效避免过拟合。考虑到视角变化可能导致的几何变形，对训练样本进行缩放，平移，以及水平、垂直、对角翻转。另外，考虑到光照差异对地物识别带来的影响，采用色彩抖动策略对训练影像的图像亮度、饱和度、对比度进行随机变化。第二种是对测试集进行扩展，即采用测试阶段增强（Test Time Augmentation, TTA）技术，包括水平、垂直、对角翻转。对原始影像和翻转后的影像对应的预测结果进行平均，得到该影像的最终结果。

除了数据增强以外，迁移学习（transfer learning）的方式也能够在数据集有限的情况下提高深度学习模型的性能。在语义分割领域，常见的是使用在 ImageNet 数据集<sup>[34]</sup>上预训练过的编码器模型作为当前模型的初始权重，然后进行微调适应当前数据集，能够有效加速训练过程和提升网络性能。

## 3.2 性能评估

所谓“工欲善其事，必先利其器”，为了对研究内容中涉及的相关技术方法进行实验验证，选择符合任务需求的编程语言和软硬件环境也是至关重要的。本文选择编程语言 Python，代码可读性高，易于维护，且拥有很多强大而便捷的图像处理、数学计算和数据挖掘模块，比如数学运算第三方库 NumPy，还有 PIL 所提供的图像处理通用功能等。在软件环境方面，实验采用 Ubuntu 18.04.2 LTS 64 位操作系统，选择深度学习框架 PyTorch 和 TensorFlow 作为搭建神经网络的主要工具，能够充分利用硬件资源，允许在多 GPU 设备上并行计算。涉及到的依赖库主要包括 CUDA、cuDNN、OpenCV 等。在硬件环境方面，实验使用一块 NVIDIA GeForce GTX 1060 6GB 显存的 GPU 显卡。

遥感影像中信息种类丰富，不同的信息内容提取形式不同。道路是一类具有拓扑结构的特殊地表目标物，因此度量道路提取算法的性能时需要考虑像素级别和拓扑级别的评价指标。像素级别的性能评估就是对道路提取结果与相应的真实标签在像素上进行比较，而拓扑级别的性能评估则偏重于道路相邻路段之间的连通性。由于道路提取可大致归纳为道路分割任务和道路中心线提取任务，在此从这两个角度分别介绍评定道路提取算法性能所需要的指标。

### 3.2.1 道路分割的评价指标

首先，道路分割任务在通常情况下可看作是二分类问题。表 3-2 展示的是二分类的混淆矩阵，通过对比真实情况与预测结果的异同，将所有像素划分为真正例 (True Positive, TP)、真反例 (True Negative, TN)、假正例 (False Positive, FP) 和假反例 (False Negative, FN) 四种，显然 TP、TN、FP、FN 对应的像素数之和等于像素总数。如果立足于道路分割问题，道路像素为正，其余非道路像素为负，那么 TP 代表被正确分类为道路的像素数，TN 代表被正确分类为非道路的像素数，FP 代表非道路被错误分类为道路的像素数，FN 代表道路被错误分类为非道路的像素数。

表 3-2 二分类的混淆矩阵

真实情况	预测结果	
	正例 (Positive)	反例 (Negative)
正例 (True)	真正例 (TP)	真反例 (TN)
反例 (False)	假正例 (FP)	假反例 (FN)

在语义分割领域，精确率 (Precision)、召回率 (Recall)、F1-Score (简称为  $F_1$ )、交并比 (Intersection-over-Union, IoU) 是像素级别评估的四个常用指标。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.2)$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.3)$$

$$\text{IoU} = \frac{TP}{FP + TP + FN} \quad (3.4)$$

结合混淆矩阵，具体计算如公式 (3.1)、(3.2)、(3.3)、(3.4) 所示。其中，Precision 是正确预测为道路的像素在所有预测为道路的像素中所占百分比；Recall 是正确预测为道路的像素在所有真实标签为道路的像素中所占百分比。Precision 和 Recall 是一对矛盾的度量，往往在一方升高时另一方却下降，通常只有在一些简单任务中才可能使二者都很高，故需要视情况综合考虑。一般的解决方法是采用如公式 (3.5) 所示的 F-Score，即求 Precision 和 Recall 的加权调和平均。当参数  $\gamma = 1$  时，便是公式 (3.3) 中的指标  $F_1$ 。IoU 指的是预测为道路像素和真实标签为道路像素的交集与二者的并集之间的比率。相比之下，IoU 和  $F_1$  是像素级别的指标中更为综合而具有说服力的。

$$F = \frac{(\gamma^2 + 1) \times \text{Precision} \times \text{Recall}}{\gamma^2 \times (\text{Precision} + \text{Recall})} \quad (3.5)$$

考虑到像素级别的评估难以判定道路网拓扑连通性的优劣，因此本文为道路分割算法的性能评估补充了一个拓扑级别的度量“道路连接性” (connectivity, Conn)，其计算方式见公式 (3.6)。首先将道路中心线标签划分为等长的多个路段，总路段数为  $N_{gt}$ ；将经过了形态学细化<sup>[56]</sup>的分割结果也划分为等长的多个路段，总路段数为  $N_{pred}$ ；将中心线标签里那些被分割结果完全覆盖的路段视为连通的道路路段，统计其数量为  $N_{conn}$ 。该指标借鉴了  $F_1$  的定义，能够在局部尺度上衡量道路网的连通性和拓扑结构准确性。

$$\text{Conn} = \frac{2N_{conn}}{N_{gt} + N_{pred}} \quad (3.6)$$

### 3.2.2 道路中心线提取的评价指标

道路中心线提取是遥感影像道路提取研究中的另一项重要任务，其目标并非关注于某个像素是不是道路，而是关注于通过道路中心线来表示的道路网的位置、结构等拓扑信息，所以评估中心线提取算法需要采取不同于道路分割算法性能评估的形式。文献[57]引入了完整性 (completeness, Comp)、正确性 (correctness, Corr) 和质量 (quality, Qual) 用来评估中心线提取算法的性能，其中涉及参考的道路中心线 (reference road data)、提取的中心线结果 (extracted road data) 以及缓冲区半径  $\rho$ 。

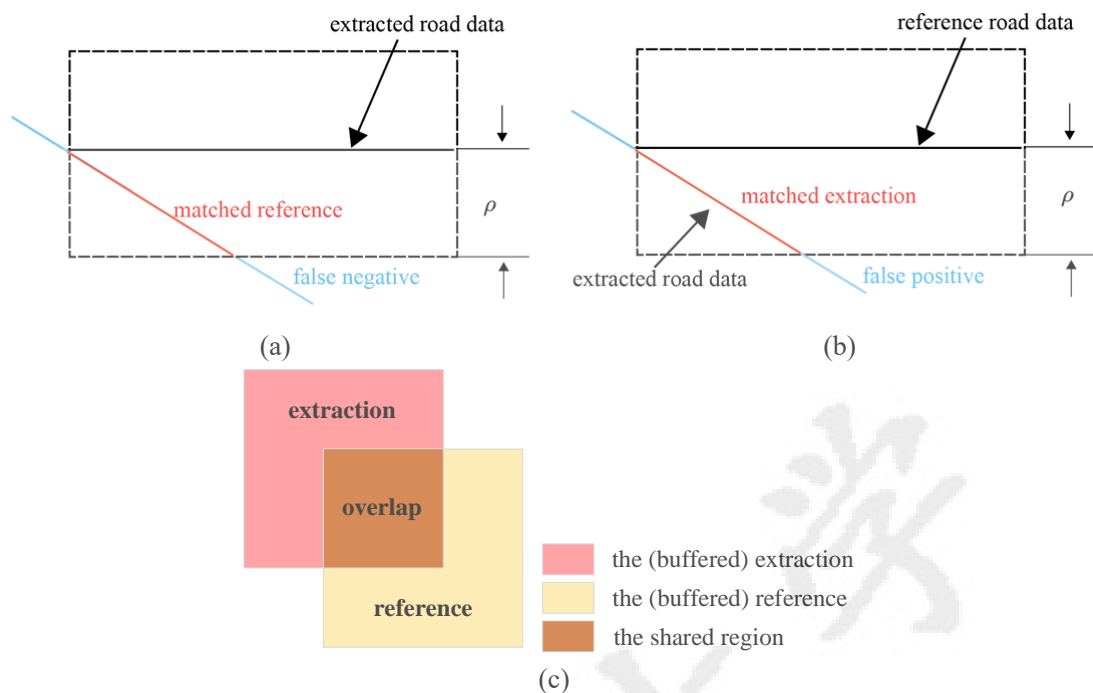


图 3-7 道路中心线提取评估指标的可视化示例

如图 3-7 (a) 所示,  $Comp$  是参考中心线位于提取的中心线的缓冲区内 (matched reference) 的百分比, 可视为宽松的 (relaxed) Recall。如图 3-7 (b) 所示,  $Corr$  是提取的中心线位于参考中心线的缓冲区内 (matched extraction) 的百分比, 可视为 relaxed Precision。如图 3-7 (c) 所示,  $Qual$  是综合了  $Comp$  和  $Corr$  的指标, 可视为 relaxed IoU。 $Comp$ 、 $Corr$ 、 $Qual$  被[4]等文献广泛应用, 具体计算可公式化为 (3.7)、(3.8)、(3.9)。

$$Comp = \frac{\text{length of matched reference}}{\text{length of reference}} \quad (3.7)$$

$$Corr = \frac{\text{length of matched extraction}}{\text{length of extraction}} \quad (3.8)$$

$$Qual = \frac{\text{length of matched extraction}}{\text{length of extraction} + \text{length of unmatched reference}} \quad (3.9)$$

### 3.3 本章小结

本章内容主要介绍了本文在进行道路提取的实验时所需的前期准备。首先介绍所收集的道路提取数据集的概况, 阐述了包括拼接、裁剪、训练集测试集划分在内的数据预处理操作, 并针对每个数据集的特点进行了说明。随后说明了本文实验的环境基本配置, 以及用于评估算法性能的指标的原理及其计算公式。



## 第 4 章 结合语义分割和拓扑追踪的道路提取算法

### 4.1 引言

#### 4.1.1 现有方法存在的问题

在前期研究中发现，基于深度学习的遥感影像道路提取方法大致有两类：语义分割方法、拓扑追踪方法。

语义分割方法一般是在道路数据集上训练全卷积神经网络或其变种模型，利用网络模型对影像进行预测，得到影像中每个像素是“道路”的概率，进而生成与影像逐像素对应的被标记为“道路”或“非道路”的二值分割图。例如文献[27]中提出的 D-LinkNet 网络模型取得了 2018 年国际计算机视觉与模式识别会议中 DeepGlobe 道路提取挑战赛（CVPR DeepGlobe Road Extraction Challenge）的第一名，结合编码-解码结构的网络 LinkNet 和空洞卷积（Dilated convolution），能够检测到影像中的大部分道路。但是，关注密集预测任务的语义分割方法往往忽略了道路的拓扑结构信息，导致提取结果的连通性较差，特别是在复杂场景下易得到零碎的路段。与之不同的是，拓扑追踪方法将道路视为由顶点（道路上的点）和边线（连接相邻点的线段）构成的图结构化数据（graph-structured data），强调其拓扑结构特性。例如文献[33]中提出的 RoadTracer 算法是从道路上的一个已知点出发，开辟窗口提取影像的局部特征，并通过卷积神经网络引导该窗口沿着道路移动，迭代式搜索道路上其他的顶点和边线添加到路径中，由此逐步追踪道路中心线，以构建全局的道路网。然而，RoadTracer 在遥感影像复杂场景中表现欠佳，一方面，这种单起点追踪的方式容易受到桥梁、周围树木房屋遮挡等影响而使得搜索中断，难以提取较为完整的道路结果；另一方面，其人工选择起始点的方式也降低了算法的自动化程度。

#### 4.1.2 技术路线和创新点

为了解决上述已有道路提取方法中存在的道路结果不完整、拓扑连通性差等问题，本章提出一种结合语义分割和拓扑追踪的道路提取算法，简称为 Seg-Tracer，其总体技术路线如图 4-1 所示。本算法采用多阶段的框架，包括三个主要阶段：1) 基于全卷积神经网络的道路提升分割；2) 基于卷积神经网络的道路多起点追踪；3) 道路分割和追踪的融合。首先，通过主流的全卷积神经网络对遥感影像

进行初始分割，然后利用所设计的另一个全卷积神经网络进行道路提升分割。其次，通过角点检测算子在上一阶段预测的道路分割图上提取多个特征点作为后续追踪的起始点，由此进行基于卷积神经网络的多起点追踪构建道路网。最后，对上述两类方法提取出的道路特征进行互相约束，融合道路的语义信息和拓扑信息，以生成更精细的道路路面以及中心线结果。

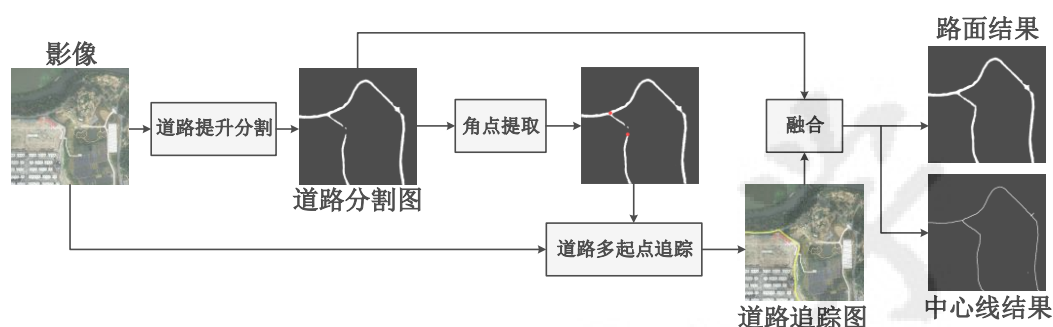


图 4-1 结合语义分割和拓扑追踪的道路提取算法 Seg-Tracer

此算法的创新点可总结为以下几点：

1) 基于多个不同的卷积神经网络，提出了一种用于从遥感影像中同时提取道路路面和中心线的多阶段框架，该框架分别在语义分割和拓扑追踪两类方法的最新模型及算法上改进，并使二者相互合作、互为补充，与仅利用语义分割或拓扑追踪的方法对比，该算法有效提高了道路提取的完整性和拓扑连通性；

2) 将经典机器学习中的 Boosting 提升策略引入用于道路分割的深度学习方法中，从先前分割的失误案例中逐步学习，以增强道路分割的连接性，减少误判；此外还在全卷积神经网络的基础上设计了一个提升分割网络模型，该模型作为提升分割过程中的“学习器”被多次使用，具有轻量高效的优势；

3) 改进了单起点追踪算法，提出一种道路多起点追踪算法，在分割结果的基础上检测多个特征点作为追踪的起点，弥补了单起点追踪算法中搜索过早停止而使道路不完整的劣势，同时通过提高特征起始点的质量来提高搜索效率及算法的自动化程度。

4) 从充分发挥语义分割和拓扑追踪两类方法各自优势的目的出发，提出了一种新的道路路面和中心线的融合策略，融合道路的语义信息和拓扑信息以进一步提高道路提取的准确性，与仅使用语义分割和拓扑追踪两类方法中任一方法的算法相比，得到了更完整和更连通的道路结果，且对复杂场景具有一定的鲁棒性。

## 4.2 基于全卷积神经网络的道路提升分割

### 4.2.1 提升分割方法

针对诸如 D-LinkNet<sup>[27]</sup> 的利用语义分割方法进行道路提取所存在的缺点，特别是其分割结果中道路拓扑连通性差等问题，本节将经典的集成学习技术引入道路分割的深度学习方法中，利用多个泛化性能相对弱的学习器构建出泛化性能较强的集成。所提出的这一“道路提升分割”受启发于 AdaBoost 算法<sup>[53]</sup>，它属于集成学习中最具代表性的 Boosting 方法，是一种串行集成方法，关注学习器之间的相关性，其关键技术有两项，分别是“训练样本的抽中概率”和“学习器的投票权重”。前者要求学习器在指定的分布下学习，在每轮训练前需根据相应的分布对每个训练样本赋权，更新各训练样本被抽中的概率；后者与各学习器的结合方式有关，本节所提出的道路提升分割采用的是基于加权投票的结合。

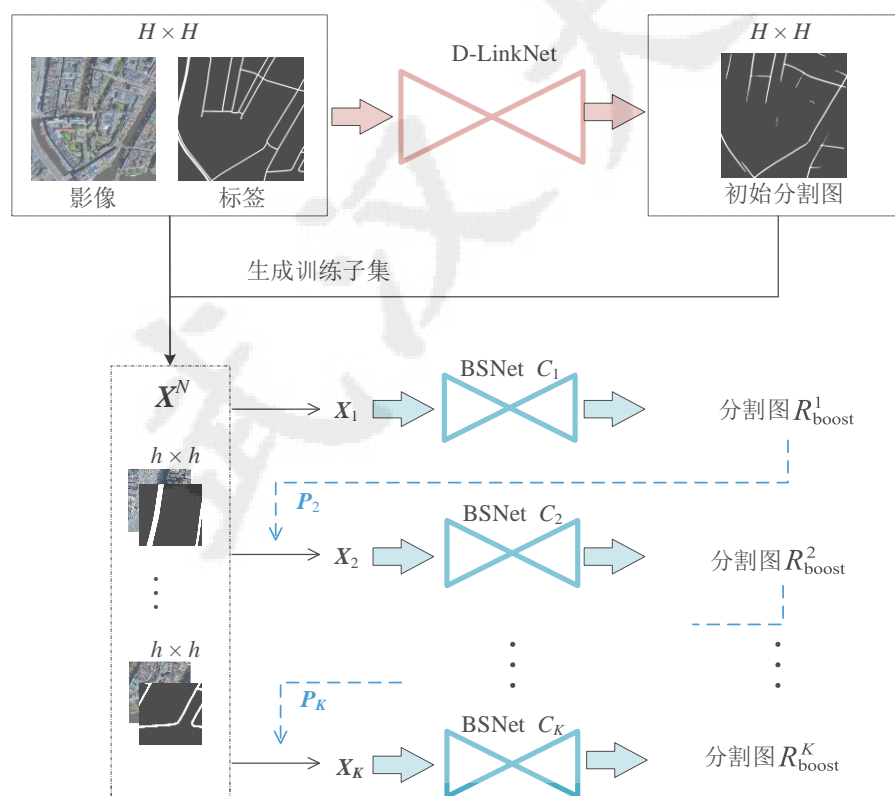


图 4-2 道路提升分割在训练阶段的流程图

道路提升分割目的是利用多个学习器进行几轮迭代，不断从先前学习器的失误案例中学习，来填补或连接先前学习器所预测的破碎路段，其在训练阶段的流程图如图 4-2 所示。首先，在道路数据集上训练网络 D-LinkNet，学习遥感影像中的道路特征，并对影像进行像素级别预测，得到道路初始分割结果。假设 D-

LinkNet 的训练样本尺寸为  $H \times H$ ，则以小尺寸  $h \times h$  裁剪原始影像及其标签，从而构建用于后续提升分割的样本子集  $X^N$  ( $N$  代表  $X^N$  中的样本块总数量)。在此，我们仅裁剪在初始分割结果中表现较差的部分，比如  $h \times h$  裁剪窗口中初始分割结果和对应路面标签的交并比小于阈值 70% 的情况。另外，将不含道路的训练样本筛掉，以减轻正负样本数量之间的不平衡。

接下来，依次训练若干轮提升分割网络 (Boosting Segmentation Network, BSNet)，该网络模型被设计作为提升分割过程中的学习器 (简记为  $C$ )，其结构将在 4.2.2 小节中详细介绍。根据公式 (4.1)，将  $X_1$  (即  $X^N$ ) 中每个训练样本  $x^i \in X^N$  的抽中概率  $P_1^i \in P_1$  进行初始化：

$$P_{j=1}^i = \frac{1}{N} \quad (4.1)$$

对于第  $j$  ( $j \in (0, K]$ ， $K$  是迭代的总轮数) 轮迭代，在  $X^N$  中根据各样本  $x^i$  被抽中的概率  $P_j^i$  获得  $X_j$  作为  $C_j$  的训练集，当  $C_j$  完成训练后在整个  $X^N$  上对其进行评估。 $C_j$  在样本  $x^i$  上的错误率可由公式 (4.2) 计算，其中 IoU 为图 4-2 中  $R_{\text{boost}}^j$  分割图与对应标签之间的交并比，计算方式可参见公式 (3.4)。

$$\varepsilon_j^i = 1 - \text{IoU}_j^i \quad (4.2)$$

那么当前模型  $C_j$  在整个  $X^N$  上的错误率  $\varepsilon_j$  为：

$$\varepsilon_j = \sum_{i=1}^N (\varepsilon_j^i \times P_j^i) \quad (4.3)$$

根据公式 (4.4) 可得到当前模型  $C_j$  的投票权重  $\omega_j$ ，

$$\omega_j = \log\left(\frac{1 - \varepsilon_j}{\varepsilon_j}\right) \quad (4.4)$$

然后由公式 (4.5) 和 (4.6) 更新  $x^i$  的抽中概率得到  $P_{j+1}^i \in P_{j+1}$ ，用于抽样生成下一轮的训练样本  $X_{j+1}$ 。其中，设置阈值  $T_{\text{IoU}}$  为 0.7，旨在关注模型的误判情况，降低能被当前模型正确预测的样本在下一轮被抽中的概率。

$$P_{j+1}^i = \begin{cases} P_j^i \times \frac{\varepsilon_j}{1 - \varepsilon_j}, & \text{IoU}_j^i > T_{\text{IoU}} \\ P_j^i, & \text{otherwise} \end{cases} \quad (4.5)$$

$$P_{j+1}^i = \frac{P_{j+1}^i}{\sum_{i=1}^N P_{j+1}^i} \quad (4.6)$$

当迭代轮数（即 BSNet 个数）达到  $K$  后，道路提升分割的训练阶段全部完成。值得注意的是，每轮迭代中都会进行一次检查，不符合要求（比如错误率大于 0.5）的 BSNet 将被丢掉，然后重新在当前训练样本上训练新的 BSNet。

在测试阶段，分别利用  $K$  个已完成训练的 BSNet 网络模型对影像进行预测，然后采用加权投票法将它们的预测结果进行结合，具体可参见 Algorithm 1 中的伪代码。

---

**Algorithm 1. Boosting Segmentation**

---

Input: training dataset  $X^N$  containing data item  $x$  with probability  $P$  and corresponding label  $y'$ . A predefined threshold  $T_{IoU}$ . The maximum number of BSNet is  $k$  and BSNet is noted as  $C$ .

**Initialize:**

**For**  $i \in [1, N]$

$P_{j=1}^i = 1/N$

**For**  $j = 1, 2, \dots, k$

    Randomly select  $X_j$  from  $X^N$  according to  $P_j^i$

**Training:**

        Train  $C_j$  on dataset  $X_j$

**Evaluation:**

$y_j^i = C_j(x^i), x^i \in X^N$

$IoU_j^i = IoU(y_j^i, y'^i)$

        Error rate  $\varepsilon_j^i = 1 - IoU_j^i$

**If**  $\varepsilon_j^i > 0.5$  **then**  $k = j - 1$ ; stop

**Updating:**

**Foreach**  $(x^i, y^i)$  in  $X^N$

            Weight update  $\beta_j = \varepsilon_j / (1 - \varepsilon_j)$

**If**  $IoU_j^i > T_{IoU}$  **then**  $P_{j+1}^i = \beta_j P_j^i$

**Else**  $P_{j+1}^i = P_j^i$

        Normalize probability;  $P_{j+1}^i = P_{j+1}^i / \sum_i P_{j+1}^i$

**Inference:**

**Foreach**  $x^i$  in test dataset:

$y^i = \sum_{j=1}^k \left( \log \frac{1}{\beta_j} \right) C_j(x^i)$

---

关于结合方式，首先根据公式(4.7)对  $K$  个 BSNet 的预测结果  $R_j (j \in (0, K])$  进行加权结合，生成的分割结果以概率图  $R_{boost}$  表示，然后根据公式(4.8)进一步结合 D-LinkNet 的初始分割结果（记为概率图  $R_{ini}$ ）最终得到提升后的分割结果（记为概率图  $R_{seg}$ ）。其中， $R[f]$  指的是保留概率图  $R$  中满足  $f$  成立的像素的值，其余像素的值设为 0。在此， $f$  的条件是  $R_{ini}$  与  $R_{boost}$  的对应像素值加和大于阈值  $T_{seg}$ ，设置  $T_{seg}$  为 1。另外，公式(4.8)中  $Norm(\cdot)$  是将概率值归一化到 0 至 1。最后通过 OTSU<sup>[58]</sup> 自适应阈值分割得到二值化的提升分割结果，0 代表“非路”，1 代表“路”。

$$R_{\text{boost}} = \frac{\sum_{j=1}^k \omega_j R_j}{\sum_{j=1}^k \omega_j} \quad (4.7)$$

$$R_{\text{seg}} = \text{Norm}(R_{\text{ini}} + R_{\text{boost}} [(R_{\text{ini}} + R_{\text{boost}}) > T_{\text{seg}}]) \quad (4.8)$$

#### 4.2.2 BSNet

由上述道路提升分割的流程可知，学习器将经过若干轮迭代，考虑到计算量问题，设计一种轻量而高效的提升分割网络 BSNet 作为提升分割过程中的弱学习器。如图 4-3 (a) 所示，BSNet 包括三个主要部分：编码部分 (encoder)，解码部分 (decoder)，以及它们之间的中间部分。其输入为  $256 \times 256$  像素的 RGB 三通道的遥感影像，输出利用 Sigmoid 激活函数逐像素判断属于道路的概率，从而得到道路分割概率图。

编码部分利用在 ImageNet 数据集上预训练过的残差网络模型 ResNet 来加速训练过程，具体由 4 组连续堆叠的卷积层及 ReLU 激活函数构成，每组中卷积核大小为  $3 \times 3$ 、步长为 1，各组之间利用步长为 2 的最大池化完成下采样，其中将倒数第二、三组的特征图采样至最后一组特征图的尺寸，然后利用  $1 \times 1$  卷积来调整特征图的通道数并级联 (concatenation) 得到  $16 \times 16 \times 256$  的特征图。

中间部分嵌入空洞率不同的、卷积核尺寸为  $3 \times 3$  的空洞卷积，以串联和并联相结合的方式构成四条路径分别对从编码部分输出的特征图进行处理，其中同一路径采用串联方式，不同路径采用并联方式。第一条路径嵌入空洞率为 1、2 和 4 的空洞卷积，其感受野为  $15 \times 15$ ；第二条路径嵌入空洞率为 1 和 2 的空洞卷积，其感受野为  $7 \times 7$ ；第三条路径嵌入空洞率为 1 的空洞卷积，其感受野为  $3 \times 3$ ；第四条路径采用  $1 \times 1$  卷积调整特征图的通道数；最后将各路径输出的特征图相加，有利于提取来自不同尺度的特征。

解码部分采用数据相关上采样模块 (Data-dependent Upsampling, DUpsampling) [59] 代替传统的转置卷积层，目的是为了在降低计算量的同时保留像素之间的空间关系，该模块是将特征图从最低分辨率转换至原始输入影像尺寸的映射矩阵，其工作原理如图 4-3 (b) 所示。输入 BSNet 的影像宽为  $w$  高为  $h$ ，编码部分的缩放因子 (采样率) 为  $r$ ，则编码器输出的特征图  $F$  的尺寸为  $h/r \times w/r \times Q$ ，利用  $1 \times 1$  卷积调整其通道数为  $Q'$  ( $Q' = r \times r$ )。将特征图  $F'$  中的每个  $1 \times 1 \times Q'$  网格重

塑 (reshape) 为对应的小块  $r \times r \times 1$ 。最终所有小块构成一个特征图  $Y'$ ，其尺寸为  $h \times w \times 1$ ，与输入影像的尺寸一致。

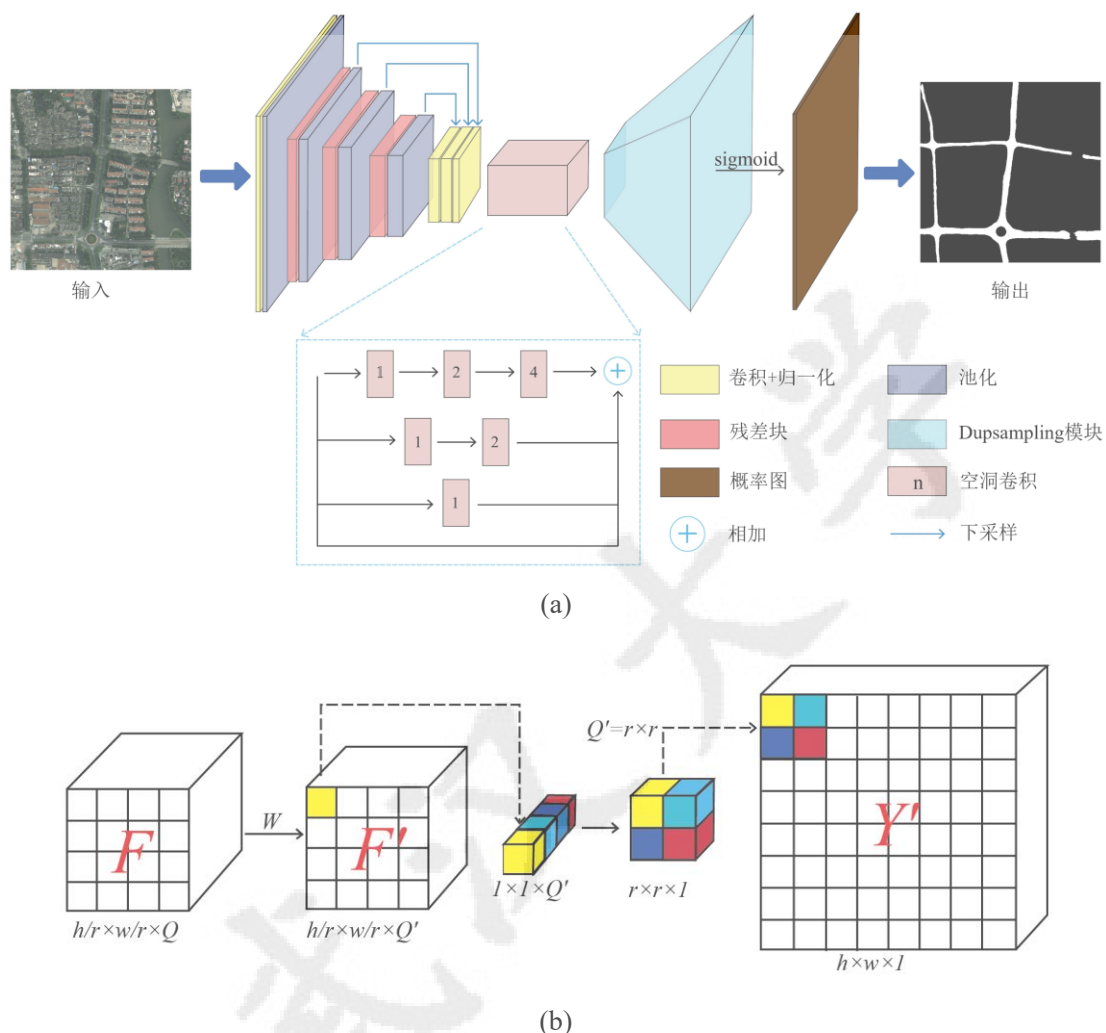


图 4-3 提升分割网络 BSNet 结构图

DUpsampling 模块通过最小化“重构损失函数”  $L_{rec}$  进行预训练，其计算方式见公式 (4.9)，其中  $W$  被定义为将特征图  $F$  转换到最终输出  $Y'$  之间的矩阵， $Z$  被定义为将  $Y$  转换为  $F$  的逆向操作， $Y$  指的是标签。在 BSNet 训练之前，预训练好的  $W$  是固定不变的。

$$L_{rec} = \sum_Y \|Y - Y'\|^2 = \sum_Y \|Y - WZY\|^2 \quad (4.9)$$

训练 BSNet 是通过最小化“分割损失函数”  $L_{seg}$  实现的，它包括“Dice 系数损失函数”(Dice Coefficient Loss, 简称为  $L_{dice}$ )和“二元交叉熵损失函数”(Binary Cross Entropy Loss, 简称为  $L_{BCE}$ )，计算方式见公式 (4.10) - (4.12)，其中 Pred 为预测结果，GT 为标签。

$$L_{\text{dice}} = 1 - \frac{\sum_{i=1}^{\omega} \sum_{j=1}^h |\text{Pred}_{ij} \cap \text{GT}_{ij}|}{\sum_{i=1}^{\omega} \sum_{j=1}^h (|\text{Pred}_{ij}| + |\text{GT}_{ij}|)} \quad (4.10)$$

$$L_{\text{BCE}} = - \sum_{i=1}^{\omega} \sum_{j=1}^h |\text{GT}_{ij} \times \log \text{Pred}_{ij} + (1 - \text{GT}_{ij}) \times \log(1 - \text{Pred}_{ij})| \quad (4.11)$$

$$L_{\text{seg}} = L_{\text{dice}} + L_{\text{BCE}} \quad (4.12)$$

### 4.2.3 提升分割与初始分割的性能比较

按照 4.2.1 和 4.2.2 中描述的提升分割流程以及网络结构, 本小节通过实验对比验证了道路提升分割的有效性。首先讨论了提升分割中迭代轮数对道路分割结果的影响, 然后分别在 Massachusetts、Shaoxing、Cities 这三个道路数据集上对提升分割的性能进行了评估, 并与基于 D-LinkNet 的初始分割相比较。

对于初始分割, 训练阶段 batch 大小固定为 2 个尺寸为  $1024 \times 1024$  像素的样本, 损失函数、网络优化器、学习率等设置均与文献[27]一致。预测阶段使用 TTA 技术, 包括在  $1024 \times 1024$  上进行图像水平翻转、垂直翻转和对角翻转 (即, 每个影像被预测  $2 \times 2 \times 2 = 8$  次), 生成概率图形式的  $R_{\text{ini}}$  用于后续提升分割,  $R_{\text{ini}}$  经过二值化后作为道路初始分割结果。对于提升分割, BSNet 的训练子集由  $256 \times 256$  像素的样本块组成, 编码部分输出的特征图尺寸为  $16 \times 16$ , DUpsampling 模块的采样率为 16。BSNet 训练阶段设置总 epoch 为 200, 如果连续 6 个 epoch 损失不下降, 则提前终止训练; 设置学习率为  $5e-3$ , 前 10 个 epoch 中采用 warm up 策略, 之后每个 epoch 采用 poly 衰减来更新学习率; 测试阶段也采用了 TTA 技术, 生成概率图形式的  $R_{\text{seg}}$ 。

表 4-1 Cities 数据集提升分割中不同  $K$  取值对比

$K$	IoU	Recall	Precision	Conn
0	0.4981	0.5998	<b>0.6708</b>	0.4769
1	0.5215	0.6726	<u>0.6307</u>	0.5391
2	<b>0.5222</b>	<b>0.6794</b>	0.6241	<b>0.5437</b>
3	0.5212	0.6736	0.6286	<u>0.5412</u>
4	<u>0.5218</u>	<u>0.6741</u>	0.6295	0.5411
5	<u>0.5218</u>	0.6738	0.6298	0.5410

提升分割中依次训练了一系列 BSNet, 为了研究 BSNet 总个数 (即总迭代轮



数  $K$ ) 对道路分割结果的影响, 在最大的 **Cities** 数据集上评估了不同  $K$  取值对应的提升分割的性能。结果见表 4-1, 以加粗字体表示最佳, 以下划线字体表示次优。与初始分割 (即  $K=0$ ) 相比, 使用一轮提升分割 (即  $K=1$ ) 后, IoU 提升了 2.3%, Conn 提升了 6.3%。当  $K$  增加时, 提升分割的性能略有变化, 表明了其对参数  $K$  取值具有一定的鲁棒性。另外, 当  $K=2$  时取得了最高的 IoU 和 Conn, 考虑到计算成本, 故在后续相关实验中设置总迭代轮数  $K$  为 2。

表 4-2 初始分割和提升分割比较

数据集	方法	IoU	Recall	Precision	Conn
Massachusetts	初始分割	0.7662	0.8121	0.8267	0.7810
	提升分割	0.7821	0.8455	0.7992	0.8110
Shaoxing	初始分割	0.6144	0.7253	0.8055	0.6411
	提升分割	0.6218	0.7798	0.7571	0.6515
Cities	初始分割	0.4981	0.5998	0.6708	0.4769
	提升分割	0.5222	0.6794	0.6241	0.5437

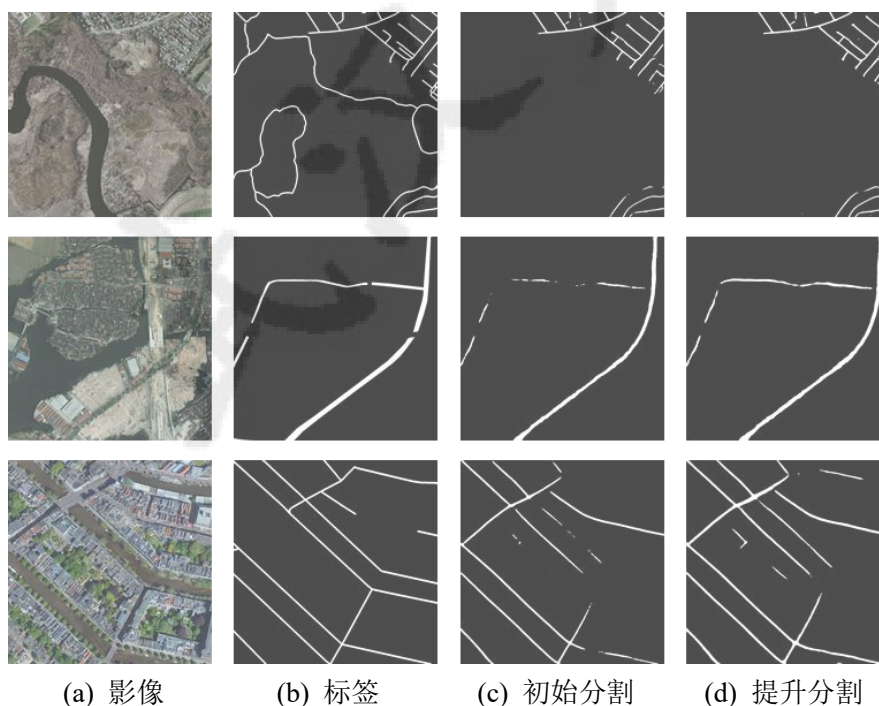


图 4-4 初始分割和提升分割的结果对比示例, 三张  $1024 \times 1024$  影像 (自上而下) 分别来自 Massachusetts、Shaoxing、Cities 数据集

表 4-2 为初始分割和提升分割在 Massachusetts、Shaoxing、Cities 三个数据集上的定量结果。可以看出, 将集成学习技术的提升策略引入道路分割后, 在像素级别以及拓扑级别的评估指标上均有提升。具体地, 在这三个数据集上 IoU 分

别提高了 1.6%、0.7%和 2.4%，Conn 分别提高了 3.0%、1.0%和 6.7%。相应的定性结果对比示例见图 4-4，提升分割的表现优于初始分割，通过从先前分割的失误案例中逐步学习，有效减少误判。

### 4.3 基于卷积神经网络的道路多起点追踪

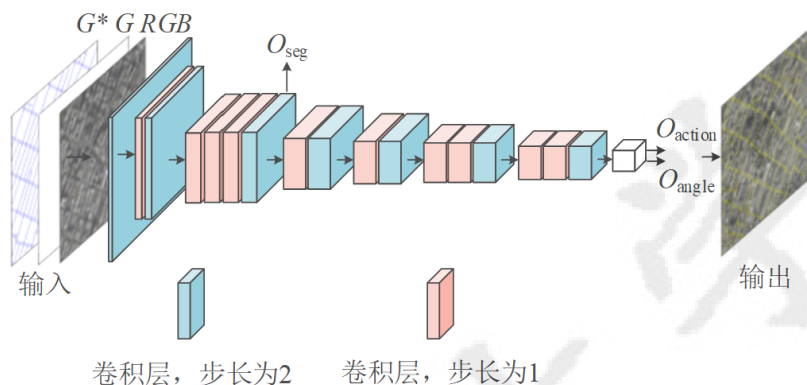
针对诸如 RoadTracer<sup>[33]</sup>的利用拓扑追踪方法进行道路提取所存在的两个缺点：人工经验性选择起始点的方式降低了算法的自动化程度，单起点追踪的方式限制了搜索的完整性，本节改进 RoadTracer 设计了一种道路多起点追踪算法 Multiple Starting Points Tracing，简称为 MSP-Tracer。首先利用道路分割图进行角点提取，自动生成若干个特征点作为后续追踪的起始点，然后基于卷积神经网络从多起点出发追踪道路网。

#### 4.3.1 MSP-Tracer

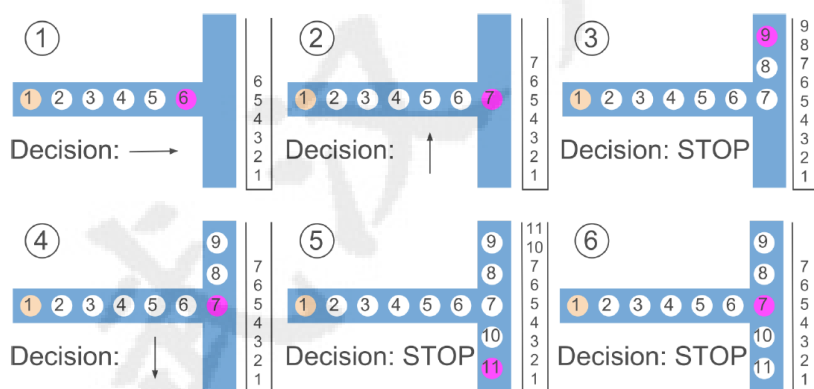
RoadTracer 中利用一个基于卷积神经网络的决策函数（CNN-based Decision Function，简称为 CNN\_decision\_func）迭代式追踪道路中心线，该决策函数是拓扑追踪方法的关键组件，其结构如图 4-5（a）所示。其中，特征提取部分由 7 组连续堆叠的卷积层及 ReLU 激活函数组成，每组中卷积核大小为  $3 \times 3$ ，步长为 1；各组之间卷积核大小为  $3 \times 3$ ，步长为 2 完成下采样。CNN\_decision\_func 的输入包括以当前点为中心  $d \times d$  像素的搜索窗口对应的 RGB 影像、当前搜索到的中心线结果  $G$ 、道路中心线标签  $G^*$ ，输出部分通过 Softmax 函数和 Sigmoid 函数分别输出“追踪动作”  $O_{\text{action}} = \langle O_{\text{walk}}, O_{\text{stop}} \rangle$ 、“追踪角度”  $O_{\text{angle}}$ 。另外，网络还输出粗略的中间分割结果  $O_{\text{seg}}$  以约束训练过程。综上，用于优化整个 CNN\_decision\_func 的损失函数包括动作、角度方面的均方误差和分割方面的交叉熵损失。

追踪的工作原理如图 4-5（b）所示，以 T 型道路交叉口为例，蓝色代表道路，橙色点代表已知属于道路的起始点，紫色点代表当前搜索点，箭头代表追踪角度  $O_{\text{angle}}$ ，Decision 代表决策的当前追踪动作  $O_{\text{action}}$ （walk 或 stop）；图中右侧还有一个“栈”用于存放追踪的顶点。栈是一种经典的数据结构，只能在栈顶进行插入和删除操作。在追踪的每一步中，以当前顶点为中心开辟搜索窗口，将其对应的 RGB 影像、 $G$  和  $G^*$  输入 CNN\_decision\_func，输出  $O_{\text{angle}}$  和  $O_{\text{action}}$ 。其中， $O_{\text{angle}}$  是可追踪的角度的集合，如果这些角度对应的边都已被追踪过，则  $O_{\text{action}}$  为停止（stop），将当前顶点从栈中删除（出栈 pop），然后将栈顶的点作为新的当

前顶点；如果  $O_{angle}$  中有某些（个）角度对应的边未被追踪过，则  $O_{action}$  为前进（walk），按照其中概率最高的一个追踪角度前进固定距离  $D = 12\text{ m}$ ，获得新的当前顶点并将它插入栈中（入栈 push）。更新以新的当前顶点为中心的搜索窗口，重复上述过程以进行下一步预测。随着搜索的进行，将追踪过的顶点和连接相邻顶点的边线添加到  $G$  中，逐步构建道路网，直到遍历完栈中的所有顶点为止。



(a) 用于追踪的决策函数 CNN\_decision\_func 结构图



(b) 追踪示例图

图 4-5 RoadTracer 追踪原理

与 RoadTracer 在预测时经验性选择一个道路上已知点不同的是，本节所提出的 MSP-Tracer 在此应用 Shi-Tomas<sup>[60]</sup>角点检测算子自动提取多个特征点。这一过程基于在 4.2 节介绍的道路提升分割。不过，角点检测算子并不是直接应用在道路分割图上，而是首先对道路分割图进行形态学细化操作后得到单像素宽的分割图，然后再应用角点检测算子。这样做目的是为了检测高质量的特征点，比如位于交叉路口等特殊位置的特征点，它们能提供更多潜在的追踪方向，更适合作为后续追踪的起始点。

角点是两条边缘的交点，表示两条边方向改变的地方，角点在任意一个方向上做微小移动都会引起该区域的梯度图的方向和幅值发生很大变化。基于梯度计

算的角点检测一般利用局部窗口在图像上进行移动判断灰度发生较大的变化,为每个窗口定义一个角点响应函数  $\mathbf{R}$ ,如果  $\mathbf{R}$  值超过阈值则被认为对应角点特征。Shi-Tomasi 角点检测算子是一种改进的 Harris 角点检测算子。Harris 算子的角点响应函数如公式 (4.13) 所示,其中  $\lambda_1$  和  $\lambda_2$  是矩阵  $M$  的特征值,如公式 (4.14)  $M$  是梯度  $I$  的协方差矩阵。利用水平、竖直差分算子对图像的每个像素进行滤波以求得  $I_x$  和  $I_y$ ,进而求得矩阵  $M$  中的四个元素的值。但是 Harris 算子的稳定性和参数  $k$  有关,而  $k$  是个经验值,难以设定最佳值。基于此,Shi 和 Tomasi<sup>[60]</sup>发现角点的稳定性与  $M$  的较小特征值有关,提出 Shi-Tomasi 算子,其角点响应函数如公式 (4.15) 所示。

$$\mathbf{R} = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (4.13)$$

$$M = \sum_{x,y} \omega(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \quad (4.14)$$

$$\mathbf{R} = \min(\lambda_1, \lambda_2) \quad (4.15)$$

将 Shi-Tomasi 算子生成的多个特征点放入“起点列表”,从其中随机选择一个点作为待追踪的起始点,将它添加到“栈”中并将其从“起点列表”中删除。然后利用训练好的 CNN\_decision\_func 从该点出发追踪道路,在此轮追踪停止前再从起点列表中随机选一个点作为新一轮追踪的起始点,继续上述过程,直到起点列表为空。考虑到计算成本,本节还提出了一个自适应起始点决策 (Adaptive Starting Point Decision, 简称为 ASPD) 策略,该策略根据先前追踪已构建的道路来动态地挑选下一轮追踪的起始点。具体而言,检查当前搜索到的道路网是否在以候选起始点为中心的边界框之外。如果边界框与当前搜索到的道路网有重叠,则将此候选起始点从起点列表中删除;如果二者没有交集,则此候选起始点可作为新一轮追踪的起始点。

需要注意,本节在 CNN\_decision\_func 的模型结构和训练方面与 RoadTracer 相似,而创新工作主要围绕在预测阶段,即在 CNN\_decision\_func 训练完成后如何从若干个特征点出发构建更完整的道路。这也是本节所提出的 MSP-Tracer 解决的问题,具体可参见 Algorithm 2 中的伪代码。

**Algorithm 2. Multiple Starting Points Tracing (MSP-Tracer)**

**Input:** starting points list  $V$ , window  $W$ , graph  $G = \Phi$ , vertex stack  $S = \Phi$ , move distance  $D$ , bounding box  $B$ .

```

while  $V$  is not empty, do
  randomly pick  $V_i$  from  $V$ 
   $S = V_i$ 
  initialize  $W_i$  centered at  $V_i$ 
  if  $G$  intersect with  $W_i$ ; break
  else
    while  $S$  is not empty, do
      action,  $\alpha = \text{decision\_func}(G, S_{\text{top}}, \text{Image})$ 
       $u = S_{\text{top}} + (D \cos\alpha, D \sin\alpha)$ 
      if action == stop or  $u$  is outside  $B$  then
        pop  $S_{\text{top}}$  from  $S$ 
      else
        add vertex  $u$  to  $G$ 
        add an edge  $(S_{\text{top}}, u)$  to  $G$ 
        push  $u$  onto  $S$ 
      end if
    end while
  end if
  remove  $V_i$  from  $V$ 
return  $G$ 
  
```

### 4.3.2 MSP-Tracer 与 RoadTracer 的性能比较

按照 4.3.1 中描述的 MSP-Tracer 流程以及网络设置，本小节分别在 Massachusetts、Shaoxing、Cities 这三个道路数据集上对 MSP-Tracer 的性能进行了评估，并与基于 RoadTracer 的单起点追踪相比较。

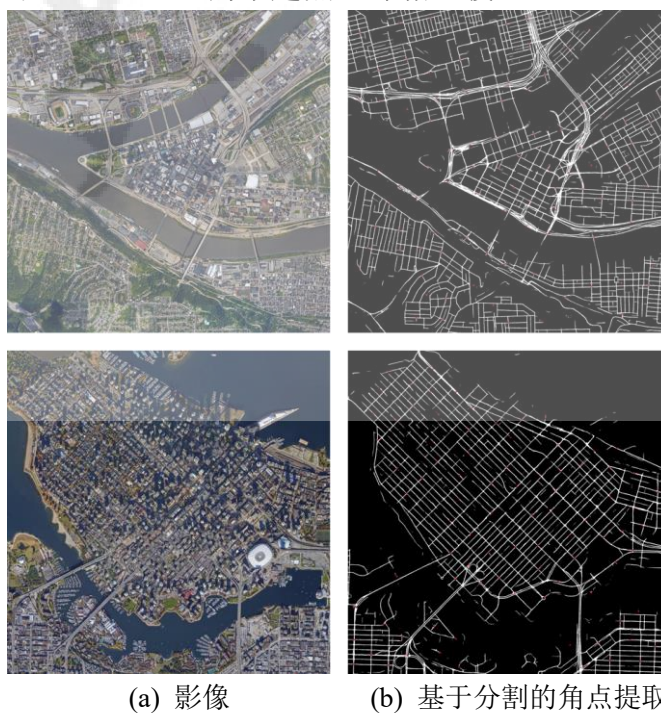


图 4-6 Cities 数据集的匹兹堡（上）和温哥华（下）可视化示例，红点为提取的多个特征点

对于 Shi-Tomasi 角点检测算子的实现，本文借助了计算机视觉库 OpenCV。针对一张 8192×8192 像素的分割图，将角点的预期数量设置为 100，角点之间的最小欧拉距离设置为 400 像素，避免得到相邻特征点。图 4-6 展示的是在 Cities 数据集的匹兹堡和温哥华两个城市上进行角点提取的可视化效果。对于 ASPD 策略，将搜索边界框的长度设置为 60 像素，这是每个路段长度（20 像素）的 3 倍。

表 4-3 RoadTracer 和 MSP-Tracer 比较

数据集	方法	Comp	Corr	Qual
Massachusetts	RoadTracer	0.435	0.513	0.308
	MSP-Tracer	0.488	0.552	0.343
Shaoxing	RoadTracer	0.168	0.413	0.150
	MSP-Tracer	0.378	0.607	0.298
Cities	RoadTracer	0.229	0.371	0.170
	MSP-Tracer	0.264	0.416	0.195

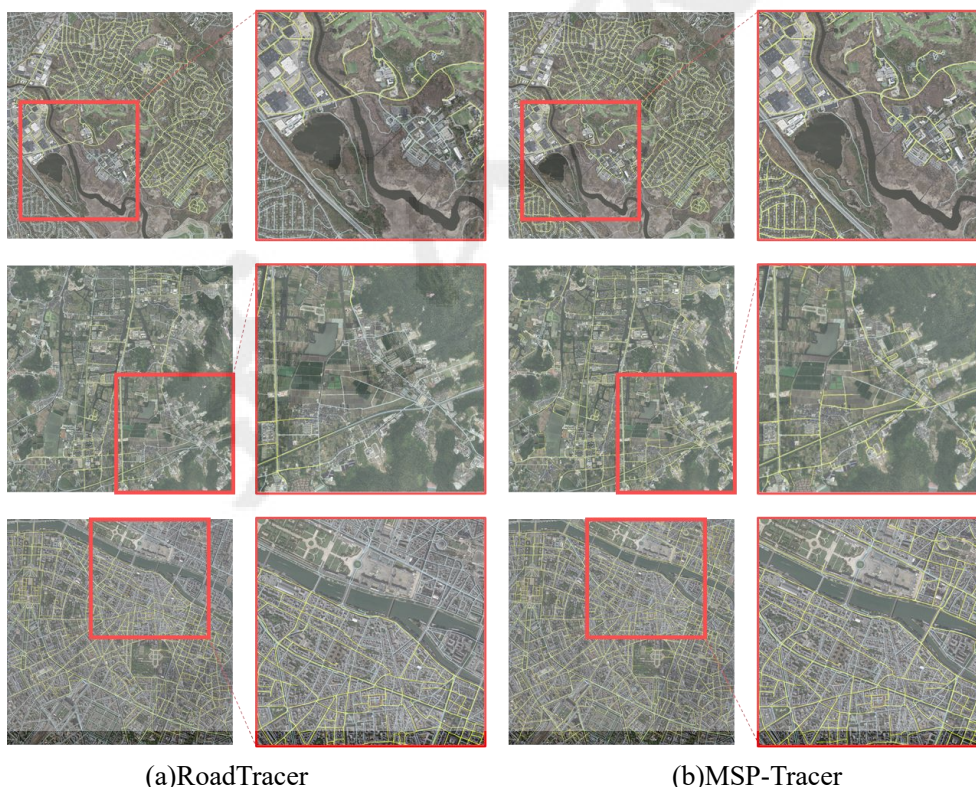


图 4-7 RoadTracer 和 MSP-Tracer 的结果对比示例，浅蓝色表示标签，黄色表示追踪结果，三张影像（自上而下）分别来自 Massachusetts、Shaoxing、Cities 数据集

对于拓扑追踪方法，视野范围较大的图像可以保留更好的道路完整性和连通性。将搜索窗口大小设置为  $d = 256$  像素。在 CNN\_decision\_func 的训练阶段，每批次训练影像数量设置为 4，训练了约 400 个 epoch。初始学习率设置为  $1e-5$ ，

随后每 100 个 epoch 更新一次。在参数更新中，采用自适应矩估计算法(Adaptive Moment Estimation, Adam)<sup>[61]</sup>进行随机梯度下降优化。与 RoadTracer<sup>[33]</sup>相似，输出角度包含 64 个均匀分布的方向，选择具有最大概率的角度作为移动方向  $O_{angle}$ ；将动作输出的阈值设置为 0.4，这意味着如果  $O_{walk}$  高于此阈值则从该点前进 20 像素，否则将停止并退回到上一个节点。

表 4-3 为 RoadTracer 和 MSP-Tracer 在 Massachusetts、Shaoxing、Cities 三个数据集上的定量结果，其中  $\rho=4$  像素。与 RoadTracer 相比，在应用多个起始点追踪 (MSP-Tracer) 之后，道路结果的精度大大改善。在 Shaoxing 数据集上，MSP-Tracer 相比 RoadTracer，Comp 提高了 21.0%，Corr 提高了 19.4%，Qual 提高了 14.8%。在 Massachusetts 和 Cities 数据集上，MSP-Tracer 的表现同样优于 RoadTracer。相应的定性结果对比示例见图 4-7，前两列为 RoadTracer 的结果，后两列为 MSP-Tracer 的结果。浅蓝色的线表示标签，黄色的线表示追踪结果。为了便于观察，在第一列、第三列的影像中随机选择红色框对应区域放大如第二列、第四列所示。可以看出，多起点追踪的 MSP-Tracer 能够构建更完整的道路，对桥梁、遮挡等复杂场景具有一定的鲁棒性，缓解了 RoadTracer 单起点追踪生成的道路覆盖范围有限的问题。

## 4.4 道路分割和追踪的融合

### 4.4.1 融合策略

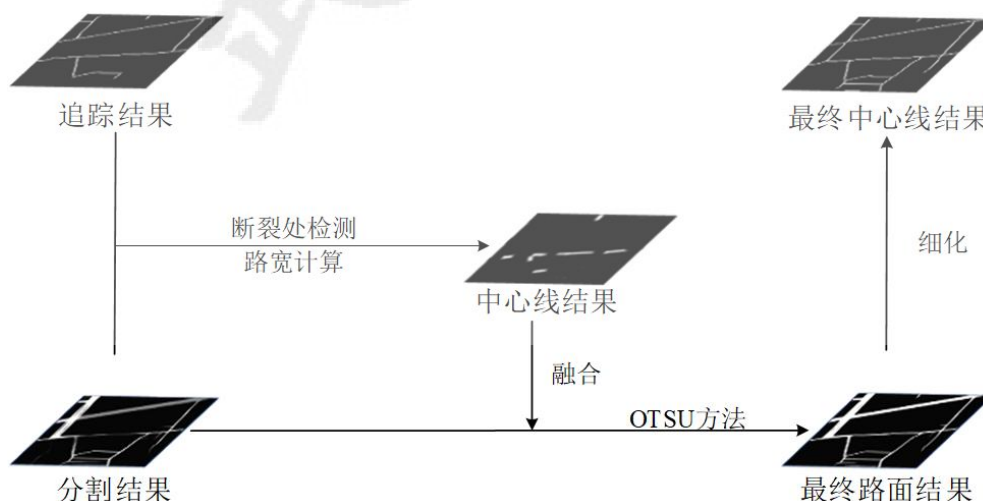


图 4-8 道路分割和追踪的融合

如前文所述及实验验证，语义分割方法逐像素预测道路，其优势是能够检测到大多数道路像素，但由于忽略了道路的拓扑结构信息，这类方法会生成许多不

连续的破碎路段；而拓扑追踪方法的优势是保留道路的连通性，但由于搜索过程受限于起始点的质量以及遥感影像的复杂性等，这类方法生成的道路完整性较差。以 D-LinkNet<sup>[27]</sup>和 RoadTracer<sup>[33]</sup>为基准方法，4.2 节介绍的道路提升分割和 4.3 节介绍的道路多起点追踪分别对两类方法做了改进，有效提升了道路提取的准确性。本节关注于语义分割和拓扑追踪两类方法的优势互补，提出了一种道路分割和追踪的融合方法。该方法旨在结合提升分割和多起点追踪提取的道路特征，实现道路语义信息和拓扑信息的融合，进一步促进算法的鲁棒性，最终同时提取更准确、更可靠的路面结果和中心线结果。

融合的流程参见图 4-8。首先借助多起点追踪结果来检测提升分割结果中的断裂路段。追踪结果是矢量形式的中心线，由顶点和边线构成。针对中心线的每个路段（即边线），逐段判断是否与分割结果存在交集，记录下那些与分割结果“部分相交”的路段位置，认为它们对应于分割结果中的“断裂”。然后，以一定宽度（例如，比道路实际宽度略宽的 11 像素）为每个断裂路段  $r$  创建缓冲区并用  $\text{Buffer}^r$  表示，根据公式（4.16）计算每个断裂路段  $r$  的合适路宽  $\text{width}^r$ 。

$$\text{width}^r = \frac{\text{area}^r}{\text{length}^r} = \frac{\text{num}(\hat{R}_{\text{seg}} \cap \text{Buffer}^r)}{\text{num}(\hat{R}_{\text{seg}} \cap R_{\text{cen}}^r)} \quad (4.16)$$

其中，分割结果用  $\hat{R}_{\text{seg}}$  表示，对追踪结果栅格化得到单像素宽的中心线，将断裂路段  $r$  对应的中心线记为  $R_{\text{cen}}^r$ ， $\text{num}(\cdot)$  是计算相交区域的像素数，将  $\hat{R}_{\text{seg}}$  和  $\text{Buffer}^r$  相交区域的面积记为  $\text{area}^r$ ， $\hat{R}_{\text{seg}}$  和  $R_{\text{cen}}^r$  相交区域的面积记为  $\text{length}^r$ （ $R_{\text{cen}}^r$  是单像素宽的中心线）。

根据合适路宽  $\text{width}^r$  将  $R_{\text{cen}}^r$  进行形态学膨胀，得到具有一定宽度的中心线结果记为  $R_{\text{cen}}$ （取值 0,1），利用  $R_{\text{cen}}$  和在 4.2.1 节介绍的提升分割结果概率图  $R_{\text{seg}}$ （取值 0-1）根据公式（4.17）进行融合，得到最终路面结果的概率图  $R_{\text{fuse}}$ 。其中  $T_{\text{fuse}}$  是 4.2.1 节中对  $R_{\text{seg}}$  应用 OTSU 算法阈值分割时的自适应阈值。该公式的工作原理是，对于  $R_{\text{fuse}}$ ，将在  $R_{\text{seg}}$  中概率值大于  $T_{\text{fuse}}$  且在  $R_{\text{cen}}$  上值为 1 的像素的值设为 1，即认为在分割概率图上置信度高且有中心线的像素一定是路。最后再结合 OTSU 算法得到二值化的最终路面结果，再通过形态学细化方法从道路路面结果中得到最终的道路中心线结果。

$$R_{\text{fuse}} = R_{\text{seg}} [(R_{\text{seg}} + R_{\text{cen}}) \leq (T_{\text{fuse}} + 1)] + R_{\text{cen}} [(R_{\text{seg}} + R_{\text{cen}}) > (T_{\text{fuse}} + 1)] \quad (4.17)$$



该方法融合了提升分割和多起点追踪的结果，图 4-9 展示了该方法的两个优点。1) 如图 4-9 (a) 所示，对于分割结果不连续的路段  $r$ ，可以通过追踪图提供的拓扑信息，利用公式 (4.16) 计算出的道路宽度  $width^r$  将分割图上的不连续路段连接起来。2) 如图 4-9 (b) 所示，根据公式 (4.17) 和形态学细化操作，分割结果可帮助连接断开的或追踪未检测出的中心线。

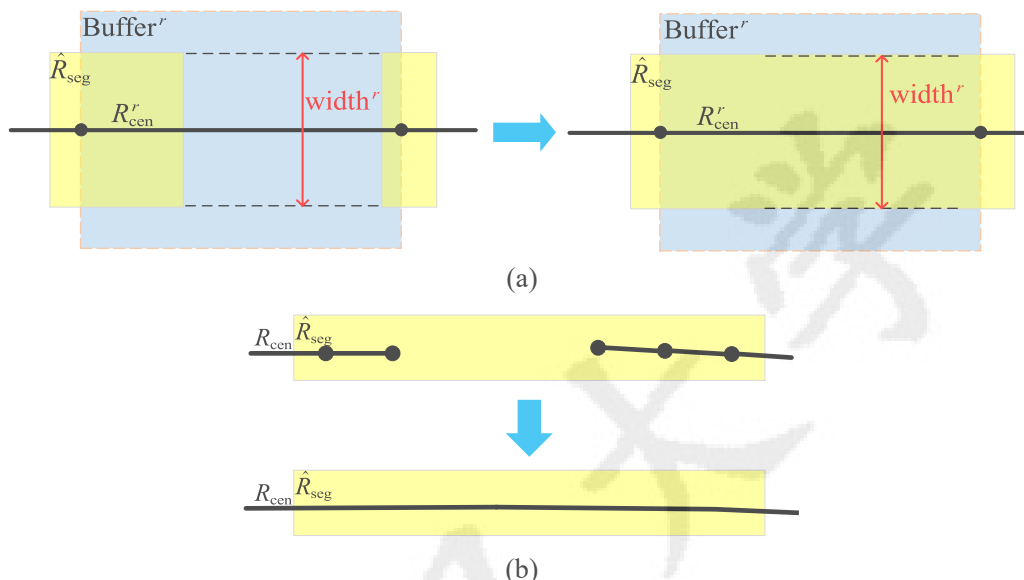


图 4-9 融合的优势：(a) 通过中心线的缓冲区（浅蓝色）增强分割结果（黄色）的连接性、(b) 通过分割图改善中心线（黑线）的完整性。

#### 4.4.2 融合前后的性能比较

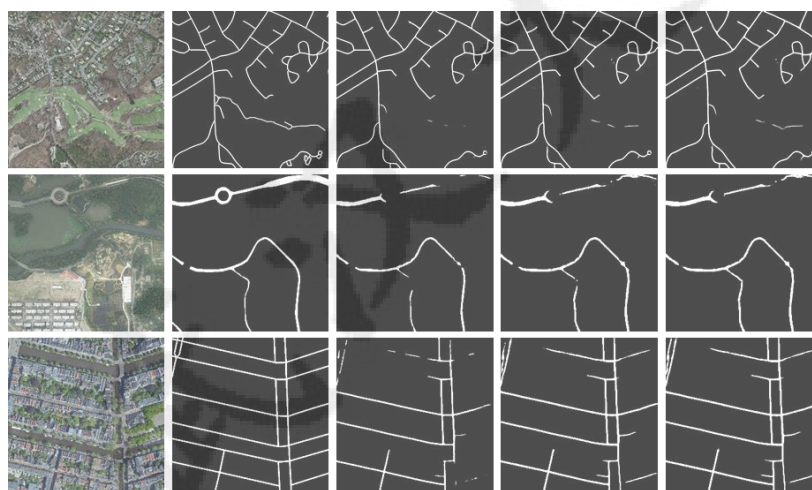
按照 4.4.1 中描述的融合流程和参数设置，本小节分别在 Massachusetts、Shaoxing、Cities 这三个道路数据集上对融合后（即 Seg-Tracer）的性能进行了评估，并分别与仅使用语义分割、仅使用拓扑追踪的方法相比较。

首先，表 4-4 展示了 Seg-Tracer 与仅使用语义分割方法对比的定量结果，可以看出本文提出的提升分割和融合策略都超过了基准方法 D-LinkNet。4.2.3 小节已分析了提升分割的有效性，所以在此主要关注引入了融合策略后给语义分割方法带来的效果，即提升分割与 Seg-Tracer 之间的对比。试验表明，虽然提升分割的表现优于初始分割，但仍受限于语义分割方法的特点（忽略道路的拓扑结构信息），而与拓扑追踪融合后的 Seg-Tracer 可以进一步提高道路分割的精度，特别是在连接性指标 Conn 上。Seg-Tracer 与 4.2 节提出的提升分割相比，Massachusetts 数据集上 IoU 提高了 0.44%，Conn 提高了 1.28%；Shaoxing 数据集上 IoU 降低了 0.4%，Conn 提高了 1.48%；Cities 数据集上 IoU 提高了 0.25%，Conn 提高了 1.08%。整体来讲，与拓扑追踪融合后的 Seg-Tracer 在 IoU 上的提升不及 Conn 提

升明显，但是道路连接性是利用语义分割方法实现道路自动化提取的关键指标，因此 Conn 的提升至关重要。相应的定性结果示例见图 4-10。

表 4-4 Seg-Tracer 与仅使用语义分割的性能比较

数据集	方法	IoU	Recall	Precision	Conn
Massachusetts	D-LinkNet	0.7662	0.8121	0.8267	0.7810
	提升分割 (Ours)	0.7821	0.8455	0.7992	0.8110
	Seg-Tracer (Ours)	0.7865	0.8588	0.7847	0.8238
Shaoxing	D-LinkNet	0.6144	0.7253	0.8055	0.6411
	提升分割 (Ours)	0.6218	0.7798	0.7571	0.6515
	Seg-Tracer (Ours)	0.6178	0.7775	0.7519	0.6663
Cities	D-LinkNet	0.4981	0.5998	0.6708	0.4769
	提升分割 (Ours)	0.5222	0.6794	0.6241	0.5437
	Seg-Tracer (Ours)	0.5247	0.6907	0.6173	0.5545



(a) 影像 (b) 标签 (c) D-LinkNet (d) 提升分割 (e) Seg-Tracer

图 4-10 Seg-Tracer 与仅使用语义分割的结果对比示例

然后，表 4-5 展示了 Seg-Tracer 与仅使用拓扑追踪方法对比的定量结果，可以看出本文提出的多起点追踪和融合策略都超过了基准方法 RoadTracer。4.3.2 小节已分析了 MSP-Tracer 的有效性，所以在此主要关注引入了融合策略后给拓扑追踪方法带来的效果，即 MSP-Tracer 与 Seg-Tracer 之间的对比。试验表明，虽然多起点追踪的表现明显优于单起点追踪，但仍受限于拓扑追踪方法的特点（提取的道路结果覆盖范围有限），而与语义分割融合后的 Seg-Tracer 可以大幅提高道路追踪的精度。Seg-Tracer 与 4.3 节提出的多起点追踪相比，Qual 在 Massachusetts、Shaoxing、Cities 数据集上分别提高了 45.3%，35.2%和 34.6%。

相应的定性结果示例见图 4-11，显然，与语义分割融合后的 Seg-Tracer 得到了更规则化的道路，且靠近影像边界的道路也能够被提取出来而不受追踪过程搜索窗口的限制。

表 4-5 Seg-Tracer 与仅使用拓扑追踪的性能比较

数据集	方法	Comp	Corr	Qual
Massachusetts	RoadTracer	0.435	0.513	0.308
	MSP-Tracer (Ours)	0.488	0.552	0.343
	Seg-Tracer (Ours)	0.889	0.882	0.796
Shaoxing	RoadTracer	0.168	0.413	0.150
	MSP-Tracer (Ours)	0.378	0.607	0.298
	Seg-Tracer (Ours)	0.823	0.757	0.650
Cities	RoadTracer	0.229	0.371	0.170
	MSP-Tracer (Ours)	0.264	0.416	0.195
	Seg-Tracer (Ours)	0.701	0.693	0.541

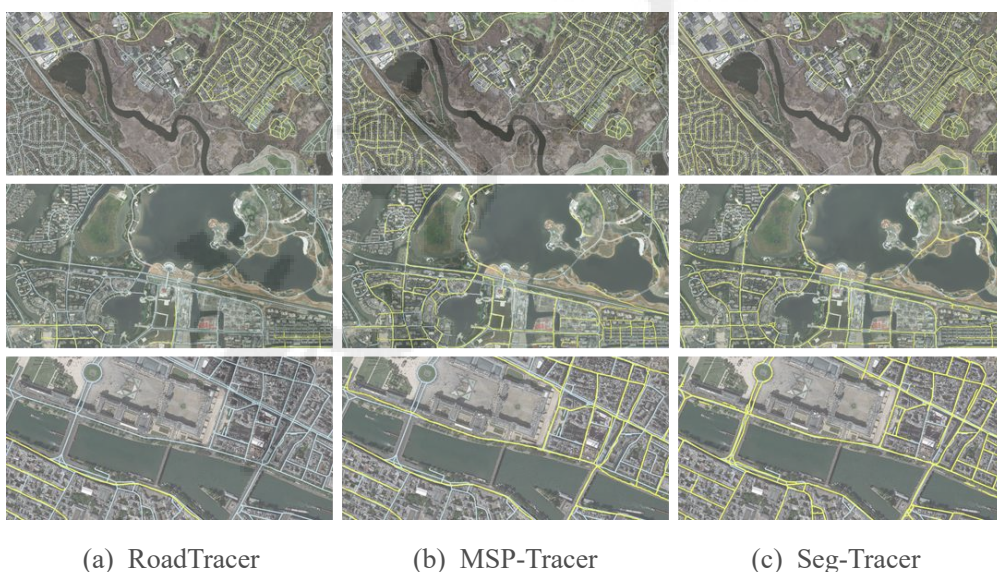


图 4-11 Seg-Tracer 与仅使用拓扑追踪的结果对比示例，浅蓝色表示标签，黄色表示结果

另外，讨论  $\rho$  的不同取值对道路中心线提取方法性能评估的影响，以 Qual 为例，图 4-12 展示了 Massachusetts 数据集不同道路中心线提取方法在  $\rho = 1$ 、 $\rho = 2$  和  $\rho = 4$  像素时的性能评估。结果表明，无论  $\rho$  取 1、2、4，与语义分割融合后的 Seg-Tracer 表现均优于 RoadTracer 和 MSP-Tracer，且相比 RoadTracer 的 Qual 提高了至少 40%。随着  $\rho$  的增加，Qual 也在增加。对于 Seg-Tracer，当  $\rho$  从 2 像素变为 4 像素时，Qual 仅有轻微提升，因此不会存在过度评估的情况。后续实验均设置  $\rho = 4$ 。



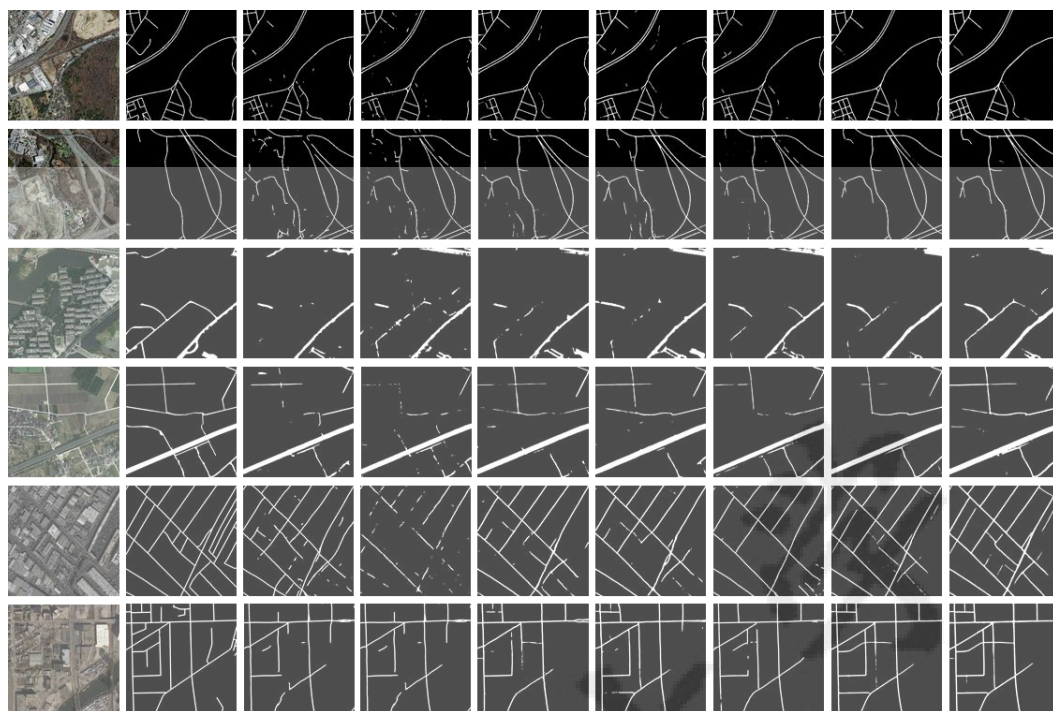
图 4-12 Massachusetts 数据集不同道路中心线提取方法在 Qual 上的性能对比

## 4.5 实验分析与讨论

为了验证本节所提出的 Seg-Tracer 方法的有效性和先进性，在此与现有的基于深度学习的道路提取方法进行比较，评估这些方法的性能。

表 4-6 Seg-Tracer 与其他的道路分割方法比较

数据集	方法	IoU	Recall	Precision	Conn
Massachusetts	U-Net <sup>[20]</sup>	0.6735	0.7518	0.7738	0.6245
	ResUnet <sup>[22]</sup>	0.6998	0.7484	<u>0.8292</u>	0.6164
	ASPP-Unet <sup>[23]</sup>	0.7061	0.7655	0.8011	0.7015
	RCNN-Unet <sup>[25]</sup>	<u>0.7067</u>	0.7786	0.7814	0.7009
	LinkNet <sup>[26]</sup>	0.7311	0.7697	<b>0.8451</b>	0.6784
	D-LinkNet <sup>[27]</sup>	<u>0.7662</u>	<u>0.8121</u>	0.8267	<u>0.7810</u>
	Seg-Tracer (Ours)	<b>0.7865</b>	<b>0.8588</b>	0.7847	<b>0.8238</b>
Shaoxing	U-Net <sup>[20]</sup>	0.4563	0.5677	0.7007	0.4276
	ResUnet <sup>[22]</sup>	0.4666	0.6173	0.6494	0.2992
	ASPP-Unet <sup>[23]</sup>	0.5087	0.6869	0.6532	0.5028
	RCNN-Unet <sup>[25]</sup>	0.5025	0.6484	0.6870	0.5002
	LinkNet <sup>[26]</sup>	0.5497	0.6428	<u>0.7882</u>	0.4961
	D-LinkNet <sup>[27]</sup>	<u>0.6144</u>	<u>0.7253</u>	<b>0.8055</b>	<u>0.6411</u>
	Seg-Tracer (Ours)	<b>0.6178</b>	<b>0.7775</b>	0.7519	<b>0.6663</b>
Cities	U-Net <sup>[20]</sup>	0.3384	0.4674	0.4961	0.2450
	ResUnet <sup>[22]</sup>	0.2615	0.2979	0.6413	0.1366
	ASPP-Unet <sup>[23]</sup>	0.4283	0.5514	0.5917	0.3858
	RCNN-Unet <sup>[25]</sup>	0.4325	0.5560	0.5925	0.3908
	LinkNet <sup>[26]</sup>	0.3926	0.4638	<u>0.6514</u>	0.3374
	D-LinkNet <sup>[27]</sup>	<u>0.4981</u>	<u>0.5998</u>	<b>0.6708</b>	<u>0.4769</u>
	Seg-Tracer (Ours)	<b>0.5247</b>	<b>0.6907</b>	0.6173	<b>0.5545</b>



(a)影像 (b)标签 (c) [20] (d) [22] (e) [23] (f) [25] (g) [26] (h) [27] (i) Ours

图 4-13 Seg-Tracer 与其他的道路分割方法的结果对比示例，自上而下六张影像每两张为一组，每组分别来自 Massachusetts、Shaoxing、Cities 数据集

利用语义分割进行道路提取的方法包括 U-Net<sup>[20]</sup>、ResUnet<sup>[22]</sup>、ASPP-Unet<sup>[23]</sup>、RCNN-Unet<sup>[25]</sup>、LinkNet<sup>[26]</sup>和 D-LinkNet<sup>[27]</sup>，表 4-6 列出了针对 Massachusetts、Shaoxing、Cities 数据集的不同道路分割方法的性能对比，以加粗字体表示最佳，以下划线字体表示次优。

首先可以看出，与 U-Net, ResUnet, ASPP-Unet, RCNN-Unet 和 LinkNet 相比，D-LinkNet 在 IoU 和 Conn 这两个指标上均表现出最佳的性能。对于标签均是由中心线等宽膨胀得到的 Massachusetts 和 Cities 数据集，Seg-Tracer 相比 D-LinkNet, IoU 分别提高了 2.03%和 2.66%，Conn 分别提高了 4.28%和 7.76%。对于标签与真实道路对齐的 Shaoxing 数据集，Seg-Tracer 超过 D-LinkNet 方法 0.34% 的 IoU 和 2.52%的 Conn。

图 4-13 展示了在不同数据集上不同道路分割方法的定性结果示例，共有 6 行×9 列的子图。前两行来自 Massachusetts 数据集，第三行第四行来自 Shaoxing 数据集，后两行来自 Cities 数据集。从第 (c) - (g) 列可以看出，U-Net、ResUnet、ASPP-Unet、RCNN-Unet 和 LinkNet 难以从真实道路区域中区分同质区域。其中，ResUnet (d) 在 Cities 数据集上表现最差。如第 (h) 列所示，D-LinkNet 消除了前面几种方法的大多数误判，但在树木遮挡等复杂场景下仍生成了破裂路段。与

D-LinkNet 相比，本节提出的 Seg-Tracer 得到了更连贯的道路区域以及更平滑的道路边界，这表明 Seg-Tracer 对复杂场景具有更强的鲁棒性，在提升分割以及与拓扑追踪的融合操作下，有效改善了道路连通性并获得更加结构化的道路网。

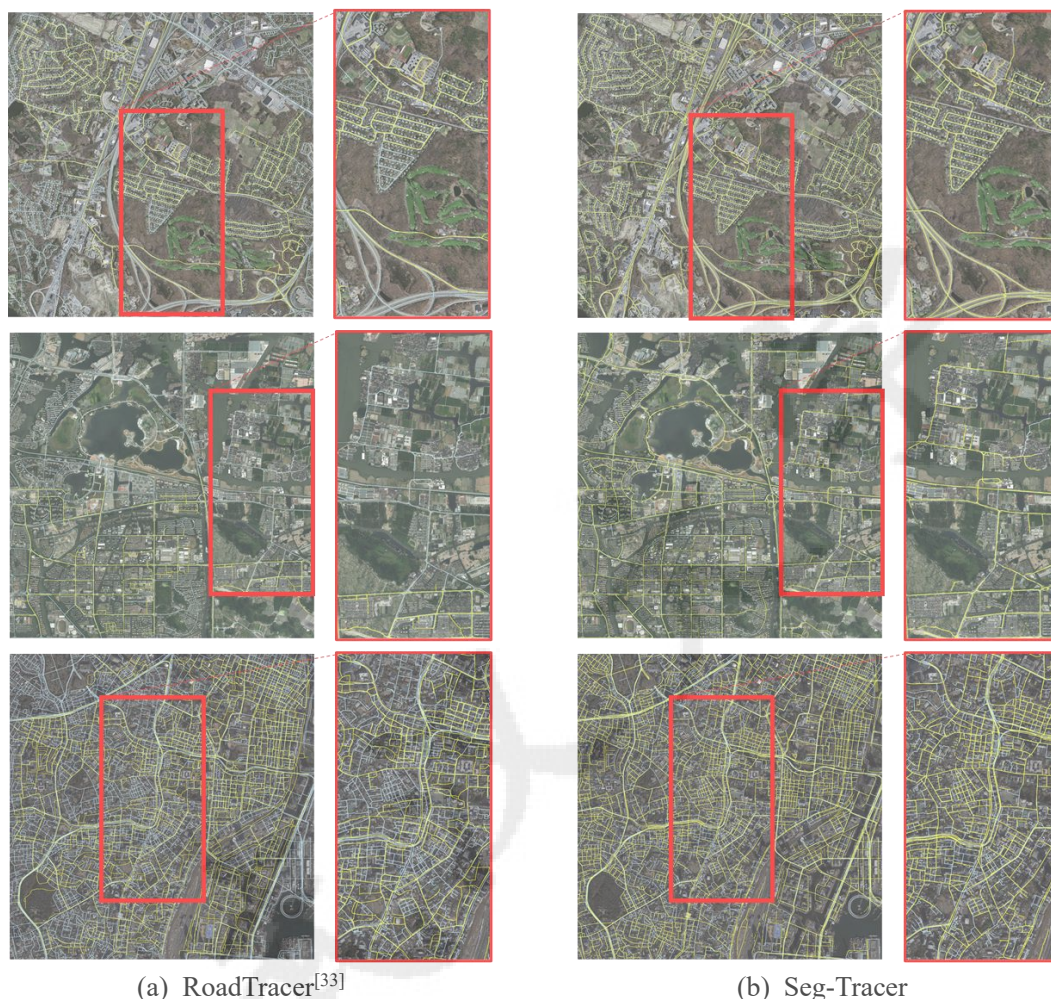


图 4-14 Seg-Tracer 与道路追踪方法 RoadTracer 的结果对比示例，浅蓝色表示标签，黄色表示结果，三张影像自上而下分别来自 Massachusetts、Shaoxing、Cities 数据集

利用拓扑追踪进行道路提取的方法目前只有 RoadTracer，定量对比结果可见表 4-5。在此展示 RoadTracer 与 Seg-Tracer 的定性结果示例，如图 4-14 所示。RoadTracer 单起点追踪的方式使得追踪容易过早中断，导致道路结果的覆盖范围有限，特别是如图 4-14 的第一、二行所示，位于靠近影像边界区域的道路难以被提取出。另外，如图 4-14 的第三行所示，拓扑追踪方法虽保留了道路拓扑结构信息，但由于遥感影像的复杂性其在道路密集的城区出现了几处误判，导致道路不规则。与 RoadTracer 相比，本节提出的 Seg-Tracer 得到了覆盖范围更广、更规则的道路，在多起点追踪以及与语义分割的融合操作下，有效改善了道路完整性并获得更加结构化的道路网。

## 4.6 本章小结

本章对结合语义分割和拓扑追踪的道路提取算法进行研究,详细介绍了基于全卷积神经网络的道路提升分割、基于卷积神经网络的道路多起点追踪,以及两类方法的融合策略。在不同的道路数据集上进行实验验证,并与仅使用语义分割或拓扑追踪的方法进行对比,探究所提出的 Seg-Tracer 算法的有效性和先进性。

武汉大学

## 第 5 章 基于弱监督深度学习的道路路面提取算法

### 5.1 引言

#### 5.1.1 现有方法存在的问题

在前期研究中发现，遥感影像道路提取任务一般围绕道路的路面和（或）中心线数据的自动化生成。道路路面提取关注于记录道路的语义特征，包括道路的宽度等；道路中心线提取关注于记录道路的结构特征，尤其是与智慧交通建设等应用息息相关的道路连通性。近年来，OpenStreetMap（OSM）等公开地图平台开源了越来越多的道路中心线数据，但开源的道路路面数据却依然比较少。

现有道路路面提取的深度学习方法多为基于全监督学习的语义分割方法，需要逐像素标注的大量数据去训练神经网络模型学习目标特征。考虑到人工标注是一个相当耗时耗力的繁琐过程且对专业领域知识要求较高，遥感影像数据量的增长以及空间分辨率的提高也使得标注数据不足或不准确的情况逐渐增多，大大限制了全监督深度学习方法的泛化能力，特别是难以实现研究成果转化及大规模推广应用。显然，利用低标注成本（即，减少对高质量标注数据的依赖）实现道路路面的准确提取，是该领域亟待解决的问题。

作为计算机视觉领域的研究热点，弱监督学习技术为解决上述全监督深度学习方法的问题提供了新方案，不过也存在一定的不足。以线状弱标签 scribble 为例，经典的 ScribbleSup<sup>[38]</sup>由于采用了交替优化方案而导致计算量增加且提取的目标边缘并不规则。另外，由于遥感影像的复杂性，直接将弱监督深度学习应用到道路路面提取的研究尚未成熟。对于极少的一个与本章研究目标相近的文献 [50]，其提出的“固定道路宽度”假设也仅能处理特定数据，难以应对道路宽度多变的更复杂的影像数据。

#### 5.1.2 技术路线和创新点

为了突破上述基于全监督深度学习的局限，并填补基于弱监督深度学习的道路提取方面的研究空白，本章提出一种基于弱监督深度学习的高分辨率遥感影像道路路面提取算法，其总体技术路线如图 5-1 所示。本算法不再依赖逐像素标注的路面标签，而是在类似道路中心线数据的线状弱标签（scribble-based）监督下即可实现接近人工绘制水平的道路路面提取结果，有效降低标注成本。该算法简



写为 ScRoadExtractor，包括两个重要组件：1) 基于 scribble 和遥感影像的道路标签传播；2) 用于弱监督道路分割的边缘感知双分支网络 (Dual-branch Network, DBNet)。一方面，为了丰富 scribble 提供的稀疏监督信号，提出一个道路标签传播方案实现将语义信息由 scribble 向未标注像素传播，生成分割代理标签。另一方面，设计一个包括道路分割分支和边缘检测分支的双分支网络 DBNet，利用代理标签和由 HED 边缘检测算子从影像上获得的边缘先验信息来联合训练 DBNet，学习道路路面的特征。

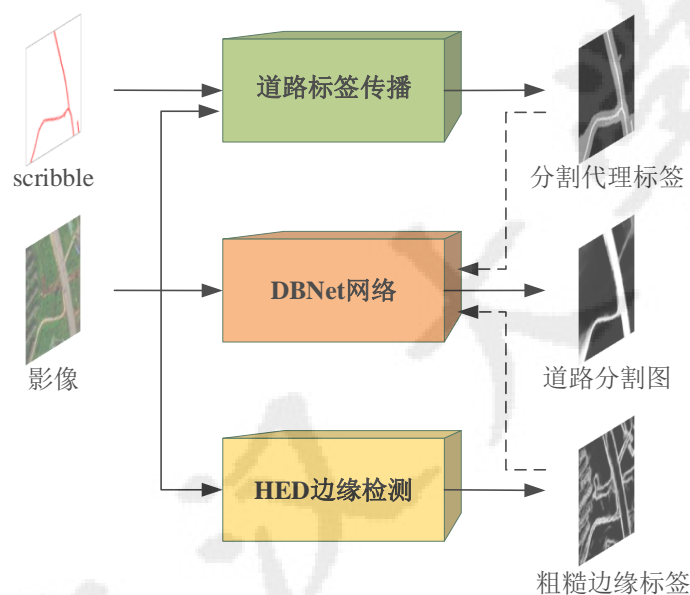


图 5-1 基于弱监督深度学习的道路路面提取算法 ScRoadExtractor

此算法的创新点可总结为以下几点：

1) 基于全卷积神经网络，提出了一种基于弱监督学习的遥感影像道路分割算法，该算法不需要路面标签，仅在类似中心线的 scribble 弱标签监督下即可实现道路路面提取，与基于全监督学习的道路提取方法对比，该算法大大降低了标注成本，具有较高的应用价值。其性能远超利用 scribble 弱监督深度学习的经典算法 ScribbleSup，也优于最近的几项相关研究工作，表现更为鲁棒。

2) 引入了基于缓冲区推理和超像素图割的道路标签传播方案，将稀疏的 scribble 提供的语义信息向未标注像素传播，生成分割代理标签。该方案通过结合缓冲区推理和超像素级别的图割，在考虑道路特性的同时还利用图模型建立超像素之间的依赖关系，对代理标签上的像素进行“道路”、“非道路”、“未知”三类别的标记，为后续分割网络的训练提供更密集的监督信号，克服了文献[50]提出的“固定道路宽度”假设的不足，能够处理道路宽度多变的复杂的影像。

3) 设计了一个用于弱监督道路分割的边缘感知双分支网络 DBNet, 在路面分割的编码器-解码器结构的基础上, 增加一个与道路分割分支平行且共享编码部分的解码分支作为边缘检测分支。边缘检测分支通过融合多尺度的上下文信息, 与道路分割分支实现联合训练, 从而辅助分割分支对路面边缘进行定位, 生成更为准确、规则的道路路面分割结果。

## 5.2 基于缓冲区推理和超像素图割的道路标签传播

### 5.2.1 道路标签传播方案

由于 scribble 可为路面分割提供稀疏的监督信息, 例如图 5-2 所展示的将矢量中心线对应的像素标记为道路, 但这样的语义信息与逐像素标注的路面标签相比仍是有限的。因此直接利用稀疏的 scribble 去训练卷积神经网络模型显然会降低道路提取的精确性。



(a) 影像 (b) 路面标签 (c) 中心线标签

图 5-2 道路中心线的 scribble 示例

以标注成本低、可从开源 OSM 上获取的道路中心线数据作为 scribble, 一个直接的方法是将 scribble 按照一定宽度进行形态学膨胀, 得到二值的“加宽版的 scribble”(简称为 expand mask) 并将其作为标签监督网络训练。这种方法最大的缺点是过早地固定标签为等宽的, 不利于全监督学习的模型准确地识别道路边缘,

尤其在处理实际情况下道路宽度多变的复杂的高分辨率遥感影像时，这种方法难以提取与地表真实道路语义对齐的道路路面结果。文献[50]中提出的“固定道路宽度”假设亦是如此，虽然有了道路、非道路、未知的初步划分，但前期等宽阈值的设定依旧限制了后续网络模型的泛化能力，而且该假设仅适用于道路中心线数据作为 scribble 的情况。

于是有了另一种思路，能否将位置不在 scribble 上但与 scribble 特征相似的像素标记为“道路”，其余像素也赋予合理的标记呢？本节提出的道路标签传播方案就是在这方面做了改进，基于缓冲区推理和超像素图割将 scribble 稀疏的语义信息向未标注像素传播，生成全像素标记的分割代理标签（proposal mask），标记也采用道路、非道路、未知三种类别，其流程如图 5-3 所示。为了便于理解，在此以道路中心线数据作为 scribble，但请注意，此方案不局限于道路中心线作为 scribble 的情况，相应的实验及分析见 5.4 小节。

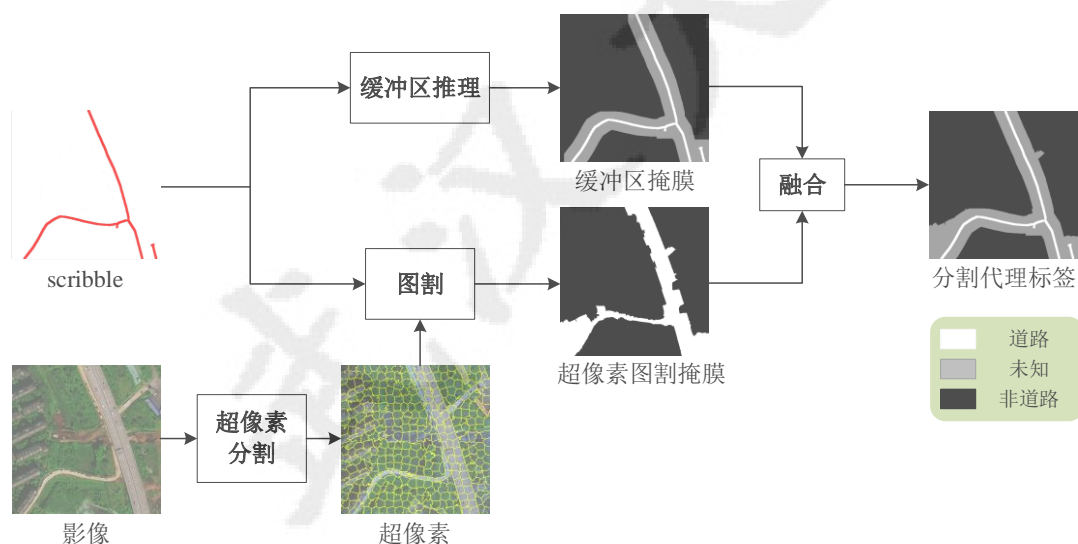


图 5-3 道路标签传播方案流程图

关于“缓冲区推理”，考虑到道路路面的边缘往往平行于道路中心线，根据每个像素与道路中心线之间的距离生成缓冲区掩膜（buffer mask）。具体来讲，创建宽度为  $a_1$  和  $a_2$  ( $a_1 < a_2$ ) 的两个缓冲区，将位于  $a_1$  的缓冲区内部的像素标记为道路，将位于  $a_2$  的缓冲区外部的像素标记为非道路，其余像素被标记为未知像素。一般情况，设置  $a_1$  小于最小道路宽度，设置  $a_2$  接近最大道路宽度。例如，对于道路宽度范围为 12 m 至 18 m 的 Cheng 数据集， $a_1=6$  m 和  $a_2=18$  m。另外，Wuhan 数据集和 DeepGlobe 数据集的  $a_1=2$  m， $a_2$  分别设置为 38 m 和 9 m。具体消融实验见 5.2.2 小节。

关于“超像素图割”，在遥感影像的超像素上构建图（Graph）模型，并通过最小化能量函数，对每个超像素进行前景或背景的划分，生成超像素图割掩膜（graph mask）。该策略受启发于图割（Graph Cut）<sup>[62][63]</sup>。首先应用简单线性迭代聚类算法（Simple Linear Iterative Clustering, SLIC）<sup>[64]</sup>对遥感影像进行超像素分割。例如，在 512×512 影像块上，SLIC 超像素的数量约为 400，将控制色彩和空间之间平衡的 compactness 参数设置为 20。将影像从 Red, Green, Blue (RGB) 色彩空间转换至 Hue, Saturation, Value (HSV) 色彩空间，利用色调 H 和饱和度 S 为所有超像素计算二维的直方图（简称为 Hist），其中 H 和 S 的范围分别为 [0, 360] 和 [0, 1]，bin 设置为 20 个，计算时所有的 bin 均经过串联并归一化。如图 5-3 所示，scribble 仅代表前景（道路）。为了满足后续图割过程对前景和背景的需要，在原 scribble 的基础上，随机绘制若干个线状弱标签代表背景（非道路），它们只能在  $a_2$  的缓冲区以外，换言之，前景和背景(line状弱标签)不会存在交集。由此得到集合  $Sc = \{s_r, c_r\}$ ，其中  $r$  是  $Sc$  中任意的一个线状弱标签， $s_r$  指的是与  $r$  对应的像素， $c_r$  指的是  $r$  的类别（前景或背景）。另外，根据超像素与  $Sc$  的重叠情况，利用 Hist 计算得到  $c_r$  的累计直方图  $Hist_r$ 。

$$E(x) = \sum_i \psi_i(x_i | Hist, Sc) + \sum_{i,j} \psi_{ij}(x_i, x_j | Hist) \quad (5.1)$$

$$\psi_i(x_i) = \begin{cases} 0, & \text{if } x_i \cap s_r \neq \emptyset \text{ and } c_i = c_r \\ \text{KLDiv}(Hist_i, Hist_r), & \text{if } x_i \cap Sc = \emptyset \\ \infty, & \text{otherwise} \end{cases} \quad (5.2)$$

$$\psi_{ij}(x_i, x_j) = \text{KLDiv}(Hist_i, Hist_j) \quad (5.3)$$

然后，以超像素为节点创建 Graph，构建如公式（5.1）所示的能量函数，利用一元项  $\psi_i$  将每个节点  $x_i$  与前景、背景的关系进行建模，利用二元项  $\psi_{ij}$  将相邻节点  $x_i$  和  $x_j$  ( $x_i \neq x_j$ ) 之间的关系进行建模。相邻节点通过构建三角网（Delaunay tessellation）实现。其中，一元项可展开如公式（5.2）所示。如果节点  $x_i$  与  $s_r$  存在重叠，那么，将  $x_i$  划分为  $c_r$  的损失为 0。如果节点  $x_i$  与  $s_r$  不存在重叠，那么，计算  $x_i$  的归一化直方图  $Hist_i$  与  $c_r$  的累计直方图  $Hist_r$  之间的 KL 散度（Kullback-Leibler Divergence，简称为 KLDiv）作为将  $x_i$  划分为  $c_r$  的损失。其含义是，KL 散度越大说明该节点与  $c_r$  差异越大，那么将  $x_i$  划分为  $c_r$  的损失越大。公式（5.3）为二元项的具体计算，通过两个相邻节点的归一化直方图的 KL 散度来评估二者

的相似性。对于越相似的两个相邻节点，在它们之间切割边的成本较高，也就是说，相似性较高的两个相邻节点更可能具有相同标签。根据 Graph Cut 中最小割最大流原理优化能量函数，为每个节点分配前景或背景标签，从而生成二值的超像素图割掩膜。

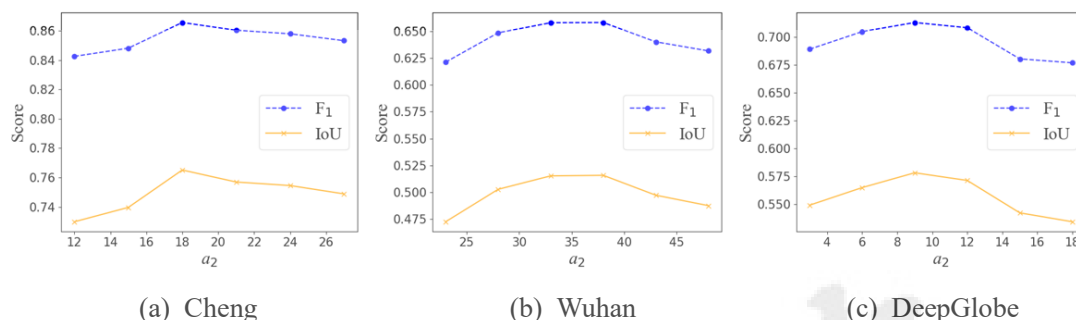
最后按照以下规则融合 buffer mask 和 graph mask: 在 buffer mask 的基础上，如果某像素在 graph mask 中是“道路”而在 buffer mask 中是“非道路”，则将该像素标记为“未知”，其余像素保持与 buffer mask 的标记一致，由此生成 proposal mask。这样，该道路标签传播方案不仅考虑了道路的条带状特点，而且还通过构建图模型获得影像的超像素之间在色彩与空间上的联系。

显然，这一道路标签传播方案弥补了缓冲区推理和超像素图割各自的不足。一方面，仅使用缓冲区推理会依赖 scribble 的准确性，例如过时的 GIS 地图或 OSM 数据中存在不正确或不完整的中心线数据，使得 buffer mask 的“非道路”像素中可能存在大量理应被划分为“道路”的像素；但是通过融合，proposal mask 将这些有争议的像素标记为未知，而未知像素将通过 5.3.1 节中描述的弱监督损失函数参与网络模型训练使其进一步被划分为潜在道路或非道路。另一方面，由于每个超像素的类别是唯一的，生成的 graph mask 并不精细，特别是在道路路面边缘处，而且标准损失函数（如二元交叉熵函数）很难区分 graph mask 中的错误标注，所以仅使用超像素图割会误导后续网络模型的训练；但是通过融合，proposal mask 中只有少量的绝对的前景和背景，而易出错的道路路面边缘则多属于未知像素，由网络模型学习得到。

### 5.2.2 与其他方法的性能比较

首先，我们评估了不同缓冲区宽度对本节介绍的道路标签传播方案的影响。参数  $a_1$  定义了道路像素的范围，可以根据数据集的最小道路宽度很好地估算。对于 Cheng 数据集，设置  $a_1$  为 6 m；由于 Wuhan 数据集和 DeepGlobe 数据集覆盖了道路狭窄的郊区、农村地区，故将  $a_1$  设置为 2 m。在此，关注能够权衡缓冲区推理与超像素图割的参数  $a_2$  的影响。在各数据集上使用不同  $a_2$  值的结果如图 5-4 所示，可以观察到，当  $a_2$  被设置为接近最大道路宽度（例如，对于 Cheng 数据集  $a_2=18$  m）时道路路面提取将获得最佳性能，当  $a_2$  逐渐增加或减少时，性能会稍变差。如果  $a_2$  值太大，则代理标签将更接近于由缓冲区推理获得的缓冲区

掩膜；如果  $a_2$  值太小，则代理标签将更接近于超像素图割掩膜。最后，将 Cheng 数据集、Wuhan 数据集和 DeepGlobe 数据集的  $a_2$  分别设置为 18 m、38 m 和 9 m。



(a) Cheng (b) Wuhan (c) DeepGlobe

图 5-4 道路标签传播方案中参数  $a_2$  不同取值的性能对比

按照 5.2.1 中描述的道路标签传播方案流程以及参数设置，本小节分别在 Cheng、Wuhan、DeepGlobe 这三个道路数据集上对 proposal mask 的性能进行了评估，并与 5.2.1 提到的 expand mask、buffer mask、graph mask 相比较。其中，expand mask 是将中心线膨胀了等宽（例如，10 米），buffer mask 和 graph mask 分别是道路标签传播方案中的中间结果——缓冲区掩膜和超像素图割掩膜。此实验均在 DBNet 上训练的。对比结果见表 5-1，以加粗字体表示最佳，以下划线字体表示次优。可以看出，在三个数据集上利用 proposal mask 训练 DBNet 能达到最佳性能，且优于利用其它形式标签训练的效果，验证了所提出的道路标签传播方案的有效性。

表 5-1 道路标签传播方案与其他方法比较

标签	Cheng		Wuhan		DeepGlobe	
	F1	IoU	F1	IoU	F1	IoU
graph mask	0.6555	0.4972	0.6165	0.4689	0.3980	0.2614
expand mask	<u>0.8280</u>	<u>0.7089</u>	0.5467	0.3912	0.6203	0.4678
buffer mask	0.8033	0.6769	<u>0.6325</u>	<u>0.4848</u>	<u>0.6666</u>	<u>0.5224</u>
proposal mask	<b>0.8482</b>	<b>0.7396</b>	<b>0.6484</b>	<b>0.5028</b>	<b>0.6805</b>	<b>0.5422</b>

### 5.3 用于弱监督道路分割的边缘感知双分支网络

#### 5.3.1 DBNet

设计一个包括道路分割分支和边缘检测分支的双分支网络 DBNet，其模型架构如图 5-5 所示，以影像块（如 512×512 像素）的 RGB 通道作为输入，网络模型在 FCN 经典的编码器-解码器结构上改进。

DBNet 使用在 ImageNet 上预训练好的残差网络 ResNet-34 作为特征编码网

络，由五组下采样组成，最小缩放比例为输入尺寸的 1/32，具体包括一组步长为 2 的 7×7 卷积层，以及四组最大池化和残差块，其中残差块由两个连续的 3×3 卷积和跳跃连接构成。从第二组最大池化和残差块开始，特征图的通道数在每组下采样后都会加倍。将编码部分生成的特征图输入 ASPP 模块，该模块由 (a) 一个 1×1 卷积、三个并行的空洞率分别为 1、2 和 4 的 3×3 空洞卷积和 (b) 图像全局平均池化组成。将 ASPP 模块生成的特征图进行级联并通过通道数为 512 的 1×1 卷积层，传递给解码部分。

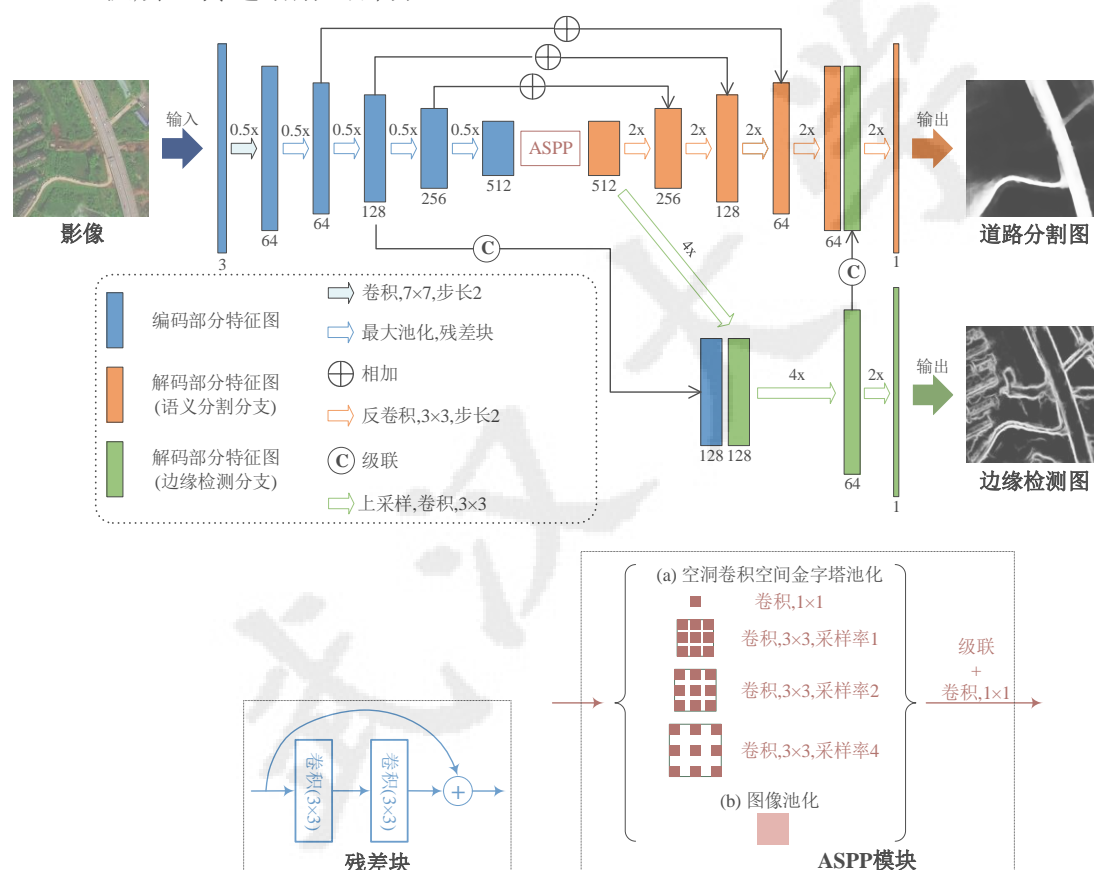


图 5-5 DBNet 网络结构图

关于解码部分，设计了两个平行的子分支：道路分割分支和边缘检测分支。其中，道路分割分支利用 5 组步长为 2 的 3×3 反卷积层<sup>[19]</sup>，将特征图的分辨率由 16×16 上采样到 512×512（与输入尺寸一致），除了最后两组之外特征图的通道数在每组上采样中均减半。编码部分和道路分割分支的特征图之间有三个跳跃连接的相加（用 ⊕ 表示）操作，分别作用在尺寸为 32×32、64×64、128×128 的特征图上。边缘检测分支通过合并多尺度上下文信息来恢复边缘，首先对道路分割分支的第一个特征图进行 4 倍双线性上采样，并由通道数为 128 的 3×3 卷积层进行处理，然后与来自编码部分的具有相同空间分辨率的低层特征进行级联（用

©表示), 再进行另一个 4 倍双线性上采样, 经过通道数为 64 的  $3 \times 3$  卷积层之后, 与道路分割分支中的相应特征图 ( $256 \times 256$ ) 进行级联, 最后对特征图进行 2 倍双线性上采样。两个分支在最后一层卷积使用 Sigmoid 激活分别输出像素属于道路和边缘的概率, 得到道路分割图和边缘检测图。DBNet 中其他卷积层均通过线性整流函数 (Rectified Linear Unit, ReLU) 激活。

DBNet 充分发挥了编码-解码结构、跳跃连接和双分支的优势, 其边缘分支检测到的边缘作为先验知识来进一步提升和指导分割分支。接下来介绍 DBNet 中双分支联合训练学习道路路面特征的过程。利用 5.2.1 小节描述的道路标签传播方案生成的代理标签 (记为  $M^{\text{seg}}$ ) 监督道路分割分支; 利用在 BSDS500 数据集<sup>[65]</sup>上预训练好的 HED 边缘检测算子<sup>[66]</sup>, 对训练集的影像进行预测生成对应的粗糙边缘标签 (记为  $M^{\text{HED}}$ ), 用以监督边缘检测分支, 进而辅助分割分支提高定位道路路面边缘的能力。注意, 粗糙边缘标签的生成无需任何微调, 仅利用 HED 边缘检测算子的预测过程, 因此并不需要准备关于道路边缘的标签。这与本章主题相符, 研究仅利用类似中心线的 scribble 线状弱标签即可实现高分辨率遥感影像的道路路面提取。由于 BSDS500 数据集多是自然影像, 所以生成的粗糙边缘标签中一般包含有遥感影像中各种显著目标的边缘信息, 旨在为分割分支提供具有丰富结构信息的粗糙边缘约束。

根据公式 (5.4) - (5.7) 计算模型输出的道路分割图  $O^{\text{seg}}$  和代理标签  $M^{\text{seg}}$  之间的“弱监督分割损失函数”  $L_{w\_seg}$ , 包括“部分二值交叉熵损失函数” (Partial Binary Cross Entropy Loss, 简称为  $L_{\text{PBCE}}$ ) 和“正则化损失函数” (Regularized Loss, 简称为  $L_R$ )。考虑到代理标记中有三类标记, “未知”像素中存在潜在的前景和 (或) 背景, 直接用类似二值交叉熵损失这样的标准损失函数显然不合适, 所以  $L_{\text{PBCE}}$  仅在代理标签中被标记为道路和非道路的“已知 (known)”像素  $p \in \Omega_k$  上计算模型输出的道路分割图  $O^{\text{seg}}$  和代理标签  $M^{\text{seg}}$  的二值交叉熵损失。正则化损失是一种弱监督损失函数, 借鉴了文献[42]和快速双边滤波<sup>[67]</sup>, 利用在 RGBXY 通道上具有密集高斯核  $W$  的条件随机场 CRF 损失来实现。参数  $\alpha$  用于权衡  $L_{\text{PBCE}}$  和  $L_R$ , 设置为 0.5。根据公式 (5.8) 计算模型输出的边缘检测图  $O^{\text{edge}}$  和粗糙边缘标签  $M^{\text{HED}}$  之间的边缘损失函数  $L_{\text{edge}}$ , 利用二者的均方误差损失 (Mean Squared Error, MSE) 实现。最后由公式 (5.9) 得到用于协同训练双分支的总损失函数  $L$ , 参数  $\beta$  用于权衡  $L_{w\_seg}$  和  $L_{\text{edge}}$ , 设置为 0.7。



$$L_{w\_seg} = L_{PBCE}(M_p^{seg}, O_p^{seg}) + \alpha L_R(O^{seg}) \quad (5.4)$$

$$L_{PBCE}(M_p^{seg}, O_p^{seg}) = - \sum_{p \in \Omega_k} (M_p^{seg} \log O_p^{seg} + (1 - M_p^{seg}) \log(1 - O_p^{seg})) \quad (5.5)$$

$$L_R(O^{seg}) = O^{seg'} W (1 - O^{seg}) \quad (5.6)$$

$$\frac{\partial L_R(O^{seg})}{\partial O^{seg}} = -2WO^{seg} \quad (5.7)$$

$$L_{edge} = \frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h (M_{ij}^{HED} - O_{ij}^{edge})^2 \quad (5.8)$$

$$L = L_{w\_seg} + \beta L_{edge} \quad (5.9)$$

### 5.3.2 DBNet 与其他双分支网络的性能比较

按照 5.3.1 中描述的 DBNet 网络结构以及参数设置，本小节分别在 Cheng、Wuhan、DeepGlobe 这三个道路数据集上对 DBNet 边缘检测分支的有效性进行了评估，并与 BPG<sup>[45]</sup>提出的边界回归网络（Boundary Regression Network, BRN）进行性能比较。

表 5-2 DBNet 中边缘检测分支的有效性验证并与其他网络比较

Network	Loss			Cheng		Wuhan		DeepGlobe	
	$L_{PBCE}$	$L_R$	$L_{edge}$	F <sub>1</sub>	IoU	F <sub>1</sub>	IoU	F <sub>1</sub>	IoU
SegB	√			<u>0.7987</u>	<u>0.6729</u>	0.6404	0.4944	0.6502	0.5070
SegB + BRN	√		√	0.7634	0.6233	<u>0.6422</u>	<u>0.4948</u>	<u>0.6712</u>	<u>0.5269</u>
SegB + EdgeB	√	√	√	<b>0.8482</b>	<b>0.7396</b>	<b>0.6484</b>	<b>0.5028</b>	<b>0.6805</b>	<b>0.5422</b>

将 DBNet 中道路分割分支简记为 SegB，边缘检测分支简记为 EdgeB。在此探讨了 EdgeB 的有效性和创新性，表 5-2 展示了在 ResNet-34 编码部分基础上使用不同分支作为解码部分的比较结果，以加粗字体表示最佳，以下划线字体表示次优。SegB 均使用相同的代理标签进行训练，文献[45]中提出的 BRN 和本节提出的 EdgeB 均使用相同的粗糙边缘标签进行训练。对于 Cheng 数据集，仅使用 SegB 在 F<sub>1</sub> 达到 79.87%，加入 BRN 后的双分支结构使降低了 3.53%，而加入 EdgeB 后的双分支结构（即 DBNet）使 F<sub>1</sub> 提高了 4.41%。对于 Wuhan 数据集，DBNet（即 SegB + EdgeB）得到了最优结果，但与其他两个结构的表现相似。对于 DeepGlobe 数据集，DBNet 超过了仅使用 SegB 道路分割分支 3.52%IoU，与

引入了 BRN 的双分支结构相比在 IoU 上提高了 1.53%。由此证明了本节所设计的 DBNet 具有良好的泛化能力和有效性。

文献[45]提出的 BRN 与本节设计的边缘检测分支存在两个主要差异。首先, BRN 中仅使用了一次上采样层, 而 EdgeB 中 (见图 5-5) 使用了多个上采样层以及与道路分割分支的特征共享。其次, BRN 缺乏正则化来捕获已知 (道路和非道路) 和未知像素之间的依赖关系。这些因素导致引入 BRN 的双分支结构表现比本节所设计的 DBNet 差, 特别是在 Cheng 数据集甚至比仅用分割分支的主干结构都要差。

关于粗糙边缘标签的生成, 本文采用在数据集 BSDS500 上预训练的 HED 作为边缘检测器, 另外也测试了其他边缘检测器, 例如经典的 Canny<sup>[68]</sup>和 Sobel<sup>[69]</sup>。在 Wuhan 数据集上, 使用 Canny 边缘检测的 IoU 为 0.493, 而使用 HED 的 IoU 为 0.503。需要说明的是, 实验发现使用这些边缘检测器的效果也令人满意, 仅比使用 HED 稍差, 因此它们也可以替代 HED。

## 5.4 实验分析与讨论

本节将 ScRoadExtractor 方法与 ScribbleSup<sup>[38]</sup>、BPG<sup>[45]</sup>、WSOD<sup>[46]</sup>和 WeaklyOSM<sup>[50]</sup>进行比较, 评估不同弱监督深度学习方法的道路路面提取性能。

表 5-3 ScribbleSup 不同迭代轮数的性能比较

数据集	迭代轮数	F1	Precision	Recall	IoU
Cheng	1	<u>0.6555</u>	<u>0.5059</u>	0.9671	<u>0.4972</u>
	2	0.5453	0.3863	0.9711	0.3823
	3	<b>0.6750</b>	<b>0.5274</b>	<u>0.9730</u>	<b>0.5190</b>
	4	0.5442	0.3872	<b>0.9732</b>	0.3832
Wuhan	1	<u>0.6165</u>	<u>0.6001</u>	0.7316	<u>0.4689</u>
	2	0.5727	0.5210	<b>0.7411</b>	0.4238
	3	<b>0.6222</b>	<b>0.6086</b>	0.7267	<b>0.4740</b>
	4	0.5923	0.5522	<u>0.7344</u>	0.4423
DeepGlobe	1	<u>0.3980</u>	<u>0.2836</u>	0.8904	<u>0.2614</u>
	2	0.2530	0.1571	<u>0.8953</u>	0.1530
	3	<b>0.4079</b>	<b>0.2951</b>	0.8813	<b>0.2694</b>
	4	0.2584	0.1602	<b>0.8965</b>	0.1695

由于经典的 ScribbleSup 方法中, 代理标签的生成和神经网络的训练是经过几轮迭代交替优化的, 考虑到其计算成本较大, 在此首先进行了 ScribbleSup 方

法的四轮迭代实验，结果见表 5-3，以加粗字体表示最佳，以下划线字体表示次优。可以看出，随着迭代轮数的增加，该方法的性能有轻微提升，但其性能并不稳定。根据  $F_1$  和 IoU 指标，第 3 轮迭代的结果是前 4 轮中表现最好的，将其作为 ScribbleSup 的结果与其他方法比较。

BPG (Boundary Perception Guidance) 方法也采用双分支网络结构，设计了边界回归网络 (Boundary Regression Network, BRN) 用以辅助语义分割任务。BRN 与本文提出的 DBNet 中的边缘检测分支的消融实验可见 5.3.2 小节。基于 scribble 的 WSOD (Weakly-supervised Salient Object Detection) 方法引入了只关注图像显著区域的门控的结构感知损失。WeaklyOSM 方法是利用 OSM 中心线数据进行弱监督道路分割的深度学习方法。

表 5-4 不同弱监督深度学习方法的性能比较

数据集	方法	$F_1$	Precision	Recall	IoU
Cheng	ScribbleSup <sup>[38]</sup>	0.6750	0.5274	<b>0.9730</b>	0.5190
	BPG <sup>[45]</sup>	0.7925	0.7179	<u>0.9085</u>	0.6627
	WSOD <sup>[46]</sup>	<u>0.8608</u>	<u>0.8468</u>	0.8852	<u>0.7594</u>
	WeaklyOSM <sup>[50]</sup>	0.8322	0.7798	0.9077	0.7170
	ScRoadExtractor (Ours)	<b>0.8657</b>	<b>0.9033</b>	0.8423	<b>0.7651</b>
Wuhan	ScribbleSup <sup>[38]</sup>	0.6222	0.6086	<b>0.7267</b>	0.4740
	BPG <sup>[45]</sup>	0.6197	<b>0.7510</b>	0.5804	0.4717
	WSOD <sup>[46]</sup>	0.5789	0.7143	0.5329	0.4298
	WeaklyOSM <sup>[50]</sup>	<u>0.6300</u>	<u>0.7509</u>	0.6020	<u>0.4805</u>
	ScRoadExtractor (Ours)	<b>0.6580</b>	0.6963	<u>0.6904</u>	<b>0.5158</b>
DeepGlobe	ScribbleSup <sup>[38]</sup>	0.4079	0.2951	<b>0.8813</b>	0.2694
	BPG <sup>[45]</sup>	0.6624	0.6681	<u>0.7638</u>	0.5157
	WSOD <sup>[46]</sup>	0.5899	0.6549	0.6265	0.4438
	WeaklyOSM <sup>[50]</sup>	<u>0.6673</u>	<u>0.7115</u>	0.7378	<u>0.5239</u>
	ScRoadExtractor (Ours)	<b>0.7132</b>	<b>0.7954</b>	0.7138	<b>0.5782</b>

与不同弱监督深度学习方法比较结果见表 5-4，以加粗字体表示最佳，以下划线字体表示次优。可以看出，ScribbleSup 在这些道路数据集上的表现欠佳。虽然采用了三轮交替训练，但相比之下，其性能远不及 BPG、WSOD、WeaklyOSM 和 ScRoadExtractor 这些仅使用一轮训练的端到端的方法。在 Cheng 数据集上，本章提出的 ScRoadExtractor 超过了 BPG 方法 7.32% 的  $F_1$  以及 10.24% 的 IoU。

同样，在 Wuhan 数据集和 DeepGlobe 数据集上，ScRoadExtractor 比 BPG 分别高出 4.41% 和 6.25% 的 IoU。BPG 的不足是，其分割分支和边界分支仅在损失函数级别上相互作用，而未考虑两个子分支的特征之间的相关性。在最小的 Cheng 数据集上，ScRoadExtractor 的性能比 WSOD 稍好，而在 Wuhan 数据集和 DeepGlobe 数据集上，ScRoadExtractor 的性能明显好于 WSOD，这表明 WSOD 仅能处理相对简单的场景，缺乏一定的泛化能力。分析 WSOD 的不足是，其过于信任显著预测结果，算法鲁棒性差。以 WeaklyOSM (次优) 的结果为基准，ScRoadExtractor 在 Cheng 数据集，Wuhan 数据集和 DeepGlobe 数据集上分别实现了 IoU 的 4.81%，3.53% 和 5.43% 增长。

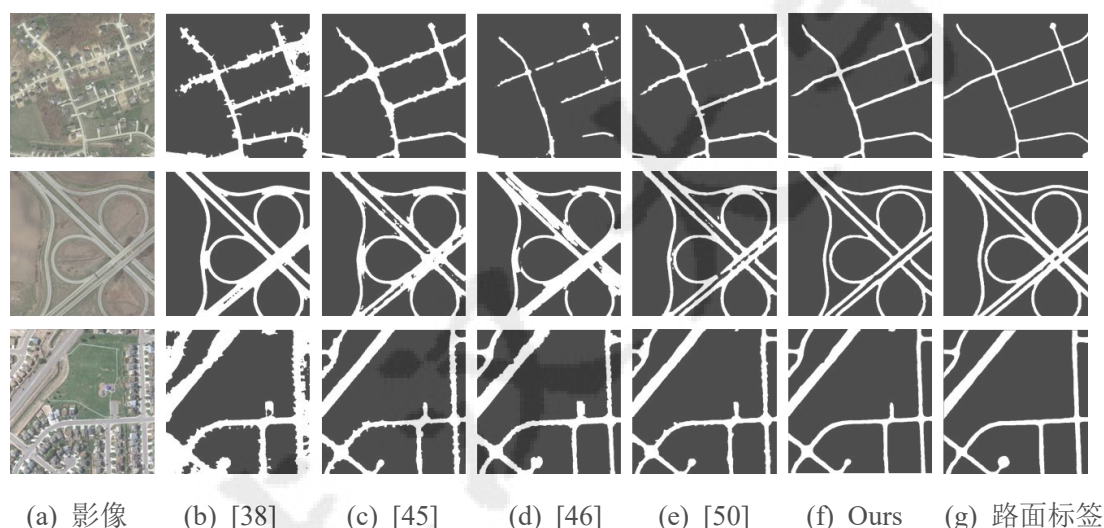


图 5-6 Cheng 数据集 ScRoadExtractor 和其他方法的结果对比示例

图 5-6、5-7、5-8 显示了不同方法在不同道路数据集上预测的道路分割结果示例。可以看出，ScribbleSup (b) 的结果包含许多路面边缘定位较差的非道路像素，这是由于仅根据超像素图割掩膜并不会达到像素级别的精细程度，特别是在道路路面边缘处，而且标准损失函数无法区分二值的超像素图割掩膜中的错误标注。如图 5-7 和图 5-8 的第一行所示，BPG (c)，WSOD (d) 和 WeaklyOSM (e) 难以从卫星影像中正确识别出被建筑物或树木遮挡的道路，从而导致提取结果中存在路段缺损、路面边缘含糊等情况。相比之下，ScRoadExtractor (f) 对遮挡情况具有一定的鲁棒性。除此之外，ScRoadExtractor (f) 生成的道路分割结果与逐像素人工标注的路面标签 (g) 最接近，具有良好的路面边缘对齐，表明 ScRoadExtractor 可从遥感影像中提取更可靠的路面数据。

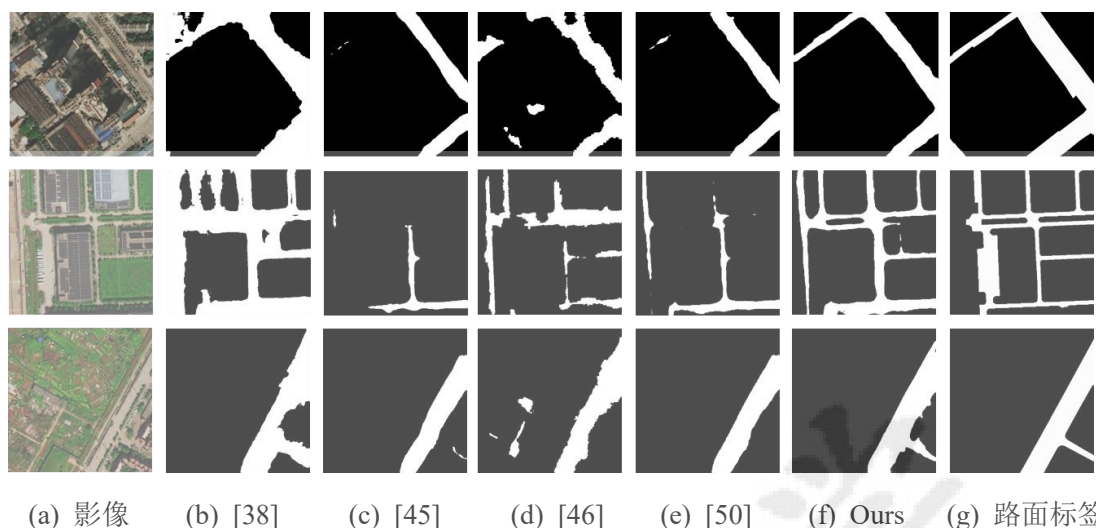


图 5-7 Wuhan 数据集 ScRoadExtractor 和其他方法的结果对比示例

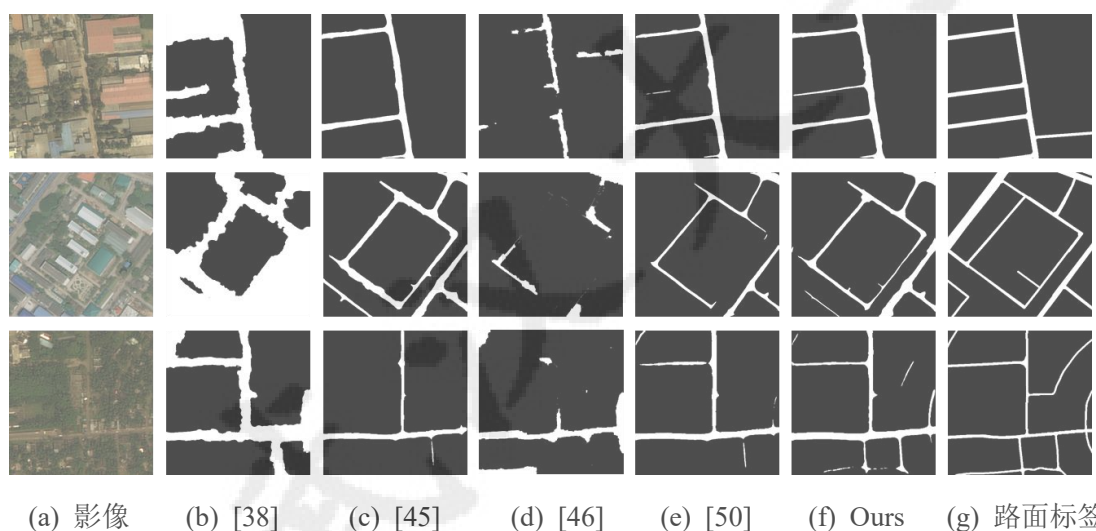


图 5-8 DeepGlobe 数据集 ScRoadExtractor 和其他方法的结果对比示例

表 5-5 Wuhan 数据集 ScRoadExtractor 在模拟 scribble 上的性能比较

方法	F <sub>1</sub>	Precision	Recall	IoU
WeaklyOSM	0.5759	0.8661	0.4687	0.4307
ScRoadExtractor	0.6403	0.8318	0.5680	0.5001

文献[50]中提出的 WeaklyOSM 具有与本章研究相似的问题表述，但是 WeaklyOSM 与 ScRoadExtractor 有以下几个方面的差异。首先，ScRoadExtractor 采用了边缘检测分支来辅助和增强弱监督道路分割的性能，该边缘检测分支的效果也可以从图 5-6、5-7、5-8 中看出，通过从连续的边缘结构信息中学习，ScRoadExtractor 能够得到更好的道路连通性，相应的消融实验见 5.3.2 小节。其次，WeaklyOSM 仅基于缓冲区的固定道路宽度假设来生成代理标签，使得代理

标签的质量依赖于固定的 scribble 形式(即道路中心线),限制了此算法的泛化能力。而 ScRoadExtractor 并不存在这种限制,因为道路标签传播方案也能够应用于道路中心线以外的其他一般形式的 scribble。为了进一步验证,在 Wuhan 数据集上引入了“模拟 scribble”,模拟的是偏移中心线位置的一般形式的 scribble。这些 scribble 的生成方式是:使用大小为 7 的十字形核和偏移锚点(3, 6)去形态学腐蚀路面标签,随后进行形态学细化得到的。表 5-5 展示的是两个方法在 Wuhan 数据集的模拟 scribble 监督下的性能对比,本章提出的方法在  $F_1$  和 IoU 上分别超出 WeaklyOSM 方法 6.44%和 6.94%。由此可见,ScRoadExtractor 能够很好地应对从道路中心线到模拟 scribble 的各种形式的线状弱标签,而不会像 WeaklyOSM 受早期硬约束的限制。

## 5.5 本章小结

本章对基于弱监督深度学习的道路路面提取算法进行研究,详细讨论了基于缓冲区推理和超像素图割的道路标签传播方案的方法与过程,以及用于弱监督道路分割的边缘感知双分支网络的模型结构和训练优化方法。并在不同的道路数据集、不同基于 scribble 的弱监督语义分割方法上进行对比试验,探究所提出的 ScRoadExtractor 算法的泛化能力。

## 第 6 章 总结与展望

### 6.1 全文总结

随着科技发展进入大数据与人工智能新时代,出现了愈来愈多的不同学科和领域的交叉融合。深度学习应用于计算机视觉领域的思想逐渐渗透到遥感领域,为遥感影像道路提取提供了新的思路。近年来,基于深度学习理论与技术的道路提取方法层出不穷,尽管这些方法取得了令人鼓舞的进展,但遥感影像远比自然影像复杂得多,基于深度学习的遥感影像道路提取研究尚未完善,仍是富有挑战性的课题。本文针对现有算法存在的道路提取精度和自动化程度有限等问题,探究了基于全监督深度学习的道路路面和中心线同时提取算法和基于弱监督深度学习的道路路面提取算法,主要研究工作包括以下几个方面:

1) 针对现有算法存在的道路提取精度有限的问题(特别是在道路连通性和完整性方面),本文创新地提出一个结合语义分割和拓扑追踪的多级框架 **Seg-Tracer**,与现有方法中分别处理的方式不同,该框架以基于多个不同 CNN 的逐级优化方式从遥感影像中同时提取道路路面和中心线,有效提高道路提取的拓扑连通性和完整性。首先,在 FCN 初始分割的基础上,引入集成学习中的 **Boosting** 提升策略,并设计了另一个轻量的 FCN 网络作为提升分割网络从先前分割的失误案例中逐步学习,以增强道路分割的连通性。其次,提出一种改进的道路拓扑追踪方法,利用从道路分割中自动检测的多个特征起始点,基于 CNN 迭代式追踪道路中心线,以弥补单起点追踪中搜索过早终止而使道路完整性差的劣势。最后,采用融合策略充分发挥两类方法各自的优势,以进一步提高道路提取的精度。在三个不同道路数据集上进行了大量实验表明,本文提出的方法在道路提取精度,尤其是在道路连通性和完整性方面,明显优于当前先进的道路分割和中心线提取方法。

2) 针对现有算法存在的道路提取自动化程度有限的问题(全监督学习依赖于大量人工标注数据),本文创新性提出一个基于弱监督深度学习的道路路面提取算法 **ScRoadExtractor**,与现有的路面提取方法需要密集标注的路面标签不同,此算法仅在易获取的稀疏 **scribble** 弱标签监督下即可实现道路路面自动提取。首先引入了基于缓冲区推理和超像素图割的道路标签传播方案,以将类似中心线的

scribble 提供的语义信息向未标注像素传播。其次,设计了一个包括道路分割分支和边缘检测分支的双分支网络,利用道路标签传播方案生成的代理标签和由 HED 边缘检测算子获得的边缘先验信息来联合训练该网络,学习道路路面的特征。在三个道路数据集上进行了大量实验发现,本文提出的方法在道路宽度多变的复杂场景下表现更为鲁棒,性能远超基于 scribble 弱监督学习的经典算法,也优于近期提出的几项相关研究工作。此外,与基于全监督学习的道路路面提取方法相比,此算法明显降低了标注成本,提高了道路提取的自动化水平。

因此,本文提出的创新性方法具有重要的理论研究意义与实际应用价值。

## 6.2 研究展望

高分辨率遥感影像道路提取是一个较为复杂的课题,本文针对道路语义分割、道路拓扑追踪、面向路面提取的弱监督学习进行了理论研究和切实可行的实验验证,但仍存在一定的不足,将在未来研究中继续跟踪国内外计算机视觉与遥感领域的前沿研究成果,探索如何基于深度学习进一步提高道路提取的精度和自动化水平。目前有以下几个方面的新思路:

### 1) 拓扑感知的道路边缘检测算法

本文提出的结合语义分割和拓扑追踪两类方法的道路提取算法,相比仅使用其中任一类方法的算法,有效提升了道路提取的拓扑连通性和完整性,但是由于采用了多阶段的框架模式,可能导致前期阶段的误差积累,影响道路提取的最终精度。因此在未来的研究中,将对此方向进行深入的研究,拟尝试以端到端的方式融合道路的语义和拓扑信息,直接提取拓扑感知的道路边缘。拟通过在卷积神经网络(CNN)、图卷积神经网络(Graph neural network, GNN)上改进,实现从遥感影像中获取边缘定位精确的道路数据,这将是与地面真实道路轮廓准确匹配的关键一步,有望为自动驾驶、智慧交通等方面的研究提供有力的数据和技术支持。

### 2) 基于半监督深度学习的道路提取算法

本文基于 scribble 的弱监督深度学习实现了高分辨率遥感影像道路路面提取,有效降低了人工标注成本,且领先于其他弱监督学习方法。尽管相比全监督学习,此算法提高了道路提取的自动化程度,但其性能仍与全监督学习有一定差距。因此在未来的研究中,将对此方向进行深入的研究,拟发展用于道路提取的半监督



深度学习算法,即同时使用少量含标签的影像和大量无标签的影像训练神经网络。拟将深度学习与传统方法的结合作为突破口,在避免需要大量手动注释工作的前提下,挖掘无标签影像数据的潜在监督信号,这将是遥感影像道路智能提取研究中的重要一步,有望为可持续发展、地理信息资源建设提供新的解决方案。

武汉大学

## 参考文献

- [1] 李德仁. 展望大数据时代的地球空间信息学[J]. 测绘学报, 45(4): 379-384, 2016.
- [2] 周成虎. 全空间地理信息系统展望[J]. 地理科学进展, 34(2): 129-131, 2015.
- [3] 龚健雅. 人工智能时代测绘遥感技术的发展机遇与挑战[J]. 武汉大学学报(信息科学版), 43(12): 35-43, 2018.
- [4] Mnih V. and Hinton G. E. Learning to detect roads in high-resolution aerial images[C]. In Proceedings of the 11th European Conference on Computer Vision (ECCV), 210-223, 2010.
- [5] 龚健雅, 季顺平. 摄影测量与深度学习[J]. 测绘学报, 47(6): 693-704, 2018.
- [6] Goodfellow I., Bengio Y., and Courville A. Deep Learning: Adaptive computation and machine learning series[M]. Massachusetts, USA:: MIT press, 2017.
- [7] Krizhevsky A., Sutskever I., and Hinton G. E. ImageNet classification with deep convolutional neural networks[C]. In Proceedings of the 26th Annual Conference on Neural Information Processing Systems, 1106-1114, 2012.
- [8] Ren, S., He, K., Girshick, R., and Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149, 2016.
- [9] 林祥国, 张继贤, 李海涛, 杨景辉. 基于 T 型模板匹配半自动提取高分辨率遥感影像带状道路[J]. 武汉大学学报:信息科学版, 34(3): 293-296, 2009.
- [10] Leninisha S. and Vani K. Water flow based geometric active deformable model for road network[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 102: 140-147, 2015.
- [11] Huang X. and Zhang L. Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines[J]. International Journal of Remote Sensing, 30(8): 1977-1987, 2009.
- [12] Yi W., Chen Y., Tang H., and Deng L. Experimental research on urban road extraction from high-resolution RS images using probabilistic topic models[C]. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium, 445-448, 2010.
- [13] 钱家航, 王金亮, 马如彪. 基于 quickbird 影像城市道路特征语义信息提取[J]. 遥感技术与应用, 29(4):653-659, 2014.
- [14] Sun Z., Fang H., Deng M., Chen A., Yue P., and Di L., Regular shape similarity index: A novel index for accurate extraction of regular objects from remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 53(7): 3737-3748, 2015.
- [15] Simonyan K. and Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. In Proceedings of the 3rd International Conference on Learning Representations, 1-14, 2015.
- [16] Szegedy C., Liu W., Jia Y., et.al. Going deeper with convolutions[C]. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 1-9, 2015.
- [17] He K., Zhang X., Ren S., and Sun J. Deep residual learning for image recognition[C]. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, 2016.
- [18] Long J., Shelhamer E., and Darrell T. Fully convolutional networks for semantic segmentation[C]. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 3431-3440, 2015.

- [19] Noh H., Hong S., and Han B. Learning deconvolution network for semantic segmentation[C]. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 1520-1528, 2015.
- [20] Ronneberger O., Fischer P., and Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]. In Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention, 234-241, 2015.
- [21] Chen, L.-C., Zhu Y., Papandreou G., Schroff F., and Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. In Proceedings of the 15th European Conference on Computer Vision (ECCV), 801-818, 2018.
- [22] Zhang Z., Liu Q., and Wang Y. Road extraction by deep residual U-net[J]. IEEE Transactions on Geoscience and Remote Sensing Letters, 15(5): 749-753, 2018.
- [23] He H., Yang D., Wang S., Wang S., and Li Y. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss[J]. Remote Sensing, 11(9): 1015, 2019.
- [24] Chen L.-C., Papandreou G., Kokkinos I., Murphy K., and Yuille A. L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848, 2018.
- [25] Yang X., Li X., Ye Y., Lau R. Y. K., Zhang X., and Huang X., Road detection and centerline extraction via deep recurrent convolutional neural network U-Net[J]. IEEE Transactions on Geoscience and Remote Sensing, 57(9): 7209-7220, 2019.
- [26] Chaurasia A. and Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation[C]. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), 1-4, 2017.
- [27] Zhou L., Zhang C., and Wu M. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction[C]. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 192-196, 2018.
- [28] Demir I., Koperski K., Lindenbaum D., et al. DeepGlobe 2018: A challenge to parse the Earth through satellite images[C]. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 172-181, 2018.
- [29] Liu R., Miao Q., Song J., et al. Multiscale road centerlines extraction from high-resolution aerial imagery[J]. Neurocomputing, 329: 384-396, 2019.
- [30] Costea D., Marcu A., Slusanschi E., and Leordeanu M. Roadmap generation using a multistage ensemble of deep neural networks with smoothing-based optimization[C]. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 210-214, 2018.
- [31] Cheng G., Wang Y., Xu S., Wang H., Xiang S., and Pan C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 55(6): 3322-3337, 2017.
- [32] Lu X., Zhong Y., Zheng Z., et al. Multiscale and multitask deep learning framework for automatic road extraction[J]. IEEE Transactions on Geoscience and Remote Sensing, 57(11): 9362-9377, 2019.
- [33] Bastani F., Songtao H., Abbar, S., et al. RoadTracer: Automatic extraction of road networks from aerial images[C]. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition (CVPR), 4720-4728, 2018.
- [34] Russakovsky O., Deng J., Su H., et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 115(3): 211-252, 2015.
- [35] Lin T.-Y., Maire M., Belongie S., et al. Microsoft coco: Common objects in context[C]. In *Proceedings of the 11th European Conference on Computer Vision (ECCV)*, 740-755, 2014.
- [36] Haklay M. and Weber P. OpenStreetMap: User-generated street maps[J]. *IEEE Pervasive Computing*, 7(4): 12-18, 2008.
- [37] Bearman A., Russakovsky O., Ferrari V., and Li. F.-F. What's the point: Semantic segmentation with point supervision[C]. In *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, 549-565, 2016.
- [38] Lin D., Dai J., Jia J., He K., and Sun J. ScribbleSup: Scribble-supervised convolutional networks for semantic segmentation[C]. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3159-3167, 2016.
- [39] Khoreva A., Benenson R., Hosang J., Hein M., and Schiele B. Simple does it: Weakly supervised instance and semantic segmentation[C]. In *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 876-885, 2017.
- [40] Ahn J. and Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4981-4990, 2018.
- [41] Tang M., Djelouah A., Perazzi F., Boykov Y., and Schroers C. Normalized cut loss for weakly-supervised CNN segmentation[C]. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1818-1827, 2018.
- [42] Tang M., Perazzi F., Djelouah A., Ayed I. B., Schroers C., and Boykov Y. On regularized losses for weakly-supervised CNN segmentation[C]. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 524-540, 2018.
- [43] Obukhov A., Georgoulis S., Dai D., and Van Gool L. Gated CRF loss for weakly supervised semantic image segmentation, arXiv:1906.04651, 2019.
- [44] Kolesnikov A. and Lampert C. H. Seed, expand and constrain: Three principles for weakly-supervised image segmentation[C]. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 695-711, 2016.
- [45] Wang B., Qi G., Tang C., et al. Boundary perception guidance: A scribble-supervised semantic segmentation approach[C]. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 3663-3669, 2019.
- [46] Zhang J., Yu X., Li A., Song P., Liu B., and Dai Y. Weakly-supervised salient object detection via scribble annotations[C]. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 12546-12555, 2020.
- [47] Yuan J. and Cheriyyadat A. M. Road segmentation in aerial images by exploiting road vector data[C]. In *Proceedings of the 4th International Conference on Computing for Geospatial Research and Application*, 16-23, 2013.
- [48] Mattyus G., Wang S., Fidler S., and Urtasun R. Enhancing road maps by parsing aerial images around the world[C]. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1689-1697, 2015.
- [49] Kaiser P., Wegner J. D., Lucchi A., Jaggi M., Hofmann T., and Schindler K. Learning aerial image segmentation from online maps[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):

- 6054-6068, 2017.
- [50] Wu S., Du C., Chen H., Xu Y., Guo N., and Jing N., Road extraction from very high resolution images using weakly labeled OpenStreetMap centerline[J]. *ISPRS International Journal of Geo-Information*, 8(11): 478, 2019.
- [51] Lecun Y., Bottou L., Bengio Y., and Haffner P., Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 86(11): 2278-2324, 1998.
- [52] Zeiler M. D., Taylor G. W., and Fergus Rob. Adaptive deconvolutional networks for mid and high level feature learning[C]. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2018-2025, 2011.
- [53] 周志华. 机器学习[M]. 清华大学出版社, 2016.
- [54] Mnih V. Machine learning for aerial image labeling[D]. Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 2013.
- [55] Shekhar S. and Xiong H. Google Earth, in *Encyclopedia of GIS*, 2008.
- [56] Zhang T. Y. and Suen C. Y. A fast parallel algorithm for thinning digital patterns[J]. *Communications of the ACM*, 27(3): 236-239, 1984.
- [57] Wiedemann C., Heipke C., and Mayer H. Empirical Evaluation of Automatically Extracted Road Axes[M]. *Empirical evaluation techniques in computer vision*, 172-187, 1998.
- [58] Otsu N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1): 62-66, 1979.
- [59] Tian Z., He T., Shen C., and Yan Y. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation[C]. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3126-3135, 2019.
- [60] Shi J. and Tomasi. Good features to track[C]. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 593-600, 1994.
- [61] Kingma D. P. and Ba J. L. Adam: A method for stochastic optimization[C]. In *Proceedings of the 3rd International Conference on Learning Representations*, 1-15, 2015.
- [62] Boykov Y., Veksler O., and Zabih R. Fast approximate energy minimization via graph cuts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11): 1222-1239, 2001.
- [63] Boykov Y. and Kolmogorov V. An experimental comparison of mincut/max-flow algorithms for energy minimization in vision[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9): 1124-1137, 2004.
- [64] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., and Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11): 2274-2282, 2012.
- [65] Arbeláez P., Maire M., Fowlkes C., and Malik J. Contour detection and hierarchical image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5): 898-916, 2011.
- [66] Xie S. and Tu Z. Holistically-nested edge detection[C]. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, 1395-1403, 2015.
- [67] Adams A., Baek J., and Davis M. A. Fast high-dimensional filtering using the permutohedral lattice[J]. *Eurographics*, 29(2): 753-762, 2010.
- [68] Canny J. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679-698, 1986.
- [69] Sobel I. E. Camera models and machine perception[D]. Doctoral Dissertation, Stanford Univ.,

Stanford, CA, USA, 1970.

武汉大学

## 攻读硕士期间主要成果

### 学术论文

- [1] Yao Wei and Shunping Ji. Scribble-based Weakly Supervised Deep Learning for Road Surface Extraction from Remote Sensing Images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, accepted and to be published, DOI: 10.1109/TGRS.2021.3061213. (SCI, 二区, 影响因子: 5.855)
- [2] Yao Wei, Kai Zhang, and Shunping Ji. Simultaneous Road Surface and Centerline Extraction from Large-Scale Remote Sensing Images Using CNN-Based Segmentation and Tracing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8919-8931, Dec. 2020, DOI: 10.1109/TGRS.2020.2991733. (SCI, 二区, 影响因子: 5.855)
- [3] Yao Wei, Kai Zhang, and Shunping Ji. Road Network Extraction from Satellite Images Using CNN Based Segmentation and Tracing[C]. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Yokohama, Japan, 2019, pp. 3923-3926, DOI: 10.1109/IGARSS.2019.8898565. (EI, 会议论文)

### 发明专利

- [4] 季顺平, 魏瑶. 一种基于卷积神经网络弱监督学习的遥感影像道路分割方法, 发明专利申请号: 202010771919.6, 申请公布日: 2020-12-11, 公布号: CN112070779A.
- [5] 季顺平, 魏瑶, 张凯. 一种同时提取遥感影像道路路面和中心线的深度学习方法, 发明专利申请号: 201911228166.8, 申请公布日: 2020-04-21, 公布号: CN111046768A.

### 软件著作权

- [6] 季顺平, 魏瑶, 张永军. 基于弱监督分割的遥感影像道路路面提取软件 V1.0, 软件著作权登记号: 2020SR1190370.

### 科研项目参与情况

- [7] 2019.01-2020.12. 超大范围光学卫星影像联合处理与信息提取. 武汉大学学科交叉类资助项目
- [8] 2019.10-2019.12. 基于高分辨率遥感影像的地物自动变化发现研究. 浙江省第二测绘院合作项目

## 获奖情况

[9] 2019-2020 武汉大学“优秀研究生”

[10] 2019-2020 武汉大学遥感信息工程学院“优秀共青团员”

武汉大学



## 致谢

2018年盛夏，我来到武汉大学开始研究生阶段的学习，至此度过了难忘的三年。临近毕业，杂感交集。虽然写下这段文字之时还未言再见，却已开始怀念在珞珈的日子。值此论文完成之际，向所有在学习、工作、生活上关心和帮助我的师长、同学、亲友致以真诚的感谢。

首先，感谢父母对我的养育和无条件的爱，在我的求学路上始终默默支持和理解。爸妈为我指点迷津时眼底的温柔，永远是我勇往直前的动力。

特别感谢我的导师季顺平教授将我引入科学研究的大门，为我提供了诸多帮助。季老师严谨的治学态度和敏锐的前瞻思想使我受益匪浅，老师带着我们赴日本横滨参加国际学术会议时，鼓励我主动提高学术交流能力，他渊博的专业知识令我印象深刻。季老师在每周的组会上认真负责地坚持与每一位同学讨论进展并提出建设性意见，正是有了季老师的启发与督促，我得以顺利完成硕士生涯中各个阶段的科研任务。十分庆幸当初保研时选择加入季老师的团队，也非常感激老师愿意收我入门下。

这三年中我认识了很多可爱的朋友们。感谢课题组这个大集体，感谢师兄：秦梓杰、张驰、魏世清、沈彦雲，感谢师姐：田思琦、朱若瑾、戴佩玉，感谢同届同学刘瑾、姜帆、王瑄、张缘、马超、张志力、王雅梦、余大文，感谢师弟：陈佳港、罗冲、胡明阳、王定盼、曾畅、高建、尉浩然、张韬，感谢师妹罗牧莹、耿笑笑、张雨姝、赵琳盈、闫晶晶。非常幸运认识张凯和我一起的道路提取研究方面共同努力，也非常开心认识室友杜影丽、魏晓冰、曹银霞。感谢大家的帮助和陪伴，我想我们都会怀念一起奋斗的那些发着光的日子。未来，希望大家以梦为马，踏遍千山万水，归来仍是少年。

犹记韶华在珞珈，人生待光华。

2021年4月21日

于武汉大学

# 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目的前提下, 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): 魏瑶

学 号: 2018202130072

学 院: 遥感信息工程学院

日期: 2021年 6 月 3日