

高效遥感图像目标检测:算法与模型规模 双视角的深度网络压缩与加速技术进展

王柳谦^{1,2}, 张菁^{1,3*}, 赵一¹, 谢笑阳⁴, 卓力^{1,3}

(1. 北京工业大学信息科学技术学院, 北京 100124; 2. 郑州大学网络空间安全学院, 河南郑州 450002;
3. 计算智能与智能系统北京市重点实验室, 北京 100124; 4. 北京工业大学计算机学院, 北京 100124)

摘要: 近年来,卷积神经网络(Convolutional Neural Network, CNN)、视觉Transformer以及大规模语言模型(Large Language Model, LLM)等深度学习模型在表示学习与任务泛化方面取得显著进展。然而,随着模型参数规模与计算复杂度迅速增加,其推理延迟与存储开销持续攀升。在遥感领域,高分辨率遥感图像与合成孔径雷达(Synthetic Aperture Radar, SAR)图像具有幅面大、分辨率高、场景复杂、目标多尺度以及小目标占比高等特点。因此,目标检测模型不仅需具备强大的表征能力以应对复杂背景与尺度变化,还必须满足星载、机载及边缘设备等受限平台对实时性与低功耗的严格要求,这使得“精度提升”与“高效部署”之间的矛盾日益凸显。本文首先系统回顾了深度网络压缩与加速技术的发展历程,涵盖从传统手工优化到剪枝、量化与知识蒸馏等经典方法,并进一步探讨了大模型时代下的条件计算与推理优化策略,包括混合专家系统(Mixture of Experts, MoE)、键值(Key Value, KV)缓存优化与混合量化等。其次,通过整合遥感目标检测领域的主流数据集与评价指标体系,本文构建了一个面向“精度-效率-资源”权衡的多维评估基准,用于分析不同压缩方案在遥感场景中的适配性与局限性。与现有综述多沿单一算法主线或仅聚焦轻量网络不同,本文从“算法视角”与“模型规模视角”两个互补维度组织研究脉络:算法视角侧重于在既定模型上通过剪枝、量化、蒸馏与稀疏化来降低计算与存储复杂度;模型规模视角则区分轻量模型与大规模基础模型在结构冗余、敏感模块与内存瓶颈方面的差异,进而分析压缩策略在不同体量模型中的作用机制与优化潜力。再次,基于这一双视角框架,本文对代表性方法的核心思想与实现要点进行对比,总结影响压缩效果的关键因素,并提炼可复用的工程化实践建议,以期对模型选型与部署配置提供参考。最后,结合遥感智能解译的应用需求与绿色低碳发展趋势,本文展望自适应压缩、多模态协同压缩、参数高效微调与软硬件协同优化等方向,为遥感目标检测模型在边缘端与工程系统中的高效落地提供系统化的理论支持与实践指导。

关键词: 遥感图像;目标检测;网络压缩;模型规模;加速

基金项目: 北京市自然科学基金(No.L247025)

中图分类号: TP751.1

文献标识码: A

文章编号: 0372-2112(2026)03-1348-16

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20251009

Efficient Object Detection in Remote Sensing Images: Advances in Deep Network Compression and Acceleration from Algorithmic and Model-Scale Perspectives

WANG Liuqian^{1,2}, ZHANG Jing^{1,3*}, ZHAO Yi¹, XIE Xiaoyang⁴, ZHUO Li^{1,3}

(1. School of Information Science and Technology, Beijing University of Technology, Beijing 100124, China;

2. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, Henan 450002, China;

3. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China;

4. College of Computer Science, Beijing University of Technology, Beijing 100124, China)

Abstract: In recent years, deep learning models such as convolutional neural networks (CNNs), vision transformers, and large language models (LLMs) have made remarkable breakthroughs in representation learning and task generalization. However, this progress has been accompanied by rapid increase in parameter scale and computational complexity, leading to rising inference latency and storage overhead. In remote sensing applications, high-resolution remote sensing images and synthetic aperture radar (SAR) images are characterized by large coverage, high resolution, complex scenes, and multi-scale variations with a high proportion of small objects. Object detection models must therefore possess strong representational

capability to handle complex backgrounds and scale variations, while also meeting stringent real-time and low-power constraints on resource-limited platforms such as satellite-borne, airborne, and edge devices. As a result, the tension between “improving accuracy” and “efficient deployment” has become increasingly pronounced. This study systematically reviews the development of deep network compression and acceleration techniques, covering traditional manual optimizations as well as classical methods such as pruning, quantization, and knowledge distillation. It further discusses conditional computation and inference optimization in the era of large models, including mixture-of-experts (MoE) architectures, key-value (KV) cache optimization, and hybrid quantization. By integrating mainstream datasets and evaluation metrics in the field of remote sensing object detection, this paper establishes a multi-dimensional evaluation benchmark oriented toward the “accuracy-efficiency-resource” trade-offs, aimed at characterizing the applicability and limitations of different compression approaches in remote sensing scenarios. Unlike existing surveys that often follow a single algorithmic thread or focus solely on lightweight networks, this review organizes the research landscape from two complementary perspectives: the algorithmic perspective and the model-scale perspective. The former emphasizes reducing computational and storage complexity through pruning, quantization, distillation, and sparsification on a given model. The latter distinguishes between lightweight models and large-scale foundation in terms of structural redundancy, sensitive modules, and memory bottlenecks, analyzing how compression strategies operate and where further optimization is possible across models of different sizes. Based on this dual-perspective framework, the paper compares and analyzes the core ideas and implementation details of representative methods, summarizes key factors influencing compression performance, and extracts reusable engineering practices to inform model selection and deployment configurations. Finally, considering the application demands of remote sensing intelligent interpretation and the trend toward green and low-carbon computing, this paper looks ahead to promising directions such as adaptive compression, multimodal collaborative-compression, parameter-efficient fine-tuning, and hardware-software co-optimization. The aim is to provide systematic reference and theoretical support for the efficient deployment of remote sensing object detection models on edge devices and in engineering systems.

Keywords: remote sensing images; object detection; network compression; model-scale; acceleration

Foundation Item(s): Beijing Natural Science Foundation (No.L247025)

0 引言

遥感技术作为地球观测、环境监测和灾害评估等领域的重要手段,在高分辨率卫星图像、无人机(Unmanned Aerial Vehicle, UAV)航拍以及多传感器数据融合等应用中展现出巨大潜力^[1]。得益于高分辨率卫星成像系统、UAV多光谱航拍平台以及多源传感器数据融合技术的突破,遥感数据获取能力已实现从宏观地表覆盖到厘米级地物特征的跨尺度感知。在此背景下,以卷积神经网络(Convolutional Neural Network, CNN)和视觉Transformer为核心的深度学习架构,通过端到端特征学习机制在遥感图像目标检测、语义分割及变化检测等任务中取得突破性进展,尤其是大规模语言模型(Large Language Model, LLM)展现出的跨模态推理能力,进一步拓展了智能解译的技术边界。

为直观展示图像目标检测中模型复杂度的增长趋势,本文统计了近年来典型检测模型的参数规模与推理成本随时间的变化情况,其增长曲线如图1所示。2017—2024年,代表性模型的参数量实现了跨越式增长,从Faster R-CNN^[2]的约42 M(百万)激增至InternVL-Chat^[3]的7 000 M(7B,十亿)级别,增幅超过150倍。与此同时,模型的计算量亦从几十亿次

浮点运算每秒(Giga Floating point Operations Per second, GFLOPs)攀升至数百GFLOPs甚至更高,呈现出指数级增长趋势。

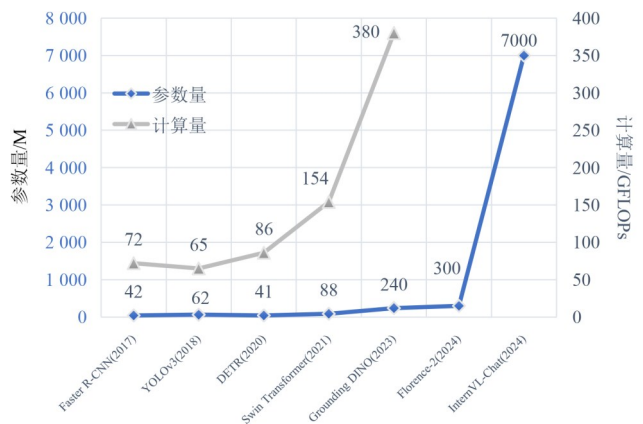


图1 检测模型参数规模增长曲线(2017—2024年)

Figure 1 Growth trend of detection model parameters (2017—2024)

模型规模的急剧扩张在提升检测精度与泛化能力的同时,也带来了前所未有的算力、存储与能耗压力。这一问题在遥感图像处理领域尤为突出。面对遥感智能处理系统在推理效率与部署成本方面的刚性约束,发展适配遥感图像特性的网络压缩与加速技

术,已成为突破当前工程化瓶颈的关键路径。目前,网络压缩与加速技术已从早期的手工优化逐步发展为以深度学习模型为核心的系统性框架。尽管已有工作对相关技术进行了综述,但大多沿单一算法主线展开,或侧重于轻量模型,对压缩技术与模型规模、遥感任务特性之间的系统关联仍缺乏整合视角。为此,本文尝试从两个互补维度重新组织相关研究:其一,从“算法视角”出发,系统梳理剪枝、量化、知识蒸馏与稀疏性优化等技术在遥感目标检测中的原理演进与任务适配方式;其二,从“模型规模视角”出发,区分轻量化模型与大模型,“大模型”主要指参数规模达到数亿至数十亿级别、通常基于Transformer或多模态架构进行预训练的通用或基础模型,进而讨论不同体量模型在设计、压缩与部署中的差异化约束及协同策略。基于此双视角框架,本文构建了一条兼顾算法路径与模型体量特征的综述脉络,为遥感目标检测网络压缩与加速研究提供更结构化、可扩展的参考体系。

1 背景

本节详细介绍了遥感图像目标检测背景与挑战、遥感网络压缩与加速技术进展以及常见的遥感图像目标检测数据集与评价指标。

1.1 遥感图像目标检测中的深度学习应用与挑战

近年来,随着深度学习在遥感图像分析中的广泛应用,遥感图像目标检测技术取得了显著进展,但同时也面临诸多挑战。这些挑战主要体现在算法设计和模型规模两个方面,为后续方法从“算法视角”和“模型规模视角”展开研究提供了理论依据。

1.1.1 算法挑战

遥感图像目标检测面临多个算法挑战,主要包括目标的多尺度性、复杂场景适应性、多模态融合难度及实时性要求等方面。首先,遥感图像覆盖范围广、空间分辨率高,目标尺度差异显著,尤其是小目标数量多、特征弱,易被背景淹没。如何提升检测模型的尺度不变性,准确捕捉不同尺度下的目标,成为关键问题^[4]。因此,近年来研究者广泛采用多尺度特征融合结构(如特征金字塔、跨层连接)以增强模型的尺度表达能力,从而提升对小目标和远距离目标的识别性能。其次,遥感图像采集过程常受光照、云层、季节变化等环境因素干扰,背景复杂、噪声较多,增加了模型提取有效特征的难度。为了提升鲁棒性,研究引入了数据增强、对比学习等方法,增强模型对不同场景的泛化能力。再次,遥感任务常涉及光学、合成孔径雷达(Synthetic Aperture Radar, SAR)、激光雷达(Light Detection And Ranging, LiDAR)等多源数据,这

些数据具有异构性和模态间差异,融合难度大。为实现互补信息的有效整合,亟须设计高效的跨模态特征对齐与融合机制。最后,遥感应用对实时性有较高要求,尤其是在卫星和 UAV 平台上,受限于计算资源和功耗,复杂深度网络难以部署。因此,在保持模型精度的同时降低计算开销,是提升遥感图像目标检测实用性的关键。

1.1.2 模型规模增长趋势

随着深度学习技术的迅速发展,遥感图像目标检测领域中常用的模型,如 CNN、Transformer 以及最近兴起的 LLM,其参数量呈现爆炸式增长。在高分辨率遥感图像的目标检测、语义分割和变化检测任务中,为了捕捉更多细节信息和丰富上下文,模型规模不断扩大。模型规模的增加虽然有助于提升检测精度,但也带来了诸多的问题。例如,计算复杂度与存储需求的增加导致超大规模模型在训练和推理时需巨额计算资源和内存支持,这在资源受限的边缘设备和实时应用场景下尤为突出。同时,能耗问题也亟须解决,高算力需求必然伴随较高的能耗,难以满足 UAV、卫星等平台对低功耗的要求。

综上所述,遥感图像目标检测在深度学习应用中既面临多尺度、复杂场景、多模态融合等算法处理挑战,又受到模型规模不断增长带来的计算和能耗压力。这些矛盾映射到方法论层面,催生了本研究的系统性分析框架:从算法视角、模型规模视角以及最新研究热点展开的讨论不仅提供了重要依据,还强调了在设计高效遥感图像目标检测系统时,压缩与加速技术的重要性。

1.2 遥感网络压缩与加速技术进展

在过去二十余年中,随着深度学习的快速发展,网络压缩与加速技术在遥感中的需求不断提升。由于遥感数据具有高分辨率、大尺寸、跨模态与实时处理等特性,通用压缩方法在适配性方面面临挑战。因此,本节结合遥感应用特点,从时间演进角度概述网络压缩与加速技术的三个阶段(图2):(1)以手工/工程化轻量化为主的早期工程化阶段;(2)将剪枝、量化、蒸馏等深度网络压缩方法引入并适配遥感任务的深度学习阶段;(3)随着大规模预训练与多模态基础模型兴起,向条件计算、稀疏激活与软硬件协同方向发展的大模型适配阶段。

早期工程化阶段(2000s—2010s)。在这一阶段,遥感领域的网络压缩以工程可用为主,多采用降维与特征工程手段。典型方法包括基于主成分分析(Principal Component Analysis, PCA)和非负矩阵分解(Negative Matrix Factorization, NMF)的特征降维与通信负载压缩。基于此,Zubko 等人^[5]率先用 PCA 剖析卫

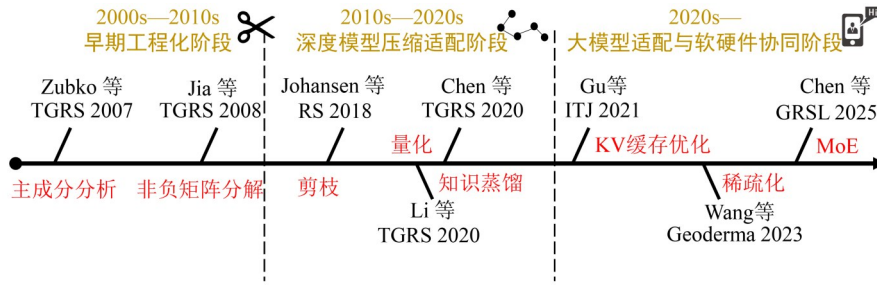


图2 网络压缩与加速技术发展历程

Figure 2 The evolution of network compression and acceleration

星海洋气溶胶观测的信息维度,为遥感载荷“少通道-多角度-加偏振”的压缩设计奠定理论基石。此外, Jia 等人^[6]将空-谱平滑与丰度稀疏先验引入 NMF 框架,为高光谱端元提取提供更鲁棒的自监督解混方案,在显著降低存储与计算需求的同时,为后续大规模任务奠定基础。

深度网络压缩适配阶段(2010s—2020s)。随着 CNN 与 Transformer 在遥感目标检测中的普及,研究重心转向在保证精度的前提下实现深度网络轻量化。代表性方向包括:①结构化剪枝,从全局权重剪枝演进到任务感知与通道敏感剪枝, Johansen 等人^[7]提出吸收式剪枝与比率调整方法,有效保留浅层与中层多尺度信息;②结合任务的量化策略,结合分块推理与拼接误差,采用通道/层级感知量化与混合精度推理, Li 等人^[8]进一步提出量化深度学习哈希,实现端到端的压缩与检索;③知识蒸馏的设计,将教师模型判别能力迁移至轻量学生模型, Chen 等人^[9]基于特征金字塔与蒸馏-多任务协同,实现无旧数据条件下的增量目标检测。

大模型适配与软硬件协同阶段(2020s—)。视觉与多模态大模型显著提升了遥感场景表征能力,同时加剧部署难度。现有研究主要从以下三方面展开:①键值(Key Value, KV)缓存优化,通过重构 KV 缓存组织形式降低大模型推理过程中的内存与带宽开销, Gu 等人^[10]在星-空-地一体化网络中提出地球同步轨道编码缓存策略,提高回传效率与能量利用;②稀疏化算法为超大模型在资源受限平台上的应用提供可行路径, Wang 等人^[11]以轻量梯度提升机为核心,实现高精度且可解释的大尺度土壤盐分反演;③混合专家策略,通过按需激活异构专家降低平均计算成本, Chen 等人^[12]构建稀疏混合专家架构,在遥感超分中按像素动态分派专家,实现细节与全局结构的协同重建。

总体而言,遥感网络压缩技术经历了从“早期工程化”到“深度网络适配”,再到“大模型适配与软硬件协同”的三阶段演进。这些进展为高分辨率遥感图

像的高效解译提供了关键的技术支撑。

1.3 遥感图像目标检测数据集与评价指标

表 1 列出了遥感图像目标检测领域常用的数据集。这些数据集按照数据源可以分为三类:光学红绿蓝三通道(Red-Green-Blue, RGB)数据集、SAR 数据集和红外数据集。这些数据集广泛应用于遥感图像目标检测任务。

光学 RGB 数据集主要来源于高分辨率卫星图像和航空摄影,该数据集具有丰富的色彩信息和较高的空间分辨率,适合多种应用场景,包括车辆和密集建筑区域识别数据集(Vehicle and Densely built Area Identification dataset, VEDAI)^[13]、高分辨率船舶数据集(High Resolution Ship Dataset, HRSC2016)^[14]、遥感图像目标检测数据集(Remote Sensing Object Detection dataset, RSOD)^[15]、航空图像中目标检测数据集(Dataset for Object deTectioN in Aerial images, DOTA)^[16]、扩展商业和军事对象数据集(extended commercial and military object dataset, xView)^[17]、视觉与无人机数据集(Vision meets Drones dataset, VisDrone)^[18]、细粒度飞机图像识别数据集(Fine-grained Aircraft Images Recognition Dataset, FAIR1M)^[19]。

SAR 数据集主要源于 SAR 图像,具有全天候成像能力,适合复杂环境下的目标检测。该数据集通常具有较高的分辨率和特定的极化方式,如 SAR 船舶探测数据集(Synthetic aperture radar Ship Detection Dataset, SSDD)^[20]、高分辨率 SAR 船舶探测数据集(High Resolution SAR ship detection Dataset, HRSID)^[21]、旋转 SAR 船舶检测数据集(Rotated SAR ship Detection Dataset, RSDD)^[22]、旋转 SAR 基准数据集(Rotated SAR benchmark dataset, RSAR)^[23]。

红外数据集源于红外成像系统,具有较高的对比度和对热源的敏感性,适合夜间成像下的小目标检测,如鸟瞰智能数据集(BIRD'S AI dataset, BIRDSAI)^[24]、单帧红外小目标数据集(Nanjing University of Aeronautics and Astronautics-Single-frame Infrared Small Target dataset, NUAA-SIRST)^[25]。

目标检测常用的评价指标包括平均精确度均值 (mean Average Precision, mAP)、精确率 (Precision)、召回率 (Recall)、 F_1 分数 (F_1 measure)、检测框重叠率 (Intersection over Union, IoU)、漏检率 (miss rate) 以及

平均检测时间 (average detection time)。这些评价指标共同评估目标检测模型的精度、召回能力、定位效果以及计算效率, 全面反映模型在不同场景下的性能。

表 1 遥感图像目标检测常用数据集

Table 1 Datasets for object detection in remote sensing images

数据集	年份	样本数	尺寸	类别	数据源
VEDAI ^[13]	2015	1 250	1 024 × 1 024	9	光学 RGB
HRSC2016 ^[14]	2017	1 061	1 500 × 900	26	光学 RGB
RSOD ^[15]	2017	976	512 × 512~1 083 × 923	4	光学 RGB
DOTA v1.0/1.5 ^[16]	2018	2 860	800 × 800~4 000 × 4 000	15/16	光学 RGB
xView ^[17]	2018	1 413	3 000 × 3 000	60	光学 RGB
VisDrone ^[18]	2019	10 209	2 000 × 1 500	10	光学 RGB
FAIR1M ^[19]	2022	39 891	10 000 × 10 000	37	光学 RGB
SSDD ^[20]	2017	1 160	500 × 500	2	SAR
HRSID ^[21]	2020	5 604	800 × 800	2	SAR
RSDD ^[22]	2021	2 884	1 024 × 1 024	6	SAR
RSAR ^[23]	2025	95 842	—	6	SAR
BIRDSAI ^[24]	2020	162 000	640 × 640	10	红外
NUAA-SIRST ^[25]	2021	427	300 × 300	1	红外

2 方法

本节从“算法视角”与“模型规模视角”两个维度对深度网络的压缩与加速技术进行探讨, 如图 3 所示。“算法视角”指针对已有模型, 通过剪枝、量化、知识蒸馏、稀疏性优化等技术降低其计算与存储复杂度的方法体系, 属于实现模型轻量化的具体技术路径。“模型规模视角”则关注模型本身的结构与参数量级, 分析不同规模模型 (如轻量模型与超大模型) 在设计、压缩与部署中的独特挑战及应对策略。

需要说明的是, 这两个视角并非严格正交, 而是从“技术路径”与“对象属性”两个层面, 对同一问题进行拆解与整合。算法的有效性通常依赖于模型的内在属性, 同一轻量化方法在不同规模模型中的作用机制与优化空间存在显著差异。例如, 同一剪枝策略在参数冗余的大模型中可能效果显著, 而在紧凑轻量模型中则容易导致精度损失。因此, 本节首先从算法视角出发, 系统梳理当前主流的压缩与加速方法; 然后转入模型规模视角, 重点探讨这些技术在不同体量模型中的适用性及演进趋势, 为读者构建一个层次清晰、结构完整且具可操作性的认知框架。

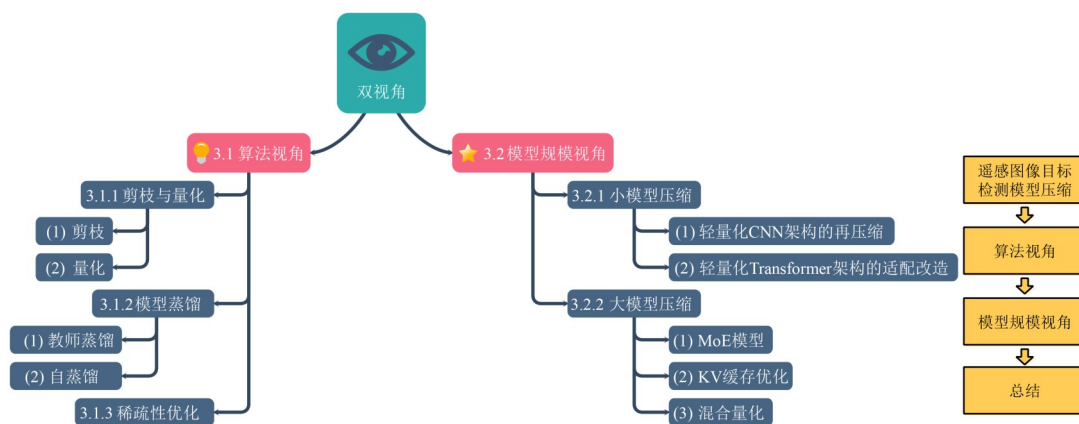


图 3 网络压缩的多视角思维导图

Figure 3 Multi-view mind map of network compression

2.1 算法视角

在遥感图像目标检测中,算法层面的优化是平衡模型性能与部署效率的核心手段。剪枝、量化、知识蒸馏及稀疏性优化等技术通过精简网络结构、降低计算精度或迁移知识分布,显著减少了模型的参数量与推理开销。

2.1.1 剪枝与量化

现代遥感图像目标检测模型往往存在过度参数化的问题,这限制了其在资源受限平台上的部署。剪枝(Pruning)和量化(Quantization)是减小模型规模、提升推理效率的两种主要算法策略。本节主要介绍基于剪枝与量化算法的压缩技术概念和结构以及在遥感任务中的典型应用。

(1)剪枝

剪枝是一类重要的网络压缩技术,广泛用于遥感目标检测中,以在有限计算与存储资源下实现对高分辨率图像的实时处理。其基本思想是移除冗余参数,降低计算复杂度,提高推理效率。在遥感场景中,由于小目标密集、尺度差异大以及多模态图像复杂,剪枝需结合任务特性进行适配。例如,针对星载边缘计算,Lu 等人^[26]提出面向稀疏 CNN 的加速框架,利用稀疏权重索引与块状稀疏调度策略提升存储访问效率,使得轻量模型能够在低功耗硬件上稳定运行。然而,遥感图像中存在大量小目标和旋转目标,若剪枝策略不加区分,极易造成小尺度特征损失。为此,Hu 等人^[27]提出了一种基于任务感知的多目标优化剪枝框架,在进行通道裁剪时引入小目标检测约束,使模型在压缩率与检测精度之间取得平衡。

(2)量化

量化技术通过将高精度浮点参数转换为低比特

整数(如 INT8),以降低计算复杂度和存储需求,从而提升推理效率。与通用计算机视觉不同,遥感图像通常具有超大尺寸和多模态特征,若直接量化,则可能导致跨块拼接误差累积、小目标检测精度下降等问题。因此,遥感场景中的量化需在速度与精度之间进行更精细的权衡。典型流程包括权重量化、激活量化、量化误差校正与推理优化。在遥感图像分割与检测任务中,量化方法显著提升了模型在边缘平台的适配性。例如,Xie 等人^[28]提出动态偏移二值量化与 Swin-ViT 结合的高效框架,利用多光谱特征融合与动态二值量化机制,在不同卫星平台数据上实现高效分割与检测,兼顾精度与计算效率。该方法特别适合 UAV 与星载设备的实时任务部署。

总体而言,剪枝与量化是遥感图像目标检测中较为成熟的两类压缩技术:前者通过移除冗余通道与卷积核来简化结构,适用于参数冗余较高且注重结构可解释性的检测网络;后者则通过降低数值表示的比特宽度来提升计算效率,尤其适合资源受限平台。针对遥感图像中多尺度小目标密集、影像幅面大等特点,剪枝需侧重保留对小目标敏感的通道,量化则需抑制分块推理引起的拼接与边缘误差。因此,在工程中常将剪枝、量化、知识蒸馏和稀疏化等方法结合使用,以在精度与效率之间取得更稳健的平衡。

2.1.2 模型蒸馏

知识蒸馏(Knowledge Distillation, KD)通过将复杂教师模型的知识迁移至轻量学生模型,在遥感目标检测中展现出独特优势^[29]。与剪枝和量化直接修改结构不同,蒸馏通过知识迁移有效保留了小目标、旋转目标和多模态任务中的判别能力,非常适合资源受限的遥感平台。其主要形式包括教师蒸馏和自蒸馏,如图 4 所示。

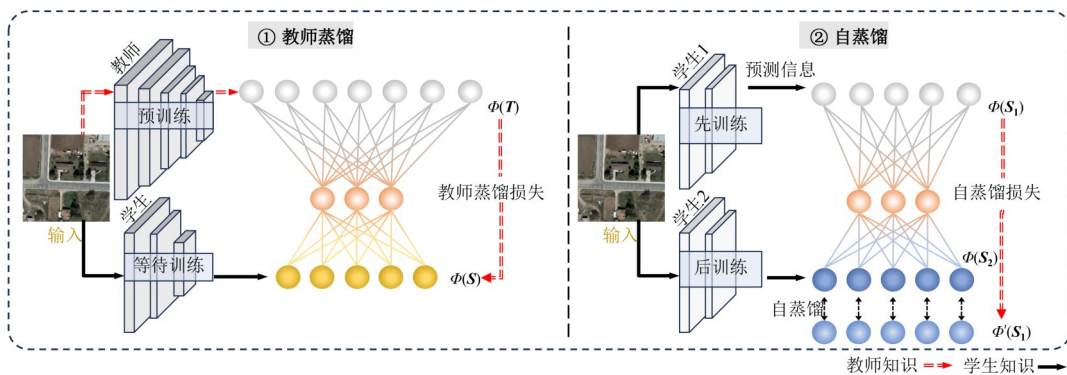


图 4 知识蒸馏示意图

Figure 4 Schematic diagram of knowledge distillation

(1)教师蒸馏

在遥感图像目标检测中,教师蒸馏常用于将高精

度的两阶段检测器或多模态大模型压缩为轻量级单阶段学生模型。Yang 等人^[30]针对遥感背景杂糅、目

标尺度悬殊导致的蒸馏噪声问题,提出自适应强化监督框架,以面积加权的模块聚焦小目标核心特征,在多个遥感目标检测数据集上使轻量模型反超重量级对手。Yang 等人^[31]提出双向辅助蒸馏框架(Two-Way Assistant distillation, TWA)。该框架包含压缩辅助模块与多尺度自适应辅助模块,相较于现有方法显著提升了单阶段与两阶段检测器的性能表现。

在之前的研究中^[32-33],同样利用了知识蒸馏方法在直播视频不良主播识别领域,针对深度网络参数量大、推理效率低的问题,提出基于多模态知识蒸馏(Multimodal Knowledge Distillation, MKD)的模型压缩方法。该方法构建视觉-语音双模态教师模型与轻量化学生模型,在保证识别准确率的前提下,有效降低了计算复杂度并提升了实时检测效率。

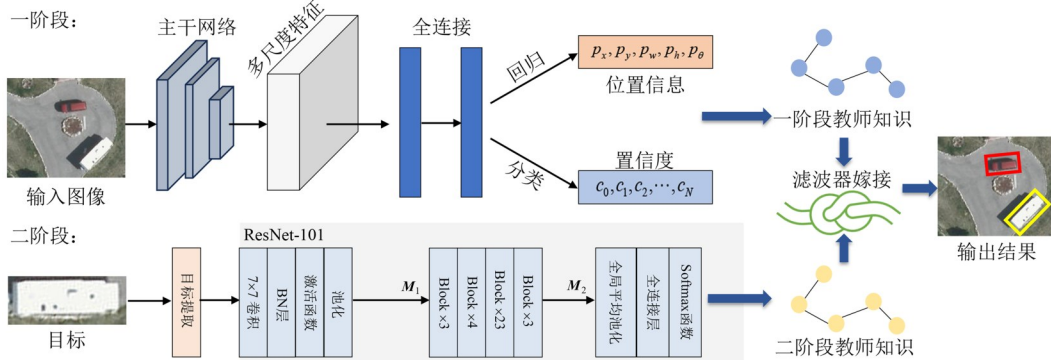


图5 两阶段遥感图像目标检测知识蒸馏流程

Figure 5 A two-stage knowledge distillation framework for remote sensing object detection

该方法通过两阶段实现细粒度目标识别:第一阶段利用主干网络提取多尺度特征,精确定位并分类地物对象,同时作为教师网络提供一阶段教师知识;第二阶段设计了一种由粗到细的目标识别网络(Coarse-to-Fine Object Recognition Network, CF-ORNet),主要通过 ResNet-101 的结构提取特征并进行类别校正进一步提升识别精度,同时获得二阶段的教师知识。在此基础上,本文提出了一种滤波器嫁接方法,通过对知识蒸馏后的学生网络进行滤波器嫁接实现高效细粒度目标识别,其关键思路如下:

(a) 鉴于遥感图像目标检测中存在各种规模、密集排列且对象高度相似的情况,设计了一个两阶段细粒度目标检测网络,从而挖掘检测器结果的精细特征和类别修正。

(b) 由于细粒度目标检测网络结构复杂且计算成本高,通过通道级知识蒸馏、加权软标签知识蒸馏和滤波嫁接,提出了一种从知识蒸馏到滤波器嫁接的轻量级模型,以在遥感图像中实现高效的细粒度目标检测。

(2) 自蒸馏

自蒸馏无需外部教师模型,直接利用模型自身预测进行优化,具有较低的额外开销^[34]。在遥感图像哈希检索任务中,Hu 等人^[35]将条件扩散去噪引入自蒸馏框架,提出(Diffusion Self-Distillation Network, DS-DNet),通过逐步去噪净化浅层特征,使其在卫星与 UAV 端几乎零增参的条件下仍能获得深层监督的细节。该工作在遥感场景分类中实现“噪声抑制-模型压缩”一体化,为星上轻量级判别模型提供了新的自蒸馏范式。

(3) 样例分析

在之前的工作中^[36-37],针对高分辨率遥感图像的目标检测和细粒度识别的效率问题,提出了一种从知识蒸馏到滤波器嫁接的高效方法,其网络结构如图5所示。

(c) 所提出的高效模型能够在 HRSC2016 和 VEDAI 等基准遥感数据集上更好地平衡精度和速度。

为了验证所提方法的性能,开展了遥感图像细粒度目标检测的相关实验,实验细节如下:

数据集。HRSC2016(High-Resolution Ship Collection 2016)数据集包含 1 061 张 300×300 到 $1\ 500 \times 900$ 像素的图像,共标注了 2 976 艘 22 类细粒度船只对象,并按照官方方案划分为 436 张训练图像、181 张验证图像和 438 张测试图像。VEDAI 数据集(Vehicle Detection in Aerial Images, VEDAI)包含 1 210 张 $1\ 024 \times 1\ 024$ 的遥感图像,共标注了 3 640 个带有定向框注释的车辆对象,识别 6 类细粒度车辆对象,并按照 4:1 的比例划分为 968 张训练图像和 242 张测试图像。

实验参数。实验操作系统为 Ubuntu 16.04,硬件设备为 NVIDIA GeForce RTX 3080 GPU,深度学习框架为 PyTorch。优化器采用随机梯度下降(Stochastic Gradient Descent, SGD)算法,初始学习率为 0.005,动量设置为 0.9,权重衰减系数为 0.000 1。旋转检测器采用 Oriented R-CNN 结构,ResNet-101、ResNet-50 和

ResNet-18 作为骨干网络,所有方法的骨干网络均在 ImageNet 上预训练。在知识蒸馏中,学生网络的最大数量设置为 2,超参数 A 设置为 0.4,超参数 c 设置为 500。

评估指标。本文采用平均精度均值 (mean Average Precision, mAP)、每秒帧数 (Frames Per Second, FPS)、模型参数量 (Parameters) 和每秒浮点运算次数 (Floating-point Operations, FLOPs) 进行评估。

本文对所提 CF-ORNet 框架在不同基线模型上的性能增益进行了系统比较,结果如表 2 所示。表 2 中

详细列出了多种主流旋转目标检测方法在 HRSC2016 和 VEDAI 数据集上的原始 mAP,以及在结合 CF-ORNet 后的性能变化。实验结果表明,集成 CF-ORNet 后,各模型的 mAP 指标均获得显著提升,增幅介于 1.35%~7.32%,有力验证了所提方法的有效性和广泛适用性。此外,本文进一步提出了一种 AutoFGOR 算法,融合了扩散模型与视觉语言大模型,以提升细粒度表征能力。其在 HRSC2016 与 VEDAI 数据集上分别达到 88.05% 与 80.25% 的 mAP,显著优于基线方法,充分验证了多模态先验在细粒度识别中的性能潜力。

表 2 遥感图像细粒度目标检测法性能比较

单位:%

Table 2 Performance comparison of fine-grained detection methods on remote sensing images

unit:%

方法	主干网络	mAP	
		HRSC2016	VEDAI
SASM ^[38]	ResNet-50-FPN	74.70	76.16
S ² ANet ^[39]	ResNet-50-FPN	77.73	72.14
SRep-RDet ^[40]	RepVGG-B1g2-FPN	81.23	75.70
ReDet ^[41]	ReResNet-50-ReFPN	83.31	77.55
Oriented R-CNN ^[42]	ResNet-101-FPN	84.34	76.47
SASM ^[38] +CF-ORNet	ResNet-50-FPN	82.02	76.34
S ² ANet ^[39] +CF-ORNet	ResNet-50-FPN	82.83	76.09
SRep-RDet ^[40] +CF-ORNet	ResNet-50-FPN	82.49	77.01
ReDet ^[41] +CF-ORNet	ReResNet-50-FPN	85.07	77.20
Oriented R-CNN ^[42] +CF-ORNet	ResNet-101-FPN	85.69	77.34
所提 AutoFGOR ^[43]	ResNet-50-FPN	88.05	80.25

注:加粗字体为最优结果。

2.1.3 稀疏性优化

除剪枝、量化和蒸馏外,稀疏性优化在遥感目标检测中也具有重要价值。遥感图像往往包含大量冗余区域(如海面、林地等),若能够利用稀疏化思想按需分配计算资源,则可显著提升推理效率。

(1) 动态感知稀疏优化

在遥感图像目标检测领域,不同输入图像的计算需求存在差异(如单舰船检测与密集小目标检测)。自适应推理方法通过动态调整计算量,以适应输入复杂度或外部资源约束,从而显著降低计算量,提高推理效率。例如,Sun 等人^[44]提出基于低秩增强的检测框架,该方法通过构建三重创新机制,在保持检测精度的同时,将 FLOPs 降低约 65%,展现了强鲁棒性,为 UAV 全天候目标检测提供了高效解决方案。

利用稀疏性优化的思路,在直播视频主播重识别研究中^[45],针对多模态生物特征融合效率与识别精度的平衡问题,提出轻量级跨模态注意力网络(Light Cross-Modal Attention Network, LCMA-Net)。该方法通过构建多模态特征提取与融合框架实现高效识别,分

别提取声纹、面部与时空特征。此外,还设计了轻量跨模态池化注意力模块,结合多层感知机实现跨模态特征对齐与融合。其轻量化设计为实时主播身份验证提供了可行方案,同时为后续开展遥感图像目标检测算法的稀疏化优化设计提供了新的思路和方向。

(2) 样例分析

在遥感图像小目标检测的前期工作中,尝试了利用稀疏性优化的思路设计网络压缩的策略,针对多源遥感图像中的小目标检测问题,提出位置引导动态感受野网络(Position Guided Dynamic Receptive Field Network, PG-DRFNet)^[46]。如图 6 所示,其网络结构主要设计了动态感受野和组合头。该方法通过建立跨特征层的小目标位置引导关系,有效缓解特征传递过程中小目标信息丢失问题;同时,设计组合头,主要包括基础头、位置引导头和动态感知头三部分,引入小目标监督信号以增强模型判别能力。为平衡检测精度与计算效率,在动态感受野中设计动态感知算法通过对特征层级与感知区域的自适应稀疏优化,实现对不同尺度目标的精准捕获。

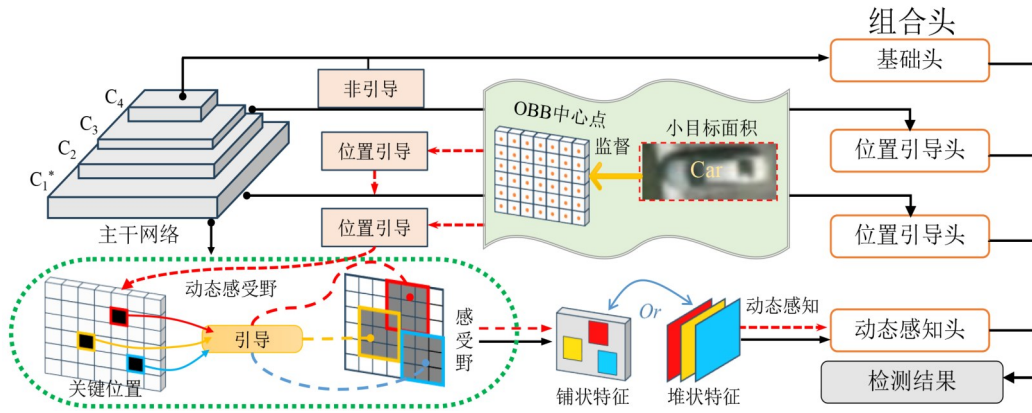


图6 遥感图像小目标检测稀疏性优化网络结构

Figure 6 Sparsity-optimized network architecture for small object detection in remote sensing images

实验结果在多个多模态数据集上验证了该方法的有效性,其关键思路如下:

(a)为应对小物体在特征中消失或被淹没这一挑战,提出了一种动态感受野,其通过在各种特征层之间建立位置引导关系来增强模型捕捉小尺寸物体的能力。

(b)设计了一种专门针对动态感受野(Dynamic Receptive Field, DRF)机制的新型组合头结构,其在平衡不同检测头的同时,结合了从小物体中提取的额外监督信息,从而使检测过程更有效、更灵活,并在多源遥感中提高了小目标检测的性能。

(c)为了最大限度发挥动态模型的优势,开发了一种基于特征构建的动态感知算法,该算法动态优化模型的感知区域和特征层次结构,以在检测精度和推理速度之间实现最优权衡,从而在各种场景中确保稳健性和效率。

为了验证所提方法的性能,开展了遥感图像小目标检测的相关实验,实验细节如下:

数据集。除前文提到的VEDAI数据集外,本实验还包括了以下数据集:DOTA-v2.0数据集、SSDD数据集、HRSID数据集以及VisDrone数据集。

实验参数。实验在Ubuntu 16.04操作系统、NVIDIA GeForce RTX 3090 GPU以及PyTorch深度学习框架上进行。优化器采用带权重衰减的Adam优化器,初始学习率设置为0.000 25,权重衰减系数为0.05。

为增强模型对小目标的感知能力,本研究在主干网络中引入通道注意力机制,并设置两个超参数 α 与 β 。 α 通过对卷积核 k 的归一化放大高响应区域权重,以强化显著特征; β 用于抑制低响应区域,从而提升特征判别性。基于特征响应统计与验证实验,最终将 α 和 β 分别设为2和1,以在小目标密集场景中实现稳定有效的注意力聚焦。

$$k = \left\lfloor \frac{\log_2(N)}{\alpha} + \frac{\beta}{\alpha} \right\rfloor \quad (1)$$

在小目标的稀疏回归分支中,为构建小目标面积的监督信息,本文找出(find)面积小于阈值 S 的所有小目标有向边界框(Oriented Bounding Box, OBB)并计算其均值矩阵 M 。据此建立子矩阵 b_i^{small} 以专门优化该类目标的回归性能。

$$b_i^{\text{small}} = b_i[M], M = \text{find}(w \cdot h \leq S) \quad (2)$$

式中, $w \cdot h$ 表示OBB的面积。本研究将 S 设置为32像素值,以平衡检测灵敏度与稀疏回归的计算效率。

在损失函数设计方面,本研究采用旋转IoU损失作为回归监督,并引入了质量焦点损失以增强定位质量与分类置信度的一致性。该损失通过软化独热标签并与IoU联合构成监督信号。

$$\mathcal{L}_{\text{BH}} = -\ln(\text{iou}) - |y - \theta|^\gamma \left((1 - y) \ln(1 - \theta) + y \ln(\theta) \right) \quad (3)$$

式中, γ 为损失函数中的调节因子,用于平衡正负样本并促进定位与分类的协同优化。其取值越大,对低质量预测的抑制越强,模型更关注高质量候选。本文遵循在质量焦点损失中的设置,将 γ 设置为2。该取值被证明能够在优化过程中有效兼顾精度提升与梯度收敛的稳定性。

评估指标。评估指标包括前文所提及的mAP、FPS、模型参数量以及FLOPs,本案例不涉及其他新的评估指标。

为了全面评估所提PG-DRFNet在遥感小目标检测任务中的综合性能,表3中提供了其与多种先进方法的详细对比结果。实验结果表明,PG-DRFNet在多数数据集上均取得了最优或接近最优的检测精度,尤其是在VEDAI数据集上mAP达到84.06%,显著优于其他方法。同时,该方法在模型参数量与推理速度之间也实现了良好平衡(26.6 M, 27.0 FPS),体现出优异的实用价值。

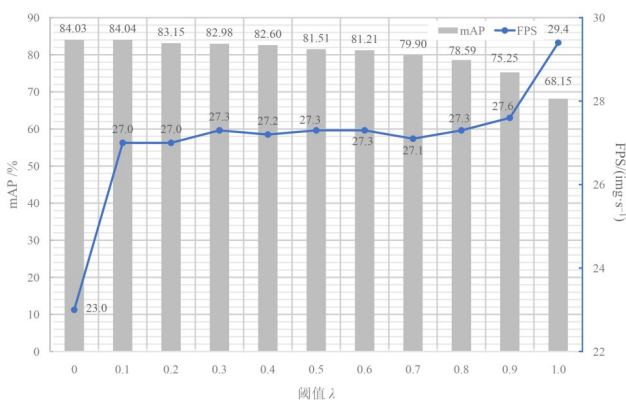
表 3 遥感图像小目标检测方法性能比较

Table 3 Performance comparison of small object detection methods on remote sensing images

数据集	VEDAI	DOTA-v2.0	SSDD	HRSID	复杂度		
评价指标	mAP/%				模型参数量/M	GFLOPs/G	FPS
S ² ANet ^[39]	72.14	52.44	89.57	79.44	36.2	197	27.0
Oriented R-CNN ^[42]	76.47	52.94	89.45	71.38	60.1	122	15.8
CF-ORNet ^[36]	77.34	57.71	89.85	74.40	52.3	93	13.7
LSKNet-S ^[47]	81.44	54.18	90.01	80.15	31.0	174	20.5
RTMDet-M ^[48]	80.44	57.71	89.50	79.85	24.7	100	20.2
所提 PG-DRFNet ^[46]	84.06	59.01	90.06	80.59	26.6	236	27.0

注: 加粗字体为最优结果。

在所提方法中, 特征稀疏程度由阈值 λ 动态调控, 实验结果如图 7 所示。当 $\lambda=0$ 时, 模型保留全部特征, 检测精度达到最高 (84.04% mAP), 但推理速度仅为 20.3 FPS, 属于“全特征、低效率”的工作模式。随着 λ 逐步增大, 冗余特征被有选择地抑制, 当 λ 增至 0.8 附近时, 推理速度提升至约 27 FPS, 而精度仅出现轻微下降。需要注意的是, 当 $\lambda>0.6$ 时, 过度稀疏化会削弱模型对小目标与边缘区域的表征能力; 当 $\lambda=1$ 时, 由于完全舍弃候选特征, mAP 下降至 68.15%, 尽管帧率上升至 29.4 FPS, 但已难以满足遥感小目标检测的精度要求。综合不同区间的性能表现可知, λ 在 0.1~0.6 时能够较好地平衡精度与速度, 在维持 84% mAP 以上时, 可实现约 26~27 FPS 的实时推理性能。该结果表明, 所提出的动态稀疏优化机制在“背景广、目标稀疏”的遥感场景下具有效率-精度协同优势, 同时也为实际部署中的阈值选取提供了可参考的工作区间。

图 7 不同阈值 λ 的敏感性分析Figure 7 Sensitivity analysis of different thresholds λ

2.2 模型规模视角

在遥感图像目标检测中, 算法与模型规模是彼此关联的两个关键维度。第 2.1 节从算法视角出发, 分析了剪枝、量化、蒸馏与稀疏性优化在遥感任务中的应用与演进, 而这些方法高度依赖于模型的体量与结

构。因此, 本节转向模型规模视角, 讨论轻量模型与大模型在压缩与加速策略上的差异及各自适配特点。

2.2.1 轻量模型压缩

压缩轻量模型的主要目标是在高性能和最小计算资源之间保持平衡, “轻量模型”主要指参数量通常在数百万至数千万级别, 且可在消费级图形处理器 (Graphics Processing Unit, GPU) 或嵌入式边缘设备上实现实时或准实时推理的网络结构。这对于边缘设备上的实时遥感处理至关。

(1) 轻量化 CNN 架构的适配与优化

对于轻量模型而言, 即使已采用 MobileNet^[49]、EfficientNet^[50] 等轻量化 CNN 架构, 在遥感图像目标检测场景下, 仍需进一步结合任务特性与硬件约束进行结构与训练策略的适配与优化。例如, 通过调整损失函数、引入任务感知模块或进行超参数搜索等方式, 在不显著增加参数量的前提下, 提升模型在高分辨率、复杂背景与小目标场景下的性能。因此, 本小节重点探讨基于轻量化 CNN 架构的任务适配与二次优化实践。例如, 在之前的工作中^[51], 将轻量化 CNN 架构的思路运用到标志检索任务中, 提出显著性引导的 CNN, 为智能交通系统的实时数据处理提供了高效解决方案。

(2) 轻量化 Transformer 架构的适配改造

将 Transformer 模型整合到遥感图像目标检测任务中已展现卓越成效, 如视觉 Transformer (Vision Transformer, ViT)^[52]、Swin Transformer^[53]。然而, 由于自注意力层的二次复杂度, 这些模型计算成本高昂。为缓解这一问题, 已开发出诸多 (如 MobileViT^[54] 和 Twins^[55] 等) 轻量级 Transformer 变体, 以适应移动和嵌入式设备。MobileViT 首先在早期层使用卷积块提取低级特征, 然后采用 Transformer 层进行高级处理, 将卷积和 Transformer 的优势相结合。这种混合方法在保持 Transformer 全局上下文优势的同时降低了计算成本。

2.2.2 大模型压缩

近年来, 大模型在跨模态理解与知识推理中展现

出卓越能力,尝试将目标检测、分割等任务整合至统一框架,并借助文本先验增强遥感场景的语义理解。早期工作多聚焦于遥感图像多模态对话模型^[56],近期则转向零样本识别与开放语义理解,以更低的监督成本构建通用遥感基础模型。然而,这类大模型同时伴随计算与存储开销的急剧增长,严重制约了其在实际遥感应用中的落地。本节从 MoE 模型、KV 缓存优化和混合量化 (Mixed-Precision Quantization, MPQ) 三个角度探讨大模型压缩技术在遥感中的适配与进展。

(1) MoE 模型

MoE 模型通过在推理过程中仅激活部分专家网络,实现了在不显著增加计算成本的前提下扩展模型规模。对于遥感多模态任务(如光学与 SAR 融合), MoE 模型可为不同模态设计专属专家,从而在高效分工的同时避免特征冗余。例如, Liu 等人^[57]提出先验增强的 MoE 融合网络,以动态门控与频段解耦策略实现环境自适应多模态特征融合,并用跨尺度交叠融合压缩计算,为星上实时多模态目标检测提供了轻量且鲁棒的新范式。此外, Lin 等人^[58]将基于损失距离的 K -专家聚类 (Loss Distance-based K -Experts Clustering, LD-KEC) 驱动的“多专家”架构引入遥感目标检测,首次以无监督方式训练多特征-多专家网络 (MultiExpert Detection Network, MEDNet),从而在单管线框架之外显著提升复杂地物识别能力,为模型压缩时代的精准轻量检测开辟了新路径。

(2) KV 缓存优化

大模型的注意力机制通常需存储庞大的 KV 缓存,这在处理高分辨率遥感图像时,尤其是全景或连续视频序列,带来了巨大的内存压力。KV 缓存优化通过减少存储与计算冗余,使模型在边缘平台更具实用性。例如, Qiu 等人^[59]提出的 KV 缓存联合优化框架在 UAV-边缘计算协同场景下,通过缓存复用与任务卸载优化,显著降低了能耗与延迟。该方法可直接应用于 UAV 编队的实时监测,在多架 UAV 在处理相似区域图像时复用缓存信息,避免重复计算,并通过轨迹调度与传输优化实现高效任务分配。这类方法有效缓解了大模型在边缘部署时的存储瓶颈,使得目标检测与变化监测等任务能够在动态干扰环境下稳定运行。

(3) 混合量化

混合量化 (Mixed-Precision Quantization, MPQ) 根据不同模块对精度的敏感性分配不同的量化位宽,在保证关键任务精度的同时大幅压缩计算与存储开销^[60]。在遥感大模型应用中,注意力层和分类头需保持较高的精度,以保障小目标检测与细粒度分类的可靠性,而特征融合或中间层则可量化至低比特以节省资源。Li 等人^[61]针对红外小目标分割提出

对称混合量化 (Symmetrically Progressive decreasing Mixed-precision Quantization, SPMix-Q), 按层敏感度渐进降比特,在模型体积压缩 13 倍的同时保持与全精度相当的 IoU,为星载/机载端实时小目标探测提供了可落地的超低比特方案。

综上所述,面向遥感场景的大模型压缩主要沿三条互补路径展开。基于 MoE 的条件计算通过按需激活异构专家,能够在保持表征能力的同时降低平均计算成本,适用于多模态、多任务一体化解译框架; KV 缓存优化致力于缓解高分辨率影像与序列处理中的内存与带宽压力,更适合卫星-UAV-地面协同观测等连续处理场景;混合量化通过为不同模块分配差异化比特宽度,可在保障注意力层与分类头精度的前提下显著压缩整体模型体积,促进其在星上及边缘设备的实际落地。这类技术并非相互排斥,未来可与剪枝、蒸馏及参数高效微调等方法进一步协同,构成面向遥感大模型的高效压缩工具箱。

3 研究热点

随着遥感技术发展以及深度学习在遥感图像处理中的广泛应用,针对遥感数据高分辨率、多模态、实时性要求等特点,近年来出现了多项前沿技术与方法^[62]。本文重点讨论两个热点方向:针对遥感多模态数据的混合专家策略,以及引入参数高效微调技术作为跨领域知识迁移与轻量化训练的重要补充手段。

3.1 针对遥感多模态数据的混合专家策略

遥感数据具有多源异构的特点,不同模态数据在信息表征上存在显著差异。这使得单一模型难以同时兼顾多模态数据的特征表达与噪声抑制,易在融合过程中出现信息冗余或有效特征丢失。为此,混合专家策略 MoE 逐渐成为遥感多模态建模中的研究热点^[63]。在混合专家策略中所指“异构专家”,特指在网络结构、感受野大小、特征维度、输入模态或关注尺度范围等方面存在显著差异的一组子网络/子模块。通过门控机制按需激活不同类型的专家,可针对多模态、多尺度及多任务的遥感场景实现更具针对性的特征提取与融合,相较同构 MoE 结构,其在适应复杂遥感数据分布方面具有更高的灵活性。

整体而言,混合专家策略在遥感多模态建模中兼顾了精度、灵活性与计算成本,为多源数据融合提供了一种具备可扩展性与可部署性的解决方案。

3.2 参数高效微调技术

为降低大模型在遥感任务中的迁移成本,并缓解灾难性遗忘等问题,参数高效微调技术 (Parameter-Efficient Fine-Tuning, PEFT) 成为近年来的重要研究方

向。该类方法通过仅更新模型中的少量参数模块,在保持原始模型主体不变的前提下,实现任务适应性训练,适用于遥感等计算资源受限的场景^[64]。典型代表如低秩适应方法(Low-Rank Adaptation, LoRA)^[65],通过在权重矩阵中引入低秩分解组件,有效模拟全量微调效果。其变体如 QLoRA^[66]通过量化与 LoRA 结合,使得在消费级硬件上也能微调超大规模模型,具备极高的实用性。此外,上下文扩展方法(如 LongLoRA^[67])通过提升输入窗口长度,使模型能够处理更大尺度的遥感图像数据。

综合来看,混合专家策略为遥感图像目标检测及多模态数据处理提供了新的思路和技术路径,而参数高效微调技术则为大规模预训练模型在资源受限平台上的部署提供了强有力的支持。这些研究热点和趋势相互交织,共同推动遥感智能处理技术向更高效的方向发展。

4 总结与展望

4.1 存在的问题

遥感图像目标检测网络压缩与加速技术虽取得阶段性突破,但仍面临三个方面的技术瓶颈。

(1) 大模型压缩效能衰减

随着深度学习模型规模的不断增大,传统的压缩方法在处理超大模型时仍存在局限。虽然这些方法在一定程度上缓解了计算压力,但对于大规模模型的处理能力有限。

(2) 精度-效率平衡问题

诸多网络压缩方法在提高推理速度的同时,会导致精度损失。尤其是在高分辨率遥感图像处理任务中,如何在保证高精度的同时加速推理,仍然是一个难题。

(3) 多模态特征坍塌风险

遥感数据常涉及光学、SAR 等多模态信息,现有压缩策略通常以单模态结构为基础,难以在降维或剪枝过程中保持跨模态对齐与互补信息,导致融合时易出现特征坍塌与性能瓶颈。

这些挑战本质上源于遥感智能解译的物理约束与数据特性间的固有矛盾:(a)地物多尺度特性要求模型具备跨层级特征保持能力;(b)时效性约束强制压缩过程需考虑图像计算稀疏性;(c)多模态异质特征对齐需要压缩过程保持模态间映射关系。

4.2 未来发展方向

未来,遥感图像目标检测网络压缩和加速技术有望在多个方向上实现进一步突破。

首先,自适应压缩将成为核心趋势。随着任务多样性与计算约束并存,自适应剪枝与量化将根据任务

复杂度、数据特性及资源条件动态调整压缩强度。可通过引入注意力机制实现跨尺度特征选择,从而精准适配小目标与复杂背景。同时,探索自适应策略在多源数据中的迁移能力,可提升异构场景下的泛化性。

其次,多模态融合与跨任务知识迁移将推动轻量化模型的高阶发展。面对光学、SAR、红外等多源数据,应探索协同压缩与跨任务蒸馏机制,实现不同模态间的高效知识共享。同时,研究混合量化与统一编码策略,在保留判别信息的同时显著降低存储与计算成本。

再次,PEFT 可与传统压缩技术形成互补协同。一方面,可采用“先压缩、后适配”的两阶段策略,即先通过剪枝、量化与稀疏化获得满足算力与存储约束的基础大模型,再叠加少量 LoRA、Adapter 等任务特定参数,实现对不同遥感任务的轻量级定制。另一方面,在多区域、多传感器、多任务并存的遥感应用中,同一压缩后的基础模型可在不同场景间共享,仅按需加载对应的 PEFT 模块即可完成任务切换,从而显著降低多任务部署与动态更新的成本。进一步地,结合量化感知训练的 PEFT 框架,有望在极低比特宽度下实现大模型的快速域迁移,为星上及机载平台提供灵活可升级的解决方案。

最后,在绿色低碳 AI 背景下,能效约束的轻量化将愈发重要。遥感系统需融合能耗感知的压缩与推理优化,在保持高性能的同时控制能源消耗。未来可建立结合碳排放度量与能耗预算的可持续压缩框架,推动遥感模型向低功耗、绿色智能方向发展。

综上所述,通过自适应压缩、多模态迁移、参数高效微调与绿色计算等方向的持续探索,遥感图像目标检测中的网络压缩与加速技术将不断向更高效、更灵活、更节能的方向演进。

参考文献

- [1] Lu S Q, Guo J L, Zimmer-Dauphinee J R, et al. Vision foundation models in remote sensing: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2025, 13(3): 190-215.
- [2] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [3] Chen Z, Wu J N, Wang W H, et al. Intern VL: Scaling up vision foundation models and aligning for generic visual-linguistic tasks[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 24185-24198.

- [4] 苏航, 徐从安, 姚力波, 等. 一种轻量化 SAR 图像舰船目标斜框检测方法[J]. 航空学报, 2022, 43(S1): 157-164.
Su Hang, Xu Congan, Yao Libo, et al. A lightweight oriented ship detection method in SAR images[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(S1): 157-164. (in Chinese)
- [5] Zubko V, Kaufman Y J, Burg R I, et al. Principal component analysis of remote sensing of aerosols over oceans[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(3): 730-745.
- [6] Jia S, Qian Y T. Constrained nonnegative matrix factorization for hyperspectral unmixing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(1): 161-173.
- [7] Johansen K, Raharjo T, McCabe M F. Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects[J]. *Remote Sensing*, 2018, 10(6): 854.
- [8] Li P, Han L R, Tao X W, et al. Hashing nets for hashing: A quantized deep learning to hash framework for remote sensing image retrieval[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(10): 7331-7345.
- [9] Chen J Z, Wang S H, Chen L, et al. Incremental detection of remote sensing objects with feature pyramid and knowledge distillation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5600413.
- [10] Gu S S, Sun X Y, Yang Z H, et al. Energy-aware coded caching strategy design with resource optimization for satellite-UAV-vehicle-integrated networks[J]. *IEEE Internet of Things Journal*, 2022, 9(8): 5799-5811.
- [11] Wang L Y, Hu P, Zheng H W, et al. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images[J]. *Geoderma*, 2023, 430: 116321.
- [12] Chen B W, Chen K Y, Yang M H, et al. Heterogeneous mixture of experts for remote sensing image super-resolution[J]. *IEEE Geoscience and Remote Sensing Letters*, 2025, 22: 6006405.
- [13] Razakarivony S, Jurie F. Vehicle detection in aerial imagery: A small target detection benchmark[J]. *Journal of Visual Communication and Image Representation*, 2016, 34: 187-203.
- [14] Liu Z K, Yuan L, Weng L B, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//*Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS - Science and Technology Publications. Porto: SCITEPRESS, 2017: 324-331.
- [15] Long Y, Gong Y P, Xiao Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2486-2498.
- [16] Ding J, Xue N, Xia G S, et al. Object detection in aerial images: A large-scale benchmark and challenges[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 7778-7796.
- [17] Lam D, Kuzma R, McGee K, et al. xView: Objects in context in overhead imagery[PP/OL]. V1.arXiv (2018-02-22)[2025-12-11]. <https://doi.org/10.48550/arXiv.1802.07856>.
- [18] Zhu P F, Wen L Y, Du D W, et al. Detection and tracking meet drones challenge[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 7380-7399.
- [19] Sun X, Wang P J, Yan Z Y, et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 184: 116-130.
- [20] Zhang T, Zhang X, Li J, et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis[J]. *Remote Sensing*, 2021, 13(18): 3690.
- [21] Xiao M, He Z, Li X Y, et al. Power transformations and feature alignment guided network for SAR ship detection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 4509405.
- [22] 徐从安, 苏航, 李健伟, 等. RSDD-SAR:SAR舰船斜框检测数据集[J]. 雷达学报, 2022, 11(4): 581-599.
Xu Congan, Su Hang, Li Jianwei, et al. RSDD-SAR: Rotated ship detection dataset in SAR images[J]. *Journal of Radars*, 2022, 11(4): 581-599. (in Chinese)
- [23] Zhang X, Yang X, Li Y X, et al. RSAR: Restricted state angle resolver and rotated SAR benchmark[PP/OL]. V1.arXiv (2025-01-08)[2025-12-11]. <https://doi.org/10.48550/arXiv.2501.04440>.
- [24] Bondi E, Jain R, Aggrawal P, et al. BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos[C]//*2020 IEEE Winter Conference on Applications of Computer Vision*. Piscataway: IEEE, 2020: 1736-1745.
- [25] Dai Y M, Wu Y Q, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection[C]//*2021 IEEE Winter Conference on Applications of Computer Vision*. Piscataway: IEEE, 2021: 949-958.
- [26] Lu J H, Liu D S, Cheng X, et al. An efficient unstructured sparse convolutional neural network accelerator for wearable ECG classification device[J]. *IEEE Transactions on*

- Circuits and Systems I: Regular Papers, 2022, 69(11): 4572-4582.
- [27] Hu Z P, Gong M G, Lu Y H, et al. Data customization-based multiobjective optimization pruning framework for remote sensing scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4508514.
- [28] Xie Y F, Hou X, Ren J C, et al. Binary quantization vision transformer for effective segmentation of red tide in multi-spectral remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4202814.
- [29] Wang R Q, Fu Y F, Xu Y, et al. Dual prediction-guided distillation for object detection in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5618817.
- [30] Yang Y R, Sun X, Diao W H, et al. Adaptive knowledge distillation for lightweight remote sensing object detectors optimizing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5623715.
- [31] Yang X, Zhang S, Yang W C. Two-way assistant: A knowledge distillation object detection method for remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5612710.
- [32] Wang L Y, Zhang J, Yao J C, et al. Porn streamer recognition in live video based on multimodal knowledge distillation[J]. *Chinese Journal of Electronics*, 2021, 30(6): 1096-1102.
- [33] 王立元. 基于多模态深度学习的网络直播视频不良主播识别技术与实现[D]. 北京: 北京工业大学, 2020. Wang Liyuan. Identification technology and implementation of bad anchor in webcast video based on multimodal deep learning[D]. Beijing: Beijing University of Technology, 2020. (in Chinese)
- [34] Han L R, Paoletti M E, Moreno-álvarez S, et al. Deep robust hashing using self-distillation for remote sensing image retrieval[C]//IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE, 2024: 7111-7114.
- [35] Hu Y T, Zhang L, Luo X Y, et al. Diffusion self-distillation for remote sensing scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5626315.
- [36] Wang L Q, Zhang J, Tian J M, et al. Efficient fine-grained object recognition in high-resolution remote sensing images from knowledge distillation to filter grafting[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4701016.
- [37] 田吉淼. 高分辨率遥感影像的细粒度目标识别技术实现[D]. 北京: 北京工业大学, 2022. Tian Jimiao. Realization of fine-grained target recognition technology for high-resolution remote sensing images[D]. Beijing: Beijing University of Technology, 2022. (in Chinese)
- [38] Hou L P, Lu K, Xue J, et al. Shape-adaptive selection and measurement for oriented object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(1): 923-932.
- [39] Han J M, Ding J, Li J, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5602511.
- [40] Tian J M, Zhang J, Li W S, et al. Structurally re-parameterized rotation detector for arbitrary-oriented objects in high-resolution remote sensing images[J]. *International Journal of Remote Sensing*, 2022, 43(1): 241-269.
- [41] Kingma D P, Salimans T, Poole B, et al. Variational diffusion models[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: ACM, 2021: 21696-21707.
- [42] Xie X X, Cheng G, Wang J B, et al. Oriented R-CNN for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 3500-3509.
- [43] Wang L Q, Zhang J, Mi G M, et al. Weak supervision makes strong details: Fine-grained object recognition in remote sensing images via regional diffusion with VLM[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2026, 233: 231-246.
- [44] Sun X, Yu Y H, Cheng Q. Low-rank multimodal remote sensing object detection with frequency filtering experts[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5637114.
- [45] Yao J C, Zhang J, Zhang H, et al. LCMA-Net: A light cross-modal attention network for streamer re-identification in live video[J]. *Computer Vision and Image Understanding*, 2024, 249: 104183.
- [46] Wang L Q, Li J F, Zhang J, et al. Position guided dynamic receptive field network: A small object detection friendly to optical and SAR images[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(8): 8265-8282.
- [47] Li Y X, Hou Q B, Zheng Z H, et al. Large selective kernel network for remote sensing object detection[C]//2023 IEEE/CVF International Conference on Computer Vision.

- Piscataway: IEEE, 2023: 16748-16759.
- [48] Lyu C Q, Zhang W W, Huang H A, et al. RTMDet: An empirical study of designing real-time object detectors[PP/OL]. V2.arXiv (2022-12-16)[2025-12-11]. <https://doi.org/10.48550/arXiv.2212.07784>.
- [49] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1314-1324.
- [50] Tan M, Le V Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning. California: PMLR, 2019: 6105-6114.
- [51] Liang X, Zhang J, Tian Q, et al. A saliency guided shallow convolutional neural network for traffic signs retrieval[C]//2018 IEEE Conference on Multimedia Information Processing and Retrieval. Piscataway: IEEE, 2018: 340-345.
- [52] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[PP/OL]. V2.arXiv (2021-06-03)[2025-12-11]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [53] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9992-10002.
- [54] Mehta S, Rastegari M. MobileViT: Light-weight, general-purpose, and mobile-friendly vision transformer[PP/OL]. V2.arXiv (2022-03-04)[2025-12-11]. <https://doi.org/10.48550/arXiv.2110.02178>.
- [55] Chu X, Tia Z, Wang Y, et al. Twins: Revisiting the design of spatial attention in vision transformers[C/OL]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event: Curran Associates, Inc., 2021: 9355-9366. <https://dl.acm.org/doi/10.5555/3540261.3540977>.
- [56] Zhang W, Cai M X, Zhang T, et al. EarthMarker: A visual prompting multimodal large language model for remote sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5604219.
- [57] Liu K W, Peng D L, Li T. Multimodal remote sensing object detection based on prior-enhanced mixture-of-experts fusion network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5405314.
- [58] Lin Q F, Zhao J H, Du B, et al. MEDNet: Multiexpert detection network with unsupervised clustering of training samples[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4703114.
- [59] Qiu J L, Kuang Z F, Huang Z Q, et al. Security offloading scheduling and caching optimization algorithm in UAV edge computing[J]. IEEE Systems Journal, 2025, 19(1): 96-106.
- [60] Kim M, Hong S, Ko R, et al. Oaken: Fast and efficient LLM serving with online-offline hybrid KV cache quantization[PP/OL]. V2.arXiv (2025-05-14)[2025-12-11]. <https://doi.org/10.48550/arXiv.2503.18599>.
- [61] Li B Y, Wang L G, Wang Y Q, et al. Mixed-precision network quantization for infrared small target segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5000812.
- [62] 曹峰, 曾科文, 李德玉, 等. 基于EIMYOLO的高分遥感图像目标检测[J]. 电子学报, 2025, 53(7): 2266-2278.
Cao Feng, Zeng Kewen, Li Deyu, et al. Object detection based on EIMYOLO for high-resolution remote sensing images[J]. Acta Electronica Sinica, 2025, 53(7): 2266-2278. (in Chinese)
- [63] Lin H, Hong D F, Ge S H, et al. RS-MoE: A vision-language model with mixture of experts for remote sensing image captioning and visual question answering[PP/OL]. V2.arXiv (2025-02-10)[2025-12-11]. <https://doi.org/10.48550/arXiv.2411.01595>.
- [64] 秦董洪, 李政韬, 白凤波, 等. 大语言模型参数高效微调技术综述[J]. 计算机工程与应用, 2025, 61(16): 38-63.
Qin Donghong, Li Zhengtao, Bai Fengbo, et al. Review of parameter-efficient fine-tuning technology for large language models[J]. Computer Engineering and Applications, 2025, 61(16): 38-63. (in Chinese)
- [65] Hu E J, Shen Y, Wallis P, et al. LoRA: Low-rank adaptation of large language models[C]//International Conference on Learning Representations. Virtual Event: ICLR, 2022.
- [66] Dettmers T, Holtzman A, Pagnoni A, et al. QLoRA: Efficient finetuning of quantized LLMs[C]//Advances in Neural Information Processing Systems 36. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023: 10088-10115.
- [67] Chen Y K, Qian S J, Tang H T, et al. LongLoRA: Efficient fine-tuning of long-context large language models[PP/OL]. V3.arXiv (2024-03-08)[2025-12-11]. <https://doi.org/10.48550/arXiv.2309.12307>.

作者简介



王柳谦 男,1997年1月生,河南平顶山人。北京工业大学信息科学技术学院博士研究生,现为郑州大学网络空间安全学院讲师。主要研究方向为遥感影像目标检测与识别等。
E-mail: wangliuqian@zzu.edu.cn



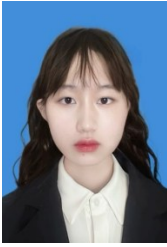
谢笑阳 女,1991年2月生,湖南宁远人。博士,现为北京工业大学计算机学院讲师、研究生导师。主要研究方向为图像处理、人工智能、遥感图像处理以及目标识别等。
E-mail: xiexiaoyang@bjut.edu.cn



张菁 女,1975年2月生,广东梅县人。博士,现为北京工业大学信息科学技术学院教授、博士生导师。主要研究方向为图像处理、图像识别、图像检索等。
E-mail: zhj@bjut.edu.cn



卓力 女,1971年10月生,江苏徐州人。博士,现为北京工业大学信息科学技术学院教授、博士生导师。主要研究方向为图像/视频的编码与传输、多媒体大数据处理等。
E-mail: zhuoli@bjut.edu.cn



赵一 女,2000年8月生,河南许昌人。现为北京工业大学信息科学技术学院博士研究生。主要研究方向为目标检测与图像处理等。
E-mail: ZZYY-@emails.bjut.edu.cn