



华南理工大学
South China University of Technology

专业学位硕士学位论文

高分辨率遥感图像目标检测方法研究

作者姓名	张 锰
学科专业	电子与通信工程
指导教师	余翔宇 副教授
	卢斌 高级工程师
所在学院	电子与信息学院
论文提交日期	2021 年 4 月

Study on objects detection algorithm in high-resolution remote sensing image

A Dissertation Submitted for the Degree of Master

Candidate: Zhang Kun

Supervisor: Prof. Yu Xiangyu

South China University of Technology

Guangzhou, China

分类号：TP391

学校代号：10561

学 号：201821011613

华南理工大学硕士学位论文

高分辨率遥感图像目标检测方法研究

作者姓名：张锐

指导教师姓名、职称：余翔宇 副教授 卢斌 高级工程师

申请学位级别：工程硕士

学科专业名称：电子与通信工程

研究方向：智能信息处理

论文提交日期：2021年4月28日

论文答辩日期：2021年5月30日

学位授予单位：华南理工大学

学位授予日期： 年 月 日

答辩委员会成员：

主席：贺前华

委员：倪江群 胡永健 余翔宇 章烈剽

华南理工大学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。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后果由本人承担。

作者签名：张巍

日期：2021年6月1日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、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华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；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。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

保密（校保密委员会审定为涉密学位论文时间：____年____月____日），于____年____月____日解密后适用本授权书。

不保密，同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文编入有关数据库进行检索，传播学位论文的全部或部分内容。

（请在以上相应方框内打“√”）

作者签名：张巍 /
手写

日期：2021.6.1

指导教师签名：

日期 2021.6.1

作者联系电话：18813298234

电子邮箱：18813298234@163.com

联系地址（含邮编）江西省新余市分宜县钤山镇行山村中社背村小组15号

邮编：336600

摘要

高分辨率遥感图像的目标检测是一项具有挑战性的任务。尽管有许多基于卷积神经网络（Convolutional Neural Network, CNN）的先进方法在自然图像中取得了不错的效果，遥感图像中的目标检测进展得却不是那么顺利。不同于自然图像，遥感图像中的目标具有任意方向、密集分布和尺度差异大的特点，导致会出现一系列如特征不对齐、误检漏检和大长宽比目标检测不佳的问题。针对上述问题做出的相应改进工作如下：

(1) 考虑到自然图像中的水平框目标检测方法在对密集分布且方向任意的目标进行检测时容易出现的漏检和误检问题，采用了相对于水平框有一定偏移角度的有方向框代替水平框，有方向框在密集分布场景下相邻框之间只包含对应目标，不会发生重叠区域过大的情况。

(2) 提出了一种改进的多方向两级级联 R-CNN 方法来更好的检测遥感图像中具有任意方向的小目标。首先将区域提议网络（Region Proposal Network, RPN）得到的水平感兴趣区域转换成旋转感兴趣区域，旋转区域可以更好地提取待检测目标的特征，之后再对旋转感兴趣区域作进一步回归得到目标的精确位置。在第一级区域转换网络设计了多方向 RoI 对齐模块从多个不同旋转方向的水平感兴趣区域获取方向敏感特征，同时在回归分支添加方向注意力模块自适应地给每个方向通道的特征赋予权重，加强特征的方向敏感性。

(3) 针对候选区域位置的变化带来的特征不对齐问题，设计了基于可形变卷积的多分支特征对齐模块来重新采样特征，同时采用了不同扩张率的空洞卷积来获取不同尺度的感受野。另外提出了基于目标长宽比的角度偏移惩罚损失函数来缓解大长宽比目标对于角度偏移更加敏感的问题，在训练过程中更加关注大长宽比目标角度偏移量的学习。

在 DOTA 和 HRSC2016 这两个公开数据集上得到的消融实验以及和其它先进方法的对比实验结果都验证了本文提出算法的有效性。

关键词：高分辨率遥感图像；目标检测；多方向 RoI 对齐；特征对齐；长宽比

目 录

摘要	I
第一章 绪论	1
1.1 课题背景	1
1.2 拟解决的关键问题	2
1.3 数据集	3
第二章 改进的多方向级联目标检测方法研究	6
2.1 引言	6
2.2 算法设计	6
2.2.1 有方向检测框	6
2.2.2 两级级联检测器结构	8
2.2.3 多方向 RoI 对齐模块	9
2.2.4 方向注意力模块	11
2.3 本章小结	13
第三章 基于特征对齐和目标长宽比的目标检测方法研究	14
3.1 引言	14
3.2 算法设计	15
3.2.1 空洞卷积	15
3.2.2 多分支特征对齐模块	16
3.2.3 基于目标长宽比的角度偏移损失函数	18
3.3 本章小结	20
第四章 总结与展望	21
参考文献	23

第一章 绪论

1.1 课题背景

随着航空航天技术的不断发展，遥感技术水平也在不断革新。由于遥感技术的优越性，遥感技术已广泛应用于多个领域。未来十年，随着遥感图像的各项指标的提高，遥感技术有望进入实时快速地提供各类地球观测数据的新阶段。在遥感图像的基础上，目标识别的信息处理技术是当今自动目标识别的关键技术之一，也是遥感信息提取的核心所在。它在军事和民用领域都具有重要的应用意义和研究价值。自动目标识别技术可以从遥感图像的复杂背景中自动提取目标特征，并根据特定区域和典型目标的特征模板数据库，或利用边缘、灰度、纹理结构等信息实现检测、拦截、识别和跟踪目标。虽然图像处理、模式识别和计算机视觉技术的快速发展为遥感图像目标识别技术的改进创造了条件，但目前遥感图像目标识别技术还不够成熟，与实际应用的要求仍存在较大差距，还有许多问题亟待解决。

首先，遥感图像中的目标具有多样性和复杂性，即遥感图像存在着丰富的信息，待检测的目标的类型和结构复杂多样。它包括河流、湖泊、森林等自然物体，以及建筑物、公路、居民区等人造物体。同时，在遥感图像，待检测物体与其他物体之间会出现重叠等现象，这给遥感图像的目标检测和识别带来了困难。其次，遥感图像中噪声、光照变化、云雾的干扰可能导致同类目标的类内差异增大，不同类型目标的类间差异减小，从而降低了目标的识别精度，给自动识别带来困难。此外，遥感图像的内容复杂，目标来源多样，仅采用低阶特征提取方法无法充分准确地表达遥感图像的目标，限制了遥感图像目标识别的准确性。最后，图像语义信息的处理技术还不够成熟，低层次特征与高层次语义信息难以结合，缺乏有效的先验信息，制约了目标识别精度的进一步提高，这对遥感图像目标识别技术的研究提出了更高的挑战。

目标检测作为计算机视觉中的重要任务，可以从高分辨遥感图像所涵盖的场景中对目标进行准确定位和类别划分，有助于更好的理解图像，在军事战争、农业生产以及城市规划建设等方面都能发挥关键作用。传统的目标检测技术依靠人工设计出的特征算子来进行特征提取，之后再通过滑动窗口的方式用分类器对图像各个子区域进行分类判断，面对简单图像场景下的特定目标检测可以达到不错的效果。但是高分辨率遥感图像中的背景复杂，目标和背景之间的区分度低，并且需要检测的目标种类多样，传统目标检测技术存在这很大的局限性。近年来深度神经网络的兴起以及在图像处理领域的卓越

表现吸引了极大的关注，深度神经网络不需要针对特定目标设计特征算子，而是依靠网络自身学习自动化地从图像中提取有用的特征用于后续处理，并且不需要对图像区域进行重复计算，相比于传统方法具有更好的泛化性能，不论是性能还是效率都获得了大量的提升。因此，将深度神经网络用于高分辨率遥感图像目标检测是实现遥感信息自动化处理的重要开端。

1.2 拟解决的关键问题

从理论的角度分析，与通用的图像目标检测问题相比，遥感图像目标检测具有以下待解决的关键问题：

1. 小目标问题：本文所提出的遥感目标检测方法，主要以可见光遥感图像作为重点研究对象。众所周知，可见光遥感图像一般借助于航空航天设备如卫星、无人机等获得，因此其拍摄位置往往位于数千米高空之中，而特别地，在常见的遥感图像待检测目标中，有诸多尺寸较小的目标如小型车辆 (Small Vehicle)、船只 (Ship) 等，其在遥感图像当中所占尺寸远小于其他目标，同时相比于通用目标检测中的类似目标，其在遥感图像中所占据的像素尺寸依然是较小的，显然地，在遥感图像的目标检测方法中，这一类目标将比网球场 (Tennis Court)、码头 (Harbor) 等大型目标更加难以检测，如何提高对此类小型目标的识别与检测精准度与效率就成为算法设计中一个不可忽视的问题。

2. 目标密集性问题：由于遥感图像取自于高空之中，其往往尺寸较大，能覆盖到非常广阔地面范围，包含有大量的地面建筑与设施，显然待检测目标并不会均匀地分散在整个图像当中，而是在某些设施中，可能会聚集着大量的待检测目标，最常见的当属停车场、码头与机场，在这三类地面建筑与设施中，分别会聚集有大量的小型车辆、舰船与飞机，体现在遥感图像当中，即这些目标会集中在一片小型区域之中，这就是遥感图像的目标密集性问题。由于现有的检测方法往往是基于“检测 + 识别”的思路，在检测阶段，过于密集的目标会使得目标检测框产生过分堆积，若不加以鉴别区分，则很有可能让原本正确的检测框在后续处理中因过多重叠被剔除，进而影响检测效率与精准度。

3. 尺度多样性问题：在遥感图像中，存在各种尺度的待检测目标，对于多类型的目标检测方法而言，待检测目标间的尺度多样性成为必须考虑的问题。在传统单一目标检测方法中，会使用预先定义好的输入图像，结合预先设置的尺度网络，得到固定尺寸的特征图像与后续矢量。但在遥感图像中，既有足球场这种大型的目标，也有十字路口、

网球场等中型的目标，更多的，还是车辆船只等小型的目标，倘若只使用单一固定尺度大小的检测网络，必然会导致网络对某一类尺度目标的响应远大于其他目标，这是由神经网络的工作方式所限制的，因此，为了得到尽可能好的目标检测效果，必须考虑网络对于多尺度目标的检测方法，即网络必须使用多尺度结构以满足不同目标的检测与识别需求。

4. 视角特殊性问题：常见的通用目标待检测图像往往是基于地面水平方向获取得到，如人脸检测、车辆检测、文字检测等，因此在设计此类通用目标检测方法之时，多会从目标的水平视角特征入手进行检测，但对于高空遥感图像而言，其为俯视角，以车辆检测为例，在水平视角车辆检测方法中，会针对车辆水平外观特征进行学习，然而在遥感图像中，这一特征是不可获取、不可知的，因此需要针对遥感图像中的待检测目标选取更加符合其俯视视角的特征进行学习，这就涉及到对网络感受野与结构，必须依据这一特殊性进行设计。高分辨率遥感影像中的目标通常具有尺度变化较大、视角特殊、方向多变和背景复杂等特点，当前通用的目标检测和识别方法对于此类任务的表现较为一般，体现在其检测精度不高、适应性差，难以胜任遥感目标检测与识别任务。因此，针对遥感图像的多尺度、多方向目标，提出具有针对性的目标检测与识别方法是非常必要的。

5. 方向多样性：遥感图像获取时的俯视视角让所有目标近乎是在同一水平线上，处于不同位置的目标会产生不同的旋转角。而自然图像中的目标如行人、人脸和车辆等都是和水平面垂直分布，一般具有确定的方向。以人脸检测为例，人脸五官在空间方向上往往具有一定的规律，而这种方向上的规律也可以被用作检测中的一类特征；行人、车辆等目标大多都是垂直于水平面的；遥感图像检测则需要对目标的不同方向具有特征适应性。

面对上述提出的关键性问题，课题组内研究人员分别进行了相关的理论研究，并对得出的解决方案进行了详细的实践验证。本文以下章节将会对提出的解决方案分别进行详细的论述。

1.3 数据集

随着深度学习方法在遥感图像目标检测之中的发展与应用，各类遥感数据集也应运而生。针对遥感图像中有方向目标的检测方法研究，本文使用了两个常见的遥感图像公开数据集 DOTA 和 HRSC2016 来进行对比实验。

在 DOTA 数据集提出之前，已有各种遥感数据集可供研究者使用，其中使用较为广泛的有 HRSC2016[44] 数据集，HRSC2016 是由西北工业大学刘子坤团队制作的专门用来进行船舶检测的数据集，图像来源于 GoogleEarth，总共有 1061 张图片，包含了海上船舶和近海船舶在内的两种不同场景。它的图片分辨率大小跨度没有 DOTA 大，从 300×300 到 1500×900 ，并且大部分图片都是在 1000×600 左右，数据集中大约有 436 张图片作为训练集，181 张图片为验证集，剩余的 444 张为测试集，比例大致为 4:2:4。数据集图片中的船舶目标也都是由旋转框来进行标注的，所有图片都进行了标注，大部分的船舶目标长宽比值都比较大，达到了 7:1。

DOTA 数据集是武汉大学夏桂松与华中科技大学白翔等人在 2018 年共同提出的多目标遥感数据集，其主要数据来源为 GoogleEarth 和两颗中国卫星 GF-2, JL-1，共包含有 15 类待检测目标与共计 2806 张遥感图像，这 15 类目标分别为飞机、船只、储蓄罐、棒球场、网球场、篮球场、田径场、码头、桥梁、大型车辆、小型车辆、直升飞机、足球场、环形路口、游泳池，在每一幅图像中都含有大量的实例，针对实例采用了矩形框进行标注，按顺时针方向依次记录标注框坐标，同时对目标的检测难易度进行了划分。在 2019 年，DOTA 数据集进行了填充与改进成为 DOTA1.5，其检测目标由原来的 15 类增加到 16 类，添加了对于集装箱起重机的标注信息，同时补全了部分在原数据集中缺漏的实例标注。

在使用 DOTA 数据集之前，首先必须针对数据集的特殊性进行分析，这样有助于对数据进行预处理，避免引入无关因素影响检测与识别方法的性能，DOTA 数据集具有如下几个差异性：

1. 样本分布差异。虽然在 DOTA 数据集中对多类目标进行了标注，但由于实际遥感场景的限制，在数据集中，部分较大尺度目标数量稀少如足球场，而小型车辆类目标确广泛地存在于每幅遥感图像之中，这就导致小型车辆目标数量过多，进而在汇集其标注信息时，容易造成训练时的溢出与显存不足 (Out Of Memory, OOM) 问题；

2. 尺度分布差异。在 DOTA 数据集中有数千幅遥感图像，但其原始尺寸从 800×1000 到 9000×6000 皆有，存在显著的尺度分布差异，而神经网络往往只对特定尺寸范围的输入具有良好响应。因此，在使用 DOTA 数据集进行实验与评估时，一种确保网络输入尺寸一致性的方法是很有必要的；

3. 采样距离差异。与 OGST(Oil and Gas Storage Tank Dataset) 数据集 [47] 等具有固定采样距离的数据集不同，由于 DOTA 使用的遥感图像来源多样，因此其标注中带有

不同的采样距离参数，另外还存在部分图像采样距离缺失。值得注意的是，采样距离与分辨率共同决定了一幅遥感图像所覆盖的实际地面区域大小，因此，在进行地面对象检测时，这两类因素都将对检测目标尺度产生影响。

针对这些差异性，在文章中分别使用了针对性的处理方法进行优化，如针对小型目标过多导致的 OOM 问题，可以使用筛选与增广的方式调节网络在一次学习中所使用的标注真实样本数量；针对尺度的多变，可以使用裁剪与拼合的方式进行一致性处理；针对采样距离的差异，则设计了对采样距离敏感的卷积网络结构来进一步引入预测信息辅助识别。

第二章 改进的多方向级联目标检测方法研究

2.1 引言

在高分辨率遥感图像中，由于获取图像时的俯瞰视角，待检测的目标在图像中是密集分布和方向多样的，并且由于获取位置往往是几千米的高空，图像中的小目标最小到只有十几个像素大小。基于传统水平框检测的目标检测方法在面对高分辨率遥感图像时会碰到很大的问题，相邻目标的水平检测框之间会存在大部分的重叠，一方面会对检测过程中候选区域框的特征提取带来干扰，另一方面在使用非极大值抑制对部分检测框进行滤除时可能会过滤掉一些相邻目标的正确预测框，极大的影响了检测目标的召回率。因此，如何应对高分辨率遥感图像中小目标密集分布且任意朝向的特点，设计改进的检测方法，提升模型表现，成为了亟待解决的问题。

本章提出了一种改进的多方向级联 R-CNN 目标检测方法，设计了一种两级级联的检测网络，第一级检测网络完成水平感兴趣区域到旋转感兴趣区域的转换，第二级检测网络对旋转区域进行特征提取做进一步的检测框的精确定位。在第一级检测结构提出了多方向 RoI 对齐结构获取方向敏感特征用于回归，同时在回归分支设计了方向注意力模块来对方向信息进行增强。另外考虑到分类和回归任务对于方向敏感的不一致性，在分类分支将方向敏感特征做了平均池化来获取适合分类的方向不变特征。在本章最后进行了对比实验来验证所提出的算法的有效性。

2.2 算法设计

图2-1是本章提出算法的总体结构，展示了在基于区域提取的 Faster R-CNN 算法上所做的改进。在特征提取阶段，将 ResNet 作为基础网络，为了应对高分辨率遥感图像中的小目标和尺度差异大的问题，使用了 FPN 结构将低层的定位信息和高层的语义信息进行融合加强特征表现。在检测阶段，将原有的水平框检测器替换为本章提出的两级级联有方向检测器，包括使用了有方向检测框代替了传统的水平检测框、设计的多方向 RoI 对齐模块和方向注意力模块。

2.2.1 有方向检测框

近年来自然图像中的基于水平框检测的目标检测方法取得了巨大的成功，在检测精度和速率上都有了长足的进步。不同于自然图像，水平框检测在面对高分辨率遥感图像会碰到这样的问题：图2-2所示为同一副图像中真实标注的水平框和有方向框之间的对

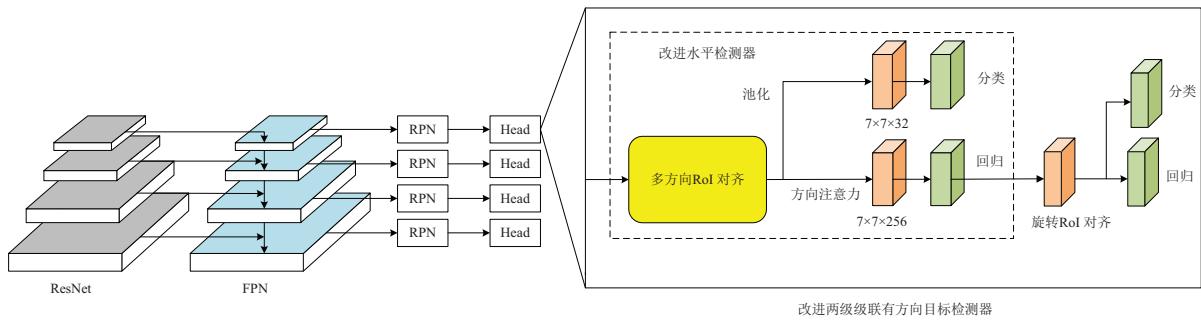


图 2-1 改进的多方向级联 R-CNN 目标检测方法结构示意图

比情况，当采用2-2 (a) 所示的水平框对密集分布的小目标进行检测时，相邻目标之间的检测框非常容易发生重叠，在检测阶段进行非极大值抑制滤除部分检测框时，阈值难以界定，过低的阈值会使得很多有效的检测框因为和相邻目标的检测框之间有重叠而被滤除，过高的阈值会使得最终的检测结果中遗留下了很多的实际效果不好的检测框，严重影响到了目标的召回率和精确率，进而降低了模型表现。而当采用图2-2 (b) 所示的有方向检测框进行检测时，相邻目标之间的检测框的重叠情况得到了很好的改善，非极大值抑制阶段的问题可以很好的解决，能够有效提升检测性能。

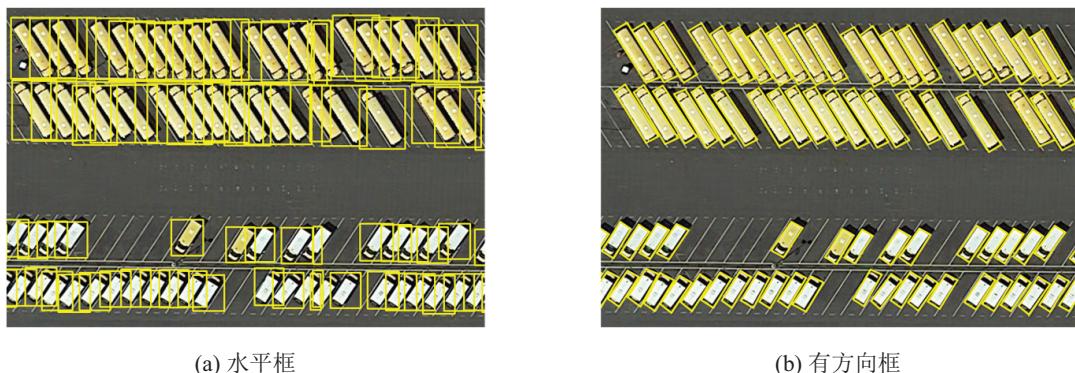


图 2-2 水平检测框和有方向检测框对比

如图2-3 (a) 所示，基于自然图像的目标检测方法采用水平框对目标位置进行定位时使用的参数为 (x, y, w, h) ，其中 (x, y) 指的是框的中心， w 和 h 指的是水平框的宽和高，训练阶段将这 4 个参数量的偏移作为回归分支的预测值进行训练。而本章的方法中使用了额外的角度参数来定义有方向检测框，如图2-3 (b) 所示，有方向检测框的参数表示为 (x, y, w, h, θ) ，其中角度 θ 表示的是水平线与检测框长边之间的夹角，此定义下的 θ 范围为 $(-\frac{\pi}{2}, \frac{\pi}{2})$ ，当 $\theta = 0^\circ$ 时有方向检测框也就变成了水平检测框，在本章的两级级联的检测网络中都采用了有方向检测框的参数作为回归分支的预测输出。

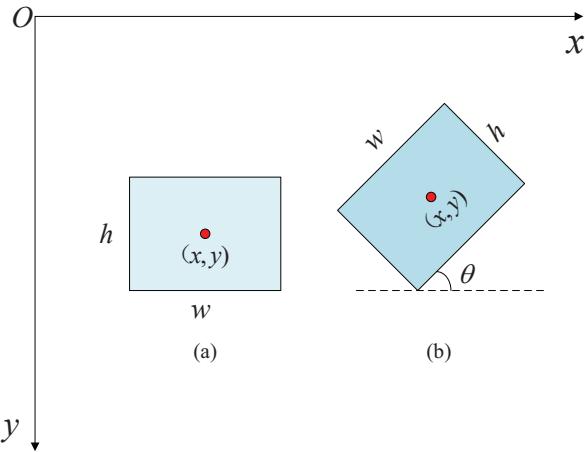


图 2-3 两种检测框表示方法: (a) 水平框; (b) 有方向框

2.2.2 两级级联检测器结构

一般情况下，一些基于水平框检测的目标检测方法首先使用深度神经网络对输入图片进行特征提取，然后在候选区域生成阶段特征图每个位置点会生成不同尺度和长宽比大小的锚点框作用于 RPN，对锚点框的位置作进一步的调整得到水平感兴趣区域，之后通过使用单级检测器在分类和回归分支对感兴趣区域中的目标位置进行预测。这样的单阶检测器在水平框检测上可以达到比较好的检测效果，但是在进行有方向检测框时，仅仅使用单阶检测器从水平感兴趣区域框中获取待检测有方向目标的特征并对目标的位置作进一步的定位效果欠佳，所提取的特征不仅不能很好地对有方向目标进行表示而且在一些情况下会被周围同类目标的特征干扰。基于这样的问题有一些研究学者采用在候选区域生成阶段生成额外的不同方向的锚点框作为候选区域来表示有方向目标框的位置^{[35][55][56]}，这些方法虽然可以提升有方向框目标的检测精度，但是在候选区域生成阶段所产生的锚点框数量较之前有成倍数的增长，又由于 RPN 网络的正负样本划分和非极大值抑制阶段需要计算两个锚点框之间的 IoU，产生的的计算量相较于之前的方法有指数级别的增长。过多计算资源的消耗导致在训练阶段对一些包含较多检测目标的图片进行训练时很容易出现内存溢出的情况，一方面会使得一部分的训练数据得不到充分的学习，另一方面计算量的显著增长使得模型的训练和检测速度非常缓慢，对检测精度和检测效率都有不好的影响。

为了解决上述提到的问题，本文提出了一个两级级联的检测结构来进一步实现对有方向框的检测。这样的两级级联的检测结构不需要在候选区域生成阶段生成带不同方向的锚点框来得到有方向目标位置的准确特征，而是先采用了一个分类和回归检测网络结

构对从 RPN 网络得到的水平感兴趣区域进行调整，筛选出包含目标的兴趣区域并根据获取的兴趣区域特征进行第一次有方向框检测回归得到包含目标的旋转感兴趣区域的初步位置，然后再使用另外一个分类和回归检测网络结构，以初步获取得到的旋转感兴趣区域的特征作为输入，进一步对该区域内有方向目标所在的准确位置进行预测，通过这样一个两级级联的检测结构，不仅可以解决因为生成数目过多的锚点框导致的内存溢出问题，而且经过从水平感兴趣区域到旋转感兴趣区域再到有方向目标精确位置这样一个递进的回归过程，一步步地提升了获取的目标位置的特征质量，有效的改善了有方向检测框的检测精度。

2.2.3 多方向 RoI 对齐模块

在基于区域提取的目标检测方法中，对候选区域的特征提取是检测环节中极为重要的一环，只有提取出能够很好的表示候选区域中所包含目标的特征才能让后续的分类和回归任务更精确地进行。在自然图像的目标检测算法中经常用到的特征提取方法为 RoI 对齐，将从 RPN 网络中获取得到的水平感兴趣区域通过划分单元格、特征点采样和双线性插值的过程输出为固定尺度大小的特征图，以此作为每个感兴趣区域的特征表示，再将其输入到分类和回归分支对每一个感兴趣区域是否包含目标以及目标的具体位置进行学习预测。

然而，不同于自然图像，由于高分辨率遥感图像中目标密集分布的特性，在进行有方向检测框的任务时对于感兴趣区域的特征提取有了更高的要求。如图2-4 (a) 所示，红色实线框为标注的有方向检测框，黄色虚线框为标注的水平检测框，绿色虚线框为 RPN 网络中输出的水平感兴趣区域框，结合图2-4 (a) 和图2-4 (b) 能够看出，RPN 输出的兴趣区域框与实际标注的水平框之间存在偏移，一方面没有将待检测目标完全包括在内，另一方面由于目标的密集分布特性，待检测目标相邻的同一类的目标也被包含在内。如果使用之前提到的传统 RoI 对齐提取特征会忽略掉这样的问题，导致提取出的特征并不能很好的表示目标并且会被相邻目标的特征所干扰，输入到分类和回归分支的特征受到影晌不能进行充分的学习，进而影响了对感兴趣区域中有方向目标的进一步定位。

针对上述的问题，在本章的方法中提出了多方向 RoI 对齐模块，如图2-4 (c) 所示，不直接对 RPN 网络中输出的水平感兴趣区域进行特征提取，而是先将该感兴趣区域进行多个不同角度的旋转，得到形状大小一样但是方向不同的带旋转角度的区域，也就是图2-4 (c) 中的蓝色实线框。可以看到经过一定角度的旋转之后，带旋转角度的区域可以更好的包括待检测目标，对于相邻的同类目标也有一定的滤除作用，并且不同角度方向

的区域对于同一目标能提取到的特征也不一样，使得各个方向上的旋转区域可以更好的表示出待检测目标的特征。之后再在对应通道的输入特征图单独提取每个方向上的旋转区域特征并拼接在一起组成方向敏感的目标特征表示，具体的操作流程如下：

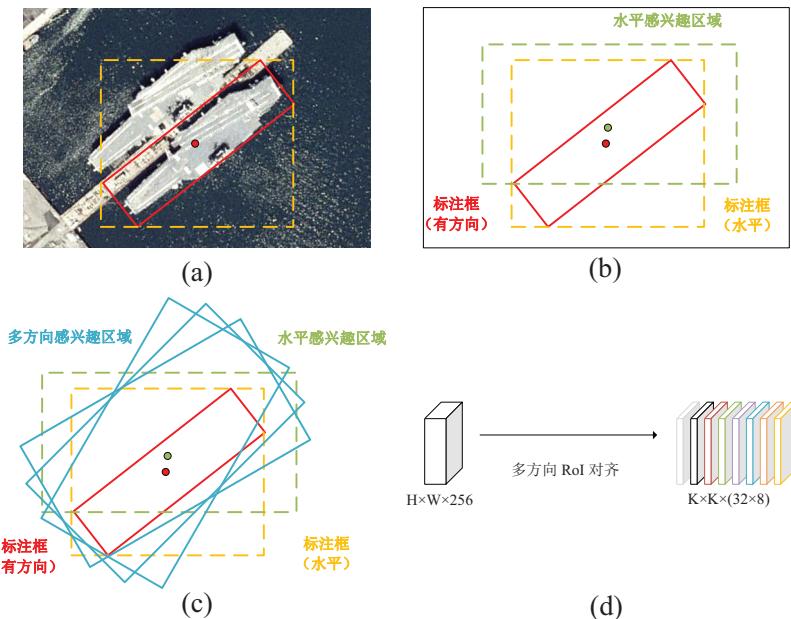


图 2-4 多方向 RoI 对齐的思路流程：(a) 带标注框的原始图像；(b) 水平感兴趣区域与有方向目标标注框之间的偏移；(c) 不同角度旋转水平感兴趣区域；(d) 多方向 RoI 对齐输出

给定大小为 $H \times W \times C$ 大小的特征图和水平感兴趣区域框 (x, y, w, h) , (x, y) 表示框的中心点坐标, w 和 h 分别表示框的宽和高, 多方向 RoI 对齐首先在此基础上生成 N (N 默认为 8) 个不同方向上的旋转感兴趣区域框, 然后使用旋转 RoI 对齐提取每个方向上框的特征, 最后的输出特征为 $K \times K \times (\frac{C}{N} \times N)$, $K \times K$ 表示的是我们将旋转框分成的单元格数目, 通常来说 K 的大小为 7, 因此, 对于每个索引为 $(i, j)(0 \leq i < K, 0 \leq j < K)$ 的单元格, 提取出的特征为

$$y_{c^n}(i, j) = \sum_{(x_h, y_h) \in bin(i, j)} F_{i, j, c^n}(\varphi(x_h, y_h)) / s_{i, j} \quad (2-1)$$

$$\theta = -\frac{\pi}{2} + n \frac{\pi}{N}, c^n \in [n \frac{C}{N}, (n+1) \frac{C}{N}), n = 0, \dots, N-1 \quad (2-2)$$

其中 F_{i, j, c^n} 表示的是对应通道为 c^n 的特征图上索引为 (i, j) 的单元格特征, 对于输入的 $H \times W \times C$ 大小的特征图, 先按特征图总的通道数 C 将输入特征图按顺序平分为 N 份, 在进行特征提取时每一份特征图对应的旋转框的方向也不同, 比如说 $C=256$, $N=8$, 则 $c^1 \in [0, 32)$, 对应的角度 $\theta = -\frac{\pi}{2}$, 即通道数为 0 到 32 的特征图提取特征时旋转框的角度为 $-\frac{\pi}{2}$ 。 $s_{i, j}$ 是每一个单元格内的采样点数目, 一般情况为 4, 然后对于水平框中每一个

采样点 (x_h, y_h) , 先要通过转换方程 φ 将其转换成对应方向的旋转框上的相应位置 (x_r, y_r) 来表示, 转换方程如下:

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \varphi(x_h, y_h) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_h - w/2 \\ y_h - h/2 \end{pmatrix} + \begin{pmatrix} x \\ y \end{pmatrix} \quad (2-3)$$

其中 (x, y, w, h) 分别表示的是水平感兴趣区域的中心位置、长和宽, θ 为对应框的旋转角度, 在得到每个单元格内需要采样的点的位置之后, 每个采样点位置的特征值通过双线性插值的方法从邻接位置获取, 最后再对所有采样点的特征值求平均得到每个单元格最后的特征表示输出。

基于上述的结果, 经过多方向 RoI 对齐模块, 我们可以从 RPN 输出的水平感兴趣区域框提取得到包含 N 个方向的方向敏感特征, 同时考虑到分类任务和回归任务对于方向敏感的不一致性, 我们只将方向敏感特征用于回归分支。而对于分类分支, 我们旨在通过平均池化方向敏感特征来获取方向不变特征, 这可以将 N 个方向通道的特征值取平均来简单完成。分类分支的方向不变特征 x_{cls} 的计算方式如下:

$$x_{cls} = \frac{1}{N} \times \sum_{n=0}^{N-1} y_{cn} \quad (2-4)$$

通过上述的步骤, 我们可以得到多方向的方向敏感特征和方向不变特征分别作为回归和分类分支的输入特征。需要注意的是方向不变特征相比于之前有着更少的参数, 假定输入的特征图大小为 $H \times W \times 256$, 采用的方向数目为 8, 则分类分支的方向不变特征的大小为 $K \times K \times 32$, 与之前的 $K \times K \times 256$ 相比显著减少了后续操作所需的计算量。

2.2.4 方向注意力模块

虽然说我们可以通过多方向 RoI 对齐模块提取多个不同方向上的旋转框的特征来更好地对待检测的有方向目标进行特征表示, 但是我们也不能保证每一个方向上的特征都是绝对有效的。比如说当待检测目标的旋转角度为 0° , 也就是水平框时, 旋转角度过大的旋转框相对于目标位置的偏移会使得提取到的特征不能很好的表示待检测目标, 如果以同等贡献来定义该特征和其它方向上的特征在一定程度上会降低对于此类目标的检测性能。

针对上述问题, 本文提出了方向注意力模块来量化每个方向上的特征的贡献, 对于不同方向的待检测目标, 每个方向上的特征所做的贡献也不相同。图2-5展示本文提出的方向注意力模块的结构图, 受 SE-Net (Squeeze and Excitation Network)^[57] 的启发, 我

们想对每个方向通道上的特征进行 0-1 之间的量化，以此来表示每个方向上的贡献。与 SE-Net 不同的地方在于，我们将全局平均池化结构替换了组卷积结构，把属于同一个方向的通道特征分为一组，每一组分别进行卷积得到该方向上的特征表示。例如对于输入到回归分支的 $K \times K \times C$ 大小的特征图，将 C 个通道的特征分为 N （与方向数目相等）组，每一组的通道数为 $\frac{C}{N}$ ，组卷积的卷积核大小为 $K \times K$ ，再将 N 组经过组卷积之后的特征拼接在一起，这样处理之后输出的特征图大小为 $1 \times 1 \times C$ 。之后再经过两层全连接层获取每个方向通道上的量化贡献值，并将量化贡献值和输入的方向敏感特征相乘获取最后的输出，最终经过方向注意力模块得到的输出可以表示为

$$F_{out} = s * F = \sigma(W_2 \delta(W_1 \underset{K \times K, group=N}{groupconv}(F))) * F \quad (2-5)$$

其中 σ 表示的是 Sigmoid 激活函数， δ 表示的是 ReLU 激活函数， W_1 表示的是维度下降的全连接层，维度衰减参数为 r （默认为 16），与之相反 W_2 表示的是维度上升的全连接层，最终维度保持一致， $*$ 表示的是将每个方向通道的贡献 s 和输入特征图 F 相乘，也就是最后的输出特征图。

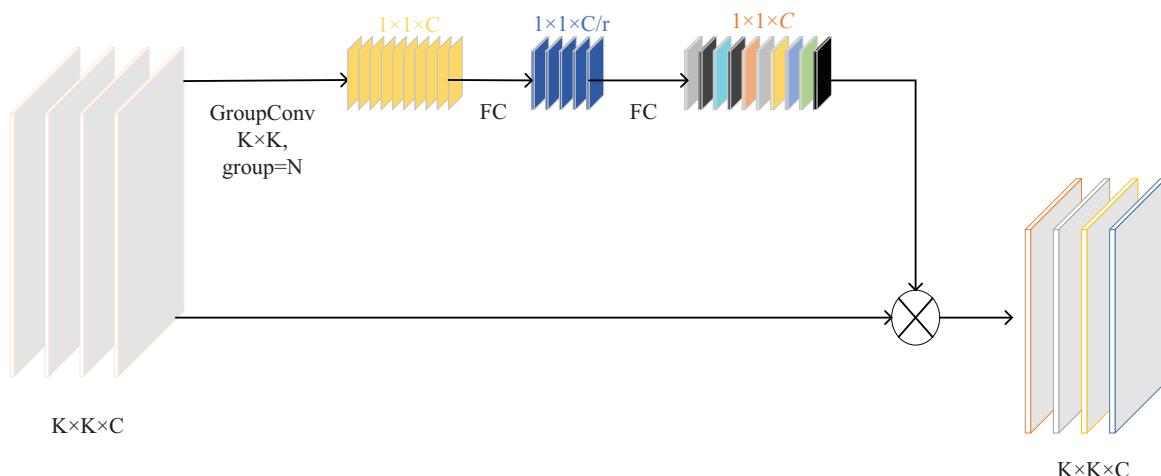


图 2-5 方向注意力模块网络结构

通过上述的方式，方向注意力模块从本质上可以实现对输入特征进行动态的选择，可以看成是对每个方向通道上的特征做了自注意力机制。将多方向的方向敏感特征作为输入，每个方向的特征自己进行卷积得到该方向的贡献表示来量化每个方向上的重要性，赋予重要性低的方向小的权重，而赋予重要性高的方向相对大的权重，可以根据待检测目标在图像中的方向自适应地调整每个方向的贡献，进一步提升对于方向多样的目标的检测精度。

2.3 本章小结

由于遥感图像获取时的特殊视角，待检测目标在图像中具有密集分布并且任意朝向的特点，影响了模型的检测表现。针对这样的问题，本章提出了一种两级级联的多方向目标检测方法，增加了一个额外的检测网络来转换得到能更精确表示有方向目标位置的兴趣区域，缓解了密集分布导致的直接检测时分类分支得分和回归分支定位精度不一致的问题，同时在该检测网络设计了多方向 RoI 对齐模块获取不同方向区域的池化特征，并在回归分支结合方向注意力模块自适应地对每个方向上的特征进行选取，最后得到的方向敏感特征对于方向多样的目标的回归更加友好。最后通过各个模块之间的消融实验分析以及和现有的几种方法之间的对比实验结果验证了本章提出方法的有效性。

第三章 基于特征对齐和目标长宽比的目标检测方法研究

3.1 引言

在基于区域提取的目标检测方法中，检测过程由两部分组成，第一部分先在特征图每个位置生成不同尺度和长宽比的候选区域，经过 RPN 调整候选区域位置得到可能包含目标的感兴趣区域，第二部分是由分类和回归分支组成的检测网络，逐个对 RPN 得到的感兴趣区域提取特征，输入检测网络进一步对是否存在目标和目标精确位置进行判断。在一般情况下 RPN 和检测网络的输入图像特征图是一样的，这样的设定没有考虑到生成的候选区域和 RPN 调整得到的感兴趣区域之间的位置发生了偏移，导致特征点位置也发生了变化，使用相同的特征图会使得提取的感兴趣区域的特征不够准确，影响后续的检测网络。这样的问题在第三章提到的改进的多方向级联 R-CNN 目标检测方法中更为明显，该方法采用了两级级联的检测结构，第一级检测网络负责将 RPN 得到的水平感兴趣区域转换为旋转的兴趣区域，第二级检测网络再对得到的旋转区域进行池化。旋转区域和水平区域之间特征点位置的变化远大于候选区域和水平感兴趣区域之间，特征不对齐问题更加严重。

另外，我们在基于区域提取的有方向目标检测基础方法 Faster R-CNN 和第三章提出的方法的研究中发现，模型对于不同长宽比的检测目标得到的有方向检测框的结果有明显的差距，在其它参数保持不变的情况下密集分布的小长宽比的目标的漏检率会远远小于大长宽比的目标，而且小长宽比的目标得到的检测框和数据集中标注的框之间的 IoU 值也比大长宽比的目标要好。这意味着模型在训练过程中没有对不同长宽比之间的角度偏移参数进行充分平等的学习，小长宽比目标的角度偏移量学习在实际情况下的表现更好。

针对上述的问题，本章提出了一种基于特征对齐和目标长宽比的方法，首先利用了可形变卷积具有显著地几何形变建模能力的特性，基于可形变卷积设计了一个多分支结构来进行特征对齐，并且通过不同扩张率的空洞卷积对感受野进行扩增以适应不同尺度的检测目标，其次在训练模型对损失函数进行了改进，以目标长宽比比值作为控制系数，让模型在训练阶段在角度偏移的学习上更加注重大长宽比的目标，以此来提高模型对于大长宽比目标的检测精度。

3.2 算法设计

图3-1所示为本章提出的方法的整体结构图，所示的结构是基于第三章所提出的方法做出的改进，特征提取阶段采用 ResNet 和 FPN 提取多尺度特征，为了应对旋转感兴趣区域和水平感兴趣区域之间的特征不对齐问题，在输入第二级检测网络之前，先采用了多分支特征对齐模块对特征图调整，使得提取的旋转感兴趣区域的特征能够更好的表示被检测的有方向目标，同时在两级的检测网络结构中都采用了基于目标长宽比的损失函数来加强模型对于不同长宽比目标的检测精度。

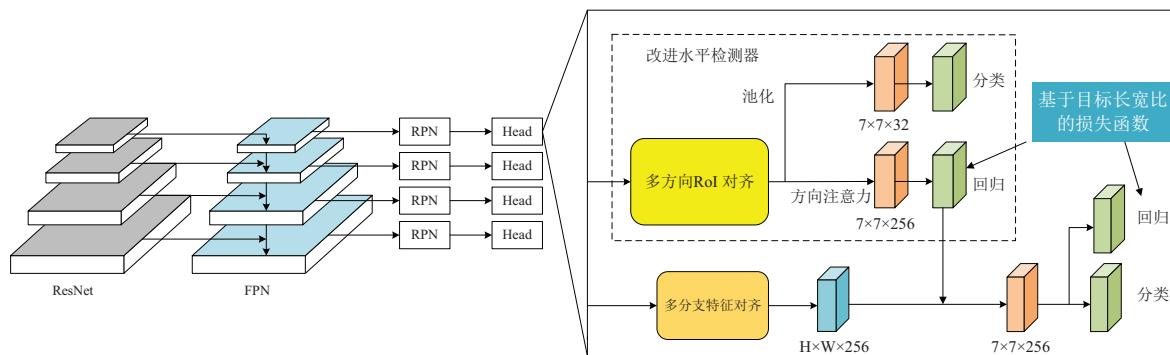


图 3-1 第 4 章算法总体结构

3.2.1 空洞卷积

在 CNN 中，感受野的定义是 CNN 中每一层的输出特征图上的特征点在输入图像中映射的区域大小，即特征图上的一个点对应输入图像上的一个区域。在传统的 CNN 网络结构设计中，主要是通过在卷积操作中采用大的卷积核或者多个小的卷积核进行堆叠以及池化操作降低特征图的尺寸来达到增大感受野的作用，这样的做法一方面大卷积核和堆叠的小卷积核都产生了额外的计算量，另一方面池化操作会使得空间信息丢失，尤其是小目标来说，在经过几次池化操作之后目标的特征信息甚至不能由特征图上的特征点来表示。

为了能够在不产生额外的计算量和损失空间信息的基础上进行感受野的扩增，Yu 等人提出了空洞卷积（dilated convolution）^[59]，和标准的卷积操作相比，空洞卷积在卷积时多定义了一个表示扩张率的超参数 d 来对卷积核进行间隔采样， d 表示的是卷积核进行卷积时各个位置的间隔距离，当 $d = 1$ 时即为正常卷积，图3-2分别展示了卷积核尺寸为 3×3 ， d 分别为 1、2、3 时的空洞卷积的示意图，红色的圆点为卷积时卷积核在特征图上的采样位置，图3-2 (a) 所示的 $d = 1$ 时的空洞卷积和普通卷积一样，图3-2 (b) 所

示为 $d=2$ 时的空洞卷积，可以看到通过对卷积核间隔采样， 3×3 大小的空洞卷积核可以达到 5×5 大小的普通卷积的效果（红色圆点所围成的区域大小），同理图3-2 (c) 中的 $d=3$ 时的空洞卷积可以达到 7×7 大小的普通卷积的效果。在进行空洞卷积的过程中，特征图的大小自始至终都没有产生变化，没有造成空间信息上的损失，并且在达到 7×7 的感受野也只是使用一个 3×3 大小的卷积核，没有产生额外的计算量。在应对遥感图像中目标的多尺度问题上，可以使用多个并行的不同扩张率大小的空洞卷积层来获得不同的感受野。

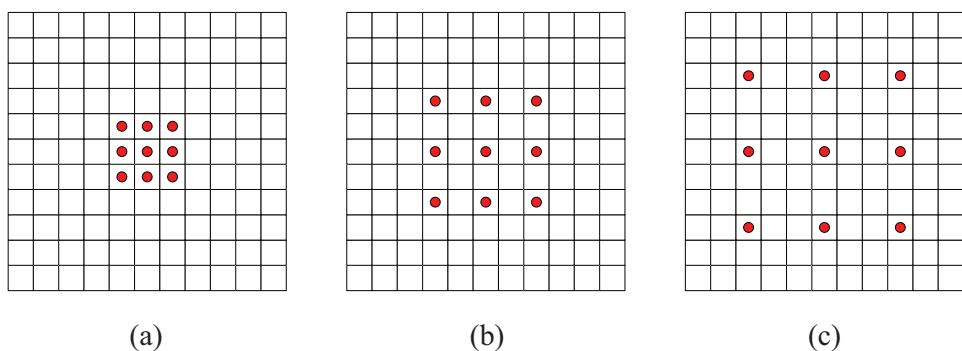


图 3-2 不同扩张率空洞卷积示意图：(a) 扩张率 $d=1$; (b) 扩张率 $d=2$; (c) 扩张率 $d=3$

3.2.2 多分支特征对齐模块

在第三章的改进的多方向级联 R-CNN 目标检测方法中，第一级检测结构对水平感兴趣区域进行了调整，从水平区感兴趣域转变成了旋转感兴趣区域，此时第二级检测结构需要对旋转感兴趣区域进行特征提取做进一步分类和回归，相对于第一级检测结构的水平感兴趣区域，旋转感兴趣区域需要采样的特征点在水平和垂直方向上都有了一定的偏移，如果仍然使用在第一级检测结构中所用到的特征图，采样特征点上的特征值不能很准确的来表示旋转感兴趣区域内的待检测目标。当然，这样的一个位置点的偏移可以看成是几何形变的一种，而可形变卷积^[32] 在形变问题上有着显著地表现，因此可以将可形变卷积用来处理这样的特征不对齐问题。

可形变卷积在模块中增加了对于空间采样位置偏移量的学习，来提高对于几何形变的建模能力。如图3-3所示，对于大小为 $H\times W\times C$ 大小的输入特征图，假定卷积核的大小为 $K\times K$ ，可形变卷积通过一个平行的卷积结构分支对卷积核在输入特征图上的采样点的偏移进行学习，该分支输出的参数大小为 $H\times W\times(K\times K\times 2)$ ，每一次进行卷积得到输出特征图上的点时，卷积核中每个采样点的位置相对于普通卷积在水平和垂直方向都有了一个偏移量，可以实现在当前位置附近随意采样而不是局限于之前的规则格点（±

K), 能够自适应地去解决因为形变带来的特征点位置偏移的问题。

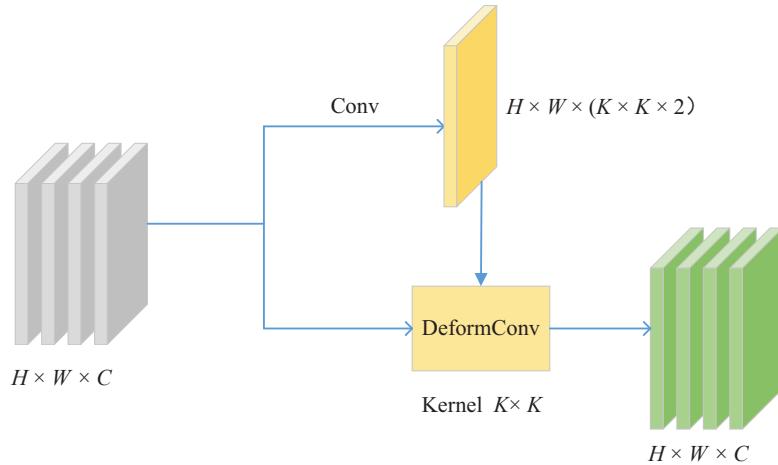


图 3-3 可形变卷积结构图

基于以上描述, 图3-4展示了本节提出的多分支特征对齐模块的结构。对于从 ResNet 和 FPN 主干网络每一层中提取出的特征图, 首先通过多分支特征对齐模块自适应地对特征进行重新采样, 该模块由三条平行的可形变卷积分支组成, 每个分支的卷积核大小为 3×3 , 为了应对遥感图像中目标的多尺度特点, 分别使用了扩张率为 1、2、3 大小的空洞卷积来获取不同大小的感受野, 之后将三个分支得到的对齐后的特征拼接在一起, 最后再通过一个 1×1 大小卷积核的结构将通道数调整成与主干网络特征图通道数目相同的输出, 将该输出作为新的特征图对感兴趣区域进行特征提取, 得到准确的有方向目标的特征表示, 进一步提升网络的检测表现。

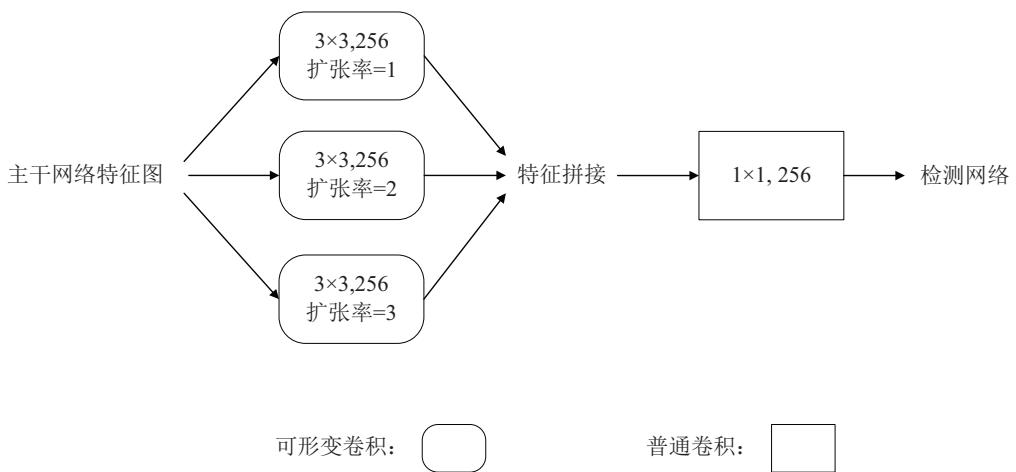


图 3-4 多分支特征对齐模块总体结构

3.2.3 基于目标长宽比的角度偏移损失函数

在传统的基于区域提取的有方向目标检测方法中，为了得到预测的有方向框的中心坐标和尺寸，一般不直接在回归网络对坐标和尺寸大小进行直接预测，而是将标注的有方向框和 RPN 网络生成的兴趣区域之间的偏移量作为预测，这样的做法一来可以得到更稳定的损失函数，二来可以加快模型的收敛。在回归分支，兴趣区域不参与模型的训练过程，只在计算与标注框之间的偏移量编码以及对预测偏移量解码得到最后的输出结果用到，以第一级检测网络为例，对于每一个标注的有方向框 $(x_g, y_g, w_g, h_g, \theta_g)$ ，通过计算与 RPN 生成的兴趣区域之间的 IoU 确定得到对应的作为正样本的兴趣区域 $(x_a, y_a, w_a, h_a, 0^\circ)$ ， 0° 表示为水平的兴趣区域，此阶段的偏移量编码为

$$\begin{aligned} t_x &= (x_g - x_a)/w_a, t_y = (y_g - y_a)/h_a \\ t_w &= \log(w_g/w_a), t_h = \log(h_g/h_a), t_\theta = \theta_g - \theta_a \end{aligned} \quad (3-1)$$

其中 $t_x, t_y, t_w, t_h, t_\theta$ 分别为回归分支需要预测的中心点位置，长，宽和角度的偏移量，假定最后得到的预测框为 $(x_p, y_p, w_p, h_p, \theta_p)$ ，则解码得到的回归分支的偏移量输出为

$$\begin{aligned} t'_x &= (x_p - x_a)/w_a, t'_y = (y_p - y_a)/h_a \\ t'_w &= \log(w_p/w_a), t'_h = \log(h_p/h_a), t'_\theta = \theta_p - \theta_a \end{aligned} \quad (3-2)$$

为了可以更加稳定地进行训练，一般来说回归网络的损失函数都是使用 Smooth-L1 函数，该函数的定义为

$$L_{SmoothL1loss} = \begin{cases} 0.5(t - t')^2 & if |t - t'| < 1 \\ |t - t'| - 0.5 & otherwise \end{cases} \quad (3-3)$$

其中 t 和 t' 分别为真实的偏移量和模型预测的偏移量。然而在使用上述 Smooth-L1 损失函数进行有方向目标检测方法的实验中，存在了这样的一个问题：由于不同长宽比目标对于角度偏移的敏感性不一致，导致在对训练好的模型进行测试时小长宽比目标的检测精度要好于大长宽比的目标。图3-5展示了不同长宽比目标之间发生的不平衡现象，黄色的框表示的是标注的有方向框，蓝色的框表示的模型最后输出的预测框，红色的框表示的是两个框之间的交集区域，图中对应的黄色框和蓝色框之间有着相同的中心点，长和宽，唯一不同的地方在于角度值，并且图3-5 (a) 和图3-5 (a) 中的两个框之间角度的偏移量也是一样的。在这样的情况下，通过 Smooth-L1 损失函数计算得到的训练损失是

一样的，但是图3-5 (a) 和图3-5 (b) 之间的 IoU 值相差却极大，很大程度上影响到了最终的检测表现，图3-5 (a) 中的小长宽比的目标有更高的 IoU 值，在测试阶段该预测框可以被很好的检测到，而图3-5 (b) 中的大长宽比的目标的 IoU 值过小导致在测试阶段该预测框会被过滤掉，对于此类目标得不到最终的预测结果，影响了目标的检测精度。

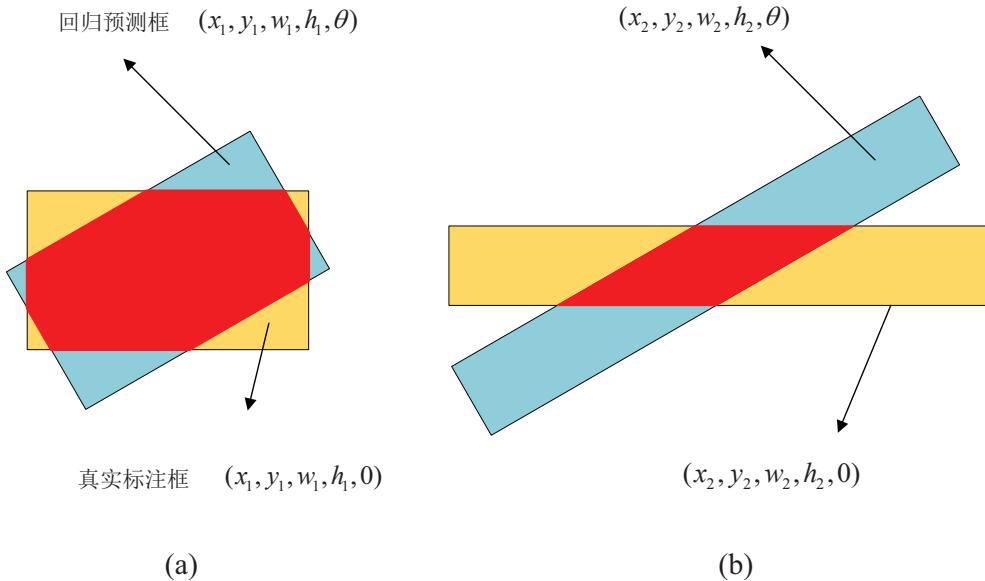


图 3-5 不同长宽比目标对角度偏移的敏感差异：(a) 长宽比为 2:1 的目标；(b) 长宽比为 7:1 的目标

为了解决上述的问题，依据大长宽比目标对角度偏移更加敏感的特点，本章提出了一种基于目标长宽比的损失函数，旨在训练阶段让模型更加关注大长宽比目标的角度偏移量的学习，所以在原有的损失函数的基础上添加了额外的对于角度偏移的惩罚项，最终的损失函数定义如下：

$$L_{reg} = L_{SmoothL1loss} + |t_\theta - t'_\theta| * (\ln r - 0.5) \quad (3-4)$$

其中 r ($r \geq 1$) 表示的是目标的长宽比值，我们观察到对于长宽比特别小的目标 (r 接近于 1)，只要中心点位置、长和宽的预测准确，角度偏移的准确程度对于最终的检测结果并没有那么重要。因为对于长和宽近乎相等的目标来说，即使有一些角度偏移不会很大影响到最后的 IoU 值，该目标同样可以在测试阶段被很好的检测到，结果没有受到影响，所以在惩罚项的设计里设定了一个 0.5 的阈值来确定长宽比值的临界条件，对于长宽比小于该阈值的目标，降低了模型对于其角度偏移量预测的关注度，让模型更加关注于其他偏移量的预测，而对于长宽比大于该阈值的目标，在训练阶段让模型更加关注对于其角度偏移量的学习。

由于遥感图像中的目标长宽比值跨度极大，数据集中的一些目标的长宽比最大可以达到 30:1，为了平衡网络对于不同长宽比目标的检测精度，将上述的损失函数应用到了之前提出的两级检测网络的两个回归分支当中。除此之外，分类分支采用的损失函数为常见的交叉熵损失函数。基于此，本章提出的网络结构的总体损失函数定义如下：

$$L = \frac{1}{N_1} \left(\sum_{n=1}^{N_1} L_{cls} + s'_n L_{reg} \right) + \frac{1}{N_2} \left(\sum_{n=1}^{N_2} L_{cls} + s'_n L_{reg} \right) \quad (3-5)$$

N_1 和 N_2 分别表示两级检测网络中的判定为正样本的感兴趣区域的数量， s'_n 是一个二进制值， s'_n 为 0 当该感兴趣区域分类为背景区域， s'_n 为 1 当该感兴趣区域分类为目标区域，即对于背景区域不需要计算回归分支的损失。

3.3 本章小结

本章从基于区域提取的目标检测方法中的“两步走”策略带来的特征不对齐问题出发，设计了基于可形变卷积的多分支特征对齐模块来对特征重新采样，同时考虑了不同长宽比目标对于角度偏移的敏感差异，设计了基于目标长宽比的角度偏移惩罚损失函数，让模型在训练过程中更加关注大长宽比目标角度偏移的学习。实验结果证明，本章设计的模块显著地提升了遥感图像中一些密集分布和大长宽比目标的检测精度。

第四章 改进的深度级联遥感目标检测网络

4.1 引言

在将通用目标检测中的两步法方法迁移至遥感目标检测任务中时，会受到来自遥感图像多种特殊性质的影响：其一是遥感图像的尺度特殊性，其往往具有极大的分辨率，但传统检测却只能针对小型尺度图像进行分析，无法直接处理具有特大尺度的遥感图像；其二是待检测遥感目标的类间多尺度特性，由于人类活动和建筑的限制，存在诸多小目标聚集的部分，同时这些目标往往在整幅遥感图像中只占有非常小的尺度，而与此同时却还存在着占用较大像素区域的部分场地型目标，显然一个单一尺度的特征提取网络是无法同时应对同一图像中不同尺度的目标的；其三则是遥感图像中目标的密集性，以车辆目标为例，这一类目标会大量存在于公路路口、停车场等位置，这导致目标相互间隔极小，如果在检测时定位不够精细，非常容易将不属于目标的其他噪声与背景作为输入，影响了网络的泛化性能与定位精度。因此不考虑图像和目标尺度适应性的网络往往难以达到与通用目标检测任务中的接近的检测性能与检测效率，其卷积性能也受到了遥感目标多方向性、密集性的影响而更难以提取出具有强鲁棒性的特征图，而鉴于输入图像所具有的前背景信息的迅速增加，在通用目标检测中使用的网络深度难以完全达到最佳检测效果。因此，在本章中，对通用目标检测网络中难以适应遥感检测任务的场景进行分析，并探讨网络性能受到限制的原因，并最终设计了一种深度级联的遥感目标检测网络，以适应尺度多样性的遥感多目标检测任务，同时还基于遥感图像与遥感目标特点，添加了更为针对多方向性目标的可形变卷积结构，从而实现了对遥感图像具有针对性的目标检测网络设计，进而提升检测精度。

4.2 算法介绍

本章提出了一种改进的深度级联遥感目标检测网络，其整体框架如图??所示。在 Faster-RCNN 网络的基础上，改进了特征提取、候选框选取与分类检测的网络结构，使得其对遥感图像具有良好的适应性。网络的具体检测动作流程为：首先将遥感图像进行尺度标准化预处理后，输入 backbone 网络中进行特征提取，为了确保能够保有图像尽可能多的位置信息与语义信息，使用了一个深度网络 ResNet-50 作为 backbone 网络；在进行特征提取时，进一步地利用可形变卷积 DCN^[?] 进行修正，进而获得更加贴合目标的卷积感受野，减少区域内无关背景的特征噪声干扰；然后将 ResNet-50 网络中 Conv2_

`x`、`Conv3_x`、`Conv4_x`、`Conv5_x`的输出层的作为不同尺度的卷积特征图输入 FPN 结构中，FPN 经过自上而下与自底向上的过程后，保留了特征的浅层位置信息与深层语义信息，经由 RPN 网络进行候选框提取与池化；最终，在检测网络中，构建了深度级联的网络结构，层层级联提高门限值并最终输出检测结果。

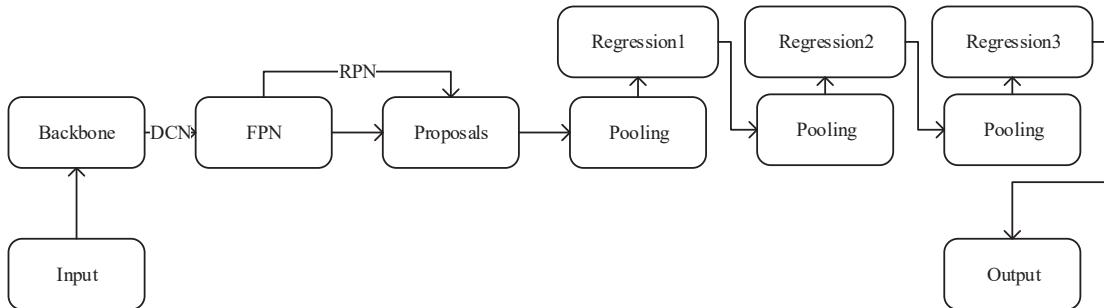


图 4-1 网络整体框架

4.2.1 特征金字塔

FPN 特征金字塔结构是一种适用多尺度目标检测的网络结构，在传统目标检测网络中往往使用多层卷积层逐层卷积的结构，其最终输出为一个较小尺度的多通道特征图块，由于通用目标检测中场景往往为局部小区域，同时目标间尺度差异较小，因此这一结构可以取得良好表现。但在遥感图像中，由于遥感图像覆盖了广阔的地面区域，同时其中的待检测目标尺度差异极大，因此，在最终卷积块中，每一像素映射到原图像中都将包含一片较大的区域，显然地，这不利于网络对待检测目标进行精确的定位，更重要的是，在经过多层卷积后，原本尺度极小的部分目标在特征图块中难以保留足够的特征信息，这大大影响了网络对小目标的检测准确率。

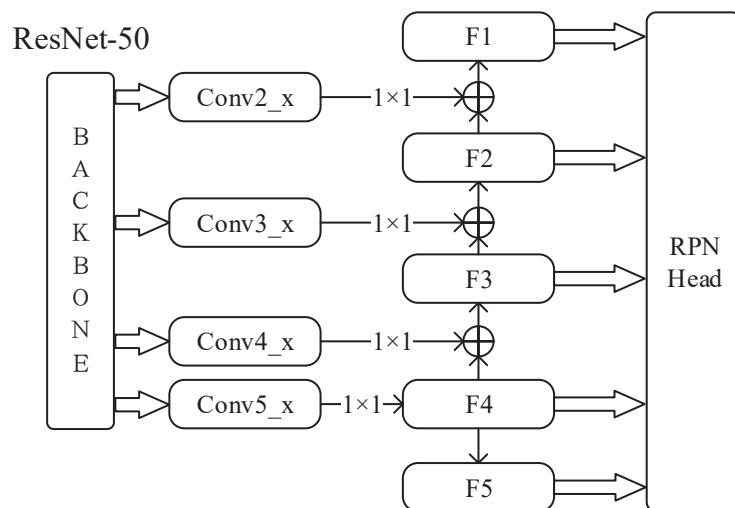


图 4-2 特征金字塔结构

因此，为了能够同时获取到图像中不同尺度目标的特征信息，需要对卷积的输出进行改进，其改进方法是基于不同卷积层次的输出构建特征金字塔 FPN 网络，其原理在于卷积网络在卷积过程中会随着网络的深入与像素中感受野的增大，逐渐学习到更为抽象的语义特征，对于不同尺度的目标来说，这一过程的最优深度是不一致的，一个较大目标可能需要经过多层的卷积与池化后，才能提炼出核心特征；但对于一个较小目标而言，仅需要较少的卷积便已经得到了高度抽象且准确的特征，如果再进行更深层的卷积池化操作，反而会导致最终特征被过度精炼失去了表征意义。而使用 FPN 网络可以将单一层级的输出改为多层输出的聚合体，并且同时利用网络不同层之间的信息，来增强检测网络对多种尺度各异的目标的检测能力。图??给出了 FPN 网络的结构设计，在经过 backbone 网络 ResNet-50 的特征提取过程后，取出 Conv2_x, Conv3_x, Conv4_x, Conv5_x 层（以下简称 C2、C3、C4、C5 层），从 C2 到 C5 特征图尺度逐渐减小，C2、C3、C4、C5 层输出通道数分别为 256、512、1024、2048；FPN 网络接收这四层输出作为输入，通过一个 1×1 卷积核建立连接，定义了 F1 至 F4 层，其中 F1 层尺度与 C2 层相同，以此类推，F4 层与 C5 层尺度一致，需要说明的是，由于 1×1 卷积核的加入，让 F1 至 F4 层通道数被限制为 256，F4 层由 C5 层连接得到，而其上层 F3 则由 C4 层经由 1×1 卷积与 F4 层经由上采样后融合得到，同理 F2、F1 层；而 F5 层则是由尺度最小的 F4 层经由下采样方法得到，一般地，可以使用最大池化或卷积的方式进行，在此网络中使用了最大池化方式得到 F5 层。就此得到了 FPN 的五层输出 F1 至 F5，其尺度逐层降低，这五层将分别地送入 RPN 阶段提取候选框。

FPN 的加入，有效保留了遥感图像中的多尺度特征，通过自顶向下的卷积过程提取出多尺度目标特征，随后经由自底向上的融合过程，将底层获取的语义特征与原有层特征融合，在保有高层次语义特征的同时增强了低层次位置信息，使网络获得了对遥感多尺度性的适应能力。

4.2.2 可形变卷积

在遥感目标检测任务中，由于小目标特性与目标密集性的存在，对特征的提取过程受到了极大关注，如何提取到更贴近目标、更鲁棒的特征，且减少其他无关目标与背景的噪声干扰成为切入点。而回到卷积过程中，通用的卷积核设计往往为 $n \times n$ 的正方形，这就意味着在层层深入的卷积中，最终特征图中的一点，其对应的感受野始终保有这一矩形特性，但在遥感图像中，目标往往并不具有类似的规则性，且目标还具有密集多方向的显著特性。因此，为了增强卷积特征提取过程，适应遥感目标的特殊性，将可形变

卷积结构加入了特征提取网络中，通过更改卷积核感受野，加入偏移量，增强了特征的鲁棒性与识别的准确率。

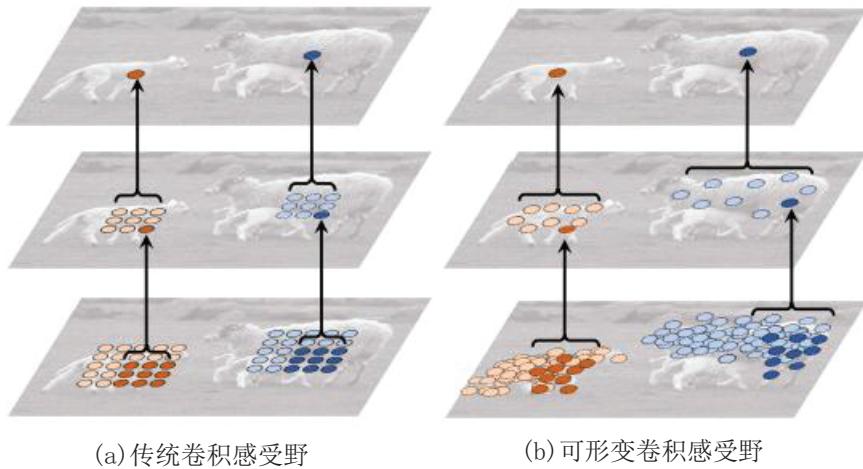


图 4-3 可形变卷积感受野

可形变，即卷积核感受野的可形变性，如图??所示，图 (a) 为传统卷积过程中感受野的逐层映射关系，而 (b) 为可形变卷积感受野的逐层映射关系。显然的，在可形变卷积过程中，卷积核所选取的卷积区域不再是一个规则的矩形区，而是由若干分散点像素组合而成，在经过训练学习后，可以使得这些点贴近到目标存在前景的不规则区域内。

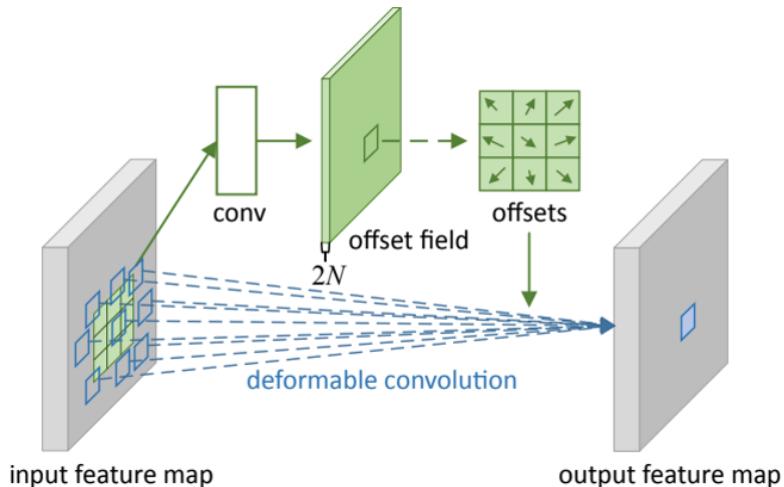


图 4-4 偏移学习网络

要达到可形变的卷积效果，需要在卷积时给被卷积图像中的每一像素点施加一个偏移量，而要使得偏移量与目标区域相匹配，则需要利用无监督的方法进行学习，图??给出了这个无监督学习的网络结构。网络的输入为上一层卷积结构的输出特征图块，在进行下一次卷积之前，将其送入一个卷积结构中，与特征提取中的卷积结构不同，这一卷

积结构对偏移量敏感，输出的是与输入同等尺度的偏移预测图，当输入的特征图维度为 $h \times w \times N$ 时，得到的偏移预测图则为 $h \times w \times 2N$ ，其中 $2N$ 代表了对原特征图 N 通道上同位置像素点在原始特征图像中 x 轴、 y 轴上的偏移量 (x_offset, y_offset)，在此之后，代入偏移量进行特征提取的卷积过程如式??所示：

$$y(P_0) = \sum_{P_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (4-1)$$

其中 Δp_n 由式??计算得到：

$$\Delta p_n = \sqrt{x_offset^2 + y_offset^2} \quad (4-2)$$

在计算过程中，有两点需要加以限制：

1. 由于偏移量是通过预测网络学习得到，因此其值不为整数，显然，代入偏移量后得到坐标 $(x+x_offset, y+y_offset)$ ，如图??所示，对该点坐标值上下分别取整，可以得到一个四点矩阵，而小数位置在图像的像素操作中没有实值，因此，需要使用插值法计算出实际得到像素值大小，在本网络中使用双线性插值实现，因此其计算方法可以由式??表示，展开为式??，通过周围四个整数像素点的值加权计算得到结果。其中 $f(Q)$ 表示该点的像素值大小；

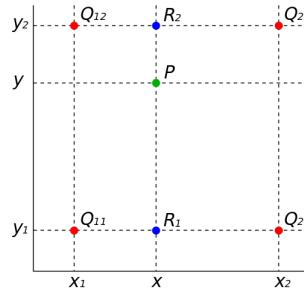


图 4-5 双线性插值

$$f(P) = \begin{bmatrix} (x_2 - x) & (x - x_1) \end{bmatrix} \begin{bmatrix} f(Q_{11}) & f(Q_{12}) \\ f(Q_{21}) & f(Q_{22}) \end{bmatrix} \begin{bmatrix} (y_2 - y) & (y - y_1) \end{bmatrix} \quad (4-3)$$

$$\begin{aligned} f(P) = & f(Q_{11})(x_2 - x)(y_2 - y) + f(Q_{21})(x - x_1)(y_2 - y) \\ & + f(Q_{12})(x_2 - x)(y - y_1) + f(Q_{22})(x - x_1)(y - y_1) \end{aligned} \quad (4-4)$$

2. 其次，当偏移量超出原特征图分辨率尺度范围或为负值时，对应原特征图像上得到的点也是无意义的，因此在预测网络中加入需要加入边界限制，确保其计算结果在特

征图尺度之内，是可以计算的且有意义的。

在本网络中，基于对网络中各卷积层尺度及其对应感受野的大小进行分析，结合需要检测的目标其尺度特征在卷积层中的分布确定修改方案。在 backbone 的 ResNet-50 的 Conv3_x, Conv4_x, Conv5_x 层中加入了可形变卷积结构，能够在不占用过多计算资源的同时增强对遥感图像中密集性、多方向性目标的特征提取能力。

4.2.3 级联网络

级联网络的思路最早在 Cascade RCNN^[?] 中被提出，其关注的是网络中设置的 IoU 门限值对检测结果的影响。在常见的两步法目标检测方法中，RPN 网络通过多比例与多尺度的锚点 anchor 来得到候选框，在遍历图像时，会在同一像素处置生成多个形状的检测框，为了尽量避免其中重复检测框问题的出现，在检测网络中往往设置 IoU 门限对候选框进行筛选，IoU 的计算方式由式??给出，其中 A、B 表示独立的两个检测框，显然当两个检测框越接近时，它们交集越大、并集越小，IoU 计算值也越大；反之当两个检测框非常疏远时，IoU 计算值会趋近于 0，IoU 值在某种程度上表征了两个检测框重合度的大小。

$$IOU = \frac{A \cap B}{A \cup B} \quad (4-5)$$

因此，在网络训练过程中，通过更改 IoU 门限大小，可以人为地影响网络学习效果。在 IoU 门限值较低时，大量与真实检测框相关性较低的检测框作为正样本被保留，导致检测结果中出现无关背景与其他噪声；当 IoU 门限升高时，只有和真实框非常相近的检测框得以保留，此时网络能学习到精确的目标信息，但由于检测产生的正样本数量急剧减少，网络存在过拟合的风险。对于一个目标检测任务而言，检测框预测与真实值越接近，在检测结果上体现为网络对目标的定位能力越强，可以更准确地预测出目标位置。因此级联的思想也就有了立足之地，其核心在于，当候选框经过检测器后，其输出预测框与目标真实标注 (Ground Truth, GT) 的 IoU 会增大，可以得到一个具有良好数据分布的结果，那么只要将输出再次作为输入，此时更为精确的检测框可以导致普遍的 IoU 提升，进而避免了在 IoU 过高时，检测正样本过少导致的过拟合问题。因此当使用多个检测网络级联时，通过使用前一阶段的检测结果去训练下一阶段的检测器，可以人为地提升 IoU 门限而不产生过拟合问题，同时也提高了网络的检测能力。在本章提出的网络中，便以此思路设计了三级级联的检测网络 S0、S1、S2，对其 IoU 门限值设置分别为

0.5、0.6、0.7 逐层递增, 将 S0 网络检测输出框的仍含有较多噪声的结果送往 S1 中继续训练, 并同样地取输出送往 S2 进行检测, 最终将 S2 输出作为网络最终输出, 这一结构提升了检测网络的学习能力, 给遥感图像中密集目标与小目标的检测带来助益。

第五章 基于 GSD 预测的超分辨遥感目标检测网络

5.1 引言

在第三章节中，改进的深度级联网络在遥感图像目标检测上获得了较大进步。但值得关注的是，虽然网络针对遥感图像的小目标性、密集性和多方向性进行了初步的优化与设计，却尚未利用到遥感图像中的特殊信息。与通用目标检测的图像不同，遥感图像往往会标注地面采样距离 GSD，也称地面采样间隔。它是存在于遥感与航拍影像中的一个参数，其值的大小说明了图像中的一个像素点所对应的实际地面距离的大小。在 DOTA 遥感数据集中，由于遥感图像取自于多个卫星图像源，且同一卫星在不同情况下得到的遥感图像也具有不同的拍摄参数与拍摄高度，因此具有不同的 GSD。显然的，当遥感图像分辨率相同时，GSD 的差异将导致图像所包含的实际区域的大小不同，如图??图??所示即为不同 GSD 下的两张图像表现，图??的 GSD 大小为 4.47，分辨率为 4526×2708 ，图??的 GSD 大小为 0.12，分辨率为 9089×6473 ，可以明显看出，虽然图??的分辨率尺寸远小于??，但由于具有较大的 GSD，图??中包含有更大、更密集的区域与目标，且这一差异源自于图像拍摄时的参数设置，难以直接通过常用网络的尺度后处理方式进行规避，这直接导致了在大 GSD 图像中，目标尺度过小较难以被检测的问题较为严重。因此，针对 GSD 这一特殊信息，本章设计了一个基于图像纹理复杂度的 GSD 预测网络，试图通过将 GSD 引入检测过程中，结合图像超分辨思路，对前述遥感目标检测网络进一步优化，从而获得对极小目标的检测能力及在高 GSD 图像中目标检测准确率的提升。

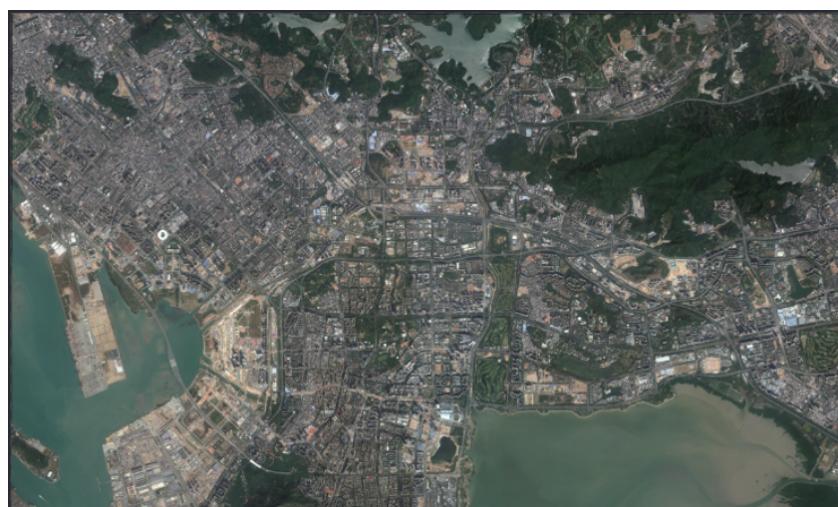


图 5-1 大 GSD 遥感图像



图 5-2 小 GSD 遥感图像

5.2 算法介绍

网络整体结构如图??所示，网络主要由三个子模块网络构成，分别是 GSD 预测网络、超分辨网络及目标检测网络。GSD 预测网络可以接收完整的待检测遥感图像，通过对图像纹理复杂度的分析，输出对图像 GSD 大小的判断结果；而超分辨网络则接收来自 GSD 预测网络的输出与输入的遥感图像，按照预测结果对图像进行分割和选择性的超分辨，得到若干图像序列；最后，目标检测网络接收前端网络标准化之后的输出作为输入，基于两步法思想执行目标的定位与检测任务，它的输出即为网络的最终输出检测结果。

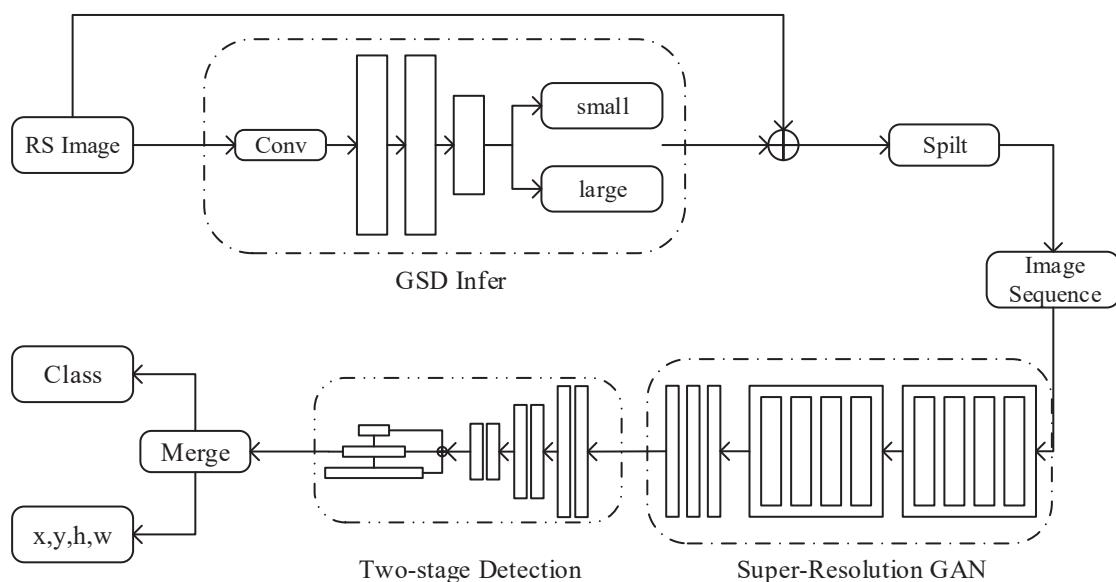


图 5-3 网络整体结构

5.2.1 GSD 预测网络

在计算机视觉领域, 图像复杂度 (Image Complexity) 可以从多个维度进行评估, 如色彩复杂度 (Color Complexity)、纹理复杂度 (Texture Complexity) 与形状复杂度 (Shape Complexity)^[?]。在遥感图像中, GSD 的大小决定了像素所代表实际地面距离, 可以推断出, GSD 间接地决定了在遥感图像中一个单位像素块所包含的实际区域大小, 显然地, GSD 越大, 对于同样的地理区域而言, 图像所涵盖的区域内建筑、目标、背景环境就越复杂, 而这一复杂性可以通过图像的纹理复杂度得以体现, 因此一种基于遥感图像纹理复杂度的 GSD 预测网络应运而生, 其整体算法流程如图??所示, 主要由预处理、纹理提取和距离估计三部分构成。

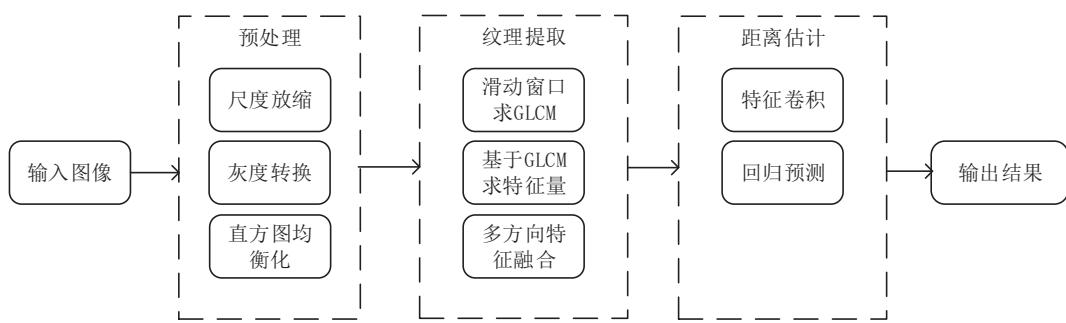


图 5-4 GSD 预测网络

5.2.1.1 预处理

预处理环节在提取遥感图像纹理特征之前先对输入图像进行一定的处理, 使其能够符合后续处理的尺度、色彩要求, 避免结构性误差的引入。一般地, 基于输入图像特性, 从以下几个方向进行预处理:

1. 尺度放缩。遥感图像往往具有常规意义上的过大分辨率, 而在进行图像的纹理特征提取与 GSD 预测时, 需要确保输入的完整性, 防止因局部平滑导致的误判和错漏。因此对图像的尺度进行放缩是很有必要的, 这样就能利用图像的尺度不变性特征, 在计算量较小的情况下进行纹理提取工作, 降低了不必要的资源消耗, 提升了网络整体运算效率;

2. 灰度转换。多通道图像会使得图像的纹理提取工作更为复杂, 但与此同时却不能显著增强提取出的纹理特性, 因此基于优化网络结构、提升网络计算效率的目标, 可以将输入的彩色图像以一定方式转化为灰度图像, 具体计算方式如式??所示, 其中 R、G、B 分别代表图像 RGB 通道对应像素值, 该转换方式是由计算机图像处理及人眼视觉感

知的通用准则规定而成，以认知心理学对于 RGB 图像色彩的研究为基础，可以最大程度的确保灰度图像与原图像在视觉感知、图像处理中的一致性。

$$Y = 0.299R + 0.587G + 0.114B \quad (5-1)$$

3. 灰度直方图均衡化。这是一种传统的灰度图像处理方法，其目标是使得图像灰度分布更为均衡地分布在整個灰度域之上，使图像的亮暗都更加明显，增强了对比度。在图??中，图 (a) 为原灰度图像，图 (b) 为经过直方图均衡化之后的灰度图像，相比于原图，均衡化使得图像中明度对比度得到增强，图像细节进一步凸显，纹理更加清晰，便于后续处理。

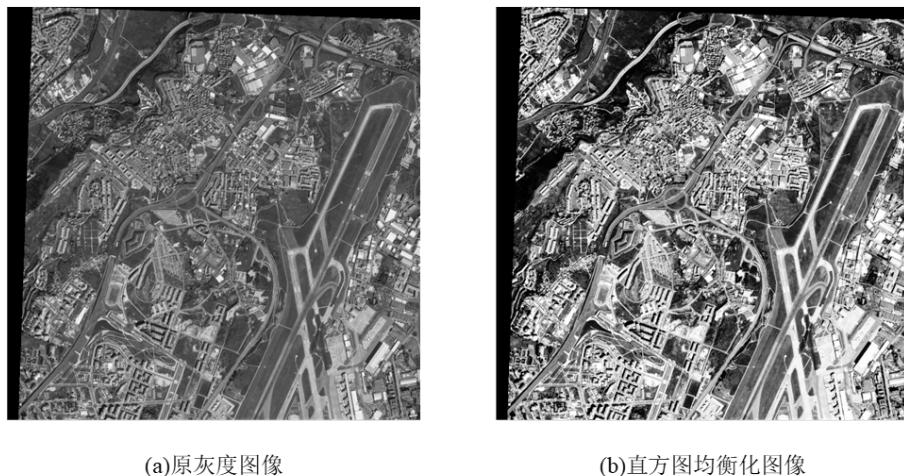


图 5-5 直方图均衡化

5.2.1.2 纹理提取

在图像分析领域，常常通过计算灰度共生矩阵及其特征值 (Gray-Level Co-occurrence Matrix, GLCM)^[?] 来对图像的纹理复杂度进行评估。在文章中，灰度共生矩阵被定义如下：设一灰度图像具有 N 个灰度值，那么计算得到的灰度共生矩阵的大小就为 $N \times N$ 阶，规定一个像素对偏移量为 δ ，它意味着计算时需要不断地在图像中按角度 θ 去遍历距离为 δ 的像素对 a 、 b ，可以得到其灰度值对 (a_k, b_k) ，随后将具有相同灰度值对的像素对数量累加，对应着灰度共生矩阵中 (a_k, b_k) 的元素 $G(a_k, b_k)$ 的值。当遍历完整个图像之后，就得到了该图像对应的灰度共生矩阵 G 。

如图??所示为一幅具有 4 阶灰度级的灰度图像及以水平方向进行遍历得到的灰度共生矩阵，显然地，灰度共生矩阵是基于像素对灰度值的统计量，它记录了在图像当中不同灰度匹配模式的出现次数，而 δ 则代表了该灰度共生矩阵对于灰度纹理精细级别的

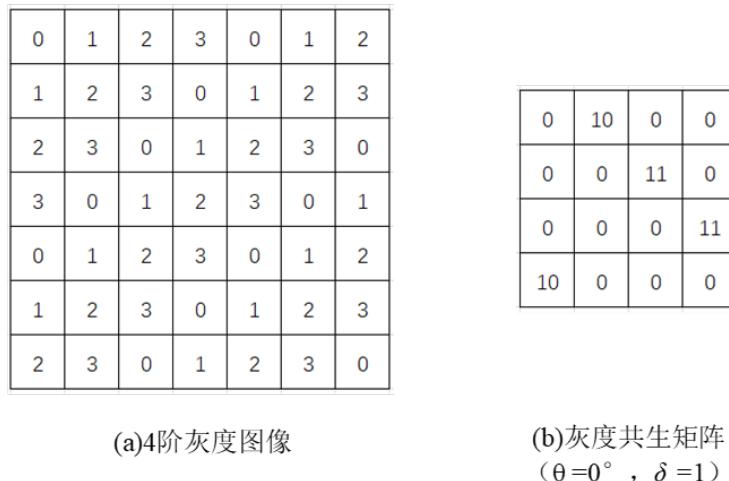


图 5-6 灰度图与灰度共生矩阵

敏感度，当 δ 越小时，像素对间距较小，则检测的纹理较细致，反之则纹理越粗糙。

进一步的，当使用不同的距离 δ 与遍历角度 θ 时，可以得到设置不同方向与不同距离时的灰度共生矩阵如图??所示。这体现了灰度共生矩阵对于纹理方向的敏感性。一般地，为使提取到的纹理特征具有一定的旋转不变性，会特别地使用 0° 、 45° 、 90° 、 135° 这四个角度进行计算，距离则由图像中主要纹理的粗糙程度来定义。

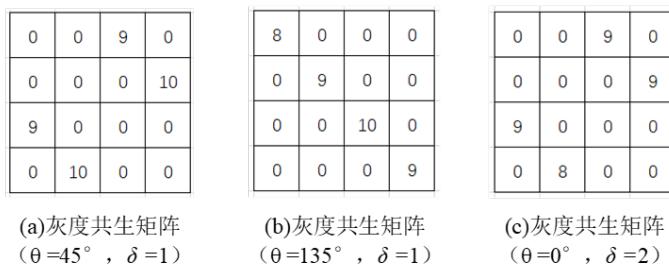


图 5-7 不同条件下的灰度共生矩阵

灰度共生矩阵具有着图像的纹理统计信息，但就作为图像纹理特征而言，其鲁棒性尚不能满足要求，不够直观也是一个巨大的不足，研究者们难以直接利用灰度共生矩阵进行比较和分析。因此，在提出灰度共生矩阵的同时，Haralick 团队也给定了 14 个用于纹理分析的计算统计量，也即纹理描述子，分别为：二阶矩、熵、对比度、均匀性、相关性、方差、和平均、和方差、和熵、差方差、差平均、差熵，同质性及最大相关系数，但彼时他并未给出这些统计量的有效性与相关性结论。随后，Baraldo^[2] 等人经由研究计算论证，说明在 Haralick 给出的 14 个统计量中，其中 4 个统计量之间不相关性最强，对于遥感图像具有较好的特征表现能力。分别是：

1. 二阶矩 (Angular Second Moment, ASM)，也被称作能量，其计算方法由式??给

出，其中 $P(i, j)$ 为灰度共生矩阵中位于 (i, j) 处的元素归一化概率。ASM 对灰度共生矩阵内值的和大小敏感，当图像纹理分布均匀、纹理粗细程度接近时，ASM 可以获得较大值；

$$Asm = \sum_i \sum_j P(i, j)^2 \quad (5-2)$$

2. 熵 (Entropy, ENT)，在物理学领域中，熵被用来描述一个系统的混乱程度。在此处，其计算方法由式??给出，它具有类似的描述功能，可以用于表现图像中纹理的复杂程度或分布的非均匀程度，当熵值越大时，意味着图像中纹理随机性较强、分布较为复杂，显然地，对于一张纯色图像，其熵值为 0；

$$Ent = \sum_i \sum_j P(i, j) \log P(i, j) \quad (5-3)$$

3. 同质性 (Homogeneity, HOM)，也被称为逆方差。其计算方法由式??给出，正如其名，同质性关注的是图像中局部区域的纹理变化情况，当不同区域间的图像纹理变化较为缓慢平滑时，其值较大，反之较小的同质性意味着不同区域间纹理变化剧烈；

$$Hom = \sum_i \sum_j P(i, j)^2 / [1 + (i - j)^2] \quad (5-4)$$

4. 非相似性 (Dissimilarity, DIS)，这是一类用于度量图像局部或整体区域内灰度差异区分度的统计量，其计算方法由式??给出，通过公式可以发现，DIS 与灰度间的度量具有线性关系。

$$Dis = \sum_i \sum_j P(i, j) |i - j| \quad (5-5)$$

除此之外，基于对图像信息的提取需求，还有另外几类较为直观的统计量也常常被用于计算之中，如均值、对比度及相关性等，其中均值表征了图像中灰度的整体或局部明暗度，当图像中出现规则、易于描述的纹理时，均值将具有较大值；对比度是对图像中局部像素与其周围像素的灰度差距的度量，灰度共生矩阵对比度与图像对比度表征类似，越大对比度的图像说明在图像中纹理越清晰，边界较明显^[?]；相关性则具有方向敏感性，其与计算时遍历图像的方向有关，它揭示了在图像中沿着某一特定方向上灰度纹理的延伸长度，与非相似性 DIS 类似，它也是一个灰度关系的线性度量统计量。

需要更进一步指出的是，上述的所有统计量都是标量形式，在一些简易灰度图像处理方法中，往往将这类统计量进行联合，构建一个多维矢量对图像进行纹理的表征，称之为特征矢量，然后利用特征矢量代入计算之中执行相似性对比、分类等任务。但在遥感图像中，这一方法存在若干不足之处，其一是遥感影像更为复杂，由于其覆盖广阔，

影像中会存在多种地块与丰富的目标，单一统计量在复杂图像中显得较为乏力，区分维度不够，无法表征出图像局部纹理特征，如水域与城市共存时，即无法体现出水域的统计特性，也难以表征城市建筑的纹理特征，使得不同地域、不同纹理下的遥感图像在特征矢量中差异性较小、边界不易区分，严重降低了遥感图像纹理特征的表征能力与可识别性。因此，借由卷积神经网络的启发，在本设计中，将滑动窗口思想应用于灰度共生矩阵的计算当中，使用了基于局部灰度共生矩阵的特征图表征方法，该算法主要有以下步骤构成：

1. 设置一个 7×7 的滑动窗口，padding 为 3，步长为 1，窗口起始位于图像左上角；
2. 针对滑动窗口中的 7×7 图像区域以前述定义的计算方法进行统计，得到对应的局部图像灰度共生矩阵；
3. 计算得到的局部灰度共生矩阵中均值、方差、同质性、对比度、非相似性、熵、能量、相关性和自相关等一系列统计量，分别输出作为该滑动窗口的输出，其输出形式为多维度矢量；
4. 将滑动窗口按照既定的步长，以从左至右，自上而下的轨迹对遥感图像进行遍历，每次移动后重复第 2、3 步骤得到输出；
5. 待滑动窗口遍历结束，将期间的所有输出按统计量类别组合，可以得到多幅特征图像，图像与原输入图像具有相同的分辨率，但仅仅表征了图像的灰度共生矩阵在某一特定统计量下的表现；
6. 重复 1 至 5 的步骤，但在计算灰度共生矩阵时，分别使用 0° 、 45° 、 90° 、 135° 对纹理进行统计；
7. 将同一图像中同一纹理统计量在不同计算角度下的特征图以平均加权的方式进行加和，得到最后的特征图像，由于本网络中设计了 9 类统计量表征纹理特性，因此得到 $H \times W \times 9$ 维度的纹理特征图像，其中 H 、 W 为输入图像的高与宽。

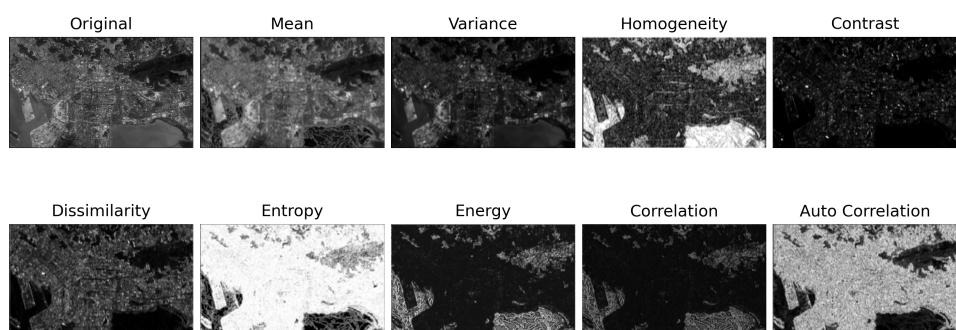


图 5-8 大 GSD 图像 GLCM

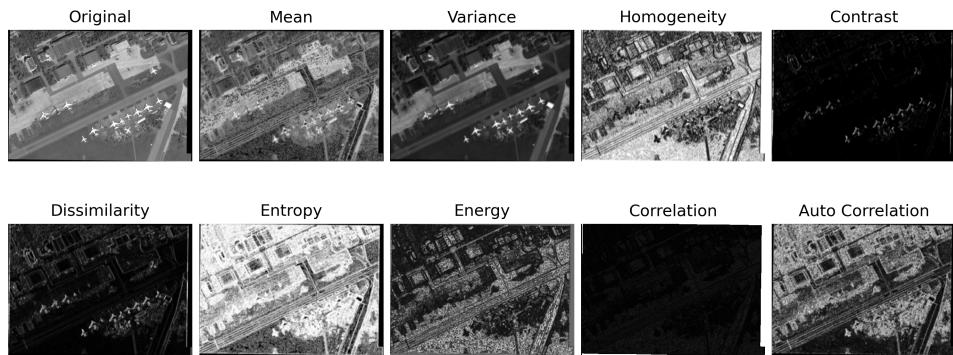


图 5-9 小 GSD 图像 GLCM

至此，获得了一幅输入遥感图像的纹理特征图，它同时具有着局部特征与整体特征的信息，且在多个方向上都具有一定的纹理敏感性，最大程度地提取了图像中由不同地块、不同目标和不同区域所产生的纹理特征，为评估图像复杂度提供了坚实依据，为后续 GSD 预测提供了可靠鲁棒的输入。图像??和图??给出了对图像??和图??的纹理特征提取图结果。

5.2.1.3 距离估计

距离估计网络的设计目标是通过学习图像的复杂度，尤其是纹理复杂度的信息，合理推测出图像拍摄时的 GSD 参数状态。各领域研究者也曾提出过相应的方法来利用灰度共生矩阵信息，如 Wed 等人便利用 GLCM 与 K-means 聚类算法相结合，试图以无监督的方式对医学影像中的病症点进行分析^[?]；还有 Naveena 等人则将 GLCM 与 SVM 分类器相结合，在树叶分类问题中取得了显著表现^[?]。然而，在遥感图像领域，这类方法效果却并不尽如人意，这是因为遥感图像具有较大分辨率与较为丰富的图像内容，在简单线性分类方法中会带来过高维的复杂度，占用过多计算资源。

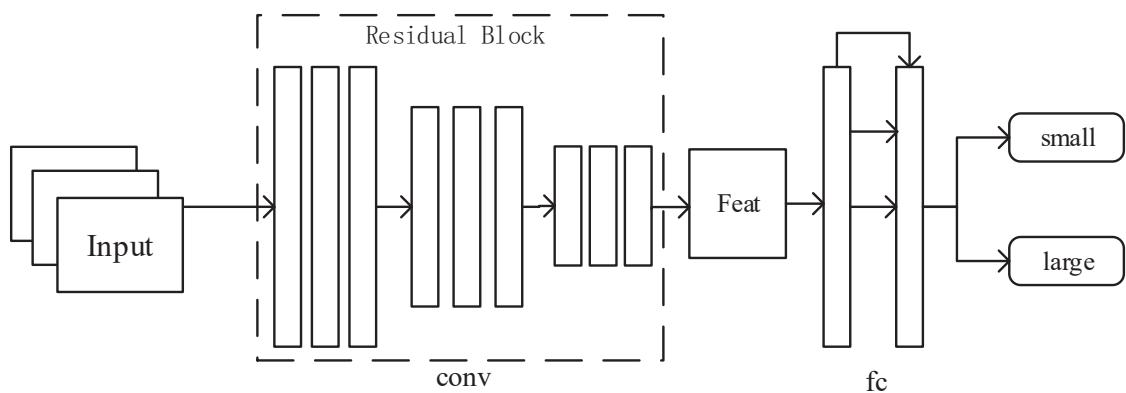


图 5-10 预测网络结构

因此，基于对输入图像内容的复杂性进行估计的需求出发，本章节中提出了一种使

用深度卷积网络提取特征并回归分类预测遥感图像 GSD 的方法。网络结构以 ResNet-50 为基础进行修改，因为其深度能够满足特征提取与学习的需求。具体结构如图??所示，输入图像为前述计算所得到的纹理特征图像块，在输入网络前，为满足网络卷积输入条件，需要先将其尺度调整为 224×224 ，随后设计了多个残差卷积块，每个残差块由 1×1 和 3×3 的多个卷积核叠加而成，最终展平后得到 $1 \times 1 \times 4096$ 维的特征图像，在池化层利用平均池化方式进行降维后，直接与全连接层 FC1 相连，FC1 具有 2048 个神经元，它的输出经由密集全连接输入全连接层 FC2 的 512 个神经元，最终汇集为两类输出，完成回归分类并输出标签。基于本次使用数据集 DOTA 中的图像 GSD 实验特性，当 GSD 较小时，检测网络对目标仍能具有较好的适应性，而当 GSD 过大时，网络无法识别出其中的极小目标，因此在 GSD 预测输出中将图像分类为正常 GSD 尺度与过大 GSD 尺度两类。网络使用 Adam 优化器进行反向梯度传播，损失函数为交叉熵损失，如式??所示，在本网络设计中 N 为 2， L_i 为单一类别的交叉熵损失。

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (5-6)$$

在进行网络设计时，曾经存在着另一方案的设计，即将 GSD 准确值或近似值作为回归的预测目标。但经过实验证明，这一方案存在两个不足之处：其一是 GSD 与图像复杂度特征间为非线性关系且不具有强烈相关性，对于遥感图像这类具有较高复杂度的图像而言，要将其与单一精确数值进行关联是非常困难的；其二是由于后续需要使用超分辨等图像尺度变换操作，为确保其输出效果准确，对于图像输入尺度与放缩比例具有严格要求，无法直接利用预测出的精确值进行操作，防止在超分辨过程中引入不必要的噪声干扰检测结果。因此，综上两点所述，使用分类预测方案，对遥感图像的 GSD 值进行预测是一个能在检测效率、识别准确率及资源利用率上达到平衡的方法，具备一定的优势。

5.2.2 图像超分辨

前述对图像的分析与预测工作，其最终目的都在于提取遥感图像的 GSD 信息，并将其引入网络之中。在本节中，正是出于对 GSD 信息的利用方法进行思考，设计了一个遥感图像超分辨网络，其逻辑意义在于：基于图像??和??的分析可以发现，当在不同 GSD 图像中取同样大小像素块时，对于图像中的同类目标，其体现在像素尺度之上的差距正是两图像 GSD 的比值，此外，研究者们的一致认知是，对于现有的遥感目标检

测网络而言，由于其训练时大部分目标都分布于正常 GSD 尺度之下，网络学习到了这一尺度敏感性，所以难点之一便是在大 GSD 图像中出现的极小目标，其尺度比正常尺度下的同类目标缩小了数倍，故难以被网络检出。因此，一个利用 GSD 信息的方法便自然地产生，即将大 GSD 图像通过某种方式进行放大，人为地调整图像 GSD 的大小，使之接近正常 GSD 图像，从而增大对大 GSD 图像中极小目标的检测能力。

传统的图像放大方法主要有插值类方法，其主要原理是参考原低分辨率图像中的相邻像素值，对得到的高分辨率图像中空白像素进行填充，根据插值内核处理方法不同可以分为线性插值方法如最邻近插值、双线性插值和三线性插值等方法，和非线性插值方法如基于边缘信息的插值方法、基于小波系数的插值方法等^[?]；而超分辨网络方法则是基于神经网络学习的方法，其所利用来超分辨的信息并不局限于单一输入图像，而是基于多幅图像的学习结果来估计目标高分辨图像的像素值，超分辨网络得到的图像像素与原低分辨图像间并不存在严格的位置映射关系，严格来说，新图像中每一个像素都是重新计算得到，不会保留原有图像中像素的所有信息。传统方法较为简单，运算速度快，且能适应更为灵活的图像放大倍率，但它只能基于输入图像中所具有的像素信息进行放大，对于大型图像或较为复杂的图像，其输出图像在视觉效果及噪声检测指标上都存在不足；而超分辨类方法则能适应复杂图像的放大任务，但需要使用大量已知标定数据进行预先训练，且计算时间相对较长，由于网络结构的限制，对放大倍数和输入尺度存在限制，只有在预设倍数和尺度下工作时才能发挥网络的最佳性能。综合这两类方法的优劣，在本网络中，针对遥感图像显然使用超分辨网络更为合适，而在之后的实验中也证明了这一点。

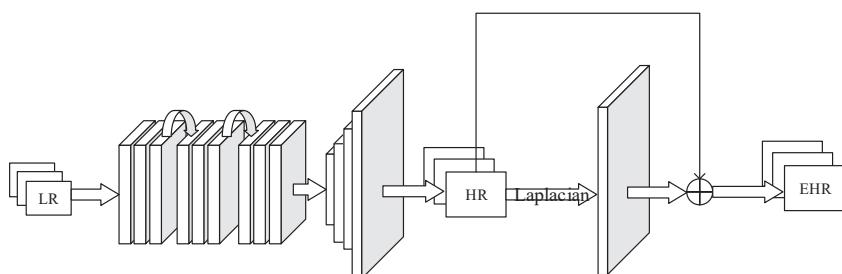


图 5-11 超分辨网络结构

本章所使用网络结构为对抗生成网络 (Generative Adversarial Network, GAN)，最初的 GAN 结构是由 Goodfellow^[?] 提出，GAN 网络正如其名，具有“生成器”与“判别器”之间的对抗过程，生成器以某一规则不断生成与真实样本相似的样本或分布，如音频样本、图像样本和文本样本等，而判别器则被设计成为具有类似特征提取与识别功能

的网络，它需要将生成样本与真实样本进行区分，这两部分互相拮抗，通过构建相关联的损失函数进行相互优化。而在超分辨任务中，生成器的任务便是基于原始低分辨率图像 (Low Resolution, LR) 生成高分辨率图像 (High Resolution, HR)，经由判别器与原高分辨率图像对比得到误差，不断优化生成图像质量，从而实现图像超分辨。如图??所示即为本网络中使用的超分辨网络的示意结构，首先以超分辨网络 SRGAN^[?] 的生成器结构为基础，考虑到遥感图像的特殊性，泛用的超分辨方法很容易导致输出的图像中出现地面细节部分的模糊，严重影响到遥感目标的检测与定位，因此在后端加入一个边缘提取网络对得到的高分辨率图像进一步增强，并把增强后具有边缘信息的边缘增强高分辨率图像作为网络的最终输出，其具体的操作步骤如下所述：

1. 首先将需要超分辨遥感图像进行尺度裁剪至预定大小的多张图像，获得尺度一致的输入低分辨率图像 LR；
2. 利用多个残差块构建生成器网络，每个残差块结构如图??所示，包含有两个 $3 \times 3 \times 64$ 的卷积层，其步长为 1，在卷积层之后加入批归一化处理 (Batch Normalization, BN) 与 PReLU(Parametric Rectified Linear Unit) 作为激活函数；由多个残差块进行连接，在此过程中特征特图块的通道数逐渐增长，从广义上看便是逐渐学习到了高分辨图像所缺失的内容，最后连接 $3 \times 3 \times 256$ 卷积层并进行维度变换，得到 3 通道放大尺度图像，达到了对图像进行上采样的目的，并输出高分辨率图像 HR；

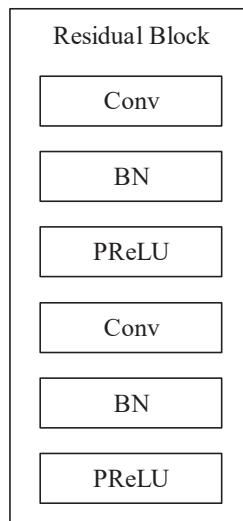


图 5-12 残差块

3. 将高分辨率图像 HR 再次作为输入，首先进行灰度转换得到灰度图像，随后使用一个低通滤波器对图像进行滤波，消除原本存在的部分高频噪声；最后再使用拉普拉斯算子提取高频边缘信息，得到边缘图像；

4. 最后将边缘图像与生成的高分辨率图像 HR 融合，融合方法为将 HR 图像的三通道图像分别与得到的边缘图像进行加权求和后归一化，获得具有强边缘效果的高分辨率图像 EHR，其尺度与最初生成的高分辨率图像一致，并具有更清晰的边缘信息，虽然无法完全恢复原图像中的边缘信息，但后续实验证明，其输出 EHR 在检测效果上已优于 HR。

5.2.3 目标检测网络

目标检测网络采用了第三章节中所构建的基于 Faster-RCNN 加入可形变卷积与级联结构的两步式遥感目标检测网络，其在实验中取得了最佳表现，非常适合作为后端的目标检测网络，其具体结构原理与检测能力已经在第三章中进行了详细实验说明，此处不再赘述。

第六章 总结与展望

遥感技术的飞速发展带来了内容越来越丰富和分辨率越来越高的遥感图像，图像中包含了大规模信息，有效地处理大规模信息能够给社会发展带来极大的便利。本文以高分辨率的遥感图像为研究对象，针对现有一些方法在目标检测过程中遇到的遥感图像小目标密集分布、方向任意、背景复杂和尺度差异大的问题，提出了相应的改进方案，有效提升了算法整体检测精度。本文的主要研究工作由以下三点概括：

(1) 考虑到自然图像中的水平框目标检测方法在对密集分布且方向任意的目标进行检测时容易出现的漏检和虚警问题，在研究过程中采用了有方向框来定位遥感图像中的目标，有方向框可以更好的说明相邻目标检测框之间的相似程度。

(2) 提出了一种改进的多方向两级级联 R-CNN 方法来更好的检测遥感图像中的有方向小目标。在第一级检测网络设计了多方向 RoI 对齐模块从多个不同旋转方向的兴趣区域获取方向敏感特征，同时在回归分支添加方向注意力模块自适应地给每个方向通道的特征赋予权重，增强了特征的方向敏感性，然后第一次回归得到初始的可能包含目标的有方向框，这样的有方向框可以更好的表示密集分布的目标的特征。最后在第二级检测网络对该有方向框做进一步回归得到目标的精确位置。

(3) 基于区域提取的目标检测方法在检测过程中由于候选区域位置的变化造成了特征不对齐问题，针对这一问题，设计了基于可形变卷积的多分支特征对齐模块来重新采样特征，同时采用了不同扩张率的空洞卷积来获取不同尺度的感受野。另外针对在研究过程中发现的网络模型对于不同长宽比目标的检测倾向不一致问题，提出了基于目标长宽比的角度偏移惩罚损失函数，在训练过程中更加关注大长宽比目标角度偏移量的学习。

在两个公开数据集 DOTA 和 HRSC2016 上进行的消融实验以及与其它方法的对比实验都验证了本文提出算法的有效性。然而本文算法在一些地方仍然存在着局限性，下一步的工作重点可以集中在以下方面：

(1) 采用更简单的模型进行研究。本文的研究基础是基于区域提取的目标检测方法 Faster R-CNN，同时又采用了两级检测结构来完成，导致整体的模型结构比较复杂，降低了检测效率，可以考虑替换为基于回归的目标检测方法进行研究。

(2) 遥感图像中目标任意方向分布的特性使得检测器需要更多的网络参数来编码方向信息，效率降低，可以考虑在主干网络提取特征时将方向信息编码，在检测时根据编

码到的方向信息分别进行处理，能够很大程度提高效率。

(3) 利用监督信息来对特征不对齐问题进行分析。本文在特征不对齐问题的研究上是基于可形变卷积的，可形变卷积在卷积时的采样特征点偏移是从一个单独的卷积分支得到，带有一定的随机性。而两级检测结构可以在第一级得到每个水平感兴趣区域的偏移量，可以在后续研究中将其作为监督信息引导可形变卷积特征点偏移的方向。

参考文献

- [1] 翁祖平, 刘桂云, 马岚华. 小流域系留气球低空遥感初步试验研究 [J]. 中国水土保持, 1993, 11.
- [2] 孟执中, 何正华. “风云一号”气象卫星 [J]. 宇航学报, 1989, 3.
- [3] 刘斐, 吕大旻. 我国成功发射高分一号卫星 [J]. 中国航天, 2013, 5:10–13.
- [4] 潘腾, 关晖, 贺玮. “高分二号”卫星遥感技术 [J]. 航天返回与遥感, 2015, 36(4):16–24.
- [5] Padwick C, Deskevich M, Pacifici F, et al. WorldView-2 pan-sharpening[A]. In: Proceedings of the ASPRS 2010 Annual Conference, San Diego, CA, USA[C], 2010. 2630:1–14.
- [6] Kruse F A, Perry S L. Mineral mapping using simulated Worldview-3 short-wave-infrared imagery[J]. Remote Sensing, 2013, 5(6):2688–2703.
- [7] Harris C G, Stephens M, et al. A combined corner and edge detector.[A]. In: Alvey vision conference[C], 1988. 15:10–5244.
- [8] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2):91–110.
- [9] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. IEEE transactions on pattern analysis and machine intelligence, 2005, 27(10):1615–1630.
- [10] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer vision and image understanding, 2008, 110(3):346–359.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[A]. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)[C], 2005. 1:886–893.
- [12] Viola P, Jones M, et al. Robust real-time object detection[J]. International journal of computer vision, 2001, 4(34-47):4.
- [13] Noble W S. What is a support vector machine?[J]. Nature biotechnology, 2006, 24(12):1565–1567.
- [14] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2014. 580–587.
- [15] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278–2324.
- [16] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recog-

- nition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9):1904–1916.
- [17] Girshick R. Fast r-cnn[A]. In: Proceedings of the IEEE international conference on computer vision[C], 2015. 1440–1448.
- [18] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. arXiv preprint arXiv:1506.01497, 2015.
- [19] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. arXiv preprint arXiv:1605.06409, 2016.
- [20] Li Z, Peng C, Yu G, et al. Light-head r-cnn: In defense of two-stage object detector[J]. arXiv preprint arXiv:1711.07264, 2017.
- [21] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2017. 2117–2125.
- [22] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[A]. In: Proceedings of the IEEE international conference on computer vision[C], 2017. 2961–2969.
- [23] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2018. 6154–6162.
- [24] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2018. 8759–8768.
- [25] Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv preprint arXiv:1312.6229, 2013.
- [26] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2016. 779–788.
- [27] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[A]. In: European conference on computer vision[C], 2016. 21–37.
- [28] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[A]. In: Proceedings of the IEEE international conference on computer vision[C], 2017. 2980–2988.
- [29] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[A]. In: Proceedings of the European conference on computer vision (ECCV)[C], 2018. 734–750.
- [30] Cheng G, Zhou P, Han J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016,

- 54(12):7405–7415.
- [31] Cheng G, Han J, Zhou P, et al. Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection[J]. IEEE Transactions on Image Processing, 2018, 28(1):265–278.
- [32] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[A]. In: Proceedings of the IEEE international conference on computer vision[C], 2017. 764–773.
- [33] Xu Z, Xu X, Wang L, et al. Deformable convnet with aspect ratio constrained nms for object detection in remote sensing imagery[J]. Remote Sensing, 2017, 9(12):1312.
- [34] Zhong Y, Han X, Zhang L. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery[J]. ISPRS journal of photogrammetry and remote sensing, 2018, 138:281–294.
- [35] Yang X, Sun H, Fu K, et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. Remote Sensing, 2018, 10(1):132.
- [36] Ding J, Xue N, Long Y, et al. Learning roi transformer for oriented object detection in aerial images[A]. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition[C], 2019. 2849–2858.
- [37] Yang X, Fu K, Sun H, et al. R2CNN++: Multi-dimensional attention based rotation invariant detector with robust anchor strategy[J]. arXiv preprint arXiv:1811.07126, 2018, 2:7.
- [38] Yang X, Yang J, Yan J, et al. Scrdet: Towards more robust detection for small, cluttered and rotated objects[A]. In: Proceedings of the IEEE/CVF International Conference on Computer Vision[C], 2019. 8232–8241.
- [39] Xu Y, Fu M, Wang Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE transactions on pattern analysis and machine intelligence, 2020.
- [40] Yang X, Liu Q, Yan J, et al. R3det: Refined single-stage detector with feature refinement for rotating object[J]. arXiv preprint arXiv:1908.05612, 2019.
- [41] Han J, Ding J, Li J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021.
- [42] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain.[J]. Psychological review, 1958, 65(6):386.

- [43] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088):533–536.
- [44] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7):1527–1554.
- [45] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25:1097–1105.
- [46] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[A]. In: 2009 IEEE conference on computer vision and pattern recognition[C], 2009. 248–255.
- [47] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1):1929–1958.
- [48] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[J]. 2015.
- [49] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. nature, 2016, 529(7587):484–489.
- [50] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[A]. In: European conference on computer vision[C], 2014. 740–755.
- [51] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8):1735–1780.
- [52] Wiering M, Van Otterlo M. Reinforcement learning[J]. Adaptation, learning, and optimization, 2012, 12(3).
- [53] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[A]. In: European conference on computer vision[C], 2016. 630–645.
- [54] Bodla N, Singh B, Chellappa R, et al. Soft-NMS—improving object detection with one line of code[A]. In: Proceedings of the IEEE international conference on computer vision[C], 2017. 5561–5569.
- [55] Li K, Cheng G, Bu S, et al. Rotation-insensitive and context-augmented object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(4):2337–2348.
- [56] Xia G S, Bai X, Ding J, et al. DOTA: A large-scale dataset for object detection in aerial images[A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C], 2018. 3974–3983.
- [57] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C], 2018. 7132–7141.

- [58] Liu Z, Wang H, Weng L, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(8):1074–1078.
- [59] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.

谢裕麟

2021 年 10 月 6 日