

学号 2015302590155

密级                     

# 武汉大学本科毕业论文

## 基于 CNN 的遥感影像道路提取方法研究

院（系）名 称：遥感信息工程学院

专 业 名 称 ：遥感科学与技术

学 生 姓 名 ：张凯

指 导 教 师 ：季顺平 教授

二〇一九年五月

# 郑 重 申 明

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名：\_\_\_\_\_

日期：\_\_\_\_\_

BACHELOR'S DEGREE THESIS  
OF WUHAN UNIVERSITY

Research on Road Extraction Method  
of Remote Sensing Image Based on CNN

College: School of Remote Sensing and Information Engineering

Subject: Remote Sensing Science and Technology

Name: Kai Zhang

Director: Shunping Ji Professor

May 2019

## 摘要

道路网在城市规划、导航、自动驾驶等多个方面发挥着重要作用。然而，构建和更新道路网是一个耗时又耗力的工作。随着遥感技术的发展，高质量的最新的高分辨率卫星图像可供越来越多的人使用。利用遥感图像提取道路网为道路网建设与更新提供了更好选择。本文提出了三种不同的方法，分别是基于分割、中心线追踪方法和将二者综合的方法进行道路自动提取。最新的深度学习方法如卷积神经网络和传统的图像处理方法被用来完成这项任务。所有这些方法都在三个公开以及私有数据集中评估。实验表明，所有提出的经过改进的方法都比原始方法获得更好的结果，本文最后提出的集成方法具有最佳的 IoU 和道路连通性。

**关键词：**道路提取；卷积神经网络；遥感影像

## **ABSTRACT**

Road networks play an important role in many aspects, such as city planning, navigation, autonomous driving. However, constructing and updating road networks are both time-consuming and labor-intensive. With the development of remote sensing technology, the latest and qualified high-resolution satellite images are available to more and more people. Utilizing remote sensing images to extract road networks is a better choice for road network construction and updating. In this paper, three different approaches, segmentation based, centerline tracing method, and integrated method, are proposed for automatic road extraction. The latest deep learning method like convolutional neural networks and traditional image processing methods are used to accomplish this task. All these methods are evaluated in three public and private datasets. All these improved methods obtain better results than original methods and the novel integrated methods have the best IoU and road connectivity.

**Key words:** Road Extraction; Convolutional Neural Networks; Remote Sensing Images

# 目录

1 绪论 .....	1
1.1 研究背景.....	1
1.2 国内外研究现状.....	2
1.2.1 道路分割.....	2
1.2.2 道路中心线提取.....	3
1.3 本文内容安排.....	4
2 基础知识.....	4
2.1 卷积神经网络.....	4
2.2 几种特殊卷积.....	5
2.2.1 空洞卷积.....	5
2.2.2 转置卷积.....	6
2.3 集成学习.....	7
2.4 本章小结.....	7
3 研究内容.....	8
3.1 道路分割.....	8
3.1.1 基于 D-LinkNet 的遥感影像道路提取.....	8
3.1.2 基于 RoadNet 的遥感影像道路提取.....	10
3.1.3 基于修补策略的遥感影像道路提取.....	13
3.2 道路中心线追踪.....	15
3.2.1 单起点道路中心线追踪.....	16
3.2.2 多起点道路中心线追踪.....	19
3.3 分割追踪方法结合进行道路提取.....	21
3.4 本章小结.....	23
4 实验结果与分析.....	24
4.1 数据及评价标准.....	24
4.1.1 数据集.....	24
4.1.2 软硬件环境配置.....	27
4.1.3 评价指标.....	27
4.2 实验结果.....	30

4.2.1 道路分割.....	30
4.2.2 道路中心线追踪.....	31
4.2.3 结合道路分割与中心线追踪方法.....	32
4.3 对比分析与讨论.....	33
4.3.1 分割方法.....	33
4.3.2 道路中心线追踪.....	35
4.4 本章小结.....	37
<b>5 总结与展望 .....</b>	<b>38</b>
5.1 总结.....	38
5.2 展望.....	38
<b>参考文献 .....</b>	<b>40</b>
<b>致谢.....</b>	<b>43</b>

# 1 绪论

## 1.1 研究背景

道路网作为一种基础地理数据，在城市规划，行车导航，地图绘制及更新等方面承担重要的作用。然而，由于道路网数据量大，更新快，精度要求高，使得道路网数据的构建和更新成为一种繁琐、昂贵并且需要大量人工操作的任务。目前，世界上许多地区的道路网地图覆盖范围差甚至已经过时，很多地图供应商无法及时制作高精度、广覆盖的地图。例如在多哈，商业供应商的地图始终落后于实际的道路网状态，即便在经过地图更新之后依旧如此。同时，随着基于位置的移动服务的日益普及以及自动驾驶汽车的快速发展，高精度以及更新及时的道路图显得尤其重要。

遥感技术作为一种大范围，远距离感知的对地观测技术，具有客观、准确、快速的特点。随着卫星遥感技术的发展，获取高分辨率、最新的遥感影像更加方便快捷。因此，利用遥感影像进行道路网的提取成为一种高效的手段。然而，在实际中，从高分辨率影像中提取道路依旧困难重重。例如，在高分辨率遥感影像中，由于光照的影响，城市中高大的建筑物形成的阴影，路旁行道树的遮挡，道路上行驶的车辆等等都使得提取的道路有着很多断裂块。许多学者提出了增强道路连接性的方法，但是大部分方法都是在道路提取的结果基础上进行了一系列后处理操作，而且它们只能在一定程度上提升最终结果，当基础的提取结果比较差或者有较多的噪声的时候，这类方法往往没有提升或者提升效果很小。实际情况中，道路的周围是十分复杂的场景，所以单单使用后处理操作来解决道路连接性问题不太可能，因此，需要找到具有更强鲁棒性的算法来进行道路的提取。

深度学习和人工智能是当今时代计算机领域最为火热的话题之一。深度学习起源于人类大脑的神经连接结构，由加拿大科学家 Hinton 提出，它通过连接多个神经元，从原始数据中逐层提取更加抽象的高层语义特征，从而发现并总结数据的特点，进而对数据进行分类。其中，卷积神经网络(CNN, convolutional neural network)是深度神经网络中的一种，利用离散卷积的操作对数据进行处理。卷积神经网络因其独特的结构，适用于处理结构化的数据，如图像，语音等。它在计算机视觉领



域中被广泛使用，如图像分割，图像分类，目标检测，超分辨率重建等等，都取得了良好的效果。

本文利用航空、卫星影像，结合卷积神经网络，分别从道路分割以及道路中心线提取方面进行道路网的提取及算法的优化改进，最后结合两种策略提出道路网提取的创新性框架，提高道路提取结果的连接性以及准确性，获得更加精确的道路网结果。

## 1.2 国内外研究现状

从上世纪七十年代以来，国内外的一些学者就开始利用遥感影像进行道路的提取。最初的提取策略都是基于道路的基本特征，如(a) 几何特征：一条道路通常呈条带状，宽度基本没有变化，而且具有明显的方向特征，维持一个较为稳定的长宽比。但是由于周围地物的遮挡，道路的边缘在遥感影像中可能表现不明显或者形状不规则。(b) 辐射特征：一条道路上的纹理以及颜色变化不大，但由于周围地物的遮挡容易产生一定范围的明暗变化。(c) 拓扑特征：道路的拓扑结构通常呈网状，具有强连接性，极少出现孤立的道路。

如今，利用深度学习的方法进行道路网提取主要可分为两大类别，一类是通过语义分割的方法提取道路，获得像素级的道路掩膜；另一类是通过 CNN 追踪道路的中心线，直接从影像中获取道路的拓扑结构。

### 1.2.1 道路分割

道路分割的方法通常是 will 遥感影像道路提取视为二分类问题，即道路为前景，其他非道路为背景，通过逐像素分类提取出道路。由于遥感影像中道路的特点，要求道路分割网络在保持空间信息的同时也要有较大的感受野，才能获得更好的连接性的道路。

Wand 等人[1]提出一种基于深度卷积神经网络和有限状态机的框架，用于从航空和卫星影像中提取道路网。深度卷积神经网络识别来自复杂和混乱环境的模式，有限状态机则是转换其识别的模式状态，从而捕获其追踪行为。结果表明，他们的方法可以获得比传统方法更加精确的路网信息。Satio 等人[2] 训练单个深度卷积神经网络同时提取道路和建筑物，所采用的的标签数据由建筑物、道路和背景组成。Sarhan 等人[3]提出了一种使用细胞神经网络的新型自动道路提取方法，利用

图像中道路的光谱和几何特性，构建一个名为“CNN-细胞神经网络”的框架。初步结果表明，该算法的精度高，速度快，甚至可以在硬件芯片组上实现。

除了使用卷积神经网络，全卷积神经网络也在遥感影像道路提取中广泛应用。卷积神经网络会受到卷积核大小的限制使得其感知区域有限，而全卷积神经网络则是拥有更大的视野与感知范围，所以能更加完好地从高级语义特征中判断像素的类别，获得更好的分类结果。Zhong 等人[4] 等人利用全卷积神经网络，从高级语义特征中恢复并且判断每个像素的类别，并在此基础上对不同超参数以及训练策略进行分析，最终确定了全卷积神经网络在航拍图像上提取道路的能力。Cheng 等人[5]提出一种级联的端到端的卷积神经网络(CasNet)，可以同时处理道路检测和中心线提取任务，即一个网络进行道路检测，另外一个与前者级联，利用生成的特征图进而生成道路的中心线。Fu 等[6]提出一种改进的 FCN 模型的高分辨率遥感影像精确分类的方法，其中引入空洞卷积来扩大感受野并且引入连接层(skip layer)增加细节，最后使用条件随机场(Conditional Random Field, CRF)进一步后处理输出类映射。Panboonyue 等[7]将深度卷积编码器-解码器网络(Deep Convolutional Encoder Decoder, DCED)的激活函数由整流线性单元 ReLU 替换为指数线性单元 ELU，之后添加测量(LMs)和 CRFs 来提高提取结果的准确度。

尽管在过去的几十年中，道路分割方法被广泛研究，但从高分辨率航空影像分割道路依旧是一件具有挑战性的任务。

### 1.2.2 道路中心线提取

道路中心线提取的策略通常是基于初始的分割结果，然后利用一些提取算法从分割的结果中提取中心线。常用的方法有形态学细化、霍夫变换及回归。形态学细化的方法总是会产生马刺状的分支；利用霍夫变换则会经常产生一些错误的中心线，而且对于非直线状的道路不能提取；基于回归的方法在提取交叉口的中心线时表现不佳。除了通过处理分割的结果获取中心线，Bastani[8]提出了一种基于 CNN 的，迭代搜索以及追踪道路中心线的方法。从一个道路已知点出发，通过输入以该点为中心的 RGB 影像，利用 CNN 推断的角度移动一定的距离或者退回到上一个节点，不断迭代循环最终遍历搜索栈的每个节点。

### 1.3 本文内容安排

第一章主要介绍本文的研究背景以及意义，分别从道路分割和道路中心线提取两方面简单介绍国内外遥感影像道路提取的研究现状，并指出两类方法的缺陷以及问题。

第二章是本文的理论基础部分，介绍卷积神经网络的基础知识以及主要特点，并结合本文研究任务，也简单介绍卷积神经网络的结构以及一些特殊的卷积。此外，本章也简单的介绍了集成学习的策略，主要为提升(Boosting)编码解码器的结构以及相似的网络结构变种。

第三章是本文的研究内容部分，介绍本设计中采用的框架，分别从三个方面介绍：道路分割、中心线追踪以及两者结合的策略，具体使用的算法会在该章节中详细的介绍。

第四章首先介绍本文中涉及的数据集、软硬件环境配置和评价指标，重点展示了第三章中所阐述方法对应的实验结果，以及在不同策略中使用不同算法的得到的结果以及讨论对比。

第五章是对本次设计的总结，并提出将来改进的方向。

## 2 基础知识

### 2.1 卷积神经网络

上世纪 60 年代，Hubel 等人通过对猫视觉皮层细胞的研究，首次提出感受野的概念，到 80 年代，Fukushima 在感受野概念的基础之上提出了神经认知机的概念，可以看作是卷积神经网络的第一个实现网络，神经认知机将一个视觉模式分解成许多子特征，然后进入分层递阶式相连的特征平面进行处理，它试图将视觉系统模型化，使其即使在物体有位移或轻微变形时也保持鲁棒性。

卷积神经网络是多层感知机(MLP)的变种，由纽约大学的 Yann Lecun 于 1998 年提出，其本质是一个多层感知机，成功的原因在于其所采用的局部连接和权值共享的方式，减少权的数量使得更容易优化同时也降低了模型的复杂度，减少了过拟合的风险。这些特点在网络的输入是图像时表现得更加明显。图像可以直接作为网

络的输入，避免了传统图像处理算法中复杂的特征提取和数据重建的过程，在二维图像的处理过程中，卷积神经网络能够自行抽取图像的特征包括颜色、纹理、形状及图像的拓扑结构，因此在处理二维图像的问题中，特别是识别位移、缩放及其它形式扭曲不变性的应用上具有良好的鲁棒性和运算效率等。

随着 2006 年深度学习的提出，人们普遍认为深度神经网络比浅层网络具有优异的学习性能。深度神经网络的基本思想是通过构建多层网络，对目标进行多层表示，以期通过多层的高层次特征来表示数据的抽象语义信息，获得更好的特征鲁棒性。与此同时，深度卷积神经网络作为一种带有卷积的深度神经网络模型，被广泛用于处理结构化数据，并且获得了较好的表现。

深度卷积神经网络主要包括三部分，卷积层、降采样层和全连接层，每一层有多个特征图，通过不同的卷积核进行不同特征的提取，然后通过网络的连接传递给下一层，这样逐层提取，最后获得具有高级语义信息的特征。

在利用卷积神经网络进行图像处理任务时，通常由输入层读入经过规则化（统一大小）的图像，每一层的每个神经元将感受野内的前一层的一组小的局部近邻的单元作为输入，神经元抽取一些基本的视觉特征，如边缘、角点等，这些基本特征之后被更高层的神经元使用。卷积神经网络通过卷积操作获得特征图，每个来自不同特征图的单元得到各自不同类型的特征。一个卷积层中通常包含多个具有不同权值向量的特征图，从而保留图像更丰富的特征。卷积层之后会连接池化层进行降采样操作，一方面可以降低图像的分辨率，减少参数量，另一方面可以获得平移和形变的鲁棒性。卷积层和池化层的交替使用，使得特征逐步增多，并且分辨率逐渐降低，是一个双金字塔结构。

## 2.2 几种特殊卷积

除了普通的卷积操作，深度卷积神经网络中还有一些异状卷积操作，它们拥有各自的特点，适用于处理不同的图像处理任务。本节将简单介绍其中的两种，空洞卷积和转置卷积。

### 2.2.1 空洞卷积

空洞卷积<sup>[9]</sup>的诞生背景是为了解决由于池化层的下采样方式造成了信息损失的问题。池化操作使得后面一层神经元的像素能够观察到前面神经元的更大

的部分，即增加了其感受野，但是在下采样的过程中，一些细节信息被丢弃掉，使得在后续的上采样过程得到的最终效果有一个上限。为了保留细节信息并且增大神经元的感受野，空洞卷积就被研究人员设计而出。

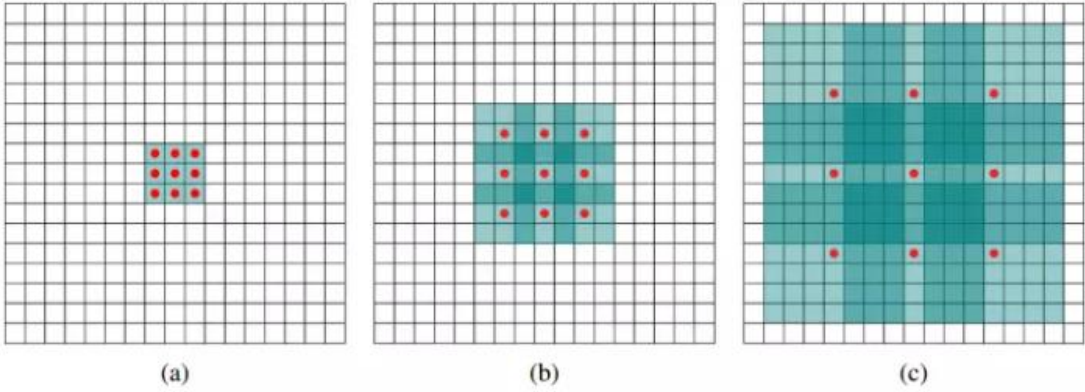


图 2.1 空洞卷积

如图 2.1，(a)对应  $3 \times 3$  的空洞率为 1 的卷积，和普通的卷积操作一样，(b)图对应  $3 \times 3$  的空洞率为 2 的卷积，实际卷积核的大小还是  $3 \times 3$ ，但是空洞为 1，也就是对于一个  $7 \times 7$  的图像切片，只有 9 个红色的点和  $3 \times 3$  的卷积核发生卷积操作，其余的点略过。虽然卷积核的大小只有  $3 \times 3$ ，但是卷积的感受野已经增大到了  $7 \times 7$ ，(c)图是空洞率为 4 的卷积，放在空洞率为 1 的卷积和空洞率为 2 的卷积后面，能达到  $15 \times 15$  的感受野。

### 2.2.2 转置卷积

转置卷积<sup>[10]</sup>也称反卷积，作用与卷积相反，主要从特征图中还原出原矩阵，即与原影像相似的低级语义特征。

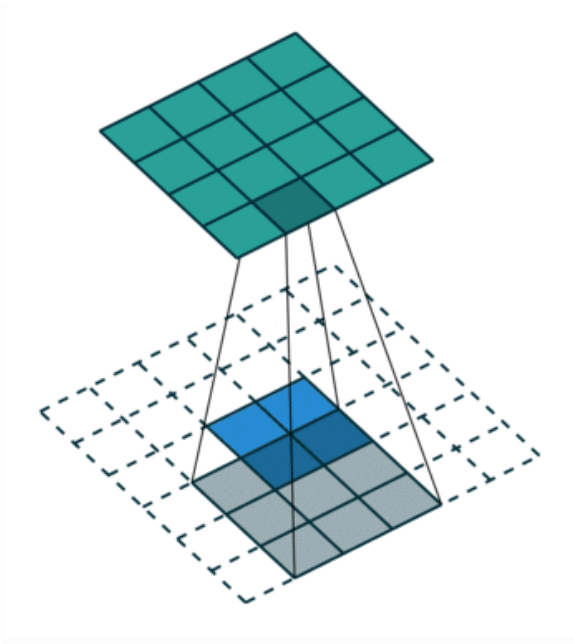


图 2.2 转置卷积

如图 2.2，转置卷积依旧采用卷积的策略，但是在卷积之前首先对输入的特征层进行边缘填充，填充的部分通常为 0，然后在填充的结果上进行卷积操作，最后获得比输入特征尺寸更大的计算结果。

转置卷积通常有三种用途，一种是图像重建，与卷积的步骤相反，即确定一种卷积核以及相应的特征图之后，利用反卷积的操作重建原图像。第二种是对 CNN 结果进行可视化，将特征图恢复到像素空间，从而发现卷积神经网络关注的图像区域，可以为目标检测提供初始的位置信息。第三种则是利用其上采样的特点，在图像语义分割中，往往需要与原始输入影像一样大小的预测结果，因为卷积会利用池化层或者步长等方式降低图像的尺寸，最后都要用转置卷积来把预测结果恢复到与输入影像一样大小的尺寸。

## 2.3 集成学习

集成学习<sup>[11]</sup>不是指一个单纯的机器学习算法，而是一种策略，通过构建并结合多个机器学习器来完成学习任务。组合多种学习器的关键在于不同的学习器要有差异性，即拥有不同超参数，不同的算法，不同的输入表示，不同的训练集等。

存在两种不同的组合方式获得最终的输出结果。(a)多专家组合。让多个学习器并行工作，然后获得输出结果。其中存在两种策略，一种是给定一个输入，让所有的分类器都产生一个输出，然后基于投票的策略或者层叠的策略进行结果的融合，另外一种是根据输入来选择一个或者几个学习器来产生输出。(b)多级组合。使用一种顺序的方法，其中下一个基学习器只在前一个基学习器预测不够准确的实例上进行训练或者检验。基本思想是，基学习器按照复杂度递增的方式排序，使得除非前一个更简单的基学习器的结果是不足够可信的，否则就不使用复杂的基学习器。

## 2.4 本章小结

本章主要介绍了卷积神经网络的基础知识，从卷积神经网络的历史出发，简单讲解了深度卷积神经网络的结构。除了普通卷积之外，还存在一些特殊的卷积方式，空洞卷积和转置卷积。本章简单地介绍了这些特殊的卷积方式，指明了它们的用途。

在本章的最后，简单阐述了集成学习的定义以及策略，强调了集成学习与普通机器学习方法的不同之处。

## 3 研究内容

本设计主要基于卷积神经网络进行遥感影像道路提取，分别使用三种策略进行道路提取。(a)基于图像语义分割的策略进行道路提取；(b)利用中心线追踪的方法进行道路中心线提取；(c)分割和追踪方法结合的策略进行道路提取。

### 3.1 道路分割

通过对遥感影像分割进行道路提取的策略源自于图像的语义分割。道路提取可以理解为图像的二分类问题，即遥感影像上的道路为前景，其它地物为背景，对遥感影像上所有的像素分别进行分类，把代表道路的像素分类出来，就完成了遥感影像的道路提取。如今，影像的语义分割通常采用编码器解码器结构，首先对原始影像进行编码下采样，通过卷积逐步提取高级语义特征，然后利用解码器进行特征的映射，逐级恢复各个像素的分类结果，最终，原始影像上每一个像素都会对应一个分类结果。

本章节主要介绍了两种遥感影像道路提取的分割网络，Dlinknet 网络以及本文改进的 RoadNet 网络。此外，本章节还提出了一种修补的策略，通过两种网络的融合提升道路分割的结果。

#### 3.1.1 基于 D-LinkNet 的遥感影像道路提取

在 2018 年的计算机视觉与模式识别挑战赛上(computer vision and pattern recognition workshop)，北京邮电大学的参赛队伍提出了 Dlinknet[12]网络用于 Deepglobe 道路提取挑战赛中，并且取得了第一名的成绩。Dlinknet 网络使用的是编码器-解码器结构，输入的是 RGB 遥感影像，输出结果是与输入影像一样大小的二值图。

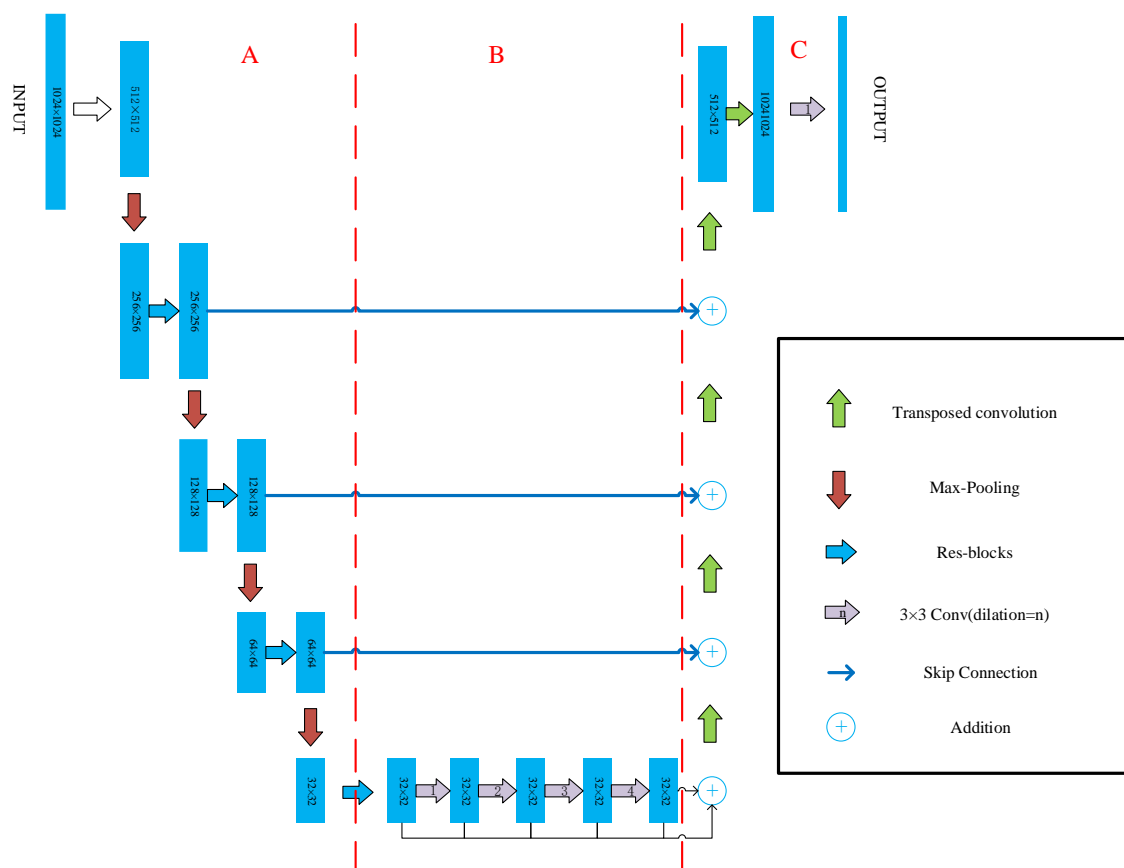


图 3.1 Dlinknet 网络图

如图 3-1，Dlinknet 由三部分组成，分别是编码器部分、中间部分和解码器部分。其中，编码器使用预训练的 ResNet34 模型，它是一种为中等尺寸的图像编码设计的编码器，由四个残差块组成，能够将图像下采样至原来的  $1/32$ 。

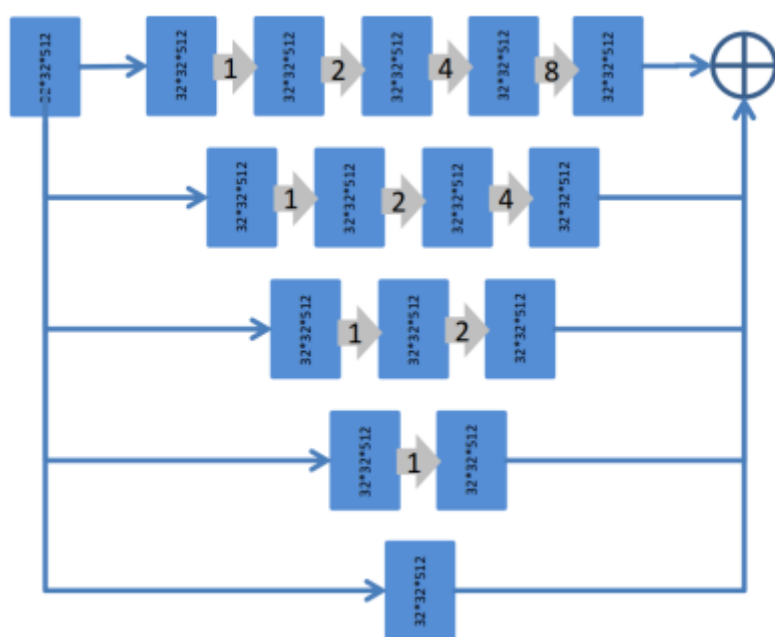




图 3.2 级联结构

中间部分是高级语义信息融合部分，Dlinknet 采用串联以及并联的方式把各个特征图融合在一起，如图 3.2，相邻的特征图之间相互串联，各个特征图还与最后一个进行并联。串联过程中，下一层的特征图是使用空洞卷积在上一层的结果上处理得到的，使用不同扩张率的空洞卷积使得卷积具有不同的感受野，获得多尺度的信息。

解码器是使用的转置卷积，从中间部分的最后一个融合的特征图开始，使用转置卷积逐层上采样，在每一层与对应的编码器层使用连接层互联，重复该步骤直至获得与输入图像同样尺寸大小的上采样结果，最后一层使用激活函数进行非线性映射，获得最终的分类结果图。

在训练过程中，Dlinknet 采用了迁移学习 (transfer learning) 策略。使用从 ImageNet<sup>[13]</sup>数据集预训练的 ResNet34 模型参数作为初始权重，利用小学习率进行微调适应新的数据集，完成新的学习任务。

### 3.1.2 基于 RoadNet 的遥感影像道路提取

尽管 Dlinknet 已经可以较好地提取道路结构，但是由于其使用了编码器解码器结构，网络参数较多，计算量大，训练速度慢。基于这些方面的问题，本文提出了一种新的网络模型 RoadNet，它具有计算量小，参数少，训练快的特点，是一个轻量级高效的网络模型。

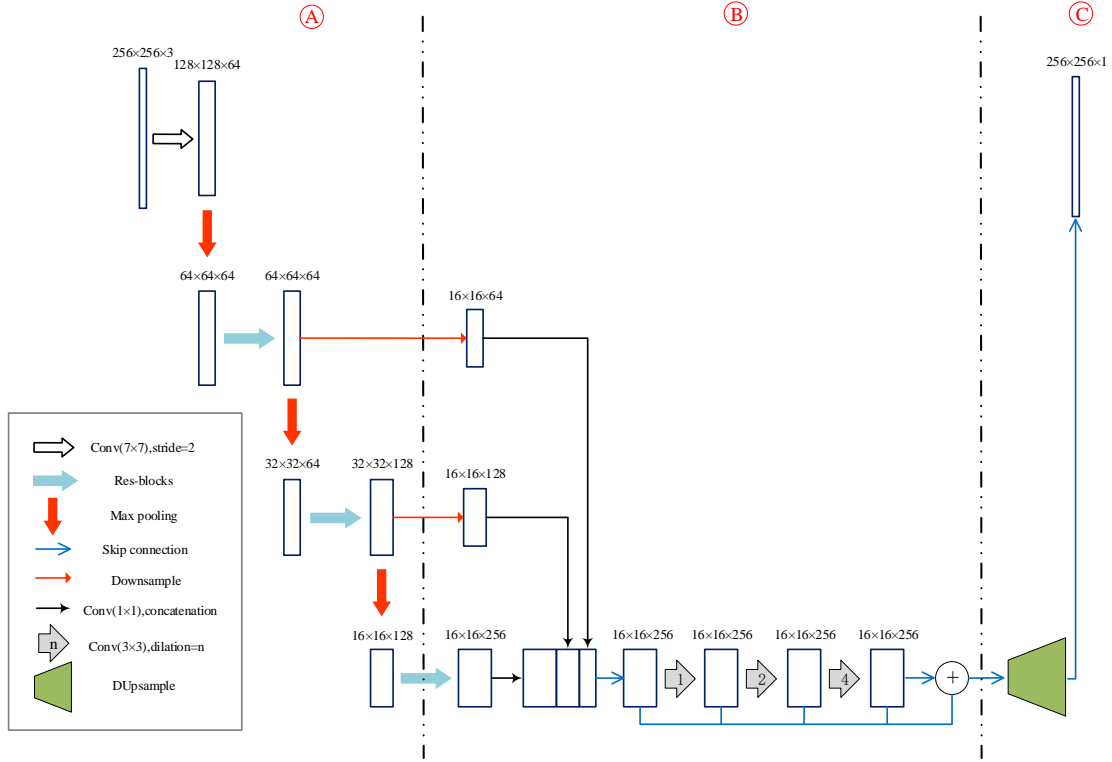


图 3-3 RoadNet 网络结构

如图 3-3，与 Dlinknet 相似，网络也是由三部分组成，编码器，中间部分和上采样部分。编码器采用的预训练的 ResNet34 模型进行编码，逐步提取图像的高级特征。

在中间部分，由编码器提取的特征没有直接利用空洞卷积进行融合，而是首先对最后三层的特征图进行叠加。不同尺度大小的特征图则是利用采样的方式转换到和编码器同样的大小，进行特征的叠加。然后对叠加后的特征图进行空洞卷积处理以及不同尺度特征的融合，获得最终的特征图。

上采样部分与 DlinkNet 不同，没有使用计算量大的转置卷积逐步上采样，而是使用 Dupsampling<sup>[18]</sup>模块直接从特征融合层恢复到与原始输入层一样的大小。

Dupsampling [18]是通过建立特征层和分类结果的映射关系，将融合后的特征层转换为最终的分类结果。令  $F \in R^{\hat{H} \times \hat{W} \times \hat{C}}$  表示经过中间部分空洞卷积以及不同特征图融合的结果， $Y \in 0, 1^{(H \times W)}$  代表真实值标签，其中 0 代表背景，1 代表道路。F 的尺寸大小是真实值 Y 的 1/16，即  $\frac{\hat{H}}{H} = \frac{\hat{W}}{W} = \frac{1}{16}$ 。为了在训练过程中计算损失函数以及获得每个像素的分类信息，需要把 F 上采样到和 Y 一样大小的尺寸。因为 Y 是独立分布的，每个位置的值都是相互独立的，所以 Y 可以被压缩为  $\hat{Y} \in R^{\hat{H} \times \hat{W} \times \hat{C}}$ ，

然后计算  $F$  和  $\hat{Y}$  之间的损失。

从  $Y$  压缩到  $\hat{H}$  的过程中, 用  $r = \frac{\hat{H}}{H} = \frac{1}{16}$ 。然后把  $Y$  分成  $\frac{H}{r} \times \frac{W}{r}$  个方格, 每个方格的大小为  $r \times r$ , 令  $S \in \{0,1\}^{r \times r \times C}$ , 把  $S$  铺平成一个一维向量  $v \in \{0,1\}^N$ , 其中  $N = r \times r \times C$ , 然后把向量  $v$  压缩为一个更低维的向量  $x \in \mathbb{R}^{\hat{C}}$ , 然后纵向以及横向堆叠所有的  $x$  从而获得  $\hat{Y}$ 。从  $v$  到  $x$  的压缩过程可以用一个线性变换  $P \in \mathbb{R}^{\hat{C} \times N}$  表示, 如公式 (3.1)。

$$x = Pv \quad (3.1)$$

相反, 从  $x$  到  $v$  的步骤可以表示为公式 (3.2)。

$$\hat{v} = Wx \quad (3.2)$$

其中  $W$  是  $P$  的逆矩阵, 将  $x$  重建到  $v$ 。  $\hat{v}$  是重建后的  $v$ 。矩阵  $P$  和  $W$  是通过在训练集上最小化重建  $v$  和  $\hat{v}$  之间的损失得到的。重建公式如式子 (3.3)。

$$P^*, W^* = \operatorname{argmin}_{P, W} \sum_v \|v - \hat{v}\|^2 = \operatorname{argmin}_{P, W} \sum_v \|v - WPv\|^2 \quad (3.3)$$

通过随机梯度下降来循环迭代, 直至收敛, 从而获得  $P$  和  $W$ 。然后可以依据这种映射关系, 将  $F$  映射到和  $Y$  同样大小尺寸, 然后利用损失函数计算损失。图 3.4 展示了该步骤的流程。

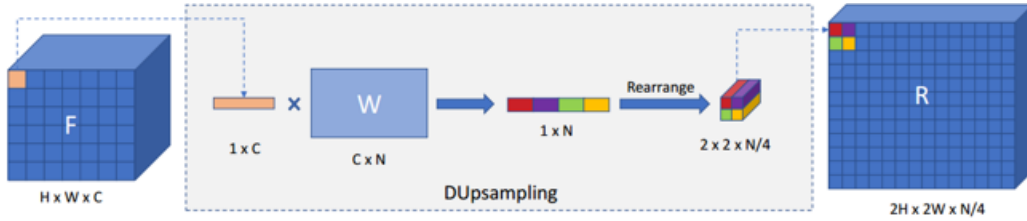


图 3.4 Dupsampling 结构

在计算最终的分割的损失时, 使用了二分类交叉熵损失函数和 dice 二分类交叉熵损失函数(Dice Binary Cross Entropy Loss, Dice BCE loss), 如公式(3.4)和公式 (3.5)。

$$L = 1 - \frac{\sum_{i=1}^N |P_i \cap GT_i|}{\sum_{i=1}^N (|P_i| + |GT_i|)} + \sum_{i=1}^N \text{BCELoss}(P_i, GT_i) \quad (3.4)$$

$$\text{BCELoss}(P, GT) = - \sum_{i=1}^W \sum_{j=1}^H [gt_{ij} \cdot \log p_{ij} + (1 - gt_{ij}) \log (1 - p_{ij})] \quad (3-5)$$

其中P代表预测结果，GT 代表真实值，N 代表批量大小。W是图像宽度，H是图像高度，gt 是 GT 中的一个像素，p 是 P 中的一个像素。

在预测过程中，使用了预测结果增强模块(Test Time Augmentation)，即对预测图像进行上下翻转，水平翻转，旋转等方式分别预测，然后把每次的预测结果进行加权取平均，从而获得最终的预测结果。

由于 Dupsampling 的结构，RoadNet 网络适合于对中小型影像进行分割处理，在本设计中其被用来在修补策略中的修补工具，详细信息见章节 3.1.3。

### 3.1.3 基于修补策略的遥感影像道路提取

考虑到基于 D-LinkNet 进行遥感影像道路提取的分割结果中存在断裂、误判、不完整的道路路段，在此设计针对在初始分割中效果较差的小区域进行精细分割的修补策略，以提高道路结果的连通性及准确性，从而实现对初始分割结果的优化改进。具体流程如图 3.5 所示，

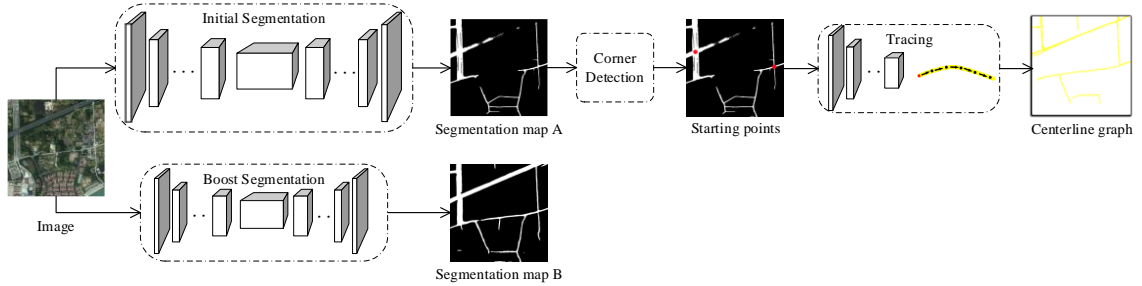


图 3.5 修补方法结构图

该方法主要分为三个步骤，首先对原始影像进行初始的分割，获得粗糙的分割结果，然后基于该分割结果进行种子点提取，为道路中心线追踪提供起始点，利用追踪的方法获取道路的中心线；第二步则是利用中心线找到初始分割结果中道路断裂的部分，将断裂部分从图像中裁出作为训练用于修补道路的网络，在预测过程中，由于修补网络的输入大小和原始分割的大小不同，因此将原始影像裁剪成多个小块，然后利用修补网络进行预测，最后将结果拼接形成和原始影像一样大小的分割结果图；第三部分则是将初始分割的结果，道路中心线的结果以及修补网络的分割结果进行融合，得到最终的道路分割结果。

第一步中对道路进行粗分割，本文采用的 Dlinknet 的网络模型，因为它拥有连接层，空洞卷积以及串联并联的结构，在编码解码部分保留了大量的信息，当对尺

寸较大的航空影像进行分割时具有优势。提取种子点以及利用种子点进行中心线提取的方法将在 3.2 节中详细介绍。

第二步中，修补网络采用的是本文提出的 RoadNet，它具有计算量小，训练速度快的特点，而且当训练数据检测较少的时候依旧效果较好。在训练 RoadNet 之前，需要把 Dlinknet 粗分割的结果以及道路中心线的真实值相叠加，遍历中心线的每一个结点，选取周围区域为一个小窗口，判断该区域道路是否已经被分割出来，如果已经分割出来，进行下一个结点的判断，如果没有，就以该结点为中心，创建一个  $256 \times 256$  的窗口，把该窗口内的影像和真实值保存下来作为 RoadNet 训练的数据集。最终形成的数据集几乎都是粗分割，即 Dlinknet，分割效果不好的区域，RoadNet 以此为基础进行学习，能够学习到互补的信息，从而在与粗分割结果进行融合的时候获得更好的道路分割效果。

第三步是将三种结果进行融合，本设计实验了两种融合策略，分别介绍如下。融合策略一为加权融合方式，如图 3.6 所示。

由于道路中心线的结果是矢量格式，在进行融合之前需要转化为栅格格式。本文采用的策略是将矢量格式的中心线首先转成 1 像素宽的栅格图，考虑到道路具有一定的宽度，对单像素宽的栅格图膨胀一定的宽度然后利用高斯滤波生成概率图，即从道路中心向两边呈高斯概率分布。在融合粗分割结果和 RoadNet 修补结果的过程中，采用的是加权融合的方式，为了增强分割结果的连接性，对融合结果进行道路中心线增强，即与加权后的道路中心线栅格概率图融合，进行平均操作。最后利用最大类间方差法 (OTSU) 自动确定阈值从而对概率图上进行二值化，获得最终的道路分割结果。

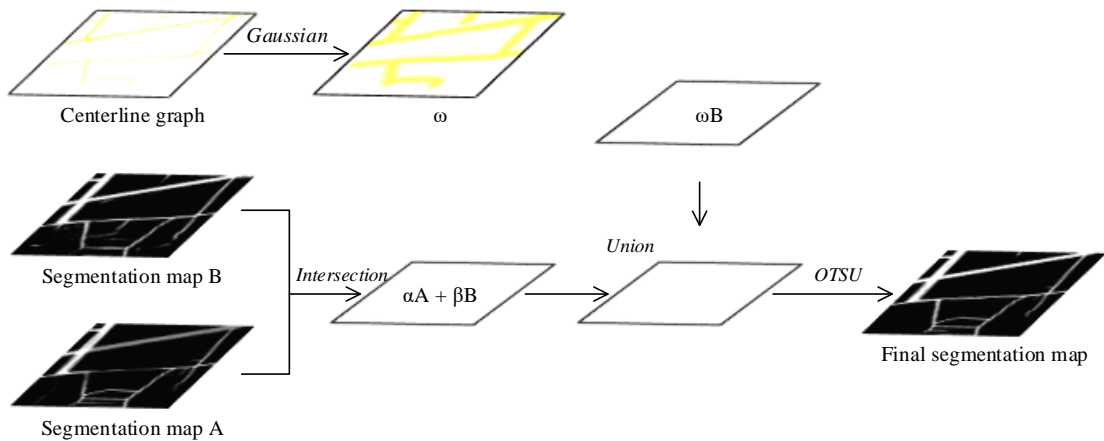


图 3.6 加权融合方式

OTSU 方法[14]是一种自动确定阈值的方法，是按照图像的灰度特征，将图像分为前景和背景两部分。因为方差是灰度分布均匀性的一种度量，前景和背景的两类之间方差越大，代表构成图像的两部分的差别越大，当部分背景被分为前景或者前景被分为背景，都会导致两部分的差别变小。因此，使用类间方差最大的分割能够让错分概率最小。在本设计中，采用 OTSU 代替直接取 0.5 作为分割阈值进行分割能够获得连接性更好的结果。

融合策略二为基于缓冲区的融合策略，如图 3.7 所示。

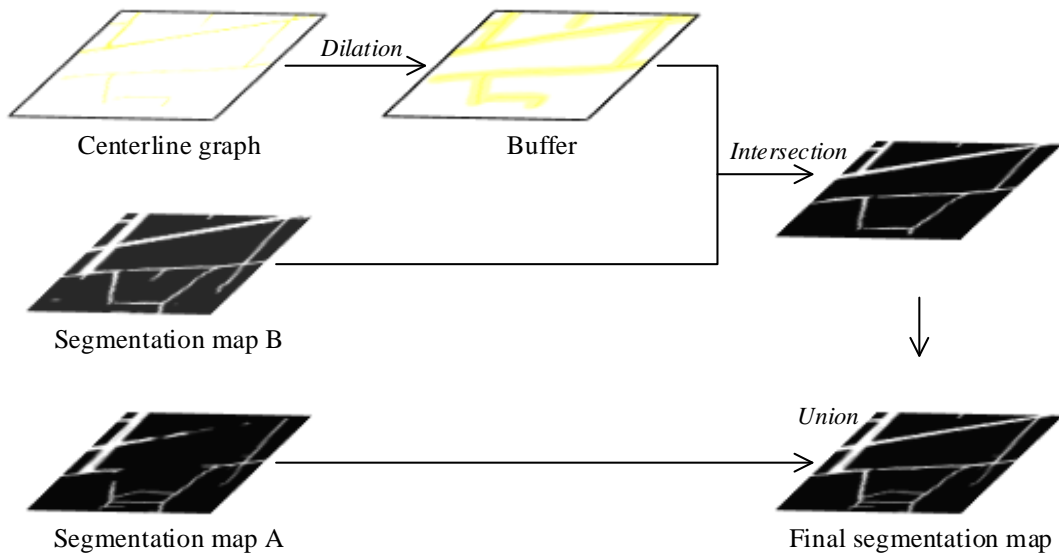


图 3.7 基于缓冲区的融合方式

与策略一不同，基于缓冲区的融合策略是把以道路中心线为中心，取一定范围的缓冲区，将缓冲区和修补网络的分割结果做交集，去除修补网络分割引入的噪声，把交集的结果与粗分割的结果进行融合，最后利用 OTSU 的策略对融合结果进行二值化，获得最终的道路分割结果。

### 3.2 道路中心线追踪

在遥感影像中，道路是具有复杂拓扑关系的条状地物，道路中心线为道路提取研究提供了重要的连通性及完整性信息。提取道路中心线的方法有很多，其中追踪的方法是一种直接从遥感影像获取道路网的方法，而不需要在初始道路分割的结果上进行中心线的提取。

本章介绍目前国际上道路中心线追踪的最新方法 RoadTracer [8]，并分析其优势与不足，结合道路网的特点，在此基础上提出一种新颖的多起点道路中心线追踪的改进策略，从而得到更完整的道路网结果。

### 3.2.1 单起点道路中心线追踪

Bastany 等人[8]于 2018 年计算机视觉与模式识别会议(Computer Vision and Pattern Recognition 2018, CVPR 2018)上提出 RoadTracer 方法，在作者的实验中，它能够从遥感影像中正确提取 59%的道路，并且具备完全连接性。RoadTracer 从一个已知是道路的起始点出发，利用卷积神经网络决策函数作为引导，搜索式、启发式地构建道路，从而实现对道路网的追踪，较好地保留了道路的拓扑连续性。它主要包含 CNN 决策函数和迭代搜索算法。

决策函数是基于卷积神经网络实现的，其网络结构如图 3.8 所示。

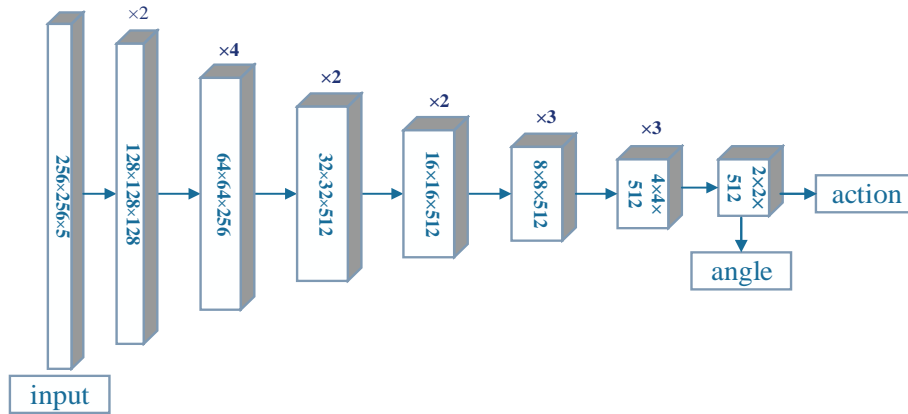


图 3.8 RoadTracer 决策网络

网络输入层总共由五个部分组成，分别是以搜索栈中的当前点  $S_{top}$  为中心的尺寸为  $256 \times 256$  窗口中的 RGB 影像、道路网中心线真值图以及用于存储所搜索的道路结果图。输出主要由两部分组成，分别为  $O_{action} = \langle O_{walk}, O_{stop} \rangle$  决定动作是 walk 还是 stop，和  $O_{angle}$  决定动作是 walk 时的搜索角度。



---

**Algorithm 1** Iterative Graph Construction

---

**Input:** A starting location  $v_0$  and the bounding box  $B$   
 initialize graph  $G$  and vertex stack  $S$  with  $v_0$   
**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**  $action = \text{stop}$  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**

---

图 3.9 迭代搜索算法

迭代搜索算法的流程如图 3.9 所示。输入一个区域( $V_0, B$ )， $V_0$ 是一个已知的起始点， $B$  是一个窗口，包含了要预测道路网络的图像。搜索算法要维持一个道路网络图  $G$  和一系列预测出来的道路的结点  $S$ ，在搜索开始时，两个都初始化为 $V_0$ 。 $S_{\text{top}}$ 表示  $S$  堆栈中最上端的结点元素，也代表当前搜索的位置。在每一步中，把  $G$ ， $S_{\text{top}}$ 和以 $S_{\text{top}}$ 为中心的影像输入到决策函数中，由决策函数决定是否从 $S_{\text{top}}$ 向某个方向移动一个固定的距离  $D$ ，或者停止向前移动并且返回到上一个搜索到的道路结点  $S_{\text{top}}$ 。当移动的时候，决策函数会选择一个最优的移动角度 $\alpha \in [0, 2\pi)$ 。然后，搜索算法会把新的结点  $u$  加入到记录道路结点的堆栈中，并且把边 $(S_{\text{top}}, u)$ 加入到边的堆栈中，将当前点 $S_{\text{top}}$ 转移到新的位置  $u$ 。如果决策函数决定停止，在道路结点的堆栈  $S$  中会把当前点 $S_{\text{top}}$ 出栈。停止向前搜索代表前面的道路都被搜索过了，所以搜索算法会向前返回直至到另外一个节点，从该结点重新搜索道路。迭代以上步骤直至存放点的搜索栈为空。这种方法有效地消除了生成道路中心线过程的中间图像表示步骤，并且由于其按路段添加的策略使得最终的追踪结果保持绝对的连接性，还避免了复杂的后处理操作。



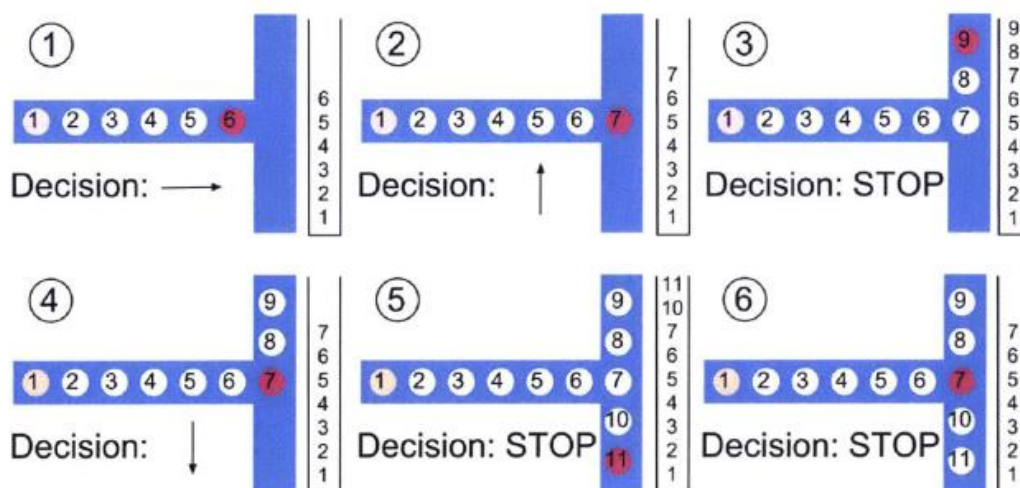


图 3.10 道路中心线搜索算法

图 3.10 是一个搜索算法运行的例子，蓝色代表在遥感影像上的道路，圆代表道路图  $G$  中的结点，其中红色圆代表搜索算法当前的结点。当搜索算法搜索到道路交叉口的时候，决策函数决定向上搜索，当搜索到该岔路的末尾的时候，因为没有另外的道路可以移动，决策函数决定停止搜索，当前点就逐渐向前返回，当再次到达交叉路口的时候，决策函数发现可以向下搜索，当前点会向着下方移动，直至再次到达道路的末尾，此时决策函数停止搜索，当前点逐步的向前返回，随着存放道路当前点的栈的出栈，当前点返回到最初的起点的时候，存放当前搜索点的栈变为空，此时，道路中心线追踪提取的任务就结束了。

在这个过程中，采用融合的方式将所有搜索过的道路的边结合到一起，在每一次添加新的道路结点的时候，都会先判断在范围  $3D$  内是否有已经搜索到的道路，避免搜索到重复的道路或者形成搜索的循环。

在训练决策函数的过程中，真实值标签是动态生成的，在每一次训练的步骤中，根据当前点所在的位置，然后从真实道路网判断是否还有未检测到的道路，如果有，则该处的  $O_{stop} = 0$ ,  $O_{walk} = 1$ ，如果没有，则  $O_{stop} = 1$ ,  $O_{walk} = 0$ 。角度标签是从真实道路网上找到离当前点最近的点，然后按照移动的方向在真实道路网上计算角度，最后，损失函数是通过计算角度之间的差距获得，进而反向传播更新决策网络的参数。

### 3.2.2 多起点道路中心线追踪

RoadTracer 能够追踪得到连接性较好的道路网，但同时也存在一定的问题。首先，它是在道路中心线真值中选取一个已知位置的种子点作为追踪的起始点，对中心线真值的依赖阻碍了其在实际生产中缺少标签数据的情况下的泛化应用；其次，由于遥感影像的复杂性，单起点道路中心线追踪受限于桥梁、高架桥等特殊地物环境的影响，往往会提前终止搜索，或者在某些区域不能搜索到，仅仅适用于连接程度比较明显的道路。

因此，本文结合道路路面分割方法和道路中心线追踪方法，提出了一种多起点道路中心线追踪的改进策略，先进行语义分割获得追踪的种子点，后由追踪方法实现道路中心线的提取，具体流程如图 3.11 所示。

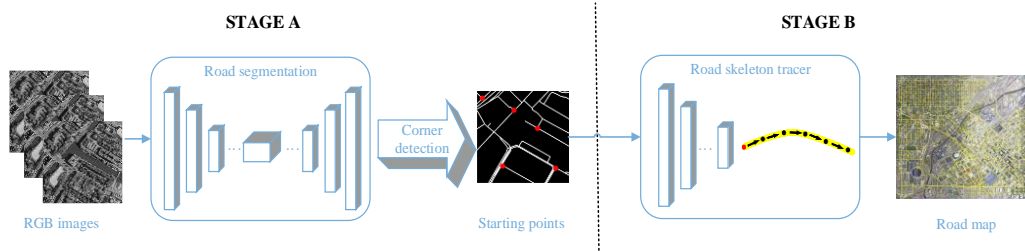


图 3.11 多起点道路中心线追踪方法流程

首先，对输入的原始 RGB 影像使用语义分割的网络进行初始分割，得到分割图之后，对分割结果进行形态学细化等操作。为了得到高质量的路网图，种子点应满足均匀分布且具有较强的潜在搜索方向。例如，如果种子点位于十字路口，则它有四个潜在的搜索方向。通过改进的 Harris 角点检测[15]算子来定位这些特征点，并作为后续追踪道路中心线的起始点。

在 Harris 算子中，角点评分计算方式如式(3.6)，

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (3.6)$$

而在 good feature to track 算子中，评分计算方式如式(3.7)，

$$R = \min(\lambda_1, \lambda_2) \quad (3.7)$$

其中  $\lambda_1 \lambda_2$  是图像的加权协方差矩阵  $M$  的特征值， $M$  计算方式如式(3.8)。

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \quad (3.8)$$

在使用算子进行种子点生成之前,先进行形态学细化处理分割结果,使得生成的种子点位于道路的中心,具有更高的置信度。

生成多个种子点之后,从一个种子点出发,使用 RoadTracer 的策略进行道路中心线跟踪以及道路网提取,然后从未遍历过的道路区域挑选新的种子点,再次进行道路中心线的提取。由于从多个起点出发分别进行追踪,最终构建的道路网具有重叠部分,本文提出了一种对多个道路网络融合的方法,删除重叠的部分,进而得到完整性、准确性较高的道路中心线结果。

从种子点列表中挑选位于没有遍历过的道路区域的种子点,是通过在种子点周围创建缓冲区实现的。下一次跟踪的起始点由先前探索过的道路是否在以当前点为中心的边界框之外决定的,如果边界框中有已经搜索到的道路结点,那么该种子点会被标记为已经遍历过,如果没有,该种子点会被挑选为下一个道路搜索的种子点,进行道路中心线构建。如图 3.12(a),其中黄色的线代表已经搜索到的道路,黑色的点代表道路网上的结点,蓝色的虚线框代表包围框,红色的点形成的包围框与道路有交点,则该点会被舍弃,绿色的点形成的包围框与道路没有交点,则该种子点会被保留,作为下次进行道路中心线追踪的起始点。

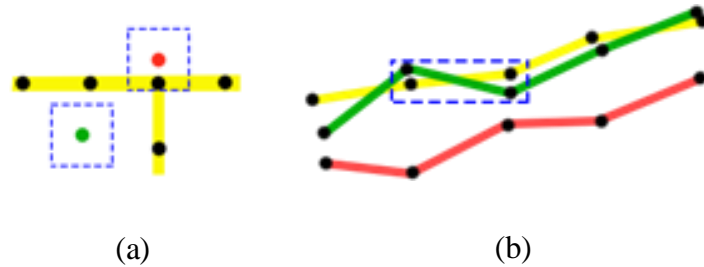


图 3.12 种子点挑选算法及中心线合并算法

在多起点追踪结果的合并流程中,首先,从多个起点追踪提取的道路网中挑选一个覆盖范围最大道路网的作为 $G_{base}$ ,然后把其它的道路网 $G_i$ 与 $G_{base}$ 交叠,遍历 $G_{base}$ 中的每一条边 $E_{base}$ ,构建包围盒,然后从 $G_i$ 中搜索与该包围盒的交叉的边,分别计算它们的角度,然后与 $E_{base,j}$ 求角度差,如果角度差小于阈值,则认为两条边可以合并,否则不能合并。如果两条边角度差为 0,则认为这两条边平行。如图 3.12(b)黄色的线代表 $G_{base}$ 中的边,黑色的点是道路网中的结点,绿色的和红色的线代表从 $G_i$ 中检索到的边,以 $G_{base}$ 中的一个边为中心,设置包围盒,按照以上的合并规则,绿色的边会被合并,红色的边则会被保留下来。

图 3.13 显示了多起点中心线追踪的整体流程。

---

**Algorithm 1 Road Centerline Tracer with multiple starting points**

---

Input: starting points list  $C$ , an initial graph array  $G_{array}$ , window  $W_i$  centered at  $C_i$ , threshold for line merge  $T$ ,

While  $C$  is not empty do

- random choose  $C_i$  from  $C$
- Remove  $C_i$  from  $C$
- Initialize  $W_i$  centered at  $C_i$
- if  $G_{array}$  intersect with  $W_i$ ; break
- else

  - $G_i = \text{road\_inference}(C_i, \text{Image})$
  - add  $G_i$  to  $G_{array}$

random choose  $G_{base}$  from  $G_{array}$

for  $G_i$  in  $G_{array}$

- for edge in  $G_i$ :

  - if  $\text{angle\_difference}(\text{edge}, G_{base}) > T$  then
  - add edge to  $G_{base}$

return  $G_{base}$

---

图 3.13 多起点中心线追踪流程

### 3.3 分割追踪方法结合进行道路提取

利用语义分割的方法能够分割提取到更多的道路，但是往往在道路连续性上表现不佳，即结果中存在断裂的、孤立的道路段；中心线追踪的方法可以获得连接性高的道路拓扑信息，然而会受到影像的限制导致其搜索到的道路有限。本设计中提出一种将两者结合的框架进行遥感影像道路提取，结合道路分割的优点和中心线追踪的优点，提取出具有高精度，广覆盖的道路网络。

该方法的流程如图 3.14，对 RGB 影像利用 Dlinknet 网络进行初始的分割，然后依据分割结果中较差的区域生成 Boosting 网络的训练集，利用该训练集训练 Boosting 网络，学习互补信息。在预测过程中，将初始分割的结果和 Boosting 网络的结果进行融合，并基于融合的结果使用多起点中心线追踪的方法提取道路中心线，最后将中心线的结果与分割网络的结果融合，形成最终的道路分割结果，然后从分割结果生成道路中心线。

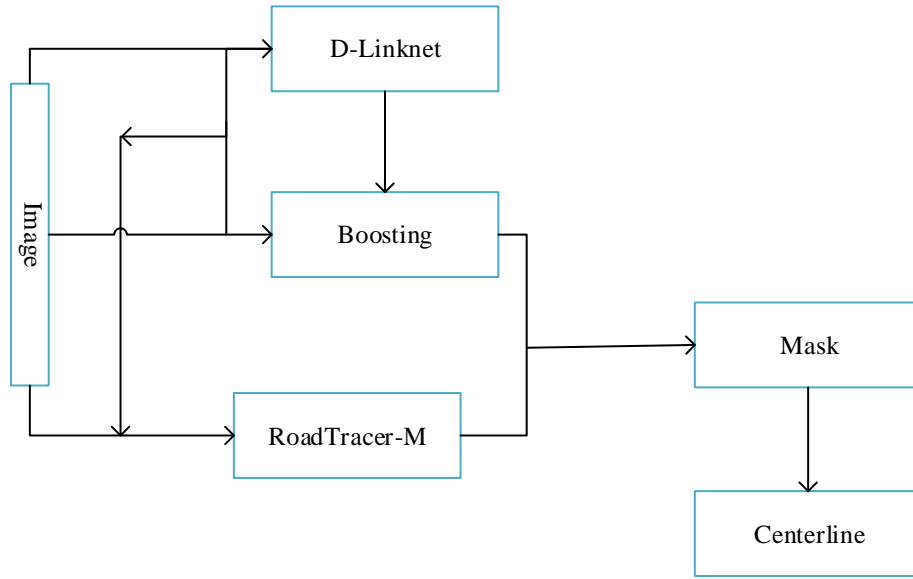


图 3.14 结合方法的主要流程

Boosting 网络是采用了集成学习中的提升(Boosting)策略。在提升中，下一个分类器是基于上一个分类器的结果训练的，下一个分类器学习到的信息是上一个分类器的互补信息，通过组合多个弱分类器得到强分类器。弱分类器是指错误率低于 0.5 的分类器，通常是简单的分类器，强分类器则是具有任意小错误率的分类器。在 Boosting 网络中，将 RoadNet 作为弱分类器，采用 AdaBoost 的理念进行构建。网络的流程如图 3.15。

训练：

```

For 所有的  $\{x^t, r^t\}_{t=1}^N \in X$ , 初始化  $p_1^t = 1/N$ 
For 所有的基分类器  $j=1, \dots, L$ 
  按照概率  $p_j^t$  随机地从  $X$  中抽取  $x_j$ 
  使用  $x_j$  训练  $d_j$ 
  For 每个  $\{x^t, r^t\}$ , 计算  $y_t = d_j(x^t)$ 
  计算错误率:  $\varepsilon_j = \sum_t p_j^t \cdot 1(y_j^t \neq r^t)$ 
  If  $\varepsilon_j > 1/2$  then  $L = j - 1$ ; stop
   $\beta_j = \varepsilon_j / (1 - \varepsilon_j)$ 
  For 每个  $(x^t, r^t)$  //如果正确, 降低概率
    If  $y_j^t = r^t$  then  $p_{j+1}^t = \beta_j p_j^t$  Else  $p_{j+1}^t = p_j^t$ 
  //规范化概率
   $Z_j = \sum_t p_{j+1}^t$ ;  $p_{j+1}^t = p_{j+1}^t / Z_j$ 
预测：
给定  $x$ , 计算  $d_j(x)$ ,  $j=1, \dots, L$ 
计算类输出,  $i=1, \dots, K$ :

$$y_i = \sum_{j=1}^L \left( \log \frac{1}{\beta_j} \right) d_{ji}(x)$$

  
```

图 3.15 Boosting 策略流程

构建的主要思想是每一个弱分类器的训练集是通过加权随机抽取原训练集构成的。令  $p_j^t$  表示每张图片  $x^t$  被抽取用于训练第  $j$  个分类器的概率。最初，所有的  $p_1^t = 1/N$ ，即所有图片被抽取的概率是相同的，然后从  $j=1$  开始，令  $\epsilon_j$  表示分类器  $d_j$  的错误率，设定  $T$  为产生最多分类器的数目，如果在训练过程中，出现了某个分类器的错误率  $\epsilon_j > 1/2$ ，则停止构建新的分类器。定义  $\beta_j = \frac{\epsilon_j}{1-\epsilon_j} < 1$ ，当  $d_j$  正确的对  $x^t$  分类时，更新图片  $x^t$  的概率  $P_{j+1}^t = \beta_j p_j^t$ ，如果分类错误，则  $P_{j+1}^t = p_j^t$ 。在每一次更新结束概率之后，都以  $\sum_t p_{j+1}^t$  进行规范化，使得概率的和为 1。这样可以使得正确分类的图像的概率降低，而错误分类的图像的概率提高，其中，采用阈值  $T$  与 IoU 比较来表示图像分类是否正确。然后，根据修改后的概率，从原样本中有放回的抽取相同大小的样本集，用于训练  $d_{j+1}$ 。这样做的效果就是使得分类器  $d_{j+1}$  更加专注于被  $d_j$  误分类的图像。一旦训练结束，就采用投票的方法，即对所有的  $d_j$  的分类结果进行加权取平均，其中每个分类器的权与准确率成正比： $w_j = \log(\frac{1}{\beta_j})$ 。

另外，在中心线结果与分割结果融合的过程中，将矢量的中心线变成栅格图像是通过一种自适应膨胀的操作实现的。在分割的结果中根据中心线提取的道路结点击开启小窗口，然后统计窗口内分割出来的道路的面积，并且计算道路的长度，将道路的面积除以道路的长度获得该窗口内道路的平均宽度。基于此宽度，对该结点所在的边膨胀一定的宽度，最终产生栅格形式的道路图。

### 3.4 本章小结

本章主要介绍了本设计中采用的方法，从语义分割道路提取，道路中心线追踪以及两者结合的方法分别进行阐述。针对每一种方法，本章对其整体流程进行描述，并针对流程中特殊的部分进行详细的讲述。另外针对不同方法得到的结果，在本章中提出了三种不同的策略进行结果的融合，从而提升道路分割的结果。

## 4 实验结果与分析

### 4.1 数据及评价标准

#### 4.1.1 数据集

本文的主要研究内容是基于遥感影像利用卷积神经网络方法实现对道路网的提取,需要遥感影像数据及道路标签数据。在此选取三套数据集作为主要实验数据来对本文的研究方法进行科学的实验验证,其中一套是马萨诸塞州的开源道路数据集(在下文中记为马萨诸塞州数据集),一套是非公开的我国浙江省绍兴市的道路数据集(在下文中记为浙江数据集),还有一套是同文章[8]一致的道路数据集(在下文记为 RoadTracer 数据集)。

马萨诸塞州数据集[16]由美国马萨诸塞州政府发布,包括分辨率为 1 米的 1711 张  $1500 \times 1500$  像素的航空影像,覆盖广泛的的城市、郊区、乡村地区,总计面积约为 2600 平方公里。标签样本是通过栅格化从 OpenStreetMap[17] 项目中所获取的道路中心线而生成的,线条采用 7 像素宽且没有平滑。其中,影像中有很多数据缺失的空白区域,导致影像与道路标签数据不匹配影响实验准确度和效率。在此进行数据预处理操作,剔除质量不佳的部分数据,并将其随机分为 192 张  $1024 \times 1024$  训练集和 64 张  $1024 \times 1024$  测试集。图 4.1 为该数据集的示例图像。





图 4.1 马萨诸塞州数据集

浙江数据集是中国浙江省绍兴市区域的卫星影像及道路标签数据，分辨率为 0.2 米，有 4788 张影像。为了与其它方法对比，对数据集进行了拼接和下采样至 0.6 米分辨率，并随机分为 372 张  $1024 \times 1024$  训练集和 160 张  $1024 \times 1024$  测试集。图 4.2 为浙江数据集的示例图像。



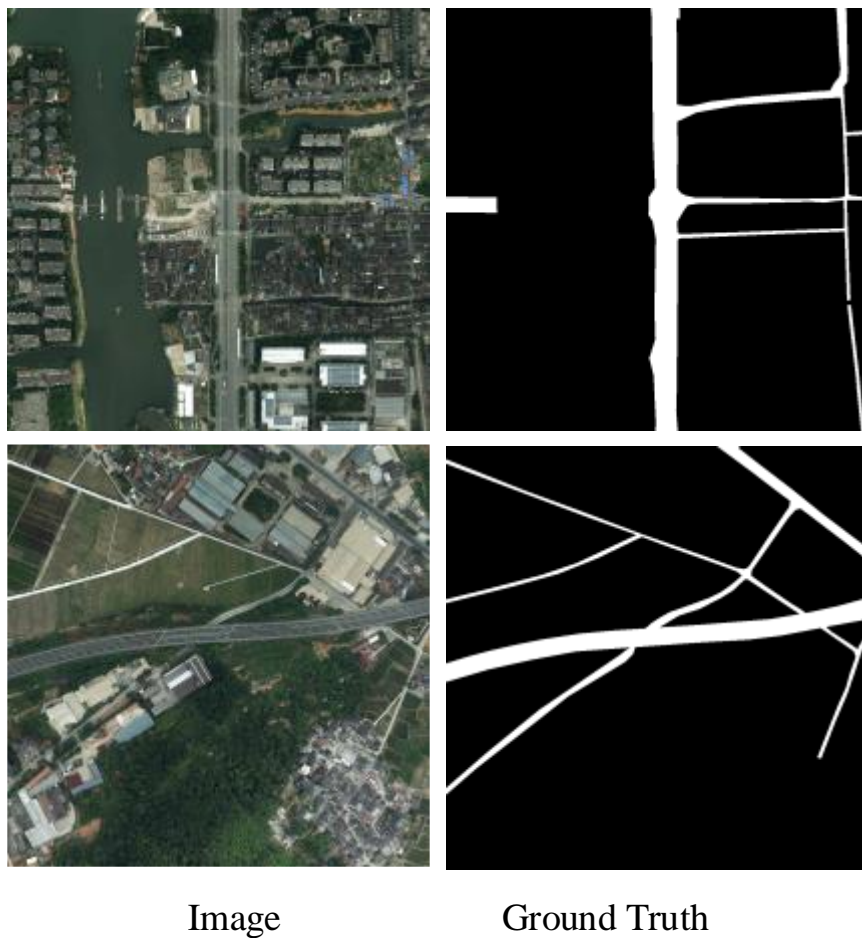
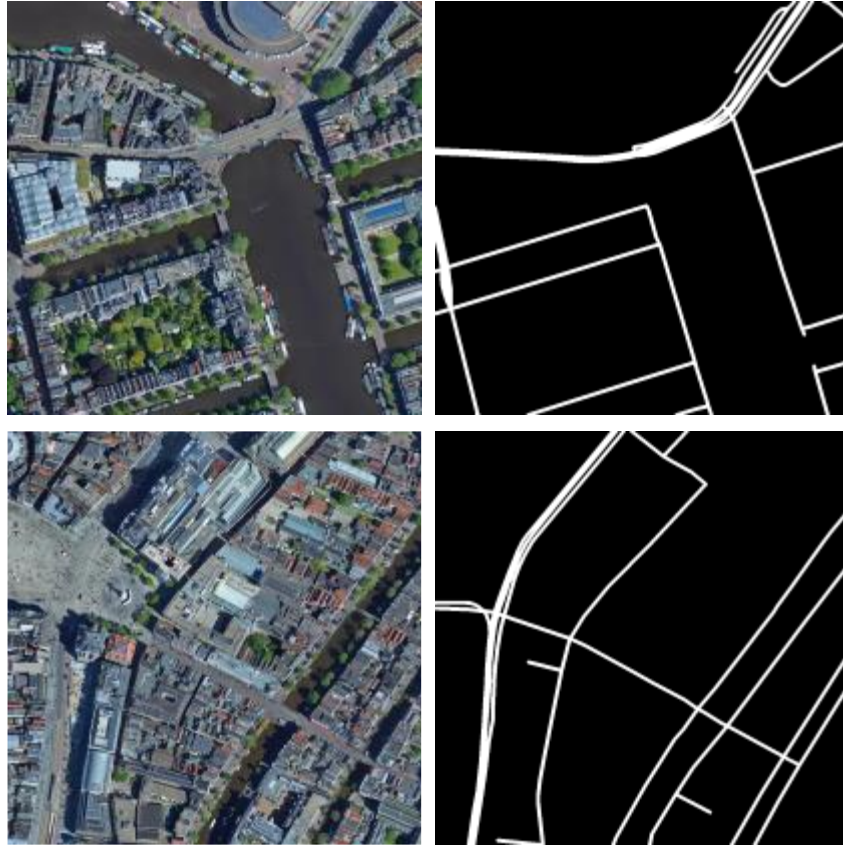


图 4.2 浙江数据集

RoadTracer 数据集包含来自 Google Earth 的全球 40 个城市的卫星影像数据，分辨率为 0.6 米，影像尺寸为  $1024 \times 1024$  像素，共 3268 张影像。标签数据有两种格式，一种是来自于 OpenStreetMap 的矢量数据文件，一种是将矢量数据进行栅格化，膨胀成等宽的 8 像素宽的栅格数据。本文选择 25 个城市的数据作为训练集，其它 15 个城市作为测试集评估模型的精度及预测能力。图 4.3 为 RoadTracer 数据集示例图像。



Image

Ground Truth

图 4.3 RoadTracer 数据集

这些航空、卫星影像包含城市、乡村、山区、沿湖沿海地区，覆盖地区范围大且场景类型多，构成巨大的样本空间，在这些数据集上进行训练与测试，能够科学地保证结果具有更好的泛化性以及更高的可信度。

#### 4.1.2 软硬件环境配置

本文中的实验都是在 Ubuntu 18.04 LTS 64 位操作系统下进行的，采用的编程语言为 Python 和 Go 语言，深度学习框架是 TensorFlow 和 PyTorch。

硬件环境为使用华硕 Z10PE-D8 WS 主板，8GB 内存，NVIDIA GeForce GTX 1060 6GB 显存的显卡。

#### 4.1.3 评价指标

本文的实验结果评估采用两类指标，分别是基于混淆矩阵的像素级评价指标，以及基于道路拓扑信息的完整性和连接性指标。

首先，道路提取可以从二分类任务的角度出发，即道路目标为前景，其余非道路目标为背景。如表 4.1 所示为二元分类器的混淆矩阵，其中 True Positive (TP) 表示被正确预测为道路的像素数，True Negative (TN) 表示被正确预测为非道路的像素数，False Positive (FP) 表示非道路被误判为道路的像素数，False Negative (FN) 表示道路被误判为非道路的像素数。根据研究内容，本文选取精度 precision、召回率 recall、综合指标 F1 值、交并比 IoU (Intersection-over-Union) 作为像素级精度评定指标，其计算公式分别如式(4.1)、(4.2)、(4.3)、(4.4)。

表 4.1 二元分类器的混淆矩阵

		预测	
		Positive	Negative
实际	True	TP	TN
	False	FP	FN

$$\text{precision} = \frac{TP}{TP + FP} \quad (4.1)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (4.2)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4.3)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (4.4)$$

精度 precision 是正确预测为道路像素在所有预测为道路像素中所占百分比，召回率 recall 是正确预测为道路像素在所有实际真实为道路像素中所占百分比。综合指标 F1 值综合了 precision 和 recall，通过计算二者的加权调和平均来均衡它们出现矛盾的情况。交并比 IoU 是模型预测出的目标像素和实际中真实目标像素的交集与二者的并集之比，是语义分割和目标检测研究中常见的评价指标。

其次，考虑到道路这一特殊地物的拓扑结构信息，本文引入两个指标 Road Completeness 和 Road Connectivity 分别用于统计道路结果的完整性与连接性。

如式(4.5)， $L_{pred}$ 为预测出来的路的长度， $L_{truth}$ 是真实标签中路的长度，其中，路的预测结果与真实标签都是单像素的栅格数据。 $L_{pred}$ 是通过统计 $L_{truth}$ 在预测出来的道路路面范围内的长度。

$$\text{Road Completeness} = \frac{L_{pred}}{L_{truth}} \quad (4.5)$$

图 4.4 表示了计算 Road Completeness 的方式。其中黄色为预测出来的道路，蓝色为真实道路，经过细化之后为蓝色的细线，即 $L_{truth}$ ，在第三幅中，经过叠加，落在预测结果上的绿色的代表预测成功的长度，即 $L_{pred}$ ，红色的为缺失的长度。

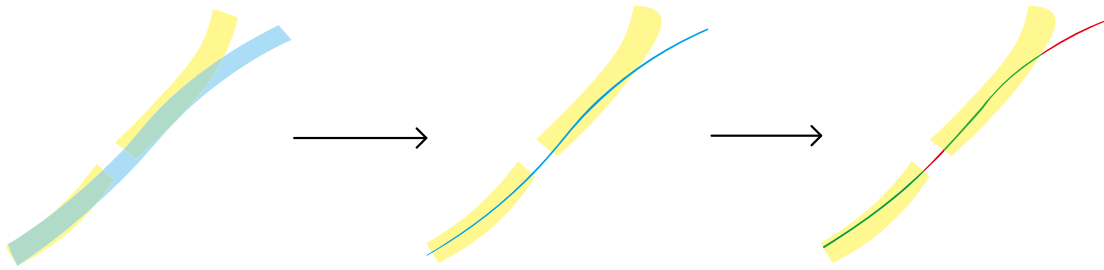


图 4.4 Road Completeness 统计方式

式子(4.6)表示 Road Connectivity 的计算方法。 $N_{connected}$ 代表预测结果中连续的道路片段的数量， $N_{total}$ 代表真实值中道路的片段的数量。统计的方法是将真实值的矢量数据分割成等长的片段，然后统计这些片段是否被完整预测出来，如果完整预测出来，则记为 connected，没有完全预测出来记为 unconnected。

$$\text{Road Connectivity} = \frac{N_{connected}}{N_{total}} \quad (4.6)$$

图 4.5 表示 Road Connectivity 的统计过程。黄色为预测出来的道路，蓝色和红色的细线为矢量格式的道路真实标签，由很多片段组成。红色的表示断裂的部分，即 unconnected，蓝色的为 connected。

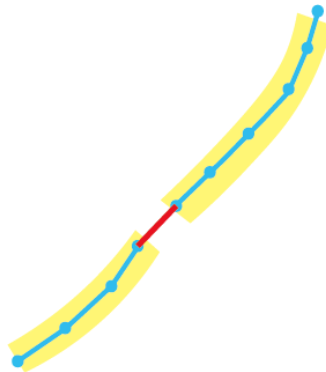


图 4.5 Road Connectivity 统计方式

## 4.2 实验结果

### 4.2.1 道路分割

将 Dlinknet 网络(D)和 Dlinknet 与 RoadNet 结合的方法在浙江数据集的 372 张  $1024 \times 1024$  训练集中进行训练, 然后在 160 张  $1024 \times 1024$  大小的测试影像上进行预测, 最后与真值进行对比评估, 分别从像素角度和拓扑结构角度计算评估结果。其中, Dlinknet 与 RoadNet 结合的策略有两种, 分别为基于缓冲区的融合方式(D+R+B)和基于加权的融合方式(D+R+W), 两种融合方式也进行了评估, 定量结果如表 4.2。

表 4.2 三种方法在浙江数据集上统计结果

	Precision	Recall	F1score	IoU	completeness	Connectivity
D	0.76	0.55	0.58	0.46	0.59	0.31
D+R+B	0.73	0.64	0.65	0.51	0.70	0.52
D+R+W	0.70	0.69	0.67	0.54	0.76	0.65

在马萨诸塞州数据集中, 将三种方法在 192 张  $1024 \times 1024$  的训练集上进行训练, 然后在 64 张  $1024 \times 1024$  的测试集上进行预测, 定量评估结果如表 4.3

表 4.3 三种方法在马萨诸塞州的数据集统计结果

	Precision	Recall	F1score	IoU	completeness	Connectivity
D	0.7	0.84	0.76	0.62	0.87	0.68
D+R+B	0.74	0.83	0.78	0.64	0.88	0.73
D+R+W	0.81	0.79	0.79	0.67	0.88	0.75

由表 4.2 和表 4.3 可知, 相对于单独使用分割的方式提取道路, 采用与中心线结果融合的方式能够获得更高质量的道路结果, 采用缓冲区融合以及采用加权方式融合的策略获得的结果在 IoU 上分别比单纯分割的结果提升了 2%和 5%, 在 F1score 上分别提高了 2%和 3%。另外, 在道路完整性以及道路连接性上面, 使用了融合策略的方法结果明显比单纯分割结果好。其中, 使用加权方式融合策略的获得的道路结果比分割方式以及采用缓冲区方式获得的结果分别高 7%和 2%。

三种方法在两套数据的预测集上的预测结果如图 4.6, 其中 A、B 是浙江数据集的测试集上的定性结果, C、D 是在马萨诸塞州数据集的测试集上的定性结果。从左至右分别为影像, 真值, Dlinknet 的预测结果, 使用缓冲区方法融合 Dlinknet 和 RoadNet 的结果, 使用加权的方式融合 Dlinknet 和 RoadNet 的结果。从定性结果可以看出, 融合策略相对于 Dlinknet 分割能够提取更多的道路, 并且提取的结果具有更好的连接性。分析原因是因为 RoadNet 更加关注图像的细节, 将细节融合

到最终结果中，显著提升了分割结果的质量。在两种策略的对比中，加权的方式表现更好，因为缓冲区方式虽然少了很多的噪声，但是过于依赖中心线结果使得它舍弃掉很多分割出来的细节。而加权的方式使得细节信息被加入到最终结果，更加贴近真实值。

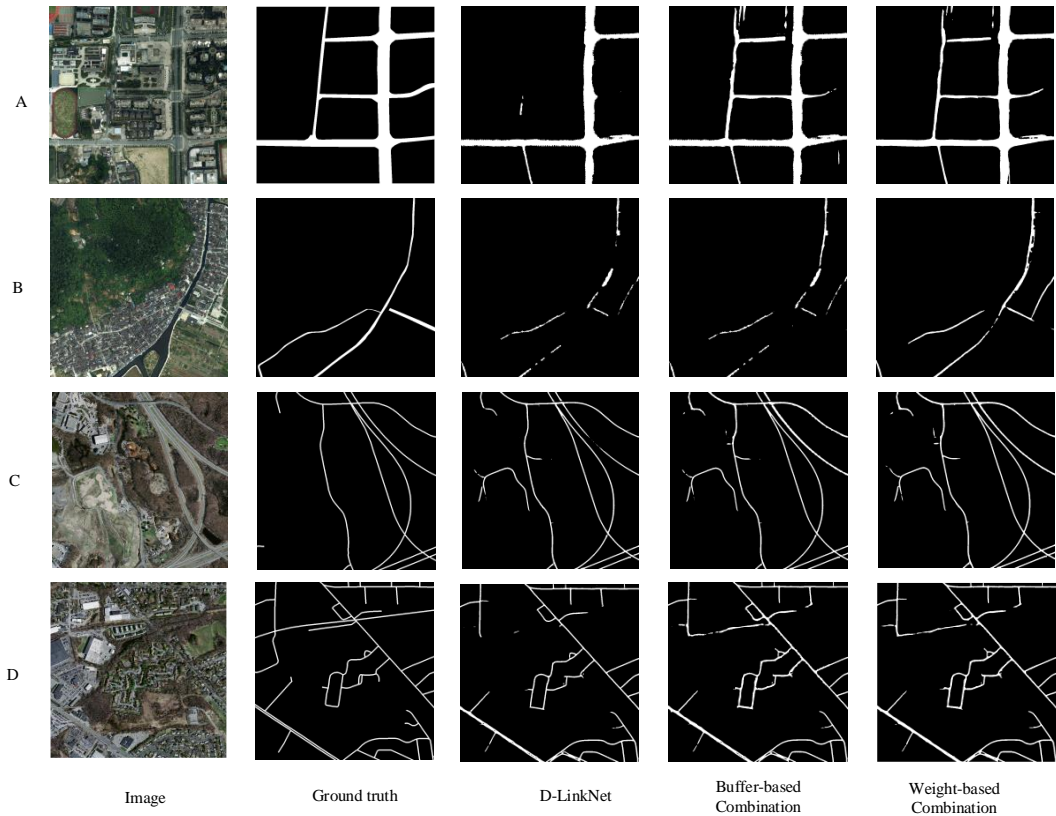


图 4.6 分割方法的定性结果

#### 4.2.2 道路中心线追踪

在 RoadTracer 数据集上进行中心线追踪算法的训练以及测试。选取其中的 25 个城市作为训练集，剩余 15 个城市作为测试集。在测试过程中，分别使用单起点的 RoadTracer 算法和经过改进的多起点的 RoadTracer 算法(RoadTracer-M)进行测试，定量结果如表 4.4。其中，F1score 和 IOU 的计算是将矢量结果栅格化之后膨胀 8 像素宽进行统计的。表 4.5 是在测试集中部分城市的结果。

表 4.4 道路中心线追踪结果

	F1score	IoU	Completeness
RoadTracer	0.2717	0.1725	0.26
RoadTracer-M	0.3733	0.2575	0.40

表 4.5 部分城市道路中心线 IoU 结果

	Chicago	Paris	Pittsburgh	Toronto
RoadTracer	0.15	0.15	0.07	0.26



RoadTracer-M	0.31	0.24	0.28	0.48
--------------	------	------	------	------

图 4.7 为道路中心线追踪方法在测试集中的定性结果，从左到右三列分别为影像、单起点追踪结果、多起点追踪结果。其中，蓝色的为真值，黄色的为预测出来的结果。从定性结果可以观察到，多起点追踪方法可以提取更多的道路，尤其在有很多桥梁以及高架桥的地区，多起点方法可以较好的跨越这些障碍物，而单起点往往被迫提前终止，受起始点影响较大。

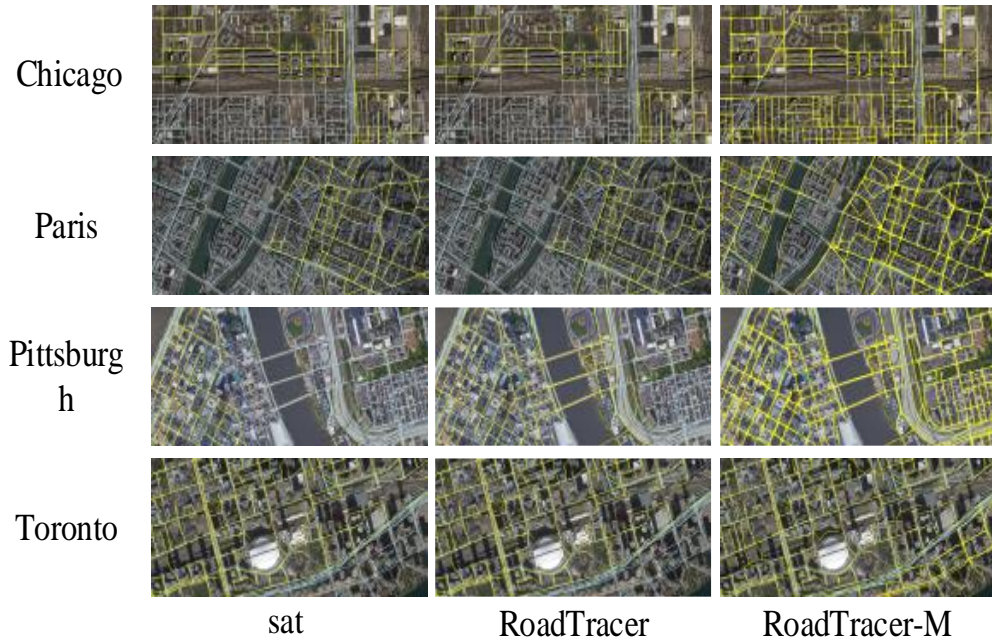


图 4.7 道路中心线追踪方法定性结果

#### 4.2.3 结合道路分割与中心线追踪方法

将分割方法（Dlinknet）和 Boosting 与中心线追踪结合的方法(D+B+G)在浙江数据集的 372 张 1024×1024 训练集中进行训练，然后在 160 张 1024×1024 大小的测试影像上进行预测，最后与真值进行对比评估，分别从像素角度和拓扑结构角度计算评估结果。定量结果如表 4.6。

表 4.6 结合策略在浙江数据集评估结果

	Precision	Recall	F1score	IoU	Completeness	Connectivity
D+B+G	0.76	0.76	0.75	0.62	0.87	0.80

在马萨诸塞州数据集中,将该方法在 192 张 1024\*1024 的训练集上进行训练，然后在 64 张 1024\*1024 的测试集上进行预测，定量评估结果如表 4.7。

表 4.7 结合策略在马萨诸塞州数据集评估结果

	Precision	Recall	F1score	IoU	Completeness	connectivity
D+B+G	0.69	0.89	0.78	0.74	0.90	0.89

将分割方法加 Boosting 策略方法及再与中心线追踪结合的方法在两套数据的预测集上的预测结果如图 4.8，其中 A、B 是浙江数据集的测试集上的定性结果，C、D 是在马萨诸塞州数据集的测试集上的定性结果。从左至右分别为影像，真值，Dlinknet 加 Boosting 再结合道路中心线融合策略预测结果。

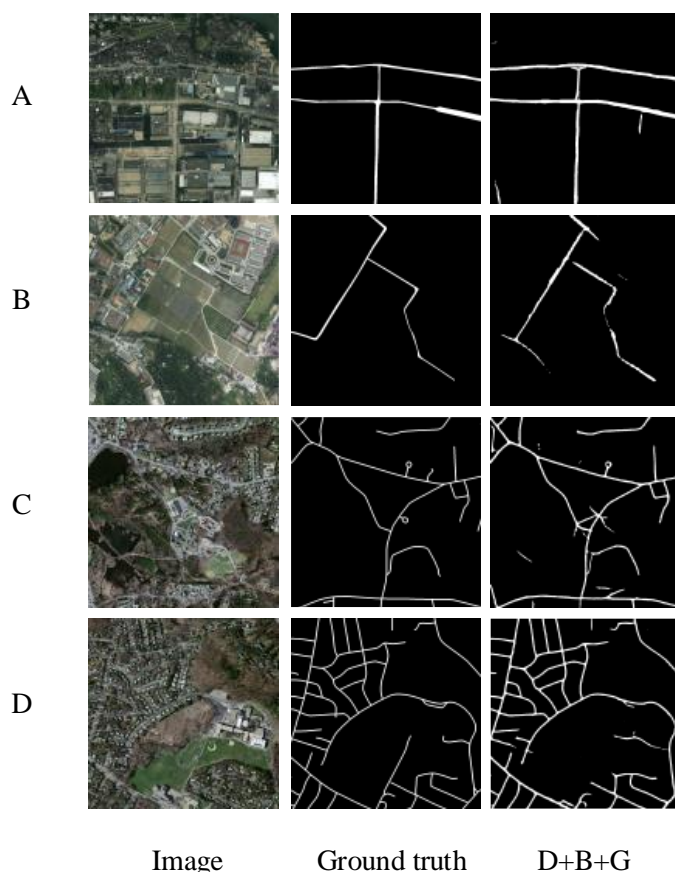


图 4.8 组合策略的定性结果

### 4.3 对比分析与讨论

本节将从分割方法和中心线提取方法两个方面对本设计中的方法进行对比与分析。所有方法都是基于 4.1.1 节说明的数据集，评价指标用到了 IoU，Road Completeness，Road Connectivity 三种指标。

#### 4.3.1 分割方法

分别将 Dlinknet 直接分割(D)，Dlinknet 和 RoadNet 结合的修补策略(D+R)，Dlinknet 和采用 Boosting 集成学习再与道路中心线结果进行融合 (D+R+G) 这三种方法在训练集上进行训练，然后在测试集上测试，利用本章介绍的评估方法进行精度评估。



表 4.8 在浙江数据集分割方法精度评估

	F1score	IoU	Completeness	Connectivity
D	0.58	0.46	0.59	0.31
D+R	0.65	0.51	0.76	0.65
D+B+G	0.75	0.62	0.87	0.80

表 4.9 在马萨诸塞州数据集分割方法精度评估

	F1score	IoU	Completeness	Connectivity
D	0.76	0.62	0.87	0.68
D+R	0.78	0.64	0.88	0.75
D+B+G	0.78	0.74	0.90	0.75

表 4.8 和表 4.9 分别是三种分割方法在浙江数据集和马萨诸塞州数据集上的评估结果。通过对比，Dlinknet 与 RoadNet 结合的修补网络以及分割追踪方法结合的策略在两个数据集上的测试效果都超过了原始分割的结果。其中，在 IoU 评价指标上，本文提出的分割追踪结合策略的提取结果在浙江数据集和马萨诸塞州数据集上分别比 Dlinknet 提升 17%和 12%，在道路完整性指标中，分别提升 28%和 3%，在道路连接性指标中，分别提升 28%和 7%。

三种方法在两套数据的预测集上的预测结果如图 4.9，其中 A、B 是浙江数据集的测试集上的定性结果，C、D 是在马萨诸塞州数据集的测试集上的定性结果。从左至右分别为影像，真值，Dlinknet 直接分割策略，Dlinknet 加 RoadNet 修补策略提取道路，Dlinknet、Boosting 及中心线追踪融合策略预测结果。从定性结果可以看出，融合策略相对于 Dlinknet 分割能够提取更多的道路，并且提取的结果具有更好的连接性。本文最后提出的分割追踪融合方法在道路的完整性以及连接性上都具有最好的表现。

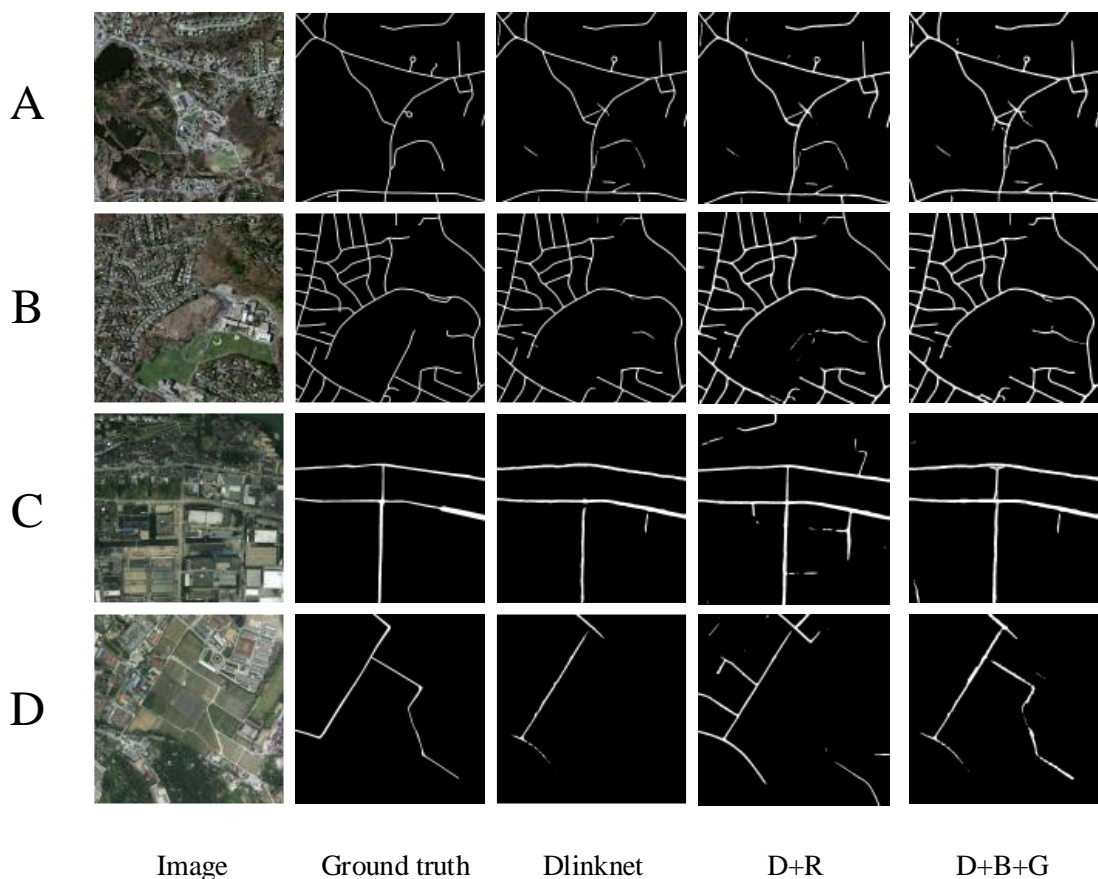


图 4.9 分割方法定性结果

#### 4.3.2 道路中心线追踪

分别使用 RoadTracer 单起点追踪方法，多起点追踪方法，以及利用分割中心线结合方法对测试集进行预测，统计评估结果。

表 4.10 浙江数据集中心线追踪方法结果评估

	F1score	IoU	Completeness	Connectivity
RoadTracer	0.38	0.26	0.18	0.11
RoadTracer-M	0.37	0.26	0.31	0.17
D+B+G	0.59	0.46	0.67	0.47

表 4.11 马萨诸塞州中心线追踪方法结果评估

	F1score	IoU	Completeness	Connectivity
RoadTracer	0.52	0.38	0.43	0.25
RoadTracer-M	0.52	0.38	0.49	0.31
D+B+G	0.79	0.70	0.86	0.75

表 4.10 和表 4.11 分别是三种中心线提取方法在浙江数据集和马萨诸塞州数据集上的评估结果。由表可知，分割与追踪结合方法在中心线提取上有着明显的优势，其中在浙江数据集上，经过等宽膨胀后的预测的中心线结果的 IoU 比 RoadTracer 和 RoadTracer-M 提升了 20%，在道路完整性上提升了 49%和 36%，在道路连接性

上提升了 36%和 30%。在马萨诸塞州数据集中，经过等宽膨胀后的预测的中心线结果的 IoU 比 RoadTracer 和 RoadTracer-M 提升了 32%，在道路完整性上提升了 43%和 37%，在道路连接性上提升了 50%和 44%。

三种方法在两套数据的预测集上的预测结果如图 4.10，其中 A、B 是浙江数据集的测试集上的定性结果，C、D 是在马萨诸塞州数据集的测试集上的定性结果。从左至右分别为 RoadTracer 单起点追踪方法，多起点追踪方法，以及利用分割追踪结合方法的测试结果。由定性结果可知，分割追踪结合方法获得的中心线明显在精度和范围上优于 RoadTracer 的中心线追踪结果。多起点追踪方法 RoadTracer-M 相对于单起点追踪方法 RoadTracer 能够搜索到更多的区域，但是其也受到起始点的位置的影响。另外，由于地面地形的多样性，其追踪过程很容易受到区域变化的影响而提前终止，而在结合策略中，分割方法则对环境的变化有着更好的适应性，所以取得了更好的结果。

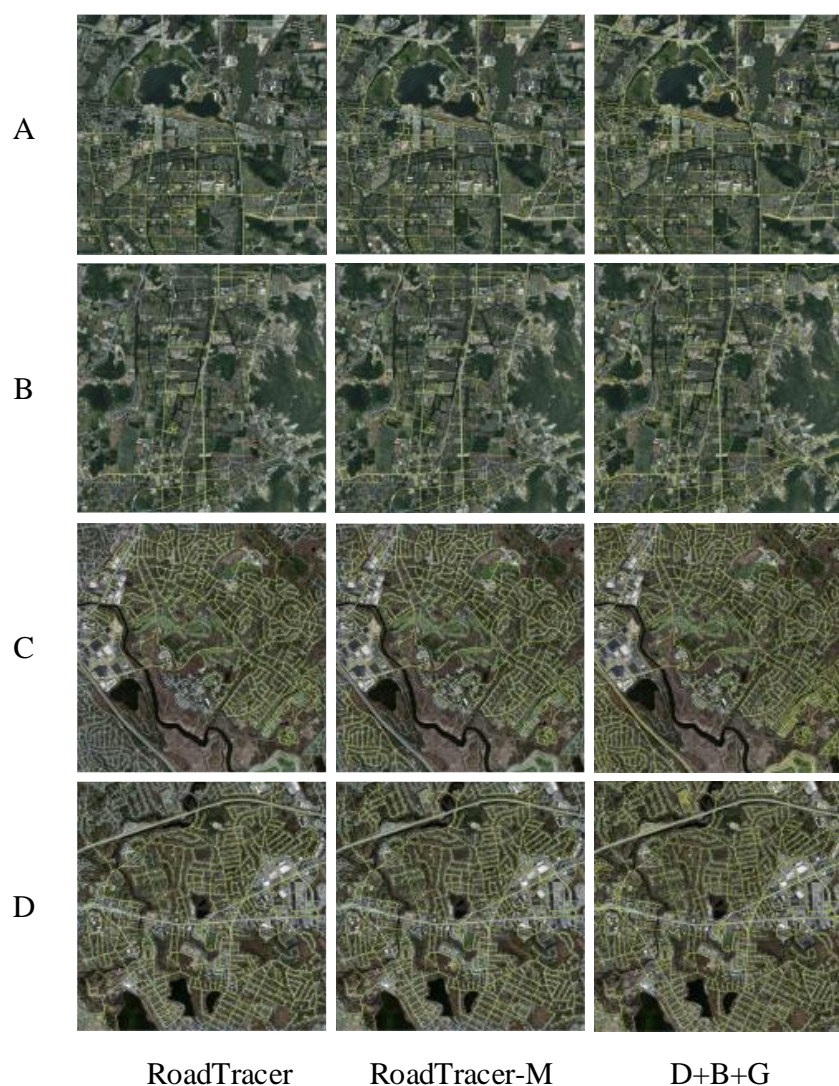


图 4.10 道路中心线追踪方法定性结果

#### 4.4 本章小结

本章主要展示了实验结果以及精度评定。以三套数据集为基础，分别将分割方法，中心线追踪方法，分割与追踪结合方法在数据集上训练并且测试，然后从 IoU、道路完整性和道路连接性三种评价指标评估结果。除了定量结果以及分析之外，还从定性方面直观地进行结果的对比，根据定性结果进行了方法的分析以及总结。最终结果表明，本文提出的改进策略以及结合方案是有效的，可以获得更高质量的道路网提取结果。

## 5 总结与展望

### 5.1 总结

本文基于深度学习中的卷积神经网络结构开展对遥感影像中道路自动化提取方法的研究。首先回顾了现有国内外关于遥感影像道路提取的研究，然后从道路路面分割以及道路中心线追踪两类方法开展试验。对一些现有方法进行了复现，然后根据现有方法的缺点，提出了改进方案，并且在数据集上进行了实验与对比。除此以外，基于分割方法与追踪方法的优缺点，本文也提出一种分割方法与追踪方法结合的遥感影像道路提取策略。通过该方法可以同时提取道路的路面分割栅格数据以及道路中心线，并且在道路完整性以及道路的连接性上有了很大的改进。实验结果表明，该方法比其他方法能获取更加高质量的道路网结果。

从应用上，本文提出的改进方法以及结合的策略为遥感影像道路提取提供了更优的解决方案，为自动化道路网数据提取的发展提供了一些帮助。此外，本设计中提出的多种道路数据融合的方式也为提取高精度，广覆盖的道路网络提供了建议与启发。

### 5.2 展望

以下是本文方法以及实验中的一些问题和下一步的工作计划。

第一，道路数据集质量欠佳。本文虽然采用了多个数据集来减少单一数据集对方法以及实验结果的影响，但是从实验结果中可以发现，采用的数据集依旧存在着诸多问题。如在浙江数据集，高等级道路数据存在真实标签，但是在一些非等级道路上不存在人工的标注，导致影像中的道路网稀疏，不能十分准确的评定实验结果，而在马萨诸塞州数据集中，道路标签为等宽的。因此寻找以及构建精确的道路网数据集是一件十分重要以及迫切的任务，采用质量更优的数据集进行实验能更加客观与准确地评估方法的表现。

第二，尝试多种数据融合。除了道路路面栅格数据以及中心线数据，还有其它的与道路相关的数据，如出租车的 GPS 轨迹数据。将多种数据进行融合，能够提

取出覆盖范围更加广泛的道路网，并且也能减少单一数据中的噪声导致的错误数据。

第三，尝试新方法。除了使用语义分割的方法进行道路像素级的分割，是否可以结合道路的特点，首先对道路进行对象提取，然后进行分割，从而减少噪声。另外，能否将中心线追踪以及道路路面分割同时进行，类似多任务网络的结构，进行互相约束，从而取得更好的结果。

第四，拓宽研究课题。从道路提取到遥感影像的其它地物提取，或者由车载影像出发，对道路及车道进行提取。在现有的策略上进行修改，使其适应不同的任务，经过多种道路提取任务的实验，从而提出一种通用的框架，为道路自动化提取提供方便。

## 参考文献

- [1] Wang J, Song J, Chen M, et al. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine[J]. International Journal of Remote Sensing, 2015, 36(12):26.
- [2] Saito S, Yamashita T, Aoki Y. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks[J]. Electronic Imaging, 2016, 60(1):10402-1/10402-9.
- [3] Sarhan E, Khalifa E, Nabil AM. Road extraction framework by using cellular neural network from remote sensing images[C]. In 2011 International Conference on Image Information Processing. IEEE, 2011 Nov 3 (pp. 1-5).
- [4] Zhong Z, Li J, Cui W, et al. Fully convolutional networks for building and road extraction: Preliminary results[C]// Geoscience & Remote Sensing Symposium. IEEE, 2016.
- [5] Cheng G, Wang Y, Xu S, et al. Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017:1-16.
- [6] Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network[J]. Remote Sensing, 2017, 498.
- [7] Teerapong P, Kulsawasd J, Siam L, et al. Road Segmentation of Remotely-Sensed Images Using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields[J]. Remote Sensing, 2017, 9(7):680-.
- [8] Bastani, Favyen, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, Sam Madden, and David DeWitt. RoadTracer: Automatic Extraction of Road Networks from Aerial Images[A]. In Computer Vision and Pattern Recognition[C]. 2018.
- [9] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[C]// ICLR. 2016.
- [10] Noh H, Hong S, Han B. Learning Deconvolution Network for Semantic Segmentation[J]. 2015.
- [11] Ethem Alpaydin. Introduction to Machine Learning[M]. 第二版, 中国, 机械工业出版社, 2014 年, 258-265
- [12] Zhou L, Zhang C, Ming W. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2018.
- [13] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2014, 115(3):211-252.
- [14] Otsu N. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1):62-66.
- [15] Shi J, Tomasi. Good features to track[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2002.
- [16] Volodymyr Mnih, Machine Learning for Aerial Image Labeling[D],

University of Toronto, 2013

[17] Carsten Keßler. OpenStreetMap[J]. 2017.

[18] Tian Z, He T, Shen C, et al. Decoders Matter for Semantic Segmentation: Data-Dependent Decoding Enables Flexible Feature Aggregation[J]. 2019.



## 相关科研成果

Yao Wei, **Kai Zhang**, Shunping Ji. ROAD NETWORK EXTRACTION FROM SATELLITE IMAGES USING CNN BASED SEGMENTATION AND TRACING. Accepted by IGARSS2019.

## 致谢

四年的本科生涯转瞬即逝，在武汉大学读书的生活即将接近尾声，十分幸运在武汉大学接受了四年的教育，它帮我找到了自己的将来规划，也让我从之前的青涩变得成熟。在四年时间中，受到了很多人的帮助和鼓励，正是这些让我在成长的道路上少了几分困难，在此，我衷心地向他们表示感谢。

首先感谢武汉大学，她为我提供了一个学习与交流的平台。在这里有很多优秀的老师和同学，他们热情善良，乐于助人，在我学习和生活中提供了很多帮助，帮助我确定了人生的方向。另外，丰富的学习资源让我开阔了眼界，对人生，对世界有了更深的认识和理解。

尤其感谢季老师在我的科研道路上的帮助。作为班主任，季老师在大二的时候就引导我学习做科研，十分细心的给我安排了一个入门科研的过程，让我逐步感受到科研的乐趣，也正是因为季老师的引导，让我决定将来继续在科研的道路上发展。在本次毕业设计中，季老师从选题到任务规划以及最终的毕业论文的撰写都提供了很多的建议和帮助，十分感谢季老师。

另外感谢魏瑶学姐在毕业设计研究中对我的建议和帮助，她在毕业论文撰写过程中给我提供了很多意见。

感谢父母在我大学四年中对我的关心和爱护，感谢所有的亲人朋友对我的鼓励和帮助，为我的成长提供了动力。

谨以此文献给所有给予我关心、支持、帮助的领导、老师、同事、同学以及亲友们。