

北京师范大学

# 硕士学位论文

论文题目：基于模糊聚类及生成对抗网络的遥感影像  
分类方法研究

作 者：江涛

导 师：余先川教授

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：\_\_\_\_\_

# 基于模糊聚类及生成对抗网络的遥感影像分类方法研究

## 摘要

遥感影像分类是遥感领域重要的研究内容，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。高分辨率影像提供高质量细节信息时其同类地物类内差异大，使得地物分类更容易错分，且难以预测出明确、清晰的地物边界。

面向对象的模糊聚类分割方法因其优秀的不确定信息表征能力，被广泛用于处理高分影像中数据不确定性。现有的面向对象分割方法中对影像分割单元多采取均值数据建模或区间值数据建模，然而，这两种信息表达模型无法区分具有相同均值和区间值但内部分布不一致的分割单元。因此，设计一种新的数据模型表征影像分割单元内部特征有利于提高模糊聚类方法的分类能力。面向对象的聚类分割方法属于无监督分类，需要人工后处理。随着深度学习的发展和全卷积分割网络的出现，有监督的图像分割方法也被广泛应用于遥感影像分类中。现有全卷积分类网络中上采样操作造成影像特征的损失，导致地物分类边界模糊，设计一种新的网络模型提高影像分类同类地物的空间一致性，也具有研究意义。基于上述两点思路，本文将从影像分割单元特征建模和地物一致性处理两个角度，对遥感影像地物分类问题加以分析和讨论。具体的研究内容如下：

- (1) 从影像分割单元特征建模角度改进现有面向对象模糊聚类分割方法。文中设计了三角形模糊集值信息表达模型来表征影像分割单元，其次提出一种新的区间值距离来度量两个三角形模糊集值数据的相异性，最后结合新提出的三角形模糊集值数据模型和区间值距离提出基于三角形模糊集值的区间二型模糊聚类方法。该方法在 SPOT 5 和高分二号高分影像数据上的聚类分割结果验证了其能更好的刻画影像数据的不确定性特征。
- (2) 从地物一致性的角度改进现有全卷积有监督分类方法。文中借鉴生成对抗网络优秀的图像生成能力，将条件生成对抗网络框架应用到现有全卷积网络分割模型中，提出基于条件生成对抗网络的全卷积影像分割方法。新方法在 Vaihingen 影像数据集上的分割结果有更准确的地物边界，同类别地物空间具有一致性。

**关键词：** 遥感影像，地物分类，模糊聚类，神经网络，全卷积网络，生成对抗网络

# **Remote Sensing Image Classification Based on Fuzzy Clustering and 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.

Object-oriented fuzzy clustering segmentation methods are widely used to handle uncertainty in high-resolution images because of its excellent ability to characterize uncertain information. In the current object-oriented segmentation methods, the image segmentation unit adopts mean data modeling or interval value data modeling. However, accurately distinguishing two objects with the same mean or interval values and different internal distributions is difficult. Therefore, we designed a triangular fuzzy set modeling to describe objects and designed an interval distance metric to measure the dissimilarities between triangular fuzzy sets. The object-oriented clustering methods belong to unsupervised classification methods, which required manual post-processing. With the development of deep learning and the emergence of a full convolutional network, supervised image segmentation methods are also widely used in remote sensing image classification. In full convolutional network, the upsampling operation causes the loss of image features, which leads to the blurring of the boundary of the feature classification. Designing a new network model to improve the spatial consistency of image classification of similar features also has research significance. Based on the above two ideas, this paper will analyze and discuss the remote sensing image feature classification from the perspective of image segmentation unit data modeling and ground object consistency processing. The specific research contents are as follows:

- (1) Improves the existing object-oriented fuzzy clustering segmentation method from the perspective of image segmentation unit feature modeling. In this paper, a triangular fuzzy set modeling

was designed to describe objects, and an interval distance metric was proposed to measure the dissimilarities between triangular fuzzy sets. Finally, the triangular fuzzy set-valued data interval type 2 fuzzy c means clustering method was proposed. The clustering results of SPOT-5 and Gaofen-2 high-resolution image data verified that the proposed method results in improved classification quality and accuracy.

- (2) Improves the existing full convolutional supervised classification method from the perspective of feature consistency. This paper draws on the excellent image generation ability of the generative adversarial network, applies the conditional generative adversarial network to the existing full convolutional network segmentation model, and proposes a full convolutional image segmentation method based on conditional generative adversarial network. The segmentation results of the proposed method on the Vaihingen image dataset have more accurate feature boundaries and consistency with the same type of feature space.

**KEY WORDS:** Remote Sensing Imagery, Land Cover Classification, Fuzzy Clustering, Convolutional Neural Network, Fully Convolutional Network, Generative Adversarial Network

# 目 录

摘要 .....	I
ABSTRACT .....	II
插图索引 .....	VI
表格索引 .....	VII
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 基于深度学习框架的影像分类研究现状 .....	4
1.2.3 存在的主要问题 .....	5
1.3 本文的组织结构 .....	5
1.4 本文主要创新点 .....	6
第2章 影像分类相关理论基础 .....	7
2.1 遥感影像的模糊聚类分割基础 .....	7
2.1.1 影像分割方法 .....	7
2.1.2 模糊聚类方法 .....	10
2.2 基于神经网络的影像分割相关理论 .....	11
2.2.1 卷积神经网络基础 .....	12
2.2.2 基于全卷积网络的图像分割 .....	14
2.3 本章小结 .....	17
第3章 基于区间二型模糊聚类的遥感影像无监督分割方法 .....	18
3.1 引言 .....	18
3.2 三角形模糊集值建模与相似性度量 .....	18
3.3 面向对象的改进型区间二型模糊遥感影像聚类方法 .....	22
3.3.1 算法整体框架 .....	22
3.3.2 面向对象的 TFSV 模型 .....	23
3.3.3 基于 TFSV 数据模型的面向对象模糊聚类方法 .....	24
3.3.4 后处理 .....	28
3.4 实验结果与分析 .....	29
3.4.1 神湾地区实验 .....	30

3.4.2 横琴地区实验 .....	32
3.4.3 三门峡地区实验 .....	35
3.5 本章小结 .....	36
<b>第 4 章 基于生成对抗网络的遥感影像分割方法 .....</b>	<b>37</b>
4.1 引言 .....	37
4.2 基于生成对抗网络框架的影像分割方法 .....	37
4.2.1 生成对抗网络框架 .....	37
4.2.2 基于 CGAN 的全卷积影像分割方法 .....	39
4.3 实验数据介绍与预处理 .....	41
4.3.1 Vaihingen 数据介绍 .....	41
4.3.2 数据预处理 .....	41
4.4 实验结果与分析 .....	44
4.4.1 实验环境 .....	44
4.4.2 评价指标 .....	44
4.4.3 网络参数 .....	45
4.4.4 结果与分析 .....	47
4.5 本章小结 .....	49
<b>第 5 章 总结与展望 .....</b>	<b>51</b>
5.1 本文的主要内容 .....	51
5.2 未来的期望 .....	51
<b>参考文献 .....</b>	<b>53</b>
<b>学术成果 .....</b>	<b>58</b>
<b>致谢 .....</b>	<b>59</b>

## 插图索引

图 1	Watershed 分水岭算法示意图	8
图 2	SLIC 算法分割成尺寸为 64, 256 和 1024 的超像素单元	9
图 3	图像卷积示意图	13
图 4	最大池化与平均池化	14
图 5	典型的卷积网络结构	15
图 6	影像分割网络结构示意图	15
图 7	反卷积上采样过程	16
图 8	基于 FCN 网络的遥感影像分割示意图	17
图 9	区间值相同但分布不同的数据比较	18
图 10	三角形模糊集 $\tilde{A}$ 示意图	19
图 11	$\tilde{A}$ 和 $\tilde{B}$ 位置关系示意图。(a) 相交; (b) $\tilde{A}$ 包含集合 $\tilde{B}$ ; (c) 不相交.	20
图 12	不同位置下, $\tilde{A}$ 和 $\tilde{B}$ 两者间各种距离比较图。	21
图 13	面向对象的改进型区间二型模糊遥感影像聚类算法框架图	23
图 14	TFSV-IT2FCM 算法流程图	26
图 15	神湾研究区不同方法分类结果示意图	30
图 16	神湾研究区 TFSV-IT2FCM 与 IV-FCM 算法聚类分割图	33
图 17	三门峡研究区不同方法分类结果示意图	34
图 18	GAN 结构示意图	37
图 19	CGAN 结构示意图	38
图 20	基于对抗训练框架的全卷积语义分割模型示意图	39
图 21	Vaihingen 影像数据	42
图 22	图像裁剪方式	43
图 23	影像分割的可视化结果	48
图 24	DSM 波段对影像分类结果的影响	49

## 表格索引

表 1	实验区土地覆被类别表 .....	29
表 2	神湾地区影像地物类别抽样 Ground truth 表 (pixels).....	32
表 3	神湾研究区不同方法分类精度表.....	33
表 4	横琴地区分类结果表 .....	34
表 5	横琴地区分类精度表 .....	35
表 6	三门峡地区分类结果表 .....	35
表 7	三门峡地区分类精度表 .....	36
表 8	Vaihingen 数据类别标签颜色对照表 .....	42
表 9	基于 CGAN 框架的全卷积分割模型参数表 .....	46
表 10	Vaihingen 数据分割精度评估表 .....	48
表 11	DSM 波段对分割精度的影响 .....	50

# 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<sup>[1]</sup>。近年来，随着卫星遥感技术的发展和信息技术的完善，遥感影像分辨率不断提高，高分辨率影像信息量越来越丰富。同时全球遥感数据呈爆发式增长，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<sup>[2]</sup>，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；分类器也从单一的分类器发展为层叠或多个分类器相结合的方法<sup>[3]</sup>。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，由于遥感影像存在混合像元，同物异谱和同谱异物等问题<sup>[4]</sup>，遥感影像数据固有的不确定性成为影像分类亟需解决的问题，如果能构建适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<sup>[5]</sup>。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。

模糊聚类方法（Fuzzy clustering methods）<sup>[6]</sup> 是上世纪八九十年代提出的非监督聚类方法，它的核心优势在于构建的影像数据的不确定信息表达模型，具有良好的特征表达能力。随着遥感影像分辨率的提高，近几年，基于像素的模糊聚类方法发展为面向对象的模糊聚类方法，被广泛应用到高分影像的聚类分割任务中。相关学者对面向对象的影像分割单元采用均值数据建模<sup>[7]</sup> 或区间值数据建模<sup>[8]</sup>，均取得了不错的地物分类效果。

深度学习（Deep learning, DL）<sup>[9]</sup> 和全卷积分割网络（Fully convolutional network, FCN）<sup>[10]</sup> 模型是近几年兴起并迅速发展的有监督影像分类方法。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。基于全卷积网络的图像分割方法相比传统机器学习方法具有更大的优势，被广泛用于遥感影像有监督分类中。

本文将从遥感影像不确定性数据建模和地物分类一致性两个角度，对高分影像地物分类问题进行研究，并改进相关地物分类方法。一方面设计三角形模糊集值信息表达模型来表征影像分割单元特征，提出基于三角形模糊集值的区间二型模糊聚类分割方法用于高分影像聚类分割处理。另一方面，将深度学习中生成对抗网络的思想应用到全卷积分割模型中，借助生成网络优秀的图像生成能力，新提出的基于生成对抗网络的分割方法能生成更准确的地物分割边界，且同类别地物空间具有一致性。

## 1.2 国内外研究现状

为了方便介绍，本文中将深度学习方法之前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<sup>[9,11,12]</sup>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<sup>[13]</sup>。遥感影像分类方法依据是否使用地物类别先验知识分为监督分类和非监督分类。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；非监督分类是指在缺乏样本类别先验知识的前提下，只根据样本数据本身特性进行分类，根据样本相似度划分类别，如聚类<sup>[14]</sup>。根据分类单元不同，遥感影像的分类方法可划分为基于像元和面向对象的分类。基于像元的分类方法以像元的光谱信息作为主要依据进行分类，常见的基于像元的遥感影像分类方法有：最小距离分析法<sup>[15]</sup>，最大似然分类法<sup>[16]</sup>，K-均值聚类法<sup>[17]</sup>，ISODATA 聚类法<sup>[18]</sup>和模糊 C 均值聚类方法<sup>[6]</sup>等。随着遥感技术不断发展与成熟，遥感影像空间分辨率不断提高。一般地，我们将空间分辨率高于 5m 遥感影像称作高分辨率遥感影像<sup>[19]</sup>。高分影像相比低分辨率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<sup>[20]</sup>。

面向对象的高分辨率影像分类方法将影像中邻域同质像元组成的对象作为分类单元，

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<sup>[19]</sup>。早在 1976 年, Kettig 和 Robert<sup>[21]</sup> 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<sup>[22]</sup> 将面向对象分类方法应用到遥感影像分类中,通过实验证明了在高分影像识别任务中面向对象的分类方法比基于像元的方法识别速度更快,分类精度更高。Baatz<sup>[23]</sup> 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到高分影像分类识别任务中,发展迅速。Geneletti<sup>[24]</sup> 和 Guo<sup>[25]</sup> 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。文献<sup>[7,8]</sup> 则从刻画遥感数据不确定性的角度提出面向对象模糊聚类方法,且都取得了不错的分割结果。在工业应用领域,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的高分影像分类方法的商业发展,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三个部分:影像分割,特征提取和分类识别。高分影像因其空间分辨率高,纹理、形状等空间信息相对丰富,分割方法和精度便成为影像分割的关键要素。Canny 通过提出 Canny 算子<sup>[26]</sup> 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<sup>[27]</sup>。Vincent<sup>[28]</sup> 等结合沉浸模型提出影像的分水岭分割。Achanta 和 Radhakrishna<sup>[29]</sup> 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,该方法目前被广泛应用到影像分割中。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<sup>[30]</sup> 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<sup>[31]</sup> 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<sup>[8]</sup> 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的相关方法,设计一种区间值特征来提取目标地物的特征,用于面向对象的非监督分类。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证。分类识别阶段是针对特征提取阶段获得的分割对象的特征,利用分类器对待识别目标进行分类识别。目前,常用的机器学习分类方法包含随机森林<sup>[32]</sup>,支持向量机<sup>[33]</sup>,决策树<sup>[34]</sup> 和神经网络模型<sup>[35]</sup> 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<sup>[36]</sup> 的方法也被应用到高分影像分类中。

然而,传统的高分影像的分类与识别方法只应用到影像中、低层特征,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,无法充分学习和表达遥感影像复杂的数据结构和特征信息。因此,研究结构更复杂,表达能力更强的影像分

类模型具有重要的意义。

### 1.2.2 基于深度学习框架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<sup>[9]</sup> 教授于 2006 年提出。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，从大量数据中自动学习数据的特征表示。深度学习利用多层网络模型学习抽象概念完成自我学习<sup>[37]</sup>。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<sup>[11]</sup>。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<sup>[12]</sup> 模型，对 ImageNet 数据集上千万级的自然图像进行分类，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<sup>[38]</sup> 和牛津大学学者提出的 VGGNet<sup>[39]</sup> 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<sup>[40]</sup> 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，深度学习算法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)<sup>[9]</sup>。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)<sup>[41]</sup>、降噪编码器 (Denoise Autoencoder, dAE)<sup>[42]</sup> 和稀疏自编码器 (Sparse AutoEncoder, SAE)<sup>[43]</sup> 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)<sup>[44]</sup> 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<sup>[45]</sup>、DCGAN<sup>[46]</sup>、InfoGAN<sup>[47]</sup> 和 WGAN<sup>[48]</sup> 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<sup>[49]</sup> 利用迁移学习知识，首次将深度卷积神经网络应用到高分辨率遥感影像场景分类中，能有效学习影像的高级特征表示。Marco 等人<sup>[50]</sup> 将预训练的 GoogLeNet 网络参数，迁移到 UC Merced 土地利用数据集<sup>①</sup> 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<sup>[51]</sup> 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<sup>[52]</sup> 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<sup>[53]</sup> 基于全卷积网络（Fully convolutional network，FCN）结构，对影像进行四层卷积下采样提取特征，接着四层反卷积对特征图上采样回初始影像分辨率，输出所有像素点类别，实现遥感影像的像素级分类。U-Net<sup>[54]</sup> 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<sup>[55]</sup> 基于 U-Net 网络完成对海陆影像水域 - 陆地分割识别。文献<sup>[56]</sup> 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。

### 1.2.3 存在的主要问题

结合高分影像研究现状可知，影像分类与识别方法发展迅速。传统机器学习分类方法只能提取浅层特征，且分类器结构相对简单。基于深度学习理论的影像分类方法具有很大的应用潜力，但是其需要大量标记好的训练样本，模型难以训练。高分影像分类中遥感影像不确定信息的表达和预测同类地物的一致性均是可研究和改进的方向，恰当的特征建模和分类方法可以进一步提升影像分类的精度。

## 1.3 本文的组织结构

本文主要从遥感影像不确定性数据建模和地物分类一致性两个角度对高分影像分类识别展开研究。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。

---

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

**第 2 章：**详细介绍了遥感影像分类相关的理论基础。首先介绍了面向对象的模糊聚类分割方法中常用的影像分割方法和经典的模糊聚类方法。然后介绍了深度学习卷积神经网络的基本结构和相关概念，同时介绍了基于全卷积网络的影像分割模型，包含全卷积网络的结构、模型以及网络特点。

**第 3 章：**首先针对遥感数据的不确定信息建模，设计三角形模糊集值模型，并提出新的区间值度量方法，并通过理论推导证明新提出的数据模型和距离度量的合理性。然后结合新提出的数据模型和距离度量提出改进的区间二型模糊聚类方法。最后，将新提出的方法应用到多来源数据的聚类分割实验中，对比其他模糊聚类分割方法，验证新提出方法的有效性。

**第 4 章：**首先介绍生成对抗网络的基本框架和模型方法。然后将条件生成对抗网络模型应用到全卷积分割方法中，提出基于条件生成对抗网络的影像分割方法，并详细介绍了新提出方法的模型结构和目标函数。接着对实验中使用到的 Vaihingen 数据集介绍和预处理，获得实验需要的训练集和测试集数据。最后，比较新提出的方法与对比方法分割结果，验证新提出方法的有效性。

**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四个创新点：

- (1) 设计三角形模糊集值信息表达模型来表征影像分割单元，能表征分割单元数据内部特征信息。
- (2) 针对两个三角形的模糊集值数据，提出一种新的区间值距离度量，能够表征数据间的相异性。
- (3) 结合新提出的数据模型和区间距离度量，改进现有的面向对象的模糊聚类分割方法，并应用到高分影像分割实验中。
- (4) 将生成对抗网络的思想应用到全卷积影像分割模型框架中，提出的基于 CGAN 的分割方法能生成更好的地物边界，且保持空间一致性。

## 第 2 章 影像分类相关理论基础

本文主要从遥感影像不确定性数据建模和地物分类一致性两个角度对高分影像分类识别展开研究。第 3 章研究内容是对面向对象的模糊聚类影像无监督分割方法改进，这部分工作基于遥感影像的模糊聚类分割基础知识。第 4 章则是针对全卷积分割网络中边界模糊、类内不一致性问题，提出基于条件生成对抗网络框架的影像分割方法，该章节中大量工作基于卷积神经网络和全卷积语义分割网络模型相关知识。所以本章将依次介绍遥感影像模糊聚类分割方法，深度神经网络，全卷积分割网络相关的理论知识。

### 2.1 遥感影像的模糊聚类分割基础

本课题中处理的遥感影像数据均为高分辨率遥感影像数据，文中主要研究面向对象的遥感影像模糊聚类方法。面向对象的模糊聚类方法主要包含影像分割，模糊聚类和后处理三个部分。后处理部分包含借助辅助软件人工处理相关的内容，不是本文研究重点。下面依次介绍影像分割和模糊聚类方法相关知识与算法原理。

#### 2.1.1 影像分割方法

面向对象的影像分割是指对遥感影像分割划分，将邻域内同质性的像元当作一个整体，形成的影像分割单元区域内部的颜色、灰度、纹理等特性相似性较大，区域间的颜色、灰度、纹理等特性相异性较大。常见的影像分割根据分割类型可以划分为：边缘检测分割方法、阈值分割方法和基于区域的分割方法。边缘检测分割方法是通过 Sobel、Canny、LOG 等滤波算子提取影像边缘，再将边缘连接起来形成边界，一般适用于灰度影像的分割。阈值分割方法通过比较像素与设定阈值大小关系，划分影像的前景和背景，该方法简单但不适用于较复杂的影像。基于区域的分割方法包含生长法和区域分裂合并法，其原理依据是影像中距离近的像素点相似性大，距离远的相似性小。基于区域的分割方法应用广泛，且分割效果较好。文中详细介绍两种常见的分割方法：Watershed 分水岭算法<sup>[28]</sup> 和 SLIC 超像素分割算法<sup>[29]</sup>。

## 1. Watershed 分水岭算法

Watershed 分水岭分割算法是一个非常经典的区域生长分割方法。算法的思想是把待分割图像想象成一个地理上的地形图，图像灰度值大小可看作地形高程值，图像每个局部极小值及其影响区域对应地形上的低洼处，即集水盆。两个集水盆汇合边界则为分水岭。分水岭算法如图 1 所示。

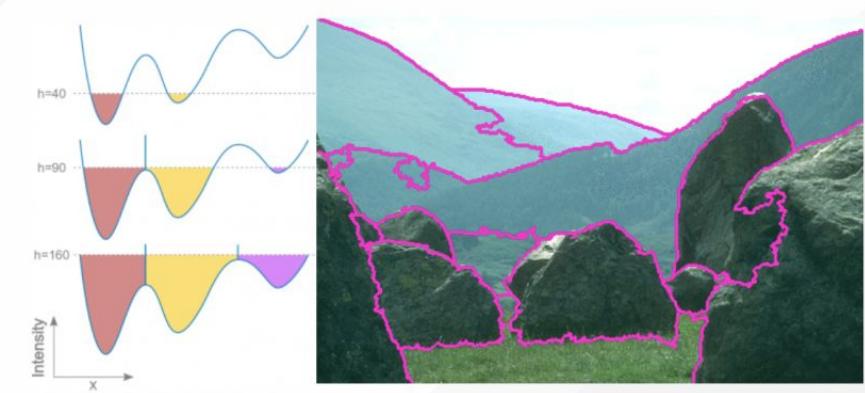


图 1 Watershed 分水岭算法示意图

根据 Watershed 分水岭算法原理，为得到图像的边缘信息，通常把梯度图像作为输入图像，即

$$g(x,y) = \nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2-1)$$

其中， $f(x,y)$  表示原始图像， $g(x,y)$  表示原始图像的梯度运算。假设  $R_1, R_2, R_3, \dots, R_m$  表示待分割图像的极小区域的集合， $C(R_i)$  表示与极小区域  $R_i$  相关的流域， $n$  表示溢流的增加数值（即在第  $n$  步时溢流的深度）， $T[n]$  表示满足梯度  $g(x,y) < n$  的所有像素点  $(x,y)$  的集合。对于一个给定流域，假设在第  $n$  步时极小区域  $R_i$  发生溢流，令  $C_n(R_i)$  为与极小区域  $R_i$  相关流域的一部分，即在溢流深度  $n$  时，在流域  $C(R_i)$  中形成的水平面构成的区域， $C_n(R_i)$  为二值图像，可表示为：

$$C_n(R_i) = C(R_i) \cup T[n] \quad (2-2)$$

为防止分水岭算法产生的过度分割，通常设置阈值对梯度函数进行修改，以消除灰度的微小变化产生的过度分割。即

$$g(x,y) = \max(\nabla f(x,y), g_\theta) \quad (2-3)$$

式中,  $g_\theta$  表示阈值。

## 2. SLIC 超像素分割算法

SLIC(Simple linear iterative clustering), 即简单的线性迭代聚类超像素分割方法。它采用  $K$  均值聚类方法高效地生成超像素, 较以前的算法可以更好地获取分割边界, 且有着更快的运行速度。

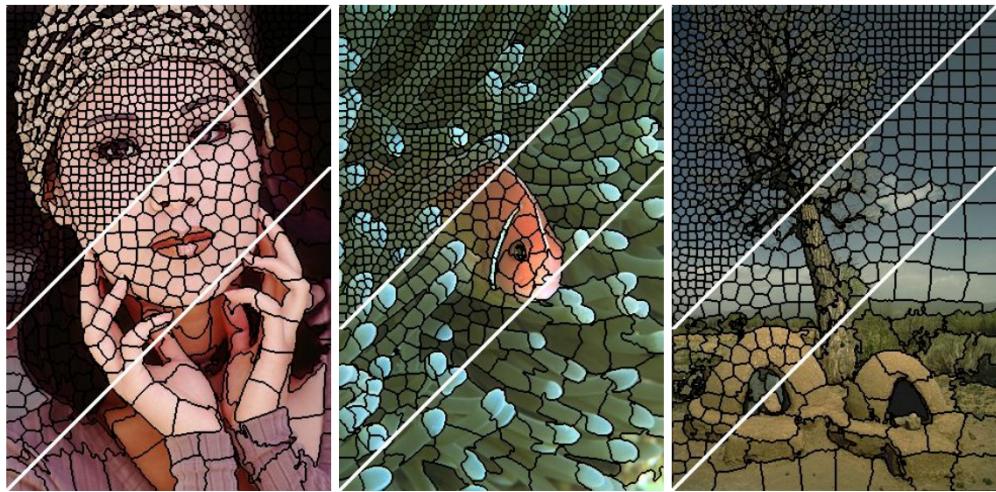


图 2 SLIC 算法分割成尺寸为 64, 256 和 1024 的超像素单元

SLIC 算法将彩色图像转化为 CIELAB 颜色空间和 XY 坐标下的 5 维特征向量, 即位于  $(x, y)$  处的像元在 CIELAB 颜色空间表示为  $[l, a, b, x, y]$ , CIELAB 空间由 LAB 色彩模型表征,  $l$  为亮度,  $a, b$  为色彩值。对有  $N$  个像素点的图像, 预分割为  $K$  个相同尺寸的超像素, 则每个超像素单元的大小为  $N/K$ , 则相邻种子间的步长为  $S = \sqrt{N/K}$ 。SLIC 的搜索范围限制为  $2S * 2S$ , 对于每个搜索到的像素点, 分别计算它和该种子点的距离, 包括颜色距离和空间距离。距离计算方法如下

$$\begin{aligned} d_c &= \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \\ d_s &= \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\ D &= \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \end{aligned} \quad (2-4)$$

式中,  $d_c$  表示颜色距离,  $d_s$  表示空间距离,  $N_s$  表示类内最大空间距离, 定义为

$$N_s = S = \sqrt{N/K} \quad (2-5)$$

$N_c$  为最大的颜色距离, 取值随具体的图片而定, 实验中一般取一个固定常数  $m$  (取值范围为  $(1 \leq m \leq 40)$ , 一般取 10。  $D$  表示总距离, 即最终的距离度量为:

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad (2-6)$$

通过迭代优化直至误差收敛即可完成图像的超像素分割。图 2 所示为基于 SLIC 算法分别将图像大约分成 64, 256 和 1024 三个尺寸的结果。

## 2.1.2 模糊聚类方法

模糊 C 均值 (fuzzy c-means, FCM) 聚类方法是结合模糊数学不确定性理论对 K-means 聚类方法的改进。与 K-means 算法中依据欧式距离作为类别划分依据的硬分类不同, FCM 算法是以隶属度为类别划分依据的软分类方法。FCM 方法是一种柔性的模糊划分, 因此 FCM 比 K-means 算法更能突出遥感影像不确定性的特点。在复杂、高维遥感影像实际应用中, FCM 算法效果显著优于 K-means 算法。

介绍 FCM 方法前, 先引入模糊集合中隶属度函数的概念。隶属度函数表示一个对象  $x$  属于集合  $A$  的程度的函数, 记作  $\mu_A(x)$ 。其自变量是可能属于集合  $A$  的所有元素, 因变量是  $[0, 1]$  区间内的值, 即满足  $0 \leq \mu_A(x) \leq 1$ 。 $\mu_A(x) = 0$  表示元素  $x$  不属于集合  $A$ , 等价于普通集合中  $x \notin A$ ;  $\mu_A(x) = 1$  则表示  $x$  完全隶属于  $A$ , 等价于  $x \in A$ 。一个定义在论域  $X$  上的隶属度函数就定义了一个模糊集, 即模糊集  $A$  可由其成员  $x$  的隶属函数  $\mu(x)$  来刻画, 如式 2-7:

$$A = \{(x, \mu(x)) | \forall x \in X, \mu(x) \in [0, 1]\} \quad (2-7)$$

有了模糊集合与隶属度的概念, 一个元素属于模糊集合就不是硬性的了。对应到聚类问题中, 可以把聚类生成的簇看作模糊集合, 因此, 每个样本点隶属于簇的隶属度就是  $[0, 1]$  区间内的值。

假定数据集  $X = \{x_1, x_2, \dots, x_n\}$ , 把数据划分为  $m$  个类别, 聚类中心为  $C = \{c_1, c_2, \dots, c_m\}$ , 样本  $x_i$  属于类别  $c_j$  的隶属度为  $\mu_{ji}$ , 则 FCM 方法优化的目标函数为式 2-8:

$$J = \sum_{j=1}^m \sum_{i=1}^n \mu_{ji}^k \|x_i - c_j\|^2 \quad (2-8)$$

其中  $1 \leq k \leq \infty$  为模糊指数，目标函数  $J$  的约束条件为式 2-9：

$$\sum_{j=1}^m \mu_{ij} = 1, i = 1, 2, \dots, n \quad (2-9)$$

通过拉格朗日乘子法将约束条件写入目标函数式中，则：

$$J = \sum_{j=1}^m \sum_{i=1}^n \mu_{ji}^k \|x_i - c_j\|^2 + \lambda_1(\sum_{j=1}^m \mu_{j1} - 1) + \dots + \lambda_i(\sum_{j=1}^m \mu_{ji} - 1) + \dots + \lambda_n(\sum_{j=1}^m \mu_{jn} - 1) \quad (2-10)$$

令式 2-10 分别对  $\mu_{ji}$  和  $c_j$  求偏导并让求导结果等于 0，化简求解，有：

$$\mu_{ji} = \frac{1}{\sum_{s=1}^m (\frac{\|x_i - c_s\|}{\|x_i - c_s\|})^{(\frac{2}{k-1})}} \quad (2-11)$$

和

$$c_j = \frac{\sum_{i=1}^n (x_i \mu_{ji}^k)}{\sum_{i=1}^n \mu_{ji}^k} = \sum_{i=1}^n \frac{\mu_{ji}^k}{\sum_{i=1}^n \mu_{ji}^k} x_i \quad (2-12)$$

依据式 2-11 和 2-12 中的求导结果对 FCM 方法聚类中心迭代优化，直到误差小于阈值或到达最大迭代次数，迭代停止。

迭代停止时，样本  $x_i$  对所有聚类中心  $[c_1, c_2, \dots, c_j, \dots, c_m]$  的隶属度矩阵向量为  $U_{\cdot i} = [\mu_{1i}, \mu_{2i}, \dots, \mu_{ji}, \dots, \mu_{mi}]$ ，样本  $x_i$  最终所属类别依据隶属度最大化原则划分为类别  $c_K$ ，则有：

$$K = \arg \max_j U_{\cdot i} = \arg \max_j (\mu_{1i}, \mu_{2i}, \dots, \mu_{ji}, \dots, \mu_{mi}) \quad (2-13)$$

依据式 2-13 对每个样本找到其隶属度最大的类别，即可完成整个数据集的模糊聚类划分。

## 2.2 基于神经网络的影像分割相关理论

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<sup>[57]</sup>。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。

### 2.2.1 卷积神经网络基础

CNN 是一种具有卷积结构的深度前馈神经网络，使用反向传播（Backpropagation, BP）算法进行训练，主要用来处理图像信息。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、下采样等特性。全连接层起到图像“分类器”的作用，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。因为图像是二维结构数据，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：

$$Y = X \otimes W \quad (2-14)$$

式中， $\otimes$  是卷积运算。输出特征图上某点  $(i, j)$  的值可由式2-15计算得到：

$$y_{ij} = \sum_{u=1}^m \sum_{v=1}^n w_{uv} \cdot x_{i-u+1, j-v+1} \quad (2-15)$$

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)。同一特征图内神经元共享权重和偏置项，且每一个神经元与上一层的区域局部连接。CNN 中共享的权重称为卷积核 (Convolutional kernel)，利用卷积核对上一层的特征图进行卷积运算可以提取影像特征产生下一层网络的输出层。如图 3 所示，图中使用大小为  $3 \times 3$  的卷积核，输入图像大小为  $5 \times 5$ ，卷积核从左上方以 1 个像素点的步长 (Stride) 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2-16 计算得到：

$$X_l = \lfloor \frac{X_{l-1} - K + 2 \times P}{S} \rfloor + 1 \quad (2-16)$$

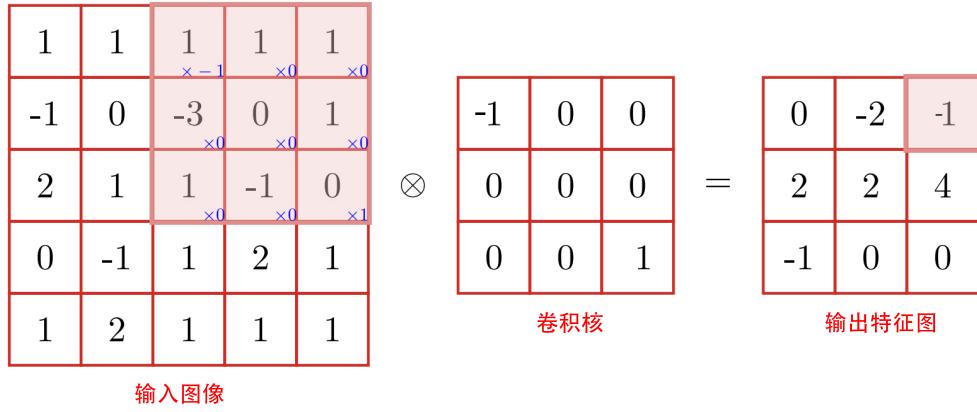


图 3 图像卷积示意图

卷积操作是线性的，线性模型的特点为任意线性模型的组合仍然是线性模型。线性神经网络无法解决实际复杂的非线性问题。为每个卷积输出添加一个非线性函数，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 函数的表达式如下：

$$f(x) = \max(x, 0) \quad (2-17)$$

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。卷积层虽然可以显著减少网络中连接参数的数量，但特征映射组中的神经元个数并没有显著减少。在卷积层后面加上一个池化层，可以降低下一层待处理的数据量，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，常用的池化函数有以下两种：

1. 最大池化（Maximum pooling）最大池化是选取区域内所有神经元的最大值作为池化输出。

$$Y_{m,n} = \max_{i \in R_{m,n}} x_i \quad (2-18)$$

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。

$$Y_{m,n} = \frac{1}{|R_{m,n}|} \sum_{i \in R_{m,n}} x_i \quad (2-19)$$

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  $Y = \{Y_{m,n}\}, 1 \leq m \leq M, 1 \leq n \leq N$ 。

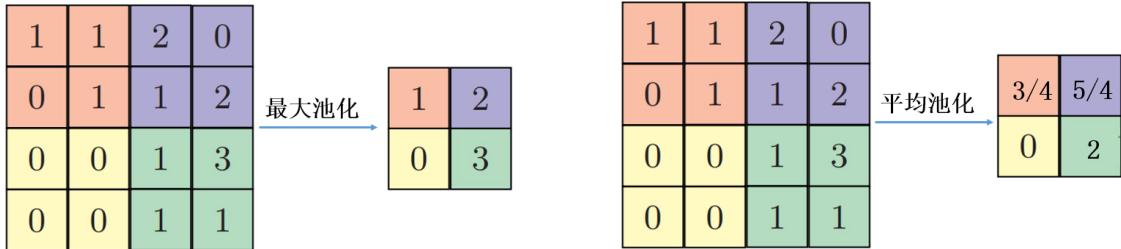


图 4 最大池化与平均池化

图如 4 所示, 输入特征图大小为  $4 \times 4$ , 池化函数核为  $2 \times 2$ , 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 5 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对不同形态的图像具有优秀的适应能力, 它可以拟合高分影像中因地物尺度不一、拍摄角度不同等原因形成的复杂特征, 从而有效提高分影像的分类识别精度。

#### 2.2.2 基于全卷积网络的图像分割

传统的卷积神经网络最右端结构通常连接全连接层, 它会将原来二维矩阵压缩成一维的, 从而丢失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

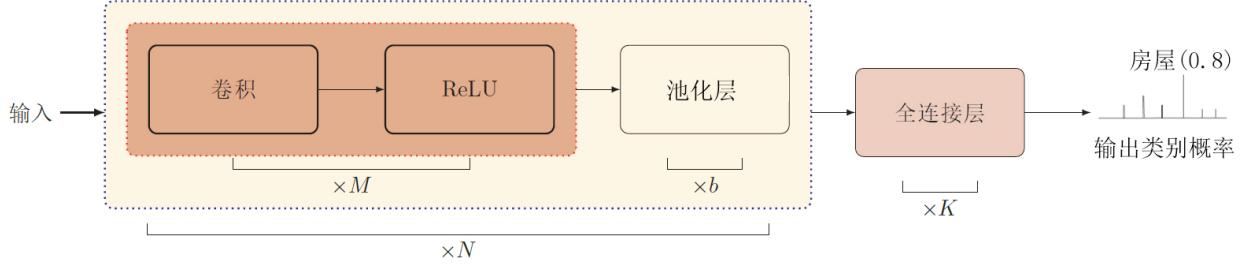


图 5 典型的卷积网络结构

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语义分类，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）<sup>[10]</sup>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分割图，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，实现图像语义分割像素级的预测。图像语义分割应用到遥感领域即为遥感影像的分类。图 6 为遥感影像分割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。

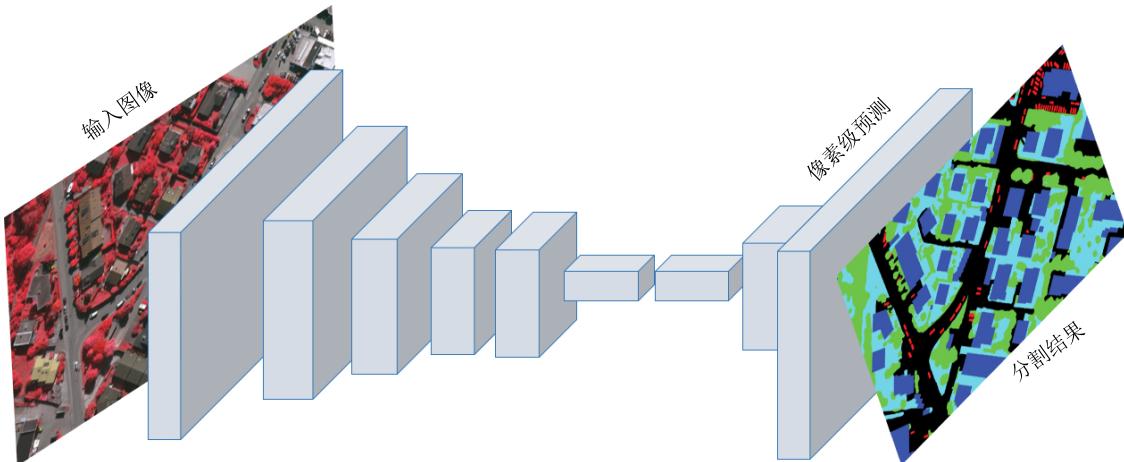


图 6 影像分割网络结构示意图

## 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  $4 \times 4$  的二维特征，用大小  $3 \times 3$  的核，做步长为 1 的卷积运算得到  $2 \times 2$  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  $2 \times 2$  的二维特征，使用

$3 \times 3$  的核操作得到  $4 \times 4$  的输出。如图 7 所示，输入特征大小  $2 \times 2$ ，核大小  $3 \times 3$ ，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  $4 \times 4$ ，图中显示了反卷积上采样的计算过程。

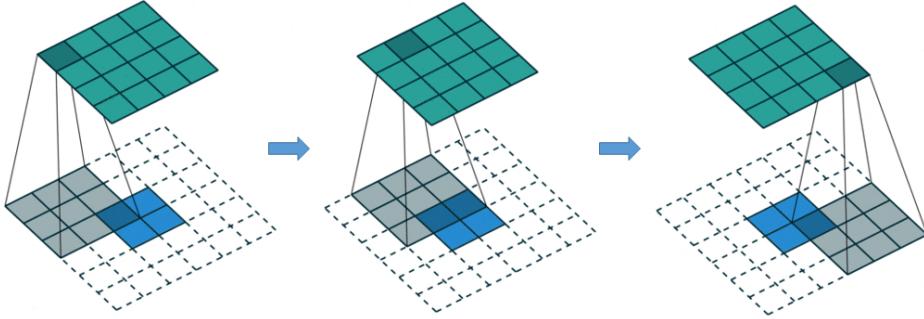


图 7 反卷积上采样过程

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：

$$O = S \times (X - 1) + K \quad (2-20)$$

## 2. FCN 网络图像分割

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 ImgaNet 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<sup>[57]</sup> 或 VGG<sup>[39]</sup> 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割结果，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 8 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的特征图进行上采样，使它恢复到输入图像的相同尺寸，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。

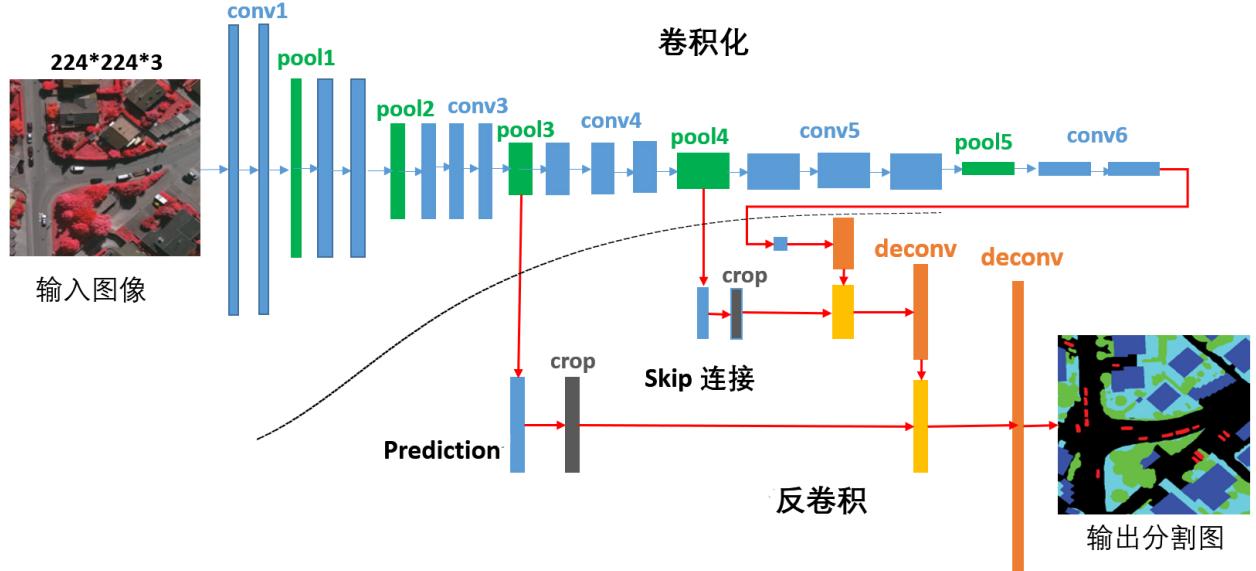


图 8 基于 FCN 网络的遥感影像分割示意图

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，通过层层反推，CNN 网络的输入图像尺寸大小也必须固定，所以去除全连接层的 FCN 网络可以接受任意尺寸的图像。另外一方面卷积化将全连接层输出的一维向量转换为二维矩阵，有利于后续反卷积上采样对图像进行像素级预测。反卷积上采样操作能将特征图恢复为原始图像大小，有利于保留图像空间位置信息，实现图像像素级预测。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割结果。

### 2.3 本章小结

本章主要介绍遥感影像分类相关的理论研究基础。首先从刻画影像不确定性角度介绍了面向对象遥感影像的模糊聚类分割相关理论。另外介绍了神经网络中 CNN 网络的结构和原理，并介绍了 FCN 网络模型的结构和特点，最后介绍了基于 FCN 的遥感影像分割方法思想。

# 第3章 基于区间二型模糊聚类的遥感影像无监督分割方法

## 3.1 引言

由于遥感影像数据具有同物异谱、同谱异物等固有的不确定性，模糊 C 均值聚类 (Fuzzy c-means clustering, FCM) 分割方法是利用模糊集理论数据建模，因此被广泛应用到遥感影像解译中<sup>[6]</sup>。随着遥感影像空间分辨率的提高，高分影像数据呈现信息多样性和复杂性，遥感影像聚类方法由基于像元发展为面向对象的聚类分割方法。本章内容从遥感影像特征信息表达和目标地物类别关系两个角度来表征遥感影像分类中的不确定性信息。首先设计了三角形模糊集值信息表达模型来表征影像分割单元信息，其次提出一种新的区间值度量方法计算两个三角形模糊集值数据的相异性，最后，改进了已有的二型模糊集合聚类分割方法来对影像数据建模，以刻画遥感影像数据的不确定性。

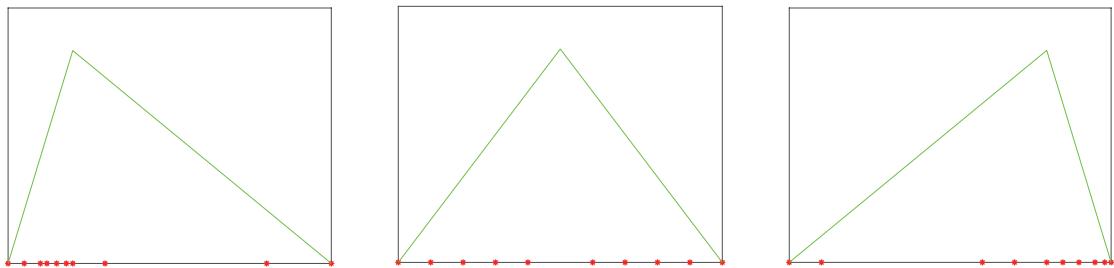


图 9 区间值相同但分布不同的数据比较

## 3.2 三角形模糊集值建模与相似性度量

当前，面向对象分割方法中对影像单元多采取均值数据建模<sup>[7]</sup> 和区间值数据建模<sup>[8]</sup>。然而，这两种信息表达模型无法区分具有相同均值和区间值但内部分布不一致的分割单元。如图 9 所示，每组模拟数据内的点集可以看作一个影像分割单元像素点集合，具有相同的均值和区间值的影像单元内像素点的分布差异明显。

模糊集 (Fuzzy sets, FS) 是 Zadeh 教授 1965 年提出的概念，通过建立适当的隶属度函数 (Membership function, MF) 来描述对象的不确定性<sup>[58]</sup>。常见的隶属度函数有：三角形 MF，梯形 MF，截断高斯 MF 和钟形 MF 等。FS 最常用和最基本的 MF 是三角形 MF。因此，文中利用三角形模糊集来定义三角形模糊集值数据模型。

定义 3.1：三角形模糊集值（Triangular Fuzzy Set Valued, TFSV）模型的定义

三角形模糊集值数据由以下三个关键参数组成： $(a^-, 0)$ ,  $(a^m, 1)$  和  $(a^+, 0)$ 。如图 10 所示，几何上， $(a^-, 0)$  和  $(a^+, 0)$  组成三角形 MF 的下边缘；代数上， $(a^-, 0)$  和  $(a^+, 0)$  形成一个区间值，确保一定的变化范围。 $(a^m, 1)$  是 TFSV 数据的顶点，表征 FS 的最高置信度。

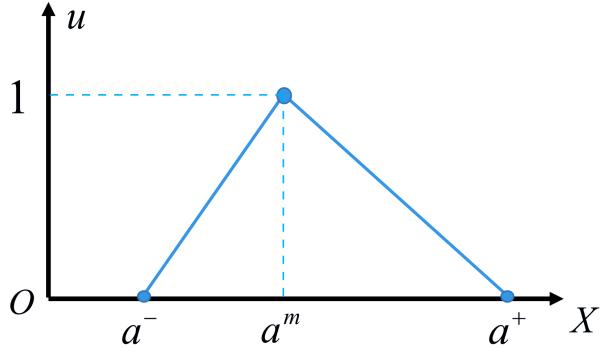


图 10 三角形模糊集  $\tilde{A}$  示意图

聚类是根据某种相似性或距离将相似的对象聚为一类的无监督学习方法，相似性度量是聚类的核心要素。对于两个 TFSV 数据  $\tilde{A}$  和  $\tilde{B}$ ，常见的相似性度量方法有以下几种：

(1) 欧式距离 (Euclidean distance)

$$d_E(\tilde{A}, \tilde{B}) = \sqrt{\sum_{x \in X} |\tilde{A}(x) - \tilde{B}(x)|^2}, x \in X \quad (3-1)$$

$\tilde{A}$  和  $\tilde{B}$  的欧式距离  $d_E(\tilde{A}, \tilde{B})$  被看作是集合对应元素差值平方和的平方根。

(2) 城市距离 (City-block distance)

$$d_C(\tilde{A}, \tilde{B}) = \sum_{x \in X} |\tilde{A}(x) - \tilde{B}(x)|, x \in X \quad (3-2)$$

$\tilde{A}$  和  $\tilde{B}$  的城市距离  $d_C(\tilde{A}, \tilde{B})$  被看作是集合对应元素差值绝对值的和。

(3) 豪斯多夫距离 (Hausdorff distance)

豪斯多夫距离最开始为区间或普通集合设计，两个普通集合  $A$  与  $B$  的豪斯多夫距离为：

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} |a - b|, \sup_{b \in B} \inf_{a \in A} |a - b| \right\} \quad (3-3)$$

将其推广到模糊集，可以考虑模糊集  $\tilde{A}$  和  $\tilde{B}$  的一个  $\alpha - cut$  截集<sup>[58]</sup>  $d_H^\alpha(\tilde{A}, \tilde{B})$ ，则有：

$$d_H^\alpha(\tilde{A}, \tilde{B}) = \max \left\{ \sup_{a \in \tilde{A}_\alpha} \inf_{b \in \tilde{B}_\alpha} |a - b|, \sup_{b \in \tilde{B}_\alpha} \inf_{a \in \tilde{A}_\alpha} |a - b| \right\} \quad (3-4)$$

其中， $\inf$  和  $\sup$  分别表示取集合的最大下界和最小上界。模糊集是度量空间的非空紧致和有限子集，所以等式 3-4 中的  $\inf$  和  $\sup$  操作可分别替换为  $\min$  和  $\max$  操作，即

$$d_H^\alpha(\tilde{A}, \tilde{B}) = \max \left\{ \max_{a \in \tilde{A}_\alpha} \min_{b \in \tilde{B}_\alpha} |a - b|, \max_{b \in \tilde{B}_\alpha} \min_{a \in \tilde{A}_\alpha} |a - b| \right\} \quad (3-5)$$

然后对  $\tilde{A}$  和  $\tilde{B}$  所有可能的  $\alpha - cut$  截集积分，就可得到模糊集  $\tilde{A}$  和  $\tilde{B}$  的豪斯多夫距离：

$$d_H(\tilde{A}, \tilde{B}) = \int_0^1 d_H^\alpha(\tilde{A}, \tilde{B}) d\alpha = \int_0^1 \max \left\{ \max_{a \in \tilde{A}_\alpha} \min_{b \in \tilde{B}_\alpha} |a - b|, \max_{b \in \tilde{B}_\alpha} \min_{a \in \tilde{A}_\alpha} |a - b| \right\} d\alpha \quad (3-6)$$

那么，如何选择合适的距离来度量两个模糊集的相似性呢？对于模糊集  $\tilde{A}$  和  $\tilde{B}$ ，如图 11 所示，我们考虑在同一坐标系内， $\tilde{A}$  和  $\tilde{B}$  只存在三种位置关系：相交、包含和不相交。

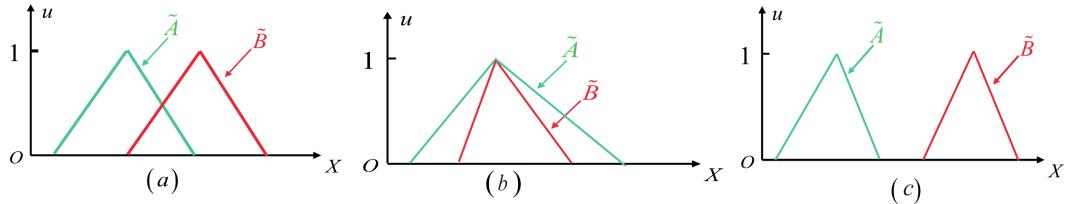
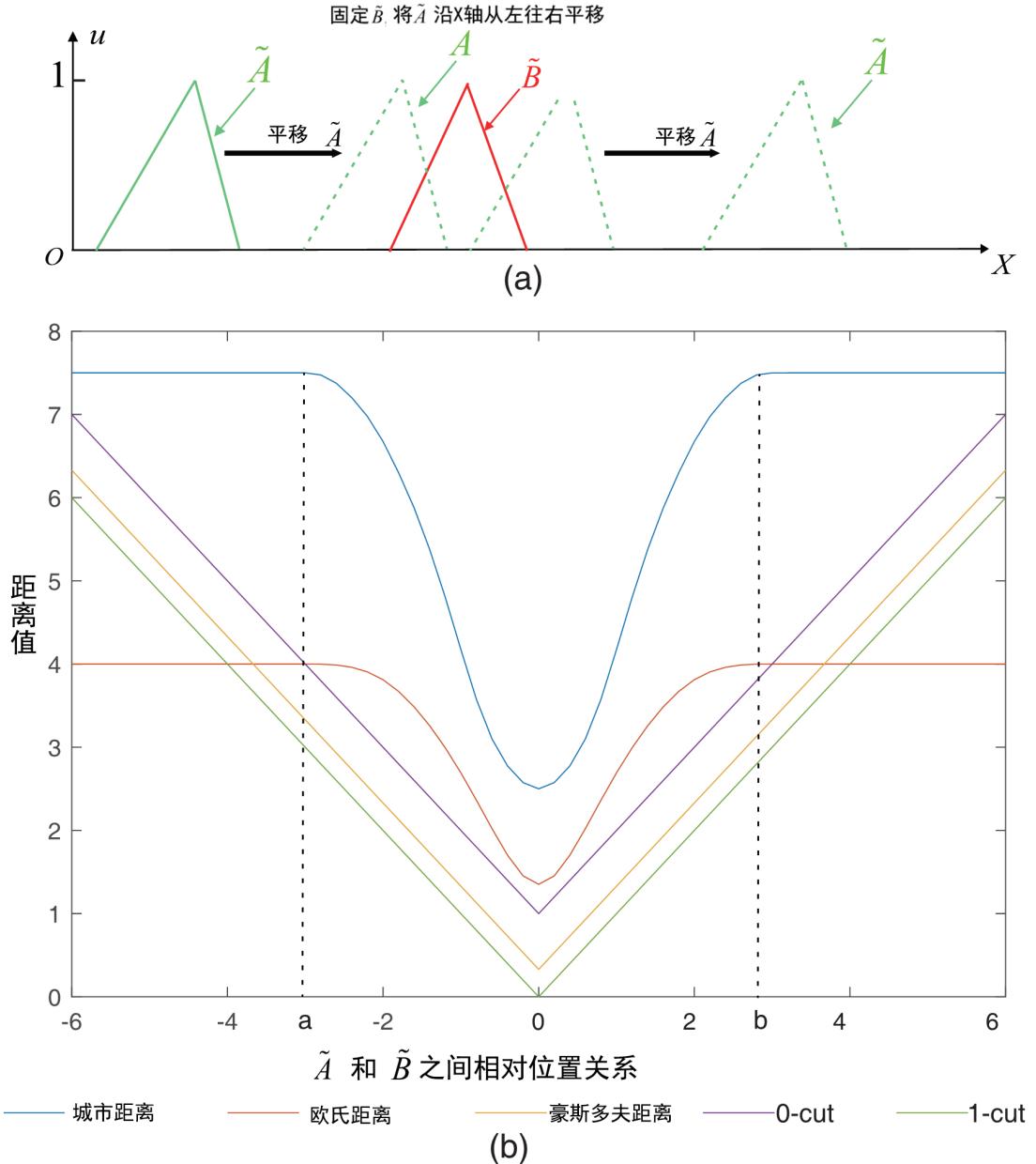


图 11  $\tilde{A}$  和  $\tilde{B}$  位置关系示意图。(a) 相交; (b)  $\tilde{A}$  包含集合  $\tilde{B}$ ; (c) 不相交.

为了比较各种位置关系下两个模糊集间上述各种距离的大小关系。文中设计以下实验：如图 12(a) 所示， $\tilde{B}$  固定不动，将  $\tilde{A}$  沿 X 轴从左向右移动，分别计算各相对位置下  $\tilde{A}$  和  $\tilde{B}$  对应位置的距离。结果如图 12(b) 所示（图中  $0 - cut$  和  $1 - cut$  截集的豪斯多夫距离后面会讨论）， $\tilde{A}$  位于  $(a, b)$  区间内时， $\tilde{A}$  和  $\tilde{B}$  相交，以上四种距离均可度量  $\tilde{A}$  和  $\tilde{B}$  的相似性。然而，在两者不相交时（图中  $\tilde{A}$  位于  $(-\infty, a)$  和  $(b, +\infty)$  区间内）无论  $\tilde{A}$  和  $\tilde{B}$  相距多远，欧式距离和城市距离均为一固定常量，只有豪斯多夫距离可以精确度量上述三种位置下两个模糊集的相似性。

然而，计算两个模糊集之间的豪斯多夫距离具有较高的计算复杂度，导致其在模糊聚类迭代更新中难以适用。另外，Zadeh 教授在可能性分布理论<sup>[59]</sup> 中提到：模糊集作为一个弹性约束值，若使用区间距离度量，而不是一个单一固定值来度量两个模糊集间相似性可

图 12 不同位置下， $\tilde{A}$  和  $\tilde{B}$  两者间各种距离比较图。

以获取产生更高的识别能力。因此，文中尝试引入一种新的基于豪斯多夫距离的区间距离度量方法来度量三角形模糊集的相似性。

首先，考虑模糊集合的特性，文献<sup>[59]</sup>中定义了模糊集的  $\alpha - cut$  截集并提出了模糊集的  $\alpha - cut$  截集表现定理，其中提到任何模糊集都可以由它指定的  $\alpha - cut$  截集来表示。在所有的截集中，有两个截集是最重要的与最具有代表性的，它们就是  $0 - cut$  和  $1 - cut$  截集，分别代表模糊集和的支撑集和置信集合。模糊集  $\tilde{A}$  的支持集  $0 - cut$  包含  $\tilde{A}$  中非零的所有

元素，体现了模糊集的可能性；而  $\tilde{A}$  的置信集  $1 - cut$  包含元素的隶属度均为 1，这表明  $1 - cut$  截集中所有元素都有最高的隶属度和置信度，体现了模糊集的必要性。此外，最近一些关于  $\alpha - cut$  截集的研究对任一模糊集，仅使用  $0 - cut$  和  $1 - cut$  截集就可近似拟合模糊集的质心；此外，任意其他  $\alpha - cut$  截集 ( $0 \leq \alpha \leq 1$ ) 都可以表示为  $0 - cut$  和  $1 - cut$  截集的广义线性组合<sup>[60]</sup>。

根据公式3-5，我们分别可以获得  $\tilde{A}$  和  $\tilde{B}$  的  $0 - cut$  和  $1 - cut$  的豪斯多夫距离为  $d_H^0(\tilde{A}, \tilde{B})$  和  $d_H^1(\tilde{A}, \tilde{B})$ 。基于上面提到的可能性分布理论和模糊集的  $\alpha - cut$  截集表现定理，我们可以定义两个三角形模糊集间  $\tilde{A}$  和  $\tilde{B}$  间一种新的区间值距离  $d_I(\tilde{A}, \tilde{B})$  为下式：

$$d_I(\tilde{A}, \tilde{B}) = [\min\{d_0(\tilde{A}, \tilde{B}), d_1(\tilde{A}, \tilde{B})\}, \max\{d_0(\tilde{A}, \tilde{B}), d_1(\tilde{A}, \tilde{B})\}] \quad (3-7)$$

图 12 中的结果 ( $0 - cut$  和  $1 - cut$ ) 也表明  $d_I(\tilde{A}, \tilde{B})$  能够度量两个模糊集不相交时的相似性；另外，还可以看出文中新定义的距离  $d_I(\tilde{A}, \tilde{B})$  是  $d_H(\tilde{A}, \tilde{B})$  的弹性膨胀，即有下式3-8<sup>①</sup>：

$$d_H(\tilde{A}, \tilde{B}) \in d_I(\tilde{A}, \tilde{B}) \quad (3-8)$$

一个相似性度量能够定位为距离，当且仅当其能满足距离度量三个条件，即非负性、对称性和三角不等式。假定  $\tilde{A}$ 、 $\tilde{B}$  和  $\tilde{C}$  是任意三个三角形模糊集，则需要满足以下特性<sup>②</sup>：

- (1) 非负性： $d_I(\tilde{A}, \tilde{A}) = 0$
- (2) 对称性： $d_I(\tilde{A}, \tilde{B}) = d_I(\tilde{B}, \tilde{A})$
- (3) 三角不等式： $d_I(\tilde{A}, \tilde{C}) \leq d_I(\tilde{A}, \tilde{B}) + d_I(\tilde{B}, \tilde{C})$

综上所述，本节针对面向对象分割单元定义了 TFSV 数据模型，同时，结合模糊集与可能性分布定理特性，针对新提出的 TFSV 数据类型，本文新提出了一种区间值的距离度量方法，来度量两个 TFSV 数据间的相似性。

### 3.3 面向对象的改进型区间二型模糊遥感影像聚类方法

#### 3.3.1 算法整体框架

本章提出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) 主要用于提高高分影像聚类分割精度。图 13 展示了该面向对象分类方法的总体处理流程，具体可分为以下几步：

① 公式 3-8 的证明步骤由于篇幅过大，可参考本文作者已发表文章<sup>[61]</sup> 的附录一。

② 证明过程较复杂且非文章研究重点，可参考本文作者已发表文章<sup>[61]</sup> 的附录二。

### Step 1: 影像分割与分割单元的 TFSV 数据建模

高分影像被分割为具有同质性的像素单元集合。对分割单元提取特征，并构建 TFSV 数据模型；

### Step 2: 模糊聚类分析

使用 TFSV-IT2FCM 算法对高分影像分割单元的 TFSV 模型数据聚类，其中 IT2FCM 算法的距离度量使用式3-7中新提出的区间距离  $d_I$ ；

### Step 3: 聚类结果的后处理

使用类别组合方法对聚类分割的结果处理，得到高分影像最终得分割结果。

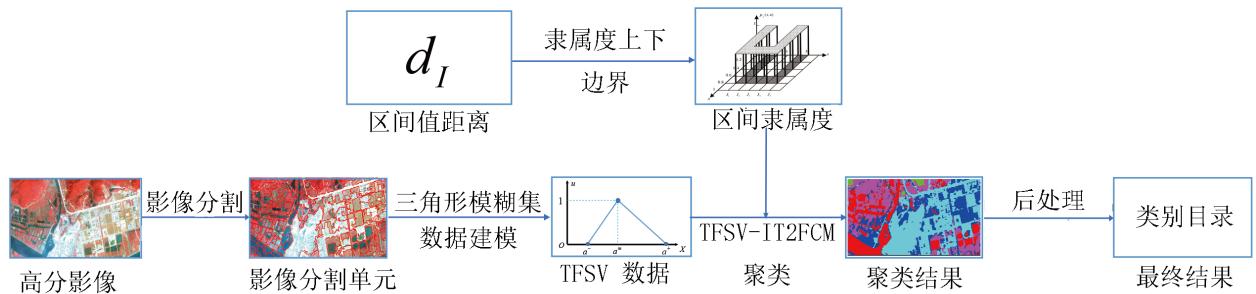


图 13 面向对象的改进型区间二型模糊遥感影像聚类算法框架图

### 3.3.2 面向对象的 TFSV 模型

#### 1. 影像分割

目前，遥感影像常用的像元分割算法包含基于分水岭的算法，基于直方图的方法和聚类方法等<sup>[62]</sup>。简单线性迭代聚类（SLIC）算法因具有高计算效率和可选数量的分段单元<sup>[29]</sup>，文中使用 SLIC 超像素分割算法提取遥感影像的分割单元。SLIC 算法的关键步骤如下：

- (I) 计算遥感影像梯度获取影像梯度图；
- (II) 初始化梯度图中的聚类中心；
- (III) 将遥感影像从 SPOT5 格式转换为 CIELAB 颜色空间计算像元间的相似度；
- (IV) 使用 SLIC 算法对 CIELAB 彩色图像进行分割，以获得影像同质性分割单元。

对高分影像  $\mathbf{I}$ ，使用 SLIC 分割算法获得影像分割单元  $\mathbf{SS}$ ，为：

$$\mathbf{SS} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n\} \quad (3-9)$$

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。

## 2. 影像单元的 TFSV 数据建模

尽管文献<sup>[8]</sup> 中已经证明: 对于影像分割单元, 区间值的特征比均值特征更有效, 但是影像分割单元的不确定性不能被充分的表达。因此, 文中使用定义 3.1 的 TFSV 数据模型  $\tilde{\mathbf{A}}_i$  来提取影像单元  $\mathbf{B}_i$  的特征:

- 1) 基于影像分割单元  $\mathbf{B}_i$  内像素的均值和方差特性, 可以获得一个  $p$  维的区间值向量, 记为  $\mathbf{X}_i$ , 如下:

$$\mathbf{X}_i = [\mathbf{X}_i^{\text{down}}, \mathbf{X}_i^{\text{up}}] = [\max\{0, \boldsymbol{\mu}_i - \alpha \times \boldsymbol{\sigma}_i\}, \boldsymbol{\mu}_i + \alpha \times \boldsymbol{\sigma}_i] \quad (3-10)$$

其中  $\boldsymbol{\mu}_i = [\mu_1, \mu_2, \dots, \mu_p]^T$ ,  $\boldsymbol{\sigma}_i = [\sigma_1, \sigma_2, \dots, \sigma_p]^T$  是第  $i$  个分割单元  $\mathbf{B}_i$  的均值和方差,  $\alpha$  是控制区间值大小的超参数,  $p$  是遥感影像的波段数。

- 2) 分割单元是邻近同质像素点的集合, 基于统计学特性, 等式3-10 中的向量  $\mathbf{X}_i = [\mathbf{X}_i^{\text{down}}, \mathbf{X}_i^{\text{up}}]$  除了少数异常点外包含分割单元  $\mathbf{B}_i$  中的绝大部分点。因此, 从  $\mathbf{X}_i$  中导出参数  $(\mathbf{X}_i^{\text{down}}, 0)$  和  $(\mathbf{X}_i^{\text{up}}, 0)$  来构造  $\tilde{\mathbf{A}}_i$  的底边。
- 3)  $\tilde{\mathbf{A}}_i$  的顶点由  $(\mathbf{med}_i, 1)$  组成。因中值不受极值的影响, 并且对噪声点和异常值具有很高的鲁棒性, 这里  $\mathbf{med}_i$  取  $\mathbf{B}_i$  的中值。

类似地, 影像的每个分割单元都可以被表征为一个 TFSV 数据模型, 分割单元的集合  $\mathbf{SS} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n\}$  可以被表示为:

$$\mathbf{SS} \rightarrow \{\tilde{\mathbf{A}}_1, \tilde{\mathbf{A}}_2, \dots, \tilde{\mathbf{A}}_n\} \quad (3-11)$$

其中  $\mathbf{SS}$  是一个  $n \times p$  矩阵,  $\tilde{\mathbf{A}}_i$  是  $\mathbf{B}_i$  对应的  $p$  维 TFSV 数据,  $n$  表示分割单元的数目。

### 3.3.3 基于 TFSV 数据模型的面向对象模糊聚类方法

#### 1. 区间二型模糊隶属度

区间二型模糊聚类算法 (Interval type 2 fuzzy clustering method, IT2FCM) 使用一个区间值来表示隶属度值。文献<sup>[63]</sup> 中使用两个模糊化指数  $m_1$  和  $m_2$  得到区间二型隶属度函数 (Interval type 2 membership function, IT2MF) 的上界和下界, 从而将一型模糊集 (Type 1 fuzzy sets, T1FS) 扩展为二型模糊集 (Type 2 fuzzy sets, T2FS)。然而, IT2FCM 算法对模糊

指数  $m_1$  和  $m_2$  的取值敏感。与传统 FCM 算法相比，不恰当的  $m_1$  和  $m_2$  取值会导致更差的实验结果。因此，基于公式 3-7 中新定义的区间值距离度量，类似 FCM 算法，本文仅使用一个模糊指数  $m$  来表达 IT2MF。

假定样本  $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p)^T$  和聚类中心  $\tilde{\mathbf{V}} = (\tilde{V}_1, \tilde{V}_2, \dots, \tilde{V}_p)^T$  是两个形如公式 3-11 的  $p$  维 TFSV 数据，其中  $\tilde{X}_i$  ( $1 \leq i \leq p$ ) 是一个一维 TFSV，由  $(X_i^{down}, 0)$ ,  $(X_i^{up}, 0)$  和  $(X_i^{med}, 1)$  这三个参数构成；类似地， $\tilde{V}_i$  由  $(V_i^{down}, 0)$ ,  $(V_i^{up}, 0)$  和  $(V_i^{med}, 1)$  三个参数组成。 $\tilde{X}_i$  和  $\tilde{V}_i$  的  $0-cut$  和  $1-cut$  豪斯多夫距离分别为：

$$d_0(\tilde{X}_i, \tilde{Y}_i) = \max \left\{ |X_i^{down} - Y_i^{down}|, |X_i^{up} - Y_i^{up}| \right\} \quad (3-12)$$

和

$$d_1(\tilde{X}_i, \tilde{V}_i) = \left\{ |X_i^{med} - V_i^{med}| \right\} \quad (3-13)$$

$\tilde{X}_i$  和  $\tilde{V}_i$  的区间值距离  $d_I(\tilde{X}_i, \tilde{V}_i)$  即为：

$$d_I(\tilde{X}_i, \tilde{V}_i) = [\min\{d_0(\tilde{X}_i, \tilde{V}_i), d_1(\tilde{X}_i, \tilde{V}_i)\}, \max\{d_0(\tilde{X}_i, \tilde{V}_i), d_1(\tilde{X}_i, \tilde{V}_i)\}] \quad (3-14)$$

从而， $\tilde{\mathbf{X}}$  和  $\tilde{\mathbf{V}}$  间的区间值距离  $d_I(\tilde{\mathbf{X}}, \tilde{\mathbf{V}})$  可以被表示为：

$$d_I(\tilde{\mathbf{X}}, \tilde{\mathbf{V}}) = \max\{d_I(\tilde{X}_i, \tilde{V}_i)\}, i = 1, 2, \dots, p \quad (3-15)$$

与 FCM 算法类似，新提出的面向对象的 TFSV-IT2FCM 算法求解需要最小化以下目标函数：

$$J(\mathbf{U}; \mathbf{V}) = \sum_{j=1}^n \sum_{i=1}^K (U_{ij})^m d^2(X_i, V_j) \quad (3-16)$$

其中  $m(m > 1)$  是模糊指数， $d^2(X_i, V_j) = d_I(\tilde{X}_i, \tilde{V}_j)$  是依据公式 3-7 定义的样本  $X_i$  和  $V_j$  间的区间值距离。为了最小化目标函数  $J$ ，有：

$$U_{ij} = \frac{1}{\sum_{k=1}^K \left( \frac{d_{ji}}{d_{ki}} \right)^{\frac{2}{m-1}}} \quad (3-17)$$

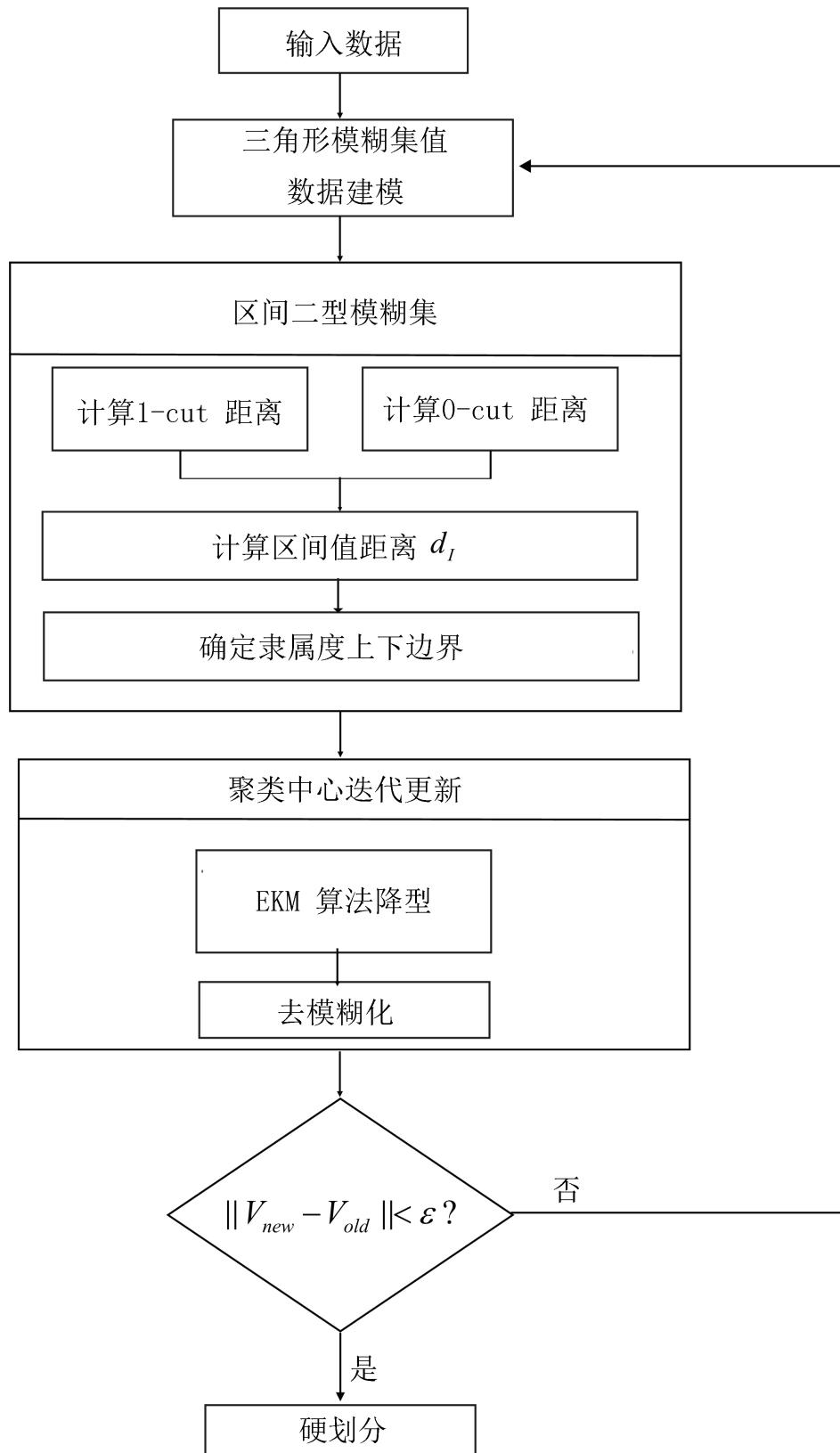


图 14 TFSV-IT2FCM 算法流程图

在 TFSV-IT2FCM 算法中，模糊指数  $m$  的上界和下界以及区间值  $d^2(X_i, V_j)$  用来描述不确定性。区间隶属度的上界  $\bar{u}_{ij}$  和下界  $\underline{u}_{ij}$  分别为：

$$\bar{u}_{ij} = \max \left\{ \frac{1}{\sum_{k=1}^K \left( \frac{d_{ji}^0}{d_{ki}^0} \right)^{\frac{2}{m-1}}}, \frac{1}{\sum_{k=1}^K \left( \frac{d_{ji}^1}{d_{ki}^1} \right)^{\frac{2}{m-1}}} \right\} \quad (3-18)$$

和

$$\underline{u}_{ij} = \min \left\{ \frac{1}{\sum_{k=1}^K \left( \frac{d_{ji}^0}{d_{ki}^0} \right)^{\frac{2}{m-1}}}, \frac{1}{\sum_{k=1}^K \left( \frac{d_{ji}^1}{d_{ki}^1} \right)^{\frac{2}{m-1}}} \right\} \quad (3-19)$$

其中  $d_{ji}^0$  和  $d_{ji}^1$  分别是  $\tilde{X}_i$  和  $\tilde{V}_j$  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。

## 2.TFSV-IT2FCM 算法

与 FCM 算法不同，文中提出的 TFSV-IT2FCM 算法核心是一个 IT2FS，无法直接通过降型去模糊化，需要先计算 IT2FS 的质心将 IT2FS 转换为 T1FS，再去模糊化得到明确集<sup>[64]</sup>。文中使用 EKM 降型算法<sup>[65]</sup> 降型和去模糊化，从而获取精确的聚类中心。降型后的聚类中心点可表示为

$$V_j = [V_j^L, V_j^R] \quad (3-20)$$

通过去模糊化后得到的明确集聚类中心为：

$$V_j = \frac{V_j^L + V_j^R}{2} \quad (3-21)$$

整个 TFSV-IT2FCM 算法框架如图 14 所示，算法的详细描述如下：

### (1) 初始化

初始化实验聚类参数个数  $K(2 \leq K \leq N)$ ，在实验 1、2、3 中分别设置  $K = 5$ 、5 和 6。初始化模糊指数  $m = 2.0 (1 < m < \infty)$ ，迭代阈值  $\varepsilon = 0.0001 (\varepsilon > 0)$ ，最大迭代次数  $T = 500$ ，初始  $t = 1$ ，初始化区间膨胀参数  $\alpha = 0.8$ 。其中超参数  $m$  和  $\alpha$  依赖具体的问题和实验数据。在本实验中，当  $m = 2.0$  和  $\alpha = 0.8$  能取得最优实验效果。

## (2) TFSV 数据建模

使用 SLIC 算法获取影像的分割单元  $\mathbf{SS} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n\}$ ，然后转化为 TFSV 类型数据  $\{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_n\}$  作为训练样本。初始化 TFSV 的聚类中心  $\tilde{\mathbf{V}} = \{\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2, \dots, \tilde{\mathbf{V}}_K\}$  为  $\tilde{\mathbf{V}}^0$ ，其中  $\tilde{\mathbf{V}}_k^0, (1 \leq k \leq K)$  由参数  $(v_k^l, 0)$ 、 $(v_k^r, 0)$  和  $(v_k^{mid}, 1)$  构成。

## (3) 聚类迭代更新

在 TFSV-IT2FCM 算法迭代中得到模糊划分矩阵  $\mathbf{U} = [U_{ij}]_{n \times K}$ ，其中  $U_{ij} = [\underline{u}_{ij}, \bar{u}_{ij}]$  如公式3-18 和3-19 所示。聚类中心  $[\tilde{v}_l, \tilde{v}_r]$  由 EKM 算法降型获取。然后，将降型后的聚类中心去模糊化，得到明确集的聚类中心为

$$\tilde{v} = \frac{\tilde{v}_l + \tilde{v}_r}{2} \quad (3-22)$$

## (4) 迭代中止

如果迭代过程满足  $\|\tilde{\mathbf{V}}^t - \tilde{\mathbf{V}}^{t+1}\| \leq \varepsilon$  或  $t \geq T$  即停止迭代；否则，令  $t = t + 1$  并跳到步骤(3)。距离  $\|\tilde{\mathbf{V}}^t - \tilde{\mathbf{V}}^{t+1}\|$  可由  $d_I(\tilde{\mathbf{V}}^t, \tilde{\mathbf{V}}^{t+1})$  取均值得到，即

$$\|\tilde{\mathbf{V}}^t - \tilde{\mathbf{V}}^{t+1}\| = \frac{d_0(\tilde{\mathbf{V}}^t, \tilde{\mathbf{V}}^{t+1}) + d_1(\tilde{\mathbf{V}}^t, \tilde{\mathbf{V}}^{t+1})}{2} \quad (3-23)$$

## (5) 硬划分

区间二型的模糊划分矩阵  $\mathbf{U} = [U_{ij}]_{n \times K}$  可以降型为

$$U_{ij} = \frac{\underline{u}_{ij} + \bar{u}_{ij}}{2} \quad (3-24)$$

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

### 3.3.4 后处理

经过像元分割与非监督模糊聚类，可以得到遥感影像的初始分类结果。使用 CORINE 地物覆被处理系统<sup>①</sup> 对初始聚类结果进行人工类别合并，实现基于高分影像识别的高级地物覆被分类<sup>[66]</sup>。

<sup>①</sup> CORINE (Coordination of Information on the Environment) 是欧洲委员会于 1985 年提出的环境信息协调系统，其包含 44 个类别的地物覆被，常用于地物覆被的后处理。官方网址：<https://land.copernicus.eu/pan-european/corine-land-cover>.

### 3.4 实验结果与分析

本章提出的面向对象的 TFSV-IT2FCM 遥感影像聚类分割方法是对文献<sup>[8]</sup>中引进的面向对象的区间值模糊 C-均值 (Interval valued fuzzy c-means, IV-FCM) 聚类方法的改进。因此，文中选择三个具有复杂土地覆盖的研究区，设计土地覆盖分类实验来验证新提出的 TFSV-IT2FCM 算法相比已有方法的性能效果，重点与 IV-FCM 算法比较。同时，分别通过可视化的聚类分割结果与量化数据来验证分类结果。三个实验研究区分别为中山市神湾地区、珠海市横琴岛和河南省三门峡地区。三个研究区详细地物覆盖介绍如表格1所示。

表 1 实验区土地覆盖类别表

研究区	地物类别	地物数据描述
神湾地区	林地	自然森林和植被等
	水域	河流
	草地	草场，长草的田地
	建筑	建筑物，居民区等
	养殖水域	水库，池塘等
横琴地区	水域	河流，池塘，水库等
	农业区	蔬菜种植区，农田等.
	林地	自然森林和植被等
	农场	果园，农场等
	建筑	建筑物，居民区，村庄等
三门峡地区	道路	公路，小径，柏油路，桥梁等。
	林地	天然林，人工林等
	水域	河流，池塘等.
	草地	杂草，长满草的农田等。
	农场	果园，农场等
	建筑	建筑物，居民区等

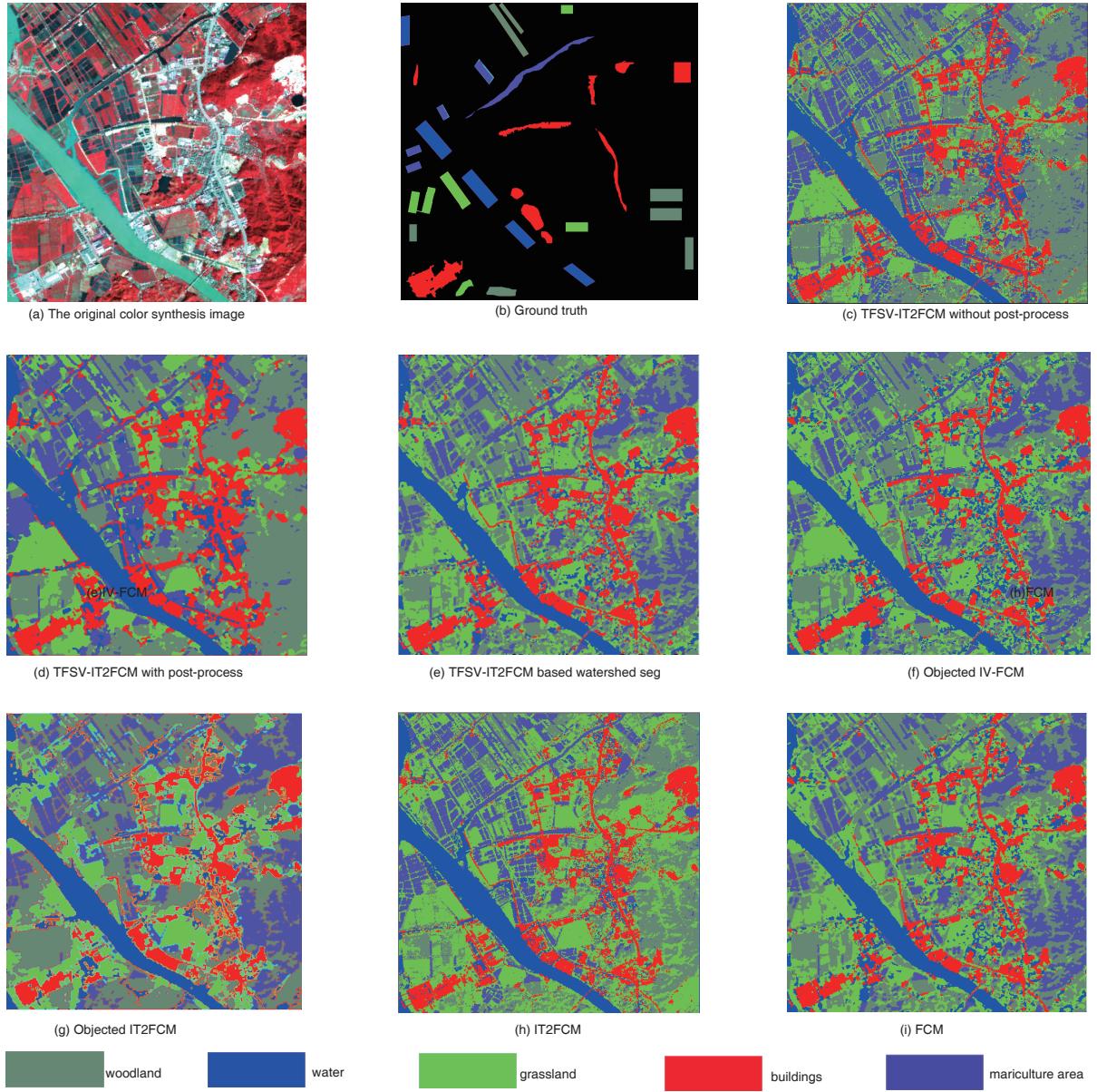


图 15 神湾研究区不同方法分类结果示意图

### 3.4.1 神湾地区实验

神湾研究区位于广东省中山市境内，是一个复杂的种植区域，该区域的高分影像由 SPOT 5 遥感卫星于 2008 年拍摄。SPOT 5 遥感卫星影像包含四个波段，空间分辨率大约 10 米，光谱范围 0.4 – 0.9 微米。神湾地区地物覆被可划分为 5 大类，分别是：水域、草地、建筑和养殖水域。实验中选择影像  $B1$ 、 $B2$ 、 $B3$  波段合成 RGB 图像，如图 15(a) 所示，其包含  $400 \times 400$  个像素。

在该研究区，文中提出的面向对象的 TFSV-IT2FCM 算法分别与其他聚类分割算法进行比较，包含基于像素的聚类方法 (FCM 算法，IT2FCM 算法) 和面向对象的聚类方法 (面向对象的 IT2FCM，面向对象的 IV-FCM 算法)，聚类结果分别由图15(f)-(i) 展示。另外，文中提出的 TFSV-IT2FCM 算法包含三个主要处理步骤：影像分割，聚类和后处理。图15(c) 和15(d) 分别展示是否使用后处理的分割结果；图15(e) 是获取影像分割单元时使用分水岭分割算法替换 SLIC 超像素分割算法得到。通过目视各方法对研究区的聚类分割结果，可以得到以下结论：

- (1) 图15(c) 和15(d) 无论是否使用后处理操作，两者均有大致相似的分割类别区域，图15(d) 使用后处理操作具有较少的奇异点。
- (2) 图15(e) 中使用分水岭算法获取分割单元，与图15(d) 相比，一些建筑类别的分割单元被错误划分为草地。因此，获取影像分割单元的分割算法在一定程度上会影像最后聚类分割的精度。
- (3) 比较图15(f)-(i) 和图15(d) 中的聚类计结果可知，这些聚类方法在水域的识别上都有不错的效果。然而，对于 IV-FCM 算法 (图15(f))，一些草地类别被错分为林地类别；对于面向对象的 IT2FCM 算法 (图15(g))，由于建筑、草地和林地这三类有相近的光谱波段，图中左下角区域无法明确区分三者；另外，与面向对象的方法相比，基于像素的 IT2FCM 算法和 FCM 算法 (分别为图15(g) 和 (h)) 在同类别大区域中产生许多异类别的异常小块。

综上所述，神湾研究区可视化的分类结果图表明文中提出的 TFSV-IT2FCM 聚类分割算法相比其他已有算法能获得更好的遥感影像分割结果。

为了得到实验中精确量化的聚类结果图，实验中使用 CORINE 土地覆被分类系统生成地物类别 ground truth 图，如图15(b) 所示。表 2 为神湾研究区各类别地物抽样点数据表。为了对分类结果进行评价，确定分类的精度和可靠性。本实验中使用以下评价指标度量分类效果：

### 1. 生产者精度 (Producer's accuracy, PA)

生产者精度是指分类器将整个影像像元正确分为 A 类的像元数（对角线值）与 A 类真实像元参考总数的比值。

### 2. 用户精度 (User's accuracy, UA)

用户精度是指正确分到 A 类的像元总数（对角线值）与分类器将整个影像的像元分为 A 类的像元总数（混淆矩阵中 A 类行的总和）比率。

### 3. 总体分类精度 (Overall accuracy, OA)

总体分类精度是指所有类别中正确分类的像元总和除以总像元数。

#### 4. Kappa 系数

Kappa 系数基于混淆矩阵，用于衡量分类精度。计算公式为

$$Kappa = \frac{\text{总体准确度} - \text{期望准确度}}{1 - \text{期望准确度}} = \frac{N \sum_{i=1}^n (X_{ii}) - \sum_{i=1}^n (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} \times X_{+i})} \quad (3-25)$$

式中， $n$  表示类别， $N$  表示总类别数， $X_{ii}$  表示混淆矩阵对角线元素， $X_{+i}$  表示类别的行总和， $X_{i+}$  表示类别的列总和。

表 2 神湾地区影像地物类别抽样 Ground truth 表 (pixels)

林地	水域	草地	建筑	养殖水域
10124	46250	15687	23697	6942

表 3 为各方法在神湾地区影像上的地物类别识别精度。文中提出的 TFSV-IT2FCM 算法获得最好的 OA 精度，为 90.33%，相比无后处理的 TFSV-IT2FCM 算法、基于 Watershed 分割的 TFSV-IT2FCM 算法、IV-FCM 算法、面向对象的 IT2FCM 算法、基于像素的 IT2FCM 算法和 FCM 算法分别提升了 0.55%、0.52%、2.22%、4.81%、5.43% 的分类精度。分类精度也表明对于文中提出的聚类分割算法，聚类是最重要的一步，在方法的三个部分（分割、聚类、后处理）中最大化的决定最终分类结果精度。Kappa 系数结果也表明，文中提出的 TFSV-IT2FCM 算法有效题目了影像地物的特征，获得了最佳的聚类分割结果。表中数值量化的结果基本与前文可视化的结果一致。

#### 3.4.2 横琴地区实验

第二个研究区横琴岛位于广东省珠海市，岛内具有复杂的农业与养殖区，该区域高分影像由 SPOT 5 遥感卫星于 2010 年拍摄。横琴地区地物类别可分为水域、农业区、林地、农田和建筑这五大类。实验区选择影像图像如图 16(a) 所示，大小为  $1203 \times 1055$  个像素。

在神湾研究区我们比较了文中提出的 TFSV-IT2FCM 算法与其他算法的聚类分割结果，初步验证了 TFSV-IT2FCM 算法的改进效果。文献<sup>[8]</sup> 中证明了 IV-FCM 算法相比其他聚类分割算法可以取得更好的实验精度，因此，在横琴地区和三门峡地区的实验中，我们着重比较提出的 TFSV-IT2FCM 算法与 IV-FCM 算法的实验效果。

IV-FCM 算法和 TFSV-IT2FCM 算法的可视化结果分别如图 16(c) 和 (d) 所示。结合目视解译可知，两种方法对具有明显边界的地物类别都有良好的划分，基本实现同一类地物划分的一致性。参照图 16(b) 中的真实地物分类，对比 16(c) 和 (d) 中可视化结

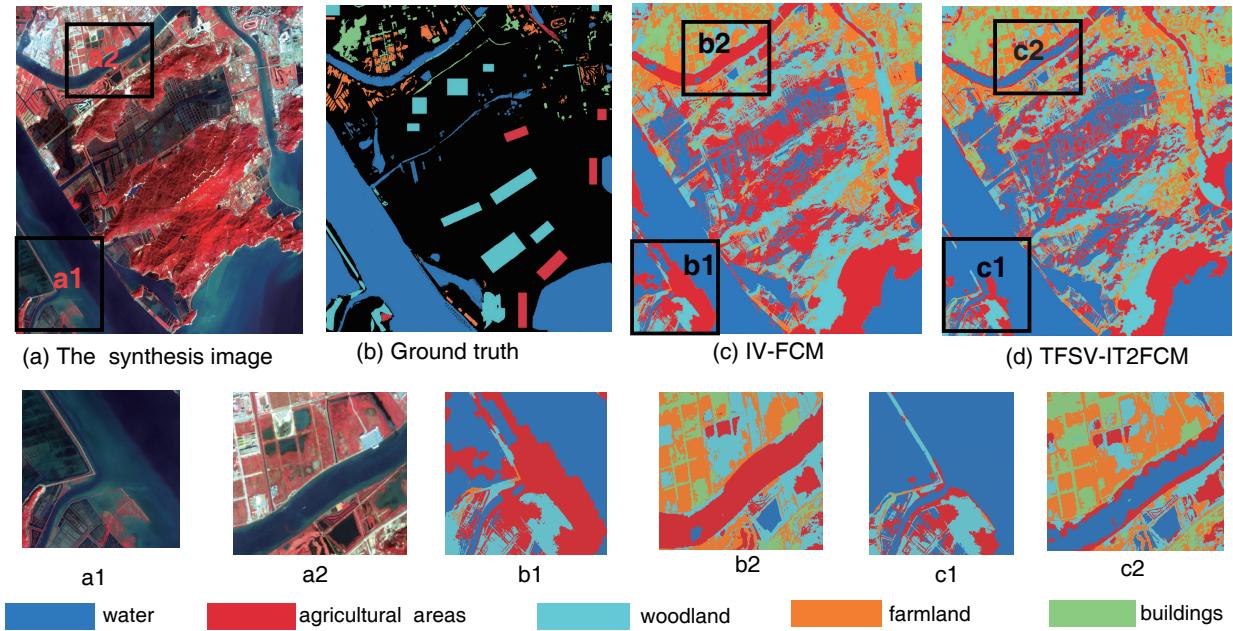


图 16 神湾研究区 TFSV-IT2FCM 与 IV-FCM 算法聚类分割图

表 3 神湾研究区不同方法分类精度表

地物类别	TFSV-IT2FCM <sup>1</sup>	TFSV-IT2FCM	TFSV-IT2FCM <sup>2</sup>	IV-FCM	IT2FCM <sup>2</sup>	IT2FCM <sup>4</sup>	FCM
	生产者精度 (%)						
林地	89.39	89.44	89.44	89.52	79.70	89.32	79.70
水域	94.08	94.58	94.18	93.59	94.06	94.09	94.21
草地	83.69	84.23	84.78	79.74	81.28	77.91	80.43
建筑	87.88	88.45	88.14	88.21	87.73	83.58	81.00
养殖水域	80.32	81.29	76.01	64.08	50.83	48.33	45.73
用户精度 (%)							
林地	82.98	83.39	83.39	84.10	77.46	82.40	77.46
水域	95.54	97.34	97.34	97.34	95.18	92.58	90.85
草地	82.96	82.96	82.96	80.69	70.21	72.32	63.84
建筑	91.95	90.68	87.97	86.04	84.60	86.04	86.46
养殖水域	69.27	69.27	70.71	56.31	70.71	61.93	69.27
总体分类精度 (%)	89.78	90.33	89.81	88.11	85.52	84.90	82.93
Kappa 系数	0.8544	0.8621	0.8546	0.8302	0.7954	0.7871	0.7605

<sup>1</sup> TFSV-IT2FCM 算法, 没有做后处理<sup>2</sup> 该方法中影像分割阶段使用 Watershed 方法获取影像分割单元<sup>3</sup> 面向对象的 IT2FCM 方法<sup>4</sup> 基于像素的 IT2FCM 方法

果, 可以看出 TFSV-IT2FCM 算法对于一定范围内不同光谱特征的同一类别具有较大的容忍性。例如, 图 16(a1) 左上角水域的光谱特征是不均匀的, 图 16(b1) 和 (c1) 结果表明 TFSV-IT2FCM 算法将该区域正确划分为水域, 由于水域的几个光谱与农业区相近, 而 IV-FCM 算法则将其错分为农业区。类似地, 对于图 16(a2) 中地物覆盖复杂的区域, TFSV-IT2FCM 算法能正确地将水与农田、建筑物等其他地物区分开来, 而 IV-FCM 算法则存在一定的分类错误。例如, 在图 16(b2) 中, 河流中的水被错误地划分为农业区; 然而, 在图 16(c2) 中, TFSV-IT2FCM 算法的结果清晰地识别出了河流两岸的水和其他物体。

表 4 横琴地区分类结果表

分类结果	TFSV-IT2FCM					IV-FCM						
	采样点					用户精度 (%)	采样点					用户精度 (%)
	水域	农业区	林地	农场	建筑		水域	农业区	林地	农场	建筑	
水域	241231	10022	44	0	0	95.99	196301	54380	557	59	0	78.12
农业区	21025	70848	4755	57	1	73.28	9810	78463	8338	74	1	81.15
林地	1108	11672	66320	2167	18	81.59	167	8971	71578	569	0	88.06
农场	0	43	4032	28234	724	85.47	0	7	10192	22785	49	68.98
建筑	0	0	0	1132	36236	96.97	0	0	0	4310	33058	88.47
合计	263364	92585	75151	31590	36979		206278	141821	90665	27797	33108	
生产者精度 (%)	91.60	76.52	88.25	89.38	97.99		95.16	55.33	78.95	81.97	99.85	

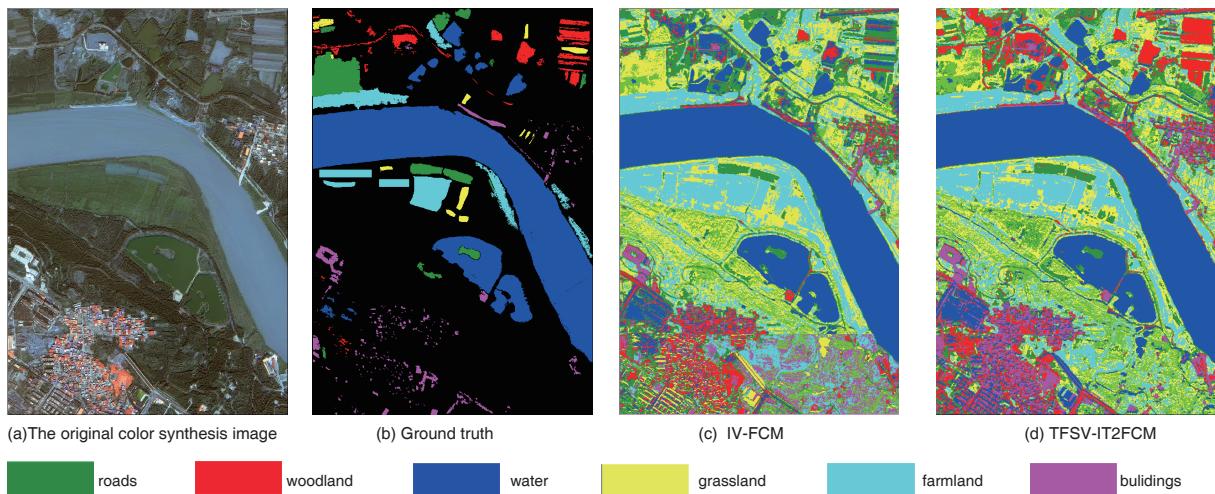


图 17 三门峡研究区不同方法分类结果示意图

同样, 为了获取横琴地区精确的分类结果, 文中使用 CORINE 土地分类系统随机采样 499,669 个像素点作为 ground truth 图, 如表 4 所示, TFSV-IT2FCM 方法中 10,022 个

像素由“农业区”被错分为“水域”，而在 IV-FCM 方法中被错分的像素点为 54,380 个。TFSV-IT2FCM 方法相比 IV-FCM 方法，“水域”类别的用户精度从 78.12% 提升到 95.99%。表 4 中的用户精度和生产者精度也表明 TFSV-IT2FCM 聚类方法的性能要比 IV-FCM 方法更好。

表 5 横琴地区分类精度表

研究区	方法	OA(%)	Kappa 系数
横琴	TFSV-IT2FCM	88.63	0.8290
	IV-FCM	80.49	0.7210

表 5 中比较了 TFSV-IT2FCM 方法和 IV-FCM 方法在横琴地区的总体分类精度和 Kappa 系数，TFSV-IT2FCM 方法的 OA 从 80.49% 提升到 88.63%，同时 Kappa 系数由 0.7210 提升到 0.8290。

### 3.4.3 三门峡地区实验

神湾地区与横琴地区遥感影像均由 SPOT 5 遥感卫星拍摄，为了验证不同卫星影像的聚类分割结果，文中选取高分二号遥感卫星拍摄三门峡地区的遥感影像，影像空间分辨率 1 米。三门峡研究区位于河南省三门峡市，具有复杂的城市地形。实验中选择影像 1、2、3 三个波段合成 RGB 图像，如图 17(a) 所示，其包含  $1625 \times 2320$  个像素。

表 6 三门峡地区分类结果表

分类结果	TFSV-IT2FCM						IV-FCM					
	采样点						采样点					
	用户精度 (%)						用户精度 (%)					
	道路	林地	水域	草地	农田	建筑		道路	林地	水域	草地	农田
道路	72564	4598	0	2965	368	3770	86.11	62754	12385	341	3169	513
林地	591	197835	6542	39638	456	376	80.60	891	176238	25324	31528	301
水域	0	0	310128	1360	2436	2625	97.97	0	358	310524	1257	2613
草地	3421	16329	865	249431	405	0	92.23	2598	25341	516	231265	5639
农田	0	673	1539	16234	291035	2391	93.32	0	408	2538	14840	289024
建筑	8589	23768	356	3568	4997	283703	87.30	6589	85961	0	2167	3219
合计	85165	243203	319430	313196	299697	292865		72832	300691	339243	284226	301309
生产者精度 (%)	85.20	81.35	97.09	79.64	97.11	0.97		86.16	58.61	91.53	81.37	95.92
												88.95

高分二号卫星拍摄的三门峡城市中心区域如图 17 所示，由原始合成图 17(a) 中可知，建筑、道路与农田交错复杂。与 TFSV-IT2FCM 算法相比，IV-FCM 算法无法精确的区分

图中右下角区域的建筑、道路与农田类别。图 17(d) 中可视化结果表明 TFSV-IT2FCM 算法对这三类可以产生明确的边界，分割效果更好。

同样地，随机采样 1553556 个像素点生成参考 ground truth 图，在三门峡地区对六个地物类别上统计生产者精度和用户精度，结果如表 6 所示，TFSV-IT2FCM 算法对“草地”的用户精度有近 8% 的绝对值提升。其他类别精度数据也有不同幅度的提升。表 7 中所示 TFSV-IT2FCM 算法在三门峡研究区获得的 OA 精度为 90.42%，插死 Kappa 系数达到 0.8827，整体比 IV-FCM 算法提升近 7%。

表 7 三门峡地区分类精度表

研究区	方法	OA(%)	Kappa 系数
三门峡	TFSV-IT2FCM	90.42	0.8827
	IV-FCM	83.48	0.7978

为验证文中提出的 TFSV-IT2FCM 算法的有效性，文中分别在神湾、横琴和三门峡三个研究区进行高分影像模糊聚类的地物分割 sh 识别实验。与 IV-FCM 算法和其他模糊聚类方法比较得到的可视化结果和量化结果均表明文中提出的 TFSV-IT2FCM 算法的有效性，其更能区分遥感影像相似光谱不同类别地物的边界。同时，TFSV-IT2FCM 算法在 SPOT 5 影像（神湾和横琴地区）和高分二号影像（三门峡地区）上的聚类分割均能取得一定的识别优势，可见其不受影像数据源的影像。

### 3.5 本章小结

本章从遥感影像不确定性特征表达角度展开研究，新定义 TFSV 数据模型来提取影像分割单元内在特征信息，同时提出一种新区间值距离来度量两个三角形模糊集间的相似性。结合新的数据模型和距离度量提出 TFSV-IT2FCM 模糊聚类分割算法。在 SPOT 5 和高分二号影像数据上的实验结果均证明了文中提出的 TFSV-IT2FCM 算法的有效性，其确实能更好刻画遥感影像数据的不确定性，实现更好的分割识别结果。

# 第4章 基于生成对抗网络的遥感影像分割方法

## 4.1 引言

深度学习方法通过组合低阶特征形成更加抽象的高阶表示，基于全卷积的影像分割方法其多层的卷积网络结构可以完成对输入图像特征的自动学习。FCN模型反卷积结构将特征图恢复到原始图像尺寸，实现影像像素级分类。然而，上采样过程造成了特征的损失，导致预测结果边界模糊的问题。另外，高分影像由于细节特征复杂，全卷积网络分割结果同类地物不够连续，导致地物边界模糊、歧义性等问题更加严重，且同类地物内部像素难以保持类别一致性。

生成对抗网络（Generative Adversarial Networks, GAN）训练时生成器（Generator）与判别器（Discriminator）不断对抗博弈，网络迭代好时判别器无法判断来源为生成器生成结果还是真实标签图，此时生成网络具有强大的图像生成能力<sup>[67]</sup>。本章将对抗网络应用到全卷积图像分割中，使得分割结果能够产生更好的地物边界，同类别物体空间上具有一致性。

## 4.2 基于生成对抗网络框架的影像分割方法

### 4.2.1 生成对抗网络框架

GAN是一种生成式的对抗网络，即通过不断对抗博弈，模拟数据分布的生成式模型。生成网络目标生成逼近真实数据的结果，判别网络则需要判别输入是生成的假样本还是真实数据。

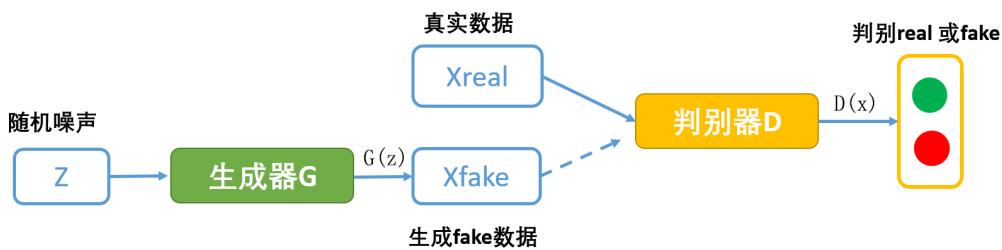


图 18 GAN 结构示意图

如图18所示，假定变量 $z$ （通常为服从高斯分布的随机噪声）通过生成器 $G$ 生成

$X_{fake}$ , 判别器 D 需要决策模型输入数据是真实样本  $X_{real}$  还是生成的样本  $X_{fake}$ 。GAN 优化的目标函数为4-1:

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (4-1)$$

式中  $x \sim p_{data}(x)$  表示  $x$  取自真实的分布数据。求解判别器 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。生成器 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4-2:

$$D^*(x) = \frac{p_g(x)}{p_g(x) + p_{data}(x)} \quad (4-2)$$

文献<sup>[44]</sup> 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, 此时  $p_g(x) = p_{data}(x)$ , 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。

传统的 GAN 结构是无监督模型, 只能生成真实的数据, 不能生成我们想要的某一种类型的数据。文献<sup>[45]</sup> 提出条件生成对抗网络 (Conditional Generative Adversarial Networks, CGAN), 网路中加入条件约束  $y$  引导模型迭代方向, 生成实验中需要的目标类型数据,  $y$  可以是任何种类的辅助信息, 如类别标签, 影像 Ground truth 图或其他不同领域模态的数据等。

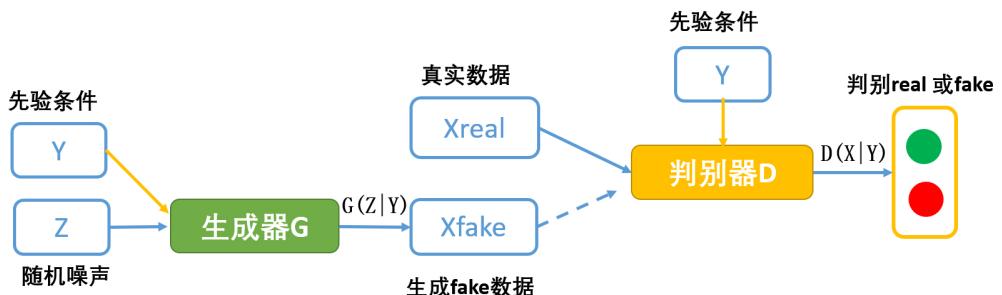


图 19 CGAN 结构示意图

CGAN 模型中, G 中随机噪声  $z$  与先验知识  $y$  拼接组成联合输入特征; D 中  $X$  和  $y$  通过词嵌入共同作为模型输入。类似式4-1, 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：

$$\min_G \max_D V(D, G) = \min_G \max_D E_{x \sim p_{data}(x)}[\log D(x|y)] + E_{z \sim p_z(z)}[\log(1 - D(G(z|y)))] \quad (4-3)$$

图19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。

#### 4.2.2 基于 CGAN 的全卷积影像分割方法

基于 CGAN 的影像分割模型主要包含两个阶段：生成网络的影像分割模型和对抗训练阶段的判别模型。整个方法处理流程如图20所示，左边生成网络是一个全卷积影像分割模型，右边对抗网络是一个二元分类判别模型。

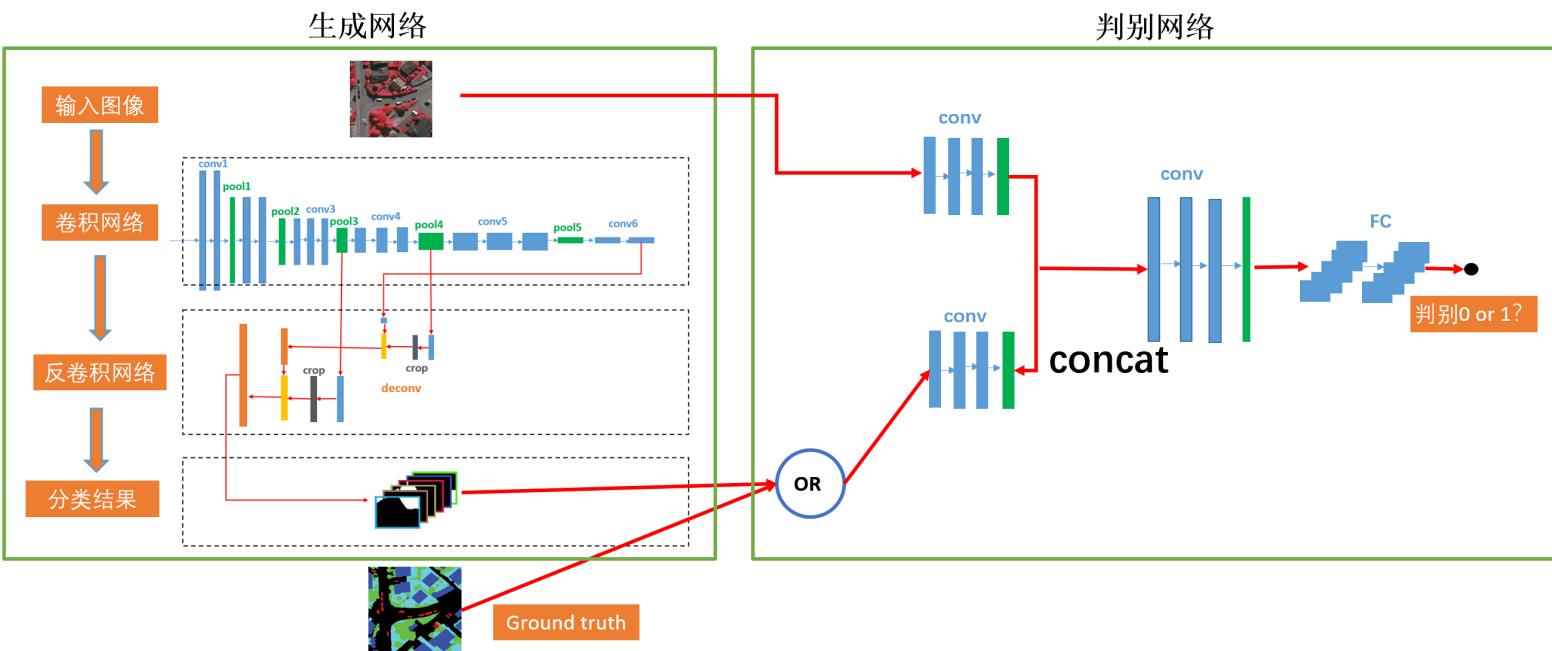


图 20 基于对抗训练框架的全卷积语义分割模型示意图

对抗网络的输入存在两种情况：一种是原始图像和 Ground truth 图拼接输入，另一种是原始图像与左边模型输出分割结果拼接输入。本文中对抗网络模型为一个六层卷积神经网络模型，一个卷积块由三个卷积层后接一个最大池化层组成，整个卷积网络由两个卷积块和两个全连接层堆叠形成。对抗网络的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）。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<sup>[68]</sup>, 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:

$$l_{bce}(\hat{z}, z) = -[z \log \hat{z} + (1 - z) \log(1 - \hat{z})] \quad (4-4)$$

生成网络是全卷积网络模型, 其是一个多分类网络模型。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), 则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:

$$l_{mce}(\hat{y}, y) = - \sum_{i=1}^{H \times W} \sum_{c=1}^C y_{ic} \log \hat{y}_{ic} \quad (4-5)$$

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:

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad \forall j \in 1, 2, \dots, K \quad (4-6)$$

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4-7:

$$L(\theta_g, \theta_d) = \sum_{n=1}^N \{l_{mce}(g(x_n), y(n)) - \lambda [l_{bce}(d(x_n, y_n), 1) + l_{bce}(d(x_n, g(x_n)), 0)]\} \quad (4-7)$$

式中  $\theta_g$  和  $\theta_d$  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,  $\lambda$  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4-7 中的代价函数，从而迭代求解得到网络参数  $\theta_g$  和  $\theta_d$ 。

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4-6 中 Softmax 函数求解，再利用  $\text{argmax}$  函数求出最大概率值对应类别标签维度，即为当前像素的类别标签，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4-8。

$$C_k = \arg \max_{k \in K} p_k(x^{(i)}), k = 1, 2, \dots, K \quad (4-8)$$

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。

## 4.3 实验数据介绍与预处理

### 4.3.1 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<sup>①</sup>。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21(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21(b-d) 分别为该影像某一区域的正射影像、DSM 高程、Ground truth 图。

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8 所示。

### 4.3.2 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

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

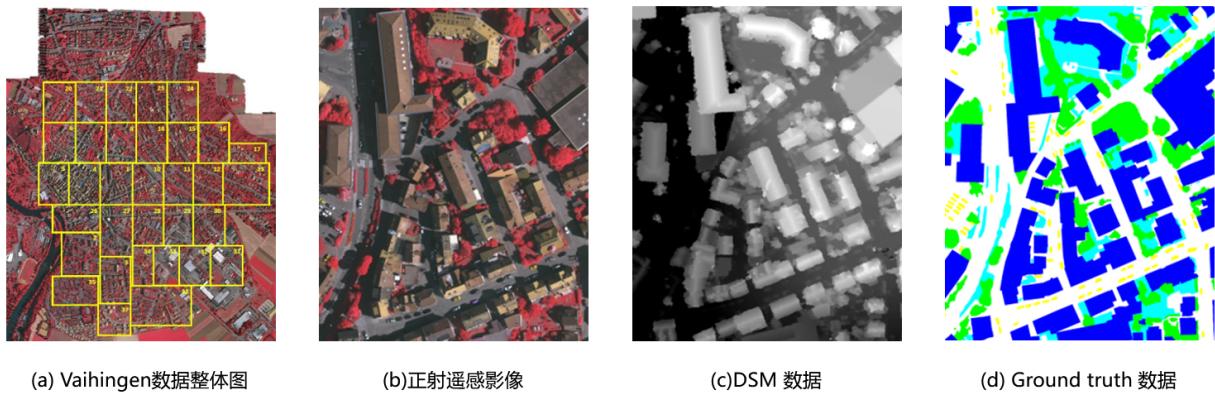


图 21 Vaihingen 影像数据

表 8 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

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21 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4-9 所示，

$$z = \frac{x - \mu}{\sigma} \quad (4-9)$$

式中  $x$  为原始输入数据， $\mu$  和  $\sigma$  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4-10 所示：

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4-10)$$

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  $2563 \times 2049$ ，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  $256 \times 256$  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。

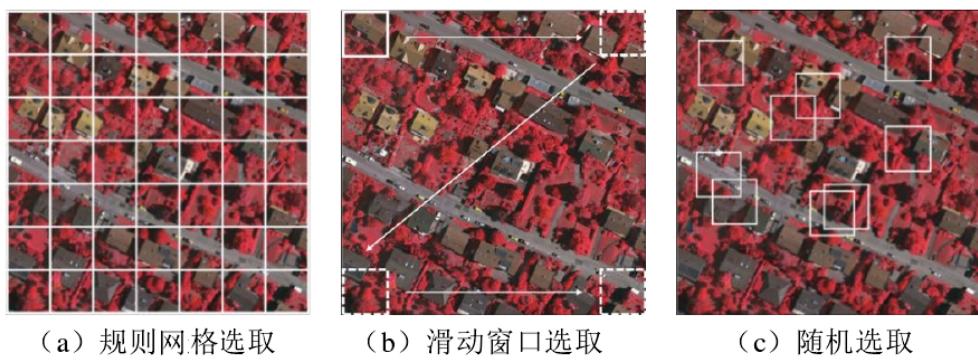


图 22 图像裁剪方式

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  $256 \times 256$  的网格切分影像，这种划分方式得到的样本量有限。滑动窗口裁剪方法是使用  $256 \times 256$  的滑动窗口以一定的滑动间隔在原始影像上滑动得到训练样本，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  $256 \times 256$  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22 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  $256 \times 256$  的测试图像合计 320 张。

## 4.4 实验结果与分析

### 4.4.1 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。

### 4.4.2 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：

$$OA = \frac{1}{A_{\text{总}}} \sum_{k=1}^K a_{kk} \quad (4-11)$$

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4-12：

$$IoU = \frac{\text{预测结果} \cap \text{真实标签}}{\text{预测结果} \cup \text{真实标签}} \quad (4-12)$$

影像所有类别地物的 mIoU 的计算方式如式4-13：

$$mIoU = \frac{1}{K} \sum_{i=1}^K \frac{a_{ii}}{\sum_{j=1}^K a_{ij} + \sum_{j=1}^K a_{ji} - a_{ii}} \quad (4-13)$$

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。

#### 4.4.3 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 VGG16 的全卷积网络权值初始化，式4-7 中 GAN 模型中的目标函数权衡因子为  $\lambda = 2$ ，使用 Adam<sup>[69]</sup> 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  $2 \times 2$  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  $1 \times 1$  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 9 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	G-输入	level 0		256 × 256 × 64			
	卷积压缩	level 1	conv1_1	256 × 256 × 64	3 × 3/64	1	ReLU
			conv1_2	256 × 256 × 64	3 × 3/64	1	ReLU
		level 2	pool2_1	128 × 128 × 64	2 × 2/-	2	--
			conv2_1	128 × 128 × 128	3 × 3/128	1	ReLU
			conv2_1	128 × 128 × 128	3 × 3/128	1	ReLU
		level 3	pool3_1	64 × 64 × 128	2 × 2/-	2	--
			conv3_1	64 × 64 × 256	3 × 3/256	1	ReLU
			conv3_2	64 × 64 × 256	3 × 3/256	1	ReLU
			conv3_3	64 × 64 × 256	3 × 3/256	1	ReLU
	level 4	pool4_1	32 × 32 × 256	2 × 2/-	2	--	
			conv4_1	32 × 32 × 512	3 × 3/512	1	ReLU
		conv4_2	32 × 32 × 512	3 × 3/512	1	ReLU	
			conv4_3	32 × 32 × 512	3 × 3/512	1	ReLU
	level 5	pool5_1	16 × 16 × 512	2 × 2/-	2	--	
			conv5_1	16 × 16 × 512	3 × 3/512	1	ReLU
		conv5_2	16 × 16 × 512	3 × 3/512	1	ReLU	
			conv5_3	16 × 16 × 512	3 × 3/512	1	ReLU
	反卷积恢复	level 6	deconv6_1	32 × 32 × 512	4 × 4/512	2	ReLU
			conv6_1	32 × 32 × 512	3 × 3/512	1	ReLU
			conv6_2	32 × 32 × 512	3 × 3/512	1	ReLU
		level 7	deconv7_1	64 × 64 × 256	4 × 4/256	2	ReLU
			conv7_1	64 × 64 × 256	3 × 3/256	1	ReLU
			conv7_2	64 × 64 × 256	3 × 3/256	1	ReLU
		level 8	deconv8_1	128 × 128 × 128	4 × 4/128	2	ReLU
			conv8_1	128 × 128 × 128	3 × 3/128	1	ReLU
			conv8_2	128 × 128 × 128	3 × 3/128	1	ReLU
		level 9	deconv9_1	256 × 256 × 64	4 × 4/64	2	ReLU
			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			

(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
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 生成预测结果），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9 所示，通过交替训练生成器 G 和判别器 D 实现网络各权值参数的学习。

#### 4.4.4 结果与分析

本章提出的方法是将全卷积分割方法与条件生成对抗网络模型结果，利用 CGAN 模型优秀的生成能力获得更精细、准确的地物分割边界，实验中采用的是融合 DSM 数据的四波段影像数据，参考的实验方法为经典的 FCN 语义分割方法。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10 比较了新提出的基于 CGAN 的分割模型和 FCN 模型的分割精度，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）的分割精度，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 的分割方法在“建筑物”这个类别地物识别精度最高，为 87.64%，相比 FCN 分割方法的 83.96%，有约 4% 的绝对精度提升。而对 FCN 分割中精度较低的“低矮植被”地物，分割精度由 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 的分割方法对地物的边界有更准确的生成能力。

本文提出基于 CGAN 的分割方法 OA 为 80.15%，mIoU 为 61.83%，相比 FCN 分割算法中 78.48% 的 OA 和 58.42% 的 mIoU，均有一定程度的提升。

表 10 Vaihingen 数据分割精度评估表

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

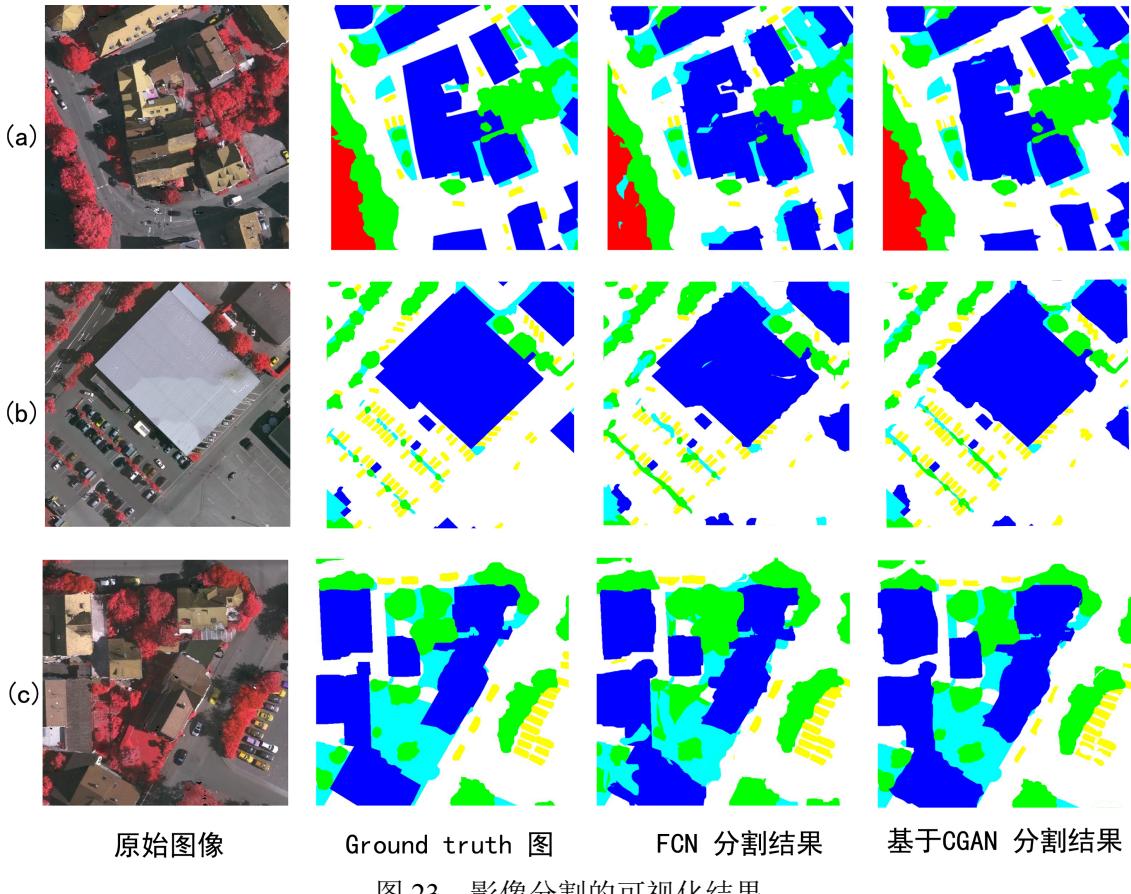


图 23 影像分割的可视化结果

图23 为本文提出的基于 CGAN 的分割方法和 FCN 分割方法分别在三组测试图像上分割的可视化结果。图23(a) 为一居民住宅区，参照真实地物类别的 Ground truth 图，相比 FCN 分割方法，基于 CGAN 的分割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，从易混淆的阴影中区分开。图23(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 分割算法无法识别出这些“车辆”，直接将车错分为背景，而基于 CGAN 的分割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23(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割的难点。如图正上方区域房屋环绕的树木与低矮植被区域，基于 CGAN 的分割方法将图中的“树木”与“低矮植被”进行了区分，而 FCN 的分割方法则将“低矮植被”识别为“树木”。综合上面三组图像的分割结果，文中提出的基于 CGAN 的分割方法相比 FCN 分割方法能够区分特征相近地物类别，生成更准确的分割边界。

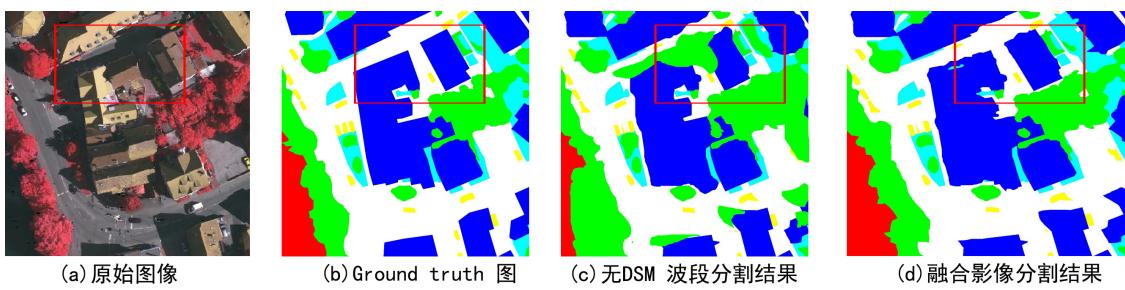


图 24 DSM 波段对影像分类结果的影响

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24 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”误别错分为“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11 则统计了有无 DSM 波段的数据源对实验分割精度的影响，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割精度和分割效果。

## 4.5 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，同时，全卷积语义分割上采样过程会损失影像的特征细节，导致遥感影像地物边界更加难以划分。本章提

表 11 DSM 波段对分割精度的影响

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%

出的方法将生成对抗网络的思想应用到全卷积分割模型中，借助 CGAN 网络强大的图像生成能力，使得影像分割结果具有更准确的地物边界，同类别地物空间上具有一致性。将文中提出的基于 CGAN 的分割方法应用到 Vaihingen 影像数据分割实验上，可视化的分割结果和量化的评估精度均表明文中提出的方法相比传统 FCN 语义分割方法有着更好的分割效果。另外，融合 DSM 波段高程信息的输入数据，可以进一步识别特征相近的不同地物类别。

## 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。针对遥感影像数据同物异谱、同谱异物等固有不确定性的特点，设计三角形模糊集值信息表达模型来表征影像数据特征，提出 TFSV-IT2FCM 模糊聚类方法用于高分影像的地物无监督聚类分割识别。针对全卷积影像语义分割中高分影像地物类别一致性问题，文中结合 CGAN 优秀的图像生成能力，提出了基于 CGAN 的全卷积分割方法来得到清晰、明确的地物分类边界。将文中提出的两种方法分别用于遥感影像地物分类实验中，实验结果均表明相比已有方法，新提出的影像分割方法提高了地物分类的精度，能得到更准确的地物分类边界。具体工作如下：

- (1) 针对遥感影像数据固有的不确定性，设计三角形模糊集值表达模型提取影像特征。对于新设计的 TFSV 数据模型，提出新的区间值距离度量两个 TFSV 数据间的相似度。利用新提取的数据模型和距离度量改进现有面向对象的遥感影像模糊聚类分割方法，提出 TFSV-IT2FCM 聚类分割算法。在 SPOT 5 和高分二号影像上进行地物聚类分割均证明了文中提出的 TFSV-IT2FCM 算法的有效性，其确实能识别近似光谱特征的不同地物，实现更好的分割识别结果。
- (2) 针对全卷积分割方法中上采样特征损失的问题，结合对抗网络的思想，提出基于 CGAN 的全卷积语义分割方法。另外，正射影像波段和 DSM 高程波段数据融合用于模型训练。在 Vaihingen 影像数据上的实验结果表明新的分割方法能够得到更好的地物分割边界，且同一类别的地物具有更好的空间一致性。

### 5.2 未来的期望

本文提出的两种方法，一定程度上均能改进遥感影像地物分类边界划分问题，对提高影像分类的识别精度有所提高。之后的研究可从以下方面开展：

- (1) 第3章提出的模糊聚类分割方法使用最简单的三角形模糊集对数据特征建模，使用其他类型模糊集（如梯形、钟形模糊集）对影像单元建模可以展开进一步的研究。
- (2) 第4章提出的基于CGAN分割方法在识别“车辆”时相对其他几类地物精度较低，原因主要是训练集影像中含有“车辆”的像素点占比比较低，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。
- (3) 第3章提出的模糊聚类分割方法能够保证类内地物整体一致，同时能较好的识别不同地物边界，后续工作中可以将该模糊聚类分割方法作为后处理过程添加到第4章基于CGAN分割方法中，期望同类地物内分割效果更好。

## 参考文献

- [1] 李石华, 王金亮, 毕艳, et al. 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, et al. 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, et al. 地理信息系统数据的不确定性问题 [D]. 2006.
- [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. 604408.
- [6] James C Bezdek, Robert Ehrlich, William Full. FCM: The fuzzy c-means clustering algorithm[J]. Computers & Geosciences, 1984, 10(2-3):191--203.
- [7] XC Yu, WJ An, Hui He. A method of auto classification based on object oriented unsupervised classification[J]. Progress in Geophysics, 2012, 27(2):744--749.
- [8] 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.
- [9] Geoffrey E Hinton, Simon Osindero, Yee-Whye Teh. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7):1527--1554.
- [10] 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.
- [11] Yoshua Bengio, et al. Learning deep architectures for AI[J]. Foundations and trends® in Machine Learning, 2009, 2(1):1--127.
- [12] 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.
- [13] John R Jensen, Kalmesh Lulla. Introductory digital image processing: a remote sensing perspective[J]. 1987.

- [14] 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.
- [15] AG Wacker, DA Landgrebe. Minimum distance classification in remote sensing[J]. 1972.
- [16] 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.
- [17] Peter M Atkinson, P Lewis. Geostatistical classification for remote sensing: an introduction[J]. Computers & Geosciences, 2000, 26(4):361--371.
- [18] 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.
- [19] 张永生, 巩丹超, 刘军, et al. 高分辨率遥感卫星应用 [J]. 科学出版社, 北京, 2004.
- [20] Thomas Blaschke. Object based image analysis for remote sensing[J]. ISPRS journal of photogrammetry and remote sensing, 2010, 65(1):2--16.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] John Canny. A computational approach to edge detection[C]. Proceedings of Readings in computer vision. Elsevier, 1987: 184--203.
- [27] Nobuyuki Otsu. A threshold selection method from gray-level histograms[J]. IEEE transactions on systems, man, and cybernetics, 1979, 9(1):62--66.
- [28] 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.

- [29] 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.
- [30] Lior Weizman, Jacob Goldberger. Urban-area segmentation using visual words[J]. IEEE Geoscience and Remote Sensing Letters, 2009, 6(3):388--392.
- [31] 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.
- [32] Mahesh Pal. Random forest classifier for remote sensing classification[J]. International Journal of Remote Sensing, 2005, 26(1):217--222.
- [33] Johan AK Suykens, Joos Vandewalle. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3):293--300.
- [34] 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.
- [35] Simon Haykin. Neural networks[M], volume 2. Prentice hall New York, 1994.
- [36] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm[C]. Proceedings of icml, volume 96. Citeseer, 1996. 148--156.
- [37] Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep learning[J]. nature, 2015, 521(7553):436.
- [38] 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.
- [39] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] Andrew Ng, et al. Sparse autoencoder[J]. CS294A Lecture notes, 2011, 72(2011):1--19.

- [44] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets[C]. Proceedings of Advances in neural information processing systems, 2014. 2672--2680.
- [45] Mehdi Mirza, Simon Osindero. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.
- [46] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
- [47] 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.
- [48] Martin Arjovsky, Soumith Chintala, Léon Bottou. Wasserstein gan[J]. arXiv preprint arXiv:1701.07875, 2017.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] 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.
- [55] 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.
- [56] 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.

- [57] 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.
- [58] Lotfi A Zadeh. Fuzzy sets[J]. Information and control, 1965, 8(3):338--353.
- [59] Lotfi Asker Zadeh. Fuzzy sets as a basis for a theory of possibility[J]. Fuzzy sets and systems, 1978, 1(1):3--28.
- [60] Feilong Liu. An efficient centroid type-reduction strategy for general type-2 fuzzy logic system[J]. Information Sciences, 2008, 178(9):2224--2236.
- [61] 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.
- [62] Jos BTM Roerdink, Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies[J]. Fundamenta informaticae, 2000, 41(1, 2):187--228.
- [63] Cheul Hwang, Frank Chung-Hoon Rhee. Uncertain fuzzy clustering: Interval type-2 fuzzy approach to  $c$ -means[J]. IEEE Transactions on Fuzzy Systems, 2007, 15(1):107--120.
- [64] Nilesh N Karnik, Jerry M Mendel. Centroid of a type-2 fuzzy set[J]. Information Sciences, 2001, 132(1-4):195--220.
- [65] Dongrui Wu, Jerry M Mendel. Enhanced karnik--mendel algorithms[J]. IEEE Transactions on Fuzzy Systems, 2009, 17(4):923--934.
- [66] Jing-hua Zhang, Zhiming Feng, Lu-guang Jiang. Progress on studies of land use/land cover classification systems[J]. Resources Science, 2011, 33(6):1195--1203.
- [67] Pauline Luc, Camille Couprie, Soumith Chintala, et al. Semantic segmentation using adversarial networks[J]. arXiv preprint arXiv:1611.08408, 2016.
- [68] 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.
- [69] Diederik P Kingma, Jimmy Ba. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

## 学术成果

1. **Tao Jiang**, Dan Hu, and Xianchuan Yu. 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 Xianchuan Yu. 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, Xianchuan Yu, 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