

密级:\_\_\_\_\_



**中国科学院大学**  
University of Chinese Academy of Sciences

# 硕士学位论文

**基于多示例学习的弱监督遥感图像目标检测**

作者姓名: 戴蔚群

指导教师: 韩振军 副教授 中国科学院大学

学位类别: 工程硕士

学科专业: 工业工程

研究所 : 中国科学院大学工程科学学院

2016年05月



**Object Detection in Remote Sensing Images by  
Weakly Supervised Learning based on Multiple  
Instance Learning**

**By**

**Dai Wei Qun**

**A Thesis Submitted to**

**University of Chinese Academy of Sciences**

**In partial fulfillment of the requirement**

**For the degree of**

**Master of Industrial Engineering**

**School of Engineering Science**

**University of Chinese Academy of Sciences**

**May, 2016**



## 中国科学院大学直属院系 研究生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。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

## 中国科学院大学直属院系 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：



## 摘 要

遥感图像具有丰富的信息和独特的视角,因此,对遥感图像特定目标进行检测的研究在国防军事、城市规划、灾难检测、交通建设、生态研究等方面具有广阔的应用场景。近几年,大多数的遥感图像目标检测方法都是基于监督学习的。监督学习虽然是机器学习中的经典场景,但其存在需要完备标注信息的缺点,需要耗费巨大的人力物力对数据进行标定。本文就此问题提出基于弱监督学习场景的遥感图像目标检测框架。弱监督学习作为近年来热门的新兴研究方向,包含多种学习类型,其最大特点是使用弱标注数据便可达到模型学习目的,更适用于实际问题场景。本文的研究仅使用极弱标注信息,通过样本挖掘、分类器训练、样本优化及迭代更新等过程最终达到学习目的。为了解决遥感图像中复杂场景或目标模糊等问题,本文引入了样本种子点来初始化训练样本。在本课题 UCAS-AOD 数据集进行实验证明本框架仅使用少量标注,所提出方法的性能便可达到接近监督学习方法的水平。

本文主要贡献如下:

1) 将遥感图像对应到多示例学习中的包概念。以图像提取的窗口作为包内的示例;通过改进后的扩充式多示例学习方法,迭代地从包内选择正例训练样本;为了保证样本挖掘的准确率,本框架加入样本优化步骤对选取的训练样本进行筛选。

2) 为了进一步保证样本的多样性和准确率,将协同训练方法与多示例学习进行结合作为研究的核心框架。按照协同训练方法的思想,将训练数据分为两个视图,视图内的分类器各自为对方选择训练样本。

3) 引入种子点样本来解决遥感图像常见的复杂场景和目标模糊等问题。以极少量的标注代价,高效地引导样本挖掘和分类器训练等过程。

**关键词:** 弱监督学习, 遥感图像, 目标检测, 多示例学习, 协同训练





## Abstract

Object detection in remote sensing images plays a significant role in the field of computer vision, because it contains abundant information in expanse of regions with a special perspective which is distinct from traditional image detection. Robust object detection in remote sensing images is potentially applicable in security, military, urban planning, large scale damage detection, traffic surveillance and ecological studies. In recent years, a majority of object detection approaches in remote sensing images based on supervised learning are proposed. Although the supervised learning is a classical task in machine learning, its major drawback is requiring a large amount of manual annotation effort. In this thesis, we propose a detection framework in remote sensing images based on weakly supervised learning that only require weakly labeled data. It consist of sample mining, model training, sample refining and iterative optimization, etc. Considering the complex scenes and indistinct objects, we import less seed samples to initialize training data. With experimental results on the UCAS-AOD dataset, we demonstrate that our method achieves performance close to supervised learning with far less annotations.

The primary works in this thesis are as follows:

1) In this thesis, each weakly labeled image is considered as a collection of regions which form a bag. The regions in each image are considered as instances in the bag. The extended-multiple instance learning approach is used to select positive training samples from each bag. We use the sample refining to ensure the accuracy of sample mining.

2) Ensure the diversity and precision of samples we mined, this thesis propose the combination of multiple-instance learning and co-training as the core of framework. As the assumption of co-training, we split training data into two parts as different views. Each model selects the training samples for the one in the other view.

3) To solve the complex scenes and indistinct objects problems in remote sensing images, we import the seed samples. A small amount of labeled data are used to initialize the sample mining and classifier training.

**Key Words:** weakly supervised learning, remote sensing images, object detection, multiple-instance learning, co-training

## 目录

摘 要.....	I
Abstract.....	III
图目录.....	VII
表目录.....	IX
<b>第 1 章 绪论</b> .....	<b>1</b>
1.1 课题背景和研究意义.....	1
1.2 国内外研究进展.....	3
1.2.1 基于监督学习的研究.....	3
1.2.2 基于无监督和半监督学习的研究.....	5
1.2.3 基于弱监督学习的研究.....	5
1.3 本文研究内容.....	6
1.4 本文的组织结构.....	7
<b>第 2 章 相关技术研究</b> .....	<b>9</b>
2.1 弱监督学习.....	9
2.1.1 弱监督学习的概念.....	9
2.1.2 弱监督学习分类.....	9
2.1.3 弱监督学习的常用算法.....	12
2.2 区域选择与特征提取.....	15
2.2.1 Selective Search.....	15
2.2.2 深度学习.....	16
2.3 多示例学习.....	18
2.3.1 多示例学习的概念.....	19
2.3.2 多示例学习的分类.....	19
2.3.3 基于支持向量机的多示例学习.....	22
2.4 协同训练.....	23
2.4.1 协同训练的概念.....	23
2.4.2 协同训练的分类.....	24
2.5 本章小结.....	26
<b>第 3 章 基于多示例学习的弱监督遥感图像目标检测</b> .....	<b>27</b>
3.1 问题描述及难点分析.....	27
3.1.1 问题描述.....	27
3.1.2 难点分析.....	28
3.2 基于多示例学习的研究框架.....	29
3.2.1 整体研究框架.....	29
3.2.2 区域提取-Selective Search.....	31

3.2.3	特征提取-CNN.....	32
3.2.4	多示例学习 .....	32
3.2.5	协同训练 .....	34
3.2.6	样本优化 .....	36
3.3	实验结果与分析.....	36
3.3.1	数据集介绍 .....	36
3.3.2	实验结果与分析 .....	38
3.4	本章小结.....	45
<b>第 4 章</b>	<b>基于样本种子点的弱监督遥感图像目标检测 .....</b>	<b>47</b>
4.1	问题描述.....	47
4.2	基于样本种子点的研究框架.....	48
4.2.1	基于样本种子点的初始化方法 .....	48
4.2.2	基于样本种子点的难例挖掘方法 .....	49
4.2.3	整体实验框架 .....	50
4.3	实验结果与分析.....	51
4.4	本章小结.....	58
<b>第 5 章</b>	<b>结论与展望.....</b>	<b>61</b>
	<b>参 考 文 献.....</b>	<b>63</b>
	<b>致    谢.....</b>	<b>69</b>
	<b>个人简历、在学期间发表的论文与研究成果 .....</b>	<b>71</b>

## 图目录

图 1-1	遥感图像目标检测大致流程	2
图 1-2	本文研究整体框架	6
图 2-1	半监督学习示意图	10
图 2-2	多示例学习示意图	11
图 2-3	多标记学习流程	11
图 2-4	传统 SVM 与 S3VM 分类面区别	14
图 2-5	深度学习特征与人脑层级	16
图 2-6	卷积神经网络概念示范	17
图 2-7	稀疏连接示意图	18
图 2-8	Weidmann 的广义多实例学习层级结构图	20
图 2-9	标准协同训练流程图	24
图 3-1	本章问题描述	27
图 3-2	遥感复杂场景示例	28
图 3-3	目标相似物干扰	29
图 3-4	目标旋转范围变化大	29
图 3-5	框架模块流程图	30
图 3-6	本研究流程图	30
图 3-7	Alex-Net 结构图	32
图 3-8	本课题中的多示例学习问题	33
图 3-9	本课题 MI-SVM 的改进	34
图 3-10	基于数据集划分的协同训练方法	35
图 3-11	基于属性集划分的协同训练方法	35
图 3-12	PAMView 数据图像示例	37
图 3-13	OIRDS 数据集图像示例	37
图 3-14	中科院大学高清航拍目标数据集 (UCAS-AOD) 示例	38
图 3-15	距离度量方式实验结果	39
图 3-16	每轮迭代中训练样本的准确率	40
图 3-17	每轮迭代中训练样本的召回率	41
图 3-18	协同训练实验比较	42
图 3-19	每轮迭代中正例训练样本的挖掘结果	42
图 3-20	本框架与监督学习性能对比	43
图 3-21	本框架在飞机数据集检测结果示例	44
图 4-1	复杂场景数据集示例	47
图 4-2	图像区域的 K-means 聚类过程	48
图 4-3	两种聚类选择方法示意	48
图 4-4	难例样本示意	50

---

图 4-5	基于种子点样本弱监督遥感图像目标检测流程图.....	51
图 4-6	聚类数目实验结果.....	52
图 4-7	两种聚类选择准确率比较.....	53
图 4-8	两种聚类选择召回率比较.....	53
图 4-9	训练样本初始化两种聚类选择方法结果示例.....	54
图 4-10	不同种子点样本个数比较.....	55
图 4-11	难例挖掘样本准确率比较.....	55
图 4-12	难例挖掘样本召回率比较.....	56
图 4-13	弱监督学习与监督学习的汽车实验性能对比.....	56
图 4-14	本框架在汽车数据集检测结果示例.....	57

## 表目录

表 2-1 自学习算法 .....	12
表 2-2 EM 算法 .....	13
表 2-3 Selective Search 算法 .....	15
表 2-4 mi-SVM 算法 .....	22
表 2-5 MI-SVM 算法 .....	23
表 2-6 标准协同训练算法 .....	24
表 3-1 数据集信息统计 .....	38
表 3-2 距离度量方式比较结果 .....	39
表 4-1 种子样本选取数量表 .....	54





## 第1章 绪论

### 1.1 课题背景和研究意义

遥感航拍技术是 21 世纪国家重点发展的高新技术之一。遥感航拍图像是指卫星或飞机从空中视角通过光学或微波成像技术拍摄地球样貌获得的俯视图。随着高分辨率光学成像和微波遥感技术的发展，成像精度越来越高、覆盖范围越来越大，遥感航拍成为获取大面积地面信息的重要手段。目前，美国的 WorldView-2、GeoEye-1 和 QuickBird-2 卫星分别达到了 0.5 米、0.4 米、0.6 米的分辨率。我国在 2015 年正式投入使用的“高分二号”卫星使得我国卫星分辨率首次进入 1 米以内，标志着我国在卫星遥感领域进入亚米级时代。随着成像技术、卫星系统、航空航天侦查系统等技术的发展，未来必定会出现分辨率更高、重访周期更短、性能更高的卫星。这意味人类将获得更加丰富而庞大的遥感航拍图像数据，因此，如何充分将图像数据转高效地化为有价值的信息是一个非常值得关注的热点问题。

在计算机视觉领域，基于遥感图像的研究方向可以包括目标检测和识别、图像场景描述以及目标空间语义分析等。目标检测和识别技术一般是在遥感所摄范围内，确定目标的候选区域，从区域内有效的检测出目标位置甚至通过细节识别具体目标；图像场景描述通过找出图像中的特定区域，自动标注出区域类别，可以为进一步的图像理解提供有效的上下文信息；目标空间语义描述的是背景与目标间的空间位置关系和语义关联信息，能够更有效地对目标及背景间进行建模表达，使得系统做出更加智能和准确的判断。

近几年的国内外研究中，遥感图像目标检测是遥感图像处理 and 计算机视觉领域结合的热门研究方向。借助遥感图像所包含的内容信息的独特性、多样性、复杂性，目标检测技术在国防军事、交通建设、水利工程、农业林业、城市规划、灾害监测、生态研究等领域发挥重要作用。在国防军事领域，利用目标检测技术可以在区域内确定军机、坦克、船舶等军用设施的数量及具体位置，达到军情侦查和精准打击的目的；对于城市的航拍图像或视频监控中，可以利用检测技术确定建筑的范围和位置，车辆的分布及走向从而更好的进行城市规划和交通建设；针对湖泊、农林、农田等大面积区域的智能监测，借助遥感的同步性和时效性可

以迅速发现异常情况，降低汛灾、火灾、病虫害等带来的人员伤亡和财产损失。

“5.12”汶川特大地震发生后，遥感检测技术还应用于灾情监测和评估工作中，为抗震救灾提供有力的技术支撑。这都得益于遥感技术的不断精进带来的丰富信息和计算机视觉技术的发展。遥感图像目标检测如图 1-1 所示大致分为以下几个步骤：1) 卫星、航天飞机、无人机等拍摄的图像，经过成像、去噪、拼合等处理；2) 通过拍摄的地理位置信息、已有模型等信息可以确定出目标的大致区域；3) 通过算法和应用背景上的设计有效地从区域上检测出目标。



图 1-1 遥感图像目标检测大致流程

美国如南加州大学、康奈尔大学都在积极开展遥感图像理解技术研究。国内高校如中国科学院大学、西北工业大学、西安电子科技大学等也在做类似的研究。目前常用的遥感检测算法分两种：一种是传统的由下而上的数据驱动型，另一种是由上而下的知识型驱动。前者是飞机、工件等团块目标常用识别方法；后者则应用在复杂场景中对机场跑道、桥梁、道路的等目标的识别。大多数目标检测技术都是基于监督学习的方法，传统监督学习面对的是监督信息明确且完备的数据对象。训练过程中需要借助大量的标记数据来保证结果的准确性和泛化性。随着研究的不断深入，大量的标记数据成为了检测技术的阻碍和必须克服的挑战。现实应用场景中，检测任务往往面临的是不明确、模糊的监督信息对象。以遥感图像检测为例，航拍遥感影像技术很容易获取指数爆炸级数量的图像数据，而监督学习的标注工作是个令人头疼的问题。特别是针对遥感图像这种覆盖面积大、目标较小、场景较模糊的应用场景中，需要耗费大量的人力物力。在此情况下，弱监督学习渐渐成为近几年研究的突破方向。

弱监督学习是近几年逐渐形成的新型学习问题，其主要面对的是缺乏明确且完整的监督信息的数据对象，在不增加人工标注工作量的条件下，扩大学习样本集合，提高模型的泛化能力<sup>[1]</sup>。在现今大数据时代，获取未标注的样本代价极

小而标注样本的代价又极其昂贵，弱监督学习显然更加适用于当下大多数的机器学习应用场景和发展趋势。

综上所述，在当下的社会背景下，遥感图像的目标检测具有非常实用的研究价值。基于弱监督学习场景的遥感图像目标检测技术不仅是未来遥感图像目标检测研究的发展方向，同时也是一个颇具吸引力和挑战性的研究课题。

## 1.2 国内外研究进展

遥感检测技术研究中面临的挑战主要体现在以下几个方面<sup>[2]</sup>：1)目标类别的多样性。分辨率越来越高的遥感图像包含越丰富的地面信息，从而使背景和地物的结构和纹理等信息更加精细化，大大增加了可辨识目标的种类和数量。2)特征信息的可变性。遥感图像中，同一目标呈现出任意大小、形状、位置等多样化，甚至在光谱、纹理、轮廓等方面都有不同。3)复杂的干扰因素。遥感成像受光照、云雾、噪声等因素的干扰容易造成表现的异常，给分类、识别等工作增加难度。上述的几个特点无疑为遥感检测技术增加了些困难，但是挑战与机遇往往并存。遥感场景下的目标检测作为运动目标跟踪的基础研究结合着计算机视觉的突飞猛进在各国研究员的不懈努力下取得了许多成果。遥感图像根据获得的来源可分为 SAR 图像和光学图像，根据检测类型不同可分为异常检测和目标检测。本文主要针对光学遥感图像场景中的目标检测进行讨论。在众多遥感航拍图像下的目标检测研究中按照学习方法可分为监督学习、无监督学习、半监督学习和弱监督学习。

### 1.2.1 基于监督学习的研究

本小节按照语义层次可分为基于像素点级别的研究和基于区域级别的研究。

#### 1.2.1.1 像素级别的研究

在较早遥感航拍图像车辆检测的研究中<sup>[3],[4],[5]</sup>，将目标检测问题作为 3D 物体目标识别问题来解决。利用光照造成车辆的阴影来构建三维立体轮廓实现快速定位。Wang<sup>[6]</sup>中也采用阴影快速定位可能目标，代替传统扫窗减少了计算量。但是这种方法的问题很明显在于建筑的影子容易被误报，此外被遮盖的目标以及在建筑阴影下的目标则会被漏检。摒弃了阴影检测法，Tsai<sup>[7]</sup>利用车辆的颜色统计，对颜色进行分类，再利用边缘特征进行车辆假说验证。Cheng<sup>[8]</sup>的方法则是同时使用角特征和边缘特征，这两者可以有效提高命中率。对于图像先进行背

景颜色去除（有效去除 **False Positives**），提取特征后再用动态贝叶斯网络进行分类。为了利用车辆本身的属性进行快速定位，Choi<sup>[9]</sup>中使用聚类算法提取候选带有汽车形状对称属性的 blob，它利用到了汽车的对称属性。但是，Choi 的方法明显缺点是对带有对称性的建筑或公路标志容易误检。且其所使用的形状描述子由固定车辆模型获得，所以使得算法不够灵活。一旦镜头进行上下角度偏移，则模型产生错误很大。

上述的 Wang<sup>[6]</sup>、Tsai<sup>[7]</sup>和 Choi<sup>[9]</sup>属于基于像素级上进行操作，Cheng<sup>[8]</sup>是像素点及其邻域上的操作。这类型操作抽象度低，数据量大。主要是建立在基于像素级别的光谱信息分析基础上，通过提取图像的灰度、纹理和上下文特征等，实现特征向目标的转换。早期研究中遥感图像分辨率较低，目标较小，像素级别的操作可以克服这个问题。但是随着各项技术的发展，遥感图像也迈向了高分辨率行列。地物的结构和纹理等细节信息变得非常丰富，单一依赖于像素为单位的统计方式不足以准确提取地物的本质属性，而且单纯依靠光谱特征也无法反映图像中所包含的空间细节信息。

### 1.2.1.2 基于区域级别的研究

基于区域层级操作时，Kembhavi<sup>[10]</sup>，在特征提取时使用三种特征叠加。提取颜色概率、边缘方向的空间分布和结构属性，这利用到了汽车的颜色、轮廓等属性。这样得到的特征维数很高，计算复杂。Zheng<sup>[11]</sup>在预处理阶段使用了形态学获得候选车辆，然后使用神经网络进行分类。以上三种都是利用目标本身的属性来进行特征的提取和目标预定位。

伴随着图像分割技术的发展，Eikvil<sup>[12]</sup>提出一种两段式的策略用于遥感图像的车辆检测。第一阶段对可能的车辆、道路、植被进行区域分割。利用多频谱信息来识别植被区域和地理信息系统（GIS）来获得道路网。第二阶段使用面积、时刻等几何信息来构造区域分类算法。这种分割方式在遥感图像下特别是城市复杂场景下，存在许多直线结构会干扰分割的效果。

Zhu<sup>[13]</sup>的研究结合当下计算机视觉领域热门的 RCNN 框架，探究了卷积网络的特征和目标角度之间的关系，并提出采用基于分割的感兴趣区域提取算法以改进目标检测框架。

上述几种方法都是基于监督学习方法，思路与传统的目标检测无异。使用已标注数据提取有效特征从而训练出模型，再对待检测图像上进行适合的分类算

法。但是监督学习需要前期对测试数据进行标注，对于遥感图像这种覆盖面积大、单幅目标众多的场景，人工数据标注的代价十分巨大。所以，研究人员也在无监督和半监督上尝试寻求一些解决方法。

### 1.2.2 基于无监督和半监督学习的研究

Cheriyadat<sup>[14]</sup>提出了一种基于无监督在遥感图像中的场景分类方法。在特征提取阶段，采用了三种不同的特征提取策略。一种是简单的使用像素强度值作为特征；第二种是在各个像素点上测量方向过滤器的响应来构建基于过滤器能值的特征向量；最后一种是像素块对应的 SIFT 描述子。虽然实验结果显示 SIFT 特征的准确度相比其他两种能获得较高的准确度，但这仅仅是无监督上的区域划分，用来做检测还远远不够。

Yao<sup>[15-16]</sup>提出了使用半监督学习的遥感图像目标检测的理论基础。基于这个理论考虑了目标周围相对稳定的背景区域，提出了一种基于上下文的航拍图像目标检测方法。Zhong<sup>[17]</sup>提出使用半监督学习的方式，在 LiDAR 数据集上充分使用光谱、海拔、几何和强度等特征进行分类。上述两种研究虽然使用了半监督学习方法，但是其成果多是在理论方法上的研究，还不能满足实际应用的需求。

### 1.2.3 基于弱监督学习的研究

弱监督学习作为近几年计算机视觉热门的研究方向，在国内外已有部分研究学者将遥感图像目标检测问题结合弱监督学习场景进行研究。例如，Han<sup>[18]</sup>的研究使用 SIFT 作为低级特征，使用 LLC 对低级特征进行编码转为中级特征，再用深层玻尔兹曼机进行学习成高级特征子。训练过程基于贝叶斯框架，在弱监督学习基础上转为分类器的迭代优化过程。Zhang<sup>[19]</sup>的研究采用基于显著性和自适应的分割和反例挖掘方法初始化训练样本，在迭代过程中优化训练样本和分类器。Zhou<sup>[20]</sup>的研究采用 Caffe 框架从感兴趣区域中提取特征，迭代地选取最具区分性分样本作为训练样本。同时，采用 Bootstrapping 算法选取最容易误化分类器的样本反例来优化分类器。

目前，遥感图像目标检测在弱监督场景下的研究还比较少。在计算机视觉的目标检测领域中，弱监督学习是未来研究的发展趋势。其以极弱的标注数据和容易获取的海量未标注样本为输入更加适用于当下信息时代的应用场景，更符合人工智能和模式识别的初衷，具有广阔的发展前景。

### 1.3 本文研究内容

本课研究的目的是在弱监督学习场景下即使用极弱标注信息数据的前提下完成在遥感航拍图像的特定目标检测工作，既要达到模型的准确率同时保证模型的泛化性。具体来说，在光学遥感图像数据集中只需极弱的图像级别的标注（标注对象为图像而非样本）或辅以极少量样本级别标注，达到接近监督学习方法的目标检测性能。与常见的弱监督学习研究类似，本课题的主要任务是准确地挖掘富有信息价值的训练样本，迭代地优化训练样本和分类器。为了实现本课题研究的目标，本课题采用多示例学习和协同训练的结合，同时加入训练样本集优化的步骤保证挖掘出样本的准确性和多样性。本文研究整体框架如图 1-2 所示：

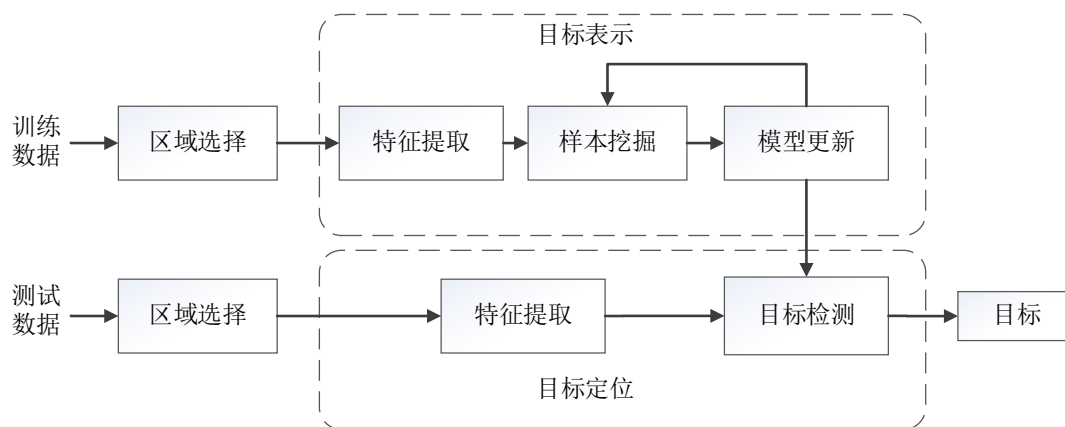


图 1-2 本文研究整体框架

1) 结合遥感图像的特点，本研究采用基于分割的算法提取感兴趣区域并用卷积神经网络对区域进行特征表示。提出一种基于 K-means 聚类以及反例筛选的初始化方法，有效选取初始的正反例训练样本。

2) 针对样本挖掘策略，本文提出以多示例学习为主的方法框架。将正反例遥感图像数据转化成多示例学习中的“正反包”概念，在迭代中从包内选择最优样本，扩充到当前训练集中，从而达到挖掘样本的目的。

3) 除了多示例学习，本课题研究将协同训练方法与之相结合。协同训练方法把数据从不同视图进行描述训练分类器并互相进行样本选择，以保证样本挖掘的准确性和多样性。

4) 为了进一步保证挖掘出的样本准确性，避免误化分类器导致迭代向错误方向收敛，本文提出一种简洁而有效的方法进行样本优化（sample refine）。该方

法能快速有效地从候选训练样本中删除噪声样本，提高正例训练样本的准确度。本课题研究在遥感图像数据集中的飞机、车辆数据集进行了实验，实验结果表明本文的算法框架能有效挖掘样本，并使模型达到接近监督学习性能的效果。

### 1.4 本文的组织结构

第一章，绪论。本章介绍了遥感图像目标检测的研究背景和研究价值。系统的总结了当前国内外在此方面的研究，同时比较了现有方法研究的优缺点。对本文基于多示例学习的弱监督遥感图像目标检测算法框架进行了简单的阐述和说明。

第二章，相关技术研究。本章是承上启下的章节，主要包括对弱监督学习、区域提取、特征选择、多示例学习、协同训练等概念进行介绍，同时以分类的形式列举现有的方法。

第三章，基于多示例学习的弱监督遥感图像目标检测。本章主要介绍本课题研究框架以多示例学习和协同训练算法的结合为主要核心，样本优化、样本初始化等步骤。本章还展示了框架各细节的实验，证明其设计的有效性。同时，通过与监督学习方法进行比较，证明了本框架在弱标注数据基础上仍能完成分类器学习任务。

第四章，基于样本种子点的弱监督遥感图像目标检测。本章以第三章框架为基础的，介绍了引入种子点样本的弱监督遥感图像目标检测框架。其目的在于解决遥感图像中更为复杂的场景中初始化和样本挖掘的问题。本章重点简述了基于种子点样本的初始化方法和难例挖掘方法，并通过对比实验和结果分析验证了环节设计和整体框架的有效性。

第五章，结论和展望。本章主要对本文的研究工作进行总结，探讨尚未解决的问题以及对下一步工作的开展进行展望。





## 第2章 相关技术研究

前一章简要介绍了本文的问题背景、研究意义、当前相关领域的研究综述、主要研究内容和结构安排。本章将系统的讲述本课题研究所涉及的相关算法和技术研究，作为第三章和第四章研究工作的基础知识。

### 2.1 弱监督学习

本节对本框架的核心弱监督学习进行介绍，包括弱监督学习的基础概念、弱监督学习的类别、弱监督学习的常见算法。

#### 2.1.1 弱监督学习的概念

监督学习在机器学习中是主流方向，能够有效地利用“经验”来系统地训练和改善机器性能。但其显著缺点在于需要借助足够多的有标记训练样本来提高模型泛化能力。然而，在大数据时代，随着存储技术和数据收集技术的发展，每个时刻都有难以量化的数据爆炸式的产生，收集大量无标记样本已是格外容易的事，而获得专家标注的样本却需要耗费大量人力物力。随着研究者们不断的探索和现实场景应用的发展，近年来不断涌现出新的学习问题。它们共同的特点都是面临缺乏明确和完整的监督信息的问题，弱监督学习在这种情况下应运而生。

弱监督学习属于近几年提出的新概念，国内外研究还进行得比较少而且没有统一的定义。按照文章<sup>[1]</sup>的总结，弱监督学习是一系列针对所有弱标定（即监督信息不完整或不明确）对象问题的统称。从实质上来讲，弱监督学习是半监督学习的一种发展和延伸，将学习问题扩展到样本的标注信息更为复杂的情形中。弱监督学习在信息检索、文本分类、生物信息、自然语言处理、互联网应用等领域已经有广泛的应用。从实际不同应用场景来讲，弱监督学习可以细分到很多类别，本文将在下一节进行阐述。

#### 2.1.2 弱监督学习分类

本文在此参照文章<sup>[1]</sup>的思路，将弱监督学习问题划分为以下几类。

##### 2.1.2.1 半监督学习

半监督学习(Semi-Supervised Learning)在机器学习领域存在已久，其主要是指样本集中存在少量已标注样本和大量未标注样本的学习问题。在此情形下，让

模型不依赖于外界交互，自动地利用未标注样本来提升学习性能。半监督学习在现实生活中的应用场景很多。例如：在网络购物时，用户会对自己感兴趣的商品进行收藏(判为正例)。但就海量商品而言，用户浏览过的商品仅仅占一小部分，大部分是未被浏览过的（即未标注）。系统如果能够自动地根据这些小部分标注数据和大部分未标注数据进行学习，便可以向用户推荐哪些未浏览过的商品可能是用户感兴趣的。

事实上，未标注样本虽然不包含标注信息，但学习前提是与标记样本从同样数据源而来并服从同样的分布。这就涉及半监督学习问题中针对未标记样本的数据分布信息与类别标记相联系的两个基本假设。

### 1) 聚类假设

聚类假设（Cluster Assumption）是最常见的假设。它是指假设样本间存在聚类(cluster)结构，即同一聚类的样本属于同一类别。按照该假设，决策边界应该尽量通过样本较为稀疏的地方，而避免将稠密的聚类划分到分界面两侧。

### 2) 流形假设

另一种常见假设是流形假设（Manifold Assumption）。它是指当样本分布在同一流形结构上，相邻近的样本应该性质相似从而拥有相似的输出值。

流形假设可以说是聚类假设的推广，前者注重局部特性，后者更加注重整体特性，但二者的实质都是相似的样本有相似的输出。

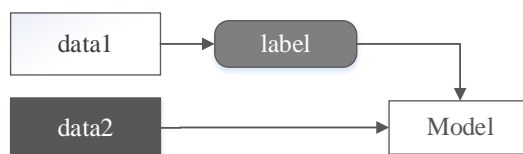


图 2-1 半监督学习示意图

## 2.1.2.2 多示例学习

多示例学习(Multiple-Instance Learning)是以“包”(Bag)的概念来看待数据。数据集中每个数据表示为一个包，每个包内由若干个示例(Instance)组成。多示例学习的特点在于只有包级别上的标定，而没有示例级别上的标定。标注的规则为：如果包内至少存在一个正示例，则该包为“正包”。如果包内不存在任一正示例，则为“反包”。在传统机器学习定义中，样本即对应多示例学习中的“示例”。但在多示例学习中，样本为包，一个包内可以含有多个示例。多示例学习的目的是

使用包级别标定的数据训练出一个分类器能对未知的包(Unseen Bag)进行正确分类。多示例学习是弱监督学习中的典型问题,在理论、算法以及应用方面已硕果累累。在本课题研究中,多示例学习是核心关键算法技术。本文将在下一节对多示例学习进行详述。

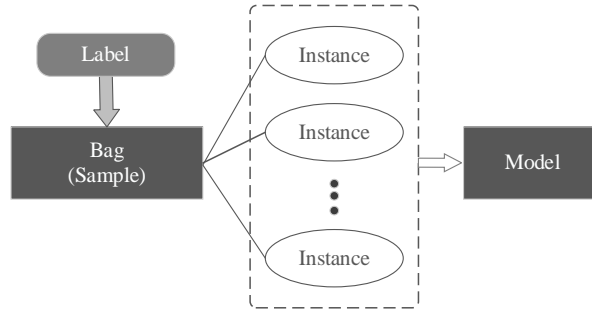


图 2-2 多示例学习示意图

### 2.1.2.3 多标记学习

多标记学习(Multi-label Learning)又称多标签学习,其考虑的是一个对象与多个类别标记相关联的情况。在传统的机器学习框架中,样本一般都与明确的、单一的类别相关联。但是在现实场景中,该假设往往都不能满足。例如,《福尔摩斯探案集》可以被标上“小说”、“推理”、“悬疑”、“英国”等标签。多标记学习是为了满足生活中对象的多义性和歧义性而产生的。如今,多标记学习已在场景分类、文本分类、生物信息学等领域得到了广泛使用。

从样本对应标记的角度看,传统的单标签问题(包括二类或多类问题)是多标记学习的特例。所以在解决多标记学习时,经常直观地将它转换为多个独立的二类问题求解。但这样往往容易忽略各标签之间的关联性,而使准确性和泛化性降低。从样本歧义性来讲,多标记学习与多示例学习都属于处理对象歧义性的学习框架<sup>[21]</sup>。前者关注的是输出空间,即样本概念的歧义性。后者从输入空间关注样本内容的歧义性。

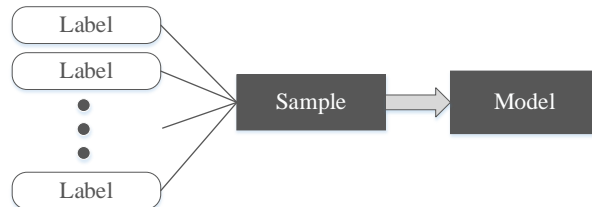


图 2-3 多标记学习流程

#### 2.1.2.4 其他学习问题

模糊标记学习，是指当对样本的标注并非完全正确的监督信息，而只是表示样本属于某类别的概率。传统机器学习框架中，不论样本属于单标记或多标记，标注的监督信息属于完全正确、真实可靠的。而模糊标记中，由于数据来源或标注时的问题，并不能保证标注的完全正确，只能包含属于某类别的可能性。

噪声标记学习，是指在标记过程或数据收集过程中，存在噪声标签混入数据中。标记噪声与属性噪声同属数据噪声，但后者指的是样本属性上的错误，容易造成离群点的问题。而标记噪声对模型学习的影响更大，容易造成决策算法的分类边界变形。

实际场景中，网络获取标定数据很容易造成上述两种问题。例如，大众标注，很容易混入恶意标注或标注用户对信息理解的不确定都可能造成标注错误或标注模糊。而通过计算机自动标注技术也容易在数据收集过程中遇到数据缺失、数据不明确等问题。

#### 2.1.3 弱监督学习的常用算法

目前基于弱监督学习的常用算法以出现已久的半监督学习方法居多，多示例学习在近几年也逐渐成为弱监督学习中的典型代表。相较之下，其他学习问题的研究还相对较少。本文在此总结几类针对弱监督学习的常用算法。

##### 2.1.3.1 自学习方法

自学习(self-training)<sup>[22]</sup>方法是最简单的半监督分类模型。它表示的是使用已有的标注样本学习一个分类器，然后分类器对未标注样本进行分类。再使用分类后的标注与已有标注样本进行学习，所以通常也称为 self-teaching 或 bootstrapping。自学习方法的基本假设是对于预测的结果，至少置信度最高的样本是倾向于正确的。如果该假设不成立，那么自学习挑选出的样本将会严重误化模型的学习。

表 2-1 自学习算法

输入：标注样本 $L = \{(x_i, y_i)\}_i^l$ ，未标注样本 $U = \{x_j\}_{j=l+1}^{l+u}$
重复： 1. 从 $L$ 中训练分类器 $f$ ； 2. 使用 $f$ 对 $U$ 进行分类； 3. 从 $U$ 移除子集 $S$ ，将 $\{(x, f(x))   x \in S\}$ 加入到 $L$ 中

### 2.1.3.2 生成式方法

生成式方法(generative methods)是基于生成式模型的方法。此类方法通常假设所有数据都由同一个潜在模型生成的。一般的做法是将无标记样本的类别信息当作模型的缺省参数，通过假设可以用潜在模型参数将数据与模型联系起来。通常是使用 EM 算法进行极大似然求解。

EM(Expectation Maximization)算法是机器学习中的经典算法。它是一种迭代算法，用于含有隐变量(Hidden Variable)的概率参数模型的最大似然估计或极大后延概率估计。EM 算法在生成式中的算法如下：

表 2-2 EM 算法

输入：已有数据 $\mathcal{D}$ ，隐藏数据 $H$ ，初始化参数 $\theta^{(0)}$
初始化： $t=0$
重复以下步骤直至 $\rho(\mathcal{D} \theta^{(t)})$ 收敛：
1. E 步：计算 $q^{(t)}(H) \equiv \rho(H \mathcal{D}, \theta^{(t)})$ ；
2. M 步：找出能使 $\sum_H q^{(t)}(H) \log \rho(\mathcal{D}, H \theta^{(t+1)})$ 最大化的参数 $\theta^{(t+1)}$ ；
3. $t=t+1$
输出： $\theta^{(t)}$

### 2.1.3.3 半监督 SVM

半监督 SVM (Semi-Supervised Support Vector Machine, 简称 S3VM) 是机器学习领域经典算法支持向量机在半监督学习领域的扩展。传统的支持向量机算法是在已知标注的样本中寻找最大间隔超平面。但是在半监督学习中，大部分样本是没有标记信息的。S3VM 的目标是寻找能将已有标记样本分开并且穿过数据稀疏区域的超平面 (如图 2-4)，这也是聚类假设与支持向量机结合的推广。直推式支持向量机 (Transductive Support Vector Machine, TSVM) [23] 是 S3VM 中典型的算法。TSVM 采用局部搜索来迭代求解。迭代过程首先利用有标记样本学习一个 SVM，然后用它赋予未标记样本“伪标记”，求解标准 SVM 间隔最大化平面。关于 TSVM 的详细求解过程可参考文章 [24]。

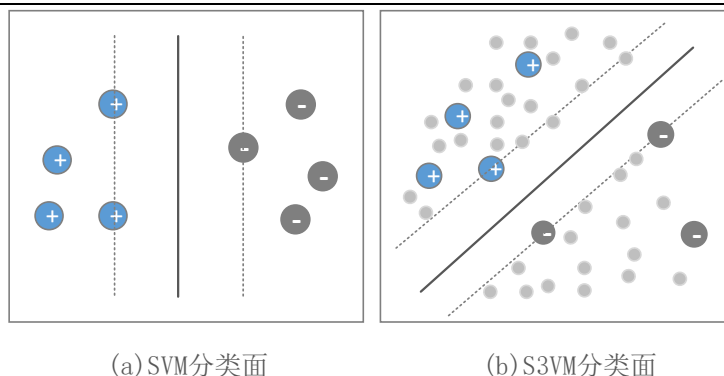


图 2-4 传统 SVM 与 S3VM 分类面区别

### 2.1.3.4 基于图的半监督方法

基于图的半监督方法是一种以流形假设为基础的方法即相似性越大的样本，对应的标识也相似。图的方法是将每个训练样本表示为一个节点，节点间的边表示为其两端节点间的相似度。常用的图构建方法有：

1)全连接图；对于图内的每两个节点都有对应相连的边。边的权重随节点间的欧氏距离增加而减少。但是，欧式距离并不是唯一用来衡量节点间相似度的方法，常用的也有使用 RBF(Radial Basis Function)核方法。

2)kNN 图；该方法令每个节点按照欧氏距离连接其最近的 k 个节点。当图内一条边存在表示其中一端节点是另一端节点的 k 近邻，而不需要两端节点互为近邻。因此，每个节点可能存在超过 k 条边。同时，对于节点密集的图，k 可设置较小；当图节点较稀疏时，可以选择较大的 k。

3) $\epsilon$ NN 图； $\epsilon$ 为选择的阈值。当两个节点(样本)间的距离小于设定的阈值时，将两节点连接。

以上是常用的图构建方法。对于图的标识函数必须满足以下两个要求：①函数对于已标注样本的标识结果应该与该样本的标注相近。该要求可以用损失函数表示；②函数对于整个图是平滑(Smooth)的，即局部变化是缓慢的。该要求使用基于图的正则项表示。上述两个要求组成了半监督学习的图方法的基本框架。典型的方法有：最小割(Mincut)、调和函数(Harmonic Function)、流形正则化(Manifold Regularization)等<sup>[25]</sup>。

上述几种方法主要是基于半监督学习。一方面是由于半监督学习发展时间最长，存在方法最多，而其他相应的学习问题研究较少。另一方面是由于其余学

习问题的解决方法多是在上述几种方法上进行扩展和变形。有关弱监督学习的研究还比较少，相应的解决方法还局限于特定的场景问题中，缺乏统一且具有普适性的方法。

## 2.2 区域选择与特征提取

图像感兴趣区域提取是图像处理中的重要步骤，是一种低层次的基础技术。其主要思想是将图像以某种规则分割成不同的区域。一般来说，提取的区域内部满足一定的规则一致性，最简单的如灰度值相近或纹理特征的一致性。被提取的区域称为感兴趣区域（ROI）原因是区域提取之后的操作将会以这些区域为单位而非在原图像上进行操作。作为一项基础性的操作，区域提取在计算机视觉框架是关键步骤之一，区域提取的质量将直接影响到诸如特征提取、目标检测等步骤的结果。

特征是一个物体区分与其他物体的特点。在计算机视觉领域，图像处理中常见的特征包括底层特征和高层特征。底层特征更简洁直观，所含的信息量较少。常见的底层特征有颜色、边缘、角点、纹理、形状、空间关系等。底层特征包含的信息最直观但也较少。高层特征一般需要人为设计，所含信息量较复杂。目前最热门的研究是手工特征，例如 HOG 特征、SIFT 特征、LBP 特征、Haar 特征等。这些特征都是研究者经过手工设计提取的，并且在各自研究领域中发挥了重要作用。本节将简述研究中所使用的区域选择和特征提取的方法。

### 2.2.1 Selective Search

选择性搜索（Selective Search）是 J.R.R. Uijlings<sup>[26]</sup>在 2012 年提出的。与传统扫窗法不同，Selective Search 考虑了物体的形状、大小、颜色、纹理特征等信息，以图像分割为基础，结合图像中物体布局的层次(Hierarchical)关系，有效地对图像进行区域提取。

表 2-3 Selective Search 算法

输入：彩色图像
输出：分割后区域位置
使用快速图割的方法获取原始分割区域 $R = \{r_1, r_2, \dots, r_n\}$ ;
初始化相似度集合 $S = \emptyset$ ;
计算两两相邻区域之间的相似度，将其添加到相似度集合 $S$ 中；
当 $S \neq \emptyset$ 时：
从相似度集合 $S$ 中找出相似度最大的两个区域 $r_i, r_j$ ，将其合并成一个区域。

从  $S$  移除  $r_i, r_j$  对应的所有相似度。计算两者合并区域与其邻域的相似度，并更新到  $S$  中。

从  $R$  中移除  $r_i, r_j$ ，加入二者合并的新区域；  
获取  $R$  中每个区域的 Bounding Boxes 及其对应位置

### 2.2.2 深度学习

深度学习(deep learning)的概念是 Hinton 等人在 2006 年提出的<sup>[27]</sup>，来源于神经网络的研究。它是一种含有多隐层的多层感知器结构，通过组合底层特征形成更加抽象的高层特征。

深度学习的提出得益于神经科学、生物学等学科的发展。如图 2-5 所示，在人类视觉系统研究中，人的视觉系统是分级处理信息的。从低级区域提取最基本特征，较高级特征是前一层特征的组合形式。越高级特征抽象度越高，所含的语义信息越多。

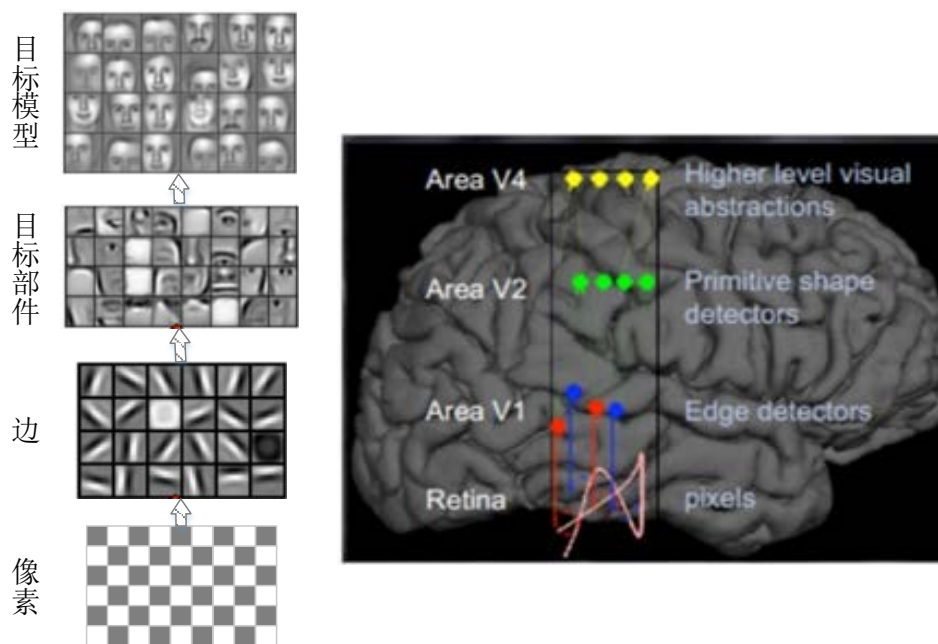


图 2-5 深度学习特征与人脑层级<sup>[28]</sup>

深度学习的基本思想是在输入层和输出层之间设置多层的隐藏层，通过训练样本在尽可能减少输入输出之间的误差过程中，学习模型参数。近几年，深度学习在人工智能众多方法中脱颖而出，在各个研究领域都获得极大的关注。

Hinton 发表的文章<sup>[27]</sup>描述深度学习的实质是通过构建含有多层隐藏层的模型学习和海量的训练数据来学习有用的特征，从而提升分类和预测的准确性。因



此，可以总结为“深度学习”是手段，“学习特征”是过程，“提升性能”是目的<sup>[28]</sup>。与浅层学习方法相比，深度学习注重于特征的学习过程，利用大量的数据来学习特征，更能够刻画数据的本质信息。另一方面，侧重于模型结构的深度，以更深的网络模型结构来达到更深层次特征的目的。其中，卷积神经网络是深度学习中最热门的研究之一。

卷积神经网络（Convolutional Neural Network, CNN）是多层感知机（Multi-layer Perceptron, MLP）的一种变型。Hubel 和 Wiesel<sup>[29]</sup>研究猫脑皮层中的视觉神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，继而提出了卷积神经网络中的“感受野”概念。K. Fukushima<sup>[30]</sup>在 1980 年提出的神经认知机是卷积神经网络第一次被实现。时至今日，卷积神经网络已在众多研究者的辛勤工作下获得重大发展，比如 LeNet-5<sup>[31]</sup>、AlexNet<sup>[32]</sup>、VGG-Net 等，已成为当前语音分析和图像识别领域的研究热点。

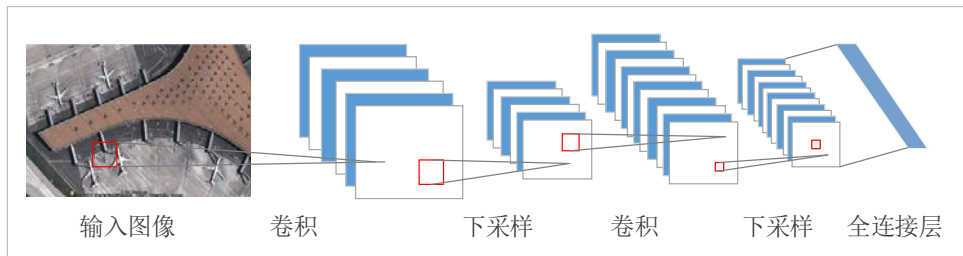


图 2-6 卷积神经网络概念示范

如图 2-6 所示，卷积神经网络的基本结构包括两层：特征提取层和特征映射层。在特征映射层即卷积层中，可训滤波器和可加偏置对前一层输入进行卷积，产生对应的特征映射图。卷积即为特征提取步骤，每个神经元与前一层的局部区域相连并提取该局部的特征，此时该局部被提取的特征与其他特征的位置关系也确定下来。特征映射层将前一层的局部进行下采样，降低计算复杂度，通过激活函数使得特征映射具有位移不变性。

#### 1) 稀疏连接

卷积神经网络通过相邻层之间节点的局部连接来挖掘图像中的空间局部关联信息。每一层隐藏层的节点由前一层节点的局部子集作为输入。这部分节点在前一层中是相邻的（空间局部连续）。如图 2-7 所示，第  $m+1$  层的神经元节点由第  $m$  层 3 个相邻节点作为输入，第  $m$  层由  $m-1$  层的 3 个相邻节点作为输入。所

以，对应到原输入中，第  $m+1$  层有宽度为 5 的非线性特征进行编码<sup>[33]</sup>。

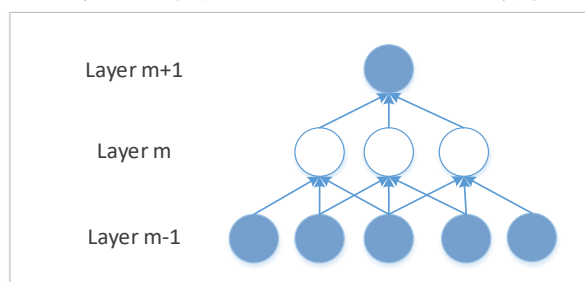


图 2-7 稀疏连接示意图

通过局部感知，神经元能够抽取一些基本的视觉特征，比如边缘、断点、拐角等，这些初级特征被更高层的神经元所组合从而获得更高级的特征。局部感受客观上减少了需要训练的权值数目。

### 2) 权值共享

卷积神经网络设置分布于图不同位置的神经单元共享相同的参数，包括相同的权值矩阵和偏置向量。这些每层中具有相同权值的神经单元组成了二维矩阵称为特征图 (Feature Map)。一个图中的不同位置的局部区域使用同样的核进行卷积，一个卷积层包含多个具有不同权值向量的特征图，使得同一位置的数据能获得多种不同的特征，同时降低大幅减少需要训练的权值数目。

### 3) 下采样

下采样模拟初级视觉皮层中的复杂细胞，在空间上对相邻连续子区域内取值，目的是降低分辨率使得卷积神经网络对输入的局部变化具有一定的不变性。

卷积神经网络出了具有旋转、平移、缩放不变性外，还具有训练参数少、速度快的特点。可以直接输入原始图像，建立逐层复杂的特征提取过程。基于上述特点，本课题采用卷积神经网络进行特征提取。

## 2.3 多示例学习

多示例学习 (Multiple Instance Learning, MIL) 问题最早是由 Dietterich 等人于上世纪 90 年代研究药物分子是否为麝香分子时提出的<sup>[34]</sup>。该研究的背景是有效的药物分子存在许多种的同分异构体，研究者们尝试从众多同分异构体分子中判断哪种是有效的。这个问题抽象成了多示例学习的概念。

### 2.3.1 多示例学习的概念

多示例学习是一种归纳式机器学习的变体。在此类学习中，训练集由若干个具有标记的“包”组成，每个包内含有若干个“示例”代替了原本单一的特征向量。在传统机器学习定义中，样本即对应多示例学习中的“示例”。但在多示例学习中，样本为包，一个包内可以含有多个示例。训练集中仅有包级别的标注(Bag-level Label)而没有示例的标注(Instance-level Label)。对于标签的定义为：当包内含有至少一个正示例时，定义为“正包”；当包内全是反示例时，定义为“反包”。这表明，在正包中可以确定包内含有至少一个正示例，但是哪个示例是正示例是未知的。多示例学习的目的在于对只有包级别标定的训练集进行学习，学习出的分类器能够对新来的包进行正确分类。

### 2.3.2 多示例学习的分类

多示例学习适用于很多真实世界的学习场景，包括药物活性预测、图像分类等。如果将最早在药物活性实验中提出的多示例学习假设作为“多示例学习标准假设”的话，近几年有许多多示例学习上的研究已经不再严格遵从于标准假设了<sup>[35]</sup>。本文将在此将从这个角度进行分类。

#### 2.3.2.1 标准多示例学习假设

多示例学习标准假设(Standard MI Assumption)表示每个示例存在一个隐藏的类别标签。一个包被判为正例当且仅当其存在至少一个示例是正例。因此，包级别的类别标签由示例的类别标签决定。

如果用 $X = \{x_1, x_2, \dots, x_n\}$ 表示一个包内含有  $n$  个示例， $g(x)$ 表示示例  $x$  对应的类别标签， $f(X)$ 表示包的类别标签。则标准多示例假设可以表示为

$$f(X) = (g(x_1) \cup g(x_2) \cup \dots \cup g(x_n)) \quad (2-1)$$

#### 2.3.2.2 广义多示例学习假设

Weidmann<sup>[36]</sup>等人在 2003 年提出一种基于示例的广义多示例学习的层级概念，具体可由三种广义多示例学习概念构成。

##### 1) 基于出现的多示例学习

在基于出现的多示例学习(Presence-based MI Assumption)中，一个包被定为正包当且仅当包内存在一个或多个示例属于所需类别的示例集合。用公式表示为：

$$v_{PB}(X) \Leftrightarrow \forall c \in C: \Delta(X, c) \geq 1 \quad (2-2)$$

其中， $X$  表示示例空间， $C$  表示所有类别的集。 $\Delta$ 表示包内一个类别出现的次数。标准多示例假说是基于出现的多示例学习的特例即其只有一个类别。

### 2) 基于阈值的多示例学习

基于阈值的多示例学习 (Threshold-based MI Assumption) 表示一个包为正包当且仅当满足包内属于类别空间中的每个类的示例个数都超过指定的阈值。这就要求对于类别空间的每个类，包中的示例不仅要出现，而且个数要超过设定的阈值。其中，不同类别指定的阈值可以不同。基于阈值的多示例学习表示为：

$$v_{TB}(X) \Leftrightarrow \forall c_i \in C: \Delta(X, c_i) \geq t_i \quad (2-3)$$

其中， $t_i$ 表示类别 $c_i$ 设定的阈值。 $c_i$ 对应的是类别空间中的每一个类。

### 3) 基于计数的多示例学习

基于计数的多示例学习 (Count-based MI Assumption) 表示包被定为正包当且仅当包内属于类别空间中的每个类的示例个数都在设定的阈值范围内。这不仅要求包内属于每个类别的示例个数超过设定的最小值，还不能超过设定的最大值。其中，每个类别都可以设置不同的最大最小值范围。基于计数的多示例学习表示为：

$$v_{CB}(X) \Leftrightarrow \forall c_i \in C: t_i \leq \Delta(X, c_i) \leq z_i \quad (2-4)$$

其中， $t_i$ ,  $z_i$ 分别表示类别对应的示例个数的最小值和最大值。

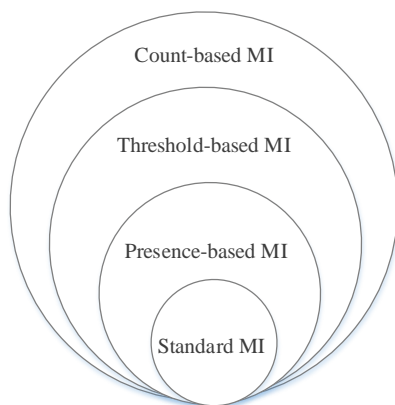


图 2-8 Weidmann 的广义多实例学习层级结构图<sup>[36]</sup>

上述三种假说组成了 Weidmann 的广义多示例的层级结构。如图 2-8 所示，广义多示例学习从假说的要求上满足这样的层级结构： $\text{Standard MI} \subset$

Presence based MI  $\subset$  Threshold based MI  $\subset$  Count based MI。按照这个结构，针对较广义的 MI 问题设计的算法同样能在比它狭义的问题上有出色的发挥。

### 2.3.2.3 GMIL

Scott 等人在 2005 年也提出一种新广义多示例假说被称为 GMIL (Generalized Multi-instance Learning)。在这个模型中存在一个目标点 (Target Points) 的集合和一个排斥点 (Repulsion Points) 集合。一个包被定为正包当且仅当存在一定数量示例足够接近目标点且不超过一定数量示例接近排斥点。在 GMIL 模型中，目标点和排斥点各自的集合被视为“理想包”。一个正例包与理想正包的距离应该在阈值  $\gamma$  内，与理想反包的距离应该在  $\gamma'$  外。两个包间点的距离使用排序式 Hausdorff 度量方式来计算。两个包 P、Q 之间的 Hausdorff 距离被定义为 P 内任一点与其在 Q 包内的最近点的距离中的最大值。

### 2.3.2.4 最近邻假说

在传统的单例学习中，kNN 算法是一种经典的分类方法。待标记样本的类别根据其 k 个最近邻的训练样本的类别决定。“最近”的度量方式通常是欧式距离。在多示例学习中，明显不能用于直接计算包之间的距离来决定包的标签。Wang<sup>[37]</sup>等人对传统 KNN 算法做了改进，其采用修正后的 Hausdorff 距离来计算包之间的距离从而选定包的近邻。针对分类方法的不同，提出了两种不同的 K 近邻方法。第一种称为贝叶斯 K 近邻算法 (Bayesian-kNN)，用贝叶斯方法在近邻中预测最可能的标签类别。第二种方法称为引用-K 近邻 (Citation-kNN)，它是一种基于信息科学领域的参考 (Reference) 和引用 (Citer) 概念的方法。在做分类决定时，不仅是参考样本的近邻样本 (Reference) 同时也考虑包的近邻包 (Citer)。

### 2.3.2.5 多示例回归

多示例回归 (Multi-instance Regression) 的概念在 2001 年时被提出<sup>[38]</sup>。它主张为包关联实数标签而不是常用的二分类标签。与多示例分类问题一样，多示例回归也来源于药物活性预测问题。研究者使用药物活性等级而非是否有效作为标签。最早提出多示例回归概念的 Ray 和 Page 认为数据是由带有高斯噪声的线性模型产生的。他们进一步认为包的实数标签应该由包内的某一示例 (Primary Instance) 的实数标签表示。

上述的几种多示例学习方法只是当前多示例学习研究的一部分。一方面，它们大多依赖于自身发现的单一场景。另一方面，正是由于多示例学习更加贴合实际场景的需要，而在多年的研究中不停地获得演变和发展。

### 2.3.3 基于支持向量机的多示例学习

Andrews 等人<sup>[39]</sup>最早将最大化超平面间隔的思想应用到多示例学习中。他指出多示例学习中最大的挑战在于解决正包中正例模式学习的模糊性并提出了两种基于支持向量机的多示例学习方法：基于包的多示例学习方法 MI-SVM 和基于样本的多示例学习方法 mi-SVM。

#### 2.3.3.1 mi-SVM

mi-SVM 是基于样本的多示例学习方法。它将示例标签作为未观察整数变量，约束条件为包标签。其目标在于最大化示例间隔。mi-SVM 的目标函数可以如下表示：

$$\min_{\{y_i\}} \min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (2-5)$$

$$\text{s.t. } \forall_i: y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \xi_i \geq 0, y_i \in \{-1, 1\} \quad (2-6)$$

按照多示例学习的假设，如果示例  $x_i$  不属于反包，那么其对应标签  $y_i$  便是未知的整数变量。mi-SVM 的策略是在初始时将包标签赋予包内示例，使用带有初始类别标签的示例训练分类器，再使用分类器对包内示例分类。迭代地从每个包内挑选示例同时更新标签并优化分类器，直至示例标签不再更新。

表 2-4 mi-SVM 算法

初始化：对于每个包内的示例标签 $y_i = Y_i$
重复：
使用数据集训练分类器 $(w, b)$
计算分类器在每个正包样本 $x_i$ 内的输出 $f_i = \langle w, x_i \rangle + b$
对正包内的每个样本设定 $y_i = \text{sgn}(f_i)$
对于每个正包：
if $(\sum_{i \in I} (1 + y_i) / 2 == 0)$
计算并找出 $i^* = \arg \max_{i \in I} f_i$
设置 $y_{i^*} = 1$
直到样本标签不再更新
输出： $(w, b)$

#### 2.3.3.2 MI-SVM

MI-SVM 是基于包的多示例学习方法，它的目标是最大化包的间隔。以包为

中心的方法中，每个正包中只有一个示例可以决定包的间隔。一旦这些“witness”示例被选中，则包内其他示例对包的间隔是不起作用的。按照这个概念，可以将 MI-SVM 定义为：

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_I \xi_I \quad (2-7)$$

$$\text{s.t. } \forall I: Y_I \max_{i \in I} (\langle w, x_i \rangle + b) \geq 1 - \xi_I, \xi_I \geq 0 \quad (2-8)$$

MI-SVM 的策略就在于从正包中选择那些“witness”的示例与反包中的示例共同训练分类器。在迭代过程中，在正包中选择最“most positive”的示例作为正示例与反包中的所有反例示例共同训练分类器，分类器在下次迭代中再次选择正示例。迭代过程在优化训练样本和优化分类器的过程中交替。MI-SVM 算法如表 2-5 所示。

表 2-5 MI-SVM 算法

初始化：对于每个正包 $B_I: x_I = \sum_{i \in I} x_i /  I $
重复：
使用反包示例和正示例集 $\{x_i: Y_i = 1\}$ 训练分类器 $(w, b)$
计算分类器在每个正包样本 $x_i$ 内的输出 $f_i = \langle w, x_i \rangle + b$
设定 $x_I = x_{s(I)}$ ：从每个正包中选择样本 $s(I) = \arg \max_{i \in I} f_i$
直到所选 $s(I)$ 不再更新
输出： $(w, b)$

## 2.4 协同训练

协同训练(Co-training) 由 A. Blum 和 T. Mitchell<sup>[40]</sup>在 1998 年提出，是半监督学习中的一种经典方法。A. Blum 和 T. Mitchell 在网页分类问题研究中发现可以直观地根据网页的内容进行分类，同时也可以根据链接到该网页的超链接所含的信息对该网页进行分类。网页内容和超链接可以视为该网页样本的两种独立视图，这便是样本的冗余视图概念。

### 2.4.1 协同训练的概念

协同训练算法假设数据集有两个充分冗余视图 (View)。这两个冗余视图满足以下两个条件：

- 1) 两个视图足以描述问题场景即任一视图上，数据都可以训练出强分类器。
- 2) 两个视图互相满足条件独立。

在 A. Blum 和 T. Mitchell 的网页分类问题研究中，网页内容和超链接各自属

性集上都可以训练出强分类器且两种属性集符合独立条件。标准协同训练在两种视图上利用已标记样本各自训练出一个分类器。在迭代过程中，每个分类器从未标记样本中标记若干个置信度高的样本加入对方分类器的训练集中。协同训练过程不断迭代直到训练集内不含未标记样本。标准协同训练算法如下：

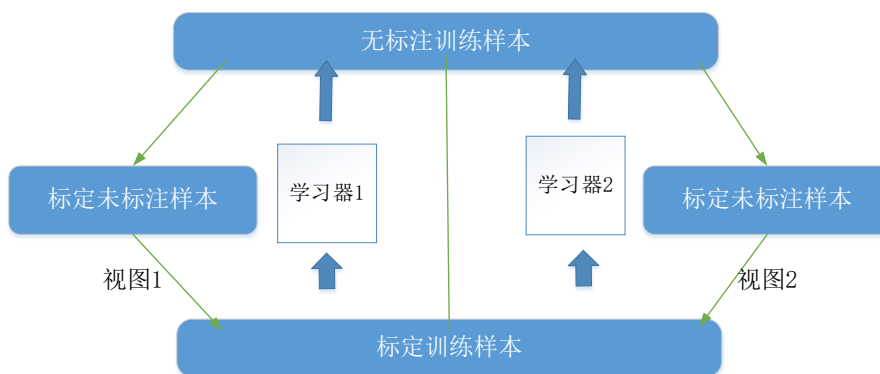


图 2-9 标准协同训练流程图

上图为标准协同训练的流程图，具体算法参见表 2-6：

表 2-6 标准协同训练算法

输入：带标记训练集 $L$ ，未标记训练集 $U$ 从 $U$ 中随机选出 $u$ 个样本组成 $U'$ 循环 $k$ 次迭代： 在样本 $x$ 的 $x_1$ 视图中，使用 $L$ 集内样本训练分类器 $h_1$ 在样本 $x$ 的 $x_2$ 视图中，使用 $L$ 集内样本训练分类器 $h_2$ 使用 $h_1$ 在 $U'$ 中标记 $p$ 个正例和 $n$ 个反例样本 使用 $h_2$ 在 $U'$ 中标记 $p$ 个正例和 $n$ 个反例样本 将这些被标记样本加入 $L$ 中 随机从 $U$ 中选择 $2p+2n$ 个样本加入 $U'$ 中
--

协同训练在网页分类上的应用取得了 95% 的正确率，超过监督学习方法 89% 的性能。文中也证明了只要满足冗余视图的两个条件，协同训练即可有效提升分类器性能。

### 2.4.2 协同训练的分类

在真实问题场景中，标准协同训练充分冗余视图的条件往往较难满足。近几年，研究者们在这方面进行了改进，本文在此角度进行分类。



### 2.4.2.1 标准协同训练

标准协同训练的假设是样本集存在两个充分冗余视图。两个视图上的样本集满足两个假设：1) 任何一个视图上的样本可以训练出强分类器；2) 两个视图标定类别时互相满足条件独立。第一个假设表明视图的有效性，只有数据在该视图上能训练出解决分类的模型，才能够使协同过程中的“训练”是有效的；第二个假设强调视图间的条件独立是保证分类器互相选择样本是有效的。如果两个视图间不满足条件独立性，则两个分类器互相选择的样本都是带有相似信息，不能起到“协同”作用。

### 2.4.2.2 基于属性集划分的协同训练

K. Nigam 和 R. Ghani<sup>[41]</sup>中研究对标准协同训练中的冗余视图条件进行放松。通过实验证明：当属性集充分大时，随机地把属性集划分成两个视图，也能达到较好的协同训练效果。对充分大的属性集进行随机划分其实也是标准协同训练中假设的弱化。当属性集可以通过随机划分组合时，一定程度上构造了条件独立的属性视图。

### 2.4.2.3 基于不同分类器的协同训练

现实问题场景中，样本的相互条件独立视图较难存在，而“充分大”属性集也不一定容易满足。S. Goldman 和 Y. Zhou<sup>[42]</sup>在研究中提出了一种不需要充分冗余视图的协同训练算法。他们将协同的重点放于分类器上。研究中采用不同的分类算法从一个属性集上训练出两个分类器。与标准协同训练算法相似，两个分类器互相从未标记样本集中选择置信度高的样本到对方分类器的训练集中。两个分类器互相选择样本并更新自身，直到达到迭代停止要求。

### 2.4.2.4 基于重复采样的协同训练

Z.-H. Zhou 和 M. Li<sup>[43]</sup>提出的 Tri-training 算法是在 co-training 思想上演化的一种既不需要冗余视图也不要求使用不同类型分类器的算法。其方法是对有标记样本集进行重复采样 (Bootstrap Sampling) 获得三个有标记样本集。每个样本集内各自训练出一个分类器。在协同训练过程中，每个分类器新加入的训练样本由其余两个分类器协同决定。当两个分类器对一个未标记样本的判别一致时才可以选为高置信度样本，从而加入到第三个分类器的训练集中，这个过程是集成学习中常用到的投票法<sup>[44]</sup>。

## 2.5 本章小结

本章从本课题最重要的弱监督学习出发，系统地介绍了弱监督学习的概念、分类及其常用算法。接着，按照本课题目标检测流程介绍了区域提取、特征描述以及本课题的两个核心技术：多示例学习和协同训练。区域选择和特征提取中简要地介绍了本研究所用的 **Selective Search** 和卷积神经网络。多示例学习是弱监督学习的热门方法。本章对传统多示例学习概念进行介绍，并根据理论假设对不同的多示例学习研究进行分类介绍。协同训练是一种经典的半监督学习算法，本章对该方法进行介绍。按照不同的视图划分方法介绍不同的协同训练方式。本章介绍了本框架相关的技术内容，为下章节的框架介绍进行铺垫。

## 第3章 基于多示例学习的弱监督遥感图像目标检测

上一章节介绍了本框架相关的基础技术内容。本章从问题描述和难点分析开始，按照整体框架顺序介绍框架的关键环节设计，最后展示实验结果及分析，系统地介绍本研究的核心框架。

### 3.1 问题描述及难点分析

#### 3.1.1 问题描述

遥感图像因其独特的拍摄视角，符合国防军事、交通建设、水利工程、农业林业、城市规划等领域的应用需求，引起了国内外研究的重视。目前大多数的遥感图像目标检测主要基于监督学习方法。监督学习的主要问题在于需要借助完备的监督信息，为了获得完全正确的监督信息需要进行人工标注。人工标注需要耗费大量的人力物力，特别是遥感图像具有覆盖面积广、目标小而多等特点，其监督信息更加“昂贵”。另一方面，遥感拍摄技术和信息存储技术的发展使得人们很容易获得大量的无标注样本。本研究结合上述两方面问题提出基于弱监督学习的遥感图像目标检测框架。弱监督学习追求仅使用极弱的标注信息达到训练分类器的目的。所以，本研究的问题背景设定为仅使用图像级别标注的训练集（即仅标注图像内是否含有目标）挖掘有效的信息以训练分类器并使得分类器满足一定的性能要求。

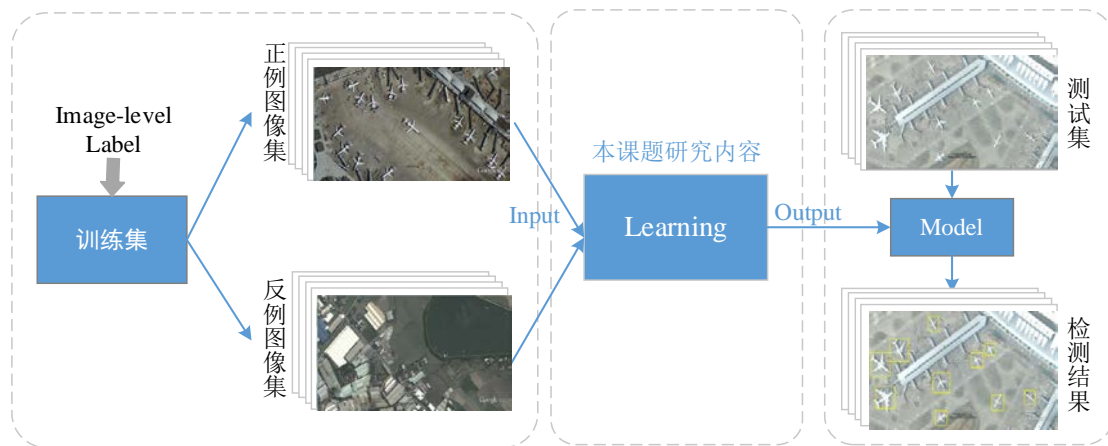


图 3-1 本章问题描述

如图 3-1 所示，本研究的输入为图像级别标注的数据。以本研究的数据集为例，数据集中每幅图像平均包含约 6 个目标区域，部分图像内甚至包含数十个目标，所以样本级别标注任务是十分艰巨的。正因如此，本研究提出弱监督学习为背景，将训练图像仅以是否含有目标作为分类依据，简单地将数据分为正反例图像。该标注过程简单易行且无需耗费过多时间和人力。所以，相对样本级别的标注来说，本文称图像级别的标注为“弱标注”。

### 3.1.2 难点分析

本研究为了避免监督学习中标定数据的过程，仅使用了图像级别标注的数据。在此情况下，为了能达到训练分类器完成目标检测工作，本研究需要思考并解决如下几个重要问题：

- 1) 如何从已有的弱标定信息中挖掘更多可用的信息？
- 2) 如何保证挖掘出的信息的准确性？
- 3) 如何保证迭代过程中信息的多样性？

本研究采用了样本挖掘的策略来代替传统监督学习中的数据标注阶段。上述问题是样本挖掘过程中将会面临的问题。因为采用样本挖掘的策略时，只有有效的样本才能够有效的训练出分类器。所以，从图像数据中挖掘出样本数据是首要问题。其次，挖掘出的样本区域必须能够正确表达所要检测的目标，否则目标检测任务无法达成。最后，挖掘出的样本必须具有代表性，能表现目标类内间的多样性。如果挖掘的样本趋于一致，则训练出的分类器泛化性较差。除了弱监督学习的问题外，遥感图像的目标检测技术本身也存在如下几个难点。

- 1) 场景复杂

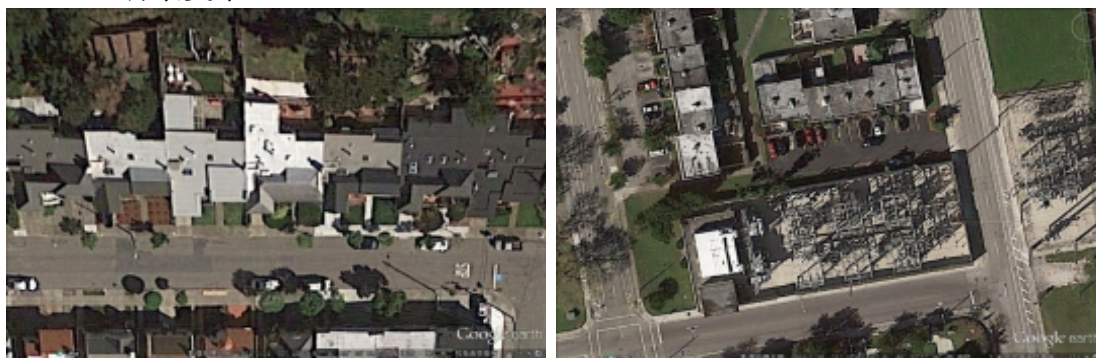


图 3-2 遥感复杂场景示例

遥感航拍摄影覆盖面积大，没有地域界线限制，所以拍摄下的场景复杂，物体多样。

2) 相似物体干扰

拍摄场景中存在着许多结构、形状、纹理与目标相似的物体，容易给样本挖掘、模型训练和目标检测带来干扰。

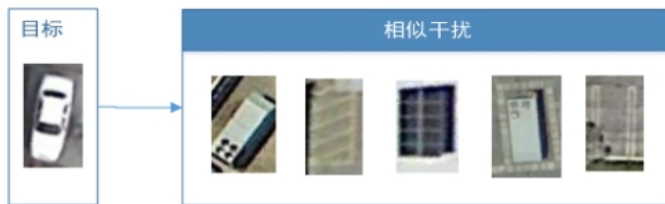


图 3-3 目标相似物干扰

3) 目标旋转角度问题

遥感视角下，目标的角度、长宽比等属性变化范围大，为样本挖掘和目标检测增加了难度。

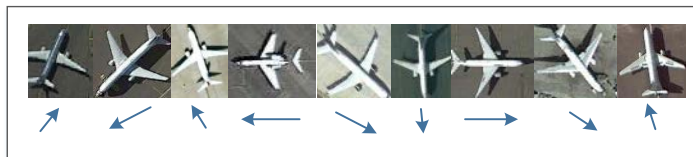


图 3-4 目标旋转范围变化大

### 3.2 基于多示例学习的研究框架

基于多示例学习的弱监督遥感图像目标检测框架是本文的核心研究。其采用多示例学习和协同训练方法的结合作为框架核心，本节先介绍整体框架，再对框架中各个步骤进行介绍。

#### 3.2.1 整体研究框架

本实验整体框架大致由区域选择、特征提取、分类器迭代训练组成。迭代优化的过程是整个弱监督学习的核心。如图 3-5 所示，该过程大致分为样本挖掘和分类器更新二者间的循环交替，训练样本和分类器在迭代过程中相互更新和优化。

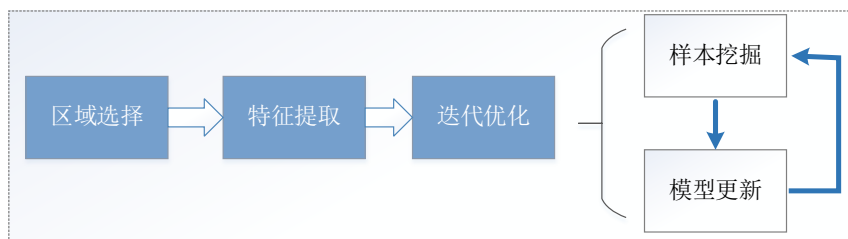


图 3-5 框架模块流程图

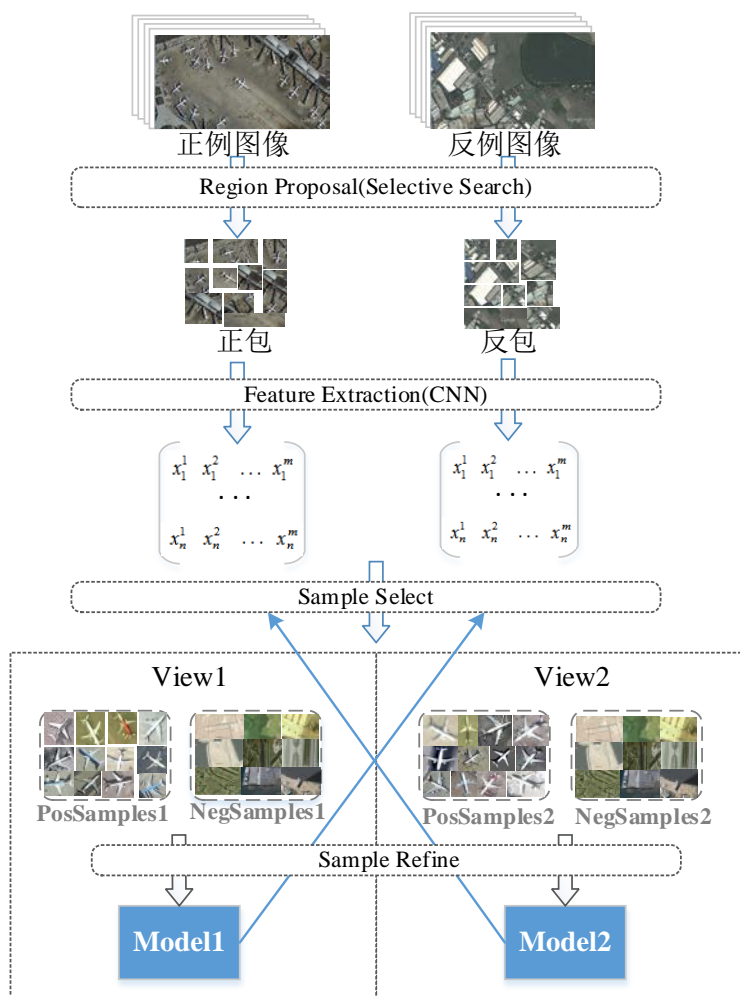


图 3-6 本研究流程图

图 3-6 所示是本章研究框图，具体研究内容如下：

- 1) 首先采用 Selective Search 从遥感图像内选择感兴趣区域；
- 2) 用已训练好的卷积神经网络对步骤 1) 获得的区域内提取特征，每个区

域提取为 4096 维的特征样本；

3) 按照协同训练的方法，将样本属性集划分两个视图，两个视图内的样本各自训练分类器并在迭代中为对方分类器选择置信度高的训练样本；

4) 分类器选择样本的方法是基于多示例学习方法从训练集的各个包内选择当轮最优样本与前一轮训练样本整合成当前训练样本集；

5) 使用样本优化的方法剔除当前训练样本集内的噪声样本；

6) 迭代地更新分类器和训练样本直至收敛。

下文将对研究框架中的每个关键环节进行分节介绍。

### 3.2.2 区域提取-Selective Search

区域选择是本框架中的第一个步骤。其过程是对分类好的正反图像集内的每一张图像进行感兴趣区域选择。本研究所采用的区域选择的方法是 Selective Search。作为一种基于图割的选择性区域提取方法，其已在多篇研究<sup>[13,45]</sup>中得到应用，证明其对图像能有效地提取候选窗口。

在本研究场景上进行区域提取需要考虑的几个问题上，对 selective search 和传统窗口扫描法进行如下对比：

1) 适应不同尺度：由于图像中目标物体的不确定性，区域的提取必须能适应物体这一变化。扫窗法的策略是采用不同窗口大小、预设步长即窗口重叠来保证能提取到目标所在区域。这一方法的问题在于产生的窗口数量众多，计算量大；选择搜索法采用图像分割及层次算法有效解决该问题。

2) 数据的多样化：扫窗法采用单一策略，不借助图像中物体的任何特征信息；Selective Search 采用了颜色、纹理、大小等多种策略对区域进行合并，更好的利用图像中的物体特征信息，能适应数据集中不同类型的图像。

3) 运行速度：扫窗法为保证能提取到物体的区域受步长、窗口大小、图像大小等因素限制，计算量大、冗余窗口多。选择搜索法自下而上结合图割算法和层次结构，速度较快。

本研究中，遥感所摄图像特点在于覆盖面积大、单幅目标众多、数据集内背景不一致，Selective Search 方法显然更适合。其优势其基于数据驱动从底层开始分割，可以满足不同尺度的目标，又省去了传统“扫窗”方式所带来的复杂计算，有效减少冗余窗口。同时，采用颜色、大小、纹理等多种策略进行合并，适用于不同场景和目标。

### 3.2.3 特征提取-CNN

图像分割成候选区域后，需要在区域内提取特征，将图像转为数值表示。本课题研究选择卷积神经网络作为特征提取的方法。其用于提取特征时除了具有旋转、平移、缩放不变性外，还具有训练参数少、速度快等特点。

在 ImageNet 上的图像分类挑战上，AlexNet 网络结构模型赢得了 2012 届的冠军。本课题研究采用经 ImageNet 训练好的 AlexNet 网络进行特征提取。下图 3-7 是 AlexNet 的结构图：

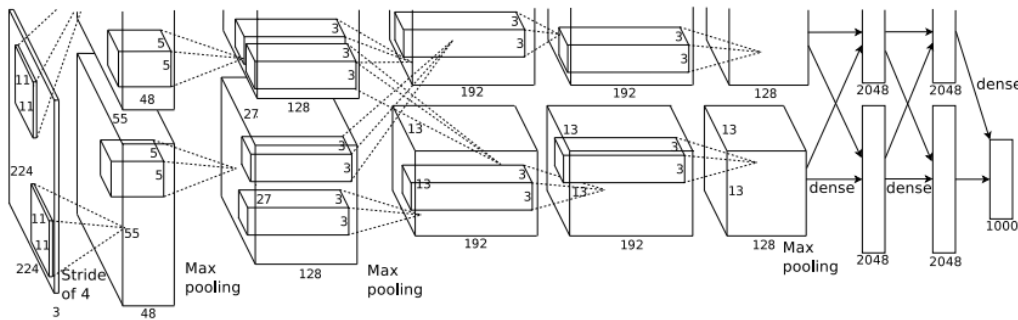


图 3-7 Alex-Net 结构图<sup>[32]</sup>

如图 3-7 所示，AlexNet 是一个八层的卷积神经网络。前五层是卷积层，后三层是全连接层。其中 conv1、conv2、conv5 层后面各跟着一个 pooling 层。最后一个全连接层的输出是 1000 个输出的 softmax。卷积神经网络在浅层时学习局部特征，深层网络对局部特征进行组合。所以层数越高，特征就越全局化。基于此特点考虑，本课题使用 AlexNet 网络的第七层结果作为样本的特征，最终每个样本提取 4096 维特征。

### 3.2.4 多示例学习

#### 3.2.4.1 基础方法

多示例学习定义“包”为多个示例的集合。该方法只对“包”作标签，若包内至少含有一个正示例，则定义为“正包”，否则为“反包”。多示例学习中本身只有包有标记概念，而包内示例无标记概念。本研究中基于多示例学习方法，如图 3-8 所示，将每副正例图像进行区域提取后定义为一个正包，每幅反例图像为反包。图像内的每个候选窗口定义为包内的“示例”。正例样本挖掘过程中，选择包内置信度最高样本作为正例样本与反包内样本进行类似监督学习上的训练，



在迭代过程中对模型进行优化。

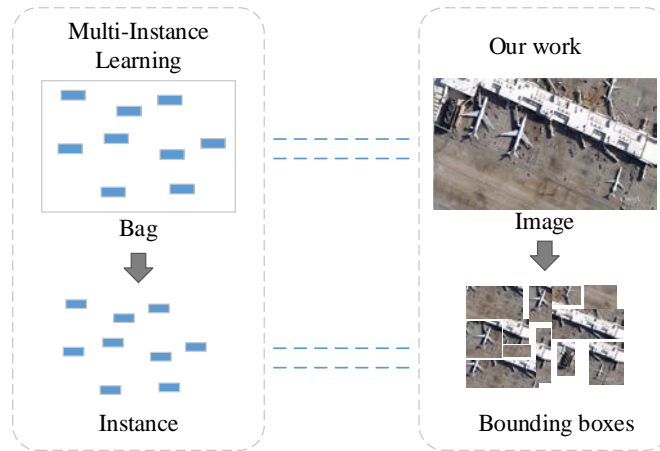


图 3-8 本课题中的多示例学习问题

多示例学习的方法本质上即是迭代优化的过程，整个优化过程分为两部分：监督学习，标记更新。该过程可以抽象为：假设所有样本的标记都是已知的，那么就可以通过某种监督学习的方法得到一个分类模型，通过这个模型可以再对每个训练样本进行预测，然后更新它们的标记。重新得到的标记可以用来重新训练分类模型，重复迭代更新优化直至收敛。

本课题采用的多示例学习方法是基于支持向量机的 MI-SVM 方法，该算法的关键是：1) 训练模型的时候，只从正样本包里挑选被预测的“最像正确”的样本。本研究中选择每个正包中与模型加权求和后置信度最高的样本。正样本包里面其他的样本，不管预测标识为正负都不被选。因为，正样本包里面“最正确”的一个样本标记是正的，跟其他样本无关。2) 挑选每个负样本包里面被预测“最像正确”的样本作为负样本进行训练。本研究中选择反包中与模型加权求和后置信度最高的反例样本，这样子的负样本也可当作难样本 (Hard Sample)。

#### 3.2.4.2 方法改进

传统多示例学习中，每次只从包内选择一个“最优”示例。在本研究中，每幅遥感图像平均生成约 8000 个 Bounding Boxes。若每次迭代在每个包内只选择一个示例，则在本课题的问题场景下将会面临样本的利用率低，样本代表性差，迭代容易陷入局部最优，分类器优化速度慢等问题。

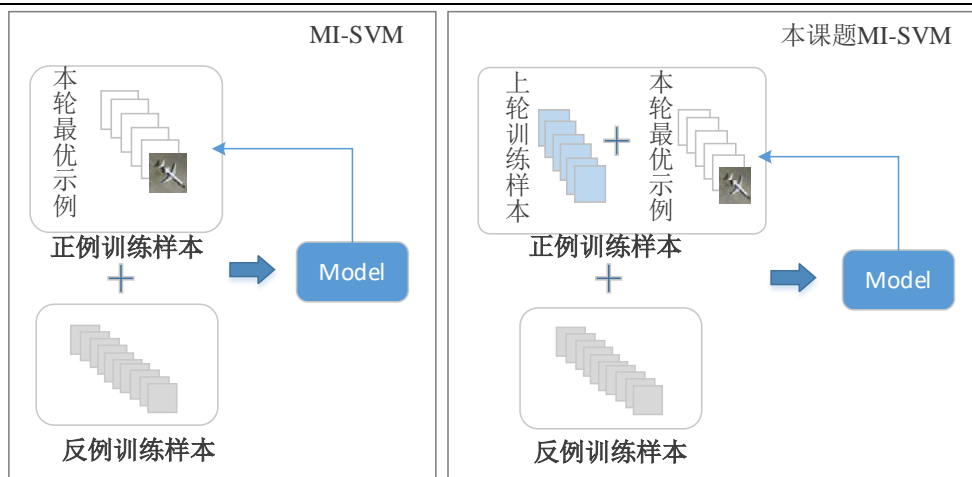


图 3-9 本课题 MI-SVM 的改进

为了适应遥感图像范围大、目标小、窗口多等特点，本研究将原有 MI-SVM 的方法进行修改。如图 3-9 所示，从原有方法每次只选择一个最优示例改为每轮迭代从包内选择“当前最优”示例加入到前一轮训练正例样本中作为本轮正例训练样本。其中，“当前最优”样本必须满足以下两个条件：

- 1) 与上一轮的训练样本的 overlap 不超过 50%
- 2) 在满足上一条件的示例内选择置信度最高的样本。

这样选择样本的方法将原本 MI-SVM 方法中的“替换”式改为“扩充”式，迭代样本数量随着迭代而增加。该方法能有效地挖掘更多的训练样本并保证一定的训练样本多样化。

### 3.2.5 协同训练

为了保证样本挖掘过程中样本的准确性和多样性，本研究中加入协同训练算法。标准协同训练算法的基本思想是假设数据集有两个充分冗余的视图满足以下两个条件：1)每个视图条件独立于另一个视图；2)每个视图内数据都能单独学习到一个强分类器。

在本课题的问题场景中，标准协同训练的假设较难满足。所以，在本课题中参考相关技术章节提到的研究，尝试两种不同的视图构建方法。

- 1) 基于数据集划分的协同训练方法

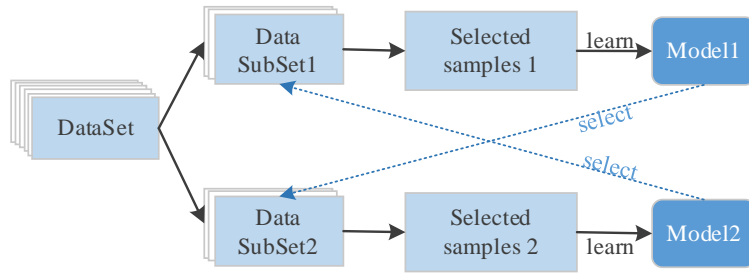


图 3-10 基于数据集划分的协同训练方法

如图 3-10 所示，基于数据集划分的协同训练方法将数据集随机划分成两个子数据集当作两个视图。两个视图的数据集各自训练分类器。每个分类器按照在对方数据集内选择置信度高的样本为对方做训练样本。

2) 基于属性集划分的协同训练方法

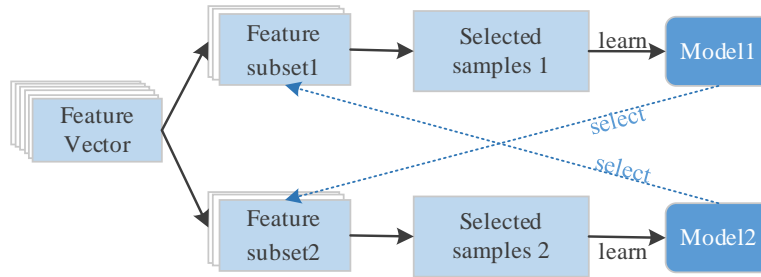


图 3-11 基于属性集划分的协同训练方法

如图 3-11 所示，基于属性集划分的协同训练方法将属性集随机划分成两个子集。以本研究的卷积神经网络第七层提取的 FC7 特征（4096 维）为例随机抽取划分当作两个视图，每个视图含 2048 维特征。两个属性集的数据各自训练分类器，每个分类器 2048 维。每个分类器从数据集内按照自身属性集选择置信度高的样本为对方分类器做训练样本。

本文将在下节实验中对上述两种方法进行比较。

本研究中将协同训练算法与多示例学习结合。针对初始分类器的训练问题，采用 MI-SVM 初始化方法使用包内样本均值作为初始训练样本。采用协同训练与 MI-SVM 方法结合，任一视图下的分类器训练样本由对方分类器挑选最优示例。在协同训练的“互选样本”和多示例学习的“扩充式最优样本选取”的结合下，每轮迭代的训练样本都会逐渐增多，为了保证过程中样本的准确性，避免分类器被误化，本研究加入针对训练样本的优化过程。

### 3.2.6 样本优化

上节提到本课题通过多示例学习和协同训练每轮迭代扩充训练样本。与传统监督学习包含完备监督信息不同，弱监督学习需要从众多无标记数据中挖掘可用样本以供训练器学习。所以挖掘出的样本准确性和多样性将决定在整个分类器学习的方向以及最终性能。为了保证训练样本这两方面的质量，本框架对挖掘出的样本加入样本优化（**Sample Refine**）步骤。

其规则为：对于在每轮新增训练样本与前一轮的训练样本共同组成了当轮候选训练样本。样本优化的方法是将计算以下两项的值：

- 1) 候选样本间进行相似度计算，每个样本取与其他样本相似度的均值；
- 2) 上一轮分类器对本轮候选样本进行置信度打分；

将上述两项得分进行归一化后求和。对正例训练样本按其求和后的值降序排列，去掉排名最末的 10% 的样本。实验证明这两者的综合考虑比单一元素考虑效果更好。

## 3.3 实验结果与分析

本节中将展示针对框架设计和改进的实验，包括细节实验和总体性能实验，同时对实验所用数据集进行介绍。

### 3.3.1 数据集介绍

本课题除实验框架外，还制作了一个遥感光学图像数据集。本节将对现有公开数据集进行介绍，再重点介绍本课题数据集。

现公开的可用遥感航拍数据集不多。通过资料整理得到目前可用于航拍图像目标检测的公开的数据集有：1) **PAMView: Providence Aerial Multi-View Dataset**<sup>[46]</sup> 2) **OIRDS: Overhead Imagery Research Data Set**<sup>[47]</sup>

#### ① PAMView: Providence Aerial Multi-View Dataset

该数据集是通过直升机飞过美国罗得岛州首府普罗维登斯城市拍摄的图片，直升机以环形轨迹飞过每个点。图像分辨率大约 30 厘米/像素，通过 **Bundler** 和 **VisualSFM** 获得标定矩阵。

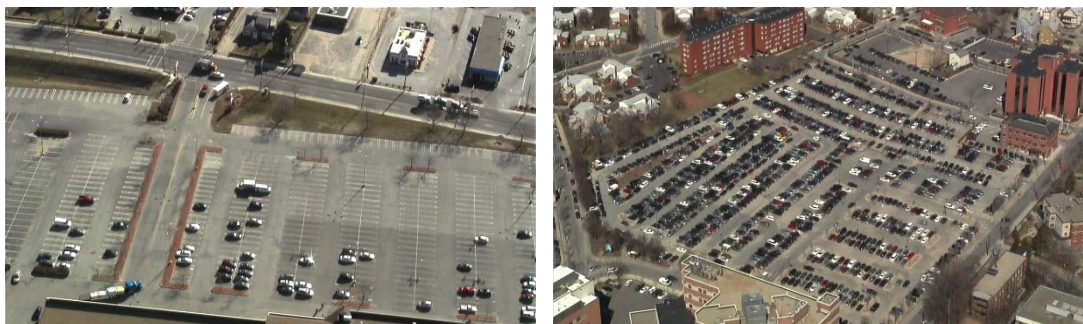


图 3-12 PAMView 数据图像示例

PAMView 数据缺点是集中目标过小，拍摄区域只局限在普罗维登斯城市，场景单一。

### ②OIRDS: Overhead Imagery Research Data Set

OIRDS 数据集里的 vehicle 包括汽车、卡车、货车等。除了目标轮廓，OIRDS 还包含图像上下文信息的对车辆量化，在主观和客观两方面的统计。



图 3-13 OIRDS 数据集图像示例

OIRDS 数据集图像小，集内图像数量少，单幅图像目标数量少。

鉴于上述两种数据集缺点明显，本课题制作了一个光学遥感航拍图像数据集名为：中科院大学高清航拍数据集（UCAS-High Resolution Aerial Object Detection Dataset, UCAS-AOD）。该数据是通过 google earth 获得高分辨率的遥感航拍图像，使用人工标定方法，对航拍图像中的目标进行标定。

数据集版本和数量参见表 3-1 每个目标结果为 13 字段格式： $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, \theta, x, y, \text{width}, \text{height}$ 。

表 3-1 数据集信息统计

数据集	飞机图像	飞机样本	汽车图像	汽车样本	反例图像
Version 1.0	600	3591	310	4475	492
Version 2.0	400	3891	200	2639	408
总计	1000	7482	510	7114	910



图 3-14 中科院大学高清航拍目标数据集 (UCAS-AOD) 示例

本数据集使用了 Google Earth 获取图像。Google Earth 数据是由遥感卫星影像与航拍的数据整合。其卫星影像部分来自 QuickBird(快鸟)商业卫星与 EarthSat 公司。航拍部分的来源有 BlueSky 公司、Sanborn 公司、美国 IKONOS 及法国 SPOT5。其中 SPOT5 可以提供解析度为 2.5 米的影像、IKONOS 可提供 1 米左右的影像。QuickBird 能够提供最高为 0.61 米的高精度影像，是目前全球商用的最高水平。

### 3.3.2 实验结果与分析

本课题在 UCAS-AOD 数据集上进行实验，为了与监督学习方法<sup>[13]</sup>进行比较，本课题采用了与其规模一样的训练集和测试集。飞机数据集包括 600 幅图像，其中包含训练集 500 幅和测试集 100 幅。训练集的 500 幅遥感图像，平均每幅图像包括 3920 个窗口和 5.8 个目标窗口(ground truth)。按照文<sup>[13]</sup>对数据集的统计，样本角度分布非常均衡。

分类器对图像的窗口进行置信度计算，当结果超过设定的阈值判为正例。判为正例的窗口以图中坐标位置的形式表现。当检测窗口与 ground truth 的 overlap 超过 50%则表示检测正确。overlap 的计算方式如下：

$$\text{overlap} = \frac{S_{\text{window}} \cap S_{\text{gt}}}{S_{\text{window}} \cup S_{\text{gt}}} \quad (3-1)$$

其中,  $S_{window}$  表示检测结果位置,  $S_{gt}$  表示 ground truth 位置。

本文下面部分将介绍本研究中的部分实验, 包括细节设计验证实验和性能对比实验。

1) 不同距离度量方式比较

由于本研究在多处应用到距离度量作为相似度计算, 所以相似度的衡量, 比较三种距离度量方式: 欧氏距离、余弦距离、卡方距离。输入为 400 幅正例图像、300 幅反例图像, 实验结果统计如下:

表 3-2 距离度量方式比较结果

度量方式:	欧式距离	余弦距离	卡方距离
平均时间:	0.1027ms	0.1371ms	1.3971ms
准确率:	41.97%	36.53%	51.7%
召回率:	15.08%	13.01%	16.03%

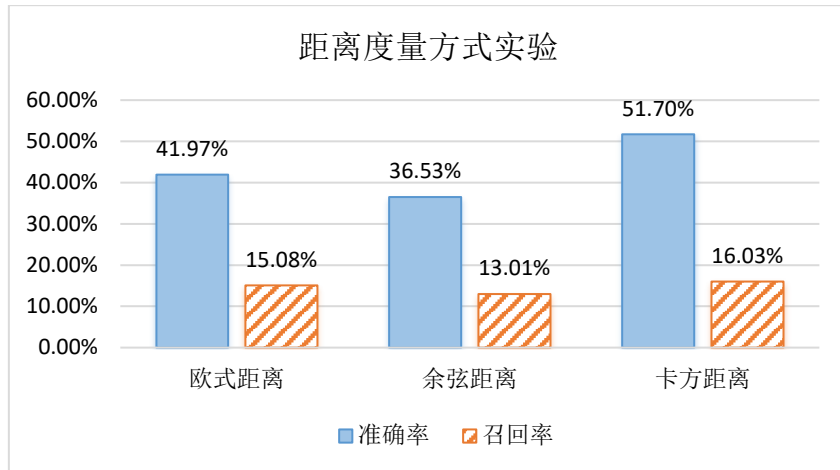


图 3-15 距离度量方式实验结果

图 3-15 所示, 卡方距离度量方式与另外两种方法相比, 耗时较长, 但准确率比最低的余弦距离高 15%, 召回率高 3%。在本研究中多处需要衡量相似度都采用卡方距离度量方式。

2) 多示例学习实验

标准的多示例学习中选出“最正确样本”即置信度最高的样本。本研究前期实验中为了最大化有效正例样本个数, 所以在每轮迭代选择包内置信度靠前且相互之间重叠率不超过 50% 的前五个样本作为候选样本。这样的问题在于: 当正

例图像的示例小于五个时，挑选出的样本必然含有反例样本，容易干扰后阶段的噪声去除步骤，误化分类器。于是，本研究将迭代过程中的样本挑选改为扩充式。具体方法为：初始化时使用包内均值作为正反例训练出初始分类器后，将正反例样本清空。之后每次迭代，都将包内置信度最高的样本添加入候选正反例样本，前提是该轮选出的样本与已选正例样本的重叠率不超过 50%，即非重复样本。然后，新旧候选正例样本一同进入样本优化步骤。

在迭代扩充式地挖掘样本中，本课题为了保证挖掘出的训练样本的准确性，在实验框架中通过设定的规则进行筛选样本。如果不在迭代过程中去除混杂在正例训练样本中的反例，将会使分类器性能始终无法提升。本实验尝试将混杂在正例样本中的噪声去除。具体步骤为：

- ①每个样本间互相计算相似度，取均值；
- ②使用前一轮的分类器对样本进行置信度计算；

对①，②两步骤获得的得分进行归一化，综合两种总分进行从小到大排序，剔除排名最末的 10% 的样本。

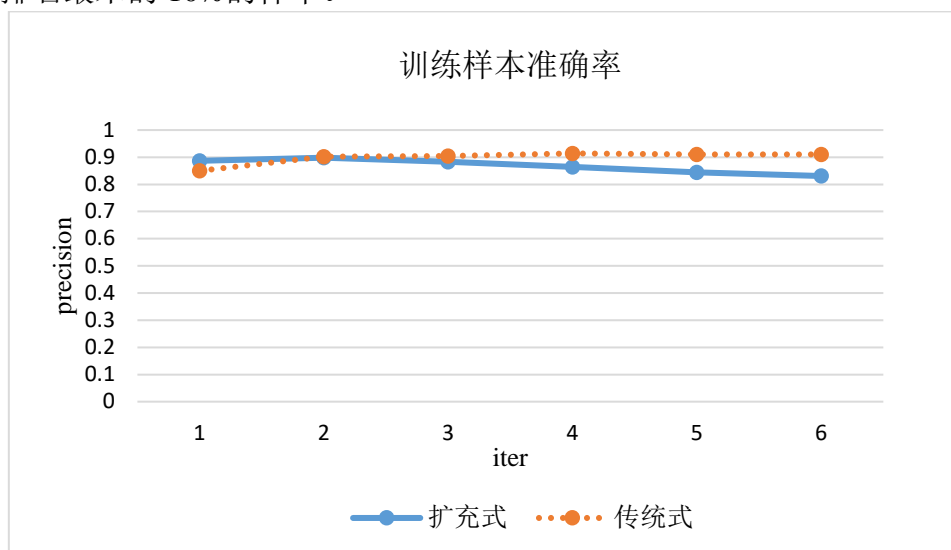


图 3-16 每轮迭代中训练样本的准确率



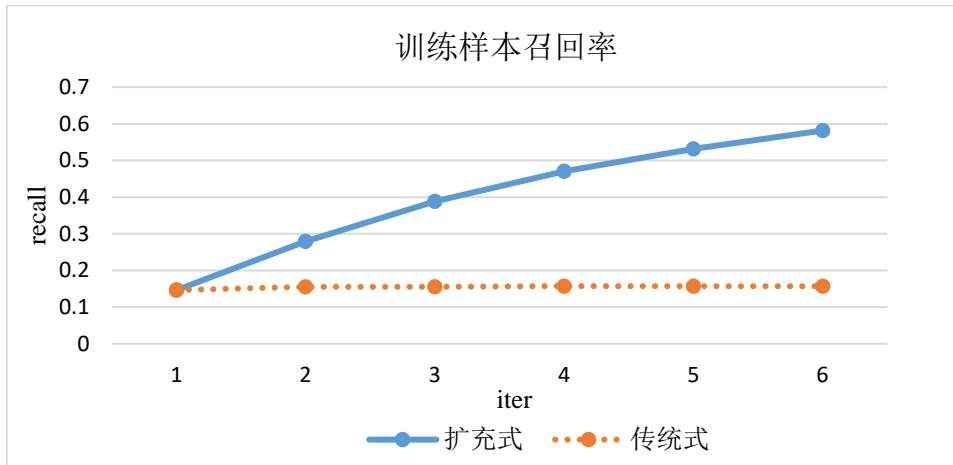


图 3-17 每轮迭代中训练样本的召回率

图 3-16 和图 3-17 表示的是在每轮迭代中挖掘出的正例训练样本与 **ground truth** 相比的准确率和召回率。传统多示例学习每次从包内只选择一个样本，所以有较高准确率，但是迭代进行到第二次后即趋于稳定。由于该方法的样本利用率较低，所以挖掘出的样本召回率非常低。本研究中的扩充式多示例学习除了以扩充的形式增加样本，在保证样本挖掘的正确性之余，提高了样本挖掘的召回率，保证了样本的多样性。当迭代次数过多，扩充的样本数量超出图中原有目标数量时容易降低准确率，所以在迭代训练过程中需留意样本多样性与准确率之间的平衡。

### 3) 协同训练实验

本节比较两种协同训练方法：一种是基于数据集划分的协同训练方法，将训练集随机划分成两份子数据集。每个子数据集各自训练分类器并为对方分类器在其数据集内选择高置信度样本作为训练样本；另一种是基于属性集划分的协同训练方法，将属性集随机地划分成两个子属性集。每个样本在两个子属性集内训练分类器，分类器在本属性集内选择高置信度样本给对方分类器在其对应属性集内作训练样本；

图 3-18 表示在原框架基础上，加入协同训练能有效提升分类器性能。这是由于通过不同视图对数据进行表示，可以学习到同一样本的不同属性。两个分类器之间进行样本挑选，除了增强样本的准确率，同时也能够增加样本集内的多样性。通过实验对比可以得出，本实验中基于数据集划分方法稍优于基于属性集划

分方法，但二者性能上差距甚小。本实验中仅使用了 400 幅训练图像，每个样本区域提取了 4096 维特征。如果不受数据集大小和提取的特征维度限制，协同训练方法的优越性将进一步得到体现。

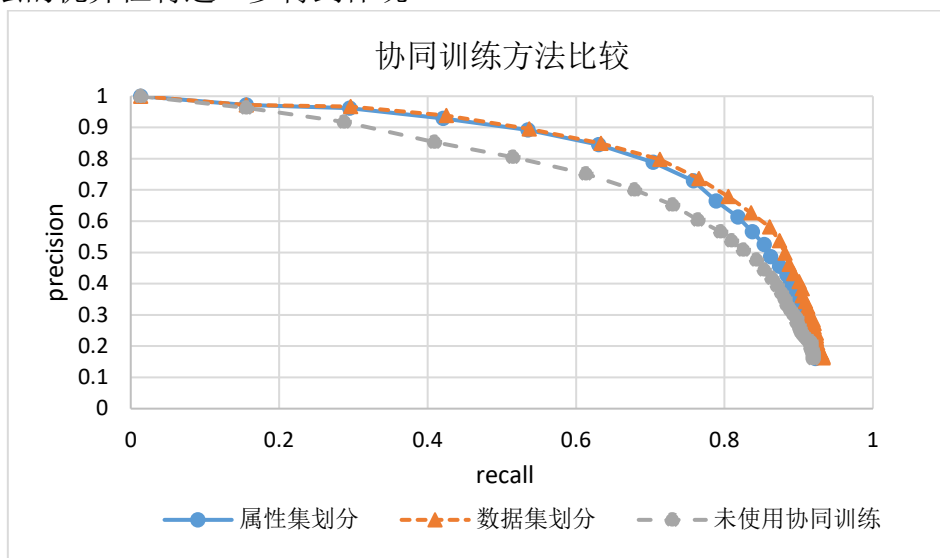


图 3-18 协同训练实验比较

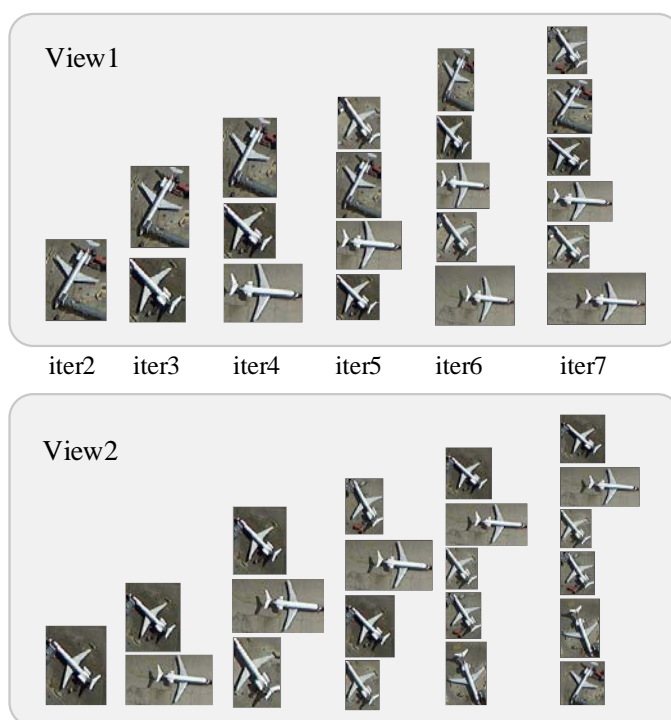


图 3-19 每轮迭代中正例训练样本的挖掘结果

图 3-19 展示的是不同视图中的每轮迭代的样本挖掘结果。上下框中表示两个视图内各自样本挖掘结果，横向表示第二轮到第七轮迭代。本实验中的扩充式多示例学习方法使得训练样本随着迭代而增多。由图中所示的结果可以看出：①本框架的样本挖掘方法能够有效地从众多窗口中挖掘出正确的目标区域；②通过扩充式的方法，每轮迭代都有新增的区域，提高了样本的利用率；③不同视图内选择出的样本稍有不同，体现了协同训练的作用。

#### 4) 整体实验结果

本实验将本文整体框架与 Zhu<sup>[13]</sup>的监督学习研究进行对比，采用相同的训练集和测试集。Zhu 的方法同样采用 selective search 提取候选窗口，卷积神经网络提取特征，使用支持向量机作为分类器。不同的是其框架中采用人工标注的样本集来训练模型，而本研究采用图像级别的图像数据。Zhu 的监督学习方法为了保证模型的准确性和泛化性使用了 400 幅训练图像中的近 3000 架飞机样本。本课题通过对训练集内 400 幅图像进行图像级别标注，有效地挖掘正例训练样本来训练分类器。下图为二者之间的性能比较结果：

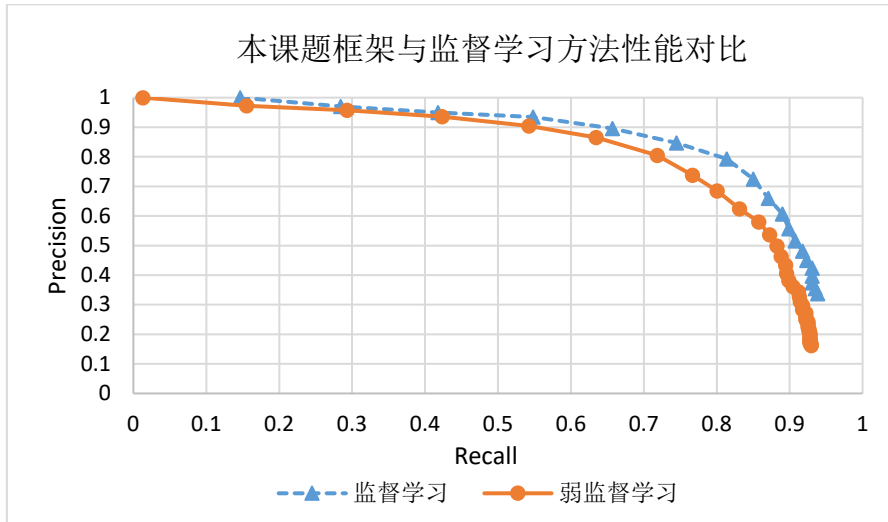


图 3-20 本框架与监督学习性能对比

如图 3-20 所示：本框架的性能与监督学习方法的性能接近。本框架通过挖掘训练样本和分类器的迭代更新能有效地从图像中进行目标模型学习。本框架重点在于仅使用了图像级别标注的数据，与监督学习大量的样本级别标注相比，极大地降低了人力物力的消耗，体现了弱监督学习的优势。

基于多示例学习的弱监督遥感图像目标检测



图 3-21 本框架在飞机数据集检测结果示例

图 3-21 展示了部分飞机数据集的检测结果。由图可见，虽然检测结果仍存在一定的误检和漏检的情况，但本框架基本能达到检测功能。后四张检测结果图

展示了一个较明显的问题是将部件区域误检为目标。其原因除了机尾部分与飞机整体在结构上具有一定相似性外，还包括弱监督学习在挖掘样本过程中迭代的方向性难以完全把握。在本实验的训练数据中，目标部件区域与目标整体区域存在共现性即在正例图像中部件与目标整体通常是一起出现的。此外，目标部件区域与目标整体区域在反例图像中都没有出现，所以仅通过反例图像难以去除部件区域的干扰。由于本框架仅使用了图像级别的标注，在迭代挖掘时去除正例训练样本集中的部件区域难度较大，所以学习到的分类器与监督学习相比还存在不足。

### 3.4 本章小结

本章介绍了本课题的实验框架，包括候选框选取、特征选取、样本挖掘和模型训练过程。样本挖掘基于多示例学习和协同训练的结合，二者能有效地在每个图像的众多窗口中挖掘出正例样本，同时兼具样本的准确性和多样性。本章还介绍了样本优化的方法，以样本间的特性进行样本筛选。此步骤进一步保证样本的准确性，避免误化分类器训练。在实验结果和分析中，展示了实验框架的细节设计对比和整体实验对比，包括：距离度量方式实验、扩充式多示例学习实验、协同训练实验、样本优化实验以及与监督学习的对比实验。

距离度量方式实验是为了选出对本数据集的 FC7 层提取出的特征有效的距离度量方式。实验比较了三种常见距离度量方式，通过实验最终选择卡方距离；扩充式多示例学习实验验证了本课题框架基于传统多示例学习改进的扩充式多示例学习的有效性，证实了扩充式多示例学习能够保证样本挖掘的准确性和召回率；协同训练实验比较了两种可行的视图划分方法的性能，实验证明两种方法都能有效提升框架性能；样本优化实验证实了样本优化步骤对挖掘出的训练样本有提高准确率的功能；整体实验是通过本课题与监督学习方法的比较，证实了本框架在使用极弱标注信息的情况下，能有效地接近当前最热门的监督学习方法的性能。



## 第4章 基于样本种子点的弱监督遥感图像目标检测

本章在上一章的基础上针对遥感复杂场景引入了样本种子点，以极小的标注代价高效地引导样本挖掘和分类器学习。本章首先阐述了复杂场景的问题，然后从样本种子点的初始化方法、难例挖掘方法和整体框架上介绍基于样本种子点的弱监督遥感图像目标检测。

### 4.1 问题描述

第三章中的基于多示例学习的弱监督遥感图像目标检测框架是基于图像级别标注的数据，通过迭代挖掘样本的方法训练分类器。在现实场景中，遥感图像还可能面临场景杂乱、目标小、形态变化多、边界模糊等问题。在这些情况下，挖掘样本的难度将变得巨大，同时迭代过程中样本挖掘和分类器训练都容易收敛于错误方向。然而，如果能够利用极少量的标注样本作为样本种子点结合原有的图像级别标注共同组成标注信息，则能够有效地进行初始化工作，进一步保证样本挖掘的正确性。

与监督学习、半监督学习相比，基于样本种子点的弱监督学习方法优势仍然在于使用极弱标注信息，减少标注耗费的人力物力。与第三章的方法相比，只增加极少量的标注样本达到更好的初始化和学习效果，适用于更为困难的遥感图像目标检测问题中。

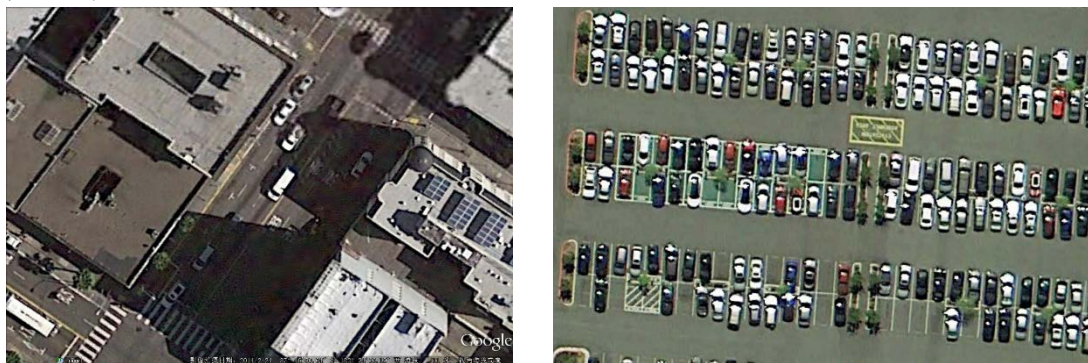


图 4-1 复杂场景数据集示例

如图 4-1 所示，本课题研究制作的汽车数据集中，存在场景物体繁多、目标较小、边缘信息模糊、目标变化大等挑战。本研究在原框架上使用尽量少种子点

样本“引导”样本挖掘过程从而进行分类器学习，同时保持弱监督学习不依赖大量人工标注的优势。

## 4.2 基于样本种子点的研究框架

本小节按照框架流程介绍本章框架的关键环节方法。

### 4.2.1 基于样本种子点的初始化方法

本课题提出一种基于 K-means 聚类的训练样本初始化方法。将遥感图像的区域集使用 K-means 方法划分成多个聚类，每个示例对应一个聚类类别，每个聚类含一个聚类中心。

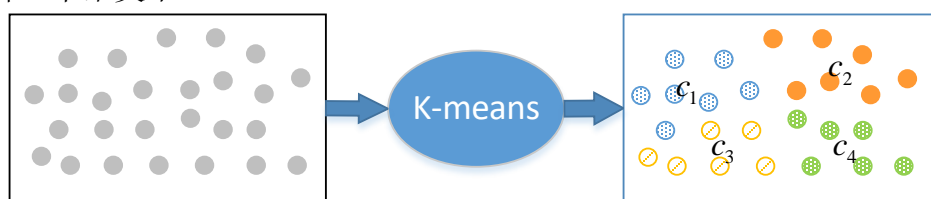


图 4-2 图像区域的 K-means 聚类过程

如上图 4-2 所示，K-means 方法在区域集内划分聚类的具体方法是：随机在区域集内分布多个指定数目的种子点，每个样本与各个种子点之间计算距离并选择最近种子点作为类别，针对每个聚类的类内样本更新聚类中心。重复上述两个步骤直至样本聚类类别不再改变。

将样本分成多个聚类后，大多数正例样本因其内部共性将能有效地聚在一类中。如何从众多聚类中正确选择出包含正例样本较多的聚类是初始化阶段的核心工作。基于是否用样本种子点，本课题实验了两种选择方法：

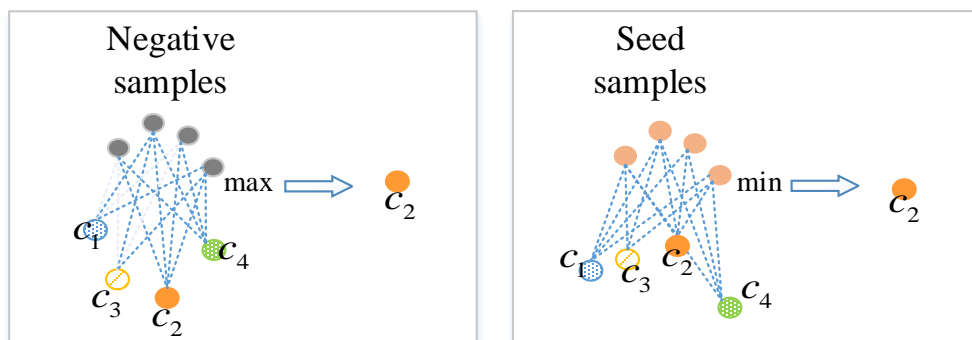


图 4-3 两种聚类选择方法示意



## 1) 基于反例样本选择聚类

在弱监督学习的场景中，正例图像内包含正例样本和反例样本，但是反例图像内的样本都是反例。本课题利用这个特性，在不使用样本种子点的情况下，使用反例图像中获取的反例样本来选择与聚类中心计算相似度，选择“最不相似”的类团。

## 2) 基于样本种子点选择聚类

样本种子点是本课题中使用的极少数的标注样本，在初始化较难的情况下，使用其来“引导”样本的挖掘。利用种子点与聚类中心计算相似度，选择“最相似”的类团。

上述两种类团选择方式都利用了未标注数据的聚类假设。在实际实验中，直接选取“最相似”或“最不相似”的类团还是容易出错的。许多反例类团在简单的距离度量中容易接近正例类团的结果。聚类数目的设定也会直接影响样本初始化的效果：当聚类数目过大时，正例样本被分成多类而不能尽可能在一个聚类类团中；当聚类数目过小时，正例样本所在类容易掺杂大量的反例样本。因此，本文在实验阶段对这两种聚类选择方式进行了放松，以反例样本选择的方法为例，当距离计算结束后，按照下式计算阈值

$$\text{threshold}_{neg} = (\max dis - \min dis) * 0.8 + \min dis \quad (4-1)$$

基于反例样本选择聚类方法按公式(4-1)所示计算阈值 $\text{threshold}_{neg}$ ，当某聚类与反例样本的平均距离大于该阈值时，则被当作被选聚类。

$$\text{threshold}_{seed} = (\max dis - \min dis) * 0.2 + \min dis \quad (4-2)$$

基于种子点样本选择聚类方法按公式(4-2)所示计算阈值 $\text{threshold}_{seed}$ ，当某聚类与种子点样本的平均距离小于该阈值时，则被当作被选聚类。

选出聚类后，将种子点与聚类中每个样本点计算距离，选出距离最近的若干个样本点当作该图像中选出的初始训练样本。将训练图像中选出的初始训练样本收集后即可与反例样本一同训练初始分类器。

## 4.2.2 基于样本种子点的难例挖掘方法

第三章中，不使用样本种子点的难例挖掘方法是用每轮训练的样本从反例图像中选择置信度较高且互相之间重叠度低的样本作为下一轮的难例样本参与分类器训练。在本章中，采用样本种子点在其图中重叠率符合一定范围内的窗口。该方法更为严格的一个限制是难例不应有包含整个样本种子点的窗口。

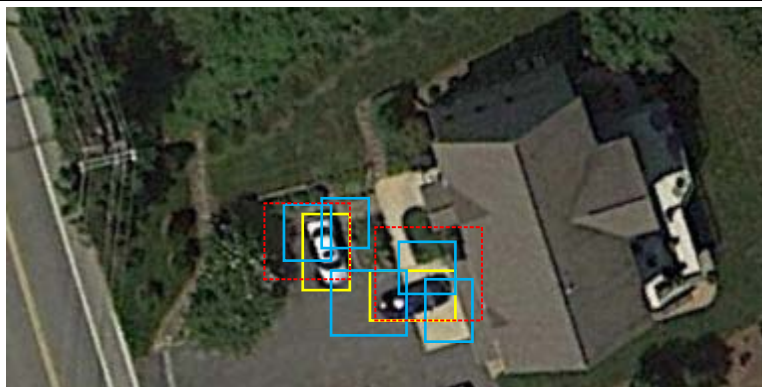


图 4-4 难例样本示意

如图 4-4 所示，该图是正例图像中截取的部分。以图中两辆车为样本种子点为例，黄框表示种子点样本区域；蓝色框是与种子点样本的 **Overlap** 在一定范围内，符合难例样本条件的窗口；红色虚线框表示该区域虽然与种子点样本的 **Overlap** 在所设置的范围内，但是该窗口包含大部分的目标，故舍去。

使用种子点样本进行难例样本挖掘的目的在于能够解决样本挖掘中的 **part** 问题。由于样本挖掘很容易陷入向目标局部区域的挖掘中去，依靠样本种子点周围区域可以有效解决这个问题。

### 4.2.3 整体实验框架

本小节以第三章的实验框架为基础，综合上述的初始化训练样本、难例挖掘等模块构成本章的整体实验框架。如下图 4-5 所示，基于种子点样本的弱监督遥感图像目标检测与第三章的实验框架核心技术相同，都是基于多示例学习和协同训练相结合的方法。具体的框架流程包括：

- 1) 使用 **selective search** 在正反例图像内提取区域，使用卷积神经网络对种子点样本、正包区域、反包区域提取特征；
- 2) 针对每一个正包，通过 **K-means** 对包内样本进行聚类。然后，使用反例图像选择聚类，使用种子点样本在被选聚类内选择初始正例训练样本，组成初始训练样本集；
- 3) 使用种子点样本进行难例挖掘，组成难例集；
- 4) 与第三章框架相同，将数据按照属性集分割成两部分，每部分各自训练分类器。每个分类器为对方挑选置信度高的样本；

- 5) 以扩充式多示例学习形式，每轮增加正例训练样本；
- 6) 训练样本经过样本优化过程后更新分类器。

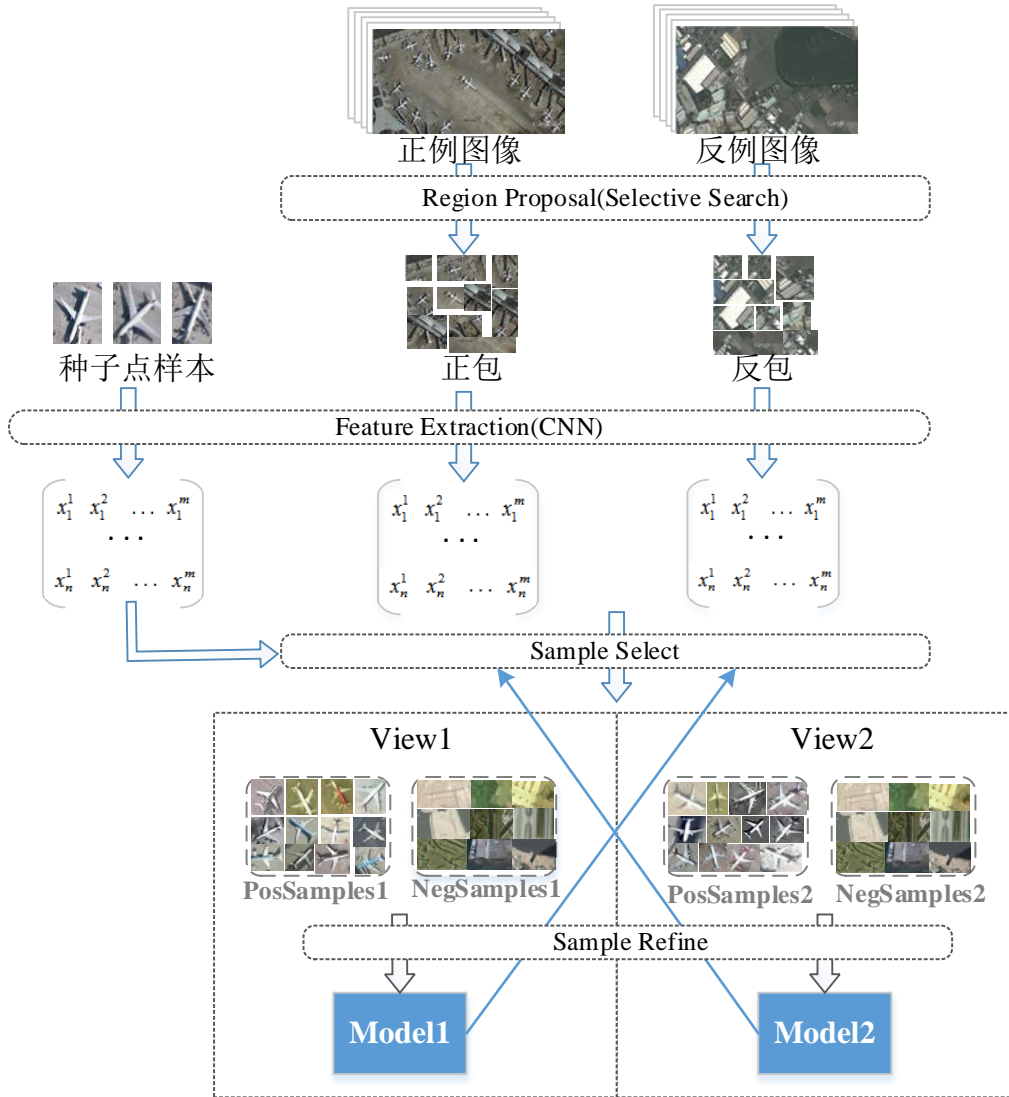


图 4-5 基于种子点样本弱监督遥感图像目标检测流程图

### 4.3 实验结果与分析

#### 1) 初始化实验

引入种子点样本后的初始化方法包含如下几个实验，包括对聚类数目的实验、基于反例样本的聚类类别选择和基于种子点样本的聚类类别选择。

### ① 聚类数目实验

初始训练样本方法中，先使用 K-means 方法将图像内的区域集进行聚类。图 4-6 实验结果表示当聚类个数变化时，聚类中样本准确率最大值。其中，样本准确率的计算方式是各个类内样本与 ground truth 计算 overlap，当重叠率超过 50% 即算为正确。下图结果所示仅对数据进行统计，不属于初始化步骤。

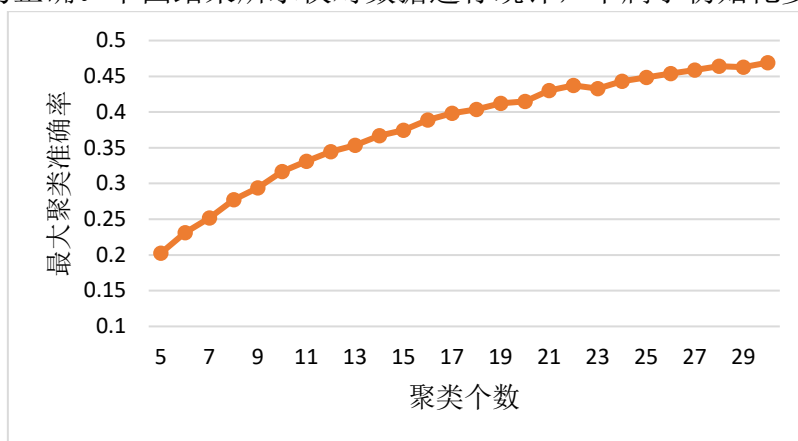


图 4-6 聚类数目实验结果

如图 4-6 所示，随着聚类个数的增加，聚类中的样本准确率最大值越高。当聚类个数过高时，类内样本准确率高。但是由于划分过细，正例样本会被划分到不同聚类中，样本多样性受到破坏；当聚类个数过低时，聚类内包含反例样本过多，容易在选择初始正例样本时增加难度。本研究设定聚类个数为 15，在类内样本的准确率与多样性取平衡。

### ② 聚类选择实验

本课题实验两种聚类选择方法：一种是基于反例样本的聚类选择方法。反例样本与各个聚类中心计算相似度，选择中心距离大于所设定阈值的聚类；另一种是基于种子点样本的聚类选择方法。种子点样本与各个聚类中心计算相似度，选择中心距离小于所设阈值的聚类。

聚类选择后，由种子点样本选择相似度最高的十个不重复样本作为该图像对应的初始训练样本。

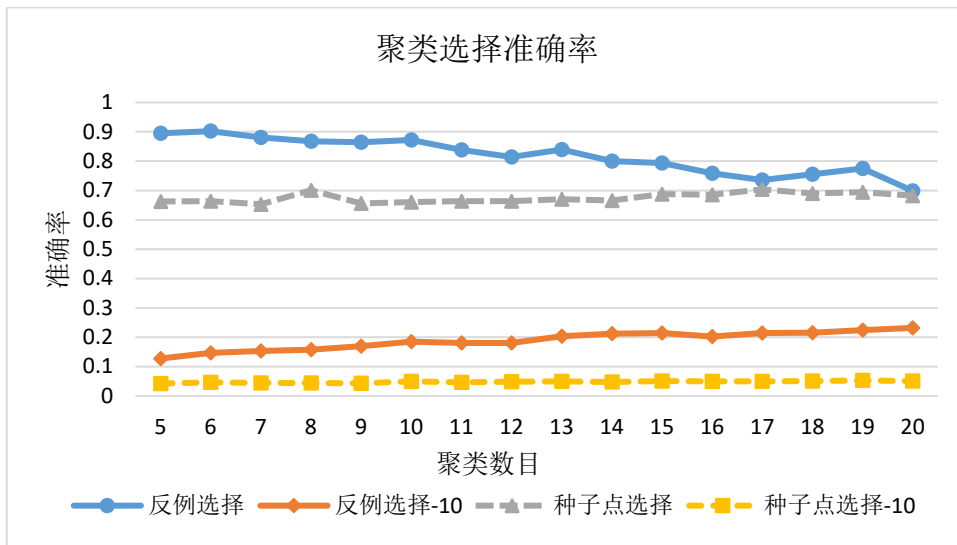


图 4-7 两种聚类选择准确率比较

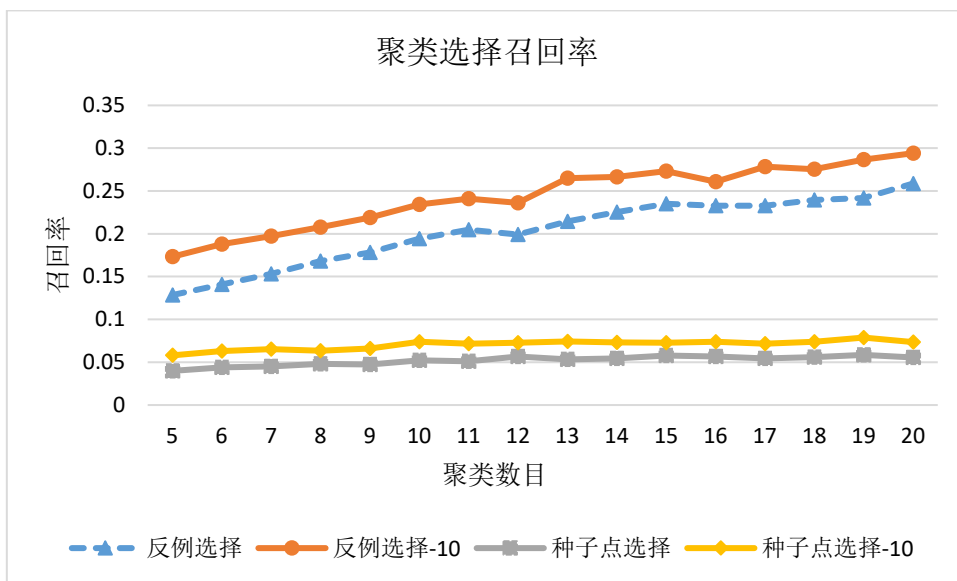


图 4-8 两种聚类选择召回率比较

图 4-7 与图 4-8 表示两种聚类选择方法的准确率和召回率。准确率的计算方法是训练图像的区域集按照聚类数目划分成多个类后，依照两种方法选择的聚类内样本与 ground truth 重叠率超过 50% 的个数占该聚类样本总数的比例。“-10”表示由种子点样本在所选聚类内选择的相似度高的不重叠的十个样本。

由上图实验结果可以看出由反例样本选择的聚类具较高的准确率和召回率，

尤其在召回率方面比种子点样本选择聚类的效果好。原因可能在于反例样本选择聚类再由种子点样本从中选择训练样本可以结合反例样本和种子点样本的正反例信息，而种子点样本选择聚类再由种子点样本从中选择训练样本过于依赖于种子点样本，当种子点样本与训练图像因环境因素等导致差别较大时，容易完全无法初步找出正例样本。两种方法的样本初始化结果如图 4-9 所示。

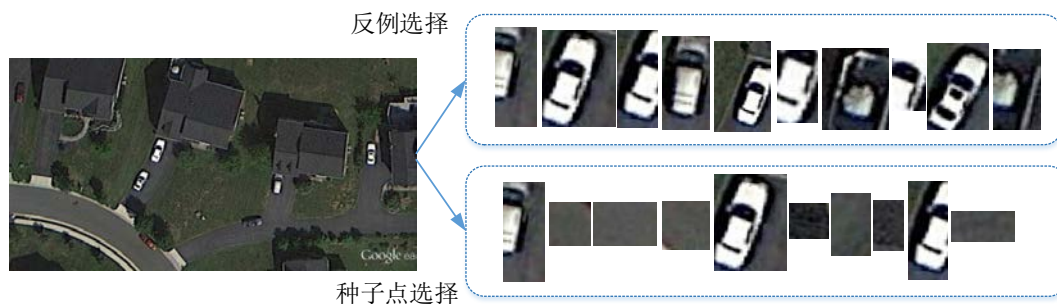


图 4-9 训练样本初始化两种聚类选择方法结果示例

## 2) 种子点样本个数实验

本实验比较种子点样本个数对实验性能的影响。实验设置如表 4-1 所示：

表 4-1 种子样本选取数量表

	图像数目	样本数目
训练集	250	3686
Seed1	1 (0.4%)	56 (1.52%)
Seed2	3 (1.2%)	95 (2.58%)
Seed3	4 (1.6%)	127 (3.45%)
Seed4	5 (2.0%)	130 (3.5%)
Seed5	6 (2.4%)	162 (4.4%)

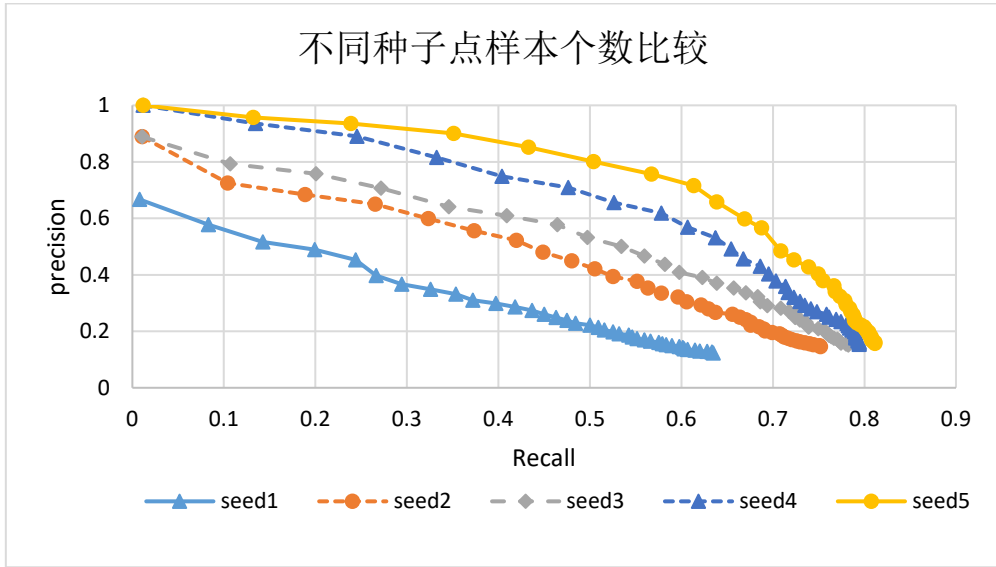


图 4-10 不同种子点样本个数比较

实验结果如图 4-10 所示，种子点样本个数对整体性能存在影响。当种子点越多，学习到的分类器性能越好。但本实验中最多仅标注了 6 幅图像，仅占监督学习所需标注量的 4.4%。

### 3) 难例挖掘实验

本实验比较基于种子点样本的难例挖掘步骤对实验框架的影响。本框架中挖掘的难例样本包含两种：一是分类器在反例图像内选择置信度较高的若干个样本；二是种子点样本在其正例图像内选择与之重叠率在一定范围内的样本。下图实验结果表示有无这两种难例样本的实验结果对比，以样本挖掘过程中的样本准确率来衡量。

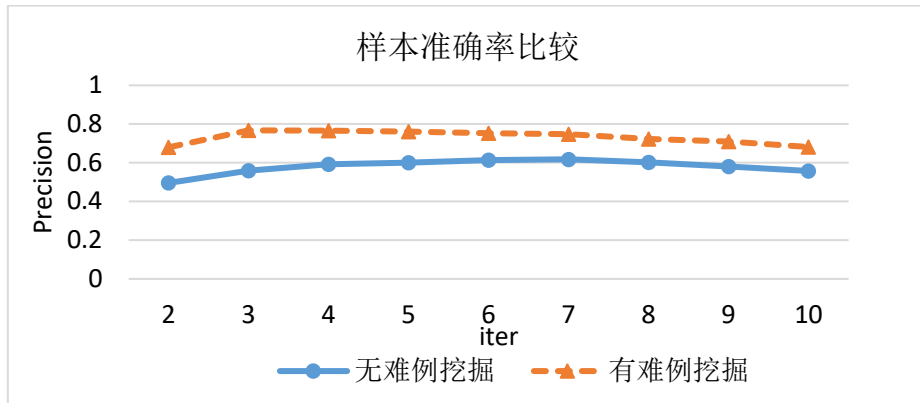


图 4-11 难例挖掘样本准确率比较

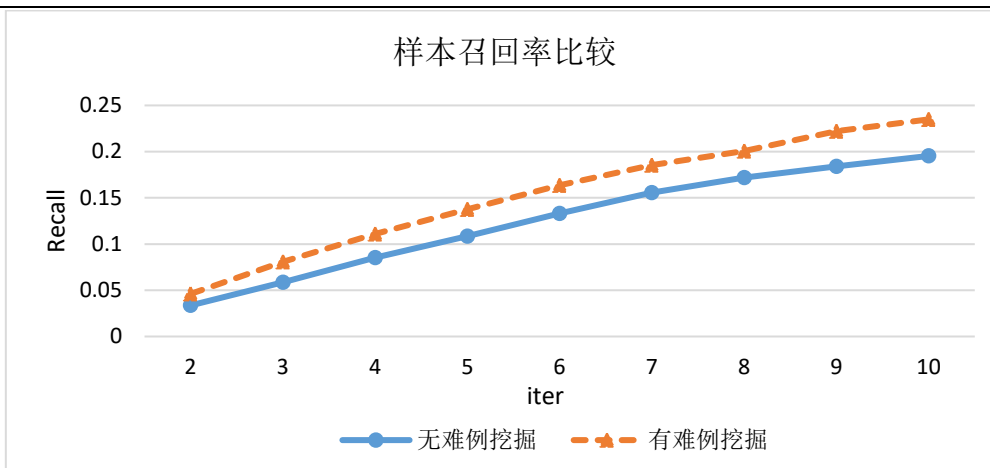


图 4-12 难例挖掘样本召回率比较

由图 4-11 和图 4-12 所示，实验结果可得，不论是从样本挖掘的多样性还是准确性来看，难例挖掘方法都有利于这两方面。

#### 4) 整体实验

将本研究框架与 Zhu<sup>[13]</sup>的监督学习框架进行比较。使用车辆数据集，包含 250 幅带图像标注正例图像和 58 幅测试图像。Zhu<sup>[13]</sup>的方法使用了所有训练集车辆样本，共 3686 个。本课题标注了数据集中 8 幅图像，占训练集图像数的 3.2%，样本数量占训练集总样本数量的 5.4%。

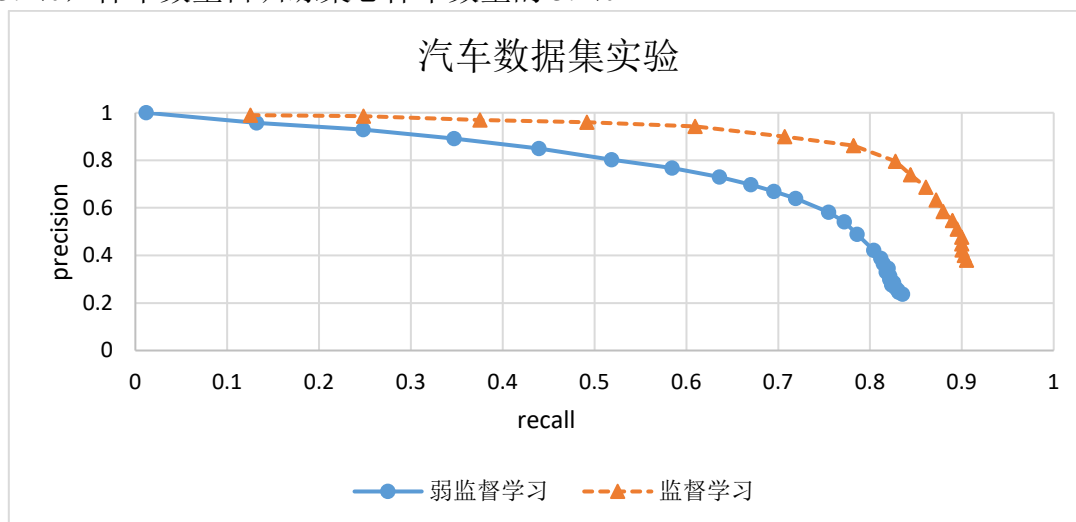


图 4-13 弱监督学习与监督学习的汽车实验性能对比

由实验结果可以见，虽然本框架的性能与监督学习方法的性能还有一段差



距，但是本框架所使用的是图像级别的标注和极少数的种子点样本。其中，本实验所使用的种子点样本数仅占监督学习方法的 5.4%，减少了 94.6% 的样本标注工作。



图 4-14 本框架在汽车数据集检测结果示例

如图 4-14 所示为本章框架在汽车数据集上的部分检测结果。结果显示框架

虽然存在部分误检和漏检问题，但总体上仍能完成目标检测工作，满足一定的检测要求。本框架优势在于仅使用了图像级别的标注，通过极少数标定的样本作为种子点样本来初始化正例训练样本并参与难例挖掘的过程，从而使得样本挖掘和分类器训练的迭代更新更加准确有效。从检测结果来看，本框架主要问题是存在一定的部件误检为目标的现象。除了前一章分析的在弱监督学习中目标部件与目标整体在正例图像的共现性和在反例图像中的非现性导致之外，汽车数据集本身也比飞机数据集难度更大。这是由于飞机出现的场景较为单一，主要以机场为主，其背景元素大多为机场跑道、机场建筑、大面积草地等；另一方面，车辆的场景较为复杂，可能出现在于城市、乡村、高速公路、停车场等地方，容易受诸如建筑、树木、道路、十字路口等物体干扰。同时，由于受车辆的用途决定，其存在的位置和方向存在巨大的不确定性，这些因素都对分类器的学习和检测的结果产生干扰。

#### 4.4 本章小结

本章介绍的实验框架是在第三章框架基础上新引入了种子点样本。在遥感图像目标检测中，由于遥感图像范围大、物体众多、不同物体间因角度、光照、形变等影响，容易互相干扰，相似物体间的干扰尤甚。这些现实问题对样本的挖掘和分类器的学习带来更大的挑战，可能使样本挖掘和分类器训练向错误的方向收敛而达不到学习目的。本章引入样本种子点概念，以极少数样本标注作为种子点，引导样本的初始化和挖掘过程，从而保证学习的有效性。本章从问题分析、框架介绍、实验分析及结果等几个部分来讲述。

问题分析中阐述了遥感图像目标检测在弱监督学习中面临的复杂问题，采取了折中的方法，引入了极少数样本种子点的概念。借用极少量的标注信息来大幅提升样本挖掘的准确性。

框架小节内容包括在引入样本种子点后将其应用到初始化、难例挖掘等步骤中去。初始化采用 **K-means** 的聚类方法，再选择最可能性最大的聚类类别，从该类中选择样本。难例挖掘阶段将种子点样本周围符合设定条件的区域当作难例参与模型训练。

实验分析与结果部分针对几个提出的方法和步骤进行了实验比较并对结果进行分析。通过实验证明本框架提出的几个改进步骤都能有效提升分类器的学

习和效果。最后通过与监督学习的方法相比，展示检测结果，证明本框架的可行性。同时，通过对检测结果的分析，从弱监督学习自身的问题和目标特性两方面总结了当前检测结果仍然存在不足的原因。



## 第5章 结论与展望

遥感图像具有覆盖范围广、信息复杂、角度独特的特点。在遥感图像中的目标检测具有广泛的应用领域和切实的研究价值。现有的遥感图像目标检测方法大部分都是基于监督学习，其缺点在于需要耗费大量人力物力在标注样本数据上。随着未标注数据越来越多、获取方式代价越来越低，如何利用未标注样本达到学习目的成为一个重要的研究方向。本文提出基于弱监督学习的遥感目标检测算法，其核心在于利用简单易得的极弱标注数据学习分类器达到遥感图像中检测目标的目的。

本文首先论述了遥感图像目标检测的研究背景和意义，总结了国内外研究进展和存在的问题。第二章中介绍了本课题相关技术，包括弱监督学习概念及算法、感兴趣区域提取、特征提取以及本文核心的多示例学习和协同训练算法。第三章中介绍了本研究的基于多示例学习的弱监督遥感图像目标检测算法框架，重点讲述了多示例学习和协同训练结合的样本挖掘过程以及二者在传统算法基础上的改进。为了保证样本挖掘的正确性，避免分类器学习过程中的误化，还加入样本优化过程。实验证明上述步骤都达到了预期的效果。第四章则是在第三章基础上针对更复杂的遥感场景和更小的目标而引入了样本种子点的概念。其不同点在于，为了正确引导样本的挖掘和分类器的学习而以极小的代价加入了标注样本作为种子点样本。在此基础上，进行了初始化和难例挖掘等步骤的改进。研究中的实验表明，以付出极弱标注代价达到接近监督学习方法的性能，这便是弱监督学习的核心目的。综上所述，本文的主要贡献在于：

1) 针对传统监督学习标注代价高的缺点，提出了基于弱监督学习的遥感图像目标检测算法。将多示例学习与协同训练方法相结合，使计算机能够从图像级标注数据中自动挖掘正例样本，迭代更新分类器和训练样本，达到分类器的学习的目的；

2) 为了更加贴合遥感图像场景应用，将传统的多示例学习方法改为扩充式训练样本挖掘，提高了图像内区域的利用率；将半监督学习中的协同训练算法改为基于属性集划分的弱监督式协同训练，能有效提高挖掘出的样本的多样性。

3) 在多示例学习中的 MI-SVM 方法与协同训练算法相结合外，加入了简单高效的样本优化步骤进一步提高了样本挖掘的准确性，避免迭代过程中误化分

类器的学习;

4) 针对更为复杂的遥感场景, 提出了基于样本种子点的遥感图像目标检测算法。将极少量的标注样本作为种子点, 应用在样本初始化和难例挖掘过程中, 保证样本挖掘的准确性和分类器迭代训练的有效性。

虽然本文提出的框架基本能够达到以极弱标注代价完成分类器训练和遥感图像目标检测功能, 但由于时间关系, 仍然存在如下不足:

1) 本课题框架研究在自制的 UCAS-AOD 数据集上实验。然而, 现实遥感图像场景受天气、地区、硬件条件等诸多因素影响互相之间有很大不同。本文只对高分辨光学遥感图像进行尝试, 没有在其他遥感图像数据集上进行实验。

2) 为了保证样本挖掘的多样性和准确性, 充分利用未标记数据。本文提出了扩充式样本挖掘和优化的方式, 但二者方法过于简单粗糙, 可以进一步细致的设计。

将来的遥感目标检测算法可以向以下两方面进行研究: 一是, 遥感图像目标检测实时性; 特别是国防军事、灾难检测领域, 需要实时快速地检测, 有时需要对新目标快速建模检测。所以, 除了模型训练和检测阶段的迅速高效之外, 还需要不能过于依赖监督信息; 二是, 同场景下的多类目标检测比单类目标问题更加复杂, 需要更优秀的模型来区分不同目标物体, 该研究在无人机、智能监测等领域都是重要的核心技术。

## 参 考 文 献

- [1]程圣军. 基于带约束随机游走图模型的弱监督学习算法研究[D]. 哈尔滨:哈尔滨工业大学, 2014.
- [2]孙显, 付琨, 王宏琦. 高分辨率遥感图像理解[M]. 北京: 科学出版社, 2011.
- [3] Zhao T, Nevatia R. Car detection in low resolution aerial images[J]. *Image and Vision Computing*, 2003, 21(8): 693-703.
- [4] Moon H, Chellappa R, Rosenfeld A. Performance analysis of a simple vehicle detection algorithm[J]. *Image and Vision Computing*, 2002, 20(1): 1-13.
- [5] Hinz S. Detection and counting of cars in aerial images[C]// 2003 International Conference on. IEEE, 2003, 3: III-997-1000 vol. 2.
- [6] S.Wang. Vehicle Detection on Aerial Images by Extracting Corner Features for Rotational Invariant Shape Matching, 11th IEEE International Conference on Computer and Information Technology ,pp171-175,2011
- [7] Tsai L W, Hsieh J W, Fan K C. Vehicle detection using normalized color and edge map[J]. *Image Processing, IEEE Transactions on*, 2007, 16(3): 850-864.
- [8] Cheng H Y, Weng C C, Chen Y Y. Vehicle detection in aerial surveillance using dynamic Bayesian networks[J]. *Image Processing, IEEE Transactions on*, 2012, 21(4): 2152-2159.
- [9] Choi J Y, Yang Y K. Vehicle detection from aerial images using local shape information[M]. *Advances in Image and Video Technology*. Springer Berlin Heidelberg, 2009: 227-236.
- [10] Kembhavi A, Harwood D, Davis L S. Vehicle detection using partial least squares[J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2011, 33(6): 1250-1265.
- [11] Zheng H, Pan L, Li L. A morphological neural network approach for vehicle detection from high resolution satellite imagery[C]// *Neural Information Processing*. Springer Berlin Heidelberg, 2006: 99-106.
- [12] Eikvil L, Aurdal L, Koren H. Classification-based vehicle detection in high-resolution satellite images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2009, 64(1): 65-72.
- [13] Zhu H, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]// *Image Processing (ICIP), 2015 IEEE International Conference on. IEEE*, 2015: 3735-3739.
- [14] Cheriadat A M. Unsupervised feature learning for aerial scene classification [J]. *Geoscience and Remote Sensing, IEEE Transactions on*, 2014, 52(1): 439-451.
- [15] Yao J, Zhang Z. Semi-supervised learning based object detection in aerial imagery[C]// *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE*, 2005, 1: 1011-1016.

- [16] Yao J, Zhang Z. Object Detection in Aerial Imagery Based on Enhanced Semi-Supervised Learning[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2010:1011-1016 vol. 1.
- [17] Zhong L, Wu J, Tang X, et al. The semi-supervised classification method of LiDAR data integrating with aerial images[C]//Information Science and Engineering (ISISE), 2010 International Symposium on. IEEE, 2010: 386-390.
- [18] Han J, Zhang D, Cheng G, et al. Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(6):3325-3337.
- [19] Zhang D, Han J, Yu D, et al. Weakly Supervised Learning for Airplane Detection in Remote Sensing Images[J]. IEEE Geoscience & Remote Sensing Letters, 2015, 12(4):155-163.
- [20] Zhou P, Zhang D, Cheng G, et al. Negative Bootstrapping for Weakly Supervised Target Detection in Remote Sensing Images[C]// IEEE International Conference on Multimedia Big Data. IEEE, 2015:318-323.
- [21] 张敏灵. 多示例学习与多标记学习的研究[D]. 南京大学, 2007.
- [22] Rosenberg C, Hebert M, Schneiderman H. Semi-supervised self-Training of object detection models[C]//Proceedings of IEEE Workshops on Application of Computer Vision (WACV/MOTION'05). Breckenridge, USA: IEEE, 2005:29-36.
- [23] Joachims T. Transductive inference for text classification using support vector machines [C] //Proceedings of ACM International Conference of Machine Learning. Bled, Slovenia: ACM, 1999 ,99:200-209.
- [24] Joachims T. Transductive Inference for Text Classification using Support Vector Machines[C]// Sixteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 1999:200-209.
- [25] Zhu X. Introduction to Semi-Supervised Learning[J]. Synthesis Lectures on Artificial Intelligence & Machine Learning, 2009, 3(1):130.
- [26] Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.
- [27] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks.[J]. Science, 2015, 313(5786):504-507.
- [28] Zouxy09. Deep Learning (深度学习) 学习笔记整理系列 [EB/OL]. 2013-04-08. <http://blog.csdn.net/zouxy09/article/details/8775360>.
- [29] Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex[J]. Journal of Physiology, 1968, 195(1):215-243.
- [30] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of



- pattern recognition unaffected by shift in position[J]. *Biological Cybernetics*, 1980, 36(4):193-202.
- [31] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- [32] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Advances in Neural Information Processing Systems*, 2012, 25(2):2012.
- [33] LISA lab. Convolutional Neural Networks (LeNet) [Z]. 2008.  
<http://deeplearning.net/tutorial/lenet.html>
- [34] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. *Artificial Intelligence*, 1997, 89(s 1–2):31-71.
- [35] Foulds J, Frank E. A review of multi-instance learning assumptions[J]. *Knowledge Engineering Review*, 2010, 25(1):1-25.
- [36] N. Weidmann, E. Frank and B. Pfahringer, A two-level learning method for generalized multi-instance problems, *Proc. Eur. Conf. Mach. Learn.* (2003) pp. 468–479.
- [37] Wang J, Zucker J D. Solving the Multiple-Instance Problem: A Lazy Learning Approach.[J]. *Proc.international Conf.on Machine Learning*, 2010:1119-1126.
- [38] Ray S, Page D. Multiple Instance Regression[C]// *In Proceedings of the 18th International Conference on Machine Learning*. 2001:425--432.
- [39] Andrews S, Tsochantaridis I, Hofmann T. Support Vector Machines for Multiple-Instance Learning[J]. *Advances in Neural Information Processing Systems*, 2002, 15(2):561-568.
- [40] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]// *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 1998:92-100.
- [41] Nigam K, Ghani R. Analyzing the effectiveness and applicability of co-training[J]. *Cikm*, 2002, 33(2):86-93.
- [42] Goldman S A, Zhou Y. Enhancing Supervised Learning with Unlabeled Data[C]// *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., 2000:327-334.
- [43] Zhou Z H, Li M. Tri-Training: Exploiting Unlabeled Data Using Three Classifiers[J]. *Knowledge & Data Engineering IEEE Transactions on*, 2005, 17(11):1529-1541.
- [44] 周志华,王珏. 半监督学习中的协同训练风范[C]//*机器学习及其应用*,北京: 清华大学出版社,2007.
- [45] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014:580-587.
- [46] Maria Isabel Restrepo. PAMView: Providence Aerial MultiView Dataset [DB/OL].  
[http://www.lems.brown.edu/~mir/helicopter\\_providence/sites.html](http://www.lems.brown.edu/~mir/helicopter_providence/sites.html)

- [47] Tanner F, Colder B, Pullen C, et al. Overhead imagery research data set—an annotated data library & tools to aid in the development of computer vision algorithms[C]//Applied Imagery Pattern Recognition Workshop (AIPRW), 2009 IEEE. IEEE, 2009: 1-8.

致谢

---



## 致 谢

转眼间，在中国科学院大学攻读硕士学位的三年学习生活即将结束。与刚入学相比，我收获了不仅仅是专业上的知识，还有良师益友，还有自信与勇气。在毕业论文的尾声，在此借这个机会衷心地感谢我的老师、实验室的兄弟姐妹以及我的家人。

本课题的研究工作及学位论文是在焦建彬教授、叶齐祥教授、韩振军副教授、秦飞副教授的悉心指导下完成的。首先，我要感谢焦老师给予珍贵的学习机会，感谢在我硕士期间每一次的指导和鼓励；感谢叶老师在我研究的选题、框架设计、论文撰写等方面提出指导和建议，至今仍记得您曾说“科研是没有标准答案的”，感谢您在我对自己毫无信心的时候给予了鼓励；感谢韩老师在理论知识和论文写作的过程中给予的指导和帮助，以及您在生活中对每个学生无微不至的照顾；感谢秦老师用严谨的思考和认真的态度以及开阔的眼界影响着我。衷心地感谢各位老师三年来对待学生的真心实意，用实际行动诠释春风化雨，使得实验室的每一位学生都在愉快而严谨的氛围中学习和成长。

在学三年，我同样要感谢朝夕相处的实验室的兄弟姐妹们。首先，感谢魏鹏旭师姐在我刚进入实验室对科研感到迷茫的时候，指导我学习许多计算机视觉和机器学习领域的基础知识，鼓励我要对自己有信心；感谢陈孝罡师兄和朱海港师兄在我课题研究一筹莫展的时候，与我一同探讨、共度难关。感谢李策师姐、崔妍婷师姐、高山师兄、柯炜师兄以及其他师兄师姐，给予我学习上的指点，生活中的无私帮助。感谢我同届好友李兆举、张天亮、万方、王忻雷，我们一同成长，一起克服困难，也一起度过美好的时光。感谢实验室的每一位成员，因为有了你们，我的三年生活才能如此多彩斑斓。

特别要感谢我的家人以及王泽伟，感谢这么多年的陪伴，你们永远是最强大的依靠！希望他们永远平安康健！

感谢参与我开题和中期评审的每一位老师，你们丰富的经验和无私地帮助我把握论文方向和研究进度。

戴蔚群  
2015年4月



## 个人简历、在学期间发表的论文与研究成果

姓名：戴蔚群 性别：女 出生日期：1992年11月15日 政治面貌：中共党员

### 教育经历：

- 2009年9月至2013年7月 厦门理工学院 软件工程 学士
- 2013年9月至2016年7月 中国科学院大学 工业工程 硕士

### 曾获荣誉：

- 中国科学院大学优秀学生（2014年）
- 国家奖学金（2012年）
- 省级高校三好学生（2012年）

### 研究成果：

- Zhu H, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]// Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015: 3735-3739.
- 《弱监督航拍图像目标检测系统》软件著作权（受理中）。  
受理号：2016R11S075802