

# 中山大学本科生 优秀毕业论文

基于多语义分割网络的高分遥感影像  
路网提取方法

学位申请人：周财霖

导师：李军 教授

专业：地理信息科学

学院：地理科学与规划学院

2021年6月

## 【摘要】

基于遥感影像的路网自动化提取在城市规划，交通建设和土地资源管理中具有重要意义。但传统的多光谱低空间分辨率影像在辨识精细类地物上不具有优势。随着遥感技术的发展，基于其空间分辨率高，地物细节信息丰富的特点，高分影像进行路网提取具有重要意义。得益于深度学习在计算机视觉领域的快速发展，通过深度学习方法进行高分影像的地物解译和分类逐渐成为相关研究的重点方向。但把自然场景下的语义分割网络应用到高分遥感的路网提取任务中的相关研究尚未得到充分的探索和挖掘。本文基于公开的道路影像数据集进行多个语义分割网络在路网提取任务上的性能对比。随后对比验证了不同样本均衡策略下的损失函数组合对于模型训练精度的影响。最后通过集成学习策略设计了一种多网络融合的路网提取模型，在结果精度上位于现有的路网提取模型前列。

**【关键词】** 高分遥感影像；路网提取；语义分割网络；集成学习

## **[ABSTRACT]**

The automatic extraction of road network based on remote sensing images has played an important role in urban planning, traffic manufacturing, and land management. However, the traditional multi-spectral image with low spatial resolution has disadvantages in distinguish tiny objects. With the development of high resolution remote sensing, having the advantages of high spatial resolution and many object details, high resolution images is in great significance of road network extraction. Thanks to the rapid development of deep learning in computer vision, utilizing deep learning technology to interpret and classify land objects on remote sensing images has been an important research. But related researches which use semantic segmentation networks on remote sensing road extraction have not been explored thorough yet. This paper compares the efficiency and precision of different semantic segmentation networks when working on road extraction tasks, based on public road images datasets. Then this paper compares the effect on the training of neural networks, of different loss function s under respective class balancing strategies. Finally, according to the theory of ensemble learning, this paper devised an road extraction model, which combines several semantic networks. The accuracy of this model list ranks high among existed high resolution remote sensing images road extraction models.

**[Keywords]** High resolution remote sensing image, Road network extraction, Semantic segmentation network, Ensemble learning

# 目 录

<b>第一章 绪论</b>	<b>1</b>
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.3 研究内容与贡献	3
1.4 技术路线	4
1.5 论文的结构安排	4
<b>第二章 基于语义分割网络的路网提取研究</b>	<b>6</b>
2.1 语义分割网络	6
2.2 骨干网络	7
2.3 损失函数	7
2.4 评价指标	8
<b>第三章 基于 OCRNet 的路网提取方法</b>	<b>11</b>
3.1 OCRNet 的网络结构	11
3.2 实验细节	14
3.2.1 数据集构建	14
3.2.2 数据预处理	15
3.2.3 训练环境与参数	18
3.3 结果分析	20
3.3.1 精度指标分析	20
3.3.2 可视化结果分析	21
<b>第四章 多模型融合的路网提取方法</b>	<b>27</b>
4.1 基于集成学习的模型融合	27
4.1.1 训练策略	27
4.1.2 结果融合	28
4.1.3 模型相对差异指数	29
4.2 基于损失函数优化的类别均衡	31

4.2.1 动态类别加权-----	31
4.2.2 在线难例挖掘-----	32
4.2.3 Lovász 损失函数-----	32
4.3 实验细节-----	33
4.3.1 数据集构建与预处理-----	33
4.3.2 实验设计-----	34
4.4 结果分析-----	35
4.4.1 模型优化对比-----	35
4.4.2 模型融合对比-----	36
<b>第五章 总结与展望-----</b>	<b>40</b>
5.1 工作总结-----	40
5.2 研究展望-----	40
参考文献-----	42
致谢-----	44

# 第一章 绪论

## 1.1 研究背景与意义

路网提取是遥感工作中的重要内容，对于城市空间规划，区域道路建设以及智能交通系统的建立具有重要意义。传统的路网提取往往需要人工在地图或高空影像上进行目视标注，不仅耗时耗力，而且难以做到实时的更新。随着高分辨率遥感影像的发展，基于地物分类的方法对遥感影像进行大尺度自动化的路网提取方法不断涌现。但道路的检测与其他类型地物往往更具难度，主要是由于：（1）道路的宽度往往受限，几何上具有细长但连通的特点，对于遥感影像的空间分辨率和提取方法精度要求更高；（2）道路作为公共开放区域，具有较多的干扰物和遮挡，例如行道树，建筑物阴影以及车辆等；（3）在影像中道路往往是像元数较少的地物类别，在数据集中常存在样本不均衡问题。综上，采用传统的地物分类或形态学方法进行路网提取效果不佳。

随着机器学习在遥感研究领域的发展，学者们开始采取基于深度学习方法进行遥感影像中的道路提取，尤其是语义分割领域中的神经网络。语义分割是计算机视觉领域重要的工作之一，旨在以像素级的精确度对不同类别的实体进行检测和标记。针对光学遥感图像的语义分割和地物分类是遥感研究工作中的关键和基础的任务，对于利用遥感图像来进行城市研究，农业监测，环境勘察，国土利用等具体的遥感实践具有重要的意义。在近年的遥感领域中，除了传统的地物分类算法外，更多学者开始关注自然场景下的语义分割方法，实现对遥感图像中地物的有效识别。

主流的语义分割模型应用在遥感图像上往往效果不佳，主要的原因有：（1）遥感图像相对于自然场景图像分辨率较低，多光谱影像分辨率仅在 30 米水平附近，全色卫星影像最高在 10 米至 1 米，且遥感数据集获取困难，缺乏标注数据；（2）在高空影像中对于地面的精细地物识别区分较困难，例如道路与建筑物，草地与耕地等空间上邻近且外观上相似的地物类别，并且存在较多的混合像元；（3）主流的语义分割模型主要关注其在自然场景下的应用，模型的训练，验证和应用均使用自然场景数据集，因此将其应用在遥感图像上是对其模型泛化能力的考验。

本文旨在将前沿的语义分割方法引入到遥感领域中，通过对相同遥感数据集

的训练，验证主流语义分割模型在遥感影像上的适用性，并通过横向对比掌握不同模型的性能差异。针对遥感图像的分类精度较低的问题，本文将探索多模型融合的可能性，包括前期数据集的多模型训练策略，以及后期的模型融合方法，从而进一步的提高地物识别和分割的精度。

## 1.2 国内外研究现状

卫星影像中的路网提取在近二三十年来一直是遥感学科的研究热点，也是计算机视觉，信息科学，几何数学等各学科的重要交叉领域。传统的道路识别方法主要有二：一类主要依靠道路的几何结构，基于形态学等理论方法进行提取<sup>[1]</sup>。Laptev I<sup>[2]</sup>等人基于形态学约束的 Snakes 模型进行道路检测。Gruen A<sup>[3]</sup>等人通过小波变换进行图像锐化后，利用快速金字塔算法进行道路识别。Maurya R<sup>[4]</sup>等采用 K 均值分类进行预提取后，进行一系列的形态学运算以保证路网提取结果的连通性和完整性。然而这类依靠几何结构的方法会受到许多噪声的影响，包括道路中的各种干扰物（树木，阴影）等。

另一类是主要依靠道路的多光谱特征，基于遥感手段进行地物的识别和分类。但这类方法对于影像的光谱分辨率和空间分辨率均有较高要求，在常用的遥感卫星影像上很难达到精细提取的水平。此外，还有部分学者采用其他数据辅助进行路网的提取，包括 DEM 数据，导航地图，道路起始点等<sup>[5]</sup>，以此为道路的识别提供更多参考信息，但适用性不高，常仅作为路网提取的精度验证或后处理手段。

基于深度学习的道路提取方法最早由 Boggess J<sup>[6]</sup>于 1993 年提出，通过对 Landsat TM 影像上的实验证明了神经网络在道路提取上的优越性，并通过对比实验证实：目标像元的上下文信息作为网络输入具有非常重要得意义。从此上下文信息的获取成为道路识别研究中学者们重点关注的环节，也是本文引进相关神经网络的重要动机。进入 21 世纪后，随着 QuickBird 等高分卫星的出现，学者开始关注高分遥感影像中的道路提取，以解决道路在传统遥感影像中像元较少，几何结构细碎，难以辨别的劣势。Mokhtarzade M<sup>[7]</sup>等首次将人工神经网络(ANN)引入到 QuickBird 和 IKONOS 高分三通道影像中进行路网提取，证实深度学习在高分影像上的精度提高效果，并通过对比实验同样强调了上下文信息的重要性。然而这些工作获取上下文信息的方式仅靠对原图进行卷积后输入神经网络，信息受卷积核大小的限制明显。针对以上局限，Mnih V<sup>[8]</sup>等人采用受限的玻尔兹曼机 RBM

对图像进行分割，并通过主成分分析 PCA 进行降维，从而实现在获取尽量多的上下文信息的同时减少运算量。

随着语义分割在计算机视觉下的发展，学者们开始发掘语义分割网络在遥感影像中的潜力，并应用到路网提取中。Zhong 等人<sup>[9]</sup>首次将 FCN 网络应用到高分遥感影像的道路和建成区提取。尽管语义分割网络的引入推进了深度学习在遥感领域的应用发展，然而由于遥感影像的复杂性和道路提取任务的难度，早期的语义分割网络在遥感影像上的直接应用精度均不太理想，主要原因有三：其一是由于遥感数据的获取和标注苦难，导致当时公开的遥感数据集影像数较少，模型的训练集明显不足；其次是语义分割的网络设计较为简单，应用在不同场景下的泛化能力较弱；最后是早期的语义分割网络没有专门对上下文信息的获取进行设计，然而上下文信息是遥感地物分类的关键要素。

### 1.3 研究内容与贡献

OCRNet 作为目前最前沿的语义分割模型，其优秀性能已经在自然场景下得到验证。但在遥感卫星图像上尚未得到使用。本文将首次把 OCRNet 应用在遥感图像数据集上，验证其在遥感语义分割上的优异性，并在同一数据集下横向对比其他语义分割模型。从而了解目前各主流深度学习网络框架在高分遥感影像中进行路网提取任务的性能差异，为后续的遥感研究与模型选择提供参考。

高分遥感道路数据集往往存在明显的样本类别不均衡问题。本文将首次引入计算机视觉领域中主流样本均衡方法，通过损失函数优化的形式，来解决遥感道路数据集中的类别不均衡。通过对比实验的方式，探索获得精度最佳的损失函数组合策略。

为进一步提高路网提取的精度，本文将会采用集成学习的策略，将性能较好且结果差异程度高的语义分割网络进行融合。采用堆栈法作为主要的集成学习策略进行基模型的构建。通过对比实验探索不同分类器作为终端模型时的精度表现，从而提出一个完整的高精度高分遥感路网提取模型，探索集成学习在路网提取任务中的优越性。



1.4 技术路线

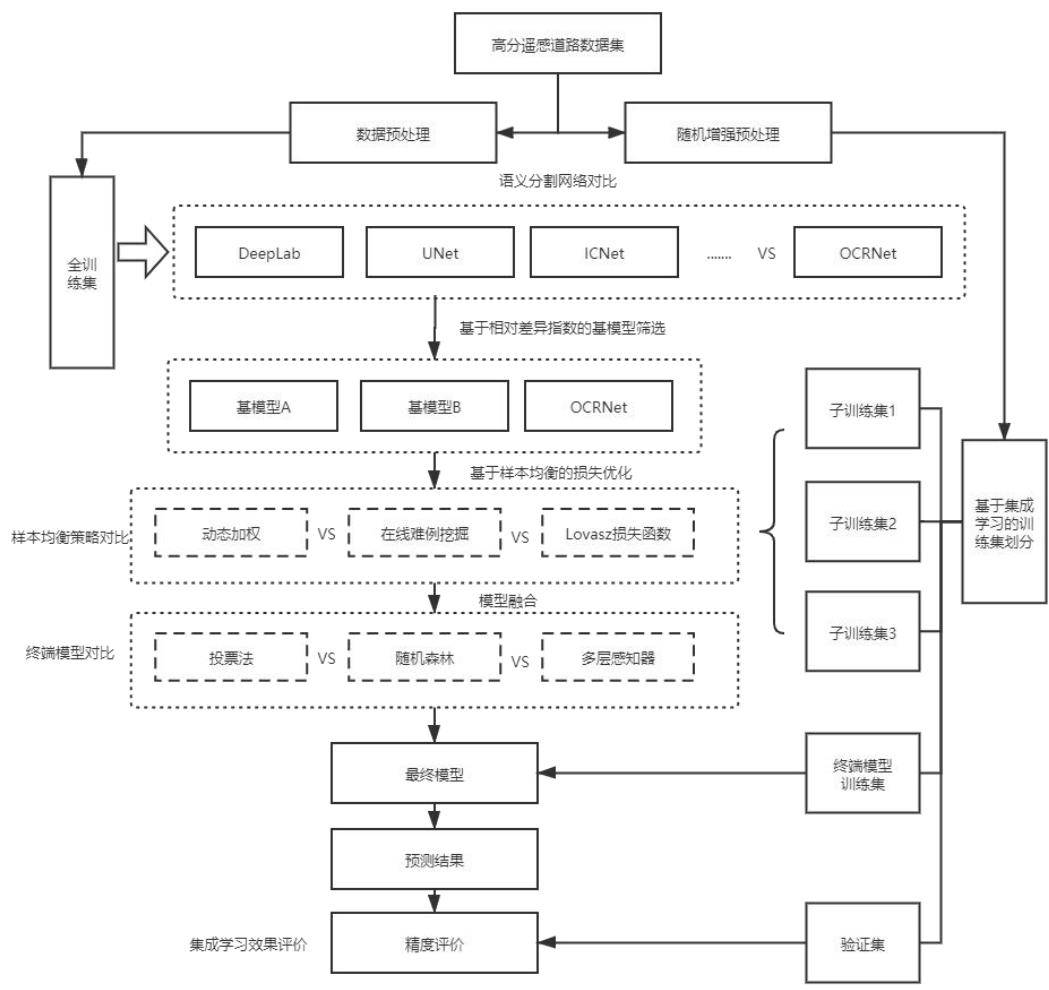


图 1.1 本文技术路线示意图

1.5 论文的结构安排

第一章为绪论，主要对遥感影像中的道路提取研究进行综述，对研究发展的现状以及存在的问题进行介绍。同时阐述本文的主要内容和创新点，以及主要结构和章节安排。

第二章为语义分割网络模型分析，主要对主流的语义分割网络进行介绍，并阐述他们的结构设计和性能差异，以及不同的应用场景和优劣势。此外介绍了在语义分割中常用的特征提取器，损失函数和评价指标，为接下来的实验部分提供理论基础。

第三章为基于 OCRNet 的路网提取研究，在实验中通过对多个模型在控制变量

和参数下的训练，将 OCRNet 与主流的语义分割网络在路网提取上的性能和精度进行对比，从而验证 OCRNet 在高分遥感影像路网提取任务中的适用性和优越性。

第四章为基于多模型融合的路网提取研究。为了解决模型融合带来的样本不均衡问题，首先对基于损失函数优化的样本均衡方法效果进行对比实验。之后基于集成学习策略对多个基模型进行融合，并对比结合不同的终端模型下的模型精度差异，最终提出了表现最优秀的多网络集成的路网提取模型。

第五章为总结与展望，对本文所完成的工作和创新点进行总结，并阐述本文研究的不足之处，以及对未来相关领域研究的展望和期待。

## 第二章 基于语义分割网络的路网提取研究

### 2.1 语义分割网络

传统的语义分割方法是基于浅层特征采用统计和几何的方法，在精度上具有一定局限性。近年来研究者通过训练神经网络学习像元的高维特征来进行预测，成为语义分割领域的发展方向。

全卷积网络<sup>[10]</sup> (FCN) 借鉴了应用在图像分类任务上的 CNN 卷积神经网络，引入了能够变换数据尺寸的反卷积层，从而使单像素输入成为可能，实现像素级的分类任务。FCN 是后续众多语义分割模型的基础。但 FCN 由于重复最大池化和下采样，生成的分割结果细节丢失严重；并且由于像素级的特征的感受野不够，对多尺度和上下文信息融合上具有局限性。在语义分割任务中，环境信息对于提高分割精度有关键性意义，因此后续模型针对上下文和感受野问题提出了不同的解决方案。

文献<sup>[11]</sup>通过采用池化层中的最大相应位置的张量来进行上采样，从而避免 FCN 中上采样对图像质量的损耗。对称语义分割模型<sup>[12]</sup> (U-Net) 是近年最流行的语义分割神经网络之一，通过扩张解码器模块的容量的方式改进实现对 FCN 的架构升级。PSPNet<sup>[13]</sup>采用金字塔池化实现多个感受野下的特征提取，从而更好的检测微小的物体。DeepLab<sup>[14]</sup>结合了概率图模型，通过扩张卷积算法来扩展感受野，并使用完全连接条件随机场来获取细节信息。总体上，各分割网络基本都在采用经典神经网络的大体结构，通过模型局部的调整和优化来追求较大感受野和上下文信息，同时尽量保持局部的丰富语义特征。

以上的语义分割模型大多数均为编码器-解码器的结构：通过编码器进行像素特征的提取和下采样；之后通过解码器进行逐像素对应的上采样进行恢复，获得增强的特征；最后输送到预测模块中获取最终结果。但这种结构会存在感受野范围狭窄，上下文信息缺失的劣势。为了解决这一问题，文献提出了一种显式学习上下文信息的神经网络<sup>[15]</sup> (Object Contextual Representation Network, OCRNet)。

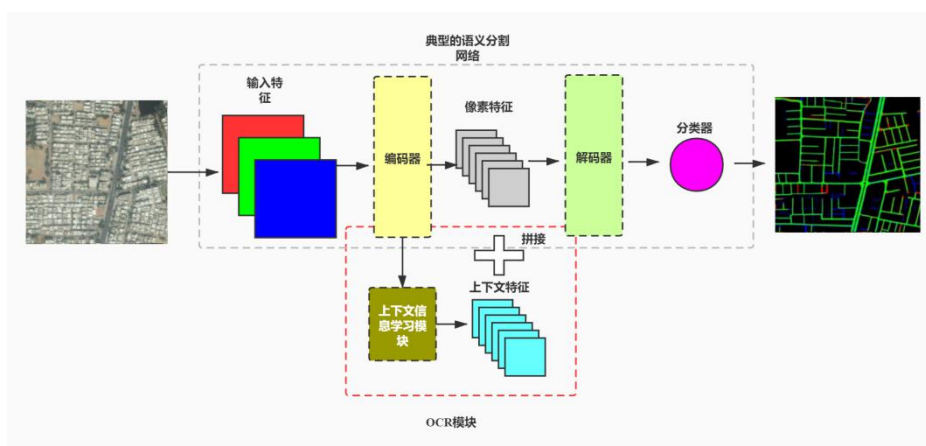


图 2.1 OCRNet 主要创新点示意

如上图所示，除了提取像素特征，OCRNet 还通过一个特别的并行模块像素提供增强的上下文特征，从而提高最终分类器的判别精度。下文将对 OCRNet 的工作原理进行详细介绍。

## 2.2 骨干网络

尽管可以直接原始图像输入到主模型中，但随着深度学习的发展，研究人员开始使用基础网络(Backbone)作为特征提取器，提取图像的高维特征信息来作为高级模型的输入。这种具有强表征力的深层特征能够更好的将不同类的像元进行区分，从而提高主模型的语义分割性能。

早期骨干网络，如 VGGNet<sup>[16]</sup>等神经网络的出现标志着深度学习进入全新的发展阶段，但存在梯度消失和网络退化问题，网络深度和精度因此受限。残差神经网络 ResNet<sup>[17]</sup>的出现使特征提取进入全新高度：通过基本残差学习和批归一化解决网络退化和梯度消失问题。高分辨率神经网络 HRNet<sup>[18]</sup>通过并行连接高分辨率与低分辨率网络，保证了多分辨率下的特征图不受下采样导致的分辨率降低的影响。针对大型神经网络无法应用在移动场景和嵌入式设备的问题，MobileNet<sup>[19]</sup>采用深度可分卷积来代替一般卷积，大幅减少参数和计算量。

## 2.3 损失函数

损失函数是在神经网络模型训练时用来测量预测结果与真实值差异的计算函数。当损失函数值越小时，说明模型在训练集中的预测性能越好。神经网络的训练过程中实际上就是反向传播的过程，而损失函数能在每次训练迭代中为反向传播调整参数提供方向和依据。损失函数的设计是基于深度学习研究中的重要环节，

一个优秀的损失函数不仅能加快模型收敛的速度，缩短训练的时间，还能防止模型过拟合，提高模型的预测精度和性能。

损失函数实质是计算预测和真实的差异，而交叉熵则是一种度量系统差异的常用工具，因此被引入到损失函数的设计中。交叉熵损失函数是语义分割任务中最常被使用的一种损失函数。二元交叉熵损失函数是一般交叉熵函数针对二元分类任务的特殊形式，具体计算公式如下：

$$Loss = \frac{1}{N} \sum_i Li = \frac{1}{N} \sum_i (-[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]) \quad (2.1)$$

式中， $N$ 为输入影像的像元数， $y_i$ 为预测像元 $i$ 的真实类别值， $p_i$ 为像元 $i$ 预测为正的的概率。概率一般由 Softmax 或 Sigmoid 函数将神经网络的线性输出进行映射。二元交叉熵损失函数在模型训练时的收敛速度上具有优势，但不利于解决样本类别不均衡下的训练情景。

$$Dice = \frac{2|Y_{gt} \cap Y_{pred}|}{|Y_{gt}| + |Y_{pred}|} \quad (2.2)$$

随着语义分割领域的发展，传统的准确率和召回率等指标已经无法满足预测结果评估的需要，因此 Dice 系数成为一种更加先进的指数，用于度量两个集合组成的相似程度的指数，具体计算方式如下：

式中， $Y_{gt}$ 为真实的目标类像元集合， $Y_{pred}$ 是模型预测的目标类像元集合。由于 Dice 系数只关注预测结果和真实影像中的目标像元，因此在计算数量较少类别时结果不会被众多的背景类淹没，在评估目标类精度时具有一定的优越性。文献为了解决医学影像中目标类像元数量过少的不均衡问题，将 Dice 系数引入到损失函数的设计中。Dice 损失函数的具体形式如下：

$$DiceLoss = 1 - Dice = 1 - \frac{2|Y_{gt} \cap Y_{pred}|}{|Y_{gt}| + |Y_{pred}|} \quad (2.3)$$

基于相关研究<sup>[20]</sup>的对比试验，Dice 损失函数在训练类别不均衡样本时对精度和收敛速度具有显著的优化作用。但在对类别不均衡程度较小的训练集时，使用 Dice 损失函数会对网络的反向传播产生负面影响，影响模型收敛和预测精度。

## 2.4 评价指标

评价指标是衡量模型在预测任务中的结果精度和性能差异。优秀的评价指标不仅能够对比模型时提供客观的量化依据，还能在模型训练过程中指导研究者

进行超参数的调整，或者表征模型训练的情况，以便在合适的时间停止训练，获得最理想的模型。

在语义分割任务中常用的评价指标为精确率，召回率，F1 值，准确度，以及交并比等。尽管不同的指标有不同的计算方法，但都离不开基本的分类混淆矩阵。对于路网提取这一二分类任务，所得到的预测结果和真实值影像均为二值图像。

表 2.1 混淆矩阵

		预测结果 Predict Result		
		0	1	合计
真实值 Ground Truth	0	True Negative (TN)	False Postive (FP)	TN+FP
	1	False Negative (FN)	True Postive (TP)	FN+TP
	合计	TN+FN	FP+TP	TN+FN+FP+TP (Sum)

精确率 (Precision Accuracy)是表示预测结果为 1 的正例中，真实值也为 1 的像元在正例像元中的比例。在路网提取任务中，精确率用以计算所提取的道路像元中确为道路的像元比例，来表征建立的模型在道路提取中的结果准确程度，计算公式如下：

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

召回率 (Recall Accuracy)是表示真实值为 1 的像元中，所得到的预测值也为 1 的像元在这部分像元中的比例。在路网提取任务中，召回率用以计算真实道路像元中被正确预测为道路的比率，来表征建立的模型在道路提取结果相对于真实影像，能够被成功召回并纳入到预测中的准确程度，计算公式如下：

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

精确率和召回率是分别从制图者和用户两个角度对模型性能进行评价，具有一定的局限性，因此更加综合全面的指标往往是语义分割和遥感地物分类中更加被常用的。F 值(F1 Score) 是精确率和召回率的调和平均值，用以对精确率和召回率进行更综合的评价，计算公式如下：

$$F1 = \frac{2 * P * R}{P + R} \quad (2.5)$$

像元准确度 (Pixel Accuracy)是另一个全面评估模型精度的指标，用以计算预

测正确的像元占有所有像元中的比例。在路网提取任务中，总体准确度表示模型正确预测为道路和正确预测为非道路的像元的比例，来表征建立的模型在道路提取中的精度，计算公式如下：

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.6)$$

交并比 (Intersection over Union, IoU)是目前语义分割中主流的评价指标，也被称为 **Jaccard Index**，常作为模型对比中的标准指数。其在遥感地物分类，目标检测等任务中同样具有重要意义。在本文的路网提取任务中，主要使用道路类交并比 **IoU**：真实道路像元集合与模型预测的道路像元集合，交集数量与并集数量的比例，计算公式如下：

$$Road IoU = \frac{TP}{TP + FP + FN} \quad (2.7)$$

以上指数均为计算机视觉领域中常用指数。而在遥感领域中常用作地物分类效果评价的指数是 **Kappa** 系数。计算公式如下：

$$Kappa = \frac{p_o + p_e}{1 - p_e} \quad (2.8)$$

式(2.8)中具体值为：

$$p_o = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

$$p_e = \frac{(TP + FP) * (FN + TP) + (TN + FN) * (TN + FP)}{Sum^2} \quad (2.10)$$

## 第三章 基于 OCRNet 的路网提取方法

### 3.1 OCRNet 的网络结构

OCRNet 是目前将上下文信息结合得最好的语义分割模型之一。精度水平在多个语义分割数据集榜单中占据前列位置。OCRNet 执行语义分割任务时的具体流程如下：

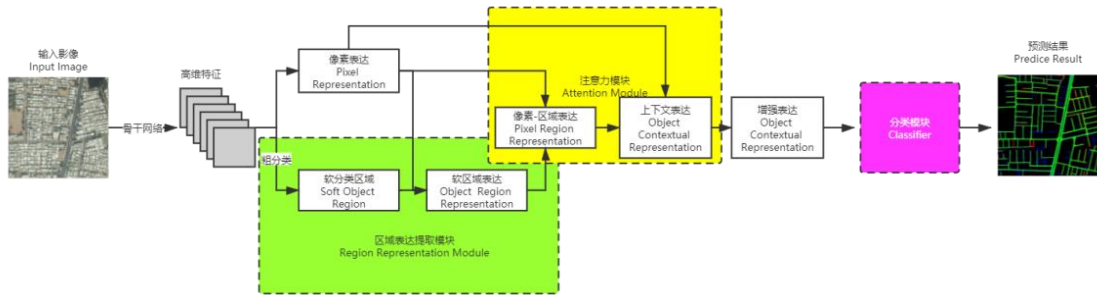


图 3.1 OCRNet 整体流程与架构

OCRNet 的核心思想在于为目标像元计算其从属类别区域的上下文特征 (OCR)，然后根据区域特征表示与像素特征表示之间的相似度来增强像元，使得像元的特征携带更多的环境信息，提高其类别划分的精度。而为了实现这一过程，模型还进行了其他的工作和计算，具体步骤可以划分为三步：区域的像素表示获取；目标的上下文表示获取；以及最终的特征增强与结果预测。下文将分步对模型结构和流程进行介绍。

#### （1）区域的像素表示获取

在基于骨干网络进行特征提取后，首先需要进行环境信息的整理，以此为接下来目标和区域的对比提高区域本身的特征依据。因为目标的像素表示可以通过直接的卷积获得，但区域目前是未定义状态。所以需要基于简单的神经网络对原影像进行粗分类，获得软对象区域。粗分类可以是多种形式的神经网络，但需要保证一定的结构简约性。



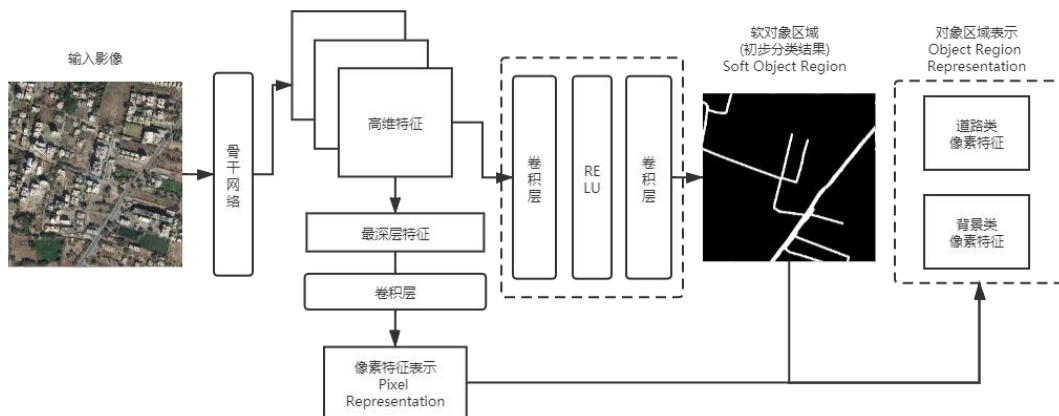


图 3.2 对象区域表示获取

如上图，在实际工作中高维特征可以是骨干网络所提取的特征的全部，也可以是部分，本实验中将采用全部特征进行初步分类。分类是通过一个简单的卷积神经网络架构，包括两个卷积层和一个激活层，由此得到软对象区域，即初步的语义分割结果。基于这一结果，以及像素特征表示，进行统计和合成，从而得到所有类别的像素表示。值得注意的是，像素特征表示是由骨干网络提取的最深层特征进行卷积得到。

## (2) 目标的上下文表示获取

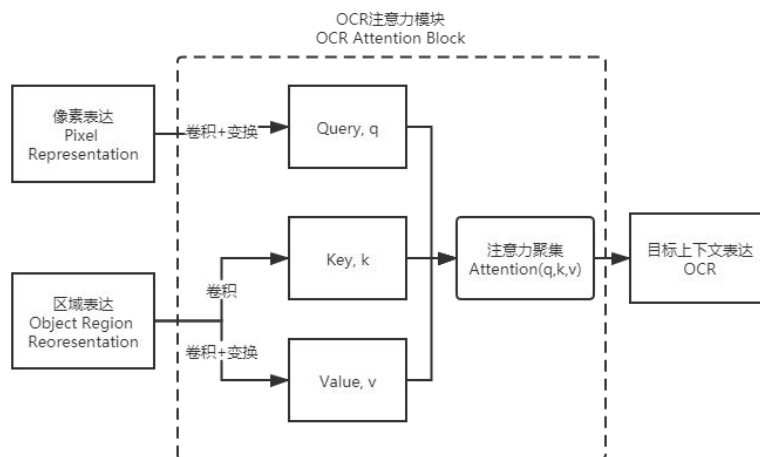


图 3.3 基于注意力机制的 OCR 获取

对于目标像元的上下文增强，实际上就是通过一定的相似度计算来寻找像元所属的区域，通过区域的一般特征来增强和改正目标像元的像素表示，从而使得目标的像素越来越趋近于从属的区域，而非从属区域差距增大。通过类似的增强，最终的分器能更好的区分不同类别的像元，提高语义分割的精度水平。实

际上这种注意力机制思想在具体操作上是集成且高效的：目标像素特征与区域像素特征的相似程度可以通过 Softmax 函数进行计算，通过相似程度的大小作为该区域对目标的权重，对目标进行增强。具体公式如下：

$$OCR = Attention(q, k, v) = softmax(\frac{qk^T}{\sqrt{d_k}}) \quad (3.1)$$

通过整合的注意力模块，模型能够进行并行的对多个目标像元进行多个区域的相似性计算和增强。将注意力机制引入到上下文表示的学习和增强是 OCRNet 在效率和性能上具有显著优越性的关键因素。

### （3）特征增强与结果预测

在获得了各目标像元的上下文表示后，即可基于此对特征进行增强。在 OCRNet 中所设计的增强方式非常简单：直接将原有的像素特征与获得的上下文特征进行拼接，从而拓展目标像元的特征维度，从而使其在训练神经网络和预测结果时能够吸收和学习上下文信息和本身的像素信息。

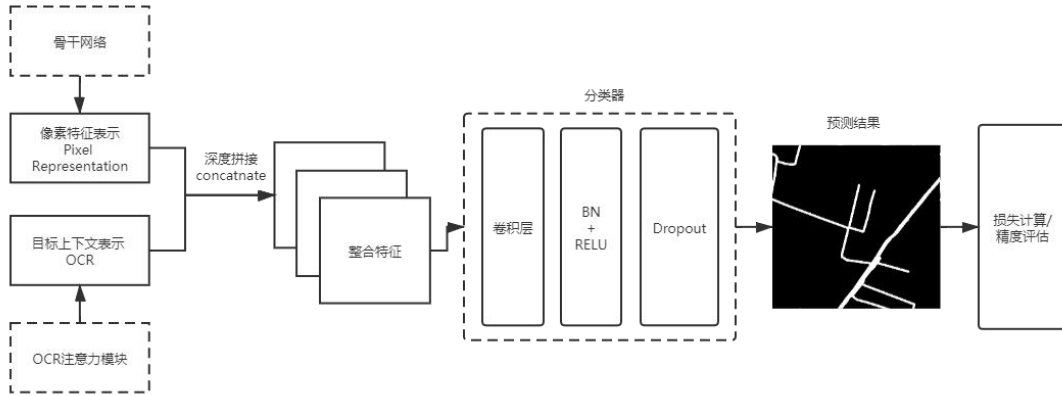


图 3.4 OCR 特征集成与结果预测

如上图所示，通过深度拼接后会获得整合后的特征，包含了目标本身的像素信息以及区域上下文信息。之后将特征图输入到分类器中进行训练，分类器是由一层卷积层，一层批归一化和 RELU 激活层，以及一层 Dropout 层构成。由于之前经过一系列的工作后获得的特征区分度大且易学习，因此最终的分器结构无需过于复杂即可保证结果能获得较高的精度水平。并且由于神经网络的结构简化，需要进行梯度优化和调整的参数减少，对于模型训练和预测的效率提升明显。接下来的实验将会验证这一点。

## 3.2 实验细节

### 3.2.1 数据集构建

语义分割作为计算机视觉和深度学习的重要领域，目前已存在较多丰富且权威的公开数据集，以供研究者们进行算法的训练和性能对比，较为著名的有 Cityscape, COCO Stuff, PASCAL 等。但这类数据集影像基本采集于自然场景下，与遥感研究所需要的数据特点差异较大，不能直接应用到遥感相关的研究中。而主要由遥感影像组成的公开的地物分类数据集一直较少，目前在学界中较常被使用的有 Gaofen Image Dataset, EVLab-SS 等。但这类数据集存在影像数较少，空间上连续集中等问题。并且由于这类地物多分类数据集非路网提取专门的数据集，在道路的标注上不够细致，且存在大量的未分类像元，会对模型的训练带来一定的干扰。

DeepGlobe 道路数据集由美国 DigitalGlobe 公司发布，是 2018 年全球卫星影像像挑战赛道道路提取分赛的标准数据集。数据集影像主要在亚洲南部国家拍摄，覆盖范围近 2 千平方千米，空间分辨率为 0.5 米。数据集公开发布的带有像素标注的标签图像有 6226 幅。标签影像中值等于 1 为道路像元，值等于 0 为非道路像元。部分影像如下图所示：

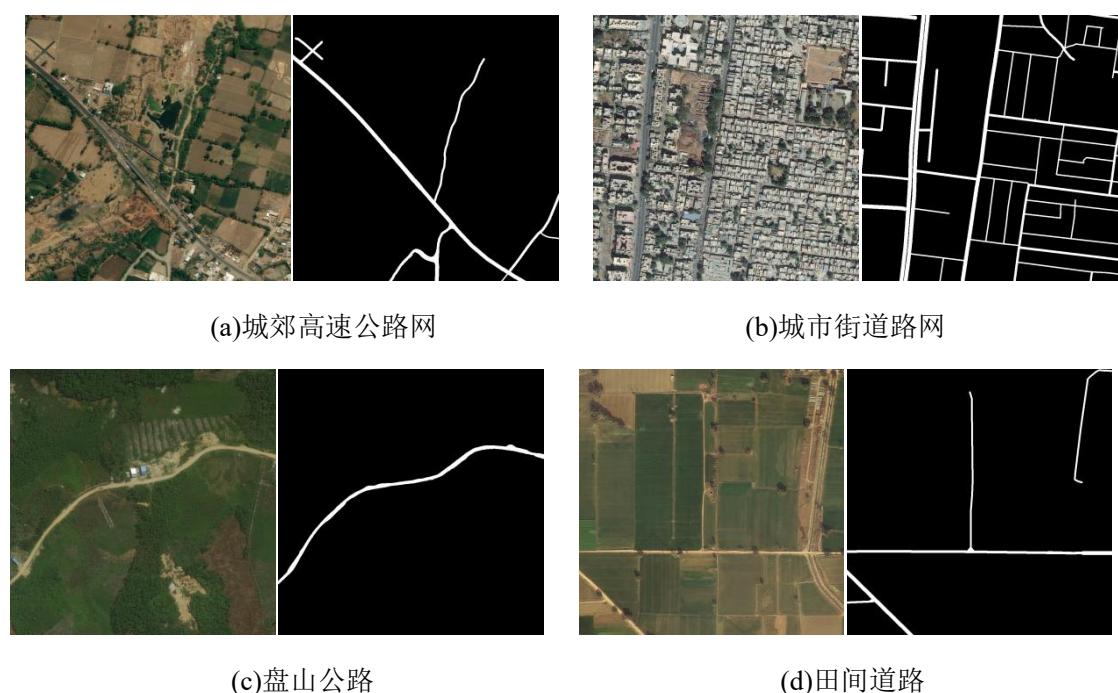


图 3.5 DeepGlobe 道路数据集

DeepGlobe 道路数据集具有覆盖范围广，影像分辨率高，数据量丰富的优势。高分影像中包括了各类道路和路网，这对于训练得到网络模型的鲁棒性和泛化能力具有重要意义。使其能够完成不同场景和要求下的道路识别任务。此外，DeepGlobe 道路数据集是目前高分遥感路网提取领域最主流的公开数据集之一，主要的路网提取模型均会基于此数据集进行精度评估和验证。因此使用这一流行的道路数据集将便于将本文实验获得的各类模型和策略进行直接的精度指标对比，从而更加客观有效的评价本文结果。本次实验将随机拆分 5604 幅作为训练影像，623 幅作为验证和测试集。

### 3.2.2 数据预处理

为了解决训练集影像数量不足的问题，防止过拟合情况的出现，在训练前对数据进行预处理是有效提高模型精度的必要方法。通过为每次迭代的相同影像进行随机的变换和增强，实际上在每次迭代中都是改变原始影像获得新影像的过程，从而能够隐式的扩增训练集，提高神经网络适应不同场景和数据源的能力。下文将介绍本次实验采用的预处理方法。

数据集由  $1024 \times 1024$  的 RGB 三波段影像组成。影像的尺寸过大会增大显存的占用，在显存大小一定的条件下，会迫使批尺寸对应的减少，从而导致同一批次输入进神经网络的影像数量减少，影响神经网络的训练时长。因而普遍的操作是对原始影像进行一定尺寸的裁剪后，再输送进模型训练。但影像的裁剪尺寸并不是越小越好，过小的训练影像会将核心地物与主体环境割裂，导致网络训练时无法获得足够的上下文信息参考，影响模型最终的预测性能。在本次实验中经测试后选取适中的尺寸： $256 \times 256$ 。

#### (1) 随机裁剪 (Random Crop)

上文阐述了在深度学习训练中影像裁剪的必要性。一般的裁剪方式有两种：随机裁剪和滑窗裁剪。滑窗裁剪是通过一定尺寸的窗口在原始影像上进行规律的滑动，以获得多张裁剪后的影像。滑窗裁剪的优势在于，即使设置的裁剪后影像尺寸较小，也能够保证影像中的所有像元都能作为训练集。这适用于地物样本分布不均衡的影像，能够保证模型训练时获得样本量较少或空间分布集中的类别样本。但劣势在于随机性不强，因为滑窗后的子影像分布相对固定，对同一原影像的多次滑窗裁剪后的结果都基本相同，因而处于边界附近的地物会受到割裂，影像训练时的上下文捕捉。



图 3.6 滑窗裁剪与随机裁剪

随机裁剪是对同一影像的每一迭代训练中，随机裁剪得到一张一定尺寸的子图。由于随机裁剪不受滑窗移动规律的限制，因而每一次对同一原影像的裁剪获得的结果都是不同的，这就能很好的为训练集增加随机性，提高模型的泛化能力。而且主要保证迭代的次数足够多，即使以前处于切割边界两侧的相邻的地物之间也总会有概率不受切割，从而还原像元的上下文信息，提高模型的预测性能。由于本次实验中道路地物在原始训练影像上的空间分布相对广泛且均匀，无需考虑样本无法获得的问题，因此采用对模型训练效果更佳的随机裁剪。

### （2）随机缩放 (Random Resize)

经过随机裁剪后获得的子图像尽管内容不同，但保持了尺寸的一致。然而如果模型接受的训练集尺寸大小一致时，则模型在预测非相同尺寸的影像时会出现效果不佳的问题。多尺度问题对于遥感影像更加重要，因为不同传感器下获得的影像空间分辨率往往不同，图幅尺寸也总是大小不一，因而随机缩放对于提升路网提取模型的鲁棒性具有重要意义。通过随机缩放预处理，每张训练影像在每一训练迭代次中会随机获得一个在一定范围内的变换系数，影像根据该系数进行等比例的放大或缩小。

### （3）随机旋转与翻转 (Random Rotate & Flip)

对原始训练影像进行随机的旋转是另一重要的数据增强和训练集扩充的方法。通过旋转，地物之间的绝对位置发生了改变，内部结构的分布位置也得到变换，从而为模型的训练提高更多样的训练信息。随机旋转对于遥感影像中的路网提取与自然场景相比更加关键，因为遥感影像在获取时往往携带地理信息，按照以上为正北的标准地图姿态。而真实分布的路网往往具有一定的方向趋势，例如按地势多以南北分布，这就会导致训练集中的路网分布多按一定方向分布，这不



利于模型解决多场景下的预测任务。因此对于遥感影像中的路网提取模型，对训练影像的随机旋转处理具有重要意义。

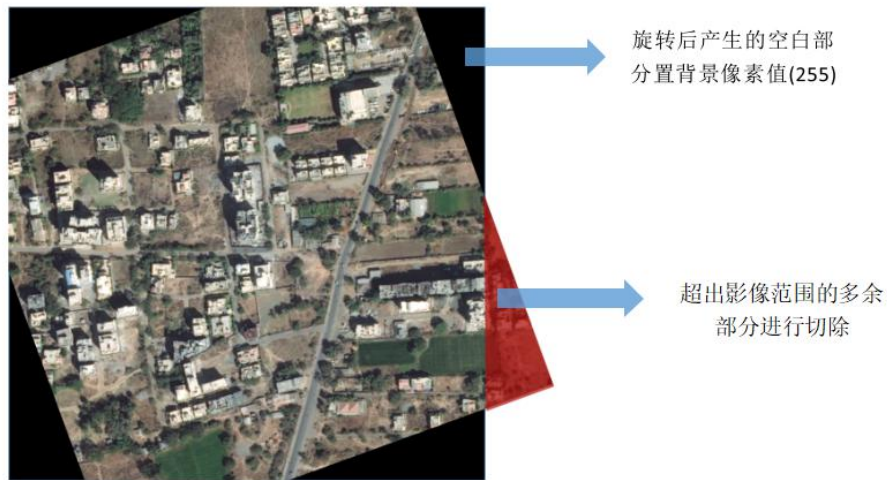


图 3.7 随机旋转示意

在处理过程中，单张影像在每一训练迭代中会得到随机的旋转角度值，图像在保证尺度和大小不变的情况下，按角度值进行顺时针方向旋转。旋转后超出图框范围的部分会被裁剪，而图框内未被旋转后图像覆盖的部分将被赋予特殊值，在网络训练过程中不进行处理。在本次实验中，旋转角度值的下限为 0 度，上限为 30 度。

随机翻转与旋转相似，对影像的尺寸和比例保持固定下进行镜像变换。包括随机水平翻转和随机垂直翻转，即对原始影像进行沿水平对称轴和垂直对称轴的镜像变换。在本次实验，在预处理阶段均增加的随机的水平和垂直翻转，即原始训练影像在进行随机旋转后，仍各有 50%的几率获得随机水平以及垂直翻转。

#### （4）随机畸变 (Random Distort)

以上的数据预处理方式都是对影像的整体几何变换，影像的像元值没有发生改变，因此需要标签影像和彩色影像进行同步的变换，保证像元之间的一一对应。而随机畸变则是单独对彩色影像的内部值进行增强，从而提高网络模型的泛化能力。随机畸变是一系列增强方法的统称，指对各种数字图像指标的随机变换。随机畸变包括：随机亮度增强，随机对比度增强，随机饱和度增强，随机色调因子增强。



图 3.8 随机增强效果

在本次实验的预处理阶段均设置以上随机畸变项，设置变换系数下限为 0，即不做任何变换，变换系数上限为 0.4。对每幅训练影像的每次训练迭代中，会获得 4 个在一定范围内的变换系数，据此对影像中的 4 个指标进行变换，从而实现训练影像的增强。

### 3.2.3 训练环境与参数

实验硬件平台 CPU 为 AMD Ryzen5 3600 3.6Ghz，GPU 为 Nvidia Geforce RTX 3060 12G，内存为 16G 配置，采用 PyTorch 和 PaddleSeg 深度学习框架实现神经网络的训练和语义分割实验。

实验将对 OCRNet，U-Net 等一众主流的语义分割模型进行同训练集的训练，并且需要保证训练环境和参数的一致，从而使各模型结果精度具有可比性，因此超参数在设定后在本次实验中不再改变。批尺寸(batch\_size)表示每次训练中同时送入神经网络中的训练影像数，数目过大会导致显存不足，数目过小会影响网络训

训练收敛的速度。根据之前的预训练的结果和其他类似实验，本次实验的批尺寸设置为 16，将在目前显存条件下保证梯度的准确性。学习率(learning rate)是影响神经网络训练性能和收敛速度的最重要因素。当学习率过大时会出现梯度在极小值处的来回摆动无法达到最优拟合点，而学习率过小则会影响训练过程的收敛速度。为了以上两种情况，本次实验设置学习率衰减，初始学习率为 0.01，之后按多项式形式进行衰减，衰减指数为 0.9。

优化器是神经网络模型在训练中如何调整参数以改变梯度，减小损失的策略。本次实验采用随机梯度下降策略(Stochastic Gradient Descent, SGD)。与传统的批梯度下降策略相比，SGD 优化器在每次训练迭代中仅采用一个样本进行梯度的更新，具有计算速度快的优势，但在 SGD 优化下的损失变化震荡较严重。为了避免采用 SGD 优化器常出现的问题：经过训练后，损失陷入局部最小值而非全局最小值，本次实验还引入了动量梯度下降法 (Momentum SGD)：使得每次参数更新的方向还受之前累积的动量影响，更容易走出损失低谷。动量系数设置为 0.9。

在上文中已提到过拟合问题在神经网络中的常见，因此在数据预处理中需要对训练影像添加随机增强。而在训练过程中，正则化则是另一重要的解决过拟合的方式。本次实验采用权重衰减策略(Weight Decay)，其等价于 L2 范数正则化，主要思路是为模型的损失函数添加 L2 惩罚项，以此减小各参数权重对于损失值的影响，从而减小模型的复杂程度。尽管进行正则化后，模型的训练集误差可能会增大，但对测试数据集的预测结果误差会减小，从而提高网络的泛化能力。本次实验设置的权重衰减参数为：4.0e-5。

为了控制变量，对实验中的语义分割网络训练：损失函数采用语义分割任务中常用的二元交叉熵损失函数；骨干网络均采用 ResNet-50。

对于神经网络的训练，训练时长和迭代次数如果设置过长，会导致模型的过拟合；而训练时长过短则会导致训练的不充分，精度提升受限。为了初步掌握网络的训练情况，设置适合迭代的次数，因此需要首先对 OCRNet 进行预训练。



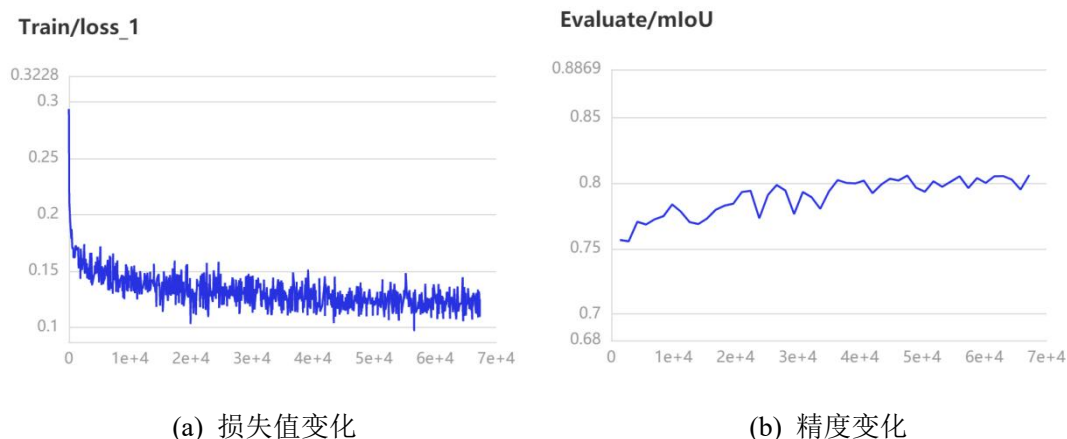


图 3.9 OCRNet 训练过程学习率变化

如上图所示，神经网络在训练初期收敛速度极快，对于精度的提升非常显著，在第一次断点评估时已经达到了较高的精度水平。但随着学习过程的不断推进，可以发现网络在 3 万到 4 万次迭代时精度已经到达最高水平。尽管损失函数仍在继续下降，后期精度无法提高，说明网络出现过拟合现象。因此接下来的实验中为了保证语义分割网络的精度最大，训练迭代次数设置为 3 万。为控制变量，所有语义分割网络均采用相同的设置和迭代次进行训练和精度评估。

### 3.3 结果分析

#### 3.3.1 精度指标分析

表 3.1 各语义分割网络路网提取精度

	Precision	Recall	F1	Road Acc.	Road IoU	Kappa
FCN	0.742	0.755	0.748	0.662	0.583	0.724
PSPNet	0.812	0.646	0.720	0.808	0.579	0.722
U2Net	0.775	0.671	0.719	0.762	0.580	0.723
DeeplabV3+	0.667	0.808	0.731	0.737	0.609	0.746
DNLNet	0.766	0.694	0.728	0.760	0.594	0.734
EMANet	0.740	0.823	0.779	0.667	0.540	0.687
OCRNet	0.739	0.848	0.790	0.741	0.665	0.789

上表列出了主要语义分割网络在本数据集下的各预测结果和精度评价指标。相对于目前主流的分割网络，OCRNet 在部分精度指标上均获得显著的提升。其中召回率相对于表现最好的 DeeplabV3+模型提升 4.95%，F1 值相对于水平最高的 FCN

模型提升 7.28%，道路类 IoU 和 Kappa 值相较于相较于最优值提升 7.39%和 4.69%。尽管部分指标 OCRNet 并不是最优的，如道路类准确度，精确度，但对于模型的精度全面性而言具有明显的优越性，各指标均在模型的前列。尤其是对于语义分割性能评价较综合和重要的指标：F1 值和道路 IoU，OCRNet 均具有明显的优势。

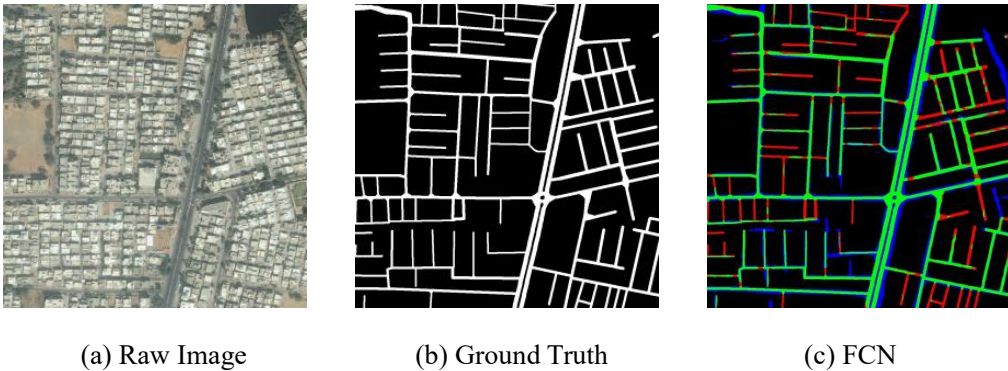
表 3.2 各语义分割网络路网提取效率

	预测总时长	单影像平均预测时	一百万像元平均预测	平均步时长
	second	长 ms/pic	时长 ms/1m. pixel	ms/step
FCN	224	360.12	343.44	361
PSPNet	282	453.37	432.37	245
U2Net	237	381.02	363.37	381
DeeplabV3+	215	345.65	329.64	291
DNLNet	252	405.14	386.37	405
EMANet	331	532.15	507.50	533
OCRNet	290	466.23	444.63	467

如上表所示，本文同时考察了在一定环境下不同模型的预测效率差异。尽管由于 OCRNet 由于自身网络结构的复杂性，在使用模型预测时耗费的时间均高于其他模型，但时间的差异非常小。对于验证集中大小为  $1024 \times 1024$  的共 622 张影像，总预测时长仅比网络结构最简单的 FCN 模型多耗费 29.5%的时间，平均每幅影像增加 12.3%的结果预测时间。因此在预测效率上 OCRNet 仍保持主流语义分割模型的水平，因此将其应用在实际路网提取任务中具有可行性。

### 3.3.2 可视化结果分析

#### (1) 城市路网提取结果



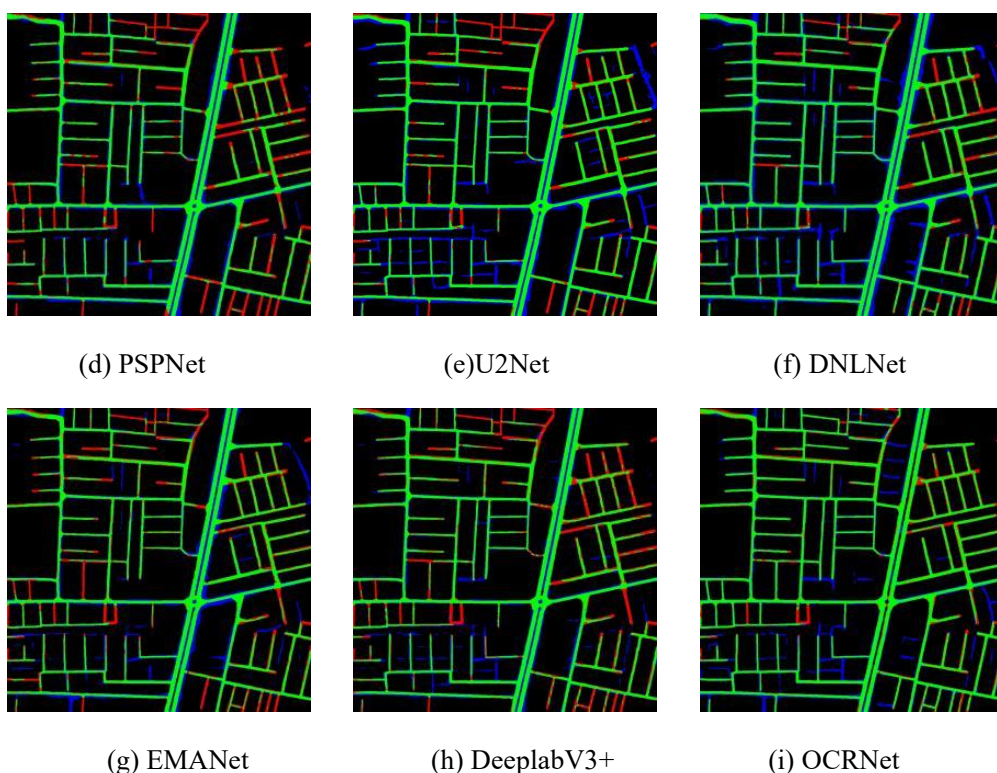


图 3.9 各语义分割网络预测结果（城市）<sup>1</sup>

城市作为人类活动最频繁的区域，也是路网分布的主要区域，因此在城市遥感影像下进行路网提取是重要的分割任务情境。城市中路网具有分布密集，占地面积大，呈格网规律分布的特点，因此从地物特征来看利于神经网络的学习和预测。由上图可见，大部分模型在城市高分影像下的路网提取结果都较好，对于城市的主干道，立交，桥梁等较大型的道路提取精度基本接近完全的分割，即使是早期的 FCN，PSPNet 等语义分割网络对于上图中的中央干道也达到了高准确率的提取分割，模型之间的提取结果差异较小。

然而对于除主干道以外，在城市中广泛分布的细小道路，不同的语义分割网络出现了明显的性能差异。因为这部分细小道路相对于主干道而言，面积较小，且常常受到行道树，建筑物的直接覆盖或阴影遮蔽。并且道路与部分建筑类地物的顶部往往具有较相似的外观，例如水泥铺设道路和水泥建造的房屋顶部，在遥感影像上区分度较小。这些可能的原因导致语义分割网络在城市细小道路的提取上精度较低。以上图为例，早期的 FCN，PSPNet 语义分割网络对于图像右侧的城市支路网出现了较多的漏判像元。而后期的 U2Net，DNLNet，DeeplabV3+ 等高精

<sup>1</sup> 绿色像元为被正确判断的真正道路像元(TP)，红色像元为被错判成背景类的真实道路像元(FN)，蓝色像元为被错判成道路像元的背景类像元(FP)，黑色像元为其他结果。

度网络减少了这部分的漏判，但仍存在较多错误判别的像元。本文所采用的 OCRNet 对于城市细小道路的处理则效果较好，没有出现整个街道漏判的情况，仅在部分道路连接处出现部分像元的误判和错判，属于可接受的错误，因为所整体预测结果仍然保留了路网的完整性。总而言之，OCRNet 在城市路网的提取具有明显的优异性，主要体现在城市支路网和细小道路的覆盖度上。

## (2) 工业区路网提取结果

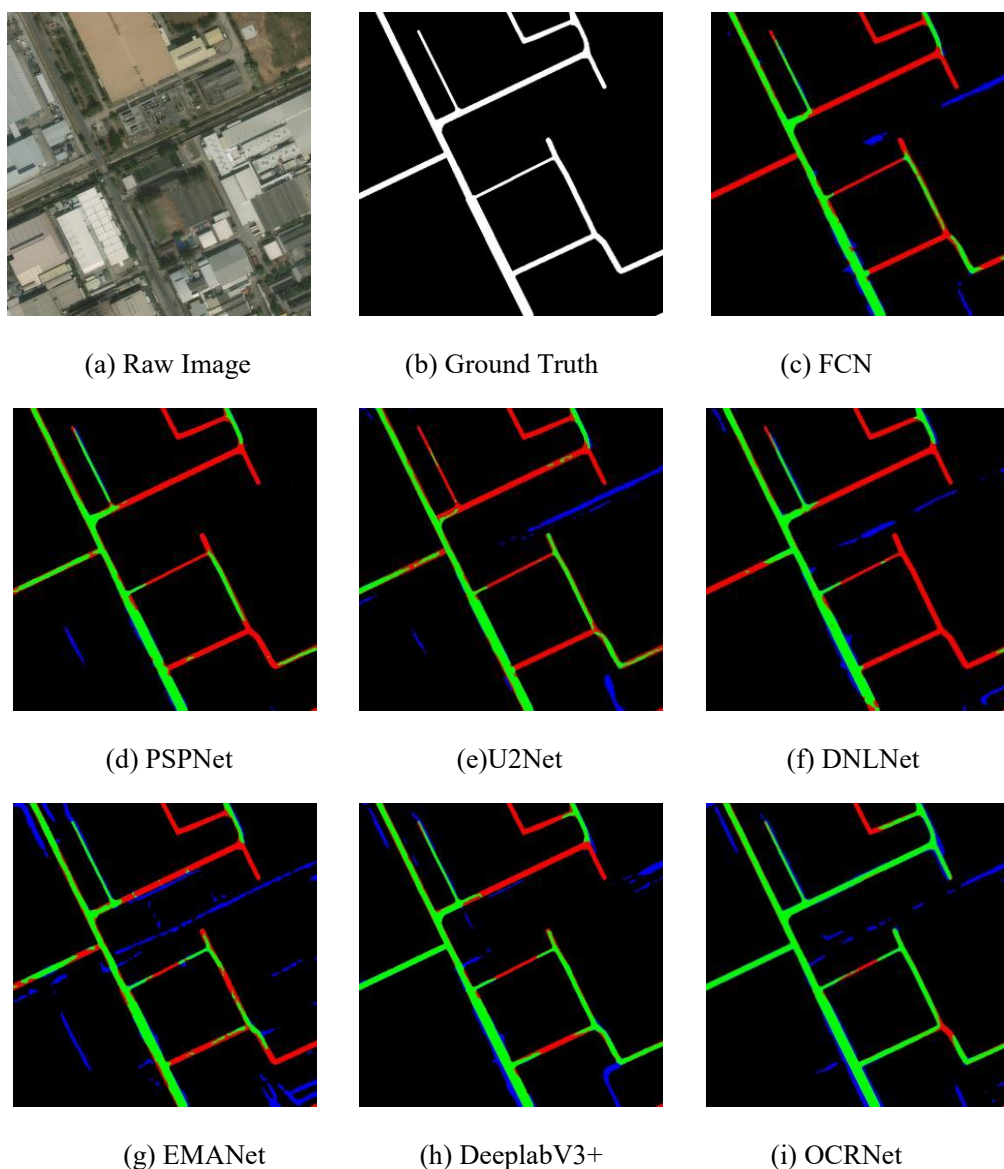


图 3.10 各语义分割网络预测结果（工业区）<sup>2</sup>

工业区是城市的重要组成部分，地物的分布与城市相似，路网与周边地物相间形成格网状分布，道路的形状相对规整。路网也常穿插在建筑物中间。但工业

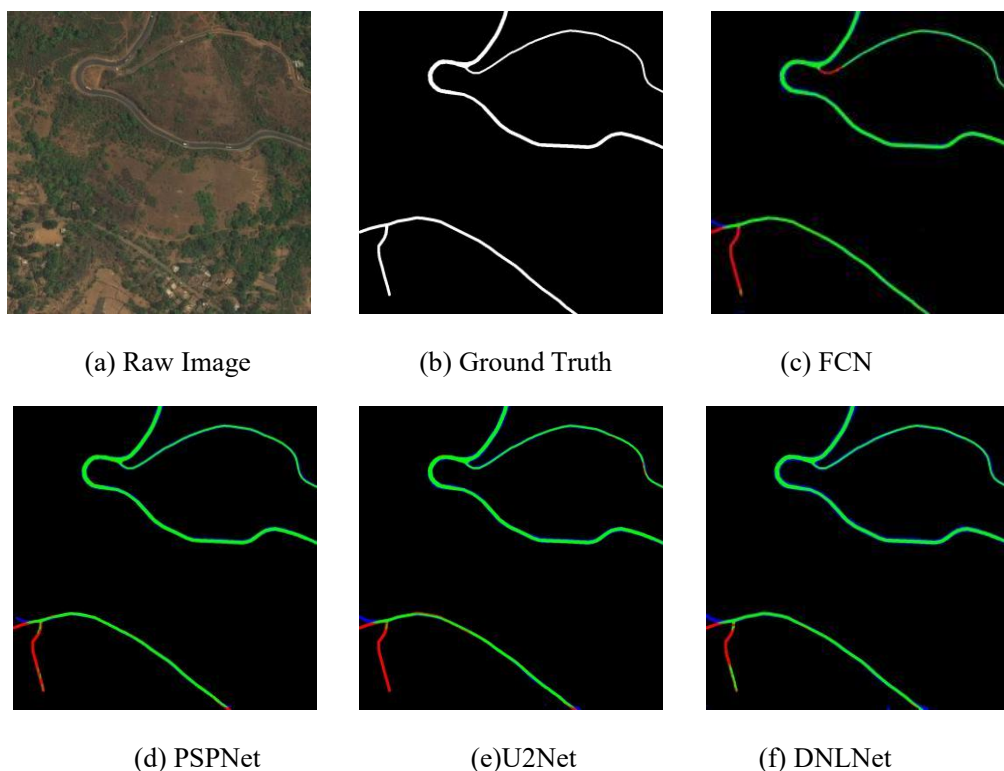
<sup>2</sup> 绿色像元为被正确判断的真正道路像元(TP)，红色像元为被错判成背景类的真实道路像元(FN)，蓝色像元为被错判成道路像元的背景类像元(FP)，黑色像元为其他结果。



路网的提取结果精度与城市路网提取结果的差距较大。由上图的各语义分割模型输出影像可见，各网络对于区域的主干道仍能保持较完整的提取结果，但对于支路网中的细小道路预测结果则不理想，直观上的漏判比城市细小道路更严重。FCN，PSPNet，U2Net，DNLNet，DeeplabV3+等语义分割网络尽管总体路网提取精度存在一定差距，但在工业区域遥感影像的路网提取的结果大致相同：主干道区域内的像元基本得到准确的预测，但影像右侧的支路网出现较多的漏判像元。从影像上看，工业区域内分布着大量与道路外观相似的空地，常分布在工业厂房之间，较难与道路区分开。此外，城市中的建筑物质地类型更加丰富，与道路相对容易区分，而工业区域中的厂房，仓库，各类工业设施等质地构造与道路相似度较高，预测模型将其区分开的难度更高。

在工业区域的路网提取任务中，相比于其他语义分割网络，OCRNet 展示了显著的优越性：对于影像区域右侧的细路网仍然保持了较高的预测准确度，尽管部分道路连接处的像元由于面积过小出现了一部分的漏判，但保持了路网的完整性。

### (3) 城郊区路网提取结果



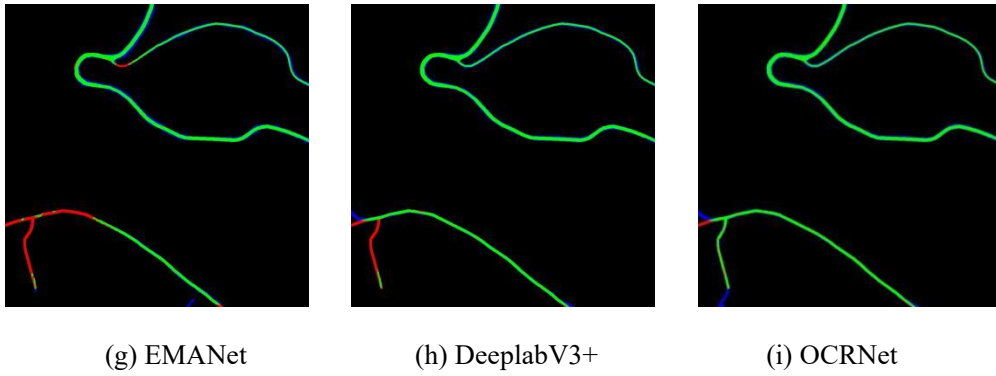


图 3.11 各语义分割网络预测结果（城郊）<sup>3</sup>

城郊区域统称除城市和工业区以外的其他区域，土地覆盖类型复杂，主要包括荒漠，裸地，林地，草地和耕地等。由于在城郊区域相对于其他区域出现的人类活动较少，交通需求较低，因此路网密度较小。城市和工业区域往往经历成熟复杂的城市规划和建设过程，因此路网的分布呈格网状规则分布；而在城郊区域，路网的分布状态在地势影响下往往呈弯曲状，或为绕开耕地林地不选择直线。尽管城郊区域路网的分布状态相对复杂，但在各语义分割模型中的路网提取结果相对较好，主要的原因是郊区路网与周边环境的区分度较高，例如道路与植被，道路与裸地，在输入的光谱信息之间已经存在较明显的差异。FCN，PSPNet，U2Net，DNLNet，DeeplabV3+等语义分割网络对于城郊主要路网的提取准确度都较高。

各模型的在城郊区域的路网提取结果主要差别在影像左下侧的细小道路。由上图可见，由于该段道路分布在树林区域，从遥感影像中可以看出明显被树木遮蔽。加上道路本身路宽较窄，面积较小，因此较难被模型辨别。而在这段道路上，大部分语义分割网络无法识别，包括 FCN，PSPNet，U2Net 模型。而 DNLNet，DeeplabV3+等精度较高的语义分割网络尽管能正确提取该段道路的部分像元，但仍存在较多的漏判区域。OCRNet 作为本次实验中精度最高的语义分割网络，成功的提取了该段道路，并保持了路网的连通性。

从以上路网提取结果可以看出，路网的宽度，面积对于模型能否预测准确该道路具有较大的影响。对于宽度较大，面积较广的主干道，各语义分割模型的路网提取精度差异较小，基本都能成功提取。而对于宽度窄，面积小的支路网，细小道路，狭窄街道，不同分割模型表现出明显的差异。在工业区域中的细窄道路

<sup>3</sup> 绿色像元为被正确判断的真正道路像元(TP)，红色像元为被错判成背景类的真实道路像元(FN)，蓝色像元为被错判成道路像元的背景类像元(FP)，黑色像元为其他结果。

与环境之间的特征差异较小，而成交区域中的细窄道路受植被遮蔽明显。OCRNet 在这类预测难度较高的道路区域时表现出显著的优越性：道路基本能够被识别和保留，连接处之间的连通性得到较大的改善。但其路网提取结果仍存在优化提升空间。

## 第四章 多模型融合的路网提取方法

### 4.1 基于集成学习的模型融合

尽管深度学习在本世纪得到爆发式的发展，各类模型和神经网络层出不穷，但现有的模型仍存在其自身的局限性，在预测精度上也不可能达到完美的效果。为了进一步提高预测精度，解决单一模型的局限性，基于集成学习的概念进行多模型间的融合逐渐成为深度学习领域的主流趋势。集成学习的主要思想就是通过各种模型融合的策略，把多个基模型进行系统的结合，使得最终的预测结果是基于多个模型得到的，从而进一步提高精度和鲁棒性，使得总体架构应用场景更广，性能更强。

集成学习是一种深度学习的思路，在具体的模型构建和解决实际问题中需要包括多个环节，在不同的环节有相应的集成策略。包括在训练阶段针对样本划分和训练集增强的训练策略，如堆栈法，装袋算法等；以及在预测阶段将多个基模型进行结果的融合策略，如投票法，加权法，随机森林法等；还有把训练和预测结果融合结合起来的整体集成策略，如 Adaboost，GBD 和 XGBoost 等。下文将结合本次实验的集成策略选择对前两类进行详细介绍。

#### 4.1.1 训练策略

集成学习的优势在于多个基模型之间的优势互补，这就要求基模型之间需要足够的差异性。除了模型本身网络结构的差异，在训练过程中往往还需要各自训练样本的差异性，从而更好的解决单模型出现的过拟合现象，降低融合后整体架构的泛化误差。

提升法(Boosting)是最早提出的模型融合方式，原理是对样本训练进行多轮迭代，每一轮训练都基于上一轮的分类结果对样本的权重值进行调整。但提升法对异常点较敏感，会逐级影响后面的产生的弱分类器。

引导聚集算法(Bootstrap Aggregating, Bagging)是常用的训练策略之一。具体步骤是针对每个基模型，通过从全体训练样本中进行随机重抽样方法来获取其子训练集。由于每个基模型获得的训练样本集是相互独立且随机的，因而其经过各自并行的训练后获得的神经网络之间的异质性较强，从而能提升模型融合后的效果。并基模型的子训练集均来源于原始训练集的随机抽样，模型之间的预测效果不会



出现较大的差异，保证融合后模型的稳定性。

堆栈法 (Stacking) 是将数据集进行  $K$  折的划分，为每个基模型选择  $K-1$  份数据进行训练，然后将各基模型对终端模型训练集的预测结果，代替原影像数值作为终端模型的训练数据的各维度特征进行输入，从而得到一个可以基于基模型结果进行最终预测的终端模型。终端模型可以是简单的逻辑回归，或是多层感知器等复杂的分类器。理论上通过堆栈可以稳健的组合较多的模型，是大型集成策略的常用方法之一。

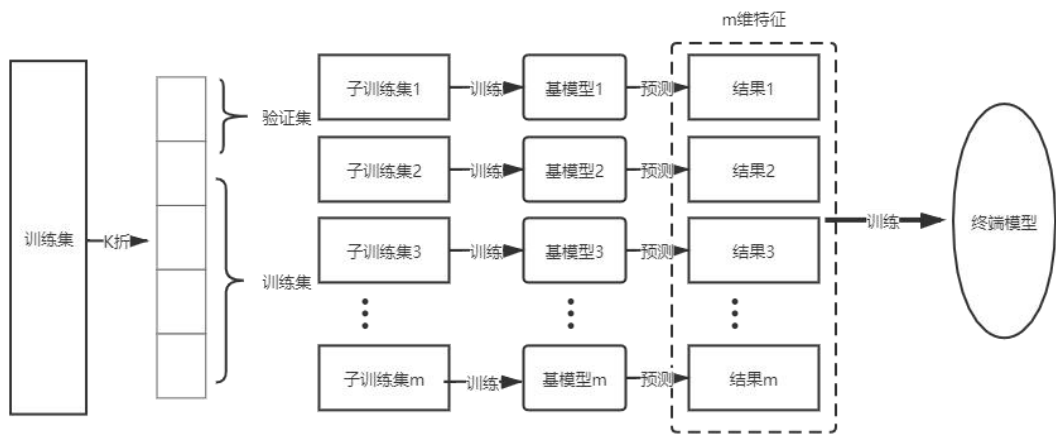


图 4.1 堆栈法模型建立模型流程

### 4.1.2 结果融合

经过集成学习的训练策略后，会获得多个训练好的基模型，能够对预测影像获得各自的结果，但这些结果仍是独立且割裂的，仍需要一个更好的融合器进行结合，以获得一个更好的最终结果。这部分工作将由基于集成学习的结果融合策略所完成，包括：投票法，回归法，随机森林，多层感知器等。

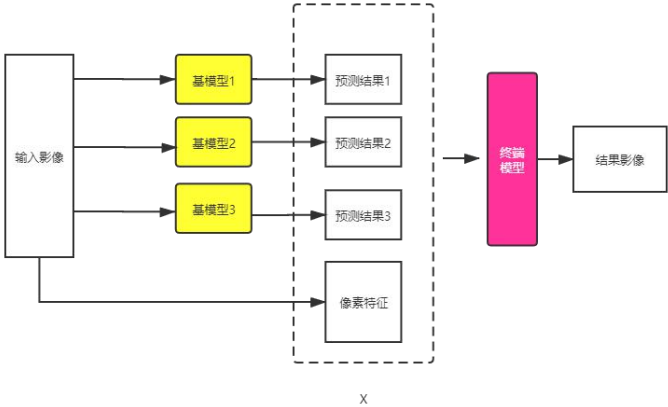


图 4.2 终端模型工作原理示意图

理论上结果融合可以使用任何一种分类模型，甚至可以使用复杂的神经网络，将多个神经网络串联起来，终端模型接收上游的预测结果作为输入特征，再进行预测结果的推断。然而这会导致整体架构过于复杂，在精度提升有限的情况下耗费的时间和硬件性能过多，使得整体架构的适用性降低。因此使用复杂的神经网络作为基模型后，精度已经很高，此时只需要简单的模型作为融合器把结果进行后处理即可。因此本次实验将会关注简单但具有代表性的几类结果融合方式。

投票法(Voting)是最简单的模型融合方式，通过多个模型的预测结果来决定像元分类的最终结果，依据是大多数基模型的结果(Hard Voting)或是依据模型可信度确定的基模型结果的加权平均(Soft Voting)。投票法的优势是简单快速且直接，但结构相对简单，对基模型的精度依赖较强。

随机森林(Random Forest)是另外一种常用的结果融合模型，在集成学习中作为由多个决策树构成的终端模型。随机森林本身可以直接作为分类器，接受图像特征作为输入进行预测。而在集成学习中，随机森林接受的特征则是不同基模型的预测结果，然后再给出最终的预测结果。与投票法不同，随机森林本身作为一个监督学习的模型还需要额外的训练数据。因此随机森林相比与投票法得到的结果是更加复杂的推断。

多层感知器(Multi Layer Perceptron)是一种特定形式的神经网络：不同的网络层之间是全连接的，是一种较为简单和早期的神经网络，但仍是一种可靠的分类器。并且由于其结构的简单，效率较高，适合作为集成学习架构中的终端模型。

#### **4.1.3 模型相对差异指数**

集成学习的核心思想是将多个“好而不同的”的分类器进行集成，基模型之间的差异越大，那么他们在融合之后结果互补的可能性则越大。如果基模型之间本身相似度较高，那么可能陷入预测结果类似的局面，无法进一步提高预测的精度。在上文中已经提到：通过训练集的随机抽取和划分，能够使得不同基模型在训练时获得样本的差异性。然而除了训练集，基模型之间本身的异质性也是挑选合理搭配的重要因素。为了量化模型之间的异质性，本文提出了相对差异指数。

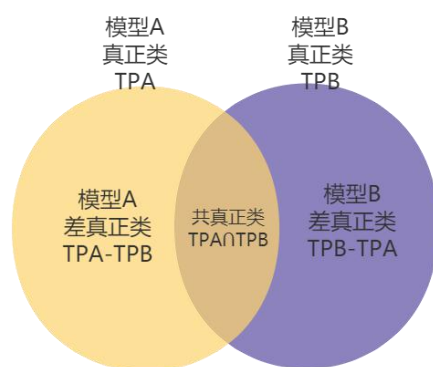


图 4.3 差真正类示意图

该指数的具体计算步骤为：对多个基模型进行训练后，计算每个模型成功预测的，但其他模型预测失败的部分真正类像元数作为其差像元数，在集合运算中是两个真正类像元集合的差集。然后计算两个模型之间各自差真正类像元数的比例作为模型相对差异指数。计算公式如下：

$$I_{diff} = \frac{|True Postive A - True Postive B|}{|True Postive B - True Postive A|} \quad (4.1)$$

其中 True Postive A 和 True Postive B 分别表示两个模型预测结果中的的真正类像元集合，Idiff 表示计算得到的模型相对差异指数。该指数实质上是从模型的预测结果出发，观察模型性能表现更好的部分相对于另一模型更好部分的像元数量，以此来判断模型能否在融合之后能够更好的补充基模型的结果。由于上文验证 OCRNet 在路网提取中精度最佳，因此在确定 OCRNet 作为基模型之一的情况下，需要考量与其异质的模型，计算各模型与基模型的相对差异指数。本文选取精度最高的五个模型进行备选对比，计算他们与 OCRNet 的相对差异指数。

表 4.1 部分模型与 OCRNet 的相对差异指数

对比模型	模型差真正类像元数	OCRNet 差真正类像元数	相对差异指数
FCN	580059	4315291	0.134
PSPNet	379841	5276897	0.072
U2Net	534440	4860768	0.110
DNLNet	1155464	2105575	0.549
EMANet	944494	4551123	0.208
DeeplabV3+	691446	3128091	0.221

如上表所示，DNLNet 和 DeeplabV3+网络，相对于 OCRNet 的性相对差异指数较高，且本身精度水平优秀，符合模型融合中对于基模型之间“好而不同”的

要求。接下来的实验将会以这两个模型和 OCRNet 作为集成学习的基模型。

## 4.2 基于损失函数优化的类别均衡

尽管模型融合能够集成多个语义分割网络，进一步提高路网提取的精度，但所采取的集成策略减少了各基模型的训练样本数。对于把基模型进行融合的终端模型，如果采用复杂的监督学习模型，如随机森林等，则需要把给基模型训练的一部分数据留出作为融合器的训练影像。而路网数据本身就有类别不平衡的问题：道路像元相对于非道路像元的数量非常的少。因此，在对基模型进行训练时，样本数不足以及样本不均衡的问题会更加明显。

目前在计算机视觉领域，解决样本不均衡问题一般有以下方向：（1）损失函数改进；（2）神经网络结构的优化和模块引入；（3）训练样本的增强重采样和构建。基于损失函数的样本均衡具有精度提高显著，操作便利，无需修改数据集和模型的优势。因此本文将基于损失函数优化的类别均衡方法，对语义分割网络的进一步优化，以此保证精度不受训练样本减少的影响。

### 4.2.1 动态类别加权

当训练影像中正负样本数量差异很大时，神经网络训练时在对负样本进行预测以及损失计算时很容易达到较低的损失值，因为负样本的预测由于自身较大的数量精度会较高，而真正需要关注的正样本，对于损失值的影响较小，因此会出现负样本“淹没”正样本的问题。在类别不平衡下样本训练得到的网络模型会在预测时针对正样本的预测精度不高。在路网提取的任务中，由于非道路像元比例过大，非道路像元在神经网络时主导了损失值的改变，从而影响参数调整梯度下降的方向，因此尽管此时总体的精度较高，但对道路的分类精度较低，不符合路网提取的训练要求。

客观条件下大幅扩增正样本数量时，解决类别不平衡的最直接方式就是在损失函数计算时为目标类别添加更大的权重，而为背景类别设置较小的权重，用以调整对损失函数的影响，防止过多的负样本导致的“淹没”问题。传统的方式是静态权重：在训练前根据先前经验和训练集的状况进行手动的设置，且设置的权重将在整个训练过程中保持不变。静态权重的优势在于计算量较小，对原本的训练速度几乎不产生影响。但静态权重的设置对于数值的准确性较高，可能需要多次的反复实验才能获得较好的权重值列表。

为了解决静态权重的局限性，动态类别加权策略在原始的 Softmax 损失函数与交叉熵损失函数的基础上，提出一种动态调整类别权重方法(Dynamic Weighted Loss)。在每次训练迭代中，动态策略会根据每张训练影像中的各类别像元的数量分布进行目标类别权重的修改和调整。目标类别权重与正样本占有所有像元数量的比例理论上成正比。但在实际训练时总像元数会增加 EPS 项以防止比例超出 0 至 1 的范围产生计算错误。本文将会采用动态加权的二元交叉熵损失函数进行对比实验。

#### 4.2.2 在线难例挖掘

难例(Hard Example)是指容易被模型错判为负样本的正样本像元，在路网提取任务中即为容易被错判为背景类的道路像元。实际上难例是样本不均衡的一种表现：由于道路像元数量少，分布零散，且夹杂在大量的非道路类像元中，因此导致道路像元常被错判。传统的难例挖掘策略(Negative Example Mining Method)为了提高网络模型对于这部分难例的判别精度，会一段训练结束后固定模型，然后进行预测以获得难例，之后调整难例重复进入训练集，以强调模型学习这部分难例。但这种两部走的传统方式不适用于端对端的深度学习网络训练，因为基于梯度下降的神经网络训练在一次迭代中会更新近千个参数，如果在几次迭代后就要固定模型输入难例，会极大的拖延模型的训练效率。

为了解决传统难例挖掘策略在神经网络上使用的局限性，文献<sup>[21]</sup>提出一种对比难例挖掘嵌入到梯度下降过程中策略，由于其实时性被称为在线难例挖掘策略(Online Hard Example Mining Method)。为了实现实时性，OHEM 将会增加一个并行的“只读”网络模块，在训练时接受原网络所传输的推荐难例区域 ROI，计算这部分区域的损失后获取通过非极大值抑制(NMS)进行排序，选取损失较高的区域传回给原网络进行真正的参数训练和梯度优化。本实验将会将 OHEM 策略结合到模型训练中进行对比实验。

#### 4.2.3 Lovász 损失函数

以上两种方法修改了损失函数的组合方式和计算结果，不涉及函数本身的改变。而 Lovász 损失函数<sup>[22]</sup>则是了一种全新的损失函数，代替以往语义分割训练常用的二元交叉熵函数，Dice 损失函数等传统形式。Lovász 损失函数的设计思路是试图将交并比指数直接利用到模型训练中。为了解决直接引入交并比形成的损失函数离散的问题，该损失函数引入 Lovász 拓展对原有的次模态函数进行延拓，从

而解决离散问题。

Lovász 损失函数在各主流的语义分割模型和数据集上的性能和精度提升均得到验证，因此已成为较主流的语义分割任务的损失函数之一。Lovász 损失根据语义分割任务中的类别数可划分为：Lovász Softmax Loss 和 Lovász Hinge Loss。前者在多分类问题上的性能更优异，而后者在二值分类中的表现更好。在文献中推荐与交叉熵损失函数或 Dice 损失函数进行加权融合。本文将会采用两种结合形式的 Lovász Hinge Loss 进行对比实验。

### 4.3 实验细节

#### 4.3.1 数据集构建与预处理

为满足模型融合中终端模型训练的需要，本实验基于 DeepGlobe 路网数据重新构建本实验数据集。验证集与上一实验保持一致：采用 622 张影像进行基模型的验证以及融合后模型的验证。原因是为了在基模型训练时能够将其与上一实验得到的各类网络模型的精度指标进行直接对比，发现其在训练样本削减但验证集保持不变下，结果的精度的损失情况，从而更好的评估基模型精度，以保证各基模型本身已具有较好的路网提取水平。除验证集外的影像按照 7 比 3 进行基模型训练集和终端模型训练集的划分。因为基模型和终端模型是两个串行的系统：终端模型需要接收各基模型的预测结果作为训练集的特征维度，所以两个系统的训练数据应相斥。

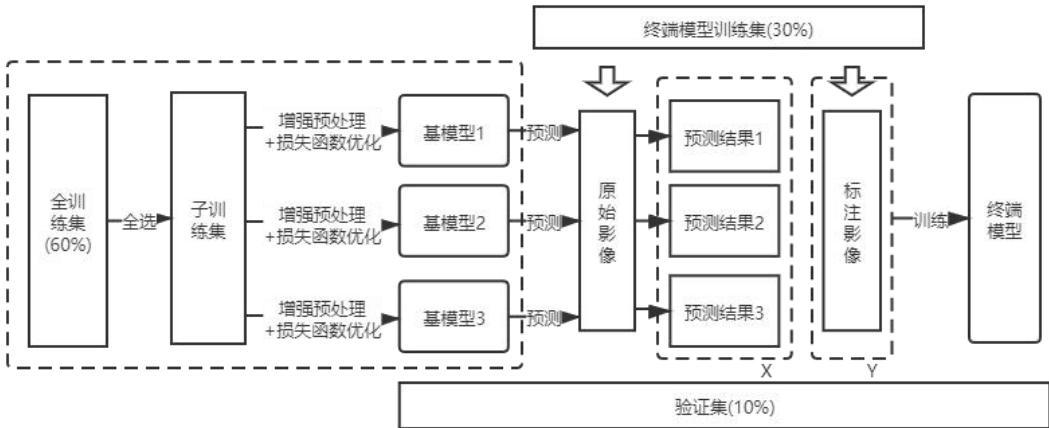


图 4.4 实验流程图

对于各基模型训练集的划分，本实验将采用特殊的堆栈法。一般的堆栈法需要是通过随机重抽样方法对训练集进行抽样，或采用 K 折的方法，以此为每个基

模型创建各不相同子训练集，提高基模型之间的差异性，从而优化融合后模型的泛化能力。但在本实验下由于训练集进一步减少，再进行子集的抽取，各基模型的训练影像数会变得过小，严重影响模型精度。实际上基模型之间的差异性除了子训练集的随机抽取能实现外，还可以通过数据的预处理来实现。回顾上一实验中对数据的预处理步骤：随机裁剪，随机旋转和随机畸变等，即使对于同一张影像，不同基模型在不同的迭代次中都能获得随机变换系数。基于这种充分随机预处理以及样本数据量过少的考虑下，本实验对每个基模型都采用全训练集，不进行子训练集的抽取。数据的预处理步骤与上一实验一致，但为了进一步增强随机性以弥补全选训练集的不足，变换系数的取值范围进行扩大，具体数值如下表。

表 4.2 随机增强预处理变换系数

	随机 缩放	随机 旋转	随机 翻转	亮度 变换	对比度 变换	色调 变换	饱和度 变换
原预处理	0.75-1.5	0° -30°	Vert.+Hori.	0-0.4	0-0.4	0-0.4	0-0.4
随机增强	0.5-2	-60° -60°	Vert.+Hori.	0-1.2	0-1.2	0-1.2	0-1.2

### 4.3.2 实验设计

实验的环境和平台与上一实验保持一致，学习率，优化方式，批尺寸等神经网络训练超参数也与上一实验相同。数据集和预处理方式则采用上文提出的方案。

实验分为两部分，第一部分将评估不同样本均衡策略下采用不同损失函数组合所带来的模型精度差异。为了在策略对比中尽量控制变量，各方案均进行 OCRNet 的语义分割网络训练，骨干网络为 HRNet-48，训练和验证数据集采用上文提出的基模型的训练集和验证集。仅改变损失函数的设置。最后得到的最佳模型可以直接作为下一部分的基模型之一，减少重复训练。

实验第二部分是进行模型融合，以及对融合后的整体模型进行精度评估。具体的实验路线已在上文提出，损失函数将会采用上一部分所得到的精度最佳的方案。在融合的基模型数上，为兼顾精度水平以及精度提升所带来的效率代价，本实验将选定 3 个基模型进行融合。基模型的选择原则按上文所提出的相对异质性指数确定，在已确定单模型精度最高的 OCRNet 作为基模型之一的情况下，挑选两个与 OCRNet 相对异质指数最高的 DeeplabV3+和 DNLNet 作为另外两个基模型。之后会基于不同的终端模型训练得到不同的融合后整体模型，通过精度指标的对

比得到最佳的模型融合方案。终端模型将会包括：投票器，随机森林模型，多层感知器。分别是统计领域，传统机器学习领域和深度学习领域所具有代表性的元分类器。

## 4.4 结果分析

### 4.4.1 模型优化对比

表 4.3 基于类别均衡方法的损失优化对比

Loss A	Loss B	$\alpha$	$\beta$	Road Acc.	Road IoU	Kappa
BCE	BCE	1	1	0.651	0.618	0.742
DWBCE	DWBCE	1	1	0.675	0.627	0.745
OHEM	BCE	0.3	0.7	0.698	0.617	0.741
OHEM	DWBCE	0.3	0.7	0.651	0.623	0.742
Lovász Hinge	DWBCE	0.5	0.5	0.635	0.614	0.749
Lovász Hinge	Dice	0.5	0.5	0.719	0.623	0.752

上表列出采用不同的损失函数组合策略下训练得到模型的精度评价指标。总损失函数的获得基于简单的加权的组合。本实验所采用的组合权重系数值均为相关文献中针对语义分割任务的推荐值。公式如下：

$$Loss = \alpha * Loss A + \beta * Loss B \quad (4.1)$$

采用动态类别加权，在线难例样本挖掘以及 Lovász 策略损失函数，相比于仅采用二元交叉熵损失函数，训练获得的模型在各精度指标上均获得提升。采用动态类别加权的精度提升不大，获得的道路 IoU 值提升 1.45%，准确率提高 3.69%，Kappa 值提高 0.4%。采用在线难例样本挖掘 OHEM 对精度的改善也不明显，甚至出现了某些指标的下降，可能的原因是数据集样本不平衡程度不高，OHEM 策略的增强效果不明显。而采用 Lovász 损失函数对于模型精度的优化效果显著，采用 Dice 与 Lovász Hinge 损失函数进行加权结合的策略对于主要指标均有大幅提升：道路准确率提升 10.45%，Kappa 值提高 1.34%，道路 IoU 提升 0.81%。说明 Lovász 损失函数对于本数据集下的类别不均衡程度适用性最高。



表 4.4 样本抽选后损失优化对比

损失函数	训练样本数	Road Acc.	Road IoU	Kappa
BCE + Dice	3735	0.741	0.665	0.789
Lovász Hinge + Dice	5604	0.719	0.623	0.752
		-0.022 (-2.97%)	-0.042 (-6.32%)	-0.037 (-4.69%)

同时将其与全训练样本训练下的模型精度进行对比，可以看到：尽管样本数量由于模型融合的要求需要大幅减少，但采用了最佳的损失函数组合后，样本减少所带来的精度降低影响较小，因此得到的基模型仍是可靠的路网提取模型，保证了模型融合“好而不同”原则中的对基模型单性能优秀的要求。基于此，本实验接下来会选定最佳组合策略作为各基模型的最佳损失函数设置，为下一部分的模型融合进行各基模型的建立和训练。

#### 4.4.2 模型融合对比

表 4.5 主要单网络和集成路网提取模型精度指标

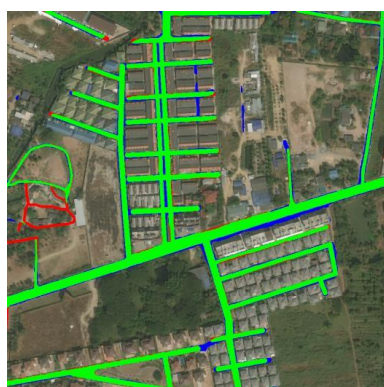
	Precision	Recall	F1	Road Acc.	Road IoU	Kappa
DeeplabV3+	0.667	0.808	0.731	0.737	0.609	0.746
DNLNet	0.766	0.694	0.728	0.760	0.594	0.734
OCRNet	0.739	0.848	0.790	0.741	0.665	0.789
Voting*	0.620	0.911	0.738	0.620	0.584	0.752
RandomForest*	0.737	0.831	0.781	0.731	0.661	0.792
MLP*	0.740	0.863	0.797	0.740	0.672	0.807

\*本文提出的多网络集成模型

上表展示了基于相同验证集下，各单一语义分割网络，以及由以上三种语义分割网络合成的，通过不同融合策略所得到的高级路网提取模型的精度水平。尽管各语义分割网络作为基模型仅获得 33% 的原训练影像数，但经过多个模型的集成以及一个可靠终端模型的训练，最终获得的整体路网提取模型相比于原单一网络组成模型具有显著的优越性。基于投票的集成模型在召回率是有明显优势，不仅远高于单网络模型，也高于基于其他融合方法的集成模型，但在精确率上水平较低，不及任一单网络模型。因此导致基于投票的集成模型在各综合的精度指标

上均较低，路网识别的准确率和交并比不太理想。由此看出通过投票法进行模型的融合逻辑上简单，并且不需要对终端模型进行训练。其预测结果趋向于各模型结果的叠加，提取路网相较于实际道路更加宽松，由于更广的覆盖导致召回率高，综合精度低。

基于随机森林和多层感知器则将一个较复杂的分类器作为终端模型，因此需要一定量的训练数据，理论上应该获得比无需训练的投票法精度较高。如上表所示，基于随机森林的集成模型在路网提取任务中各精度指标水平均接近 OCRNet，在 Kappa 值，道路类准确率上略有优势。基于多层感知器进行融合的集成模型获得了最佳的路网提取精度：F1 值，道路交并比以及 Kappa 值均为最高水平，高于 OCRNet 单网络模型 0.88%，1.05%，2.28%。因此基于关键的精度指标对比，基于多层感知器获得的集成模型在路网提取任务上具有一定的优越性。



(a) OCRNet



(b) Voting Ensemble



(c) Random Forest Ensemble



(d) MLP Ensemble

图 4.5 单网络模型与多网络集成模型的路网提取预测结果<sup>4</sup>

<sup>4</sup> 绿色像元为被正确判断的真正道路像元(TP)，红色像元为被错判成背景类的真实道路像元(FN)，蓝色像元为被错判成道路像元的背景类像元(FP)。

从预测结果影像上看，尽管基于 OCRNet 单网络的路网提取模型已经获得了非常好的效果，对于主干道和街道路网基本实现了非常准确的提取，基本不存在漏判和误判的像元。

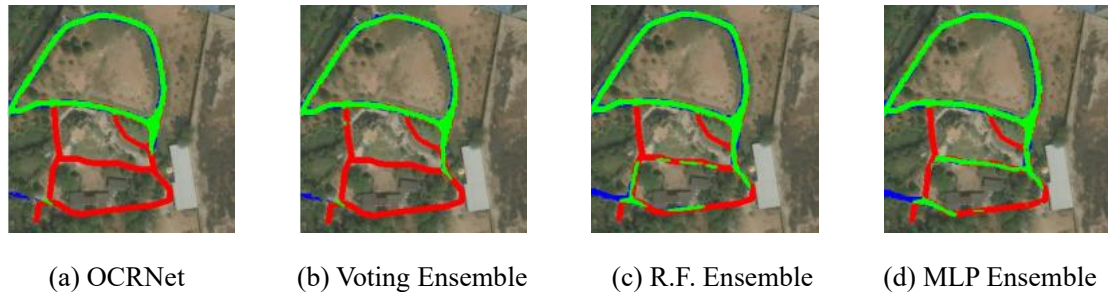


图 4.6 单网络模型与多网络集成模型的细小路网提取预测结果

但对于影像左侧的分布在裸地上的细小道路，由于面积相对较小且分布不规则，OCRNet 模型对这部分路网的识别效果较差。基于投票法的集成模型由于仅仅是模型结果之间的简单叠加筛选，因此吸收了部分低精度基模型的错误识别结果，不仅没有解决对右侧的漏判道路像元，且在影像右侧产生了大量漏判和错判预测结果。而基于随机森林和多层感知器等较高级分类器的集成模型则成功的对细小道路中单模型的漏判像元进行了正确的预测，显著的优化了仅靠 OCRNet 提取的路网结果。

此外，通过上图预测结果的对比可以发现，OCRNet 作为关注上下文信息的语义分割网络，所获得的分类结果往往在环境相似性高的影响下整体性较强，所提取得到的路网具有连通性。但经过模型融合后的预测结果，由于结合了的弱上下文增强的语义分割网络结果，获得的路网提取结果出现了较多的破碎像元，路网的连通性也有所降低。主要原因是受弱基模型预测结果影像。因此集成的路网提取模型尽管部分精度上获得提升，但对路网连通性有一定影响。

表 4.6 本文及已发表文献对 Deepglobe 道路数据集路网提取精度对比

模型	P	R	F1	Road IoU
<b>OCRNet</b>	<b>0.739</b>	<b>0.848</b>	<b>0.790</b>	<b>0.665</b>
LinkNet50-MS <sup>[23]</sup>	0.812	0.786	0.801	0.668
RoadCNN <sup>[24]*</sup>	-	-	-	0.676
CGNL <sup>[25]</sup>	0.820	0.788	0.804	0.672
<b>MLP Ensemble</b>	<b>0.740</b>	<b>0.863</b>	<b>0.797</b>	<b>0.672</b>
GCNet <sup>[26]</sup>	0.823	0.795	0.809	0.679
CBAM <sup>[25]</sup>	0.815	0.804	0.801	0.681
GAMS-NET <sup>[23]*</sup>	0.831	0.800	0.815	0.688

\*模型代码未开源

为了进一步对比和评价本实验所得到的路网提取集成模型的精度水平，上表总结了已公开发表文献使用同一数据集的路网提取方法，且仅筛选精度水平高于单 OCRNet 模型的网络模型和提取方法。可以看到，目前精度高于 OCRNet 的方法仅有 7 个，而优于通过本文提出的多层感知器集成模型的模型或方法仅有 4 个，其中有 1 个未开源可复现的代码。因此本文提出的集成模型在已有的开源模型或方法中精度名列前 3，与众多路网提取模型对比中基本在精度上均保持明显优势。此外，本文提出的集成模型在回报率水平上目前位于相关研究的最高位：高于精度最佳的 GAMS-NET 近 7.9%，而 F1 值和 IoU 仅损失 2.2% 和 2.3%。说明本文提出的模型更趋向于做大胆的预测：尽管存在较多的错判像元，但对于道路像元的覆盖率更广，较少存在漏判。因此适用于路网提取中对召回率要求高，但综合精度相对不高的情景。总之本文所提出的集成模型在精度上保持目前已有方法模型的主流，且具有自身的独特性和适用场景，具有显著的优越性。

## 第五章 总结与展望

### 5.1 工作总结

基于遥感影像进行路网提取一直是遥感领域的热点问题，因其在城市规划，道路导航，交通建设等方面具有相当高的应用潜力。但路网提取中样本不平衡，边界不清晰，像元数量少等问题一直成为研究难点与挑战。本文围绕着深度学习中的语义分割网络进行遥感影像中的路网提取，主要解决了以下问题：

（1）将应用在自然场景下语义分割神经网络引入到遥感影像的地物分类和识别中，并首次将性能优秀的 OCRNet 应用到到高分遥感地物分类和路网提取。本文基于公开的 Deepglobe 道路数据集对主流的语义分割神经网络进行训练和评估，以获得并横向比较各网络在路网提取任务中的各精度指标。最终验证了 OCRNet 在高分影像路网提取中的优越性。

（2）针对遥感影像道路数据集中常见的训练影像少，样本不均衡和道路类像元数量少等问题，本文引入语义分割中常用的样本均衡策略，包括：动态加权，在线难例挖掘，以及 Lovász 损失函数。通过基于不同策略下损失函数设计与模型训练，比较不同方案带来的精度优化差异。最终得到了高分遥感影像路网提取任务中的最佳样本均衡策略：采用 Lovász 与 Dice 损失函数结合进行神经网络的训练。

（3）集成学习是实现路网提取精度提高的有效方法。本文提出一种相对差异指数进行各网络的结果差异量化，从而为集成学习选择可靠的基模型选择策略。之后基于改进的堆栈法进行基模型的训练，并对比不同终端模型所获得的最终路网提取结果的精度。经过多模型的训练，验证了多层感知器作为终端模型下的集成学习策略能够实现更准确的预测。

### 5.2 研究展望

本文对多语义分割网络在高分遥感影像中的路网提取进行了各方面的探索，并提出了一种可行的模型优化和集成策略，但在具体工作中囿于论文篇幅和时限，仍存在如下的研究的局限性：

（1）本研究的所有路网提取模型和优化策略的训练和评估均基于单一的 Deepglobe 遥感道路数据集。尽管其是目前影像数量，标注质量和文献引用均较优

秀的高分遥感道路数据集，但在路网提取研究领域，仍有其他公开遥感路网数据集，如 Massachusetts Roads, ERM-PAIW 等数据集可待研究使用。由于本文涉及的训练的网络数量过多，且受实验的硬件平台和时间限制，未能进行多数据集下的模型评价和对比，因此所获得各类精度及结论对于其他数据集的推广有待验证。所训练得到的路网提取模型在鲁棒性和泛化能力方面也有待加强。

（2）对于路网数据集中样本不均衡问题的解决，本文仅采用了基于损失函数的优化方法。对于语义分割任务中的小目标和小样本问题，在计算机视觉领域仍有较多的优化方法可待借鉴，包括基于网络结构优化的方法，以及对样本进行增强重采样等方法。更多的样本均衡策略在路网提取任务中的可适性和效果有待研究。

（3）在路网提取的集成学习策略选择中，限于论文篇幅与时限，本文仅采用了较主流的堆栈法进行多模型的集成。而对于其他较常用的集成学习策略，如：Adaboost, GBDT 和 XGBoost 等提升方法，进行路网提取任务的结果优化尚待对比和评估。此外，在堆栈法的终端模型选择上，本文仅采用了三类具有代表性的分类器，而其他更多的高级分类器，如决策树，卷积神经网络等，在作为终端模型时的集成模型的预测精度有待对比和考察。

## 参考文献

- [1] 史文中, 朱长青, 王昱. 从遥感影像提取道路特征的方法综述与展望[J]. 测绘学报, 2001, 30(3): 257-262.
- [2] Laptev I, Mayer H, Lindeberg T, et al. Automatic extraction of roads from aerial images based on scale space and snakes[J]. Machine Vision and Applications, 2000, 12(1): 23-31.
- [3] Gruen A, Li H. Road extraction from aerial and satellite images by dynamic programming[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 1995, 50(4): 11-20.
- [4] Maurya R, Gupta P R, Shukla A S. Road extraction using K-Means clustering and morphological operations[C]. International Conference on Image Information Processing, 2011.
- [5] 李晓峰, 张树清, 韩富伟, et al. 基于多重信息融合的高分辨率遥感影像道路信息提取[J]. 测绘学报, 2008, 37(2): 178-184.
- [6] Boggess J E. Identification Of Roads In Satellite Imagery Using Artificial Neural Networks: A Contextual Approach[J]. Mississippi State University, 1995.
- [7] M., Mokhtarzade, And, et al. Road detection from high-resolution satellite images using artificial neural networks - ScienceDirect[J]. International Journal of Applied Earth Observation and Geoinformation, 2007, 9(1): 32-40.
- [8] Ranzato, Marc'aurelio, Mnih, et al. Modeling Natural Images Using Gated MRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
- [9] Zhong Z, Li J, Cui W, et al. Fully convolutional networks for building and road extraction: Preliminary results[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016.
- [10] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.
- [11] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017: 1-1.
- [12] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. Springer, Cham, 2015.
- [13] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[J]. IEEE Computer Society, 2016.
- [14] Kong Y, Liu Y, Yan B, et al. A Novel Deeplabv3+ Network for SAR Imagery Semantic Segmentation Based on the Potential Energy Loss Function of Gibbs Distribution[J]. Remote Sensing, 2021, 13(3): 454.
- [15] Yuan Y, Chen X, Wang J. Object-Contextual Representations for Semantic Segmentation[J], 2019.

- [16] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
- [17] He K, Zhang X, Ren S, et al. Identity Mappings in Deep Residual Networks[J]. European Conference on Computer Vision, 2016.
- [18] Sun K, Xiao B, Liu D, et al. Deep High-Resolution Representation Learning for Human Pose Estimation[C]. arXiv e-prints, 2019.
- [19] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [20] Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]. 2016 Fourth International Conference on 3D Vision (3DV), 2016.
- [21] Shrivastava A, Gupta A, Girshick R. Training Region-based Object Detectors with Online Hard Example Mining[C]. IEEE Computer Society, 2016: 761-769.
- [22] Berman M, Triki A R, Blaschko M B. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [23] A X L, B Y Z A, A Z Z, et al. GAMSNet: Globally aware road detection network with multi-scale residual learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175: 340-352.
- [24] Bastani F, He S, Abbar S, et al. RoadTracer: Automatic Extraction of Road Networks from Aerial Images[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [25] Batra A, Singh S, Pang G, et al. Improved Road Connectivity by Joint Learning of Orientation and Segmentation[C]. CVPR, 2019.
- [26] Cao Y, Xu J, Lin S, et al. GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond[C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2020.



## 致谢

须臾之间，在中大的五年光阴倏然而逝。首先要感谢中山大学给予我的宝贵学习机会和优良治学氛围，使我获得充实而又非凡的人生。感谢地理科学与规划学院对我的接纳与包容，给予我宝贵的转专业机会，让我踏上截然不同的学习道路，探索地理科学的奥秘。

感谢李军导师对我的悉心指导，无论是学术上的指引还是人生道理的教诲，都使我受益匪浅。李军老师是一位学识渊博，待人和善的良师益友，受其教导我深怀感恩之心。感谢陈逸敏老师和李文楷老师对我的关照和支持，不吝余力的帮助我走向人生的更上一层台阶。

感谢叶展翔师兄对我毕业论文的关心和指导，从题目的构思拟定到论文的写作修改，叶师兄都给予了非常多的帮助。

最后特别感谢我的父母，二十余载对我的细心栽培与鼎力支持，使我能够健康快乐的成长，树立独立自强的人格。养育之恩当以终身为报。