中图分类号：TN911.73

盲 审 编 号 ：

南昌航空大学

硕 士 学 位 论 文

题 目

基于改进特征融合机制的遥感图像目标检测方法的研究

学 科、专 业 计 算 机 技 术

专 业 代 码 085400

学 位 类 别 **专 业 学 位 硕 士**

2024 年 4 月

学校代码：10406

分类号：TN911.73
学号：2104085400012

南昌航空大学

硕 士 学 位 论 文

（专业学位研究生）

基于改进特征融合机制的遥感图像目标检测方法的研究

硕士研究生：陈鹏辉

导　　师：李其申

申请学位级别：硕　士

学 科、专 业：计算机技术

所 在 单 位：信息工程学院

答 辩 日 期：2024 年 6 月

授予学位单位：南昌航空大学

Research on Remote Sensing Image Object Detection Method Based on Improved Feature Fusion Mechanism

A Dissertation
Submitted for the Degree of Master
On Computer Technology

By xxx

Under the Supervision of

xxx

School of Information Engineering

Nanchang Hangkong University, Nanchang, China

June, 2024

摘 要

遥感图像目标检测技术在军事侦察、环境监测以及城市规划等诸多领域扮演了不可或缺的角色。该技术的关键之一在于多尺度特征融合，能够提升模型对遥感图像中目标的识别精度。考虑到遥感图像的特点，如较大的图像尺寸、目标的广泛分布及其尺寸多样性，以及经典的多尺度特征融合方法路径聚合网络(Path Aggregation Network, PANet)对检测模型的特征提取网络有着隐性依赖，这些因素可能会限制 PANet 的特征融合效果，甚至在特征融合阶段促进底层特征中噪声的传播，对遥感图像目标检测的性能产生不利影响。针对这一问题，本文对 PANet 进行了改进，旨在进一步提高遥感图像目标检测的准确率，主要工作内容如下：

(1) 针对 PANet 中的噪声传播问题，提出一种多尺度特征降噪与融合网络(Multiscale Feature Denoising and Fusion Network,MFDFN)。该网络由多尺度降噪模块(Multiscale Denoising Module,MDM)以及计算优化模块(Computational Optimiza-tion Module,COM)组成。MDM 旨在特征融合前使用高层的语义特征对底层的纹理特征中的噪声进行抑制，尽可能减少底层特征中的噪声对后续多尺度特征融合的影响。COM 旨在 MDM 降噪前对特征进行分组的权重计算及应用，使得在几乎不增加参数量的情况下提高 MDM 的计算精度。

(2) 提出一种基于 MFDFN 改进的增强多尺度特征降噪与自适应融合网络(Enhanced Multiscale Feature Denoising and Adaptive Fusion Network, EMFDAN)，该网络由条件变量加强的多尺度降噪模块(Conditionally Enhanced Multiscale Denoising Module, CEMDM)以及权重自适应调整模块(Weight Adaptive Adjust-ment Module, WAAM)组成。CEMDM 在 MDM 的基础上加入了条件变量，使得该模块对图像中特殊以及复杂信息得到更加有效的识别。而 WAAM 则对需要相加的特征进行局部以及整体的权重自适应调整，使得在相加的过程中重要的特征占据较大权重，减少不必要特征对后续检测流程的影响。

为了验证本文所提方法的有效性，在遥感图像目标检测数据集以及通用目标检测数据集上进行了一系列实验。结果表明，本文所提方法能够有效提升模型在遥感图像目标检测领域的检测准确度。

关键词：遥感图像，目标检测，特征融合，特征降噪，特征权重调整

Abstract

With the rapid progress of satellite and aerospace remote sensing technology, the quality and resolution of remote sensing images that can be obtained in academic research have significantly improved. This not only increases the richness of data, but also increases the technical difficulty of object detection. Traditional object detection techniques rely on manually designed features for object detection, which not only have low detection accuracy but also slow detection speed, making it difficult to meet the high-precision requirements of remote sensing image object detection. In recent years, deep learning technology has made rapid development in the field of object detection and has achieved significant results in remote sensing image object detection. Although the current mainstream remote sensing image detection methods all use pyramid feature fusion technology, considering the characteristics of large size, scattered targets, and varying sizes of remote sensing images, when the feature extraction ability of the detection model is insufficient, pyramid feature fusion technology may not be able to fully play its role, and even add noise during the feature fusion process, which can have adverse effects on detection performance. Based on this, this study focuses on improving feature fusion technology, innovatively improving the PANet structure, and conducting generalization tests on other models. The main innovative results are as follows:

(1) Propose a multi-scale feature denoising and fusion network (MFDFN) based on PANet improvement. This network consists of a Multiscale Denoising Module (MDM) and a Computational Optimization Module (COM). MDM aims to use high-level semantic features to suppress noise in low-level texture features before feature fusion, minimizing the impact of noise in low-level features on subsequent multi-scale feature fusion. COM aims to calculate and apply the weight of grouping features before calculating the multi-scale noise reduction module, so as to improve the calculation accuracy of the multi-scale noise reduction module with almost no increase in parameter quantity.

(2) Propose an improved Enhanced Multiscale Feature Denoising and Adaptive Fusion Network (EMFDAN) based on MFDFN, which consists of a Conditionally Enhanced Multiscale Denoising Module (CEMDM) and a Weight Adaptive Adjustment

Module (WAAM). CEMDM adds conditional variables on the basis of MDM, making the module more effective in recognizing special and complex information in images. WAAM, on the other hand, adaptively adjusts the weights of the features that need to be added locally and globally, so that important features occupy a larger weight during the addition process, reducing the impact of unnecessary features on the subsequent detection process.

In order to verify the effectiveness of the algorithm proposed in this paper, a series of experiments were conducted on a series of remote sensing image object detection datasets and a universal object detection dataset. The results show that the method proposed in this article can effectively improve the accuracy of target detection in horizontal and rotating boxes, and effectively enhance the detection ability of the model.
Keywords: remote sensing images, object detection, feature fusion, feature denoising, feature weight adjustment

目录

第 1 章 绪论

1.1 研究背景与意义

遥感技术的发展对于现代社会的影响深远，特别是在资源勘探[1]、环境监控[6]、农业发展[11]以及国家安全等方面提供了宝贵的信息支持。随着遥感技术的逐步成熟与卫星遥感数据的日益增加，如何有效地从大量的数据中提取有用信息，对于提升数据的应用价值具有重要意义。目前，遥感图像的自动处理和分析已经成为地理信息科学中的一个研究热点。然而，由于遥感图像自身的特点，如大尺度、高维度和复杂背景等，使得传统算法在处理这些图像时面临巨大挑战。

传统的遥感目标检测方法主要基于图像处理和特征工程，包括形态学操作、边缘检测和纹理特征提取等。然而，这传统方法的共同缺点在于它们依赖于手工设计的特征和规则，这导致它们通常缺乏对复杂环境的适应性和鲁棒性。而且，这些方法在参数调整上需要大量的人工干预，这在处理大量数据时效率低下。它们还往往难以处理图像的多尺度性和高维特征，且在实际应用中很难达到实时处理的要求。随着深度学习技术的发展，这些传统方法逐渐被具有自我学习和泛化能力的深度神经网络所取代。

近年来，深度学习的兴起，特别是卷积神经网络(Convolutional Neural Networks,CNN)的成功应用，为遥感目标检测带来了显著的突破。R-CNN[16]的引入标志着深度学习技术在目标检测领域的首次应用。然而，由于速度较慢的缺陷，Fast R-CNN[17]进一步改进了这一方法。Faster R-CNN[18]的问世进一步简化了目标检测的流程，通过引入区域建议网络(Region Proposal Network,RPN)实现了更高效的检测。而 YOLO[19]和 SSD[20]则通过将目标检测任务视为回归问题，在速度和准确度上均取得显著进展。近年来，RetinaNet[21]和 EfficientDet[22]则通过引入新的损失函数和改进网络结构，提高了在大规模数据集上的检测性能。深度学习目标检测的演进为解决遥感图像中的目标检测问题提供了强大的技术支持。在实际应用中，这些方法不仅提高了检测的精度，同时也降低了人工干预的需求，为遥感图像的自动化处理提供了新的可能性。

尽管深度学习方法在遥感目标检测中取得了令人瞩目的成果，但该领域仍然面临一系列挑战。首先，遥感图像通常具有复杂的背景和多样的目标类别，使得模型在不同场景下的泛化能力不足。其次，由于遥感图像数据的获取成本较高，样本量相对有限，导致模型容易受到小样本问题的影响。此外，不同传感器数据

的异构性和光照条件的不稳定性也给目标检测带来了额外的困难。因此，进一步研究和改进遥感目标检测方法，提高模型的鲁棒性和泛化能力，成为当前研究的重要方向。

1.2 通用目标检测研究现状

目标检测，作为计算机视觉领域的一个核心问题，旨在识别和定位图像中的各种目标。这项技术在许多应用中都至关重要，包括视频监控、自动驾驶车辆以及医学图像分析等。近年来，深度学习的兴起极大推动了目标检测技术的发展，特别是卷积神经网络的使用，已经成为目标检测算法的主流。这一趋势不仅赋予目标检测更强大的性能和准确度，也使其成为未来研究的主要方向，具备极高的研究价值。深度学习在目标检测领域的方法主要划分为两大类：一是基于候选区域的目标检测方法，也即双阶段目标检测方法；二是基于回归的目标检测方法，也即单阶段目标检测方法。

1.2.1 基于候选区域的目标检测方法

基于候选区域的目标检测方法是一类经典的目标检测方法，其核心思想是首先生成可能包含目标的候选区域，然后在这些区域上进行目标检测和分类。这样的方法通常包含候选区域生成、特征提取以及目标分类这几个步骤。在候选区域生成阶段，模型会对输入图像进行分割或提取一系列候选区域，这些区域被认为可能包含目标。候选区域的生成方法对最终检测性能有着重要影响，在这一阶段较为经典的方法是 Selective Search。对于每个候选区域，从原始图像中提取特征以供后续的目标分类使用。常用的特征提取方法包括使用 CNN 或手工设计的特征。在经过特征提取阶段后，模型对每个候选区域进行目标分类。通常使用分类器对提取的特征进行分析，以确定该区域是否包含目标，以及目标的类别是什么。

在基于候选区域的目标检测方法中，较为经典的检测模型是 RCNN，Fast RCNN 以及 Faster RCNN。R-CNN 是深度学习目标检测领域的开创性工作之一。该方法首次引入了卷积神经网络来提取图像特征，并通过选择性搜索生成候选区域，然后使用 CNN 进行目标分类。虽然 R-CNN 在准确性上取得了显著进展，但其训练和推理速度较慢，不够实用。为了提高 R-CNN 的速度，Fast R-CNN 引入了区域池化层，将整个图像输入网络，避免了对每个候选区域的独立处理。这一改进使得训练和推理速度得到显著提升，同时保持了较高的检测精度。Faster R-CNN 则进一步简化了目标检测流程，引入了区域建议网络，用于端到端生成候选区域，避免了选择性搜索的繁琐步骤。Faster R-CNN 相较 RCNN 以及 Fast RCNN

在速度和精度上都取得了巨大成功，成为深度学习目标检测的重要里程碑。Mask R-CNN[23]在 Faster R-CNN 的基础上引入了一个额外的分支网络，用于生成每个检测到的目标实例的二进制掩码。这使得 Mask R-CNN 不仅能够准确地定位目标，还能够为每个目标生成精确的分割掩码，从而实现了实例分割。同时，Mask R-CNN 则通过三个任务进行多任务学习，包括目标检测、区域建议和实例分割。这种多任务学习的方式有助于网络更全面地理解图像内容，提高了模型的综合性能。

虽然 RCNN 类方法在目标检测的准确度上有一定的优势，但也存在明显的缺点：1)高计算复杂性，最初的 RCNN 方法需要对数千个候选区域分别进行特征提取和分类，这导致了高昂的计算成本。尽管 Fast R-CNN 和 Faster R-CNN 通过共享计算和使用区域建议网络分别对此进行了改进，但计算成本相较于基于回归的目标检测方法仍然较高。2)速度较慢，由于需要两个阶段来处理图像，即先生成候选区域然后再分类，双阶段方法的处理速度通常无法满足实时应用的需求。尽管 Faster R-CNN 加入了 RPN 来加速候选区域的生成，但其速度仍然慢于基于回归的目标检测方法。3)训练过程复杂，RCNN 类方法通常需要多个阶段的训练过程，包括预训练分类网络、微调用于候选区域的网络以及训练区域建议网络等，这使得整个训练过程变得相对复杂和时间消耗大。4)内存和存储需求高，特别是在 RCNN 和 Fast R-CNN 中，需要存储每个候选区域的特征，这可能会导致巨大的内存占用和存储需求。5)难以处理大量小目标，RCNN 及其变种通常在处理图像中小目标或者密集目标时表现不佳，这部分是因为候选区域可能无法精确覆盖到所有小目标。

1.2.2 基于回归的目标检测方法

由于基于候选区域的目标检测方法的各种缺点，研究人员重新调整了网络结构，将候选区域的提取和分类检测网络合并，提出了基于回归的目标检测方法。基于回归的目标检测方法是一类利用深度学习技术的目标检测方法，其主要思想是直接回归出目标的位置和类别，而不需要生成大量的候选区域。这样的方法通常包括图像输入、特征提取、目标分类、边界框回归以及非极大值抑制等关键步骤。其中图像输入将整个图像输入网络，而非生成大量的候选区域，特征提取通过卷积神经网络等结构从图像中提取特征，这些特征将用于后续的目标分类和位置回归。目标分类对提取的特征进行分类，判断图像中是否存在目标以及目标的类别。边界框对于包含目标的图像，进行边界框回归，直接预测目标的精确位置。非极大值抑制对于重叠的候选框，使用非极大值抑制来消除冗余框，保留得分最

高的框。

在基于回归的目标检测方法中，较为经典的方法有 YOLO、SSD 以及 RetinaNet。YOLO 是一种以单一前向传播过程完成目标检测的算法，将目标检测任务转换为回归问题。它将图像划分为网格单元，并在每个单元上直接预测目标的位置和类别。YOLO 以其实时性而著称，使其在需要快速目标检测的应用中非常实用，但是其检测精度却并不理想。而 SSD 是一种多尺度目标检测算法，通过在不同层次的特征图上进行检测，实现了对不同大小目标的有效检测。它通过一次前向传播完成整个检测过程。其通过使用不同层次的特征图，能够有效地检测多尺度的目标，使其在复杂场景中表现更为出色。而 RetinaNet 引入了焦点损失，专注于解决目标检测中类别不平衡的问题。它同时使用了高级特征和低级特征来提高检测的准确性。RetinaNet 引入的焦点损失有效应对了目标检测中正负样本不平衡的情况，提高了对罕见类别的检测效果。其同时利用高级特征和低级特征，以更全面的信息来进行目标检测，提高了模型的表达能力。YOLOv2[24]引入了一系列改进，包括使用全卷积网络来提高边界框的精度，采用 Anchor Boxes 来改进目标位置的预测，以及引入了更多的数据增强技术。通过引入全卷积网络，YOLOv2 提高了目标边界框的精度，更好地捕捉目标的特征。引入 Anchor Boxes 使得模型能够更好地适应不同形状和尺寸的目标。更多的数据增强技术有助于提高模型的鲁棒性。YOLOv3[25]是 YOLO 系列中的第三个版本，引入了更深的 Darknet-53 网络结构，以及使用多尺度特征图进行目标检测。此外，YOLOv3 通过逐步预测不同尺度的边界框来提高检测性能。使用更深的网络结构提高了模型对图像特征的学习能力。引入多尺度检测使得模型能够更好地适应不同尺寸的目标。逐步预测不同尺度的边界框有助于提高检测的准确性。YOLOv4[26]进一步提高了目标检测的性能，引入了 CSPNet 结构、SAM 模块、PANet 等创新。引入 CSPNet 结构优化了信息的传递和网络的计算效率。SAM 模块引入了空间注意力机制，提升了模型对目标空间关系的学习能力。PANet 用于路径聚合，有助于更好地整合多尺度特征。PP-YOLOE 是百度所提出的目标检测模型，PP-YOLOE 使用 CSPRepResNet 作为其骨干网络，这是基于 CSPNet 的变体，集成了残差连接以提高特征提取的能力。特征融合在 PP-YOLOE 中通过 PANet 和 ET 头来实现。PANet 用于整合来自不同级别的特征图，以提高模型对不同尺度物体的检测能力。ET 头则进一步优化了任务对齐。在预测阶段，PP-YOLOE 引入了几个关键的改进，首先是锚框机制的改进，PP-YOLOE 移除了锚框依赖，并引入了锚框自由机制。其次是标签分配策略，PP−YOLOE 采用了任务对齐学习，一个动态的标签分配和任务对齐损失，以提高分类和定位之间的一致性。此外，它采用了解耦的焦点损失和分布式焦点损失，以提高损失函数的性能。这些预测阶段的优化改进了

模型对正样本的贡献权重，减少了模型在训练和推理中的不一致性。

这些基于回归的目标检测方法在速度和准确度上取得了显著的进展，相较于基于候选区域的方法更加高效。它们逐渐成为目标检测领域的主流方法，广泛应用于各种实际场景。

1.3 遥感图像目标检测研究现状

传统的遥感目标检测方法依赖于手工特征提取和机器学习分类器。这些方法通常包括图像预处理、特征选择、分类器设计和目标定位几个步骤。预处理可能包括图像去噪、增强对比度等操作，以改善图像质量。特征选择则关注于提取有助于区分不同目标的属性，如纹理、形状、颜色和光谱签名。这些特征随后被用来训练如支持向量机、决策树或随机森林等传统机器学习分类器，以实现目标的识别。最后，通过一系列的后处理步骤，例如形态学操作，来细化目标的位置和边界。

这些传统方法在一些简单或者控制良好的场景中可能效果良好，但它们在处理复杂场景时，如多尺度、不同角度或遮挡等条件下的目标检测，往往存在局限。它们通常依赖于手动调整的特征提取和严格的参数设置，这限制了它们在大规模和动态环境下的适应性和鲁棒性。随着深度学习方法在目标检测任务中显示出优异的性能，传统方法逐渐被这些更先进的自动特征提取和分类技术所取代。

基于深度学习的遥感目标检测方法主要包括两类：水平目标框检测以及旋转目标框检测。

1.3.1 水平目标框检测

水平目标框的遥感目标检测方法是指在遥感图像中识别和定位目标时，使用水平的边界框来标示目标位置的方法。这种方法在遥感目标检测中是最初和最常用的技术，由于遥感图像通常是从顶视角获取的，因此水平框通常足以覆盖目标区域。HRDNet[27]是一种针对遥感图像的高分辨率目标检测方法，通过引入高分辨率的特征图，提高了对小目标的检测精度。该方法专注于改进目标检测网络的特征表示，尤其在处理小尺寸目标时取得显著效果。一些方法通过使用更适应遥感场景的损失函数，如 IoU Loss[39]，来提高目标检测的性能。IoU Loss 专注于优化检测框的位置和形状，有助于更准确地捕捉目标的几何信息。Cascade R-CNN[28]是一种级联目标检测方法，它通过多阶段的检测过程逐渐提高检测器的准确性。在遥感图像中，这种级联结构可以帮助模型更好地捕捉目标的细节信息，提高水平框检测的精度。SCRDet[29]是专门为高分辨率遥感图像设计的目标检测

方法。它通过引入选择性上下文细化的机制，对目标的上下文信息进行精细化的建模，以提高对小目标的检测性能。

1.3.2 旋转目标框检测

使用水平目标框来对遥感图像中的目标检测会遇到许多的问题，例如 1)目标方向多变，遥感图像中的对象（如车辆、船只、建筑物）可能以任何角度出现，水平边界框可能会引入很大的背景区域，降低检测的准确性。2)空间布局复杂，遥感图像中的目标经常密集排列，传统的水平框可能会导致严重的目标间遮挡问题。为了解决这些问题，研究者们提出了旋转目标框检测方法。这种方法使用旋转的边界框来更精确地围绕目标。旋转框可以减少框内的无关区域，提高目标的定位准确性，同时减少目标之间的遮挡。

相较于水平目标框的遥感图像目标检测，旋转目标框的遥感图像目标检测增加了许多的关键技术点。例如 1)旋转框表示，旋转框通常由中心点坐标、宽度、高度和旋转角度表示。这要求算法能够预测这些参数。2)旋转框检测网络，深度学习模型需要进行调整或重新设计，以便于有效处理旋转框。这可能包括修改模型的预测头部、损失函数等。3)数据增强和预处理，由于遥感图像的特殊性，数据预处理和增强策略也需要适应旋转框的特点。

在遥感图像目标检测中，旋转目标框检测是一项具有挑战性的任务，因为一些目标在遥感图像中可能以不同的角度倾斜或旋转。Rotated Faster R-CNN[30]是对传统的 Faster R-CNN 模型的扩展，专门设计用于检测旋转目标框。它在生成区域建议时考虑了旋转，同时在后续的目标检测和分类中也考虑了目标的旋转。Rotation RPN 是一种用于生成旋转框的网络结构，它可以用于在遥感图像中生成旋转目标框的候选区域。这为后续的目标检测提供了更准确的输入。针对遥感图像中出现的旋转目标，Oriented R-CNN[31]引入了旋转框的概念，能够更好地适应水平框的检测。该方法通过考虑目标的旋转信息，提高了在遥感图像中检测目标的能力。RRPN[32]是一种专门用于检测旋转目标的方法，它在区域提议网络中引入了旋转区域，有助于更准确地提取目标的边界框。ArbOD[33]是一种用于遥感图像中任意方向目标检测的方法。它采用旋转框的思想，通过引入旋转感知特征并结合自适应的注意力机制，实现对任意方向目标的准确检测。该方法在处理旋转目标和复杂场景中表现出色。SA-AOD 是一种基于自注意力机制的方法，专注于提高遥感图像中任意方向目标的检测精度。它通过引入自注意力机制，使模型更有针对性地关注目标区域，从而提升检测性能。GWD[34]，GWD 是一种考虑几何信息的方法，特别适用于遥感图像中的旋转目标检测。它通过引入几何感知的

Wasserstein 距离，有效地处理不同形状和尺寸的目标，提高了对水平框的检测性能。RRPN++[35]是 RRPN 的进化版本，通过引入更多的旋转先验框和级联网络结构，进一步提高了旋转目标检测的性能。该方法在航拍图像等场景中具有较强的鲁棒性。R2CNN[36]是一种基于区域的旋转目标检测方法，引入了 RoI pooling 的旋转版本以及旋转边界框的回归。该方法能够有效地处理旋转目标的检测任务。R3Det[37]是一种基于特征金字塔网络的旋转目标检测方法，通过自适应的特征金字塔结构，提高了对多尺度旋转目标的检测性能。DBNet[38]是一种动态分支网络，专门用于旋转目标检测。它通过引入动态分支机制，自适应地选择不同分支网络，使得模型能够更好地适应不同旋转角度的目标。

在训练过程中，为了更好地适应旋转目标框，一些方法采用了特定的损失函数设计。这些损失函数通常考虑了角度信息，帮助网络更好地理解和预测旋转目标框。旋转目标框检测面临一些挑战，如目标形状的多样性、复杂背景和光照变化。未来的研究方向可能包括更强大的特征学习、更复杂的网络结构设计、更优秀的特征融合模块以及更有效的数据增强方法，以提高旋转目标框检测的性能。

1.4 主要研究内容

本文以改进遥感图像目标检测中所使用的特征融合机制为主要研究方向，在现有研究方法的基础上，进行了进一步的改进以及研究，具体的研究内容如下所示：

(1) PANet 虽然促进了不同层级特征的融合，但其性能对骨干网的特征提取能力有着隐形依赖。在骨干网特征提取能力较弱时，PANet 可能反而传递并放大噪声，影响检测性能。针对这些问题，提出了基于 PANet 改进的多尺度特征降噪与融合网络(MFDFN)，在特征融合的过程中对特征进行噪声抑制。

(2) 第一点中所提出的 MFDFN 在处理旋转和水平目标检测时展现出了较好的性能。然而，MFDFN 在应对遥感图像的多样场景方面存在不足，且其将原始特征以及降噪后特征进行等权相加的方法无法有效整合降噪后的特征信息与原始特征信息。为克服这些限制，本文提出了增强多尺度特征降噪与自适应融合网络(EMFDAN)，该网络进行了两方面的工作：1)对多尺度特征降噪模块进行改进，引入一个能够学习特定位置和图像信息的条件变量以增强特征降噪模块的性能；2)设计了一个权重自适应调整模块，通过自适应学习来调整两组特征中的局部与整体权重，最后进行相加操作，旨在加强重要特征的影响力，优化特征相加的过程。

1.5 主要内容与章节安排

本文主要研究了基于改进特征融合机制的遥感目标检测算法，针对遥感图像的特性以及原有特征融合策略在目标检测中存在的问题，改进了特征融合的计算方法。本文首先由人工智能和计算机视觉引出目标检测任务，并介绍了目标检测的背景、研究现状以及在遥感领域的应用。其次介绍了目标检测的相关基础理论知识，例如卷积神经网络及相关模块、激活函数等，同时对经典的多尺度特征融合算法以及注意力机制进行了介绍，最后介绍了遥感图像的特点以及经典的遥感图像目标检测算法。接着针对特征融合过程中因骨干网络特征提取能力而引起的噪声传播等问题，提出了基于 PANet 改进的多尺度特征降噪与融合网络(MFDFN)，在特征融合前对特征进行降噪，为后续的预测提供了更加准确及有效的特征。之后，为了进一步改进 MFDFN，提出了增强多尺度特征降噪与自适应融合网络(EMFDAN)，使得网络能够识别图像中的特殊和复杂信息，以及能够更加有效的融合特征信息，为后续检测流程提供更好的支持。

本文共分为五个章节，章节简介如下：

第 1 章：绪论。在本章首先简单介绍了从计算机视觉到深度学习的发展，着重介绍深度学习在目标检测领域的应用。接着阐述了目标检测的背景及国内外研究现状，并对目标检测的方法进行总结介绍，主要对基于候选框目标检测算法和基于回归的目标检测算法进行了梳理，同时对遥感图像目标检测中的水平目标框检测以及旋转目标床检测中的主流算法进行了简单的介绍。最后介绍了本文的主要工作和章节安排。

第 2 章：基础理论与方法。首先对神经网络的各组成部分和相关模块进行阐述。随后介绍多尺度特征融合网络和注意力机制的相关知识，接着介绍了遥感图像的特点以及经典的遥感图像目标检测算法，最后介绍了目标检测方法的评价标准。

第 3 章：提出基于 PANet 改进的多尺度特征降噪和融合模块(MFDFN)。首先对 Defomable Transformer 进行简要介绍，接着详细介绍了多尺度特征降噪模块(MDM)的设计细节，最后介绍了为多尺度特征降噪模块而设计的计算优化模块(COM)，使得在几乎不增加模型参数的情况下，该模块有效提高了模型的检测准确率。在进行了以上模块的介绍以及公式细节阐述之后，将 MFDFN 应用到 PP-YOLOE-R 上并在遥感图像数据集上进行了测试，之后将模块应用到其它通用检测方法上并在通用检测数据集上进行了泛化性测试。

第 4 章：提出基于 MFDFN 改进的增强多尺度特征降噪与自适应融合网络(EMFDAN)。首先介绍了一下条件变量在现有方法中的应用，接着详细介绍了条

件变量加强的多尺度降噪模块(CEMDM)的设计细节，然后介绍了权重自适应调整模块(WAAM)。在进行了模块的介绍以及公式细节阐述之后，以第三章所提方法为基线模型在在遥感图像数据集上进行了测试，最后为了更简单明了的阐述本章方法的有效性，将预测图片进行了目标检测可视化以及特征热图可视化。

第 5 章：总结与展望。针对遥感图像存在的问题和本文所提出的改进方法的优缺点进行分析，对改进方法进行总结。最后从五个方面对当前算法进行展望并提出下一步的改进方法。

第 2 章 基础理论与方法

2.1 卷积神经网络

卷积神经网络（CNN）是深度学习领域的一种核心算法，尤其在图像和视频数据的处理上展现出显著效能。该算法通过模仿人类视觉系统识别和处理图像中复杂模式的机制而设计。自 20 世纪 80 年代初的概念形成以来，CNN 经历了显著的发展。特别是 LeNet-5[40]的引入和 1998 年的开发，标志着早期的成功应用。2012 年，AlexNet[41]在 ImageNet 竞赛中取得了突破性成绩，极大地促进了该领域的研究和兴趣。此后，更深层次和复杂度的架构，如 VGG[42]、GoogLeNet[43]和 ResNet[44]等相继涌现，提高了图像识别任务的准确性和效率。CNN 的成功和广泛应用得益于硬件进步、大规模标记数据集的可用性以及持续的算法创新，已成为图像识别、自动驾驶、医疗影像分析等众多领域的关键技术。卷积神经网络主要包含卷积层、池化层及全连接层三个部分，其结构如图 2-1 所示：

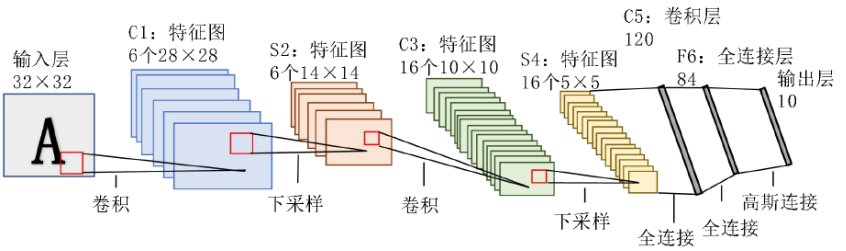

图 2-1 LeNet-5 网络结构图

2.1.1 卷积层

卷积层是深度学习中卷积神经网络的核心组件之一。它通过卷积操作对输入数据进行特征提取，具有平移不变性和局部感知性，使得 CNN 在图像、语音等领域取得了显著的成功。在深度学习中，卷积神经网络的卷积层起到至关重要的作用。在卷积层中，多个可学习的滤波器遍历输入数据，这些滤波器能够捕捉到数据中的局部特征。每一个滤波器都相当于一个特定的特征检测器，通过在输入数据上滑动并计算滤波器与数据局部区域的点积来提取特征。其运算如公式 2-1 所示：

$$F_j^l = f\left(\sum_{i=1}^{I} Y_i^{l-1} \times w_{ij}^l + b_j^l\right) \tag{2-1}$$

公式 2-1 中F_j^l是第 1 层中的第j个特征映射。在卷积层中，每个特征映射是通

过应用一个卷积核到前一层的输出来计算得到的。f表示激活函数。激活函数对卷积操作的结果进行非线性转换。I表示前一层的特征映射的数量。Y_i^{l-1}是第$l-1$层的第i个特征映射。在卷积层中，每个特征映射是通过前一层的输出与一个卷积核进行卷积操作得到的。×表示卷积操作。在这个公式中，它指的是将权重w_{ij}^l应用于相应的输入特征映射Y_i^{l-1}。w_{ij}^l是卷积核的权重。在第l层中，每个特征映射F_j^l有一个相应的权重矩阵。b_j^l是偏置项。每个特征映射有一个偏置，它被添加到卷积操作的结果上，用于调整输出。公式 2-2 描述了在卷积神经网络的卷积层中如何通过对前一层的特征映射进行加权卷积，再加上偏置，最后应用一个激活函数来计算当前层的特征映射。这是特征提取和模式识别的关键步骤。

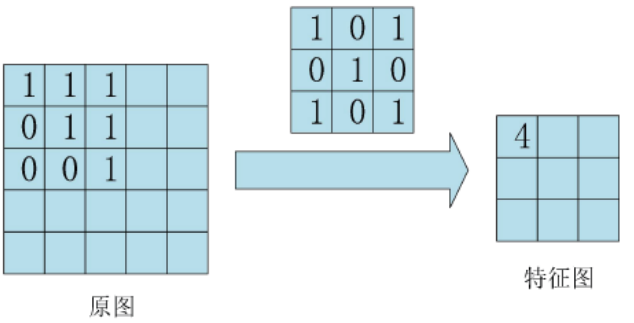

图 2-2 卷积层的卷积计算

如图 2-2 描绘了一个卷积神经网络中的卷积操作过程，其中 3×3 的卷积核覆盖输入图像的对应区域，按元素进行加权求和，生成了一个输出值。这个过程在整个输入图像上逐步重复进行，卷积核沿图像滑动，最终形成了一个新的特征映射，它揭示了输入图像中的空间特征，为深度学习模型的下一层提供了信息丰富的输入。这种卷积计算是卷积层核心功能的直观展示，它是特征提取和模式识别的关键步骤。

2.1.2 池化层

池化层是卷积神经网络中的重要组件，用于下采样和减小特征图的尺寸。池化操作通常通过对输入区域进行聚合来减少特征图的大小，有助于提取图像中的主要特征并减少计算负担。池化层在卷积神经网络中扮演着简化信息的角色。这一层通过对每个特征图进行下采样，减少数据的空间维度，从而减轻计算负担，同时使网络对小的位置变化保持不变性。池化操作通常有两种形式：1)最大池化：在每个池化窗口内选择最大的元素作为输出，用于强调输入区域的最显著特征。2)平均池化：在每个池化窗口内计算元素的平均值作为输出，用于对输入区域进行平滑处理。最大池化以及平均池化的操作如图 2-3 所示。

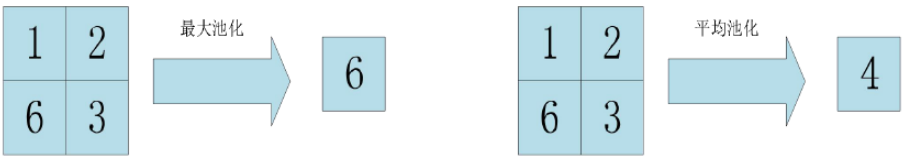

图 2-3 池化层中的最大池化以及平均池化

图 2-3 展示了卷积神经网络中的池化操作，具体来说是最大池化以及平均池化的计算过程。在最大池化中，网络从输入特征图的局部接受域中选择最大的元素作为该区域的代表，以此减小特征图的尺寸。如图所示，从左侧的 2×2 网格中选取了最大值 6 作为池化后的输出。在平均池化中，计算输入特征图的局部接受域中所有元素的平均值，并将此值用作该区域的输出，如图中的 2×2 网格平均值计算得到 4。这两种池化操作均用于降低特征维度、减少计算量以及实现对输入变化的鲁棒性。

在卷积神经网络中，通常会将卷积层和池化层交替堆叠，以构建深度网络结构。这种结构在保留图像主要特征的同时，有效地减小了计算负担。

2.1.3 全连接层

全连接层，也称为密集连接层或仿射层，是深度神经网络中的一种基本层类型。在全连接层中，每个节点都与前一层的所有节点相连接，形成一个完全连接的网络结构。在卷积神经网络中，全连接层通常位于网络的末端，起着整合局部特征并执行高级推理的作用。每个全连接层的神经元都与前一层的所有激活输出相连，因此它能够综合前面卷积层和池化层提取的局部特征，形成更为全面的全局特征表示。全连接层的参数数量较多，增加了模型的学习能力，使得网络能够在较高的抽象层次上对数据进行分类或回归。

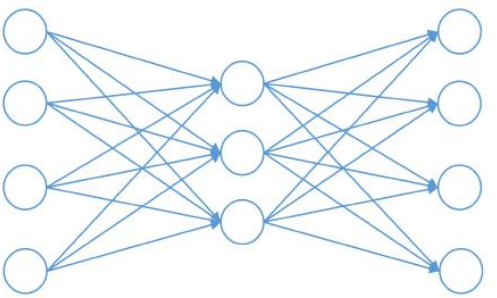

图 2-4 全连接层的网络结构

图 2-4 代表了卷积神经网络中全连接层的典型结构。在这种结构中，网络的这一部分由一系列完全互连的神经元组成，其中每个神经元与前一层的所有输出相连。具体来说，图中展示了一个三层的全连接网络架构，其中输入层的每个神

经元都与中间层（也称为隐藏层）的每个神经元通过权重连接，而中间层的每个神经元又与输出层的每个神经元全连接。

2.1.4 激活函数

卷积神经网络中的卷积以及池化等操作都只是线性变换，即使是多层卷积、池化操作叠加，最后得出的结果也是线性的。因此，研究人员在卷积神经网络中加入了非线性的激活函数，用于提高模型的学习能力。激活函数是深度神经网络中的一种非线性变换，它在每个神经元的输出上引入非线性，从而使得神经网络能够更好地学习和适应复杂的数据模式

Sigmoid 函数是一个平滑且连续可导的激活函数，它将输出限制在(0,1)范围内，适合用于二分类问题的输出层，以表示样本属于某一类别的概率。然而，在深度神经网络中，Sigmoid 函数存在梯度消失和输出不以零为中心的问题，可能导致训练速度缓慢。另一方面，ReLU 激活函数以其简单和高效而受到欢迎，尽管它在正数范围内呈线性，仍具有非线性性质，有助于实现神经网络的稀疏激活，从而提高了网络的表达能力和训练效率。

Sigmoid 函数公式为：

$$s(x) = \frac{1}{1+e^{-x}} \tag{2-2}$$

ReLU 函数公式为：

$$Relu(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2-3}$$

2.2 多尺度特征融合网络

多尺度特征融合是卷积神经网络中常用的一种技术，用于有效地捕捉不同尺度的信息。在图像中，目标物体的尺寸和位置可能因多种因素而变化。因此，单一尺度的特征提取可能无法充分捕捉到所有尺度上的有用信息。多尺度特征融合的目标在于提高网络对不同尺度物体的感知能力，从而提高模型的鲁棒性和泛化性能。

在目标检测和计算机视觉的研究领域，特征金字塔网络(Feature Pyramid Networks,FPN[45])、路径聚合网络(Path Aggregation Network,PANet[46])以及双向特征金字塔网络(Bi-directional Feature Pyramid Network,BiFPN[47])是三种关键的网络架构，它们各自的特点和应用优势为后续的研究提供了丰富的改进空间。

2.2.1 FPN

在计算机视觉领域，FPN 是一种用于增强卷积神经网络在处理多尺度物体检测任务中的能力的结构。特征金字塔网络由 Lin 等人在 2017 年首次提出，旨在解决传统 CNN 在处理尺度变化较大的物体时性能下降的问题。FPN 的模块结构图如图 2-5 所示。

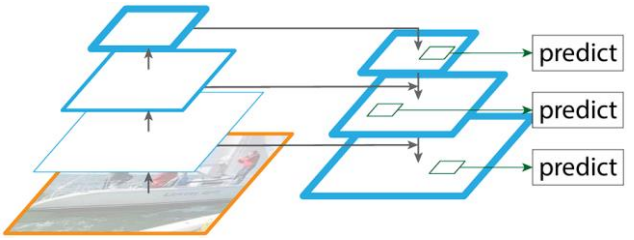

图 2-5 FPN 模块结构图

FPN 的核心思想在于构建一个由底层到高层不同分辨率的特征层次结构，并通过自顶向下的路径和横向连接有效地融合这些层次中的信息。在这一结构中，每一层都能利用来自下一层的高分辨率信息和来自上一层的高层语义信息。这样的设计使得网络在每个尺度都具有丰富的语义信息，从而更有效地处理不同尺度的物体检测任务。

具体而言，FPN 包括两个主要部分：自底向上的路径和自顶向下的路径。自底向上的路径是一个标准的卷积网络，它逐渐减小特征图的空间分辨率，同时增加其深度，捕捉更高层次的语义信息。而自顶向下的路径则是 FPN 的创新所在，它通过上采样和横向连接来增强高层特征图的空间分辨率，使之包含更多的细节信息。在自顶向下的路径中，每一层的上采样输出与来自自底向上路径相应层级的特征图通过横向连接进行融合。这种融合操作不仅为高层特征图提供了丰富的细节信息，还保留了低层特征图中的高层语义信息，从而在整个网络中实现了特征的丰富多样化。

2.2.2 PANet

PANet 是一种在深度学习和计算机视觉领域中，特别是在物体检测和实例分割任务中使用的高级神经网络结构。路径聚合网络最初由 Liu 等人在 2018 年提出，旨在进一步优化特征金字塔网络的架构，以提升物体检测和实例分割任务中对小物体的识别性能。PANet 的网络结构图如图 2-6 所示。

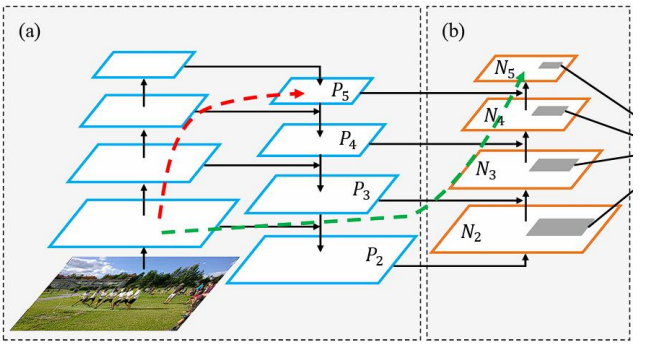

图 2-6 PANet 网络结构图

PANet 的主要创新在于增强了特征金字塔的信息流，使得低层特征能够更有效地与高层特征结合。与 FPN 相比，PANet 引入了额外的自底向上路径，以便更加充分地利用低层特征中的细节信息。在 PANet 中，自底向上的路径与 FPN 的自顶向下路径相互补充，从而形成了一个更为复杂和有效的特征聚合机制。

具体来说，PANet 在 FPN 的基础上增加了一个自底向上的路径，该路径通过逐层聚合低层特征来增强上层特征。这一过程不仅保留了高层特征中的丰富语义信息，还融合了低层特征中的细节信息，从而使得整个网络在处理细小物体时更加高效。此外，PANet 还采用了一种独特的适应性特征池化策略，进一步提高了对多尺度物体的检测能力。

相较于 FPN，PANet 的这些改进显著增强了网络对小尺度物体的识别能力，同时保持了对大尺度物体的强大检测性能。这一特点使得 PANet 在众多物体检测和实例分割任务中表现出色，尤其是在需要细粒度识别和精确边界划定的场景中。

2.2.3 BiFPN

BiFPN 是一种先进的神经网络架构，专为提高物体检测和实例分割任务中的特征融合效率和有效性而设计。BiFPN 首次由 Tan 等人在 2020 年提出，其核心目的是优化 FPN 的结构，特别是在高效性和精度之间取得更好的平衡。其中 BiFPN 结构图如图 2-7 所示。

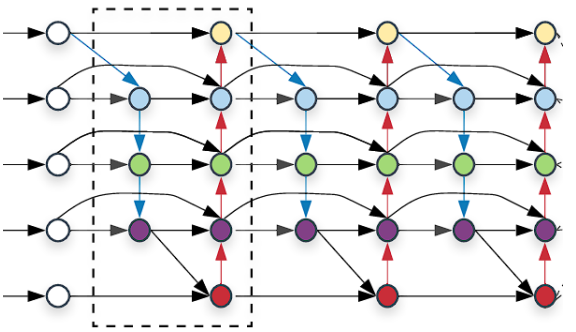

图 2-7 BiFPN 网络结构图

BiFPN 的主要创新在于引入了一个更高效的双向特征融合路径，与之前的 PANet 和 FPN 相比，这种结构更加注重在不同尺度特征间的信息交流。BiFPN 通过在特征金字塔中引入双向流动的信息路径，实现了更加有效的特征融合。这种双向流动不仅包括自底向上和自顶向下的信息流，还包括各个尺度之间的跨层连接。

具体来说，与 PANet 相比，BiFPN 的主要改进在于其使用了更加精简和高效的特征融合节点。在 BiFPN 中，通过引入称为“快速归一化融合”的方法，不同尺度的特征图在融合时能够更加高效和有效地利用计算资源。此外，BiFPN 还通过移除不必要的节点和边，减少了计算复杂度，同时提升了特征融合的效率。

这些改进使得 BiFPN 在处理多尺度特征时，不仅提高了物体检测的精度，同时也降低了计算成本。这一点在资源受限的场景下尤为重要，如在移动设备或嵌入式系统中进行实时物体检测。因此，BiFPN 在保持高精度的同时，还实现了对计算资源的高效利用。

2.3 注意力机制

注意力机制是一种使神经网络能够有选择性地关注输入的一部分信息的机制。不同变体的注意力机制，包括通道注意力和空间注意力，通过不同方式引导模型的关注，从而提高模型在处理复杂任务时的性能。

通道注意力机制关注于不同通道之间的关系，通过调整每个通道的权重，使网络更加注重对不同通道中重要信息的利用。通过全局平均池化获取每个通道的全局信息，再经过全连接层产生权重，最终将权重应用到每个通道上。空间注意力机制关注于不同空间位置之间的关系，通过调整每个位置的权重，使网络更关注图像中不同位置的重要信息。通过通道间的信息交互，学习每个位置的权重，将权重应用到每个位置上，以产生最终的输出。

CBAM[48]（Convolutional Block Attention Module）是一种在深度学习和计算机视觉领域广泛应用的注意力机制。该机制于 2018 年被提出，旨在增强卷积神经网络的特征表达能力。CBAM 通过聚焦于输入特征图的重要部分，有效地提升了网络对关键信息的捕捉能力，从而在多种视觉任务中实现了显著的性能提升。CBAM 模块主要包含两个子模块：空间注意力和通道注意力。这两个子模块按顺序对特征图进行加工，以此优化信息流, 如图 2-8 所示。

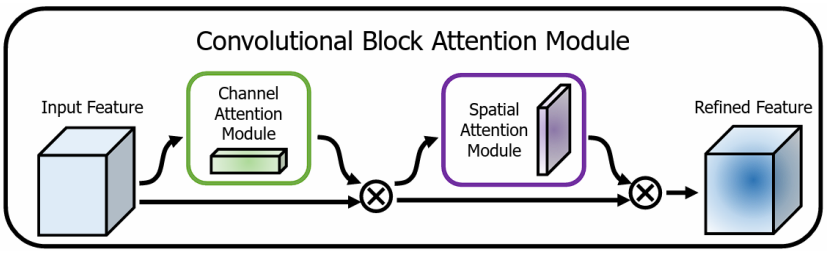

图 2-8 CBAM 结构图

通道注意力模块致力于识别特征图中哪些通道是最重要的。它通过利用全局平均池化和全局最大池化操作来聚集空间信息，然后通过共享网络层对这些信息进行处理，最后通过一个 Sigmoid 激活函数生成通道注意力图。这个图作为权重与原始特征图相乘，使得模型可以强调重要通道，抑制不那么重要的通道。CBAM 中的通道注意力结构图如图 2-9 所示。

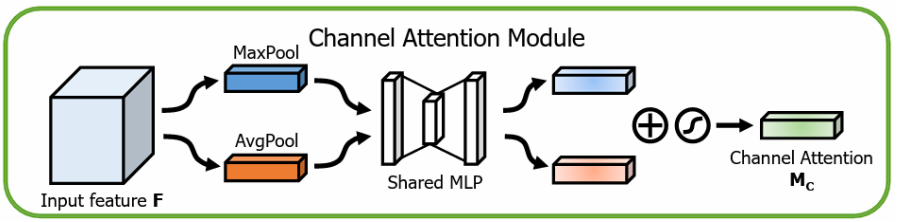

图 2-9 CBAM 中通道注意力结构图

空间注意力模块紧随通道注意力之后，它关注的是特征图中的哪些区域最为关键。该模块首先利用通道信息的全局平均池化和最大池化，然后将这些信息合并，然后通过一个卷积层，最终通过 Sigmoid 函数产生空间注意力图。这个注意力图与特征图相乘，使得网络能够集中于更有信息的区域。CBAM 中空间注意力模块结构图如图 2-10 所示。

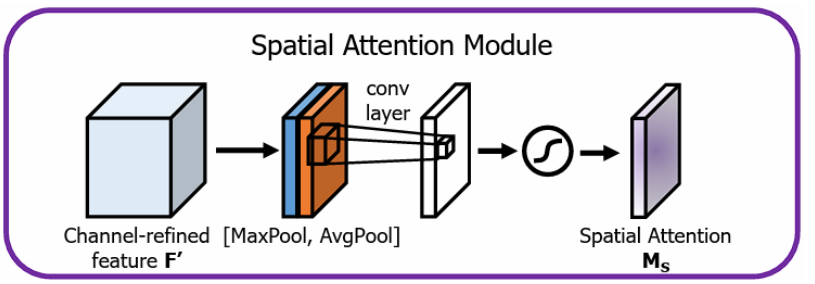

图 2-10 CBAM 中空间注意力结构图

在实际应用中，CBAM 可以无缝地集成到各种 CNN 架构中，如 ResNet、VGG 等，而不需要对现有架构做出重大修改。在图像分类、目标检测和图像分割等任务中，集成了 CBAM 的网络通常能够实现更好的性能。

2.4 Deformable Transformer 简介

Deformable Transformer[51]是一种改进的 Transformer[64]结构，旨在有效处理图像、视频和其他结构化数据。它通过引入一种名为可变形自注意力（Deformable Attention Mechanism）的机制，来改善原有 Transformer 在处理大规模数据时的性能和效率。这种机制特别适用于计算机视觉任务，如目标检测、分割以及视频分析等。

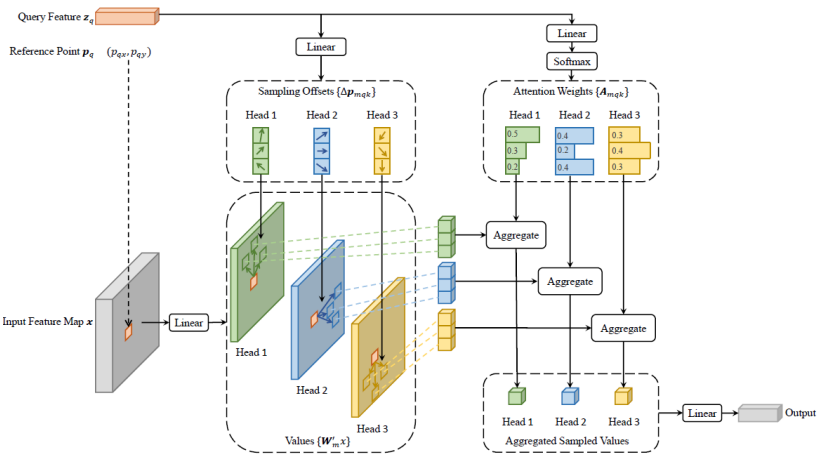

图 3-1 Deformable Transformer 计算流程

如图 3-1 所示，与传统自注意力机制不同，Deformable Transformer 不是计算输入元素与序列中所有其他元素的关系，而是只关注小部分关键元素。这种机制通过预选出对于给定任务最重要的特征位置（或像素），仅在这些选定的位置上计算注意力权重，从而大幅降低了计算复杂度。

$$\mathrm{MSDeformAttn}\left(z_q, \widehat{p_q}, \{x^l\}_{l=1}^{L}\right) = \sum_{m=1}^{M} W_m \left[\sum_{l=1}^{L}\sum_{k=1}^{K} A_{mlqk} \cdot W_{m'}\, x^l\left(\phi_l\left(\widehat{p_q}\right) + \Delta p_{mlqk}\right)\right] \quad (3\text{-}1)$$

公式 3-1 中使用的符号涉及多尺度可变形注意力机制的具体细节。这里，z_q 表示目标的查询特征向量，而$\hat{p}q$是这个目标预测的参照位置。公式使用$\mathrm{x}_{\mathrm{l}=1}^{\mathrm{lL}}$来表示一组不同尺度上的输入特征图，这些尺度从1到$L$变化。$M$代表注意力机制中使用的头的数量，其中$W_m$和$Wm'$ 分别指的是第m个注意力头的权重和用于变换特征图的线性变换矩阵。每个头为每个查询在每个特征层上采样的K个关键点中的每一个分配一个注意力权重$Amlqk$。最后，Δp_{mlqk}表示第m个头对第q个查询在第l层的第k个采样点的预测位置偏移，而$\phi_l\left(\widehat{p_q}\right)$是将预测位置映射到第$l$层特征图上的函数。

图 3-1 以及公式 3-1 展示了 Deformable Transformer 如何在多个尺度上聚合

来自不同特征图的信息，通过对特定的、学习到的位置（关键点）进动态加权求和来实现。这个过程允许模型根据每个查询的需求，灵活地关注不同特征图的不同区域，强化了对目标尺度和形状变化的适应能力。通过计算每个注意力头的输出并将它们加权求和，模型可以综合不同信息源的贡献，有效地提取和利用多尺度特征图中的目标信息。这种机制的引入大大提高了对象检测任务中的效率和准确性，展现了 Deformable DETR 在处理复杂视觉任务时的先进性。

2.5 Conditional DETR 及条件变量简介

DETR[62]利用 Transformer 的自注意力机制进行端到端的目标检测，去除了许多手工设计的环节，简化了流程，但受限于训练速度和注意力机制的性能。随后，Conditional DETR[63]通过使用条件变量进一步加速注意力机制的聚焦能力，从而提高了训练和推理速度，使得整个检测过程更加高效和精确。

Conditional DETR 引入的条件变量是其核心创新之一，旨在改进 DETR 模型的性能和效率。这一机制通过动态生成的查询对模型的注意力机制进行调整，以便更精准地定位并识别图像中的目标。以下是条件变量的几个关键方面：

(1) 动态查询生成。

在传统的 DETR 模型中，查询向量是在训练开始时随机初始化的，并且在整个训练过程中静态不变。这种静态查询方法可能导致模型在初期训练阶段效率较低，因为它需要较长的时间来调整这些查询，使其能够有效地匹配图像中的目标。相比之下，Conditional DETR 通过条件变量引入了一种基于图像特征动态生成查询的方法。这意味着每一步生成的查询都是条件化的，即它们是根据当前图像的内容和上下文信息生成的。这样，每个查询都直接关联到图像中具体的特征或区域，使模型能够更快速、更直接地关注到目标对象上。

(2) 提高注意力机制的聚焦性。

通过使用条件变量，Conditional DETR 的解码器能够更有效地将注意力集中在图像的有意义区域上。在解码过程中，根据先前解码步骤的输出和图像特征动态调整查询，使得每一步的注意力都更加聚焦于潜在的目标区域。这种聚焦性有助于模型更准确地预测目标的位置和类别，同时减少了对背景或不相关区域的注意，从而提升了检测精度。

(3) 加速模型收敛。

条件变量通过提高解码器对目标的定位能力和减少无关区域的干扰，有助于模型在初期阶段就能更快地学习到有效的特征表示。这种效率的提升直接反映在训练过程中，使得 Conditional DETR 能够在较短的时间内达到或超过传统 DETR

模型的性能水平。

总的来说，Conditional DETR 中的条件变量通过动态生成与图像内容紧密相关的查询，优化了模型的注意力分配，从而提高了目标检测的准确性和训练效率。

2.6 门控单元

门控机制是一种在深度学习领域广泛应用的技术，特别是在需要动态控制信息流的场景中。该机制通过引入可学习的参数来调节信息的传递强度，从而实现对特征表示的精细管理。在进行多源特征融合的背景下，门控机制能够为每个特征源分配一个动态权重，以优化其在最终融合特征中的贡献度。

具体地，对于来自不同特征提取路径的特征集合(F_a, F_b)，通过引入门控信号

$$\mathrm{G_i} = \sigma(\mathrm{W_i} \cdot F_\mathrm{i} + \mathrm{b_i}) \tag{4-1}$$

其中W_i和b_i 是门控层的学习参数，σ表示 Sigmoid 激活函数，每个特征F_i的贡献度被相应的门控信号G_i所加权。这种方法不仅为模型提供了在不同特征之间进行动态权衡的能力，而且也促进了特征间的有效融合，提高了模型在处理复杂数据情况下的灵活性和鲁棒性。随后，加权后的特征通过直接相加或更复杂的融合结构进行整合。例如，相加操作$F_{fusion} = F'\ a + F'\ b$或者通过卷积处理拼接后的特征 $F_{fusion} = Conv\left(\left[F'\ a, F'\ b\right]\right)$。这些融合策略不仅提升了模型对关键信息的捕捉能力，也增强了模型在不同抽象层面进行信息整合的能力。

综上，门控机制为特征融合提供了一种动态且高效的策略，使模型有能力根据不同的输入和任务要求调整特征间的相对重要性。通过精心设计的实验和细致的模型调优，可以充分发挥门控机制在多源特征融合中的优势，进一步提升模型的性能。

2.7 遥感图像的特点

遥感图像因其独特的获取方式和应用背景，在目标检测领域展现出若干特点，这些特点为旋转目标检测带来了特殊的挑战与需求。从旋转目标检测的角度来看，遥感图像的主要特点包括：

(1) 高视角

从高空拍摄的遥感图像展示了地面目标的顶视图，与人们习惯的侧视图或正面视图相比，形状和尺寸信息可能大不相同。例如，从高空看，一个圆形水塔可能只显示为一个圆圈，缺少了高度信息。这种高视角导致了目标特征的变化，要

求检测算法能够学习到不同角度下目标的特征表示。因此，深度学习模型需要在大量的遥感数据上进行训练，以学习到从高视角观察目标时的各种特征。

(2) 大尺度变化

遥感图像覆盖的范围广阔，导致图中同一类目标可能有非常不同的尺寸，这对目标检测算法是一个挑战。例如，同一张图像中既可能包含远处的小型船只，也可能包含近处的大型船只。为了应对这一挑战，旋转目标检测算法需采用多尺度处理方法，如将图像在不同尺度下进行处理或设计能够适应不同尺度目标的检测网络，进而提高对各尺度目标的检测能力。

(3) 旋转任意性

由于遥感图像的目标可能以任何角度存在，这就要求检测算法能够识别和定位出旋转的目标。传统的水平边框在这种情况下效率低下，因为它们不能精确包围旋转目标，会导致较大的空间浪费和高的误检率。因此，采用旋转边界框是一个更佳的解决方案，它能够紧密地围绕目标，减少背景干扰，从而提高检测的准确性和效率。

(4) 复杂背景

遥感图像中的背景复杂度高，目标之间的视觉差异可能非常微小，这要求检测算法具备较强的背景抑制和目标特征提取能力。为了应对这一挑战，可以采用深度学习算法，尤其是卷积神经网络，通过其强大的特征学习能力，有效区分目标和复杂背景。此外，目标检测算法还可以通过引入注意力机制，进一步提升模型对感兴趣目标的识别准确性。

(5) 密集排列与遮挡

在一些特定的遥感图像应用场景中，如港口或机场，目标（船只、飞机等）可能非常密集，并且会互相遮挡。这种密集排列和遮挡现象给旋转目标检测带来了额外的挑战，因为算法不仅需要正确检测出每一个目标，还要准确地识别每个目标的旋转角度。对此，可以通过设计更加精细的网络结构，例如使用高级别的特征融合策略和更加精确的锚框设计，以提高对密集排列和部分遮挡目标的检测精度。

如图 2-11 所示，左上角图片包含了船以及船港，其中的目标密集以及变化尺度大（船舶很小，船港很大），并且有多个方向的船只。右上角图片包含了多个球场，其中球场的颜色和周围树木的颜色非常类似。左下角的图片则是机场，其中包含了几类变化尺度非常大的目标，必去具有方向不定的飞机，以及相较飞机小非常多的汽车。而右下角则是一个工厂，其中包含了白色的储蓄罐以及小型的汽车。

图 2-11 遥感图像

针对以上特点，旋转目标检测在遥感图像中的应用需要采取特定的策略和改进方法，如采用旋转检测框架、改进的尺度自适应机制、复杂背景下的目标识别技术等，以提高检测的准确率和鲁棒性

2.8 经典遥感图像目标检测算法

遥感图像目标检测是通用目标检测的一个分支，根据标注框的类型，可以将遥感图像目标检测分为水平目标检测以及旋转目标检测。旋转目标检测在遥感图像中有许多的优势，因此本文更加关注遥感图像目标检测中的旋转目标检测。本文将在本小节对经典的遥感图像中的旋转目标检测算法，如 S2Anet、FCOSR 以及 PP-YOLOE-R 进行简单的讲解。

2.8.1 S2ANet

S2ANet[49]是一种专门针对旋转目标检测设计的深度学习模型，主要用于遥感影像中的船只、飞机等对象的检测。它结合了特征对齐和旋转框检测机制，以提高在具有复杂背景和不规则分布的场景中对旋转目标的检测精度。其整体架构如图 2-12 所示。

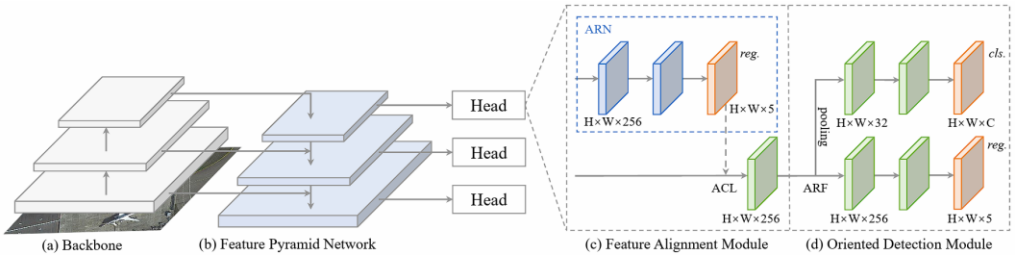

图 2-12 S2ANet 模型结构图

S2ANet 是基于候选区域的目标检测方法，也即二阶段目标检测方法，为了改进传统的水平或垂直 RoI 对齐技术，S2ANet 采用一个改进的 RoI 对齐层，它通过考虑目标的旋转框来对特征进行对齐，从而获取更加准确的目标表示。同时，为了解决由于旋转引起的目标特征不一致问题，S2ANet 引入了一种特征对齐模块。这个模块能够调整卷积特征图以匹配旋转目标的实际方向，从而为旋转检测提供更稳定的特征表示。最后，S2ANet 设计了一个专门的旋转框检测头，它能够输出目标的位置坐标以及旋转角度信息。这使得 S2ANet 不仅能够检测出目标的存在，并且还能准确识别出其朝向，对于许多应用场景（如遥感影像分析）而言至关重要。

S2ANet 通过这些设计，有效地提高了对旋转目标的检测精度和效率，在多个遥感影像数据集上显示出了优异的性能。尤其是在处理带有复杂旋转背景和高密度目标场景的遥感图像时，它能够比传统的检测框架更准确、更高效地进行目标检测。

2.8.2 FCOS-R

FCOS-R[50]是一种单阶段无锚点旋转目标检测器，其主要用于遥感图像目标检测。在 FCOS 的基础上，FCOS-R 简化了架构，使其更易于部署，并且在各种平台上都表现出良好的性能。其网络结构如图 2-13 所示。

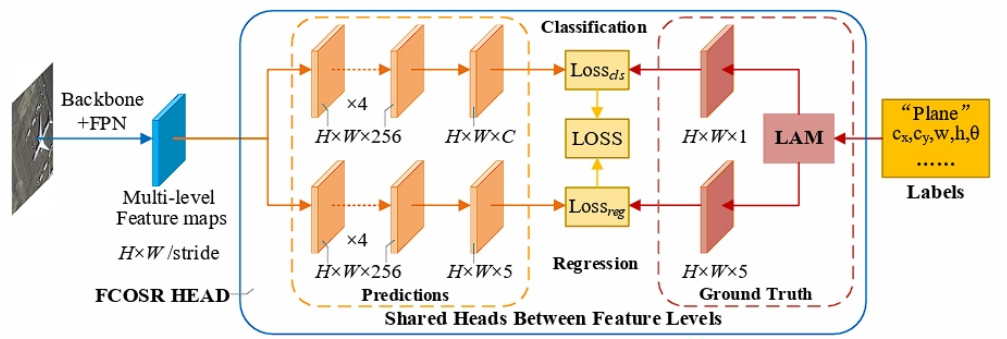

图 2-13 FCOS-R 模型结构图

FCOS-R 以一个骨干网络和特征金字塔网络（FPN）作为基础，生成多级特征图，这些特征图随后被送入 FCOSR 头部。FCOSR 头部负责生成预测，并包含了共享头部，这些共享头部在不同的特征级别之间进行预测。模型的预测部分在头部的左侧，而右侧则是训练阶段生效的组件，包括分类和回归损失计算以及标签分配模块（LAM）。LAM 负责将标签分配到每个特征图，而分类和回归分支则直接预测目标的中心点、宽度、高度和角度。

FCOSR 的主要特点是它的简洁性和易部署性，这得益于其无锚点的设计和对于旋转边界框的直接预测。通过引入椭圆中心采样和模糊样本标签分配策略，它能够有效处理重叠目标的标签分配问题，并且通过多级采样来解决大宽高比目标的采样不足问题。

2.8.3 PP-YOLOE-R

PP-YOLOE-R[57]是对 PP-YOLOE 目标检测框架的改进，主要针对旋转框（rotated bounding boxes）检测进行优化。这种改进的目标检测器在处理任意方向的对象时更为高效，特别适用于航空图像和场景文本，其中对象可能以任意方向出现。PP-YOLOE-R 采用了一系列技术和方法来提高检测精度和减少计算成本，使其在旋转对象检测任务中达到了最先进的性能，其整体架构如下图 2-14 所示。

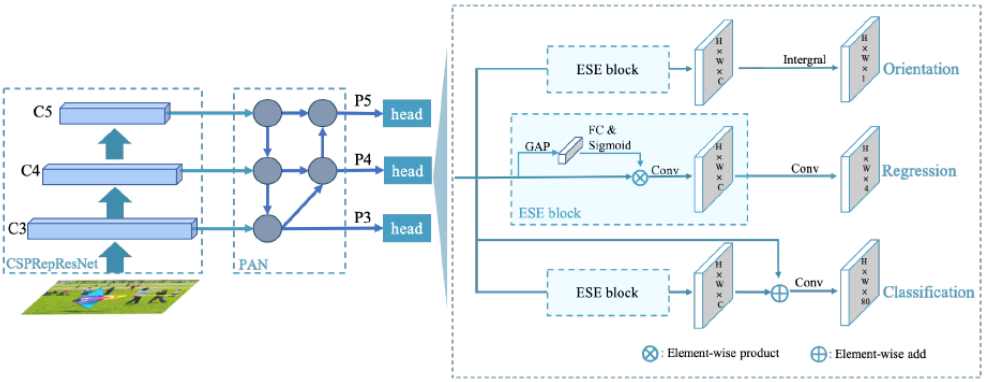

图 2-14 PP-YOLOE-R 模型结构图

PP-YOLOE-R 包含四个部分：输入、骨干、颈部以及输出。特别的，PP-YOLOE-R 在输入端集成了随即旋转以及随机翻转数据增强，以适应旋转框数据集的特性，因此网络在训练期间对遥感图像数据集有更好的拟合性。其主干网采用 CSPRepResNet，增强了特征提取能力的同时加快预测速度。此外，PP-YOLOE-R 使用 SPP 对特征图进行不同尺度的池化来增强模型对输入图像尺度的适应性。之后，网络应用路径聚合网络 PANet 对不同层级的特征进行融合，以提高小目标检测的性能。最后，PP-YOLOE-R 拥有三个检测头，分别用于角度预测、目标框

回归以及对象分类。PP-YOLOE-R 的激活函数为 ReLU 函数，这是一种非线性激活函数，用于增加模型的表达能力。

2.9 目标检测评价标准

在目标检测数据集中，研究人员将数据集按照不同比例分为训练集、验证集以及测试集。其中训练集用于训练目标检测模型，是模型需要进行拟合的数据样本，用来确定模型权重。验证集是模型训练工程中单独留出的样本集，用于调整模型的超参数，如网络层数、网络宽度、迭代次数、学习率等，和初步评估模型的检测能力。在模型训练过程中，验证当前模型泛化能力（准确率，召回率等），防止模型异常（过拟合等），决定是否停止提前训练。测试集用于评估模型的泛化能力。目标检测中常用的评价标准有准确率、召回率、mAP、FPS 等。

准确率指标度量了检测算法识别出的目标中，正确识别的目标所占的比例。它是一个关键指标，因为它直接关联到算法产生的误报数量。高准确率意味着算法在识别目标时，产生很少的误报。准确率的计算公式如公式 2-4 所示：

$$Precision = \frac{TP}{TP+FP} \tag{2-4}$$

其中 TP 代表真正例，也即正确识别的目标，FP 代表假正例，也即错误标记的非目标。

召回率评估了所有应该被检测出的目标中，算法实际检测出的比例。一个高召回率意味着算法能够捕获到更多的真实目标，减少漏检。召回率的计算公式如公式 2-5 所示：

$$Recall = \frac{TP}{TP+FN} \tag{2-5}$$

其中 FN 代表假负例，未检测到的真实目标。

平均准确率（Average Precision, AP）：AP 计算了在不同召回率水平上准确率的积分。它提供了一个单一的数值，表明在整个召回率范围内，算法的表现如何。高 AP 值意味着在所有可能的召回率阈值上，算法都能保持较高的准确率。

mAP(mean Average Precision)：当我们处理涉及多个类别的目标检测时，每个类别都会有一个 AP 值。mAP 是所有类别 AP 值的平均，因此它提供了跨类别的整体性能度量。mAP 是目标检测领域中最常用和最重要的评价指标之一。

FPS(Frames Per Second)：这一指标衡量的是算法处理输入数据的速度。在视频或实时监控场景中，处理速度是至关重要的。高 FPS 值指出算法可以实时处理视频流，而不会引起显著的延迟，这对于实时应用尤为重要。

2.10 本章小结

在第二章中，本文探讨了卷积神经网络的基础架构及其在图像处理领域的应用扩，同时对遥感图像的特点以及经典遥感目标检测算法进行了简单的介绍。章节内容详细阐释了卷积神经网络的关键组成部分，包括卷积层、池化层、全连接层，以及激活函数在内的多种元素，并对它们在特征提取和模式识别中的作用进行了分析。之后，本章还详细的介绍了目前主流的多尺度特征融合方法及其在目标检测方法中的重要性，为后续的改进奠定了理论基础。此外，本章细致讨论了注意力机制如何优化网络性能，通过突出图像的关键特征来增强模型的判别能力。通过对这些技术的评述，本章不仅揭示了它们各自的贡献，还阐明了它们如何协同工作，以实现对复杂视觉任务的有效处理。

进一步地，本章还对遥感图像的特点以及经典遥感图像目标检测算法进行了介绍，阐述了在遥感图像目标检测中可能会遇到的困难以及经典的遥感图像目标检测算法是如何解决这些困难的。综上所述，本章内容为理解卷积神经网络在计算机视觉中的应用，以及遥感图像目标检测的特点和难点提供了坚实的理论基础，并对后续实验研究的设计及其结果的分析提供了理论支撑。

第 3 章 多尺度特征降噪与融合网络

3.1 引言

PANet 在结构简单的情况下做到了较好的融合不同层级的特征，但是特征融合后的效果却主要取决于骨干网所提供的特征，当骨干网特征提取能力较弱时，PANet 会起到一定的副作用。详细来说，在 PANet 架构中，骨干网络的输出特征在不同层级间进行聚合，旨在增强多尺度特征的融合与信息流动。然而，若骨干网络的初始特征提取层缺乏鉴别力或对输入数据的噪声过于敏感，其生成的初级特征可能包含较高比例的噪声或不具代表性的信号。由于 PANet 的设计理念在于加强各层级间的特征交流，这种初始层的不足可能导致底层噪声或误差在后续层间的传递与放大。

此外，PANet 通过引入额外的上采样和下采样路径，进一步增强了特征在不同尺度间的流动。在这个过程中，如果骨干网络无法有效地过滤或纠正这些初级特征中的噪声成分，那么这些噪声元素可能会在整个网络中得到不成比例的强调，从而影响最终的检测性能。这种现象尤其复杂背景的目标检测任务中更为显著，因为在这些情况下，对初始特征层的准确性和鲁棒性要求更高。

因此，虽然 PANet 在理论上能够提高特征表达的丰富性和多尺度信息的利用，但它对骨干网络的质量有着隐含的依赖。针对以上问题，本文提出多尺度特征降噪与融合网络(Multiscale Feature Denoising and Fusion Network,MFDFN)，以此改进 PANet 中因骨干网提取特征能力而引起的特征融合过程中的噪声传播问题，从而提升其在遥感图像目标检测中的精度。本文所提出的 MFDFN 主要由两个模块组成：

(1) 多尺度降噪模块(Multiscale Denoising Module,MDM)，该模块以 Deformable Transformer 编码器为核心，主要是为了解决在 PANet 中因骨干网络特征提取能力不足而引起的特征融合过程中的噪声传播问题。该模块在特征融合前使用编码器在空间上对特征进行多尺度可变形自注意力计算，以达到降噪的目的。并在降噪后使用原特征进行信息补充，避免降噪过程中的信息丢失。从而达到在原有基础上进行进一步提升特征质量的效果，改善了在特征融合过程中由骨干网络特征提取能力而引起的噪声传播问题。

(2) 计算优化模块(Computational Optimiza-tion Module,COM)，在几乎不增加参数量的情况下对多尺度降噪模块的计算数据进行分组的权重计算，使得多尺度

降噪模块在降噪过程中关注更加重要的特征，降低干扰因素对降噪模块的影响，提高检测效率。

3.2 多尺度特征降噪与融合网络的设计与实现

本章在 PANet 的基础上提出 MFDFN，解决目前特征融合过程中由于骨干网络的特征提取能力不足而引起的噪声传播问题。该网络由 MDM 以及 COM 组成。MDM 以 Deformable Transformer 编码器为核心，在自上而下的特征融合以及自下而上的特征融合前，MDM 对待融合的多尺度特征进行降噪处理。同时，MDM 使用长短跳连将经过模块计算的前后特征相加，使得网络在不损失特征信息的情况下，对一组特征实现降噪功能，为后续的多尺度特征融合提供更加完整且准确的特征。其次，COM 以分组通道注意力为核心，在 MDM 进行降噪前对特征进行分组的权重计算，使得网络在几乎不增加参数量的情况下，为 MDM 提供更加准确且有效的特征图，从而使得后续的特征融合更加的准确以及高效。MFDFN 的结构如图 3-2 所示：

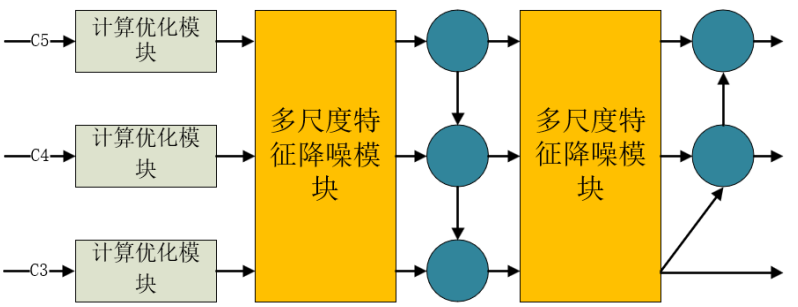

图 3-2 多尺度特征降噪与融合网络结构图

MFDFN 的整个计算流程可以分为降噪部分以及特征融合部分。降噪部分首先对 C3、C4 以及 C5 进行特征优化，随后送入 MDM 中进行降噪。降噪之后的特征将进行自上而下的特征融合，在这一步由于特征已经经过了降噪，所以在进行特征融合时高层的优质语义特征能够更好的传导至底层。经过特征融合之后又是一次降噪处理，随后进行自下而上的特征融合。由于之前的步骤已经经过了两次降噪，一次特征融合，所以在这时，底层特征的噪声已经被大大降低，自下而上的特征融合过程中优质的底层纹理特征能够正确的传导至高层，从而进一步提升检测性能。

3.2.1 多尺度特征降噪模块

受到 Deformable Transformer 在处理多尺度特征上的独特优势启发，本文结

合一些预处理以及后处理操作构建了 MDM。该模块在特征进行特征融合之前，利用 Deformable Transformer 编码器完成多尺度自注意力计算，这一步骤有效地利用了来自高层的质量较高的语义信息，对底层特征中的相对位置的特征进行了正确的强化同时对非任务相关的纹理特征（即噪声）进行了有效抑制。以 Deformable Transformer 编码器为核心，MDM 结构如图 3-3 所示：

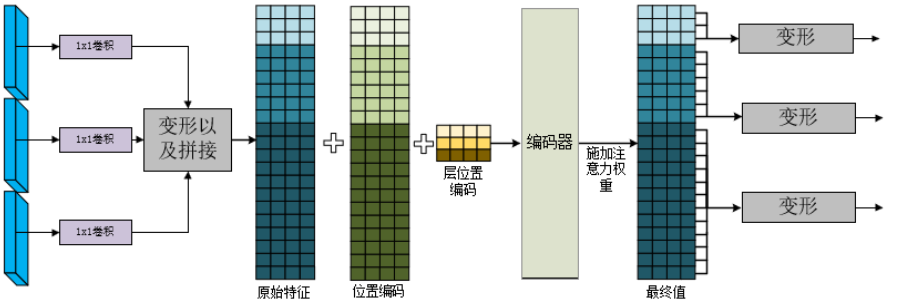

图 3-3 MDM 结构图

如图 3-3 所示，在自上而下以及自下而上的特征融合前，将三层特征进行变形、拼接后送入编码器中进行自注意力计算。同时，由于在编码器计算前将特征图通道进行压缩以及计算后又需要还原通道的这个过程中会导致部分特征信息的损失，所以本文将编码器压缩通道并计算自注意力前后的特征图相加，将编码器自注意力计算前后且未压缩通道的特征图相加，从而保证在进行自注意力计算的同时达到信息不丢失的目的。

$$\mathrm{X}_{\mathrm{u}}^{\mathrm{l}} = \mathrm{Conv}\left(\mathrm{X}^{\mathrm{l}}\right),\ l \in 1,2,3 \tag{3-2}$$

$$Z^{l} = \mathrm{MSDeformAttn}\left(Z_q^l, \widehat{p_q^l}, x^l\right) + Z_q^l,$$

$$Z_q^l = \mathrm{ShapeTransform}(X_u^l) \tag{3-3}$$

$$Y_d^l = \mathrm{ShapeTransform}^{-1}\left(\mathrm{Conv}^{-1}(Z^l)\right) \tag{3-4}$$

$$Y^l = Y_d^l + X_u^l \tag{3-5}$$

公式 3-2 中，$\mathrm{Conv}(\cdot)$表示卷积操作用于通道数统一。公式 3-3 中的Z_q^l是X_u^l形状变换后的结果，用于自注意力计算。跳连相加在自注意力操作后进行。公式 3-4 中的$\mathrm{Conv}^{-1}(\cdot)$和$\mathrm{ShapeTransform}^{-1}(\cdot)$表示将特征图恢复到原有的通道数和形状。公式 3-5 则将原特征X_u^l与经过多尺度可变形自注意力计算后的特征Y_d^l进行相加，得到最终的特征，用于算法的后续流程。

总的来说，首先将从骨干网得到的三层特征图$X^l, l \in 1,2,3$的通道数统一，假

设统一后的特征图为X_u^l。其次为了适配 Deformable Transformer 的计算格式，对特征图X_u^l进行形状变换得到x^l，并进行多头自注意力计算。计算前后，进行跳连相加操作，增加特征融合的能力。然后将计算后的特征图通道数恢复并变换回原来的形状，记为Y_d^l。最后与变换前的特征图X_u^l相加，得到最终融合特征Y^l。

整合上述步骤，得到的最终融合特征Y^l即可用于后续的自上而下特征融合等操作。此过程充分利用了 Deformable Transformer 编码器的自适应特征采样能力，以增强特征融合的效果。

3.2.2 计算优化模块

通道注意力能够对不同部位的特征分别进行聚焦，使得模型更加关注到重要的特征图。而在多头可变形自注意力中，其将特征分为多个组，分别对组进行自注意力的计算，这不仅能够降低计算量，还能提高模型的泛化性能。因此，本文设计出 COM，在 MDM 计算前将特征按照通道进行分组并以组为单位进行通道注意力的计算。其结构如图 3-4 所示：

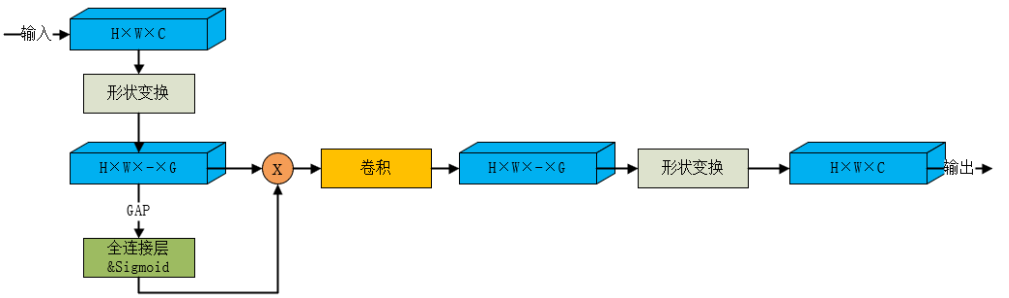

图 3-4 计算优化模块结构图

如图 3-8 所示，COM 的应用使得特征图在进入多头自注意力模块之前已经对重要的通道特征进行了强化和优化。这使得自注意力模块可以在一个更加聚焦和高效的特征表达上工作，相较于直接使用原始特征图，这种前置的分组通道注意力能够提供更丰富的信息，以支持自注意力模块捕获更细致的全局依赖关系。其计算公式如下所示：

$$F_i^{'} = \text{Reshape}(F_i, [Group \times B, C/Group, H, W]) \tag{3-6}$$

$$P_i = \text{AdaptiveAvgPool}\left(F_i^{'}, (1,1)\right) \tag{3-7}$$

$$F_i^{''} = Att\left(F_i^{'}, P_i\right) + F_i^{'} \tag{3-8}$$

$$\widetilde{F_i} = \text{Reshape}\left(F_i^{''}, [B, C, H, W]\right) \tag{3-9}$$

在公式 3-6 中，每个输入特征图首先通过重塑步骤，调整其维度以适配于注意力计算，这主要是为了将特征图分配到不同的头上，以便并行处理，其中F_i为原始特征，$Group$与可变形自注意力中的组数相对应。接着，执行自适应平均池化以提取每个通道的全局上下文信息，这一信息被压缩到一个单一数值中，反映了整个特征图的平均响应。然后，这些全局上下文信息与原始特征图结合，并通过一个注意力机制进行加权，加权操作旨在增强对重要特征的关注，同时保留那些原先就较为显著的特征，其中Att为通道注意力。最终，将加权得到的特征图重塑回其原始的维度，保证了特征图的空间结构不受影响。此过程在不同的特征图上重复执行，以实现对整个特征集合的细致调整，从而为后续的可变形自注意力计算提供了更加优质的特征。

总的来说，MFDFN 实现了特征图的多级别优化。首先在通道级别通过 COM 加权强化关键信息，然后在空间级别通过 MDM 对特征进行降噪处理。这样的分级处理既提升了模型对特征的利用效率，也增强了模型对全局信息的捕获能力，有助于提高模型对于各种任务的表现。

3.3 实验设计与结果分析

3.3.1 实验环境与实验设置

(1) 实验环境

对于本章所进行的所有实验，将采用统一的实验环境。实验环境为百度 BML Codelab，显卡采用英伟达的 TeslaV100，其最大显存为 32G；所使用的 CUDA 版本为 11.8，PaddlePaddle 版本为 2.5.2。本文的对比实验以及本章所搭建的旋转目标算法模型均使用 PaddlePaddle 所提供的 PaddleDetection 框架。

(2) 实验设置

对于本章节的实验仿真，本文将在遥感数据集 DOTA 和 DIOR-R，无人机数据集 VisDrone，通用目标检测数据集 COCO 以及 VOC 上进行。在所有的对比实验中，本文尽量采用统一的实验参数进行实验。在训练阶段，所有实验的 epoch 数量为 300，采用带动量的 SGD 随机梯度下降，初始学习率为 0.008，动量设置为 0.9，动量衰减系数为 0.0005。采用余弦退火策略以及线性 warmup 策略对学习率进行调节，其中最大 epoch 为 360，warmup 最大迭代次数为 1000，warmup 起始学习率为 0。与此同时，为了得到更好的训练效果，在训练阶段也会采用多种数据增强方式，如图片的随机翻转、图片的随机旋转、图片的随机缩放等方式。对于衡量检测精度的 mAP，使用 IoU 阈值为 0.5 时的 mAP 作为指标。

3.3.2 在遥感图像数据集上的实验结果

(1) DOTA-v1.0[54]数据集以及 DIOR-R[55]数据集简要介绍

在 2018 年，由武汉大学与华中科技大学遥感科研团队精心构建的 DOTA 数据集首次对外发布。该数据集的主要数据源是谷歌地图和国内自主研制卫星的高分辨率遥感图像。随后，该数据集分别于 2019 年及 2021 年进行了扩展与更新，形成了标记为 DOTA-v1.5 与 DOTA-v2.0 的迭代版本。本文采用的是该数据集的初始版本，即 DOTA-v1.0。该版本包含 15 个不同类别，与 2806 张分辨率从 800×800 至 4000×4000 像素不等的航空遥感图像，覆盖了总计 188282 个具有多样尺寸、朝向和形态的标注实例。该数据集被细分为三部分：其中一般的图像用作训练集，约 1/6 的图像分配给验证集，其余 1/3 则被用于测试集。

DIOR 数据集是在 2019 年由西安工业大学开发的，它涵盖了 20 个类别，包含 23663 张分辨率为 800×800 像素的航空遥感图像，以及 192472 个标注实例，其中数据集的主要数据源是谷歌地图。该数据集的特点在于目标种类的丰富性、目标尺度的广泛性以及分布的随机性。具体而言，其数据划分为 5862 张图片的训练集，5863 张图片的验证集，以及 11738 张图片的测试集。另有 DIOR-R 数据集，它基于 DIOR 数据集构建，不同之处在于所有目标均重新标注为旋转框，以适应特定的模型训练需求。在本研究中，我们采用了 DIOR-R 数据集中的训练验证集，总计 11725 张图片，来训练模型。

(2) 对比实验

为了验证本章提出的 MFDFN 在遥感图像目标检测领域的应用效果，本节将通过对比实验来展示其性能。实验选取了包括单阶段检测模型 R3Det[52]、S2ANet、FCOSR、PP-YOLOE-R，以及双阶段模型 RTMDET[53]在内的几种目标检测模型作为比较对象。由于本研究以 PP-YOLOE-R 模型作为基线进行探讨，主要集中在与其他单阶段遥感图像目标检测模型的比较。然而，为了全面评估所提方法的有效性，也将其与双阶段检测模型进行了对比分析。

表 3-1 展示主流遥感图像目标检测模型在 DOTA-v1.0 数据集上的实验成果。与其他单阶段及双阶段遥感图像目标检测方法相比，本章节引入的 MFDFN 显著提升了性能。该模块取代了 PP-YOLOE-R 中的 PANet 后，结果表明在单阶段模型中平均精度（mAP）得到了 0.62 的提升。在表 3-1 中，有一栏为多尺度，这表明模型在训练时，数据集采用的是单尺度数据增强，还是多尺度数据增强。单尺度是指将原图裁剪为 1024×1024 像素的图像片段进行训练和测试，其中每片图像间保留 256 像素的重叠区域。而多尺度则对原始图像按 0.5、1.0 以及 1.5 的比例因子进行尺寸调整，之后将其裁切为尺寸为 1024×1024，重叠区域为 500 像素的

图像片段。

表 3-1 DOTA-v1.0 test 数据集实验结果

模型	骨干网络	多尺度	$mAP_{0.5}$
R3Det	R101+FPN	-	73.79
S2ANet	R101+FPN	-	74.13
FCOSR	R101+FPN	-	77.39
RTMDet	R101	-	78.85
PP-YOLOE-R	CRN-X+PANet	-	78.28
PP-YOLOE-R	CRN-L+MFDFN	-	**78.90**
PP-YOLOE-R	CRN-S+PANet	√	78.42
PP-YOLOE-R	CRN-S+MFDFN	√	**78.91**

如表 3-1 所示，在本研究中，所提出的 MFDFN 在 PP-YOLOE-R 模型下展示了更好的性能，其最终精度超越了其他现有的目标检测方法，进而充分证实了该模块的有效性。为了进行更细致的评估，本实验设计了在单尺度与多尺度数据集上的性能测试。在单尺度评估中，相较于现有的单阶段遥感图像目标检测方法，本文的方法分别在 mAP 上相较基线模型提升了 0.62。而且基线模型所使用的骨干网为 CRN-X，而本章方法所使用的骨干网络为 CRN-L，CRN-X 在参数量上大约为 CRN-L 的两倍，即使如此，本章方法在 mAP 上依旧要高出许多。同时，在多尺度评估中，与基线模型相比在 mAP 上提升了 0.49，亦展现出了较好的性能优势。更为详细的检测结果罗列于表 3-2，其中详细列出了在单尺度 DOTA-v1.0 数据集中 15 类目标检测的精度，对比了六种不同模型的性能表现，其中 RTMDet 为目前精度最高的单阶段遥感图像目标检测方法。

表 3-2 DOTA-v1.0 test 数据集各类目标检测结果

类别 \ 模型	R3Det	S2ANet	FCOSR	PP-YOLOE-R	RTMDet	Ours
PL（飞机）	88.76	89.30	**89.50**	89.49	89.43	89.14
BD（棒球场）	83.09	80.11	**84.42**	79.70	84.21	83.82
BR（桥梁）	50.91	50.97	52.58	55.04	55.2	**55.89**
GTF（田径场）	67.27	73.91	71.81	**75.59**	75.06	74.80
SV（小车）	76.23	78.59	80.49	**82.40**	80.81	81.83
LV（大车）	80.39	77.34	77.72	85.20	84.53	**85.64**
SH（船）	86.72	86.38	88.23	88.35	**88.97**	88.83
TC（网球场）	90.78	**90.91**	90.84	90.76	90.90	90.76
BC（篮球场）	84.68	85.14	84.23	85.69	**87.38**	87.33

ST（存储罐）	83.24	84.84	86.48	**87.70**	87.25	87.55
SBF（棒球场）	61.98	60.45	61.21	63.17	63.09	**64.54**
RA（环行路）	61.35	66.94	67.77	**69.52**	67.87	61.31
HA（港口）	66.91	66.78	76.34	77.09	**78.09**	77.12
SP（游泳池）	70.63	68.55	74.39	75.08	**80.78**	79.63
HC（直升机）	53.94	51.65	74.86	69.38	69.13	**75.27**

为了进一步说明 MFDFN 在遥感图像目标检测中的有效性，选用单阶段网络 FCOSR、S2ANet、PP-YOLOE-R 模型以及纯 Transformer 模型 ViT-B、ViTAE-B 在 DIOR-R 数据集进行对比。由于 DIOR-R 数据集中图片尺度为 800×800，所以不需要对图片进行分割处理。为了适应 DOTA 数据的格式，将对 DIOR-R 中的标注信息进行转换。

表 3-3 是在 DIOR-R 数据集上的实验结果，由于 DIOR-R 数据集的目标种类为 20，以及在数据集图像数量上比分割切图后的 DOTA-v1.0 更少，所以 mAP 均比相同模型在 DOTA-v1.0 上更低。

表 3-3 DIOR-R 数据集实验结果

模型	骨干网络	$mAP_{0.5}$
S2ANet	R101+FPN	68.51
FCOSR	R101+FPN	71.43
ViT-B + RVSA-ORCN[56]	ViT	70.85
ViTAE-B + RVSA-ORCN[56]	ViT	71.05
PP-YOLOE-R	CRN-L+PANet	72.74
PP-YOLOE-R	CRN-L+ MFDFN	**72.88**

相对于同为单阶段的 S2ANet、FCOSR 以及 PP-YOLOE-R，本文改进的模型在检测精度 mAP 上分别提升了 4.37、1.45 以及 0.14。相对于纯 Transformer 模型的 ViT-B 以及 ViTAE-B，本文改进的模型在检测精度 mAP 上分别提升了 2.03 以及 1.83。

(3) 消融实验

为了验证所提出 MFDFN 中的 MDM 以及 COM 在遥感图像目标检测中的优化效果，我们在 DOTAv1.0 数据集的测试集上，以 PP-YOLOE-R 为基线模型，进行了消融实验。

如表 3-4 所示，本文中，在模型参数大小为 s 的 PP-YOLOE-R 模型中，PANet 被替换为 MFDFN。当仅使用 MDM 时，模型性能显著提高，在 mAP 上，相较基线模型提升了 2.93。而模型参数仅增加 1.08M，这验证了所提出模块的有效性。此外，当同时使用 MDM 以及 COM 时，模型的性能进一步提升，mAP 从 76.75

提升到了 77.35，而模型参数几乎没有增加。

表 3-4 消融实验

模型	参数量(M)	PANet	MDM	COM	mAP_{50}
PP-YOLOE-R	8.24	√	-	-	73.82
PP-YOLOE-R	9.32	-	√	-	76.75
PP-YOLOE-R	9.32	-	√	√	77.35

为了进一步验证 MFDFN 在遥感图像目标检测中的作用，以及本研究所提出方法相较于 PANet 的改进之处，本文采用热力图来展示特征融合前后的效果。此外，也利用热力图展现了不同尺度特征的检测头在进行回归检测时更加关注的特征位置。图 3-5 至 3-8 全面展示了不同特征融合策略在特征融合前、后，以及特征传递至检测头之后对目标检测效果的影响。

图 3-5 展示了 PANet 特征融合前后对图像区域的关注程度，其中颜色越深表示对应区域的关注度越高。图 3-5 左侧展示了未进行特征融合时的图像热力图热力图，图左中目标（如各种大小的船只）与背景（海水）之间的区分相对明显。而右侧图展示了特征融合后最顶层特征的热力图，可以看到，虽然目标中心区域获得了更多的关注，但目标与背景之间的界限变得不那么清晰，目标与背景间的特征差异减少。

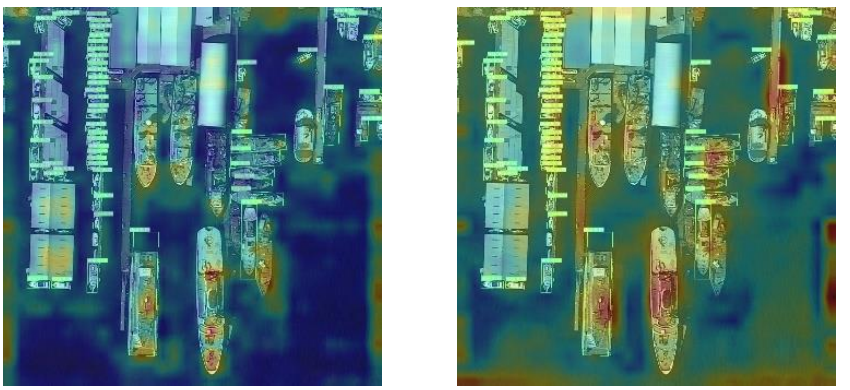

图 3-5 PANet 特征融合前（左）后（右）热力图

如图 3-6 展示了 MFDFN 特征融合前后对图像区域的关注程度，左侧为未进行特征融合时的特征热力图，在该图中，目标与背景之间已经实现了较好的分割，但背景特征并未被明显强调。相对地，右侧图展示了使用 MFDFN 特征融合后的热力图，不仅目标与背景之间的分离效果得到了进一步的提升，目标中心也成为了关注的焦点，同时背景特征也得到了有效的强调。这使得后续检测头在执行目标检测任务时，能够更容易地识别出热度明显高的特征，例如海水。

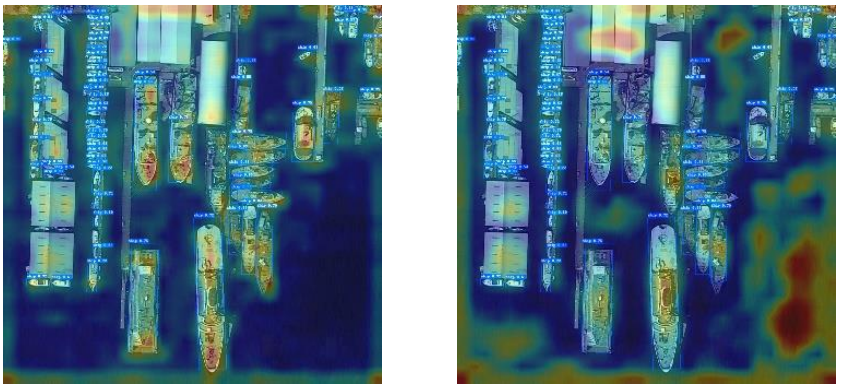

图 3-6 MFDFN 特征融合前（左）后（右）热力图

图 3-7 展示了使用 PANet 进行特征融合后，最上层特征、中层特征以及最下层特征预测结果和相应的类别热图。观察最上层特征的预测效果，可以发现它在大型目标的预测上表现并不理想，且模型似乎未能正确聚焦于关键区域。结合这些观察，我们可以得出结论：在特征融合过程中，较低层次中的不准确特征被传递到了上层，影响了模型的预测性能。

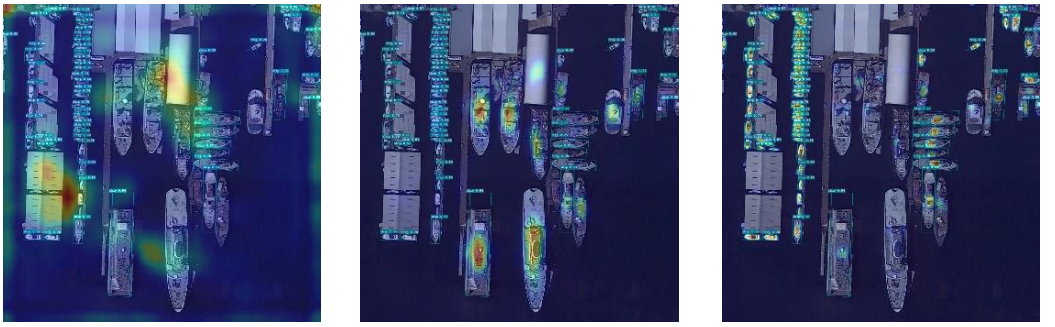

图 3-7 PANet 特征融合后送入检测头的特征热力图

图 3-8 展示了采用 MFDFN 进行特征融合之后，最上层、中层以及最下层特征的预测结果和类别热图。观察结果表明，经过本研究提出的特征融合策略处理后，最上层特征能够准确预测大型目标，且类别热图显示，这些预测未受到下层特征干扰。同时，中层特征的预测结果和类别热图表明其注意力变得更加集中。这意味着，本研究提出的模块不仅有效减少了在特征融合过程中底层特征对上层特征的负面影响，而且能够对底层特征提供有效的正向反馈，从而提升预测性能。

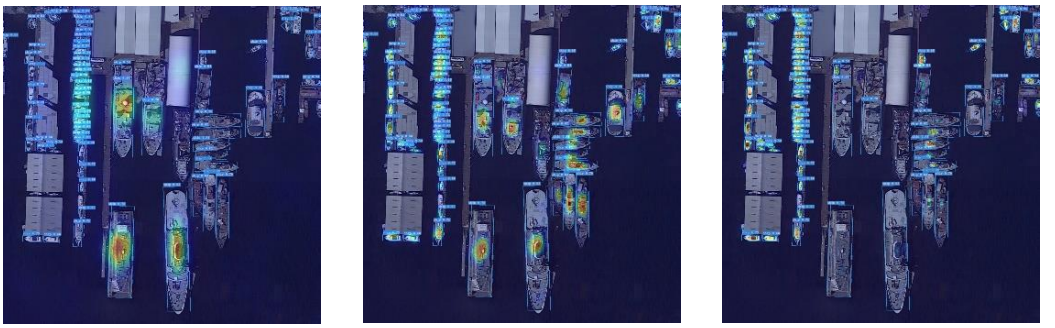

图 3-8 MFDFN 融合后送入检测头的特征热力图

3.3.3 在其它数据集上的泛化性测试实验结果

(1) VisDrone[58]数据集、COCO[59]数据集以及 VOC 数据集简要介绍

VisDrone 数据集包含 288 个视频片段，由 261908 帧视频和 10209 幅静态图像组成，这些数据来自名类无人机摄像头，覆盖范围广泛，包括不同城市(来自中国 14 个不同城市，相隔数千公里)、不同环境(城市和农村)、不同物体(行人、车辆、自行车等)和不同密度(稀疏和拥挤场景)。需要注意的是，数据集是在不同场景、不同天气和光照条件下使用不同型号的无人机平台收集的。超过 260 万个感兴趣的目标框(如行人、汽车、自行车和三轮车)已经手工标注，同时提供了一些重要属性，包括场暑可贝性，对象类别和遮挡情况，以更充分地利用数据。

COCO 是一个用于目标检测、分割和图像标注的大规模数据集。由微软研究院创建，它包含丰富的场景和多样性的物体，使得模型在真实世界的复杂环境中能够更好地泛化。COCO 数据集包含了大约 330000 张图像，其中包含超过 80 个不同类别的物体，每个图像都有详细的物体实例标注，包括边界框和像素级的分割信息。COCO 的标注质量和多样性使其成为深度学习中目标检测和图像分割任务的重要基线数据集。

VOC2007+2012 数据集是一个广泛使用的计算机视觉数据集，主要用于目标检测、图像分割和图像分类等任务。VOC 数据集最初由英国牛津大学的计算机视觉小组创建，并在 PASCAL VOC 挑战赛中使用。VOC 数据集包含各种不同类别的标记图像，每个图像都有与之相关联的目标框和对象类别的标签。数据集中包括了 20 个常见的目标类别，例如人、汽车、猫、狗等。此外，VOC 数据集还提供了用于图像分割任务的像素级标注。VOC 数据集涵盖了多个年度的发布，每个年度的数据集包含训练集、验证集和测试集，训练集用于模型的训练和参数优化，验证集用于模型的调参和性能评估，而测试集则用于最终模型的性能评估和比较

(2) 实验结果分析

如表 3-5 所示，本章通过将 PP-YOLOE+_SOD 中的 PANet 替换成 MFDFN，并在 VisDrone 数据集上进行了性能比较。相较于基线模型，我们提出的方法不仅实现了明显的检测精度提升，同时还拥有更小的模型体积。在骨干网络为 CRN-S 的基线模型中，相较基线模型，本章方法不仅在参数量下降 11.54M，而且在 mAP_{val} 以及 mAP_{test} 上分别提升了 1.4 以及 0.7。这主要是因为基线模型在特征融合前使用了原始的视觉 Transformer 结构，但是该结构不仅无法使得多尺度上的特征无法影响，而且参数量也非常大。同时在 CRN-L 的骨干网络下，相较基线模型，本章所提方法不仅在参数量上下降 27.58M，而且在 mAP_{val} 上提升了 0.2，但在 mAP_{test} 上降低了 0.2 的精度。与 PP-YOLOE 模型相比，在 CRN-S 的骨干网络下，尽管所提方法在参数量上增加了 0.91M，但是在 mAP_{val} 以及 mAP_{test} 上分别提升了 2.3 以及 1.4。同时在 CRN-L 大小的骨干网络下，参数量仅增加 5.68 的情况下，mAP_{val} 以及 mAP_{test} 分别提升了 2.9 以及 1.9。以上结果突显了本文提出的 MFDFN 在加强检测性能的同时，还优化了模型的规模和效率。

表 3-5 在 VisDrone 上进行泛化性测试

模型	骨干网络	参数量(M)	mAP_{val}	mAP_{test}
PP-YOLOE+[60]	CRN-S+PANet	7.66	23.5	19.4
PP-YOLOE+_SOD	CRN-S+PANet	20.11	24.4	20.1
PP-YOLOE+_SOD	CRN-S+MFDFN	8.57	**25.8**	**20.8**
PP-YOLOE	CRN-L+PANet	53.23	29.2	23.5
PP-YOLOE+_SOD	CRN-L+PANet	86.49	31.9	**25.6**
PP-YOLOE+_SOD	CRN-L+MFDFN	58.91	**32.1**	25.4

为了深入评估 MFDFN 的泛化能力，本研究将 YOLOv7[61]中的特征融合模块替换成了我们提出的多尺度降噪融合模块，并在 COCO 以及 VOC 数据集上进行了效果测试。根据表 3-6 的数据，可以观察到，在 COCO 数据集下，在模型参数增加 1.08M 的情况下，本章所提方法在 mAP 上提升了 0.6，而同样的条件下在 VOC 数据集下相较基线模型，mAP 的增长达到了 0.82。同时，在 VOC 数据集下，当骨干网为 CSPDarknet53-L 时，本章方法在参数量增加 3.78M 的情况下，mAP 提升了 0.26。以上不仅证明了 MFDFN 在处理复杂数据集时的有效性，也展示了其在增强模型泛化能力方面的巨大潜力，而且这一切都是在维持参数数量合理增长的前提下实现的。

表 3-6 在 YOLOv7 以及 COCO 上进行泛化性测试

数据集	模型	骨干网络	参数量(M)	mAP_{val}
COCO	YOLOv7	CSPDarknet53-Tiny+PANet	6.23	36.2
	YOLOv7	CSPDarknet53-Tiny+MFDFN	7.31	**36.8**
VOC	YOLOv7	CSPDarknet53-Tiny+PANet	6.23	66.77
	YOLOv7	CSPDarknet53-Tiny+MFDFN	7.31	**67.59**
	YOLOv7	CSPDarknet53-L+PANet	37.62	82.28
	YOLOv7	CSPDarknet53-L+MFDFN	41.38	**82.54**

3.4 本章小结

本章深入探讨了 PANet 在目标检测过程中，由骨干网提取能力不足而引发的噪声传播问题，从而导致检测精度降低的挑战，并提出了一种基于特征降噪的特征融合策略来应对此问题。初始阶段，详细介绍了 Deformable Transformer 原理及其计算流程，为接下来的改进措施奠定了理论基础。随后介绍了本文的 MFDFN，在 MFDFN 中，使用 MDM 在空间层面来对多尺度进行优化，从而达到降低特征中噪声的目的。同时，为了进一步提升 MDM 的降噪能力，本文使用 COM 来优化 MDM 运算前的特征图，实现了在几乎不增加模型参数的情况下，有效配合 MDM 运算，确保了 MDM 在进行降噪前前，每个组特征都已通过权重调节进行优化，以提升检测性能。实验结果验证了本文所提算法显著提高了模型的检测性能，并展示了良好的泛化能力。

第 4 章 增强多尺度特征降噪与自适应融合网络

4.1 引言

在第三章中，针对 PANet 的缺陷提出了 MFDFN，在旋转目标以及水平目标检测中都取得了较好的效果。但是由于 MFDFN 中 MDM 的关键部件是 Deformable Transformer 中的编码器，在适应遥感图像中的各种场景时有所欠缺。同时，MDM 降噪后所使用的特征信息补充方法仅仅是最为简单的跳连相加操作。简单的等权相加并不能充分交融降噪后的特征以及原本的特征信息。这些都对后续的检测头预测造成了一定的阻碍。为了解决上述问题，本章提出一种基于 MFDFN 改进的增强多尺度特征降噪与自适应融合网络(Enhanced Multiscale Feature Denoising and Adaptive Fusion Network, EMFDAN)，该网络主要由两个改进点：

(1) 提出一种条件变量加强的多尺度降噪模块(Conditionally Enhanced Multiscale Denoising Module, CEMDM)。MDM 模块中的核心部件是 Deformable Transformer 编码器，CEMDM 则是在多尺度可变形自注意力编码器计算过程中，增加一个可与多尺度特征共同参与计算的条件变量，该条件变量能够学习到图像中一些特殊的位置信息以及特征信息，使得自注意力的计算更加有效。同时，为了更加适应以及拟合数据集，该辅助参数会随着模型的迭代而进行迭代。

(2) 提出一种权重自适应调整模块(Weight Adaptive Adjust-ment Module, WAAM)。在经过降噪的前后特征相加前，使用所提出的 WAAM 对两组特征中的局部以及整体进行权重调整，使得特征相加时重要的特征得到增强，而不重要的特征得到削弱。同时，根据两组特征它们的信息，自适应的调节相加时的整体权重，使得重要的整体特征在相加时占据更大的比重。

4.2 增强多尺度特征降噪与自适应融合网络的设计与实现

为了进一步提升 MFDFN 对遥感图像中复杂以及特殊场景的识别能力，以及在降噪后进行特征信息补充时给予更多的关注给重要的特征以及降噪后的整体特征。本文在 MFDFN 的基础上，提出增强多尺度特征降噪与自适应融合网络(EMFDAN)。其网络结构如图 3-2 所示。

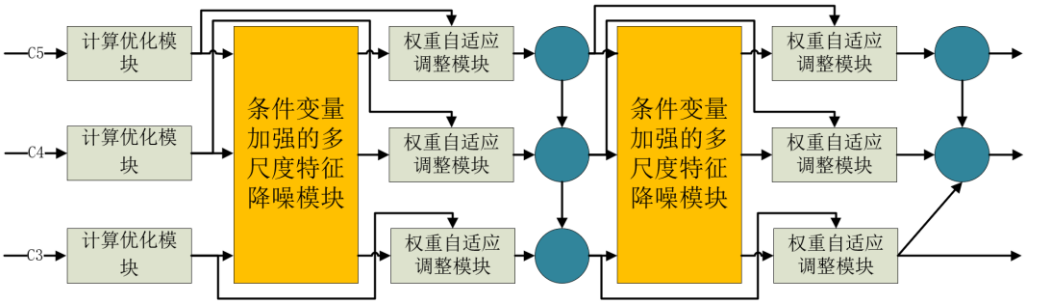

图 3-2 增强多尺度特征降噪与自适应融合网络的结构

相较于 MFDFN，EMFDAN 主要多了两个步骤，一是在进行特征降噪时有条件变量参与计算，二是在计算之后进行特征信息补充时使用了权重自适应调整模块。

4.2.1 条件变量加强的多尺度降噪模块

CEMDM 相对于 MDM 的改进在于使用条件变量辅助 MDM 中的 Deformable Transformer 编码器进行计算，在未将条件变量整合编码器前，编码器在面对遥感图像中情景复杂的场景时，对图像中不同尺度目标以及图像背景的多样性以及复杂性的感知能力可能有所减弱，导致检测模型在特定场景下较弱的表现。而引入条件变量至编码器则较好增强了遥感图像目标检测的性能。通过利用条件变量，编码器获得了额外的信息辅助计算，使得它能够根据条件变量提供的额外的动态信息计算出更加准确的注意力权重，提升了编码器对复杂背景特征的处理能力。改进后的 Deformable Transformer 编码器结构如图 4-1 所示。

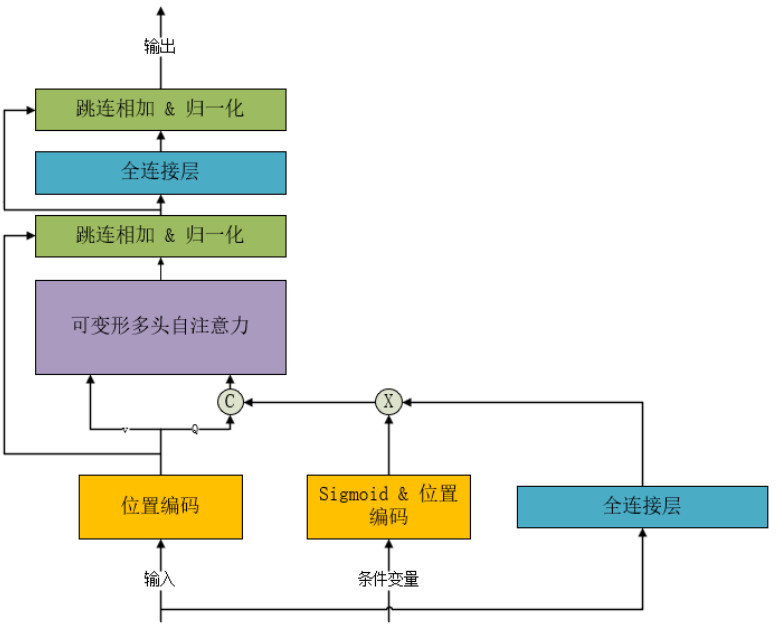

图 4-1 加入条件变量后的 Deformable Transformer 编码器

在改进后的编码器计算过程中，首先是对条件变量的二维特征进行 Sigmoid 函数处理，以便为接下来的操作做准备。其次，我们通过一个全连接层处理需要进行注意力计算的特征，得到一组新的特征。这一组新特征随后与 sigmoid 处理过的条件变量相乘，从而得到调整后的特征。最后一步是将这个调整后的特征与原始的注意力计算特征进行拼接，共同参与到后续的自注意力机制计算中。值得注意的是，尽管条件变量参与到了注意力机制的计算之中，但计算所得的注意力权重仅被应用到原始特征上，而不影响条件变量本身。这样的设计确保了条件变量仅作为一种辅助手段参与计算，而不直接影响注意力权重的分配。该策略不仅提高了模型对遥感图像中目标细节的捕捉能力，而且为遥感图像目标检测领域带来了新的视角和方法。

以上操作的公式化如下所示：

$$C_\sigma = \sigma(C) \tag{4-6}$$

$$X^{'} = FC(X) \tag{4-7}$$

$$P = C_\sigma \cdot X^{'} \tag{4-8}$$

$$F_{concat} = [X; P] \tag{4-9}$$

$$W = DeformableAtt(F_{concat}) \tag{4-10}$$

$$F_{final} = W \cdot X \tag{4-11}$$

公式 4-6 中，C代表条件变量的二维特征，σ代表 Sigmoid 函数，C_σ是经过处理后的条件变量特征，主要用于条件变量的 Sigmoid 处理。公式 4-7 中，X是原始的特征，FC是全连接层，$X^{'}$ 是转换后的特征，在原始特征的基础上进行变换。公式 4-8 中，C_σ与$X^{'}$ 相乘，得到乘积P，这一操作整合了条件变量的信息。公式 4-9 中，将原始特征X与乘积P进行拼接，得到F_{concat}，作为自注意力计算的输入。公式 4-10 中，使用F_{concat}进行自注意力计算，得到自注意力输出W。公式 4-11 中，W是从可变形自注意力计算中获得的注意力权重，但是只应用于原始特征X，以得到最终调整后的特征集F_{final}。

4.2.2 权重自适应调整模块

当经过 CEMDM 降噪后的特征与原始特征进行融合时，传统的特征融合方法（如简单相加或拼接）无法充分融合卷积特征和经过 CEMDM 降噪后的特征。这会导致在后续的任务处理中不能充分利用这些特征的潜在能力。同时，在没有专门设计的权重调整模块的情况下，模型无法动态地学习和调整不同特征的融合

比重。这可能会导致某些重要特征被忽略，而一些不那么重要的特征被过度强调。因此，普通的特征融合方式无法根据特征本身的内容和上下文自适应调整特征的权重再进行相加，使得模型在处理特征时可能过于简单，缺乏必要的灵活性和适应性。

通过引入一种结合了通道注意力机制和门控单元的权重自适应调整模块(WAAM)，模型能够更有效地捕捉和强调对当前任务更为重要的特征信息。这种选择性的特征强调，有助于提升模型的准确性和鲁棒性。在 WAAM 中门控机制的引入，使模型能够根据不同的输入动态调整原始特征和经过 CEMDM 降噪后特征的融合比重。这种自适应的调整方式，更符合实际应用中复杂变化的需求。通过精细控制特征融合的过程，模型可以更好地处理各种情境下的数据，提升泛化能力。特别是在应对未见过的或是特殊的情况时，有助于模型做出更加准确的判断 WAAM 的结构如图 4-2 所示。

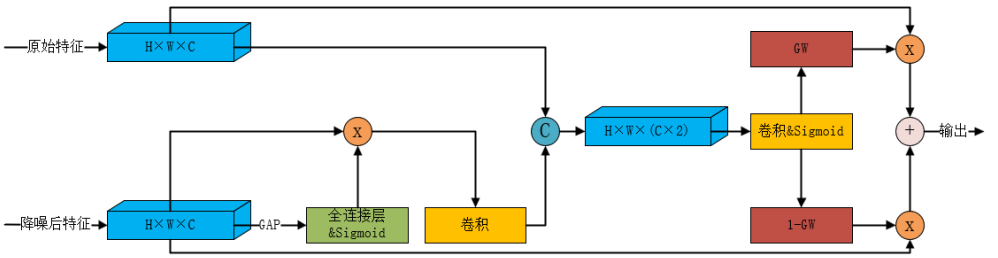

图 4-2 权重自适应调整模块

在 WAAM 的计算过程中，首先是分别对降噪后的特征应用通道注意力，因为这一组特征是由经过 CEMDM 降噪后特征扩展而来的，其中冗余特征较多，所以需要先使用通道注意力强调最为关键的通道信息，进而增强特征表达。接着，将这两组特征拼接起来并经过卷积核大小为 1 的卷积调整通道数量。然后使用门控单元计算出特征中每个位置的权重，于此同时也需要计算出一个反向权重，因为门控单元的权重总和为 1。最终，这两组经过权重调整的特征被相加，生成最终的融合特征。该融合特征综合了局部细节和全局上下文信息，能更好的服务于后续的任务。通过这种方式，本章 WAAM 不仅保留了各输入特征的优势，还通过自适应权重调整增强了特征间的互补性，较好地提升了整体模型的性能。其中门控单元的计算流程如下所示：

$$F_{combined} = \text{Concat}(F_1, F_2) \in R^{(2C)\times H\times W} \tag{4-12}$$

$$F_{intermediate} = \text{ReLU}(W_1 * F_{combined} + b_1) \tag{4-13}$$

$$G = \sigma(W_2 * F_{intermediate} + b_2) \tag{4-14}$$

$$F_{gated} = (F_1 \odot G) + \left(F_2 \odot (1 - G)\right) \tag{4-15}$$

公式 4-12 在通道维度将两个特征进行拼接中，其中F_1和F_2为输入特征，分别是原始特征以及降噪后特征，每个特征图的维度为$R^{C\times H\times W}$，对两个特征进行拼接操作后得到特征$F_{combined}$。之后在公式 4-13 中对$F_{combined}$应用卷积操作减少通道数，并使用 ReLU 激活函数，得到特征$F_{intermediate}$。在公式 4-14 中应用卷积操作并通过 Sigmoid 激活函数生成权重G。最终将得到的权重G和反向权重$1-G$，将这两个权重分别应用到特征F_1和F_2后进行相加操作，得到最终特征F_{gated}。

在这个过程中，通过门控单元的权重调整能力，我们能够有效地综合不同的特征表示，使模型能够根据任务需求适应性地调整特征的贡献度，最终得到的融合特征F_{gated}具备了丰富的信息，预计将在后续任务中发挥关键作用。

4.3 实验设计与结果分析

4.3.1 实验环境与实验设置

为了保证仿真实验的一致性，便于进行实验结果的对比，对于本章节中的实验，将采用与第三章节仿真实验相同的实验环境。与此同时，对于实验参数的设置以及数据集的处理，都将保证采用一致，严格遵从控制变量法。同时，为了更好地进行实验结果的描述，将第三章节构建的模型做为基线模型。

4.3.2 实验结果与分析

(1) 对比实验

为了验证本章提出的 EMFDAN 相较在第三章所提 MFDFN 的有效性，本节将通过对比实验来展示其性能。实验选取了包括单阶段检测模型 R3Det、S2ANet、FCOSR、PP-YOLOE-R，以及 RTMDET 在内的几种主流遥感图像目标检测模型作为比较对象。

表 4-1 展示了主流模型模型在 DOTA-v1.0 数据集上的实验成果。与其他单阶段及双阶段遥感图像目标检测方法相比，本章节所提 EMFDAN 有效的提升了模型的检测性能。

表 4-1 单尺度 DOTA-v1.0 test 数据集实验结果

模型	骨干网络	$mAP_{0.5}$
R3Det	R101+FPN	73.79
S2ANet	R101+FPN	74.13
Oriented RCNN	R101+FPN	76.28
FCOSR	R101+FPN	77.39
PP-YOLOE-R	CRN-X+PANet	78.28
RTMDET	CSP-L	78.85
PP-YOLOE-R	CRN-L+MFDFN	78.90
PP-YOLOE-R	CRN-L+EMFDAN	**79.22**

详细的来说，本章方法在 baseline 模型的基础上进一步改进了多尺度特征降噪与融合网络，使得在遥感图像目标检测上的精度进一步提高。如表 4-1 所示，相对于单阶段网络 PP-YOLOE-R 以及目前精度最高的 RTMDET，本章节所改进的多尺度特征降噪与融合网络在嵌入了 PP-YOLOE-R 中后检测精度分别提升了 0.94 以及 0.37。相较于基线模型，检测精度提升了 0.32。相较于双阶段网络 Oriented RCNN，检测精度提高了 2.94。

表 4-2 是主流遥感图像目标检测方法在 DIOR-R 数据集上的实验结果，本次对比不仅与传统的深度卷积神经网络方法进行了对比，还与 ViT 模型进行了对比。

表 4-2 DIOR-R 数据集实验结果

模型	骨干网络	$mAP_{0.5}$
S2ANet	R101-FPN	68.51
FCOSR	R101-FPN	71.43
PP-YOLOE-R	CRN-L-PANet	72.74
ViT-B + RVSA-ORCN	ViT	70.85
ViTAE-B + RVSA-ORCN	ViT	71.05
PP-YOLOE-R	CRN-L+ MFDFN	72.88
PP-YOLOE-R	CRN-L+ EMFDAN	73.10

详细来说，本章所提方法在检测精度上相对 baseline 方法提升了 0.22。相较专用于遥感图像目标检测的 PP-YOLOE-R，检测精度提升了 0.36。而相较于纯 Transformer 模型 ViT-B+RVSA-ORCN，本章所提方法在精度上提升了 2.05，如此大的精度差距主要是由视觉 Transformer 的特性引起的。在深度学习中，想要提高视觉 Transformer 的检测精度，则模型的参数需要尽可能大且训练的数据量要足够大。

如图 4-3 所示，本文对基线模型以及改进后的模型分别进行了可视化实验，进一步展示了所提方法的有效性。

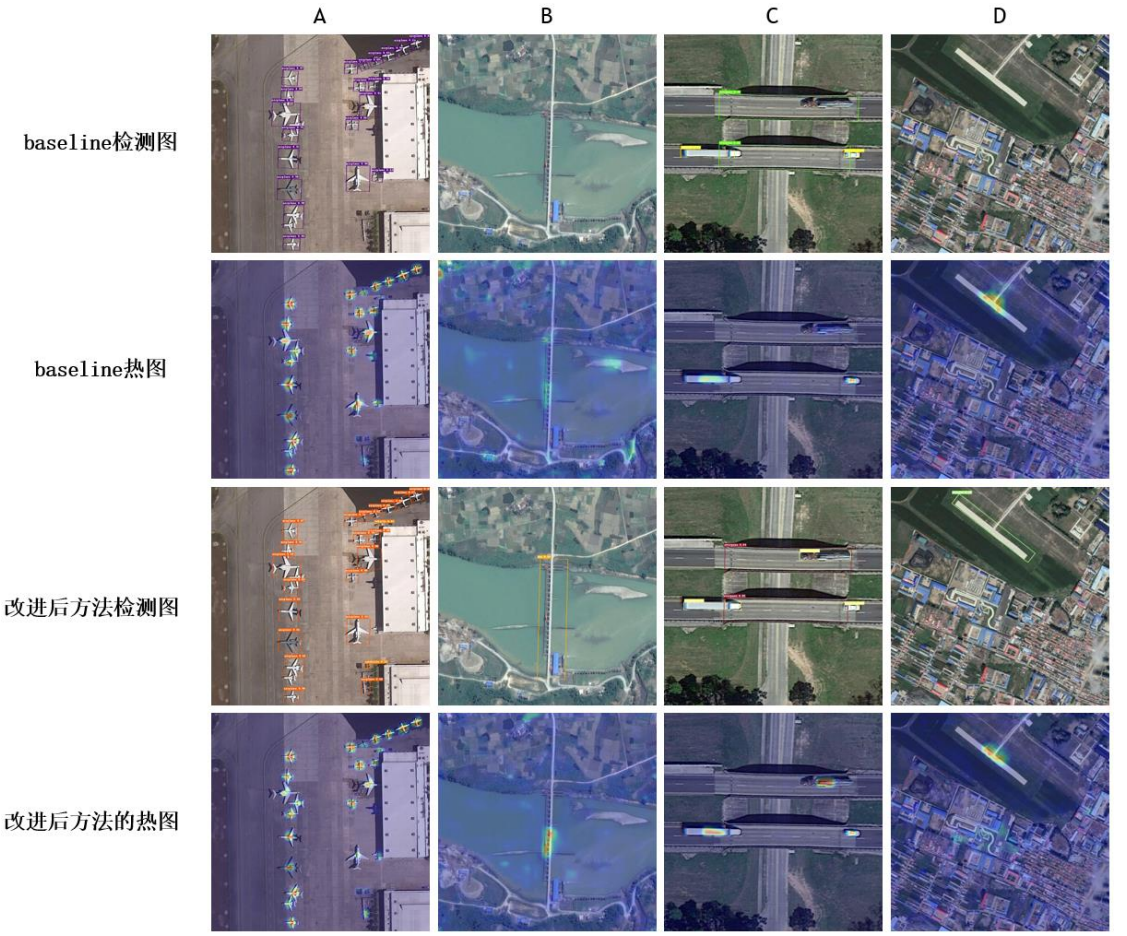

图 4-3 检测热图

详细来说，本文使用 Grad-CAM 算法进行检测头热力图展示，对倒数第二层检测头进行可视化实验。特征层中的参数根据强度回传映射倒原图上，颜色越深的部分代表网络检测的影响越大。在图 4-3 中，一共有四行四列：第一行是 baseline 检测图，这反应了模型在检测出目标的正确位置以及正确类别；第二行是 baseline 热图，这反应了在检测过程中的特征强度，用于分析目标为何会漏检；第三行以及第四行分别是改进后模型的检测图以及热图；而四列分别是四个遥感检测场景。在 A 列中，baseline 并没有将右下角的汽车检测出来，这是因为这个目标不仅小，而且与右下角的建组十分靠近。而改进后的模型将其检测出来了，从热图中也可以看出，改进后的方法对右下角的汽车给予了足够的关注。在 B 列中，baseline 模型并没有将桥检测出来，从热图中可以看出，baseline 模型的关注分散到了桥附近的小岛以及河面，而改进后的模型检测出来了，从热图中也可以看出其正确关注到了桥梁中心。C 列则更加有代表性，桥梁中还有一辆汽车，baseline 模型将桥梁检测出来了却没有检测出在桥梁中的汽车，而改进后的模型却检测出来了，这在特征热图上可以得到较明显的体现。D 列同样是一个非晶复杂的图像，不仅

仅有各种房屋，还有河流，有一座桥梁混杂在其中，对于这种场景，baseline 模型虽然注意到了桥梁，但是因为被其余复杂的背景干扰，并没有检测出来，而从改进后模型的特征热图上来看，改进后的模型在桥梁的位置给予了更加集中的关注度，是的桥梁中心部分的值更大，使得桥梁能够被正确检测出来。

(2) 消融实验

为了验证所提出 EMFDAN 的两项改进措施在遥感图像目标检测中的优化效果，我们在 DOTAv1.0 数据集的测试集上，以第三章所提方法为基线模型，进行了消融实验。

表 4-3 消融实验

模型	MFDFN	CEMDM	WAAM	mAP_{50}
PP-YOLOE-R	√	-	-	70.93
PP-YOLOE-R	-	√	-	71.16
PP-YOLOE-R	-	√	√	71.32

如表 4-3 所示，本文中，在模型参数大小为 s 的基线模型中，多尺度特征降噪与融合网络被进一步改进。当多尺度特征降噪模块在降噪过程中有条件变量参与是，在 mAP 上相较基线模型提升了 0.23。这验证了条件变量在降噪过程中的有效性，条件变量能够学习到特定、特殊的图像信息，从而使得降噪过程更加准确以及高效，最终进一步提升目标检测模型的精度。而在降噪结束后的特征信息补充过程中，使用特征权重自适应调整模块将降噪前后的信息相加，比等权相加法的 mAP 提升了 0.16。这验证了特征权重自适应调整模块的有效性，该模块相较原本的等权相加，使得重要的特征在信息补充过程中占据了更大的权重，从而进一步提高了目标检测模型的精度。

4.4 本章小结

在本研究章节中，首先对 Conditional DETR 模型及其核心组成——条件变量进行了系统性的介绍，从而理解将该技术框架应用于遥感图像目标检测的潜在价值。随后，本章详细介绍了门控单元的原理及其具体的实现方式。基于以上理论背景，本章在第三章 MFDFN 的基础上提出增强多尺度特征降噪与自适应融合网络(EMFDAN)。该网络主要包含两个创新点，一是基于 MFDFN 中的 MDM 改进而来的条件变量加强的多尺度降噪模块(CEMDM)，二是权重自适应调整模块(WAAM)。这两个模块使得本文所设计的特征融合网络在遥感图像目标检测中的实用性更大。最后，本章分别在 DOTA 以及 DIOR-R 数据集上进行了对比实验以及消融实验，为了更直观的展示相对第三章 MFDFN 的改进之处，本章还进行了

可视化实验。以上实验结果验证了本章所提方法在遥感图像目标检测中的有效性。

第 5 章 总结与展望

5.1 全文总结

遥感图像目标检测是一种基于遥感技术，应用于获取和分析从卫星或高空飞行器上携带的传感器捕获的地面图像的过程，以识别、定位和分类地面上的特定目标或对象。遥感技术能够覆盖广阔的地理区域，提供独特的视角和高效的监控手段，因而在许多领域中具有非常重要的应用价值，包括环境监测、土地使用规划、农业、森林管理、城市规划、军事侦察、灾害管理等。由于机器学习以及深度学习的快速发展，学术界中涌现了许许多多关于遥感图像目标检测的科学成果。本文针对遥感图像目标检测中特征融合过程的噪声传播问题进行了深入研究，探讨了如何更加有效的检测以及区分遥感图像目标检测中复杂的目标以及背景。主要研究内容分为三个方面，以下逐一展开说明：

(1) 首先深入探讨了目标检测领域的研究现状，展示了该领域内技术的进步及应用范围的扩大。尤其是在遥感图像目标检测的具体应用环境中，详细阐述了当前遥感图像分析在目标检测方面所面临的挑战及其研究进展，凸显了遥感图像目标检测的特殊性和其在多个领域的重要应用价值。其次系统性地介绍了该领域的基石理论和关键技术。人工神经网络作为深度学习的核心，其基本原理和结构被概述，为后续深入研究卷积神经网络奠定了基础。对卷积神经网络的介绍不仅包括其结构和工作原理，还详细探讨了其在目标检测中的应用。同时，该章节亦对多尺度特征融合技术进行了说明，指出了其在提高检测精度方面的关键性作用。此外，注意力机制的介绍则进一步展示了如何通过模拟人类的注意力集中方式，优化网络的性能。最后，本章还涵盖了遥感图像的特点以及经典的遥感图像目标检测算法，为实验验证和模型训练提供了基础。

(2) PANet 虽然能简单有效地融合不同层级的特征，但其性能极大依赖于骨干网络的特征提取能力。当骨干网络的特征提取不够强大时，PANet 可能加剧了噪声和不具代表性信号的传递与放大，尤其是在目标尺度变化大、背景复杂的遥感图像检测中，这导致了最终检测性能的下降。针对这一点，本文基于 PANet 提出了一种新的多尺度特征降噪与融合网络(MFDFN)，其在 PANet 的基础上增加了多尺度降噪模块(MDM)，这在一定程度上解决了因骨干网引起的在 PANet 中的噪声传播问题。其次，为了进一步提升性能，本文提出了一种计算优化模块(COM)，使得所提出的特征融合模块能够更加有效的进行计算。

(3) 第三章通过提出一种基于特征降噪的多尺度特征融合方法改进了旋转和水平目标检测的问题，但其主要组件 Deformable Transformer 编码器在适应遥感图像的多样性场景上存在不足，并且简单的跳连操作不能充分融合降噪后与原始特征信息，影响了检测性能。为了克服这些限制，本文采取了两项措施：一是在可变形自注意力编码器中引入一个可学习特殊位置和图像信息的条件变量，以提升计算效率；二是提出了一个特征权重自适应调整模块，在特征融合时通过调节局部和整体特征的权重，以优化重要特征的融合效果，从而提升模型的整体表现。

5.2 研究展望

本文主要从改进多尺度特征融合模块以及优化其中融合权重的角度对遥感图像目标检测算法进行了研究，虽然取得了一定程度的进展，但是仍然有许多的问题值得深入探讨，后面准备从以下几个方面进行研究：

(1) 提升算法鲁棒性和通用性

随着遥感图像分辨率的提高和覆盖范围的扩大，目标检测算法面临着越来越多变和复杂的场景。因此，未来的研究需关注算法在不同条件下的表现，包括不同季节、气象条件以及地理环境中的稳定性。此外，对于遥感图像中常见的数据不均衡问题，研究人员需要设计出新型的数据增强技术和样本平衡策略，提升模型在小样本和稀有目标检测场景中的性能。同时，交叉领域的技术，如迁移学习和领域自适应，也将被进一步探索以增强模型的通用性和适应能力。

(2) 结合多源数据和多模态学习

未来遥感图像目标检测的研究将不再局限于单一数据源，多源数据的融合能够从不同维度补充信息，增强目标检测的准确性和鲁棒性。例如，利用光学遥感图像的高分辨率特性进行细致特征提取，同时结合 SAR 图像的天气不敏感特性和夜视能力，可以有效应对复杂环境和不利气象条件。多模态学习策略，如特征融合、决策层融合等，需要进一步研究，以实现不同数据源信息的最优整合。

(3) 引入新型网络架构和算法

为应对遥感图像目标检测中的特殊挑战，开发新型结构和算法成为必要。近年来，Transformer 等注意力机制模型在视觉领域取得显著成果，其在遥感图像目标检测中的应用也将是未来研究的焦点。此外，针对遥感图像的高维特性，设计更高效的多尺度处理、特征融合方法也是未来发展的关键方向。轻量级网络的研究将有助于算法在边缘设备上的部署，为实时遥感数据处理和应用提供可能。

(4) 强化学习和自适应算法的应用

强化学习提供了一种新的视角，使模型能够通过与环境的交互学习最优策略。

在遥感图像目标检测中，强化学习可以用来动态调整检测策略，比如选择最合适的时空数据源、调整特征提取方法等，针对不同任务自动化地优化模型性能。自适应算法使得模型能够根据数据的实际分布调整其结构和参数，提高模型对新场景的适应能力。

(5) 场景理解与知识融合

达到更深层次的场景理解，不仅需在像素级别上进行目标识别，还要实现对场景的整体判断和分析，比如识别场景中的事件、理解场景的动态变化等。这要求模型能够整合和分析大量的上下文信息和先验知识。未来，通过结合遥感数据与其他类型的大数据（如社交媒体、经济统计数据等），采用深度学习、语义分析、知识图谱等技术，不仅可以提升目标检测的精确度，还可以对场景进行全方位的解读和应用，满足更广泛的实际需求。

参考文献

[1] 扶卿华,顾祝军,丰江帆.基于卫星遥感影像的水资源监测研究进展[J].中国水利,2024,(01):28-33.

[2] 马平.无人机遥感技术在森林资源管理中的应用[J].山西林业,2023,(06):18-19.

[3] 张英,郭健斌,哦玛啦等.遥感技术在西藏高原资源环境领域中的应用[J].农业与技,2023,43(21):33-35.

[4] 王永亮.基于改进遥感技术的矿产资源储量勘探研究[J].辽宁科技学院学报,2022,24(06):12-16.

[5] 孟倩.遥感技术在林业资源调查及监测中的应用研究[J].河南农业,2023,(32):42-44.

[6] 邵志东,张芳,彭康等.基于土地覆盖变化与遥感生态指数的奇台绿洲生态环境质量监测[J/OL].环境科学,1-13[2024-02-22].

[7] 杨敏,傅炜舜,聂兴信等.高光谱遥感技术在矿山地质环境调查中的应用[J].现代矿业,2024,40(01):48-52.

[8] 曹如星.基于卫星遥感的大气环境监测技术应用[J].皮革制作与环保科技,2024,5(01):66-68.

[9] 刘佳雷,钱建平,赵鹏伟.高光谱遥感在矿山环境监测中的应用浅议[J].四川环境,2023,42(06):261-266.

[10] 赵升.基于森林资源遥感动态监测研究对生态环境和生态资源的影响[J].中国林业产业,2023,(11):89-91.

[11] 宋恩泽,张颖,邵光成等.基于无人机多光谱遥感的农业园区地物分类研究[J].江苏农业学报,2023,39(09):1862-1871.

[12] 洪小丽,张语桐,王廷超等.无人机遥感技术在农业中应用的发展对策研究[J].东北农业科学,2023,48(05):140-144.

[13] 杨姝.遥感技术在农业旱涝灾害中的应用[J].大众标准化,2023,(18):142-144.

[14] 王爽.遥感与物联网技术在智慧农业灾情精准分析中的应用研究[J].产业与科技论坛,2023,22(14):31-33.

[15] 窦雅娟.基于国产卫星数据的京津冀地区农业大棚遥感监测[J].智慧农业导刊,2023,3(13):1-4.

[16] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[17] Girshick R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.

[18] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[19] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016:779-788.

[20] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, 2016:21-37.

[21] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017:2980-2988.

[22] Tan M, Le Q V. EfficientDet: Scalable and Efficient Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:10781-10790.

[23] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[24] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[25] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[26] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[27] Liu Z, Gao G, Sun L, et al. HRDNet: High-resolution detection network for small objects[C]//2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021: 1-6.

[28] Cai Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6154-6162.

[29] Yang X, Yang J, Yan J, et al. Scrdet: Towards more robust detection for small, cluttered and rotated objects[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 8232-8241.

[30] Yang S, Pei Z, Zhou F, et al. Rotated faster R-CNN for oriented object detection in aerial images[C]//Proceedings of the 2020 3rd International Conference on Robot Systems and Applications. 2020: 35-39.

[31] Xie X, Cheng G, Wang J, et al. Oriented R-CNN for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 3520-3529.

[32] Ma J, Shao W, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE transactions on multimedia, 2018, 20(11): 3111-3122.

[33] Yang X, Yan J. Arbitrary-oriented object detection with circular smooth label[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. Springer International Publishing, 2020: 677-694.

[34] Yang X, Yan J, Ming Q, et al. Rethinking rotated object detection with gaussian wasserstein distance loss[C]//International conference on machine learning. PMLR, 2021: 11830-11841.

[35] Ma J. RRPN++: Guidance towards more accurate scene text detection[J]. arXiv preprint arXiv:2009.13118, 2020.

[36] Jiang Y, Zhu X, Wang X, et al. R2CNN: Rotational region CNN for orientation robust scene text detection[J]. arXiv preprint arXiv:1706.09579, 2017.

[37] Yang X, Yan J, Feng Z, et al. R3det: Refined single-stage detector with feature refinement for rotating object[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(4): 3163-3171.

[38] Liao M, Zou Z, Wan Z, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 919-931.

[39] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.

[40] 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.

[41] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[42] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[43] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[44] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[45] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[46] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[47] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[48] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[49] Han J, Ding J, Li J, et al. Align deep features for oriented object detection[J]. IEEE transactions on geoscience and remote sensing, 2021, 60: 1-11.

[50] Li Z, Hou B, Wu Z, et al. FCOSR: A simple anchor-free rotated detector for aerial object detection[J]. Remote Sensing, 2023, 15(23): 5499.

[51] Xia Z, Pan X, Song S, et al. Vision transformer with deformable attention[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 4794-4803.

[52] Yang X, Yan J, Feng Z, et al. R3det: Refined single-stage detector with feature refinement for rotating object[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(4): 3163-3171.

[53] Lyu C, Zhang W, Huang H, et al. Rtmdet: An empirical study of designing real-time object detectors[J]. arXiv preprint arXiv:2212.07784, 2022.

[54] Xia G S, Bai X, Ding J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3974-3983.

[55] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS journal of photogrammetry and remote sensing, 2020, 159: 296-307.

[56] Wang D, Zhang Q, Xu Y, et al. Advancing plain vision transformer toward remote sensing foundation model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 61: 1-15.

[57] Wang X, Wang G, Dang Q, et al. PP-YOLOE-R: An efficient anchor-free rotated object detector[J]. arXiv preprint arXiv:2211.02386, 2022.

[58] Cao Y, He Z, Wang L, et al. VisDrone-DET2021: The vision meets drone object detection challenge results[C]//Proceedings of the IEEE/CVF International

conference on computer vision. 2021: 2847-2854.

[59] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.

[60] Xu S, Wang X, Lv W, et al. PP-YOLOE: An evolved version of YOLO[J]. arXiv preprint arXiv:2203.16250, 2022.

[61] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[62] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

[63] Meng D, Chen X, Fan Z, et al. Conditional detr for fast training convergence[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 3651-3660.

[64] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

作者攻读硕士期间发表论文及获奖情况

已发表或录用的学术论文：

[1] 作者，导师，合作者 1，等. XXXXXXXXXXXXX[C]. 2022 IEEE 2th International Conference on Algorithms, High Performance Computing and Artificial Intelligence. （已见刊）

[2] 合作者 1，导师，合作者 2，作者，等. XXXXXXXXXXXXX[C]. International Conference on Algorithms, High Performance Computing and Artificial Intelligence. （已见刊）

参与科研项目及获奖情况：

[1] 2020-2021 学年三等学业奖学金

[2] 2021-2022 学年二等学业奖学金

[3] 2022-2023 学年二等学业奖学金

致谢

时光如白驹过隙，转瞬间，我在这所令人敬仰的学府中追求学术与个人发展的日子即将画上圆满的句点。在我硕士学位的学习与研究生涯即将结束之际，我心中充满了无比的感激与不舍。每一步的成长与进步，都离不开无数人的支持与鼓励。因此，在此，我想表达我最诚挚的感谢。

首先，我必须深深感谢我们伟大的祖国。在国家的浩荡发展与强大背景下，我有幸享受优质的教育资源，追求自己的学术梦想。我的每一份收获，都凝聚着祖国的繁荣与进步，我为能生在这样一个伟大的时代而感到无比自豪。接着，我衷心感谢我的母校。这所学校不仅以其卓越的教育品质闻名遐迩，更以其温馨包容的校园文化吸引着像我这样的求学者。学校的每一处景致，每一位教师的辛勤付出，每一次学术讲座的启迪，都深深影响着我，让我不仅在知识上有所收获，在人生态度与价值观念上也有了质的飞跃。特别要感谢我的导师。在我的硕士学习旅程中，导师不仅以其深厚的学术造诣为我指导方向，更以宽广的胸怀为我解惑释惑。导师对我的严格要求是我不断前进的动力，导师的每一次鼓励都是我面对困难不退缩的勇气。我深知，没有导师的悉心指导与无微不至的关怀，就没有我的今天。我的室友，更是我人生旅途中不可或缺的伙伴。我们一同度过了无数个日夜，无论是学术上的探讨，还是生活中的点滴，都充满了对方的身影。在我遇到挫折时给予我安慰，在我取得成果时与我分享喜悦，这些温馨的记忆将伴随我一生。我也要感谢我的同门和师兄们，正是有了你们的支持与帮助，使我的学习和研究之路变得不再孤单。我们彼此鼓励，共同成长，一起面对各种挑战。这种精神的交流和知识的交锋，是一种无法言喻的宝贵经历。此外，我还要感谢所有在我硕士学习过程中曾给予我指导和帮助的教授们。您们的知识宝库、教学经验和人生智慧，对我有着难以估量的影响。每一次交流和沟通，都让我受益匪浅。

最终，我要感谢所有曾经走进我的生命给予我帮助与支持的人。在我追求梦想的路上，你们的每一份鼓励、每一句话语、每一个微笑，都是我前进的动力。我将这些宝贵的记忆深深珍藏在心。

在未来的日子里，我将带着这份深深的感激，继续努力学习，不断地挑战自己，追求卓越。愿我们都能怀揣梦想，砥砺前行，在各自的领域发光发热。最后，再次感谢所有支持我、帮助我、关心我的人，谢谢你们！