

北京师范大学

硕士学位论文

论文题目：基于生成对抗网络的遥感影像分类研究

作 者： 江涛

导 师： XXX 导师

系别年级： 信息科学与技术学院 16 级

学 号： 201621210026

学科专业： 计算机软件与理论

完成日期： 2019 年 5 月

北京师范大学研究生院

北京师范大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文使用授权书

学位论文作者完全了解北京师范大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京师范大学。学校有权保留并向国家有关部门或机构递交论文的复印件和电子版，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。保密的学位论文在解密后适用于本授权书。

本人签名：_____ 日期：_____

导师签名：_____ 日期：_____

基于生成对抗网络的遥感影像分类研究

摘要

遥感影像分类是遥感领域重要的研究内容，如何快速、准确地对遥感影像数据分类与识别是当前遥感领域的研究重点。然而，由于遥感数据固有的不确定性使得影像准确解译识别面临挑战。此外，高分辨率影像在提供高质量细节信息时其同类地物类内特征差异大，使得输出结果中类内差异大的地物会存在许多细碎的错分区域，同时正确识别明确、清晰的地物分类边界存在挑战。

近年来，随着深度学习技术的发展和全卷积分类网络的出现，基于深度学习理论的影像分类方法被广泛应用到遥感影像分类任务中。然而，当前基于全卷积网络的影像分类方法在池化处理和上采样操作中容易丢失影像低阶边界、位置等特征信息，导致地物分类边界存在混淆、歧义。此外，全卷积影像分类方法是像素级的分类，影像分类结果中同类地物区域内容易出现细碎的错分区域。于是改进现有的全卷积影像分类方法，提升影像分类质量，也具有重要的研究意义。生成对抗网络其判别器能够不断检测并纠正真实样布与生成数据间的差异，且模型的对抗损失一定程度上能够增强影像像素点间的连续性，保持整体的一致性。另一方面，对抗博弈过程中生成器能不断强化分割模型的预测能力，提升分类精度。

因此，本文将对抗训练的思想引入全卷积影像分类模型，提升影像分类精度，并对提出的基于条件生成对抗网络的影像分类方法进行改进。文中首先通过模糊聚类分割方法预处理高分影像数据，得到影像的同质性分割单元信息图。为了得到更清晰的地物分类边界，将分割单元的边界掩膜特征融合到模型上采样的特征图上，加强高阶特征的边界、位置信息。为了得到更完整的地物分类结果，文中对像素点类别预测时，考虑同质性分割单元内其他像素点间的相互关系，优化类别预测结果。基于上述思路，本课题论文对遥感影像地物分类问题加以研究和讨论。具体的研究内容如下：

- (1) 将对抗训练思想引入全卷积影像分类模型，提出基于条件对抗生成网络的全卷积影像分类方法，新方法在高分影像数据上相比原方法分类能力更强，能够提升影像分类精度。
- (2) 先利用本文作者先前提出的面向对象的遥感影像模糊聚类分割方法预处理影像数据，得到影像同质性分割单元信息图。然后通过求解图像梯度和二值化处理获取分割单元

的边界掩膜信息。接着将经过下采样得到的不同尺度的边界掩膜特征融合到分割模型对应尺度上采样模块特征图上，加强高阶特征的边界、位置信息，提出融合边界特征的影像分类方法。新的影像分类方法能够得到更准确、清晰的地物分类边界。

- (3) 本文对模型预测像素点的类别概率值更新校正，考虑像素点所属同质性分割单元内其他像素点的类别预测关系，按照一定策略更新像素点类别预测概率值，提出基于辅助信息后处理的影像分类方法。添加辅助信息后处理方法能够保持同类地物区域内像素点的空间一致性。减少类别内细碎的错分区域，分类结果更加完整。
- (4) 将融合边界特征和辅助信息后处理这两种改进思路同时引入基于条件生成对抗网络的影像分类方法中。实验结果表明新的影像分类方法分类精度提升明显，地物边界更准确，且同类地物内错分区域大幅减少。

关键词：遥感影像，地物分类，全卷积网络，生成对抗网络，深度学习

REMOTE SENSING IMAGE CLASSIFICATION BASED ON GENERATIVE ADVERSARIAL NETWORK

ABSTRACT

Remote sensing image classification is a research hotspot in the field of remote sensing. How to quickly and accurately classify and identify of remote sensing images is the focus of current remote sensing research. However, it is challenging to accurately classify remote sensing images, due to the inherent uncertainty of remote sensing images. High-resolution images provide high-quality detail information with large differences in their similar species, which leads to the misclassification, and it is difficult to get the correct classification boundary.

In recent years, with the development of Deep Learning Theory and the proposal of Fully Convolutional Networks (FCN), image classification methods based on deep learning theory have been widely applied to remote sensing image classification tasks. However, the current image classification methods based on FCN are easy to lose the low-order boundary, location and other feature information in the pooling process and upsampling operation, resulting in confusion and ambiguity of classification boundary. In addition, FCN method is a pixel-wise classification method, a finely divided area is likely to occur in the same type of feature region. Therefore, improving the FCN method quality has important research significance. The Generative Adversarial Network(GAN) can continuously detect and correct the difference between the real sample and the generated data. The adversarial loss of the model enhances the continuity between the pixels of the image to a certain extent and maintains the overall consistency. Moreover, in the process of cadversarial training game, the generator can continuously strengthen the prediction ability of the segmentation model and improve the classification accuracy.

Therefore, this work introduces the idea of confrontation training into the FCN model to improve the accuracy of image classification. The image classification method based on conditional GAN (CGAN) is improved. Firstly, the homogeneity segmentation units image is obtained by fuzzy clustering method. In order to obtain a clearer classification boundary, the boundary mask features of the segmentation units are fused into the feature map of the model to enhance the

boundary and position information of the high-order features. In order to obtain a more complete classification result of the feature, the relationship between other pixels in the homomorphic segmentation unit is considered, and the class prediction result is optimized. Based on the above ideas, this work studies and discusses the classification of remote sensing image. The specific research contents are as follows:

- (1) Combine the adversarial training model with FCN model, and propose a FCN classification method based on CGAN. The new method has stronger classification ability than the original FCN method on high-resolution remote sensing image, which can improve the image classification accuracy.
- (2) First of all, fuzzy clustering segmentation method is used to preprocess the image data to obtain the image segmentation unit information map. Then, the boundary mask features of the segmentation unit are obtained, and the boundary mask features of different scales obtained by downsampling are fused to the feature maps of the corresponding scales of the model, and the boundary and position information of the high-order features are strengthened. New method can accurate and clear feature classification boundaries.
- (3) In this work, the class probability of the model prediction pixel is updated and corrected. Considering the class prediction relationship of other pixels in the homogenous segmentation unit of the pixel, the prediction probability value of the pixel class is updated according to a certain strategy, and the post-processing based on auxiliary information is proposed. Image classification method. Adding auxiliary information post-processing methods can improve the spatial consistency of pixels in the same feature area. Reduce the subdivided areas within the category and the classification results are more complete.
- (4) The two improved ideas of fusion boundary feature and auxiliary information post-processing are introduced into the image classification method based on CGAN. The experimental results show that the classification accuracy of the new image classification method is obviously improved, the boundary of the object is more accurate, and the mis-division area within the same type of object is greatly reduced.

KEY WORDS: Remote Sensing Imagery, Land Cover Classification, Fully Convolutional Network, Generative Adversarial Network, Deep Learning

目 录

摘要	I
ABSTRACT	III
插图索引	VII
表格索引	VIII
第1章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 传统的高分辨率遥感影像分类与识别方法	2
1.2.2 基于深度学习理论的影像分类研究现状	4
1.3 本文的组织结构	5
1.4 本文主要创新点	6
第2章 深度学习相关理论	7
2.1 卷积神经网络基础	7
2.2 全卷积神经网络	10
2.3 生成对抗网络概述	13
2.3.1 生成对抗网络模型	13
2.3.2 条件生成对抗网络	14
2.4 本章小结	15
第3章 基于生成对抗网络的遥感影像分类方法	16
3.1 引言	16
3.2 基于 CGAN 的影像分类方法	16
3.2.1 算法原理	17
3.2.2 算法实现	19
3.3 实验数据介绍与预处理	19
3.3.1 Vaihingen 数据介绍	19
3.3.2 数据预处理	21
3.4 实验结果与分析	23
3.4.1 实验环境	23
3.4.2 评价指标	24
3.4.3 网络参数	24

3.4.4 结果与分析	26
3.5 本章小结	29
第 4 章 基于 CGAN 影像分类的改进方法	30
4.1 引言	30
4.2 TFSV-IT2FCM 聚类分割	30
4.2.1 方法原理	30
4.2.2 影像数据处理	32
4.3 融合边界特征的 CGAN 影像分类方法	33
4.4 基于辅助信息后处理的影像分类方法	35
4.5 实验结果与分析	37
4.6 本章小结	40
第 5 章 总结与展望	41
5.1 本文的主要内容	41
5.2 未来的期望	42
参考文献	43
学术成果	48
致谢	49

插图索引

图 1	图像卷积示意图	8
图 2	最大池化与平均池化	9
图 3	典型的卷积网络结构	10
图 4	影像分类网络结构示意图	11
图 5	反卷积上采样过程	11
图 6	基于 FCN 网络的遥感影像分类示意图	12
图 7	GAN 结构示意图	14
图 8	CGAN 结构示意图	15
图 9	基于 CGAN 的影像分类方法模型示意图	17
图 10	Vaihingen 影像数据	21
图 11	图像裁剪方式	23
图 12	影像分类的可视化结果	27
图 13	DSM 波段对影像分类结果的影响	28
图 14	TFSV-IT2FCM 算法框架	31
图 15	TFSV-IT2FCM 提取 Vaihingen 影像辅助特征信息	33
图 16	融合边界掩膜特征的分割模型	34
图 17	基于辅助信息后处理的 CGAN 影像分类流程图	37
图 18	改进方法分类结果可视化图	38

表格索引

表 1	Vaihingen 数据类别标签颜色对照表	21
表 2	基于 CGAN 框架的全卷积分割模型参数表.....	25
表 3	Vaihingen 数据分类精度评估表.....	27
表 4	DSM 波段对分类精度的影响	29
表 5	融合边界特征对影像分类的影响.....	38
表 6	后处理对 CGAN 影像分类的影响.....	39
表 7	两种优化方法共同作用对影像分类的影响.....	39

第1章 绪论

1.1 研究背景及意义

遥感技术是从各种传感器上收集地物目标的电磁辐射信息，经处理后成像，从而对地物信息进行探测和识别的一种技术。遥感影像数据被广泛应用在军事侦察、环境监测、植被分类、土地利用规划和矿产资源勘测等领域^[1]。近年来，随着卫星遥感技术的发展和信息科技技术的完善，遥感影像分辨率不断提高，高分影像蕴含信息量越来越丰富。同时全球遥感数据成爆发式增长，但相关统计表明，遥感数据 95% 是不精确的、非结构化的数据，人类能够利用的数据仅占 5% 左右^[2]，如何在有限时间内高效利用遥感数据是当前遥感技术发展所面临的挑战。

影像分类与目标地物识别是遥感影像分析和应用的重要内容，如何准确、快速地对遥感影像分类与识别是当前遥感应用领域的研究热点。传统的遥感影像分类方法从人工目视解译发展到人机交互解译，再到半自动解译，最后到当前基于机器学习模型和人工智能技术的全自动解译发展过程；影像分类模型则由传统的像元解译、局部结构特征提取发展到了面向对象识别的阶段；分类器也从单一分类器发展为层叠或多个分类器联合的方式^[3]。基于新兴理论提出的新技术、新方法在遥感影像分类与识别研究中取得了较好的识别效果，提升了影像识别的精度。然而，遥感影像存在混合像元，同物异谱和同谱异物等数据问题^[4]，影像数据固有不确定性成为影像分类亟需解决的问题，如果能建立恰当模型描述影像数据，进而提取目标地物特征信息，这将成为影像分类与目标识别的新思路^[5]。同时，因高分影像具有更详细的几何、纹理、形状等特征信息，同类地物类内特征差异大。且高分影像光谱波段较少，模型提取的地物特征有限，预测出清晰、明确的地物边界面临挑战。

深度学习（Deep learning, DL）^[6] 和全卷积分割网络（Fully convolutional network, FCN）^[7] 模型是近几年兴起并迅速发展的遥感影像分类方法。它能够自动从低阶特征中学习到复杂、抽象的高阶特征，从而更准确、高效的决策出分类结果。基于全卷积结构的影像分类方法相比传统分类方法具有更大的优势，被广泛用于遥感影像分类识别任务中。然而，FCN 分类模型在池化操作和反卷积上采样过程中会丢失影像边界、位置等低阶特征信息，导致模型分类预测的地物边界模糊、有歧义。此外，FCN 分类模型输出为各像素

点的类别概率，地物分割时容易产生许多细碎的错分区域，破坏分类结果中同类地物的完整性。

本文将从消除遥感影像地物分类混淆边界和保持分类结果中同类地物像素点的一致性两个角度，对高分影像地物分类问题展开研究，并改进现有的地物分类方法。首先将生成对抗网络中对抗训练的思想应用到全卷积分类模型中，借助生成网络优秀的图像生成能力和判别网络纠正生成结果与真实样本差异的特性，提出基于条件生成对抗网络的全卷积影像分类方法。然后，通过模糊聚类分割预处理方式得到影像同质性分割单元，在分割模型的高阶语义特征中融合对应尺度的边界掩膜特征，提出融合边界特征的影像分类方法，用于增强影像分类结果异类地物的边界识别能力。并得到更准确的地物分类边界。接着，基于像素点所属同质性分割单元内其他像素点的类别预测关系，更新像素点的类别预测概率，提出基于辅助信息后处理的影像分类方法，目的是减少分割结果中细碎的错分区域，确保同类地物分类结果的完整性。最后，综合这两种优化思路，改进基于条件生成对抗网络的影像分类方法，使得分类结果既有更清晰、准确的地物分类边界，又能确保同类地物类内像素类别一致性。本文课题来源于国家自然科学基金面上项目（11471045）和北京市自然科学基金（L172029）。

1.2 国内外研究现状

为了方便介绍，本文中将未深度学习技术前提出的遥感影像分类方法称为传统的遥感影像分类方法。本节内容主要介绍了传统的高分辨率遥感影像分类识别方法和基于深度学习技术^[6,8,9]的遥感影像识别分类方法的研究进展和现状。

1.2.1 传统的高分辨率遥感影像分类与识别方法

早在 1957 年，卫星遥感技术就应用到遥感影像分类与识别任务中。目标地物的分类与识别一直以来都是遥感影像分析中的一个基础任务，对于研究目标物体或现象的发展过程与分布规律有着重要意义^[10]。根据有无使用先验知识，遥感影像的分类方法可分为监督分类与非监督分类。监督分类是指利用样本已有先验类别训练分类模型，模型能够建立样本特征到类别标签的决策映射规则；非监督分类是指在没有类别先验知识的前提下，只能根据样本数据内在特性进行分类，根据样本间相似性度量分类，例如聚类^[11]。遥感影像分类方法还可基于分类单元不同，划分为基于像元和面向对象的分类方法。基于像元的分类方法将像元光谱特征作为主要分类依据，常见的基于像元的影像分类方法包括：最小距离分析法^[12]，最大似然分类法^[13]，K-均值聚类法^[14]，ISODATA 聚类法^[15] 和模糊 C

均值聚类方法^[16] 等。随着遥感技术迅速发展和不断成熟，影像数据空间分辨率持续增高。一般地，我们将空间分辨率高于 5m 遥感影像称作高分影像^[17]。高分影像相比低分影像来说光谱信息相对匮乏，而高分影像的几何、纹理等信息却更加丰富。基于像元的分类方法应用到高分辨率影像中会导致影像解译速度慢，同时椒盐噪声极易产生，因而其不适用于高分影像分类^[18]。

面向对象的高分影像分类方法将影像中邻域同质像元组成对象当作分类单元，充分利用影像地物的形状、纹理等特征，更适合高分影像分类与识别^[17]。早在 1976 年，Kettig 和 Robert^[19] 就将面向对象的思想引入遥感影像研究领域中。随后，Lobo 等人^[20] 将面向对象方法引入遥感影像分类领域，相关实验结论证明了在高分影像识别任务中面向对象方法比基于像元的方法识别速度更快，分类精度更高。Baatz^[21] 基于高分辨率遥感影像特性，系统地提出高分影像的面向对象分类方法。之后，面向对象分类方法被广泛应用于高分影像分类任务中，发展迅速。Geneletti^[22] 和 Guo^[23] 分别从非监督分类的研究方向表明面向对象分类方法是基于像元方法的有效替代。德国 Definiens 公司于 2009 年开发的 Ecognition 影像分析软件极大的推动了面向对象影像分类方法在工业商业领域的发展，同时也表明了面向对象的高分影像方法的成熟。

一般的，面向对象的遥感影像分类方法通常包含影像分割，特征提取和分类预测这三部分内容。Canny 通过提出 Canny 算子^[24] 检测出影像所有边缘点，并将边缘点依次连接形成边界从而实现影像边缘分割。Otsu 基于灰度直方图动态计算图像分割中的阈值范围，形成不同目标间差异最大化，实现阈值分割^[25]。Vincent^[26] 等结合沉浸模型提出影像的分水岭分割。Achanta 和 Radhakrishna^[27] 基于 K-均值聚类方法，采用简单的迭代聚类高效地生成影像分割单元，提出 SLIC 超像素分割算法，SLIC 分割效果被学界普遍认可。在特征提取阶段，最初采用影像的光谱、纹理和形状等低阶特征信息，但低阶特征无法获得较好的分类效果。文献^[28] 中引入词包模型的中层语义特征实现对遥感影像信息更好的表达，实验结果表明该方法分类效果更好。随后，Lienou 等^[29] 将主题模型应用到词包模型的单词语义分析中，改进了前者的分类精度。He 等人^[30] 针对遥感影像同物异谱的现象，结合模糊数学中不确定性理论，设计一种区间值特征对影像数据建模。目前，在特征提取的方面研究者做了大量工作，然而，高级特征的表达仍需要复杂的人工设计和反复实验验证。分类识别阶段是对特征提取得到的数据特征，利用分类器对原始数据决策识别。目前，常用的机器学习分类方法包含随机森林^[31]，支持向量机^[32]，决策树^[33] 和神经网络模型^[34] 等。在这些分类器基础上，通过结合不同分类器延伸而出的集成学习^[35] 的方法也被应用到高分影像分类中。

然而，传统的高分影像分类方法只能提取影像浅层特征，无法充分表达影像信息，而采用的影像分类器大多是只有1~2层的浅层结构模型，无法学习到遥感影像内部复杂特征。因此，探索结构更复杂、表征能力更强的分类模型具有重要的研究意义。

1.2.2 基于深度学习理论的影像分类研究现状

深度学习的概念源于人工神经网络，最早由 Geoffrey Hinton^[6] 教授于 2006 年提出。深度学习模型能挖掘数据低阶特征的内在规律，形成抽象的高阶特征或属性，从大量数据中建模数据内在规律。深度学习利用多层网络模型学习抽象概念完成自我学习^[36]。深度学习最早应用于图像处理领域，目前在自然语言处理语、音识别、搜索推荐、游戏 AI 和自动驾驶等领域广泛应用，且均表现出卓越的效果^[8]。2012 年，Alex 等人在 ILSVRC 图像识别大赛中提出了基于卷积神经网络 (Convolutional Neuro Network, CNN) 结构的 AlexNet^[9] 模型，对 ImageNet 数据集上千万级的自然图像分类识别，大幅提升了图像分类精度。AlexNet 的提出首次证明了 CNN 在复杂模型下的有效性，并极大推动了有监督深度学习领域的发展。在 2014 年 ILSVRC 大赛上，基于 CNN 结构，Google 研究团队提出的 GoogLeNet^[37] 和牛津大学学者提出的 VGGNet^[38] 分别荣获当年 ImageNet 识别大赛的一、二名。这两者在 AlexNet 的基础上均探索了网络深度与性能的关系，实验结果也证明了增加网络深度在一定程度上会影响网络最终的性能，使得分类错误率大幅下降。另外，GoogLeNet 中提出的 Inception 结构和 VGGNet 中提出的小卷积核多层网络结构也大幅优化了网络参数的数量，提升了训练学习速度的同时使得网络分类效果更优秀，且具有优秀的扩展泛化能力。之后，何凯明在 ResNet^[39] 网络模型中创造性地提出了残差学习 (Residual Learning) 的概念，解决了深度学习随着网络层数加深网络退化问题，使得更深层次网络模型得以训练，同时 ResNet 一并刷新了当年 ILSVRC 和 COCO 2015 图像识别大赛的最优记录。在非监督学习领域，深度学习模型近十年也发展迅速。2006 年，Hinton 对传统自动编码器结构进行改进，提出了深度自编码网络 (Deep AutoEncoder, DAE)^[6]。DAE 网络利用无监督逐层贪心训练算法完成对隐含层的预训练，然后用 BP 算法对整个网络参数进行调整，显著降低了深层自编码结构的性能指数，且大幅提升自编码器的学习能力。之后，基于 DAE 理论相继提出的栈式自动编码器 (stacked AutoEncoder, stacked AE)^[40]、降噪编码器 (Denoise Autoencoder, dAE)^[41] 和稀疏自编码器 (Sparse AutoEncoder, SAE)^[42] 等均取得了不错的效果。2014 年，Goodfellow 结合二人零和博弈的思想，创造性地提出了生成对抗网络 (Generative Adversarial Network, GAN)^[43] 模型，极大地促进了生成模型和计算机视觉领域（如图片生成、风格迁移和图像分割等）的发展。GAN 模型

框架由两个“对抗”模型组成：捕获数据分布的生成模型 G 和估计样本来自训练数据而不是 G 的概率的判别模型 D 。随后，基于 GAN 网络的一系列生成模型方法如 CGAN^[44]、DCGAN^[45]、InfoGAN^[46] 和 WGAN^[47] 等被相继提出，不仅提升 GAN 模型生成与识别精度，同时极大丰富了 GAN 网络的应用场景。

由于深度学习在图像分类识别的巨大成功与广泛应用，研究学者逐渐将深度学习理论引入遥感影像分类，基于深度学习理论的研究方法逐渐成为遥感影像发展的下一个趋势。文献^[48] 利用迁移学习知识，首次将深度卷积神经网络应用到高分影像场景分类中，能有效学习影像的高级特征表示。Marco 等人^[49] 将预训练的 GoogLeNet 网络参数，迁移到 UC Merced 土地利用数据集^① 上，其提出的方法在 UC Merced 数据集上获得了 10% 的分类识别精度提升，实验结果也表明了 CNN 结构在遥感影像上的成功。2016 年，Romero 等人^[50] 使用贪婪分层无监督预训练，结合稀疏表示理论，实现对高分影像土地利用和土地覆盖的无监督分类。Kampffmeyer 等^[51] 则使用 CNN 结构量化遥感影像像素尺度上的不确定性，对图像上每个像素进行分类，完成遥感影像的类别分类和语义分割。文献^[52] 基于全卷积分类网络（Fully convolutional network，FCN），将 CNN 模型中的全连接层全部替换为卷积层，模型输出影像所有像素点类别，实现遥感影像的像素级分类。U-Net^[53] 网络结合反卷积与跳跃网络的优势，对 FCN 结构加以改进。文献^[54] 基于 U-Net 网络完成对海陆影像水域 - 陆地分割识别。文献^[55] 则在 U-Net 基础上结合残差学习的思想，完成对遥感影像道路信息的提取。

1.3 本文的组织结构

本文将从消除遥感影像地物分类混淆边界和保持分类结果中同类地物像素点的一致性两个角度对高分影像分类方法展开研究。论文依据研究内容可划分为五个章节，各章节依次为：

第 1 章：系统地介绍了论文课题相关的研究背景、研究意义和国内外现状。重点对当前遥感影像分类中面临的挑战展开介绍，并针对这些问题，提出了改进方案，引出本课题主要的研究内容。

第 2 章：本章详细介绍了深度学习相关理论。介绍了卷积神经网络、全卷积神经网络和生成对抗网络这三个深度学习模型的结构和原理。然后介绍了基于全卷积网络的遥感影像分类方法。

第 3 章：针对现有全卷积影像分类方法的缺陷与不足，将条件生成对抗网络结构引入

① 数据集访问链接：<http://weegee.vision.ucmerced.edu/datasets/landuse.html>

全卷积分类模型中，提出基于条件生成对抗网络的影像分类方法。并详细介绍了新提出方法的模型原理和算法实现。在高分影像分类识别实验上验证新提出方法的有效性。

第 4 章：首先利用面向对象模糊聚类分割方法得到高分影像的同质性分割单元。再从消除地物分类混淆边界和减少细碎错分区域两个角度分别提出融合边界特征的影像分类方法和基于辅助信息后处理的影像分类方法。最后通过量化和目视实验结果验证新提出的改进方法的有效性。

第 5 章：总结本论文课题的研究内容，归纳文中研究成果，并对论文中存在的一些问题提出未来的展望。

1.4 本文主要创新点

本文研究内容主要有三个创新点：

- (1) 将生成对抗训练的思想应用到全卷积影像分类模型中，提出基于条件生成对抗网络的影像分类方法，提高影像分类精度。
- (2) 针对分割模型中池化、上采样丢失影像边界、位置特征的问题，在分割模型高阶语义特征图中融合预处理得到的边界掩膜特征信息，提出融合边界掩膜特征的影像分类方法，改进地物分类存在的歧义、模糊边界问题。
- (3) 对分割模型像素级预测的概率值进行后处理，考虑像素点同质性分割单元内其他像素点的类别关系，提出基于辅助信息后处理的影像分类方法，提升影像分类地物的完整性，减少细碎的错分区域。

第 2 章 深度学习相关理论

深度学习是以人工神经网络为架构，通过多隐层网络结构提取高层抽象特征，对数据进行表征学习的算法。常见的深度学习框架包含深度神经网络（Deep neural network, DNN）、深度置信网络（Deep belief network, DBN）、递归神经网络（Recurrent neural network, RNN）和卷积神经网络（Convolutional neural network, CNN）等^[56]。与其他网络结构相比，CNN 利用输入数据的二维结构信息，在图像分类与识别、图像语义分割等视觉任务中能够给出更好的结果。

2.1 卷积神经网络基础

CNN 是一种带有卷积结构的深度前馈神经网络，使用反向传播（Backpropagation, BP）算法进行权值更新，主要用于图像分类与识别。CNN 网络结构一般由卷积层（Convolutional layer）、池化层（Pooling layer）和全连接层（Full connected layer）交叉堆叠而成。卷积层和池化层用于提取图像高阶特征，具有局部连接、权值共享和下采样等特性。全连接层可以看作 CNN 分类网络的决策分类器，将学到的“分布式特征表示”映射到样本标记空间。。

1. 卷积层

卷积（Convolution）是分析数学中的一种运算，被广泛应用到信号处理与图像处理中。因为图像是两维结构数据，图像处理中常用二维卷积运算。给定一个图像 $X \in \mathbb{R}^{M \times N}$ ，和滤波器 $W \in \mathbb{R}^{m \times n}$ ，一般满足 $m \ll M, n \ll N$ ，其卷积输出：

$$Y = X \otimes W \quad (2-1)$$

式中， \otimes 是卷积运算。输出特征图上某点 (i, j) 的值可由式2-2计算得到：

$$y_{ij} = \sum_{u=1}^m \sum_{v=1}^n w_{uv} \cdot x_{i-u+1, j-v+1} \quad (2-2)$$

CNN 最核心结构是卷积层，每个卷积层均包含一个或多个二维平面，该二维平面称作 CNN 网络的特征图 (Feature map)。卷积层中的神经元共享权重，CNN 中共享的权重称为卷积核 (Convolutional kernel)，卷积核对上层特征图进行卷积操作提取影像特征。如

图 1 所示，图中使用大小为 3×3 的卷积核，输入图像大小为 5×5 ，卷积核从左上方以 1 个像素点的步长 (Stripe) 开始滑动，图像中像素点与卷积核之间卷积运算后的输出到特征图上的对应位置。

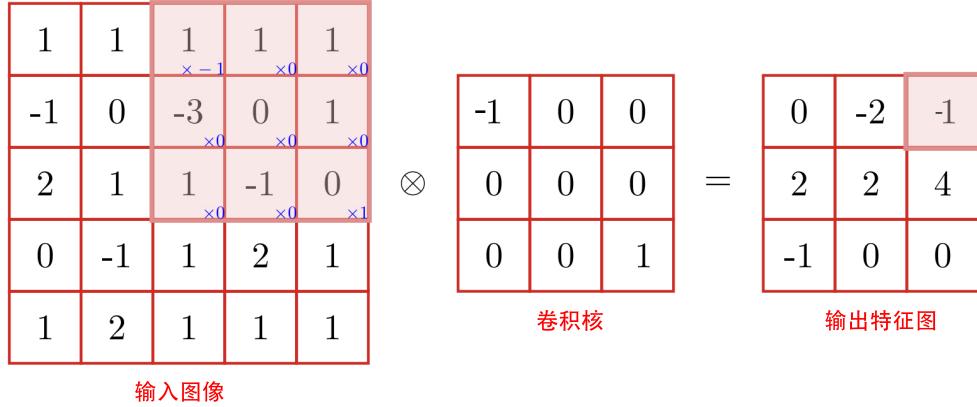


图 1 图像卷积示意图

图像卷积运算时，与全连接模式中每个像素点对应一个独立的计算权值不同，所有位置的输出神经元均使用一个卷积核计算权值，这样做的好处是既减少参数的数量，又利用了图像特征的局部性特点。图像卷积中，每个卷积核都可以提取图像的一种特征，所以一般设置多个卷积核，利用不同的卷积核来提取图像不同的特征。

卷积计算中第 l 层的输入是第 $l-1$ 层的卷积输出特征图，假定第 $l-1$ 层输入大小为 $X_{l-1} \times X_{l-1}$ ，卷积核大小为 $K \times K$ ，滑动步长为 S ，边界填充 (Padding) 大小为 P ， $l-1$ 层的输出特征图（即第 l 层的输入）的大小为 $X_l \times X_l$ ，可由式2-3 计算得到：

$$X_l = \lfloor \frac{X_{l-1} - K + 2 \times P}{S} \rfloor + 1 \quad (2-3)$$

卷积操作是线性的，即使联合多个线性模型输出依旧是线性的。实际识别任务往往是复杂、非线性的，线性神经网络实用性不强。如果为每个卷积输出加一个非线性函数，神经网络模型就不再是线性的，这个非线性函数就是激活函数。常用的激活函数有 Sigmoid、TanHyperbolic(tanh) 和线性整流函数 (Rectified linear unit, ReLU) 函数等。当前卷积网络常用的是 ReLU 激活函数，ReLU 函数表达式如下：

$$f(x) = \max(x, 0) \quad (2-4)$$

相比 Sigmoid 和 tanh 函数，ReLU 对于梯度训练的收敛有巨大加速作用，反向传播训练时不易饱和，另外 ReLU 函数简单，运算量很小，因此广泛使用到卷积网络非线性激活中。

2. 池化层

池化层又叫下采样层 (Subsampling layer)，能够缩小特征图尺寸，其作用是进行特征选择，降低特征数量，使特征更加抽象。卷积层虽然可以大幅减少网络参数的数量，但并没有显著减少特征映射组中神经元的个数。在卷积层后面加上一个池化层，可以减少网络下一层的数据量，实现特征降维，一定程度上防止过拟合。

假定池化层的输入特征图为 $X \in \mathbb{R}^{M \times N}$ ，将其划分为很多区域 $R_{m,n}, 1 \leq m \leq M, 1 \leq n \leq N$ ，这些区域可以重叠，也可以不重叠。池化层通过池化函数对特征图的每个区域 $R_{m,n}$ 进行下采样，得到一个值作为这个区域的概括。根据采样的方式不同，常用的池化函数有两种：

1. 最大池化 (Maximum pooling) 最大池化是将区域内所有神经元的最大值作为池化层输出。

$$Y_{m,n} = \max_{i \in R_{m,n}} x_i \quad (2-5)$$

其中， x_i 为区域 R_k 内每个神经元的激活值。

2. 平均池化 (Average pooling) 平均池化是选取区域内所有神经元激活值的和的平均值。

$$Y_{m,n} = \frac{1}{|R_{m,n}|} \sum_{i \in R_{m,n}} x_i \quad (2-6)$$

对输入特征图的所有区域进行下采样，就可以得到池化层的输出特征图 $Y = \{Y_{m,n}\}, 1 \leq m \leq M, 1 \leq n \leq N$ 。

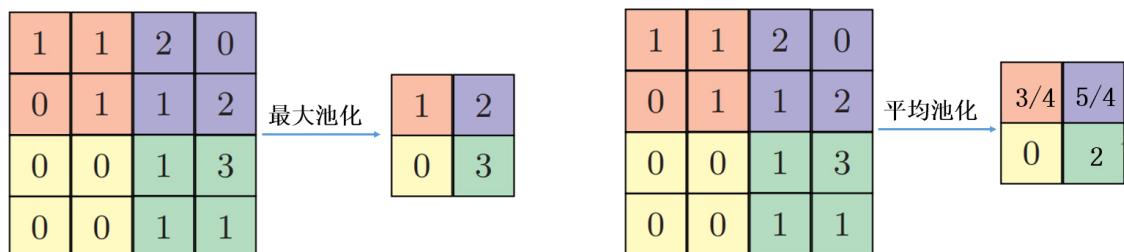


图 2 最大池化与平均池化

图如 2 所示，输入特征图大小为 4×4 ，池化函数核为 2×2 ，滑动步长为 2，分别进行最大池化和平均池化操作，得到不同的池化输出特征图。

在卷积网络的最后，往往会接入一两层的全连接层。全连接层将卷积输出的二维特征“拍平”，转化成一个一维向量。全连接层对卷积输出特征高度提取，方便交给最后的分

类器（如 SVM 分类器、Softmax 分类器等）实现图像分类。

3. 典型的卷积网络结构

一个典型的卷积网络由卷积层、池化层和全连接层交叉堆叠而成。如图 3 所示，通常一个卷积块由连续的 M 个卷积层和 b 个池化层拼接而成（ M 一般取 $1 \sim 4$, b 可取 0 或 1），一个完整的用于分类任务的卷积网络结构通常由 N 个堆叠的卷积块后接 K 个全连接层组成（ N 一般取 $1 \sim 100$, K 取 $1 \sim 2$ ）。CNN 中多层网络模型使得网络模型对不同形态的图像具有优秀的适应能力，它可以自动学习高分影像中复杂的特征，提升影像分类精度。

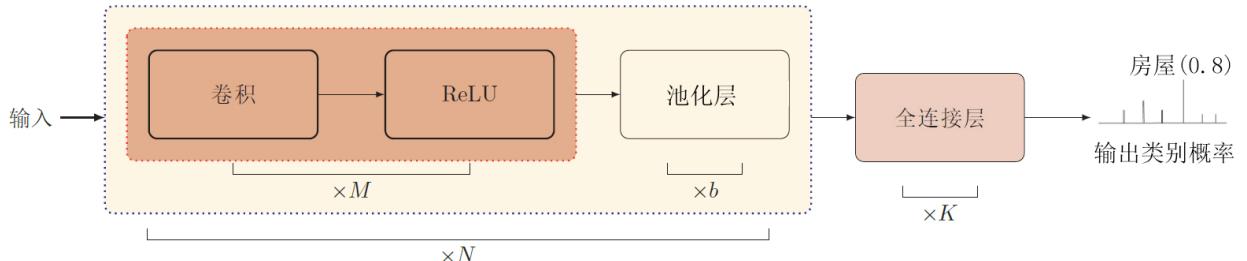


图 3 典型的卷积网络结构

2.2 全卷积神经网络

经典的 CNN 模型最右端结构通常为全连接层，它会将原来二维影像映射到一维空间，因而会失去二维图像的空间位置信息，且网络输出常为输入图像属于某一类别的概率。与分类对整张图片类别预测不同，图像的语义分割是对目标图像的每个像素点进行分类，即语义分割是从像素级对图像进行类别预测。全卷积神经网络（Fully convolutional network, FCN）^[7] 由 Jonathan Long 等人于 2016 年提出，其创造性地利用卷积层替代分类网络中的全连接层，进而保证网络输出为二维的分类结果图，使用反卷积（Deconvolution）的上采样策略，得到一个与原图尺寸大小相同的分割图，实现图像像素级的分类预测。图像语义分割引入遥感分类领域即为遥感影像分类。图 4 为遥感影像分类的全卷积网络结构示意图，全卷积网络将学习到的遥感影像判别特征解码映射到高分辨率空间，完成影像的像素级分类。

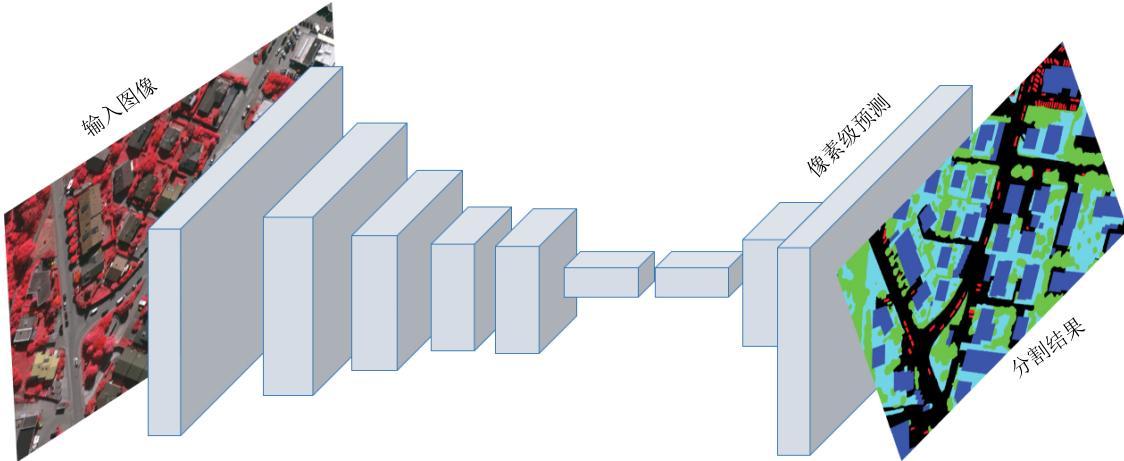


图 4 影像分类网络结构示意图

1. 反卷积

反卷积，又叫转置卷积，是一种上采样操作，可以理解为下采样的逆过程。卷积运算是一个下采样过程，一般通过卷积操作实现高维特征到低维特征的转换。如对输入 4×4 的二维特征，用大小 3×3 的核，做步长为 1 的卷积运算得到 2×2 的特征输出。反卷积则实现低维特征到高维特征的转换。与之对应，反卷积对输入为 2×2 的二维特征，使用 3×3 的核操作得到 4×4 的输出。如图 5 所示，输入特征大小 2×2 ，核大小 3×3 ，步长 $s = 1$ ，填充补 0 为 $p = 2$ ，经过反卷积处理输出尺寸上采样到 4×4 ，图中显示了反卷积上采样的计算过程。

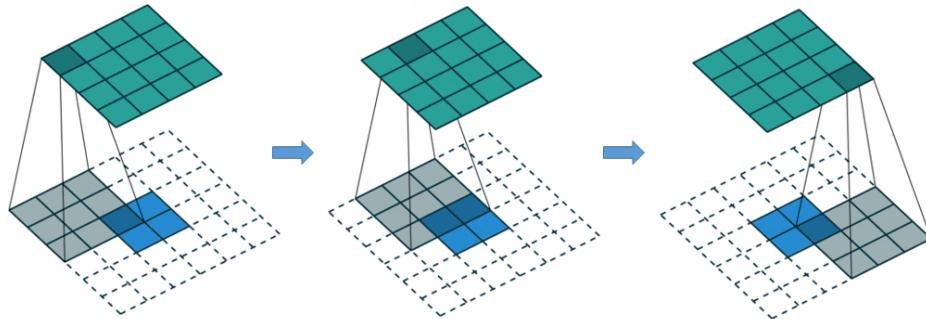


图 5 反卷积上采样过程

反卷积是卷积运算形式上转置的映射关系。对一个大小为 $X \times X$ 的图像 I ，和大小为 $K \times K$ 的卷积核，进行步长为 $S \geq 1$ 的反卷积运算，先对图像 I 进行两端补零 $P = K - 1$ ，并且在每两个像素间填充 $S - 1$ 个 0，最后进行步长为 1 的卷积操作，得到反卷积的输出。

结果，设输出特征维度为 $O \times O$ ，满足：

$$O = S \times (X - 1) + K \quad (2-7)$$

2. 基于 FCN 影像分类

FCN 网络由卷积特征提取和反卷积上采样两部分组成。FCN 特征提取阶段，为了加快网络训练速度，常使用在 ImagaNet 等数据集上训练好的网络权值初始化 FCN 网络参数，如对训练好的 AlexNet^[56] 或 VGG^[38] 网络权值，选取除全连接层的权值参数初始化 FCN 网络。输入图片经卷积池化层处理后特征图尺寸变小，所以特征图需要被上采样为输入图片相同尺寸。FCN 使用反卷积层做上采样将特征图尺寸调整为原输入图像大小。同时，为了得到更精细的分类结果，FCN 中使用跳层连接（Skip connections）将下采样阶段和上采样阶段相同尺寸的特征图融合。图 6 为基于 FCN 网络结构影像分割示意图。图中虚线上半部分为卷积池化网络结构，模型使用训练好的 VGG 16 网络权值（去除全连接层权值）初始化，堆叠的卷积层与池化层能够确保提取影像数据高阶特征。图中虚线下半部分，分别从卷积网络的不同阶段预测网络的分类结果，利用反卷积层对最后一个卷积层特征图上采样处理，使特征图尺寸还原为原始输入影像大小，保留了原始输入图像中的空间信息，从而对每一个像素都产生了一个预测，最后在上采样的特征图上进行逐像素分类，完成 FCN 网络的图像语义分割。

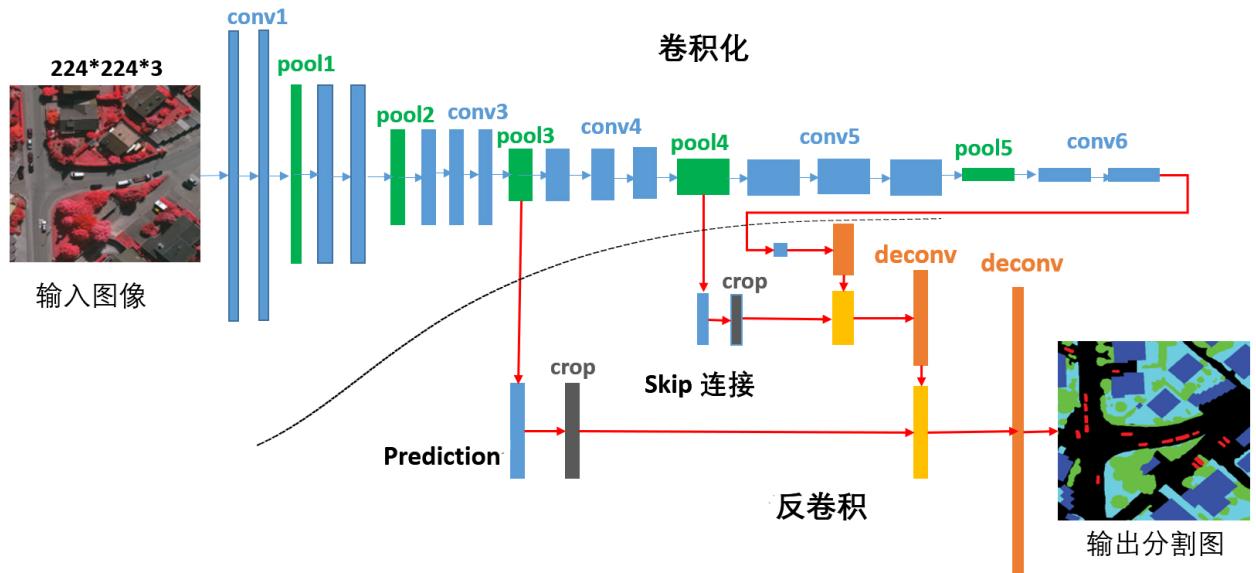


图 6 基于 FCN 网络的遥感影像分类示意图

FCN 有三个特点，分别是全卷积化、反卷积上采样和跳跃连接。全卷积化将传统

CNN 网路中的全连接层全部转化为卷积层。全连接层由于神经元个数固定，与上层连接特征图的权值也是固定，从后到前一层一层地反向推导，可知 CNN 模型的输入影像尺寸也必须相同，所以去除全连接层的 FCN 网络可以接受任意尺寸的图像。此外，卷积操作将全连接层输出由一维映射为二维特征矩阵，便于模型后半部分对特征图上采样处理。反卷积上采样操作能将特征图还原为原始影像大小，保留图像空间位置信息，实现影像的像素级分类。跳层连接融合卷积网络不同深度的特征输出，将影像低阶细节特征和高阶语义特征进行融合，能够得到更精细化的分类结果。

2.3 生成对抗网络概述

前面介绍的 CNN 和 FCN 分类模型同属深度学习理论的判别式模型，判别式模型主要思想是根据原始样本特征决策判别样本具备的性质，例如根据影像特征判别影像包含地物所属分类。与之对应的是生成模型，生成模型是对输入数据的分布规律建立模型，进而用新模型生成样本数据。随着深度学习和神经网络近几年的迅猛发展，基于深度学习的生成模型取得了重大的突破，研究基于深度学习的生成模型也具有重大意义。2014 年 GoodFellow 等人从对抗博弈的角度提出一种名为生成对抗网络（Generative Adversarial Network, GAN）^[43] 的生成模型，该模型由生成器和判别器两部分构成。生成器试图学习潜在的数据分布规律，生成新样本；判别器则决策当前数据源为真实样本还是生成样本。相关研究^[44] 证明相比其他生成模型（如自编码网络等），GAN 生成的样本更加逼真。

2.3.1 生成对抗网络模型

GAN 模型思想启发于博弈论中的“纳什均衡”，即模型训练时不断对抗博弈，模拟数据分布的生成式模型。GAN 模型中生成器目的是生成符合数据潜在分布特征的样本，尽可能欺骗判别器；判别器目的是判断网络输入是生成样本还是真实数据，不被生成器愚弄。模型交替训练时，判别器尝试区分输入样本的来源，当判别器发现真实样本与生成样本存在差异时，模型优化生成器的权值参数，消除样本分布差异，生成器样本生产能力不断提高。而判别器优化过程则是不断提高判别器对样本的判别能力。最终模型收敛时两者达到一个“纳什均衡”，此时生成器能拟合数据的分布生成逼真样本。

如图7 所示，分别用符号 G 和 D 表示 GAN 模型生成器与判别器。假定变量 z （通常为服从高斯分布的随机噪声）通过 G 生成 X_{fake} ， D 需要决策模型输入数据是真实样本

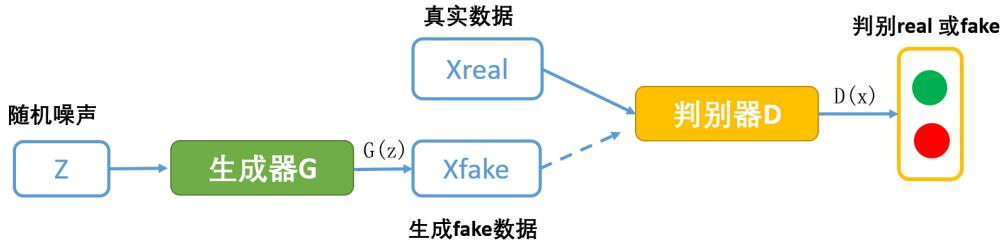


图 7 GAN 结构示意图

X_{real} 还是生成的样本 X_{fake} 。GAN 模型需要优化的目标函数为2-8:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (2-8)$$

式中 $x \sim p_{data}(x)$ 表示 x 取自真实的分布数据。求解 D 等价于求解二分类模型, $V(D, G)$ 即为二分类求解常见的目标代价函数。G 需要最大化生成结果的分类决策概率 $D(G(z))$ 干扰 D 的决策判断, 即最小化 $\log(1 - D(G(z)))$ 。

模型训练时, D 和 G 交替迭代训练, 即先固定 G 训练 D, 再固定 D 训练 G, 不断迭代, 直到模型收敛。对于 G, 最小化 $\max_D V(D, G)$, 即最小化 $V(D, G)$ 的最大值。当固定 G 时, 对代价函数 $V(D, G)$ 求导, 解出最优的判别器 $D^*(x)$ 如式2-9:

$$D^*(x) = \frac{p_g(x)}{p_g(x) + p_{data}(x)} \quad (2-9)$$

文献^[43] 中指出, 当多次往复训练后, 模型会收敛, G 与 D 达到“纳什均衡”, $p_g(x) = p_{data}(x)$, 即判别器对生成样本和真实样本的预测概率均为 $\frac{1}{2}$, 无法区分。表明生成器已经学习到数据的内在分布, 能够生成逼近真实样本的数据。

2.3.2 条件生成对抗网络

经典的 GAN 模型是无监督模型, 其生成器的输入为随机噪声 z , 通常只能生成逼近真实样本的同类型数据, 不能生成我们想要的某一种类型的数据, 例如无法根据影像原始图, 生成器生成影像分类结果图。文献^[44] 针对上述问题提出条件生成对抗网络 (Conditional Generative Adversarial Networks, CGAN), CGAN 模型生成器输入加入条件约束 y 引导模型迭代方向, 即将先验条件约束 y 和随机噪声 z 联合作为生成器的输入样本, 生成我们需要的目标类型数据。其中, y 可以是任何种类的辅助信息, 如类别标签, 影像真实 Ground truth 图或其他不同领域模态的数据等。

CGAN 模型中, G 中随机噪声 z 与先验知识 y 拼接组成 G 阶段联合输入特征; D 中将

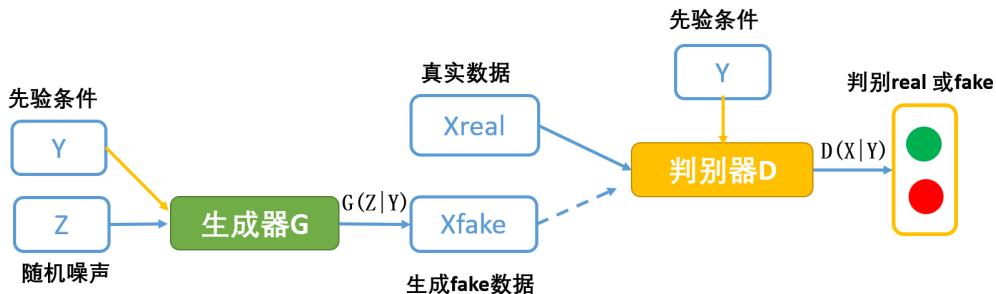


图 8 CGAN 结构示意图

真实样本/生成样本 X 和 y 通过拼接共同作为判别模型输入。类似式2-8 中 GAN 的目标函数, CGAN 模型目标函数为带有条件概率的二人极小极大值博弈函数, 即:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x \sim p_{data}(x)}[\log D(x|y)] + E_{z \sim p_z(z)}[\log(1 - D(G(z|y)))] \quad (2-10)$$

式2-10中, $D(x|y)$ 为判别器 D 对有条件约束 y 的真实样本判别为真的代价函数, $D(G(z|y))$ 为 D 对有条件约束 y 的生成样本 $G(z|y)$ 判别为假的代价函数。

图8 为 CGAN 的结构示意图, 通过将额外条件信息 y 分别输送给判别模型和生成模型组成联合隐层表征, 作为输入层的一部分, 从而指导数据的生成过程。

2.4 本章小结

本章介绍内容为论文研究内容理论基础, 围绕深度学习领域内的三大模型展开介绍。首先介绍了经典 CNN 模型的结构和原理, 包含 CNN 卷积池化、权值共享等特性。接着介绍了 FCN 模型的结构以及 FCN 在遥感影像语义分割中的应用, 重点介绍 FCN 全卷积化与反卷积上采样的处理方式。最后, 介绍了 GAN 模型和带有先验约束条件的 CGAN 模型的原理和结构。生成对抗网络在遥感影像分类中的应用将在后文中详细介绍。

第3章 基于生成对抗网络的遥感影像分类方法

3.1 引言

深度学习模型组合低阶特征提取更加抽象、复杂的高阶特征，具有高效的特征学习能力，因而取代了传统的影像分类方法，成为当前遥感影像分类领域的主要研究方法。第2章概述了深度学习领域内重要的几个网络模型，并详细介绍了基于FCN的遥感影像分类方法的结构和原理。一方面，FCN中的卷积结构能自动学习输入图像的潜在特征；另一方面，FCN反卷积结构将特征图恢复到原始图像尺寸，实现遥感影像的像素级分类。然而，FCN模型上采样过程造成了特征信息的损失，导致网络预测地物边界模糊问题。此外，高分影像细节特征复杂且FCN分割方法是影像像素点级的分类，使得预测结果中同类地物内部像素点难以保持类别一致性。进一步加剧了影像分类中地物边界模糊、歧义性等问题。

第2.3节介绍的GAN模型是一个优秀的生成模型，其生成器和判别器不断对抗博弈，收敛时的GAN模型生成器有强大的数据建模能力，判别器则会检测并纠正真实样本与生成数据之间的差异，保证整体的一致性。本章将对抗网络训练的思想引入到FCN模型中，将FCN模型像素级多分类交叉熵损失与GAN对抗目标代价结合。对抗博弈中将促使分割模型生成的地物分类标签和真实地物Ground truth图难以被判别器区分，从而提升同类地物内像素点类别的一致性。此外，对抗损失作用于整个影像数据，使得在几乎不增加模型复杂度的前提下提升影像远距离像素点间类别标签的连续性。

3.2 基于CGAN的影像分类方法

将对抗训练的思想应用到FCN影像分类模型中，提升分类结果图中同类地物像素点的类别一致性。对抗网络的生成器需要预测原始影像的分割结果，这里原始影像作为先验条件输入对抗网络模型。因此，本章将有条件约束的对抗网络思想应用到FCN影像分类模型中，提出基于CGAN的遥感影像分类方法，其中原始影像作为CGAN的先验约束条件。

3.2.1 算法原理

基于 CGAN 的影像分类方法模型主要包含两个部分：生成网络的影像分割模型和对抗训练阶段的判别模型。整个模型结构如图9 所示，左边是生成网络，是 FCN 分割模型，其卷积特征提取结构使用 VGGNet 16^[38] 前五层的卷积结构，由五个卷积块堆叠而成，每个卷积块包含 2/3 个卷积层和 1 个最大池化层；分割模型的特征图恢复部分为 4 个反卷积的上采样结构，每个上采样模块中包含一次反卷积操作，将特征图尺寸放大两倍；最后通过一个 1×1 的卷积结构将特征图通道数降为 1，生成影像分割结果。右边是判别网络，判别网络是一个二元 CNN 分类模型，其输入存在两种情况：一种是原始影像和 Ground truth 图拼接输入，另一种是原始影像与分割模型生成结果拼接输入。判别网络结构由两个卷积结构和两个全连接层堆叠，最后通过 Sigmoid 激活函数判别当前输入样本来源，输出为一个二元分类值（输出为 1 代表它判断输入是第一种情况，输出为 0 代表它判断输入是第二种情况）。

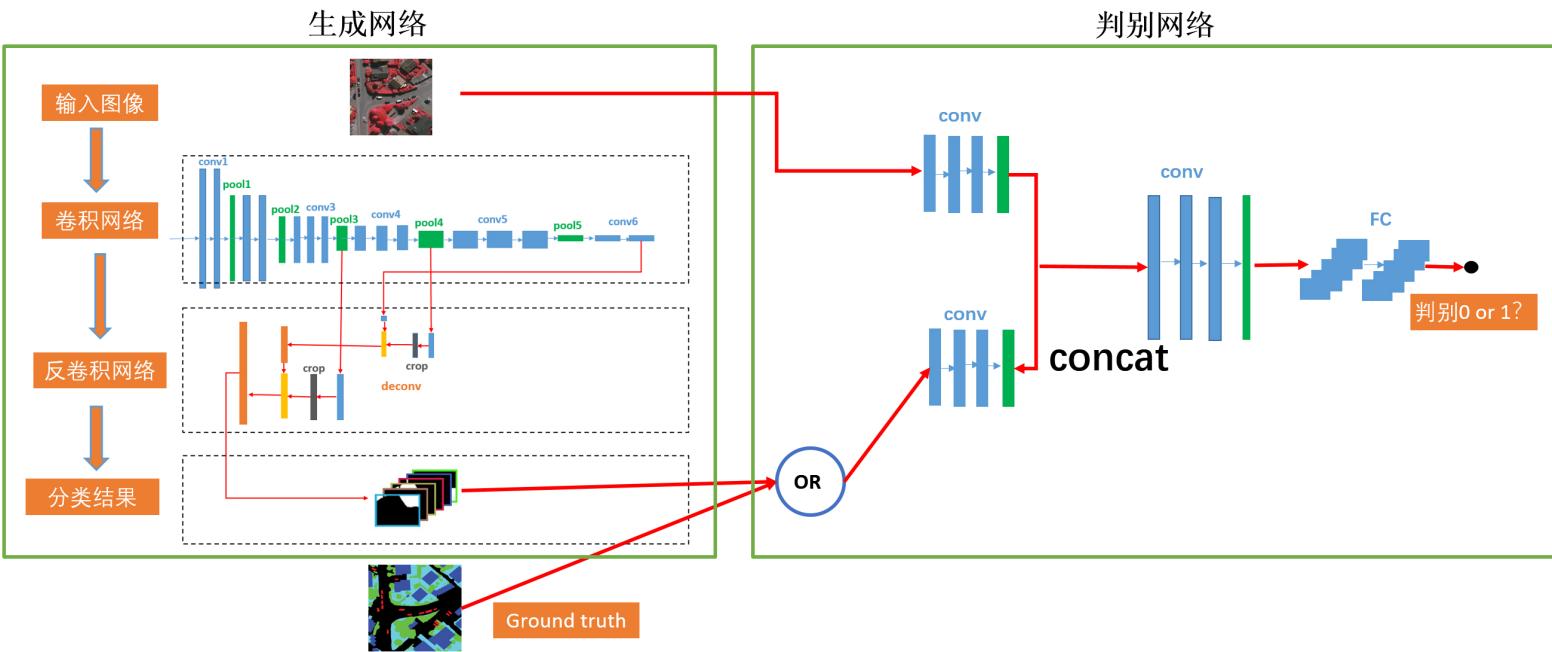


图 9 基于 CGAN 的影像分类方法模型示意图

模型中使用二元分类损失（Binary classification loss, BCE）来度量判别网络。二元分类损失为二元交叉熵（cross entropy）函数，统计学中使用 KL 散度衡量两个事件或分布中的不同，常用于计算代价损失，图像分类任务中最小化 KL 散度等价于最小化交叉熵，而交叉熵计算更简便^[57]，因此文中用交叉熵来计算二元分类损失。交叉熵函数为分类预

测概率值的负对数。二元交叉熵代价表达式如下：

$$l_{bce}(\hat{z}, z) = -[z \log \hat{z} + (1 - z) \log(1 - \hat{z})] \quad (3-1)$$

生成网络是全卷积网络模型，其本质上是一个 CNN 影像多分类模型。经典的多分类分割模型代价函数为多类别的交叉熵损失（multi-class entropy loss, MCE），分割模型输入为大小为 $H \times W \times C$ 的图像，对图像做像素级预测分类，其多元交叉熵损失为：

$$l_{mce}(\hat{y}, y) = - \sum_{i=1}^{H \times W} \sum_{c=1}^C y_{ic} \log \hat{y}_{ic} \quad (3-2)$$

其中， H 、 W 和 C 分别为图像的高度、宽度以及通道数。假定多分类问题中类标有 K 个取值，需要使用独热编码（One-hot encode）将图像类别标签编码为一个 K 维向量，借助 Softmax 函数作为分类任务的输出层。Softmax 函数把神经网络分类输出转化为一组概率，且这组概率和为 1。归属于类别 j 的概率为：

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \forall j \in 1, 2, \dots, K \quad (3-3)$$

假定有 N 张训练图片的数据集 $X = \{x_1, x_2, \dots, x_n\}$ ，对应的 Ground truth 图标签集为 $Y = \{y_1, y_2, \dots, y_n\}$ 。输入图像大小为 $H \times W \times C$ （一般 C 取 3），有 K 个类别的图像标签编码为 K 维向量，即对第 i 张图像 x_i ，对应 Ground truth 图有 $y_j = [y_j^{(1)}, y_j^{(2)}, \dots, y_j^{(K)}]$ 。我们定义 $g(x)$ 为输入图片 x 在生成网络分割模型下的预测输出，定义 $d(x, y) \in [0, 1]$ 为对抗网络判别图像标签 y 是输入图像 x 对应的 Ground truth 图还是分割模型产生的分割结果的概率输出值。基于条件对抗网络框架的全卷积语义分割模型方法需要最小化分割模型的多分类交叉熵损失，同时最大化分割模型生成样本的判别概率。优化的目标代价函数为式3-4：

$$L(\theta_g, \theta_d) = \sum_{n=1}^N \{l_{mce}(g(x_n), y(n)) - \lambda [l_{bce}(d(x_n, y_n), 1) + l_{bce}(d(x_n, g(x_n)), 0)]\} \quad (3-4)$$

式中 θ_g 和 θ_d 分别代表分割模型和判别模型网络参数，右式中第一项为分割模型的交叉熵损失，第二项为对抗网络中样本判别损失代价函数， λ 为生成阶段与对抗阶段代价权衡常数且有 $\lambda > 0$ 。模型训练时我们需要最小化式3-4 中的代价函数，从而迭代求解得到网络参数 θ_g 和 θ_d 。

当模型迭代收敛后，模型中的权值参数均得以确定，此时模型中生成网络部分已经具有优秀的图像分割识别能力。通过前向传播，利用最终分割模型对待分类测试影像求解每个像素属于各个类别的概率值，像素点所属类别概率值求采用式3-3 中 Softmax 函数求解，

接着根据 argmax 函数计算像素点最大概率值对应类别维度，即为当前预测点的分类标签，影像 x 上任一像素点 i 所属类别标签 C_k 的计算方式如式3-5。

$$C_k = \arg \max_{k \in K} p_k(x^{(i)}), k = 1, 2, \dots, K \quad (3-5)$$

类似地，对待分类影像所有像素点预测类别标签，即可实现该影像的分割。

3.2.2 算法实现

基于 CGAN 的遥感影像分类方法需要交替训练生成网络和判别网络两部分，算法采取小批次的随机梯度下降方法优化 CGAN 权值参数，Adam^[58] 优化器可以自适应调整学习率的大小，因此本算法使用 Adam 优化器更新模型权重。基于 CGAN 的影像分类模型更新迭代伪代码如**算法 1** 所示。

模型首先将影像数据划分为训练集和测试集，并赋值给对应变量。初始化 CGAN 分割模型的超参数。在每次迭代内，分批次随机抽取生成分割图和 Ground truth 图，分别与对应的原始影像对应拼接，联合输入判别模型，计算判别模型损失并更新判别模型相关权值参数。接着固定判别模型训练生成网络，计算模型损失并更新生成器参数权重，完成网络的一次迭代。模型收敛后即用生成分割模型对测试集影像进行处理，得到模型分割结果图。

3.3 实验数据介绍与预处理

3.3.1 Vaihingen 数据介绍

本章实验数据源为 ISPRS（国际摄影测量及遥感探测学会）提供的 Vaihingen 高分辨率遥感影像数据^①。Vaihingen 数据集由德国测量和遥感协会（DGRF）于 2010 年使用空中数字摄像机拍摄，拍摄区域为半农村地区的德国斯图加特市法伊欣根市镇（Vaihingen），影像空间分辨率为 $0.09m$ 。如图10(a) 所示为 Vaihingen 数据整体图，整个数据被划分为 33 幅大小不一的图像，其中带有真实地面数据的影像为 16 幅。Vaihingen 数据包含三个波段（近红外 - 红 - 绿）的正射影像数据（Orthophoto）、数字地表模型（Digital Surface Model, DSM）数据以及拍摄区域真实地物分类结果（Ground truth）。图10(b-d) 分别为该影像某一区域的正射影像、DSM 高程、Ground truth 图。

^① Vaihingen 数据集官网链接：<http://www2.isprs.org/commissions/comm3/wg4/detection-and-reconstruction.html>

算法 1 基于 CGAN 的遥感影像分类方法伪代码

输入: 原始影像 X ; 影像标签 Y ;

输出: 生成样本;

模型训练:

1: VGGNet16 前五层权值初始化生成模型, 判别器步数 $k = 1$, 权衡因子 λ , 初始学习率 α , Adam 动量 β_1, β_2 。

2: **for** 模型迭代 epoch 数 **do**

3: **判别器:**

4: **for** k 步 **do**

5: 小批次抽样 m 张影像 $\{x^1, x^2, \dots, x^m\}$; 对应影像的真实标签 $\{y^1, y^2, \dots, y^m\}$; 这 m 张影像在生成器中的输出 $\{g(x^1), g(x^2), \dots, g(x^m)\}$

6: 使用 Adam 优化器更新判别器的权重:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log(d(x^i, y^i)) + \log(1 - d(x^i, g(x^i)))]$$

7: **end for**

8: **生成器:**

9: 小批次抽样 m 张影像 $\{x^1, x^2, \dots, x^m\}$;

10: 分割模型输出这 m 张影像分割结果 $\{g(x^1), g(x^2), \dots, g(x^m)\}$

11: 使用 Adam 优化器更新生成器的权重:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m [\log(1 - d(x^i, g(x^i)))]$$

12: **end for**

模型预测:

13: 模型迭代收敛后, 用生成器分割模型对测试影像预测分割结果, 采用概率最大化预测类别标签:

$$C_k = \arg \max_{k \in K} p_k(x^{(i)}), k = 1, 2, \dots, K$$

Vaihingen 地区影像依据领域专家人工解译结果划分为地面、低矮植被、树木、建筑物、车辆、背景六类地物。Ground truth 图中六类地物的类别和对应颜色分别如表 1 所示。

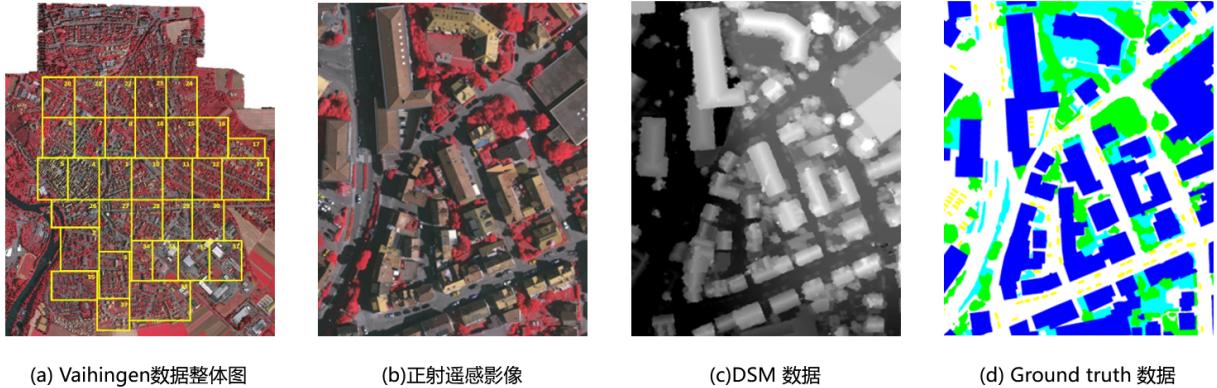


图 10 Vaihingen 影像数据

表 1 Vaihingen 数据类别标签颜色对照表

地物类别	颜色	色彩值 (R,G,B)	类别标签
地面	白色	(255, 255, 255)	0
低矮植被	青色	(0, 255, 255)	1
树木	绿色	(0, 255, 0)	2
建筑物	蓝色	(0, 0, 255)	3
车辆	黄色	(255, 255, 0)	4
背景	红色	(255, 0, 0)	5

3.3.2 数据预处理

1. 波段组合

Vaihingen 高分辨率影像空间、几何信息丰富，但正射影像光谱波段只有三个，仅使用正射影像数据无法完备有效地提取影像特征。而对于光谱相似区域的地物，如地面、建筑物、阴影等，更加难以区分，DSM 数据包含了地形、桥梁、房屋住宅和树木等高度地面高程信息，对模型区分地表建筑物、地面影像、不同高度植被一定程度上能提供帮助。Vaihingen 影像的 DSM 包含一个波段，其像素值表示高度值。因此，将 Vaihingen 数据的 DSM 作为额外的波段附加在正射影像波段后，参与模型训练。

2. 数据集划分

Vaihingen 数据中带有 Ground truth 图的影像共有 16 张，实验中选取 12 张影像（标号为 1, 5, 7, 11, 15, 17, 21, 26, 28, 30, 32 和 37）作为模型训练集，另外 4 张影像（标号为 3, 13, 23, 34）作为模型测试集。将影像 Ground truth 图由 RGB 图像转化为 1 维类别标签标签，各地物类别与对应的标签如表 10 所示。

3. 数据归一化

Vaihingen 影像数据集各通道的像素点取值在 [0, 255] 范围内，像素点的分布范围较广，如果直接输入神经网络容易引起神经元输出因输入绝对值过大而饱和的现象，从而整个网络难以收敛。所以需要对图像数据做归一化处理，即将图像像素点取值从 [0, 255] 范围映射到一个较小的变化范围，把有量纲表达式变为无量纲表达式，加快训练网络的收敛速度。

图像常用归一化方法有均值方差标准化和最大最小值归一化这两种。均值方差标准化又叫做 z-score 标准化，其处理过程是原始数据与平均值的差再除以标准差。如式3-6 所示，

$$z = \frac{x - \mu}{\sigma} \quad (3-6)$$

式中 x 为原始输入数据， μ 和 σ 分别为数据的平均值和标准差， z 为 x z-score 标准化后的输出。经过 z-score 标准化后数据服从均值为 0、方差为 1 的正态分布，该方法多用于预处理没有明显边界的数据。

最大最小值归一化方法则是通过线性函数转换，将某变化范围内数据映射到 [0, 1] 之间，线性映射过程如式3-7 所示：

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3-7)$$

式中 x 为原始输入数据， x_{max} 和 x_{min} 分别为数据的最大值和最小值， y 为 x 最大最小值归一化后的输出。

本实验中采用最大最小值归一化方法对高分影像做归一化处理，将影像像素值从 [0, 255] 映射到 [0, 1] 之间，加快神经网络的训练和收敛速度。

4. 样本选取与数据增强

高分影像单张图像尺寸通常很大，直接送入神经网络计算量太大，无法完成训练。Vaihingen 单张影像尺寸大约为 2563×2049 ，不能直接用于网络训练。本实验中对原始影像进行裁切处理，将影像裁剪为大小 256×256 的图像块作为模型训练集样本。可有效降低网络训练计算量，避免内存溢出。

常用图像裁剪方式有规则网格选取、滑动窗口裁剪和随机选取。规则网格选取是使用大小为 256×256 的网格切分影像，这种划分方式得到的样本量有限。滑动窗口裁剪方法以特定的移动间隔在原始影像上剪裁得到训练样本。文中采用滑动窗口尺寸为 256×256 ，

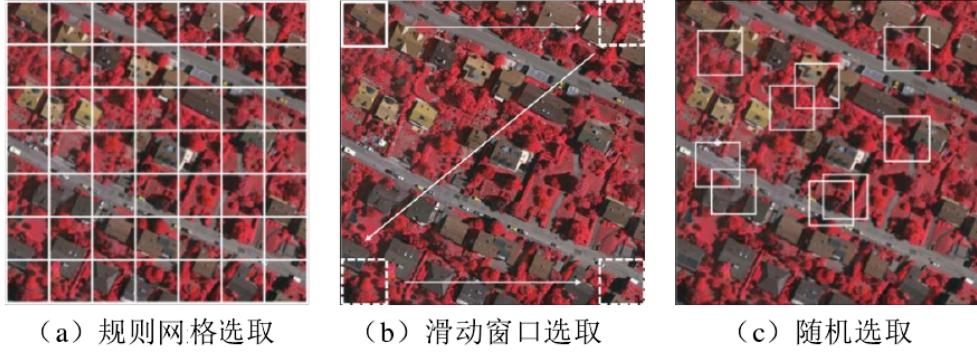


图 11 图像裁剪方式

滑动窗口样本受滑动间隔尺度影响，滑动间隔小得到的冗余影像过多，滑动间隔大则会丢失影像许多特征信息。随机选取则是使用 256×256 的裁剪窗口在原始影像上裁切获得样本。这种方法灵活便捷，能够有效利用遥感影像的信息，且裁剪出大量的训练样本。三种裁剪方式如图11所示。

本文实验中训练样本采用随机裁剪选取方式获得，测试样本采用规则网格选取方式获得。获取裁剪样本时也对同一区域的 DSM、Ground truth 图数据进行裁切，保证输入数据与类别标签的一致性。

另外，为了获取更多的有效数据，增强网络模型的泛化能力，这里对训练样本做数据增强（Data augmentation）处理。数据增强是通过一些几何变换（如平移、旋转和翻转）从已有训练样本图像生成一些新的样本，来扩大训练数据集。文中对上一步裁剪获得的训练样本做镜像对称，水平翻转处理，来扩大已有训练样本集，增强模型泛化能力。

本小节对 Vaihingen 有标签的 16 块影像数据处理，其中 12 张影像得到训练集样本图像 5760 幅，用于测试集的 4 张影像经规则网格选取法得到大小为 256×256 的测试图像合计 320 张。

3.4 实验结果与分析

3.4.1 实验环境

本章实验电脑为思腾合力 IR4200 服务器，其主要参数如下：CPU 为两块 Intel Xeon E5-2690 2.9GHz 8 核 16 线程正式版处理器，内存为 128G 容量 8 通道 DDR4 服务器内存，GPU 为技嘉 1080Ti，显存 12G。实验中使用 Ubuntu 16.04 LTS 操作系统，编程语言为 Python 3.5，神经网络模型采用谷歌开源框架 Tensorflow 编程实现，Tensorflow 版本为 1.5.0。

3.4.2 评价指标

本节实验分别通过定性和定量两种指标来评估影像分割结果。定性即人工主观对测试图像的分割结果图做出评判，定量分析则使用图像分割中常用到的两种指标：

(1) 总体精度 (Overall accuracy, OA)，即遥感影像分类精度最直观的评价指标，其值为影像中被正确分类的像元个数除以总像元个数，其计算方式如下式：

$$OA = \frac{1}{A_{\text{总}}} \sum_{k=1}^K a_{kk} \quad (3-8)$$

式中 $A_{\text{总}}$ 为真实地物像元总数， K 为地物类别数， a_{kk} 第 k 类地物被正确分类的像元数。

(2) 平均交并比 (Mean Intersection over union, mIoU)，IoU 是影像中真实标签与预测分割结果两者交集与并集的比值，mIoU 是图像语义分割领域最常用的准确度度量方法，它分别对影像每个地物类别计算 IoU，然后再对所有地物类别的 IoU 求均值。IoU 的计算方式如式3-9：

$$IoU = \frac{\text{预测结果} \cap \text{真实标签}}{\text{预测结果} \cup \text{真实标签}} \quad (3-9)$$

影像所有类别地物的 mIoU 的计算方式如式3-10：

$$mIoU = \frac{1}{K} \sum_{i=1}^K \frac{a_{ii}}{\sum_{j=1}^K a_{ij} + \sum_{j=1}^K a_{ji} - a_{ii}} \quad (3-10)$$

式中 a_{ij} 表示第 i 类地物被错分为第 j 类的像元个数， a_{ji} 表示第 j 类地物被错分为第 i 类的像元个数， a_{ii} 表示第 i 类地物正确预测的像元个数。

3.4.3 网络参数

本节实验中输入图像为 $256 \times 256 \times 4$ 的 4 波段融合影像（近红外、红、绿、DSM），生成网络中的分割模型权值使用 ImageNet 上预训练好的基于 VGGNet16^[38] 的全卷积网络权值初始化，式3-4 中 GAN 模型中的目标函数权衡因子为 $\lambda = 2$ ，使用 Adam^[58] 优化器计算梯度更新，Adam 优化器具有自适应的学习率，初始学习率为 $\alpha = 10^{-4}$ ，梯度动量一阶和二阶矩估计指数衰减率分别初始化为 $\beta_1 = 0.9$ 和 $\beta_2 = 0.9999$ ， $\epsilon = 10^{-8}$ 防止除数为 0。实验中设置批大小 (Batch size) 为 128，epoch 为 200 时能取得更好的训练效果。

模型生成器 G 中的网络压缩部分由五个卷积模块堆叠而成，其作用是逐层提取影像特征。每个卷积模块均包含 $2/3$ 个卷积层和 1 个尺寸 2×2 最大池化层，为了保证卷积前后特征图尺寸不变，所有卷积层均采用边界填充，最大池化层使得特征图尺寸缩小为池化

前的一半。生成器 G 中反卷积恢复为四个反卷积的上采样模块，其作用是逐层扩大特征图大小、恢复影像细节信息。每个上采样模块中包含一次反卷积操作，反卷积生成两倍维度的特征图，再接两层卷积层提取特征。最后通过一个 1×1 大小的 1 维卷积层将特征图维度降维 1，通过 Softmax 函数输出预测结果的像素类别。另外，在反卷积中使用跳层连接网络压缩部分与反卷积层特征图大小相同的卷积层，从而将影像低阶特征和高阶特征融合，保证更精细的分割结果。

表 2 基于 CGAN 框架的全卷积分割模型参数表

网络模型	结构	Levels	网络层	该层输出尺寸	卷积核	步长	激活函数
生成器 G	卷积压缩	G-输入	level 0	$256 \times 256 \times 4$			
		level 1	conv1_1	$256 \times 256 \times 64$	$3 \times 3/64$	1	ReLU
			conv1_2	$256 \times 256 \times 64$	$3 \times 3/64$	1	ReLU
		level 2	pool2_1	$128 \times 128 \times 64$	$2 \times 2/-$	2	--
			conv2_1	$128 \times 128 \times 128$	$3 \times 3/128$	1	ReLU
		level 3	conv2_1	$128 \times 128 \times 128$	$3 \times 3/128$	1	ReLU
			pool3_1	$64 \times 64 \times 128$	$2 \times 2/-$	2	--
			conv3_1	$64 \times 64 \times 256$	$3 \times 3/256$	1	ReLU
			conv3_2	$64 \times 64 \times 256$	$3 \times 3/256$	1	ReLU
		level 4	conv3_3	$64 \times 64 \times 256$	$3 \times 3/256$	1	ReLU
			pool4_1	$32 \times 32 \times 256$	$2 \times 2/-$	2	--
			conv4_1	$32 \times 32 \times 512$	$3 \times 3/512$	1	ReLU
			conv4_2	$32 \times 32 \times 512$	$3 \times 3/512$	1	ReLU
		level 5	conv4_3	$32 \times 32 \times 512$	$3 \times 3/512$	1	ReLU
			pool5_1	$16 \times 16 \times 512$	$2 \times 2/-$	2	--
			conv5_1	$16 \times 16 \times 512$	$3 \times 3/512$	1	ReLU
			conv5_2	$16 \times 16 \times 512$	$3 \times 3/512$	1	ReLU
	反卷积恢复		conv5_3	$16 \times 16 \times 512$	$3 \times 3/512$	1	ReLU
	level 6	deconv6_1	$32 \times 32 \times 512$	$4 \times 4/512$	2	ReLU	
		conv6_1	$32 \times 32 \times 512$	$3 \times 3/512$	1	ReLU	
		conv6_2	$32 \times 32 \times 512$	$3 \times 3/512$	1	ReLU	
	level 7	deconv7_1	$64 \times 64 \times 256$	$4 \times 4/256$	2	ReLU	
		conv7_1	$64 \times 64 \times 256$	$3 \times 3/256$	1	ReLU	
		conv7_2	$64 \times 64 \times 256$	$3 \times 3/256$	1	ReLU	
	level 8	deconv8_1	$128 \times 128 \times 128$	$4 \times 4/128$	2	ReLU	
		conv8_1	$128 \times 128 \times 128$	$3 \times 3/128$	1	ReLU	
		conv8_2	$128 \times 128 \times 128$	$3 \times 3/128$	1	ReLU	

(接下页)

(接上页)

网络模型	结构	Levels	网络层	该层输出尺寸	卷积核	步长	激活函数
			deconv9_1	256 × 256 × 64	4 × 4/64	2	ReLU
		level 9	conv9_1	256 × 256 × 64	3 × 3/64	1	ReLU
			conv9_2	256 × 256 × 64	3 × 3/64	1	ReLU
	G-输出		conv10_1	256 × 256 × 1	1 × 1/1	1	Softmax
	D-输入			256 × 256 × 5			
判别器 D	分类网络	level 11	conv11_1	256 × 256 × 32	3 × 3/32	1	ReLU
			conv11_2	256 × 256 × 32	3 × 3/32	1	ReLU
		level 12	pool12_1	128 × 128 × 32	2 × 2/-	1	--
			conv12_1	128 × 128 × 32	3 × 3/32	1	ReLU
			conv12_2	128 × 128 × 32	3 × 3/32	1	ReLU
		level 13	pool3_1	64 × 64 × 32	2 × 2/-	2	--
			FC13_1	64 × 64 × 32	3 × 3/32	1	--
			FC13_2	64 × 64 × 32	3 × 3/32	1	--
	G-输出			1 × 1			Sigmoid

模型判别器 D 中是一个经典的二分类神经网络，网络输入为 5 维数据（4 波段影像 + 1 维真实类别图或 4 波段影像 + 生成器 G 生成分割结果图），模型由两个卷积结构后接两个全连接层组成，通过 Sigmoid 函数判别当前输出结果为 1 或 0。

基于 CGAN 框架的全卷积分割模型详细的网络结构与参数权值如表2 所示，交替训练生成器 G 和判别器 D 完成网络各权值参数的学习。

3.4.4 结果与分析

本章提出的基于 CGAN 的影像分类方法将条件生成对抗的思想引入全卷积分类方法中，利用 CGAN 对生成样本与真实标签的差异纠正如能力，能够获得更准确的地物分类边界，且保持同类地物类内像素点类别一致性。实验中采用的是融合正射三波段影像和 DSM 数据的四波段数据，实验对比的基准方法为经典的 FCN 语义分类方法。下面分别从精确量化和目视评估两个角度对 Vaihingen 数据集上的地物分割识别结果进行比较。

表3 比较了新提出的基于 CGAN 的影像分类模型和 FCN 模型的分类精度，分别统计了地面、低矮植被、树木、建筑物、车辆五类地物（背景未计算）的分类精度，整幅图片中五类地物总体精度和 mIoU 指标。从分类精度可知，两种方法对地面、树木和建筑物均有较好的识别精度，低矮植被易与背景中阴影混淆，故整体识别精度较低，车辆分类精度也不高，这与样本中车辆所占像素面积少，存在该类别样本数不多相关。文中提出的基于

CGAN 的分类方法在“建筑物”这个类别地物识别精度最高，为 87.64%，相比 FCN 分类方法的 83.96%，有约 4% 的绝对精度提升。而对 FCN 分类中精度较低的“低矮植被”地物，分类精度由 63.39% 大幅提升到 74.47%，该类别识别精度提升幅度达 10%，提升原因一方面是方法中使用 DSM 波段数据能量化区分低矮植被与阴影的高度特征差异，较好得区分二者，另一方面则是基于 CGAN 的分类方法对地物的边界有更准确的生成能力。本文提出基于 CGAN 的分类方法 OA 为 80.15%，mIoU 为 61.83%，相比 FCN 分类算法中 78.48% 的 OA 和 58.42% 的 mIoU，均有一定程度的提升。

表 3 Vaihingen 数据分类精度评估表

方法	地面	低矮植被	树木	建筑物	车辆	OA	mIoU
FCN 分类方法	81.14%	63.39%	79.52%	83.96%	62.39%	78.48%	58.42%
基于 CGAN 分类方法	83.78%	74.47%	82.40%	87.64%	78.83%	80.15%	61.83%

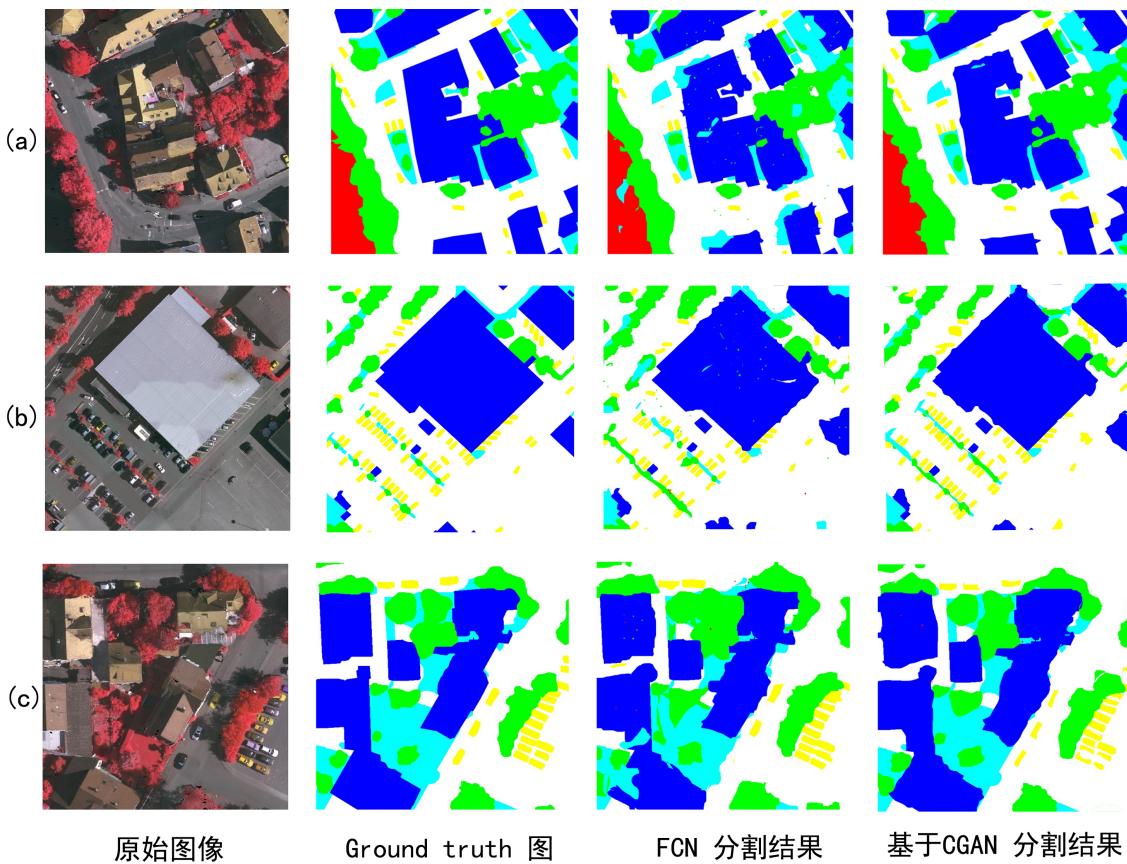


图 12 影像分类的可视化结果

图12 为本文提出的基于 CGAN 的分类方法和 FCN 分类方法分别在三组测试影像上

分类预测的可视化结果。图12(a)为一居民住宅区，参照真实地物类别的 Ground truth 图，相比 FCN 分类方法，基于 CGAN 的影像分类方法在处理图中建筑物与房屋阴影的分割边界时，能更好地将阴影划分到背景中，更加准确地将住宅等建筑物识别为一个整体，从而易混淆的阴影中区分开。图12(b)则是一处停车场周边影像图，对形态不一、位置各异的车辆进行识别预测是关键。如左下角部分停靠在树木下的两排车辆，一些车辆与地面特征相似，区分度较小，FCN 分类算法无法识别出这些“车辆”，直接将车错分为背景，而基于 CGAN 的分类方法则对停靠的车辆尽可能准确的进行了识别预测。相比前者，尽可能地找出了停靠在树木阴影下的车辆。图12(c)为一处树木与建筑物环绕区域，对“建筑物”、“低矮植被”以及“树木”三类地物的边界明确划分成为预测分类的难点。如图正上方区域房屋环绕的树木与低矮植被区域，基于 CGAN 的分类方法将图中的“树木”与“低矮植被”进行了区分，而 FCN 的分类方法则将“低矮植被”识别为“树木”。综合上面三组图像的分类结果，文中提出的基于 CGAN 的影像分类方法相比 FCN 分类方法能够区分特征相近地物类别，生成更准确的分割边界。

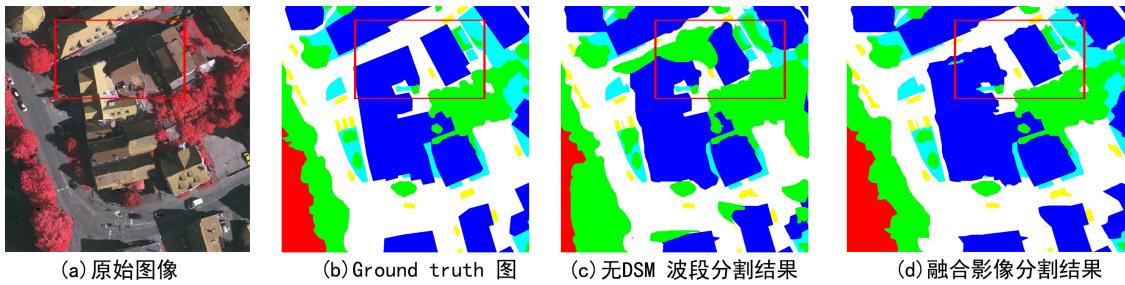


图 13 DSM 波段对影像分类结果的影响

文中使用的训练样本为三波段正射影像与一波段 DSM 的融合数据。下文实验比较了 DSM 数据对实验结果的影响。DSM 数据包含地形和地物高度等相关高程信息，“道路”、“低矮植被”和“建筑物”容易受不同地形、高度变化导致地物的特征起伏变化打，融合 DSM 特征后更利于区分这些地物。图13 中比较了正射影像是否融合 DSM 数据对分割结果的影响。如图中红框区域为“建筑物”与“地面阴影”交叉部分，没有 DSM 高程数据的分割模型中，将部分“阴影”误分为“树木”类别，同时，“建筑物”类别的边界也不完整。融合影像输入数据的分割结果则较好地解决了受高度特征影响的地物错分问题，尽可能保证“建筑物”边界的明确和完整。表4 则统计了有无 DSM 波段的数据源对实验分类精度的影响，通过对五类地物的分类精度、OA 和 mIoU 指标的对比，我实验返现输入数据加入 DSM 波段后，“低矮植被”的分类精度提升最大，相比初始的 72.68% 提升了 1.72%，达到 82.40%。整体的分类精度也由 79.52% 提升到 80.15%。因此，融合

DSM 高程信息数据能得到更优秀的分类精度和分割效果。

表 4 DSM 波段对分类精度的影响

数据源	地面	低矮植被	树木	建筑物	车辆	OA	mIoU
正射影像	83.27%	72.68%	82.61%	86.33%	78.20%	79.52%	60.96%
正射影像 +DSM	83.78%	74.47%	82.40%	87.64%	78.83%	80.15%	61.83%

3.5 本章小结

高分影像固有的不确定性和复杂的类内特征使得地物边界难以区分且同类别像素点难以保持空间一致性。此外，FCN 分类方法上采样操作会损失影像的特征细节，导致遥感影像地物边界更加难以正确识别。本章提出基于 CGAN 的影像分类方法将生成对抗网络的思想应用到全卷积分割模型中，CGAN 网络对抗损失提升影像远距离像素点间类别标签的连续性，因而影像分类结果同类别像素点更具有一致性。对抗框架下分割模型的建模能力更强，一定程度上能提升分类精度。将文中提出的基于 CGAN 的影像分类方法应用到 Vaihingen 影像数据分割实验上，分割结果和分类精度均表明文中提出的方法整体上相比经典 FCN 语义分割方法有着更优秀的分类效果。此外，文中实验结果也证明了融合 DSM 波段高程信息的影像数据，可以进一步提升识别特征相近的不同地物类别的能力。

第 4 章 基于 CGAN 影像分类的改进方法

4.1 引言

在第 3 章中，我们将对抗训练的思想应用到 FCN 分类模型中，提出了基于 CGAN 影像分类方法。模型的对抗损失一定程度上增强了影像像素点间的连续性，提高了 FCN 方法的影像分类效果。然而，因为遥感数据固有的不确定性，相比自然图像，遥感影像分类地物边界混淆、歧义性较严重。此外，模型中反卷积上采样操作将特征图恢复到原始影像尺寸，特征的损失也加剧了地物分类的边界模糊问题。因此，本章试图从增强影像分类结果边界信息和分割结果后处理两个角度改进基于 CGAN 的遥感影像分类方法。本文作者于 2017 年研究过高分影像数据不确定性特征处理相关的内容，提出一种基于三角形模糊集值的区间二型模糊聚类方法 (Triangular Fuzzy Set Valued Interval Type 2 Fuzzy Clustering Method, TFSV-IT2FCM) 方法^[59] 用于区分高分影像地物，且取得了较好的效果。影像聚类分割结果为许多同质性的分割单元，不同地物分类边界往往在分割单元间的边界线上。因此，本章将经 TFSV-IT2FCM 聚类分割得到的影像分割单元图作为辅助信息，用于改进基于 CGAN 影像分类方法。

4.2 TFSV-IT2FCM 聚类分割

4.2.1 方法原理

由于遥感影像数据存在不确定性和混合边界问题，模糊 C 均值聚类 (Fuzzy c-means clustering, FCM) 方法被广泛应用到遥感影像解译中^[16]。随着高分影像数据的普及，遥感影像模糊聚类方法由基于像元的聚类方法发展为面向对象的模糊聚类方法。本文作者提出的 TFSV-IT2FCM 方法就是面向对象的高分影像模糊聚类分割方法。TFSV-IT2FCM 算法首先定义三角形模糊集数据模型 (Triangular Fuzzy Set Valued, TFSV) 对高分影像的分割单元建模，能够表征影像分割单元内的像素点特征。然后定义一种新的区间值距离来度量两个 TSFV 数的相似性。最后基于 TFSV 数据模型和新的距离度量提出 TFSV-ITFCM 模糊聚类方法。TFSV-ITFCM 聚类分割结果为许多同质性的聚类簇。各聚类簇的像素点差异性不大，且对簇内像素点变化具有一定容忍度。即某一簇内的所有像素点可视作同类地物。

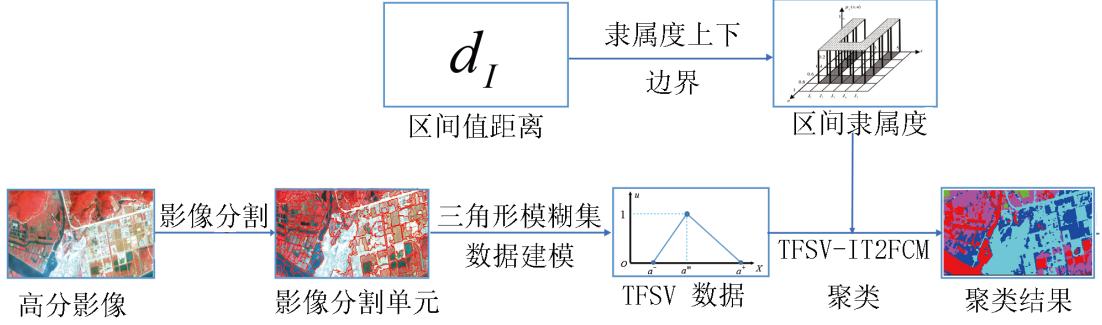


图 14 TFSV-IT2FCM 算法框架

TFSV-IT2FCM 方法的整体框架如图14 所示。首先对高分影像 I ，使用 SLIC 分割算法获得较小的影像分割单元 SS ，为：

$$SS = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n\} \quad (4-1)$$

其中 $\mathbf{B}_i (1 \leq i \leq n)$ 表示第 i 个分割单元， n 表示分割单元的总数。 $\mathbf{B}_i (i = 1, 2, \dots, n)$ 是一个 $j \times p$ 矩阵，其中 j 表示每个波段包含的像素数目， p 表示图像通道数。对任一个影像单元 \mathbf{B}_i 均可以定义为一个 TFSV 数据模型 \tilde{A}_i ，影像分割单元的数据不确定性特征可由 \tilde{A}_i 表达。即影像所有分割单元集合 $SS = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n\}$ 可以被表示为：

$$SS \rightarrow \{\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n\} \quad (4-2)$$

其中 SS 是一个 $n \times p$ 矩阵， \tilde{A}_i 是 \mathbf{B}_i 对应的 p 维 TFSV 数据， n 表示分割单元的数目。 d_I 是定义 TFSV 数据模型间的区间值距离，对于两个 TFSV 数据 \tilde{A} 和 \tilde{B} 的区间值距离 $d_I(\tilde{A}, \tilde{B})$ 为下式：

$$d_I(\tilde{A}, \tilde{B}) = [\min\{d_0(\tilde{A}, \tilde{B}), d_1(\tilde{A}, \tilde{B})\}, \max\{d_0(\tilde{A}, \tilde{B}), d_1(\tilde{A}, \tilde{B})\}] \quad (4-3)$$

式中 $d_0(\tilde{A}, \tilde{B})$ 和 $d_1(\tilde{A}, \tilde{B})$ 分别为 \tilde{A} 和 \tilde{B} 的 $0-cut$ 和 $1-cut$ 豪斯多夫距离^[60]。区间二型模糊聚类模型的隶属度上界 \bar{u}_{ij} 和下界 \underline{u}_{ij} 隶属度则由区间值距离 d_I 和模糊指数 m 求解，分别为：

$$\bar{u}_{ij} = \max \left\{ \frac{1}{\sum_{k=1}^K \left(\frac{d_{ji}^0}{d_{ki}^0} \right)^{\frac{2}{m-1}}}, \frac{1}{\sum_{k=1}^K \left(\frac{d_{ji}^1}{d_{ki}^1} \right)^{\frac{2}{m-1}}} \right\} \quad (4-4)$$

和

$$\underline{u}_{ij} = \min \left\{ \frac{1}{\sum_{k=1}^K \left(\frac{d_{ji}^0}{d_{ki}^0} \right)^{\frac{2}{m-1}}}, \frac{1}{\sum_{k=1}^K \left(\frac{d_{ji}^1}{d_{ki}^1} \right)^{\frac{2}{m-1}}} \right\} \quad (4-5)$$

其中 d_{ji}^0 和 d_{ji}^1 分别是输入样本 \tilde{X}_i 和聚类中心 \tilde{V}_j 的 $0-cut$ 距离和 $1-cut$ 距离度量，由区间值距离 $d_I(\tilde{X}_i, \tilde{V}_j)$ 计算得出。聚类迭代更新，求解区间二型的模糊划分矩阵 $\mathbf{U} = [U_{ij}]_{n \times K}$

$$U_{ij} = \frac{\underline{u}_{ij} + \bar{u}_{ij}}{2} \quad (4-6)$$

其中 $U_{ij} = [\underline{u}_{ij}, \bar{u}_{ij}]$ 。然后，求出 \tilde{X}_i 到聚类中心 \tilde{V}_k 的最大隶属度 U_{ik} ，其中 $k = 1, 2, \dots, K$ 。最后根据最大隶属度原则，将 \tilde{X}_i 划分到类别 \tilde{V}_k ，完成 TFSV-IT2FCM 聚类分割过程。

TFSV-IT2FCM 聚类分割方法可以看作是对邻域内同质性像素点关系的建模，即邻域内具有相近像素值的像素点将被划分为同一个类别。此外，因模糊不确定度量的存在，同个聚类簇内的像素点可以在一定范围内波动。将 TFSV-IT2FCM 聚类分割方法作用到高分影像，可以将各类别地物划分为不同的分类簇，而同一个簇内的像素点具有同质性，即更可能属于同一个地物类别，邻近簇之间的分界线也有很大机率为相邻的不同地物类别分类边界（也可能是同类地物的，当地物内部差异巨大时，同一地物会被模糊聚类为多个簇）。

4.2.2 影像数据处理

第3章实验用到的 Vaihingen 数据为 16 张尺寸约为 2563×2049 的高分影像。为了得到影像整体的同质性分割单元信息图，这里直接使用 TFSV-IT2FCM 聚类分割方法分别对这 16 张大尺寸影像聚类分割。因 Vaihingen 数据共有地面、低矮植被、树木、建筑物、车辆、背景这六类地物，故 TFSV-IT2FCM 聚类分割中聚类数可设置为六类。

图15 为 TFSV-IT2FCM 方法提取 Vaihingen 影像同质性分割单元结果图。图15(a) 为 Vaihingen 数据标号为 30 的影像，该影像尺寸为 1983×2556 。使用 TFSV-IT2FCM 聚类方法对影像进行分割，可以将影像依据分割单元的同质近似性，划分为许多个大小不一的簇，分割簇内的点输出标签相同，相邻簇间输出标签不同，图15(b) 即为影像聚类分割结果（聚类结果为一个波段二维矩阵，这里可视化三通道 RGB 彩色图像），对分割单元信息图上依次经过图像梯度求解和二值化处理可以得到各个簇之间的边界信息，依据 TFSV-IT2FCM 聚类分割结果提取到的影像边界信息如图15(c) 所示。图15(d) 为影响中间某区域的放大展示结果。

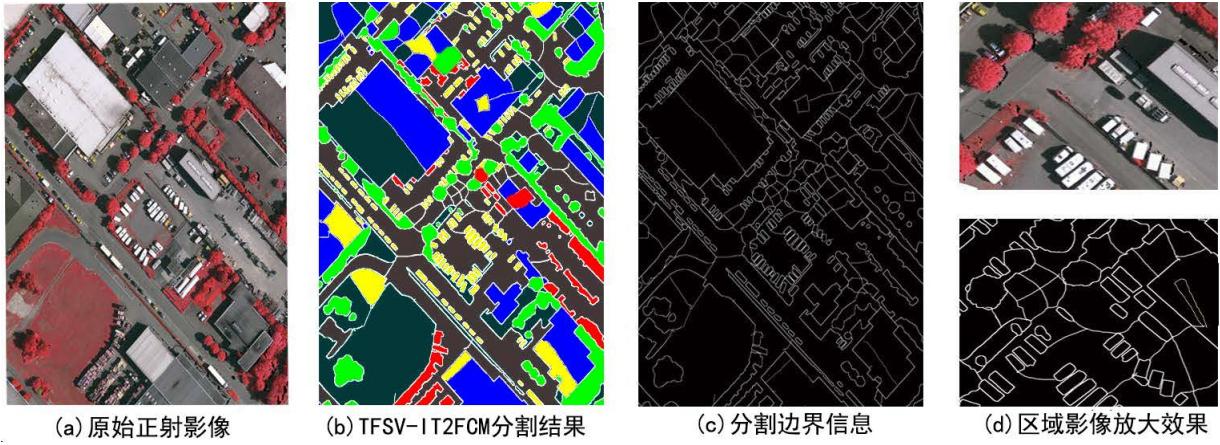


图 15 TFSV-IT2FCM 提取 Vaihingen 影像辅助特征信息

类似上面的处理方法，对 Vaihingen 数据集内所有的 16 张影像均进行 TFSV-IT2FCM 聚类分割操作，得到影像同质性分割单元图信息。数据聚类分割结果为二维的类别矩阵，这里将标签值采用最大最小归一化方法映射到 $[0, 1]$ 区间范围内。经过 TFSV-IT2FCM 聚类分割方法得到的影像边界辅助信息图与原始影像尺寸相同，且位置坐标一一对应。在第3 章随机滑动窗口裁切原始影像获取大小为 256×256 的训练样本时，也按照相同的滑窗裁切方式对同区域的影像辅助分割图进行裁切，得到训练样本边界辅助信息图。同理，待测试影像在按照规则网格裁切影像获取测试样本时也对测试影像的影像辅助分割图进行裁切。确保数据集所有样本同一区域的影像数据、类别标签图和经 TFSV-IT2FCM 方法得到的影像辅助分割信息图三者一致对应。此外，利用 TFSV-IT2FCM 方法获取影像辅助特征属于数据初始处理阶段，不会增加基于 CGAN 的分类模型的迭代训练时间。

4.3 融合边界特征的 CGAN 影像分类方法

在上一小节中，我们使用 TFSV-IT2FCM 聚类分割方法对高分影像数据做无监督聚类分割处理，使用 TFSV 模型对影像数据不确定性建模，再对影像分割单元的 TFSV 数据模型进行聚类。得到高分影像的同质性分割单元信息图。

在基于 CGAN 的影像分类方法中，生成网络的池化操作使得影像空间分辨率减小，影像像素点的位置信息不可避免的存在损失。此外，分割模型使用反卷积上采样操作扩大特征图的尺寸，像素点低阶位置信息丢失的更加严重。将影像低阶边界位置特征信息传到高层特征图中，一定程度上即可弥补因池化操作和上采样过程中像素点损失的位置信息。基于上一节得到的影像分割单元边界信息图，对分割模型中各级上采样输出的特征图融合相同尺度的边界掩膜特征信息，提升高阶语义特征的边界位置信息，对基于 CGAN 的影

像分类方法进行改进，提出融合边界掩膜特征的 CGAN 影像分类方法。

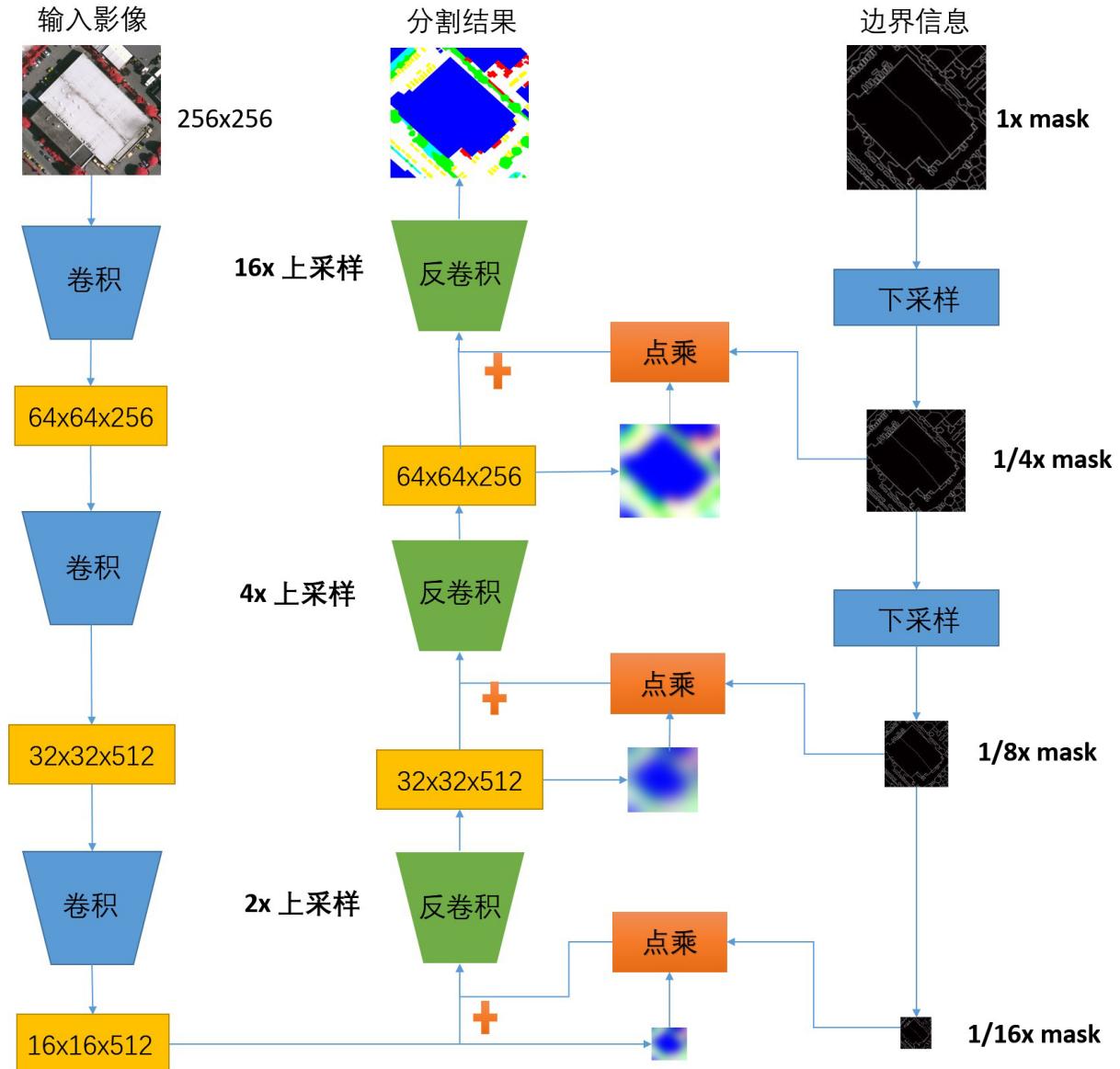


图 16 融合边界掩膜特征的分割模型

图16 所示为融合边界掩膜特征的 CGAN 影像分类方法的生成器分割模型部分示意图。假设由 TFSV-IT2FCM 聚类分割方法得到影像分割单元图为 S ，根据图像梯度求解 S 的边界信息，如式4-7 所示， S 上任一点 (x,y) 的梯度值可以求解得到， S 的梯度图表示为 S_g 。

$$|grad(S(x,y))| = [(\frac{\partial S}{\partial x})^2 + (\frac{\partial S}{\partial y})^2]^{\frac{1}{2}} \quad (4-7)$$

对梯度图 S_g 上的梯度进行二值化处理，如式4-8即可得到影像边界掩膜信息 S_e 。

$$S_e(x, y) = \begin{cases} 1, & |grad(S(x, y))| > 0 \\ 0, & |grad(S(x, y))| = 0 \end{cases} \quad (4-8)$$

边界掩膜 S_e 为单波段二值灰度灰度图，其图像初始尺寸与输入样本尺寸相同，为 256×256 。为了将掩膜融合到各级反卷积的特征输出图中，使用最大池化方法对掩膜图 S_e 按倍数进行下采样处理，分别得到 $\frac{1}{4}$ 、 $\frac{1}{8}$ 和 $\frac{1}{16}$ 倍原始尺寸大小的边界掩膜特征图。借鉴 ResNet^[39] 模型中“残差学习”组合特征的思想，我们对模型中各个层级的反卷积特征图融合对应尺度大小的边界特征信息。假设某一尺度反卷积层原始输入特征图为 X_f ，该层同尺度的边界掩膜为 S_e^m ，则添加掩膜融合后新的特征图 X_f^* 可以表示为下式

$$X_f^* = X_f + H(X_f, S_e^m) \quad (4-9)$$

式中 $H(X_f, S_e^m)$ 为矩阵哈达玛积^[61](Hadamard product)，也叫矩阵点乘。对于两个矩阵 $A, B \in \mathbb{R}^{m \times n}$ ，且 $A = \{a_{ij}\}$, $B = \{b_{ij}\}$ ，则 A 和 B 的哈达玛积为式4-10 所示。

$$H(A, B) = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & a_{13}b_{13} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & a_{23}b_{23} & \cdots & a_{2n}b_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & a_{m3}b_{m3} & \cdots & a_{mn}b_{mn} \end{bmatrix} \quad (4-10)$$

可以求出掩膜 S_e^m 和特征图为 X_f 矩阵对应位置元素的哈达玛积结果。使用融合边界掩膜的特征图 X_f^* 替代 X_f ，参与网络模型迭代训练。对于各层反卷积的特征图均采取上述融合边界掩膜的方式，即将影像像素位置、边界等低层特征组合到反卷积阶段的高层语义特征中，提升模型对影像地物边界的识别能力。

4.4 基于辅助信息后处理的影像分类方法

上文中我们提出融合边界掩膜特征的 CGAN 影像分类方法，将边界掩膜特征信息融合到高阶语义特征中，改善了影像像素点下采样过程中丢失位置、边界等低阶特征的问题。然而，基于 CGAN 影像影像分类方法中分割模型最后一层输出为影像各像素点所属类别的概率分布。对各像素点分类预测只考虑了各像素点自身的概率分布，没有考虑到该像素点邻域内像素点间的关系，这样做的后果容易导致分类结果中出现很多细碎的错分区域。特别是高分影像中复杂特征的地物，同类别地物区域内的分割结果很容易出现细碎的错分区域。

影像中像素点通常满足局部相似性原则，即相邻像素之间往往更可能具有相同的类别标签，一定空间范围内相似颜色的像素点更可能具有相同的类别标签。基于上一节由TFSV-IT2FCM聚类分割方法预处理得到的影像同质性分割单元图，我们对基于CGAN影像分类模型中生成模型输出层结果做后处理操作，当前像素点分类标签不仅依赖于当前像素点预测类别的概率，还受当前像素点所处的影像分割单元中其他像素点的影响。

设影像 X 由 N 个像素点组成，即有 $X = \{X_1, X_2, \dots, X_N\}$ ，共有 L 个类别地物的影像类别标签值为 $L = \{l_1, l_2, \dots, l_L\}$ 。对每个像素点 X_i ，基于 CGAN 影像分类方法分割模型输出是经过 Softmax 归一化的 L 维向量，如下式

$$P(X_i) = [P(X_i = l_1), P(X_i = l_2), \dots, P(X_i = l_L)] \quad (4-11)$$

其中 $P(X_i = l_c)$ 表示像素点 X_i 被分为第 l_c ($c = 1, 2, \dots, L$) 类地物的概率。模型类别决策时对式4-11 使用 argmax 函数做概率最大化划分类别维度即可。

本节研究内容就是优化这里的类别决策方式。假设预处理得到的同质性分割单元图为 S ， S 为多个影像分割单元的集合，设某个分割单元内像素点集合为 X_S ，其包含 S 个像素点。对影像分割单元 X_S 内所有像素点统计各类别维度是当前像素点概率最大的类别标签的像素点个数，即有：

$$C = \{C_{l_1}^S, C_{l_2}^S, \dots, C_{l_L}^S\} \quad (4-12)$$

且满足约束条件

$$S = \sum_{i=1}^L C_{l_i}^S \quad (4-13)$$

式4-12 中 $C_{l_i}^S$ 表示当前分割单元 X_S 中满足分割模型预测像素点类别为 l_i 时概率最大的像素点的个数，总的像素点个数之和为 S ，即有式4-13。对式4-12 中每一项除以 S ，求出分割单元 X_S 中每一个类别标签为最大概率时的像素点个数占当前分割单元 X_S 总像点个数的比值，即

$$P_s(l_i) = \frac{C_{l_i}^S}{S}, i = 1, 2, \dots, L \quad (4-14)$$

利用式4-14 中分割单元的占比关系，对式4-10 中像素点类别预测的概率值进行更新，更新方法如式4-15所示：

$$P^\star(X_i = l_c) = P(X_i = l_c) + \lambda \times \exp(P_s(l_c) - 1) \quad (4-15)$$

式中 λ 为权衡因子，取值范围为 $[0, 1]$ ，是一个超参数，在后面实验中取 $\lambda = 0.8$ 能得到最好的实验效果。 $P_s(l_c)$ 为 $[0, 1]$ 内的一个比值，比值 $P_s(l_c)$ 越大，表明该影像分割单元区域内预测为该类别的像素点占据多数，式4-15 对该类别标签的预测概率有正向加权激励作用，符合影像邻近相似像素更可能为同类标签的准则。

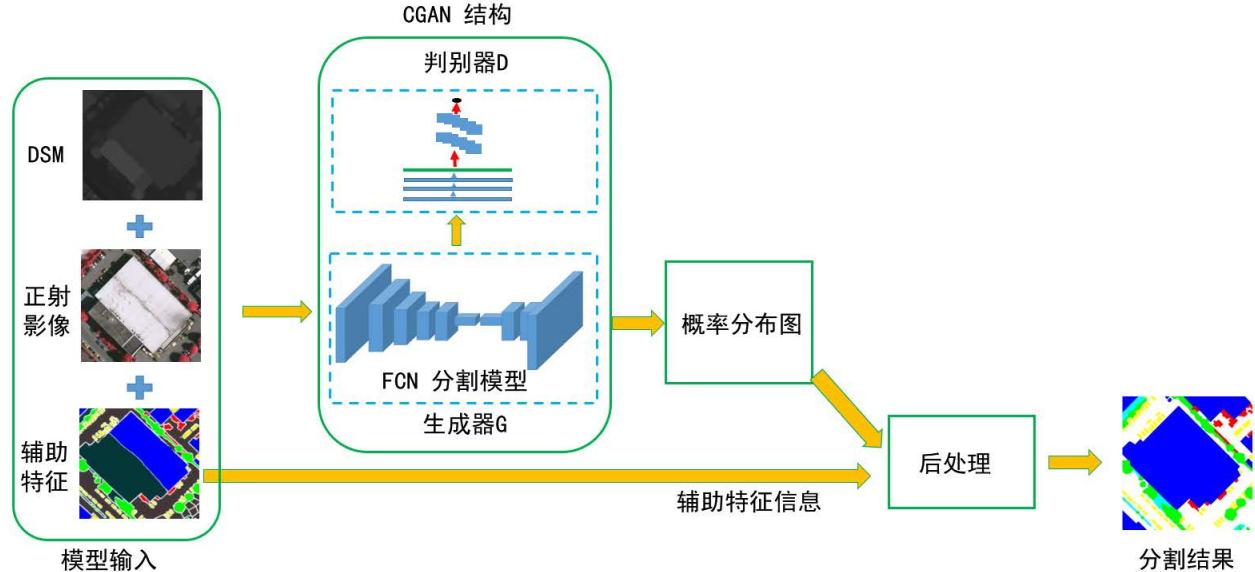


图 17 基于辅助信息后处理的 CGAN 影像分类流程图

在模型训练迭代时，使用式4-15 中概率分布结果代替分割模型预测输出概率，即式3-2 中基于 CGAN 影像分类方法分割网络的多分类交叉熵损失更新为下式：

$$l_{mce}(\hat{y}, y) = - \sum_{i=1}^{H \times W} \sum_{c=1}^C y_{ic} \log(\hat{y}_{ic} + \lambda \times \exp(P_s(l_c) - 1)) \quad (4-16)$$

将式4-16 中多分类交叉损失更新到基于 CGAN 分类模型的目标函数 (式3-4) 中，进而对网络模型进行权值更新、迭代训练。基于辅助信息后处理的 CGAN 影像分类方法流程如图17 所示，正射影像和 DSM 数据作为模型输入，同质性分割单元图作为辅助信息更新生成模型输出层的概率分布图。

4.5 实验结果与分析

基于 CGAN 的影像分类模型中网络池化操作和反卷积上采样丢失了影像低阶的位置、边界信息，文中借助 TFSV-IT2FCM 影像聚类分割方法获取影像的分割单元边界信息，然后将边界信息下采样到不同尺寸，以特征掩膜的方式融合到反卷积输入特征图中，增加高阶语义特征中的低阶边界特征。为了验证文中提出的融合边界掩膜特征的 CGAN 影像分

类方法的正确性。我们在 Vaihingen 数据集上进行实验。实验环境和参数条件等配置跟第3章的实验配置相同，分类效果的评价指标为 OA 和 mIoU。分类的量化评估结果如表5所示，可以发现整体的分类精度 OA 和 mIoU 相比没有融合边界特征的 CGAN 影像分类方法均有一定程度的提升，融合边界特征的 CGAN 分类方法 OA 提升了 3.34%，为 83.49%。mIoU 也从 61.83% 提高到 63.52%。从地面、低矮植被、树木、建筑物、车辆这五类地物正确识别的精度来看，“车辆”的分类精度提升幅度最大，为 4.75%。研究发现，相比其他类型地物，“车辆”在影像中的占比和像素区域大小均较小，融合边界特征后一定程度能够还原出影像上“车辆”这类地物的区域边界范围，因此，融合边界特征信息对影像小区域同类地物的识别改善效果最明显。

表 5 融合边界特征对影像分类的影响

方法	地面	低矮植被	树木	建筑物	车辆	OA	mIoU
CGAN	83.78%	74.47%	82.40%	87.64%	78.83%	80.15%	61.83%
CGAN + 融合特征	82.54%	77.59%	83.21%	84.48%	83.58%	83.49%	63.52%

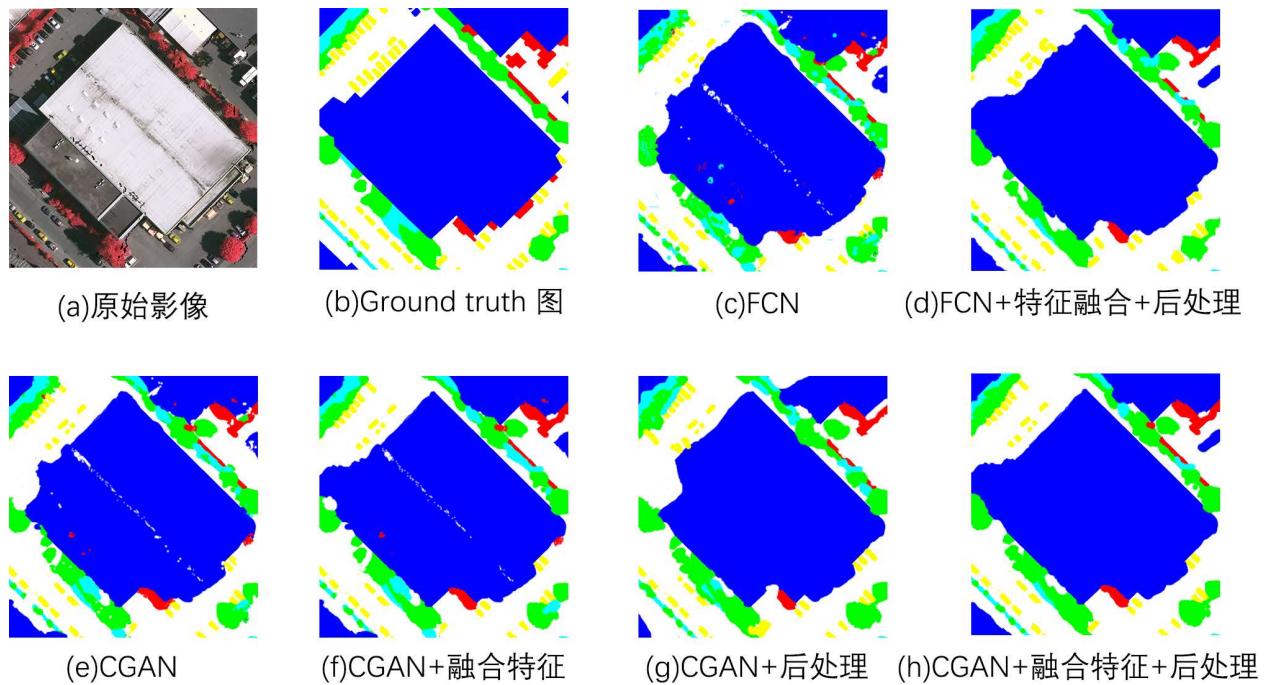


图 18 改进方法分类结果可视化图

图18 中的 (e) 和 (d) 图比较了基于 CGAN 影像分类方法和融合边界特征的 CGAN 影像分类方法的分类效果，参照原始影像和真实分类 Ground truth 图，可以看到融合了边界

特征的影像分类方法在处理图中白色建筑物与阴影区域的边界上能够得到比没有融合特征的分类方法更好的边界信息。阴影和房屋的分界线更加完整、清晰。实验结果验证了融合特征的分类方法确实能够获得更清晰、准确的地物分类边界，同时能提升 CGAN 影像分类方法的精度。

表 6 后处理对 CGAN 影像分类的影响

方法	地面	低矮植被	树木	建筑物	车辆	OA	mIoU
CGAN	83.78%	74.47%	82.40%	87.64%	78.83%	80.15%	61.83%
CGAN + 后处理	83.98%	76.92%	82.94%	87.85%	79.14%	80.97%	62.04%

为了解决高分影像中因复杂特征导致的同类别地物区域出现其他类别的细碎错分问题，在对影像各像素点类别预测时，我们考虑同质性分割单元内其他像素点的类别划分影响，提出了基于辅助信息后处理的影响分类方法，改进基于 CGAN 影像分类方法的细碎区域错分问题。验证实验同样在 Vaihingen 数据集上进行，各类别地物的分类精度结果如表6 所示。分析各类地物的分类精度、OA 和 mIoU 等指标，带有辅助信息后处理的 CGAN 影像分类方法分类精度小幅提升，OA 提升了 0.82%，后处理方法改善了一些细小区域的错分问题。

参考图18(e) 和 (g) 的分类可视化结果，辅助信息后处理方法将白色建筑物中间的灰色污垢正确识别为建筑物的一部分，保持了整个建筑物区域的整体类别一致性。基准的 CGAN 影像分类方法中则将其错分为“背景”类别，破坏了该区域的分类完整性。综合量化的分类精度和可视化的分类结果，我们可以论述基于辅助信息的后处理方法在对细碎区域的像素点进行类别预测时，结合了同质性分割单元内其他像素点的预测类别信息，一定程度上将这些细碎区域正确划分为正确的类别，保证了复杂区域的地物类别预测的空间一致性和完整性。

表 7 两种优化方法共同作用对影像分类的影响

方法	地面	低矮植被	树木	建筑物	车辆	OA	mIoU
CGAN	83.78%	74.47%	82.40%	87.64%	78.83%	80.15%	61.83%
CGAN + 融合特征 + 后处理	86.35%	81.56%	83.19%	88.36%	83.19%	85.32%	68.17%
FCN	81.14%	63.39%	79.52%	83.96%	62.39%	78.48%	58.42%
FCN + 融合特征 + 后处理	82.57%	76.87%	81.92%	85.08%	81.33%	81.87%	62.36%

前面的实验结果证明了融合边界特征和基于辅助信息后处理这两种改进方法均能提

高基于 CGAN 影像分类方法的分类精度，现在我们将这两种改进思路同时作用于基于 CGAN 的影像分类中。同时，也将改进方法引入经典的 FCN 分类模型中，比较这两种方法共同作用的分类效果。在 Vaihingen 数据集上的分类结果如表7 所示。分析表7 的分类结果，同时将融合边界特征与后处理这两种操作应用到基准的 FCN 和基于 CGAN 的分类模型中，均能提升影像的分类精度。其中，融合边界特征和带有后处理操作的 CGAN 影像分类方法效果提升最明显，OA 精度达到 85.32%，mIoU 则提升到 68.17%，相比基准的基于 CGAN 影像分类方法，分别有 5.17% 和 6.34% 的绝对精度值的提升。对于 FCN 模型，两种优化方法也提高了模型的分类精度。

图18(c)、(d)、(e) 和 (h) 分别对应 FCN 和基于 CGAN 的基准方法与优化后的方法分类的可视化结果，从图中的分割结果可以发现，融合边界特征与后处理，影像地物边界的划分更合理，此外，同类别内的地物类别更加完整，减少了细碎的错分区域。

综上所述，本章提出的融合边界特征的改进和基于辅助信息后处理的改进方法均提升了基于 CGAN 影像分类方法的分类效果。Vaihingen 数据集上量化精度和目视评估结果均验证了这两种优化方法的正确性。

4.6 本章小结

本章针对高分影像分类结果中存在边界混淆、同类地物存在细碎错分区域的问题，分别从增强分割模块中高阶特征图的边界位置特征信息和像素点分类后处理这两个角度改进基于 CGAN 影像分类方法的不足。首先使用 TFSV-IT2FCM 聚类分割方法对高分影像预处理，获取影像的同质性分割单元。然后，将影像的边界特征信息融合到反卷积上采样的特征图中，提升高阶语义特征的边界位置信息表达能力，提出了融合边界特征的 CGAN 影像分类方法，用于解决影像混淆的分类边界问题。接着，对模型像素点预测的类别分布概率进行加权优化，考虑同质性分割单元内其他像素点的类别预测概率，优化该像素点的概率预测结果，提出基于辅助信息后处理的 CGAN 影像分类方法，由于提升同类地物分类结果的空间一致性。最后，分别在 Vaihingen 数据集上进行实验验证。量化和目视的实验结果均证明了这两种优化方法的正确性和可行性。

第 5 章 总结与展望

5.1 本文的主要内容

遥感影像地物分类是遥感研究领域内一个重要问题，如何对地物类别正确识别、明确不同地物边界问题一直是一个研究热点。本文从正确预测地物边界、同类地物区域内像素保持空间一致性两个角度提出了现有高分影像地物分类方法的改进方法，且对提出的方法分别进行理论推导和实验论证。为了增强影像分割像素点间的连续性，文中将对抗训练的思想应用到 FCN 分类模型中，提出了基于 CGAN 影像分类方法。针对遥感影像混淆边界与模型中边界信息的损失问题，文中提出融合边界特征信息的 CGAN 影像分类方法。针对影像分类中细碎区域的错分问题，文中提出基于辅助信息后处理的 CGAN 影像分类方法。具体研究工作如下：

- (1) 针对全卷积分割方法中上采样特征损失的问题，文中将对抗训练网络思想应用到 FCN 模型中，提出基于 CGAN 的全卷积影像分类方法，提升影像分类效果。融合影像正射影像波段和 DSM 高程波段数据用于模型训练。在 Vaihingen 影像数据上的实验结果表明基于 CGAN 的影像分类方法能够获得更好的分类效果。
- (2) 针对因生成网络中池化操作和上采样操作导致的影像边界、位置信息损失的问题。首先由 TFSV-IT2FCM 方法预处理得到影像分割单元边界图，然后在生成模型上采样操作的特征图内融合相同尺度的边界掩膜特征信息，提升高阶语义特征的边界位置信息，提出了融合边界特征的 CGAN 影像分类方法。实验结果表明融合边界特征的 CGAN 影像分类方法能够获得更明确的地物分类边界。
- (3) 针对生成模型中各像素点未考虑邻近像素点间的关系，从而导致复杂地物类别出现错分区域的问题。文中考虑像素点所属同质性分割单元内其他像素点的类别预测关系，优化像素点类别预测概率，提出了基于辅助信息后处理的影像分类方法。添加后处理的方法在 Vaihingen 数据集上的结果提升了同类地物区域内像素点的空间一致性。
- (4) 将融合边界特征和辅助信息后处理两种改进思路同时引入基于 CGAN 影像分类方法中，改进影像分类的效果，实验结果证明了这两种方法的有效性。

5.2 未来的期望

本文提出的影像分类的改进方法，一定程度上均能提升遥感影像地物分类边界划分的效果，同时也能提高同类别地物区域像素点的空间一致性。之后的研究可从以下方面开展：

- (1) 第3章提出的基于CGAN影像分类方法在识别“车辆”时相对其他几类地物精度较低，原因是训练集影像中含有“车辆”的像素点占比较低，后续研究中构建模型目标函数时可以考虑给像素占比低的地物一个较高的权值，研究加权后的交叉熵代价能否提高像素占比小地物的识别效果。
- (2) 第4章融合边界特征方法中，融合边界掩膜时采取矩阵点乘的操作进行特征融合，其他方式的特征融合方法可以进一步研究探讨。
- (3) 第4章基于辅助信息后处理的方法中只考虑像素点同质性分割单元内其他像素点的类别预测概率。后续研究可以考虑整幅影像中所有像素点之间的相互关系，对该像素点类别预测概率进行处理、优化。

参考文献

- [1] 李石华, 王金亮, 毕艳等. 遥感图像分类方法研究综述 [J]. 国土资源遥感, 2005, 2(5):1--6.
- [2] 张俊, 于庆国, 侯家槐. 面向对象的高分辨率影像分类与信息提取 [J]. 遥感技术与应用, 2010, 25(1):112--117.
- [3] 李德仁, 童庆禧, 李荣兴等. 高分辨率对地观测的若干前沿科学问题 [J]. 中国科学: 地球科学, 2012, 42(6):805--813.
- [4] 邬伦, 承继成, 史文中. 地理信息系统数据的不确定性问题 [J]. 测绘科学, 2006, (5):13--19.
- [5] Hui He, Xianchuan Yu. A comparison of PCA/ICA for data preprocessing in remote sensing imagery classification[C]. Proceedings of MIPPR 2005: Image Analysis Techniques, volume 6044. International Society for Optics and Photonics, 2005. 60--65.
- [6] Geoffrey E Hinton, Simon Osindero, Yee-Whye Teh. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7):1527--1554.
- [7] Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully convolutional networks for semantic segmentation[C]. Proceedings of Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3431--3440.
- [8] Yoshua Bengio, et al. Learning deep architectures for AI[J]. Foundations and trends® in Machine Learning, 2009, 2(1):1--127.
- [9] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks[C]. In: F. Pereira, C. J. C. Burges, L. Bottou, et al., (eds.). Proceedings of Advances in Neural Information Processing Systems 25. Curran Associates, Inc., 2012: 1097--1105.
- [10] John R Jensen, Kalmesh Lulla. Introductory digital image processing: a remote sensing perspective[J]. 1987.
- [11] MBDY Djukanovic, Borivoje Babic, Dijan J Sobajic, et al. Unsupervised/supervised learning concept for 24-hour load forecasting[C]. Proceedings of IEE Proceedings C (Generation, Transmission and Distribution), volume 140. IET, 1993. 311--318.
- [12] AG Wacker, DA Landgrebe. Minimum distance classification in remote sensing[J]. 1972.
- [13] Alan H Strahler. The use of prior probabilities in maximum likelihood classification of remotely sensed data[J]. Remote sensing of Environment, 1980, 10(2):135--163.
- [14] Peter M Atkinson, P Lewis. Geostatistical classification for remote sensing: an introduction[J]. Computers & Geosciences, 2000, 26(4):361--371.

- [15] Frank Paul, Andreas Kääb, Max Maisch, et al. The new remote-sensing-derived Swiss glacier inventory: I. Methods[J]. *Annals of Glaciology*, 2002, 34:355--361.
- [16] James C Bezdek, Robert Ehrlich, William Full. FCM: The fuzzy c-means clustering algorithm[J]. *Computers & Geosciences*, 1984, 10(2-3):191--203.
- [17] 张永生, 巩丹超, 刘军等. 高分辨率遥感卫星应用 [J]. 科学出版社, 北京, 2004.
- [18] Thomas Blaschke. Object based image analysis for remote sensing[J]. *ISPRS journal of photogrammetry and remote sensing*, 2010, 65(1):2--16.
- [19] Robert Lawrence Kettig, DA Landgrebe. Classification of multispectral image data by extraction and classification of homogeneous objects[J]. *IEEE Transactions on geoscience Electronics*, 1976, 14(1):19--26.
- [20] A Lobo, O Chic, A Casterad. Classification of Mediterranean crops with multisensor data: per-pixel versus per-object statistics and image segmentation[J]. *International Journal of Remote Sensing*, 1996, 17(12):2385--2400.
- [21] Martin Baatz. Object-oriented and multi-scale image analysis in semantic networks[C]. Proceedings of Proc. the 2nd International Symposium on Operationalization of Remote Sensing, Enschede, ITC, Aug. 1999, 1999.
- [22] D Geneletti, BGH Gorte. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs[J]. *International Journal of Remote Sensing*, 2003, 24(6):1273--1286.
- [23] Qinghua Guo, Maggi Kelly, Peng Gong, et al. An object-based classification approach in mapping tree mortality using high spatial resolution imagery[J]. *GIScience & Remote Sensing*, 2007, 44(1):24--47.
- [24] John Canny. A computational approach to edge detection[C]. *Proceedings of Readings in computer vision*. Elsevier, 1987: 184--203.
- [25] Nobuyuki Otsu. A threshold selection method from gray-level histograms[J]. *IEEE transactions on systems, man, and cybernetics*, 1979, 9(1):62--66.
- [26] Luc Vincent, Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1991, (6):583--598.
- [27] Radhakrishna Achanta, Appu Shaji, Kevin Smith, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(11):2274--2282.
- [28] Lior Weizman, Jacob Goldberger. Urban-area segmentation using visual words[J]. *IEEE Geoscience and Remote Sensing Letters*, 2009, 6(3):388--392.

- [29] Marie Lienou, Henri Maitre, Mihai Datcu. Semantic annotation of satellite images using latent Dirichlet allocation[J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(1):28--32.
- [30] Hui He, Tianheng Liang, Dan Hu, et al. Remote sensing clustering analysis based on object-based interval modeling[J]. Computers & Geosciences, 2016, 94:131--139.
- [31] Mahesh Pal. Random forest classifier for remote sensing classification[J]. International Journal of Remote Sensing, 2005, 26(1):217--222.
- [32] Johan AK Suykens, Joos Vandewalle. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3):293--300.
- [33] Mark A Friedl, Carla E Brodley. Decision tree classification of land cover from remotely sensed data[J]. Remote sensing of environment, 1997, 61(3):399--409.
- [34] Simon Haykin. Neural networks[M], volume 2. Prentice hall New York, 1994.
- [35] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm[C]. Proceedings of icml, volume 96. Citeseer, 1996. 148--156.
- [36] Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep learning[J]. nature, 2015, 521(7553):436.
- [37] Christian Szegedy, Wei Liu, Yangqing Jia, et al. Going deeper with convolutions[C]. Proceedings of Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 1--9.
- [38] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep residual learning for image recognition[C]. Proceedings of Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770--778.
- [40] Yoshua Bengio, Pascal Lamblin, Dan Popovici, et al. Greedy layer-wise training of deep networks[C]. Proceedings of Advances in neural information processing systems, 2007. 153--160.
- [41] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, et al. Extracting and composing robust features with denoising autoencoders[C]. Proceedings of Proceedings of the 25th international conference on Machine learning. ACM, 2008. 1096--1103.
- [42] Andrew Ng, et al. Sparse autoencoder[J]. CS294A Lecture notes, 2011, 72(2011):1--19.
- [43] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets[C]. Proceedings of Advances in neural information processing systems, 2014. 2672--2680.

- [44] Mehdi Mirza, Simon Osindero. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.
- [45] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
- [46] Xi Chen, Yan Duan, Rein Houthooft, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]. Proceedings of Advances in neural information processing systems, 2016. 2172--2180.
- [47] Martin Arjovsky, Soumith Chintala, Léon Bottou. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
- [48] Fan Hu, Gui-Song Xia, Jingwen Hu, et al. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery[J]. Remote Sensing, 2015, 7(11):14680--14707.
- [49] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, et al. Land use classification in remote sensing images by convolutional neural networks[J]. arXiv preprint arXiv:1508.00092, 2015.
- [50] Adriana Romero, Carlo Gatta, Gustau Camps-Valls. Unsupervised deep feature extraction for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(3):1349--1362.
- [51] Michael Kampffmeyer, Arnt-Borre Salberg, Robert Jenssen. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks[C]. Proceedings of Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016. 1--9.
- [52] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, et al. Fully convolutional neural networks for remote sensing image classification[C]. Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016. 5071--5074.
- [53] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-net: Convolutional networks for biomedical image segmentation[C]. Proceedings of International Conference on Medical image computing and computer-assisted intervention. Springer, 2015. 234--241.
- [54] Ruirui Li, Wenjie Liu, Lei Yang, et al. Deepunet: A deep fully convolutional network for pixel-level sea-land segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, (99):1--9.
- [55] Zhengxin Zhang, Qingjie Liu, Yunhong Wang. Road extraction by deep residual u-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5):749--753.
- [56] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks[C]. Proceedings of Advances in neural information processing systems, 2012. 1097--1105.

- [57] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, et al. A tutorial on the cross-entropy method[J]. *Annals of operations research*, 2005, 134(1):19--67.
- [58] Diederik P Kingma, Jimmy Ba. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [59] Tao Jiang, Dan Hu, Xianchuan Yu. Enhanced IT2FCM algorithm using object-based triangular fuzzy set modeling for remote-sensing clustering[J]. *Computers & geosciences*, 2018, 118:14--26.
- [60] Lotfi Asker Zadeh. Fuzzy sets as a basis for a theory of possibility[J]. *Fuzzy sets and systems*, 1978, 1(1):3--28.
- [61] Roger A Horn. The hadamard product[C]. *Proceedings of Proc. Symp. Appl. Math*, volume 40, 1990. 87--169.

学术成果

1. **Tao Jiang**, Dan Hu, and XXX(导师). Enhanced IT2FCM algorithm using object-based triangular fuzzy set modeling for remote-sensing clustering[J]. Computers & geosciences, 2018, 118: 14-26. (SCI 三区收录, 检索号:GQ6TC.)
2. Dan Hu, **Tao Jiang**, and XXX(导师). The construction of non-convex fuzzy sets. 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, 2017, pp. 1175-1181. (EI 收录, 检索号:20183005590187.)
3. Haihua Xing, Hui He, XXX(导师), Dan Hu, **Tao Jiang**. An Interval Type-2 Fuzzy Sets Generation Method for Remote Sensing Imagery Classification[J]. Computers & geosciences, 2019.(SCI 三区, accept)

研究生期间主要参与项目

- (1) 自适应循优 n 型模糊系统及其应用研究, 国家自然科学基金面上项目 (11471045),
2015.01-2018.12.
- (2) 基于深度学习的遥感崩滑地质灾害信息提取, 北京市自然科学基金 (L172029),
2017.10- 2019.12.

致 谢

快乐的时间总是过得很快，转瞬之间，我在北京师范大学三年的硕士研究生生活即将结束。离别之际，我要向帮助和关照我的老师、家人、同学和好友们表示感谢，谢谢你们！

首先感谢我的两位指导老师：余先川教授和胡丹副教授。还记得三年前刚来北京师范大学学习的时候，我对实验室研究工作和科研方向都比较迷茫，是胡丹老师指导我开展科研学习的工作，是胡老师告诉我如何去看论文，怎样学习研究学者思考问题的方式。每周一到两次的讨论中胡老师也时常跟我探讨课题中的知识，对于我的困惑与科研中遇到的问题，胡老师都耐心的和我沟通、讨论。研二下学期伊始，胡丹老师去美国从事新的事业，虽然远隔重洋，但还时常接收到胡老师来自大洋彼岸的关心和问候。同时胡老师在生活上对我也非常热心，有时候生活上不如意或遇到烦心事，胡老师也体贴入微，关心并疏导我的个人问题，感谢胡老师近两年对我学习和生活上的关怀。胡老师出国后，我转到现在的导师余先川教授下继续从事科研学习。我也由衷地感谢余老师，余老师在专业学科领域有极其敏锐的洞察力，能够迅速了解当前领域内研究的前沿内容。在余老师的指导下，我能够在遥感影像数据挖掘领域继续深入研究，提升自己的科研能力。毕业论文从开题到撰写过程中，余老师都认真、耐心地指导我并指出论文中存在的问题与不足。在生活上，余老师待人亲切，关心学生，更像是实验室大家庭的家长，实验室经常性的聚餐、实验室每个小伙伴生日会的庆祝与祝福都让远离家乡，漂泊北京的我们感受到了来自家庭的温暖。此外，余老师也会带着我们去参加一些前沿学术论坛和会议，让我们接触到学术界的科研大牛的同时，也开阔拓展了我们的视野。再次对我两位导师表示致谢！

其次感谢实验室的张立保老师，张老师严谨的科研态度值得我学习，另外感谢张老师在电子楼 512 室给我提供的科研工位，让我更便利地从事科研学习。同时感谢师母的关照，感谢师母对实验室学生的关心和支持，师母也让我们体会到慈爱家长般的温暖。

感谢实验室的每一位同学，和大家在一起学习、生活的时光让我快乐、开心，感谢实验室里每一位师兄师姐师弟和师妹们。

感谢我的室友冯思博、朱云宗和戎博杰，谢谢你们三年的陪伴，大家一起度过的研究生时光也会成为我心中的美好记忆。

最后，我要感谢我的父母亲人们对我的关怀和照顾。你们在我背后默默的奉献与付出，一直是我坚持前行的动力，是你们在我迷茫的时候给我支持和力量，鼓励我面对困

难，笑对困难与挫折。

漫漫人生路，几度欢乐，几度忧愁，正因此才能体会到生命与生活的多彩。未来的路还很漫长，心怀感恩砥砺前行。

江涛

2019年5月