计算机科学
COMPUTER SCIENCE

基于小目标特征增强RT-DETR的SAR图像舰船目标检测方法

张弘森, 吴蔚, 徐建, 吴飞, 季一木

引用本文

张弘森, 吴蔚, 徐建, 吴飞, 季一木. 基于小目标特征增强RT-DETR的SAR图像舰船目标检测方法[J]. 计算机科学, 2025, 52(10): 151-158.

ZHANG Hongsen, WU Wei, XU Jian, WU Fei, JI Yimu. Ship Detection Method for SAR Images Based on Small Target Feature Enhanced RT-DETR [J]. Computer Science, 2025, 52(10): 151-158.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250100097

基于小目标特征增强 RT-DETR 的 SAR 图像舰船目标检测方法

张弘森[1,2]　吴　蔚[2]　徐　建[2]　吴　飞[1,2]　季一木[3]
1 南京邮电大学自动化学院、人工智能学院　南京 210003
2 信息系统工程全国重点实验室　南京 210003
3 南京邮电大学计算机学院　南京 210003
(zhs9586@163.com)

摘　要　在舰船检测任务中,SAR 图像因其优异的成像条件被广泛应用于海洋资源管理、海上救援等场景。然而,舰船目标尺寸较小和海面杂波等问题,导致传统目标检测算法的性能表现不佳。近年来,许多算法通过引入 Transformer 的注意力机制,实现更好的语义解释;或采用较为复杂的网络结构,以提高特征提取能力。这在一定程度上改善了检测精度,却牺牲了检测速度。对此,提出了一种基于小目标特征增强 RT-DETR 的 SAR 图像舰船目标检测方法。该方法由以下 3 部分组成:1)大模型提示生成网络:借助多模态大模型的零样本学习能力生成提示,以提取图像模态中更具判别性的信息;2)AIFI-EAA 模块:以 RT-DETR 为基线,改进尺度内特征交互模块,引入高效加性注意力机制,降低算法计算复杂度;3)轻量化小目标特征增强融合网络:在多尺度特征融合网络中加入小目标检测层,设计 CSP-OmniKernel 模块进行多尺度特征融合,提升小目标的检测性能。在 SSDD,HRSID 和 SAR-Ship-Dataset 3 个公开数据集上进行实验验证,结果表明所提方法在准确性上具有优势。

关键词:舰船检测;SAR 图像;轻量化;RT-DETR;小目标检测
中图分类号　TP751

Ship Detection Method for SAR Images Based on Small Target Feature Enhanced RT-DETR

ZHANG Hongsen[1,2],WU Wei[2],XU Jian[2],WU Fei[1,2] and JI Yimu[3]
1 College of Automation,Nanjing University of Posts and Telecommunications,Nanjing 210003,China
2 National Key Laboratory of Information Systems Engineering,Nanjing 210003,China
3 School of Computer Science,Nanjing University of Posts and Telecommunications,Nanjing 210003,China

Abstract　In ship detection tasks,SAR images are widely used in maritime resource management,search and rescue,and other scenarios due to their excellent imaging conditions. However,traditional target detection algorithms perform poorly due to issues such as the small size of ships and sea surface clutter. Recently,many algorithms have introduced the attention mechanism of Transformer to achieve better semantic interpretation or adopted more complex network structures to improve feature extraction capabilities. This has improved detection accuracy to some extent but has sacrificed detection speed. This paper proposes a ship detection method for SAR images based on small target feature enhanced RT-DETR. The method consists of three parts:1)Large model prompt generation network:Leveraging the zero-shot learning capability of multimodal large models,prompts are generated to extract more discriminative information from the image modality;2)AIFI-EAA module:Using RT-DETR as the baseline,the scale-invariant feature interaction module is improved by introducing an efficient additive attention mechanism to reduce the computational complexity of the algorithm;3)Lightweight small target feature enhancement fusion network:A small target detection layer is added to the multi-scale feature fusion network,and the CSP-OmniKernel module is designed for multi-scale feature fusion to enhance small target detection performance. Experiments on three public datasets(SSDD,HRSID,and SAR-Ship-Dataset) demonstrate that the proposed method has advantages in terms of accuracy.

Keywords　Ship detection,SAR image,Lightweight,RT-DETR,Small target detection

1　引言

近年来,在新一代卫星部署的推动下,微波成像技术取得了重大进步。合成孔径雷达(Synthetic Aperture Radar,SAR)是一种主动微波遥感器,利用散射电磁波脉冲的发射和接收来实现微波成像[1],已成为遥感领域最适合海洋监测的

到稿日期:2025-01-15　返修日期:2025-04-24
基金项目:信息系统工程全国重点实验室开放基金(05202305);国家自然科学基金(62076139)
This work was supported by the Science and Technology on Information System Engineering Laboratory(05202305) and National Natural Science Foundation of China(62076139).
通信作者:吴飞(wufei_8888@126.com)

成像方法[2]。与光学遥感图像相比，SAR 成像不受日光条件影响，具有分辨率高、穿透能力强、探测范围大等优点。因此，利用 SAR 图像进行船舶检测具有更广泛的应用，在海洋资源管理、海上救援、领海管辖等场景中发挥着重要作用。

在卷积神经网络(Convolutional Neural Network，CNN)和深度学习架构的推动下，计算机视觉领域发生了深刻的变革。目标检测作为计算机视觉的重要任务之一，旨在对图像中的特定目标进行定位与分类。近年来，研究者提出了多种目标检测方法。其中，CNN 通过对图像进行特征提取和表示，能够有效地处理不同尺寸、纹理和色彩的图像，在目标检测任务中表现出色。例如，以快速区域卷积神经网络(Faster Region-based CNN，Faster R-CNN)[3]为代表的两阶段目标检测算法，以及以 YOLO(You Only Look Once)[4]为代表的单阶段目标检测算法，在精度和效率方面都展现出较好的性能。

近年来，出现了许多基于 Transformer 的视觉模型。2020 年 Facebook 团队首次提出基于 Transformer 的端到端目标检测算法(Detection Transformer，DETR)[5]，代表着 Transformer 架构在目标检测任务中有了重要突破。DETR 不需要预定义的先验锚框，也不需要非极大值抑制的后处理策略，就可以实现端到端的目标检测。最近，百度提出的 RT-DETR(Real-Time Detection Transformer)[6]改进了编码器-解码器结构，解决了 DETR 模型训练收敛速度慢、计算成本高的问题。但是，DETR 这类算法由于自注意力机制的特性，在处理小目标时可能表现不佳，特别是场景中存在大量小目标的情况，其在局部区域的特征表达能力相对较弱。

SAR 图像中的船舶目标形状变化和尺度差异较大，远海岸以大量密集的舰船小目标为主，模型难以从背景杂波中有效提取关键特征。目前的主流算法在处理舰船小目标和复杂背景时，通常采用较为复杂的网络结构[7-9]，这虽然在一定程度上改善了检测精度，但也显著增加了计算开销，从而牺牲了检测速度和实时性。

针对以上问题，本文提出了一种基于小目标特征增强 RT-DETR 的 SAR 图像舰船目标检测方法(Small Target Feature Enhanced RT-DETR，STFE-RTDETR)。本文的主要贡献如下：

1)提出了大模型提示生成网络(Large Model Prompt Generative Network，LMPGNet)，通过创新的多模态前缀微调策略来实现轻量级微调，显著提升了多模态大模型 Paligemma[10]在舰船检测任务中的适应性，并借助大模型强大的零样本学习能力，引入多模态大模型生成提示，以获取图像模态中包含的更具判别性的信息。

2)以 RT-DETR 作为基线网络，改进基于注意力的尺度内特征交互模块(Attention-based Intra-scale Feature Interaction，AIFI)，提出了 AIFI-EAA 模块，引入一种高效的加性注意力机制(Efficient Additive Attention，EAA)[11]，用于替代传统自注意力机制的二次方矩阵乘法运算，从而降低了模型的计算复杂度，在不影响模型准确性的同时实现更快的检测速度。

3)提出了轻量化小目标特征增强融合网络(Lightweight Small-object Feature-enhancement Fusion Network，LSFF-Net)，在原本的多尺度特征融合网络中加入小目标检测层，并引入跨阶段部分连接(Cross Stage Partial，CSP)思想和全内核(OmniKernel)网络[12]，设计 CSP-OmniKernel 模块进行特征融合，以有效地学习从全局到局部的特征表征，从而提高小目标的检测性能。

4)在 3 个广泛使用的 SAR 舰船图像数据集 SSDD[13]，HRSID[14]和 SAR-Ship-Dataset[15]上进行实验，实验结果验证了 STFE-RTDETR 方法的有效性。

2 相关工作

2.1 RT-DETR 目标检测模型

RT-DETR 是一种基于 Transformer 的实时目标检测模型，主要由主干网络、高效混合编码器和解码器组成。RT-DETR 旨在消除非极大值抑制(Non-Maximum Suppression，NMS)造成的推理延迟，并且在速度和精度上都优于同规模的 YOLO 检测器。

首先将图像进行数据预处理，然后将其输入主干网络。主干网络用于特征提取。RT-DETR 采用 ResNet[16]或 HGNet[17]的 CNN 网络作为主干网络。骨干网络的最后 3 个阶段(S3，S4、，S5)的特征通过高效混合编码器提取，并融合得到多尺度融合特征。

高效混合编码器由两个主要模块组成：基于注意力的尺度内特征交互模块(Attention-based Intra-scale Feature Interaction，AIFI)和基于 CNN 的跨尺度特征融合模块(CNN-based Cross-scale Feature Fusion，CCFF)。AIFI 在主干网络的最深层特征 S5 上使用单尺度 Transformer 编码器进行尺度内特征交互，在降低计算成本的同时保持了模型对高层语义信息的敏感性。CCFF 由多个由卷积层组成的融合块所构成，这些融合块的作用是将两个相邻的尺度特征融合成一个新的特征。其中每个融合块包含两个 1×1 的卷积层和 3 个由 RepConv 组成的 RepBlocks 层[18]。

解码器迭代优化对象查询，以生成预测框和类别。首先，通过一个 IOU 感知查询模块来选择最具代表性的图像特征，该模块从编码器输出的特征序列中筛选出一组特征。然后将这些特征作为初始查询，支撑解码器生成后续的预测结果。为了提升查询的初始质量与预测精度，RT-DETR 通过评估类别与定位预测的不确定性，优先选择不确定性较低的查询作为初始对象查询。在初始查询确定后，解码器会通过多次迭代优化步骤进一步调整预测结果，最终输出一组预测框及其对应的类别置信度。

2.2 多模态大模型技术

自 2022 年 11 月 OpenAI 公布了 ChatGPT 以来，大模型技术的发展迎来了爆发式的增长。在过去一年里，大模型技术不断突破边界，实现了从文本、图像到音频等多领域的自动化和智能化生成。以 GPT-4 为例，它可以根据图像生成不同类型的文本，如描述、解释、总结和问答等，也可以根据文本生成或编辑图像，完成创意和技术写作任务[19]。而在目标检测领域也出现了一些可以用于目标检测的多模态大模型。Google 团队提出的多模态大模型 PaliGemma[10]，基于 SigLIP-So400m 视觉编码器和 Gemma-2B 语言大模型构建，

被训练为一个多功能且具有广泛知识的基础模型，能够有效地迁移到各种下游任务中。Li 等[20]提出多模态大模型 GLIP，将传统的目标检测重构为一个定位任务，使模型能够基于文本查询检测物体，通过跨模态融合层深度整合视觉和文本信息，使模型能够同时学习语言感知和物体级别的表示，提升了短语定位和开放词汇检测等任务的表现。Xu 等[21]提出的首个支持视觉和文本查询的开放集多模态大模型 MQ-Det，在已有基于文本查询的检测大模型基础上加入了视觉示例查询功能，通过引入即插即用的门控感知结构，以及以视觉为条件的掩码语言预测训练机制，使得检测器在保持高泛化性的同时支持细粒度查询，提供了更灵活的选择来适应不同的场景。

随着越来越多的多模态大模型的出现，研究者也在研究更好地将这些大模型运用到下游任务中。大模型提示学习是一种常用的手段。Zhou 等[22]首次在预训练大模型中引入提示，提出 CoOp，通过优化其语言分支上的连续提示向量集对 CLIP 进行微调，以实现少样本迁移。Khattak 等[23]提出 Maple，进一步引入视觉提示，将文本和视觉提示合并到 CLIP 中，以改善文本和图像表示之间的对齐。

对于舰船目标检测与识别任务，与目前主流的小模型相比，大模型具有更多的参数和更深的层级结构，能够提取更丰富、更复杂的特征，从而提升下游任务的准确性和泛化性能，适应多样化任务和复杂场景。此外，大模型具有更丰富的训练数据，能够更好地泛化未见过的数据。这意味着即使在面对新的任务或领域时，大模型也能表现出良好的性能。因此，本文尝试在目标检测任务中引入多模态大模型生成提示，从而提高舰船目标检测的精度。

3 STFE-RTDETR 方法

3.1 总体框架

图 1 展示了基于小目标特征增强 RT-DETR 的 SAR 图像舰船目标检测方法（STFE-RTDETR）的网络结构图。从图 1 中可以看出，与 RT-DETR 模型相比，STFE-RTDETR 模型主要有以下 3 个改进：大模型提示生成网络（LMPGNet）、AIFI-EAA 模块和轻量化小目标特征增强融合网络（LSFF-Net）。

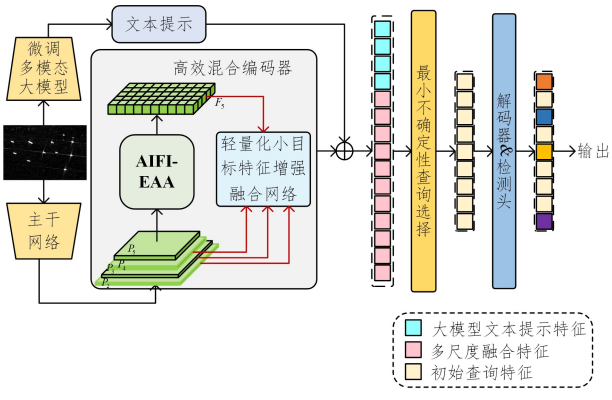

图 1 STFE-RTDETR 的网络结构

Fig. 1 Network structure of STFE-RTDETR

3.2 大模型提示生成网络

大模型提示生成网络由两阶段组成，分别是大模型微调与提示生成。所提方法采用多模态大模型 Paligemma 辅助目标检测。

大模型微调阶段：在训练样本上对多模态大模型 Paligemma 进行微调，微调架构如图 2 所示。采用前缀微调策略（Prefix Tuning），即在预训练的视觉语言模型输入序列前添加可训练、任务特定的前缀，从而实现针对不同任务的微调。前缀实际上是一种连续可微的虚拟标记，与离散的 Token 相比，它们更易于优化并且效果更佳。这种方法的优势在于不需要调整模型的所有权重，而是通过在输入中添加前缀来调整模型的行为，从而节省大量的计算资源，同时使得单一模型能够适应多种不同的任务。前缀可以是固定的（即手动设计的静态提示）或可训练的（即模型在训练过程中学习的动态提示）。在训练过程中，给定大模型提示前缀（Detect Ship），仅微调 Transformer 解码器中的 Transformer 层，冻结其他所有参数。最后得到微调后的多模态大模型 Paligemma*。前缀微调的具体步骤如下。

假设预训练多模态大模型的输入序列为 $X=[x_1,x_2,\cdots,x_n]\in\mathbb{R}^{n\times d}$，其中 d 为嵌入维度。前缀微调引入可训练参数 $P\in\mathbb{R}^{k\times d}$，其中 k 为前缀长度，构造新输入：

$$X'=[P;X]\in\mathbb{R}^{(k+n)\times d} \tag{1}$$

对于第 l 层 Transformer，键 $\boldsymbol{K}^{(l)}$ 和值 $\boldsymbol{V}^{(l)}$ 矩阵被重构为：

$$\boldsymbol{K}^{(l)}=[W_K^{(l)}P;W_K^{(l)}X],\boldsymbol{V}^{(l)}=[W_V^{(l)}P;W_V^{(l)}X] \tag{2}$$

其中，$W_K^{(l)}$ 和 $W_V^{(l)}$ 为原始预训练模型中的参数。前缀的引入实质是在注意力分布中增加可学习的偏置项，通过梯度下降调整 P 使模型关注任务相关的内容。

最后，冻结主网络参数 Θ，仅优化前缀参数 P 和指定 Transformer 层参数 Φ，目标函数为：

$$\min_{P,\Phi} L(Y,f_\Theta(X',\Phi)) \tag{3}$$

其中，Y 为舰船目标的真实标注，f_Θ 为预训练多模态大模型参数，L 为检测任务损失。

提示生成阶段：将 Paligemma* 融入 RT-DETR 模型中，用于提示生成。输入为舰船图像和提示前缀，输出为舰船预测坐标的提示文本特征。

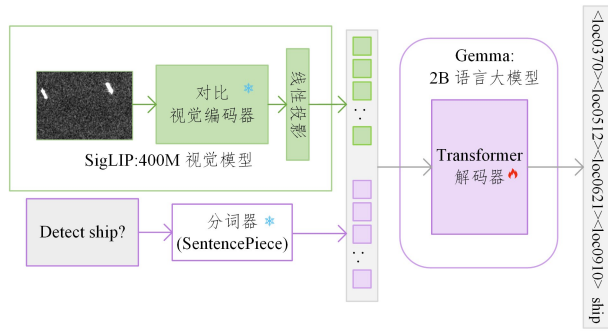

图 2 多模态大模型 Paligemma 的微调架构

Fig. 2 Fine-tuning architecture of multimodal large model Paligemma

3.3 AIFI-EAA 模块

AIFI-EAA 模块通过设计一种高效的加性注意力（EAA）来改进 RT-DETR 的尺度内特征交互模块（AIFI）。

与传统自注意力机制的点积注意力不同，典型的加性注意力通过逐元素乘法实现编码图像块之间的成对交互，从而

捕获全局上下文信息。它基于输入序列的上下文信息的3个注意力组成部分(查询$\boldsymbol{Q}$,键$\boldsymbol{K}$,值$\boldsymbol{V}$)的交互,对上下文信息的相关性分数进行编码。相比之下,本文提出的高效加性注意力机制可以在不牺牲性能的情况下去除键-值交互,仅纳入一个线性投影层来有效编码查询-键交互,就足以学习编码图像块之间的关系。高效加性注意力的网络结构如图3所示。

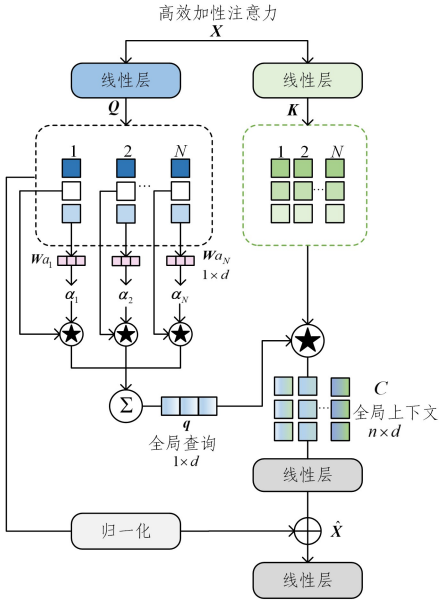

图3 高效加性注意力的网络结构

Fig. 3 Network structure of efficient additive attention

具体来说,输入嵌入矩阵$\boldsymbol{X}$通过两个矩阵$\boldsymbol{W}_q$和$\boldsymbol{W}_k$转换为查询矩阵$\boldsymbol{Q}$和键矩阵$\boldsymbol{K}$。其中$\boldsymbol{Q},\boldsymbol{K}\in\mathbb{R}^{n\times d}$,$\boldsymbol{W}_q,\boldsymbol{W}_k\in\mathbb{R}^{d\times d}$,$n$是编码图像块的长度,$d$是嵌入向量的维度。查询矩阵$\boldsymbol{Q}$乘以可学习参数向量$\boldsymbol{w}_\alpha\in\mathbb{R}^d$以学习查询的注意力权重,产生全局注意力查询向量$\boldsymbol{\alpha}\in\mathbb{R}^d$:

$$\boldsymbol{\alpha}=\boldsymbol{Q}\cdot\boldsymbol{w}_\alpha/\sqrt{d} \tag{4}$$

然后,根据学习到的注意权重对查询矩阵进行池化,得到单个全局查询向量$\boldsymbol{q}\in\mathbb{R}^d$:

$$\boldsymbol{q}=\sum_{i=1}^{n}\alpha_i*\boldsymbol{Q}_i \tag{5}$$

其中,$*$表示逐元素乘法运算。

接下来,对全局查询向量$\boldsymbol{q}\in\mathbb{R}^d$与键矩阵$\boldsymbol{K}\in\mathbb{R}^{n\times d}$之间的交互通过逐元素乘积编码,以形成全局上下文矩阵$\boldsymbol{C}\in\mathbb{R}^{n\times d}$。该矩阵与多头自注意力中的注意力矩阵作用相似,可以捕获每个编码图像块的信息,并灵活地学习输入序列中的相关性。然而,与多头自注意力相比,它的计算成本相对较低,其复杂度与编码图像块长度呈线性关系。受Transformer架构的启发,本方法采用线性变换层来处理查询-键交互,以学习编码图像块的隐藏表示。高效加性注意力的输出$\hat{X}$可以描述为:

$$\hat{X}=\hat{\boldsymbol{Q}}+T(K*\boldsymbol{q}) \tag{6}$$

其中,$\hat{\boldsymbol{Q}}$表示归一化查询矩阵,T表示线性变换。

AIFI-EAA模块一方面摒弃传统自注意力机制中的键-值交互分支,仅保留查询-键交互,通过引入线性投影层,降低了查询-键交互的复杂度;另一方面,通过逐元素乘法操作替代传统自注意力中的矩阵乘法操作,大大减少了计算资源消耗和延迟,从而在保持精度的同时,显著提升了模型的推理速度。

3.4 轻量化小目标特征增强融合网络

传统的检测层(如P_3,P_4和P_5)在有效解决SAR图像中的舰船小目标检测问题方面面临挑战。增强舰船小目标检测能力的常用方法是引入P_2检测层,但这会带来一系列问题,例如增加计算量和延长后处理时间。因此,迫切需要开发新颖的多尺度特征融合结构来有效应对舰船目标检测的挑战。本文设计了一种轻量化小目标特征增强融合网络来解决这个问题,其网络结构如图4所示。该网络在CCFF的基础上,将包含丰富小目标信息的P_2特征层经过SPDConv处理后与P_3层进行整合。整合部分借鉴CSP思想和基于OmniKernel的方法,提出了CSP-OmniKernel模块进行特征融合。

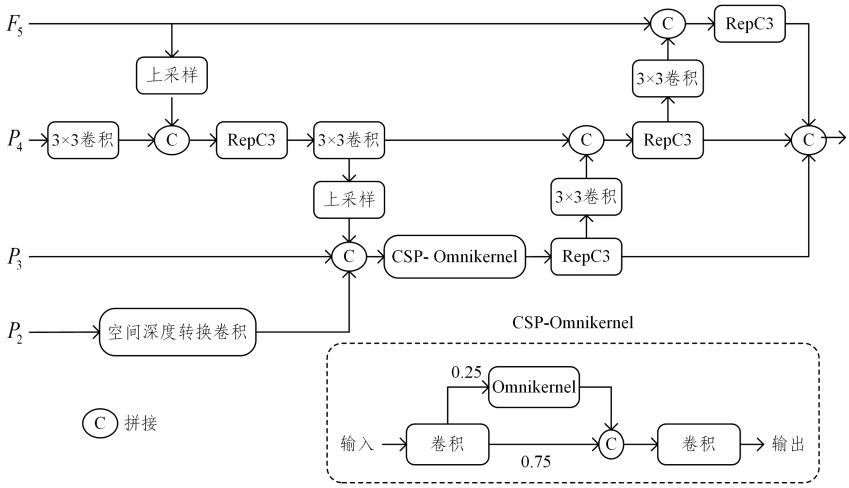

图4 轻量化小目标特征增强融合网络的网络结构

Fig. 4 Network structure of LSFFNet

3.4.1 空间深度转换卷积

空间深度转换卷积(SPD-Conv)是一种专门为改善低分辨率图像和小目标检测而设计的CNN模块。它通过替代传统步长卷积和池化操作,降低了信息损失,并加强了对细节特征的捕获,使模型在处理小目标和低分辨率图像时表现更佳。

具体来说,SPD-Conv由空间到深度(Space-to-Depth,

SPD)层和非跨步卷积层两部分组成。首先，SPD 层通过对大小为 $S\times S\times C$ 的中间特征图进行切片，生成一系列子特征图 $f_{x,y}$：

$$\begin{cases} f_{0,0}=X[0:S:scale,0:S:scale], \\ f_{1,0}=X[1:S:scale,0:S:scale],\cdots, \\ f_{scale-1,0}=X[scale-1:S:scale,0:S:scale]; \\ f_{0,1}=X[0:S:scale,1:S:scale],f_{1,1},\cdots, \\ f_{scale-1,1}=X[scale-1:S:scale,1:S:scale], \\ \quad\vdots \\ f_{0,scale-1}=X[0:S:scale,scale-1:S:scale],f_{1,scale-1},\cdots, \\ f_{scale-1,scale-1}=X[scale-1:S:scale,scale-1:S:scale] \end{cases} \tag{7}$$

其中，*scale* 为缩放因子，每个子图通过一定的缩放因子对原特征图进行下采样。为了避免缩放因子过大而导致特征损失和通道爆炸，本文选择缩放因子为 2，同时平衡信息保留与计算效率。

接下来，所有子特征图沿着通道维度连接以创建特征图 X，这一过程不仅减少了空间维度，而且增加了通道维度。

最后，通过一个非跨步卷积层来进一步处理经过 SPD 层变换后的特征图 X，该层使用可学习的参数来减少通道的数量，以防止增加的通道导致的信息冗余，同时尽可能保留判别性的特征信息。

3.4.2　CSP-OmniKernel 模块

本文设计的 CSP-OmniKernel 模块遵循 CSP 结构思想。CSP 结构的核心思想是将输入特征图分成两部分，一部分经过一个小的卷积网络进行处理，另一部分则直接进行下一层的处理。将两部分特征图拼接起来，作为下一层的输入。

如图 4 所示，通过卷积层将输入特征图划分为 4 个输入通道切片，其中只有一个通道由 OmniKernel 网络处理，将经过处理的特征图与其余 3 个通道特征图进行拼接，最后经过一次卷积操作调整通道数。这样可以显著地减少网络的参数和计算量，同时提高特征提取的效率，从而加快模型的训练和推理速度。

OmniKernel 网络由 3 个分支组成，即全局分支、大核分支和局部分支，旨在有效地学习从全局到局部的特征表征，提升小目标的检测性能。OmniKernel 网络结构如图 5 所示。

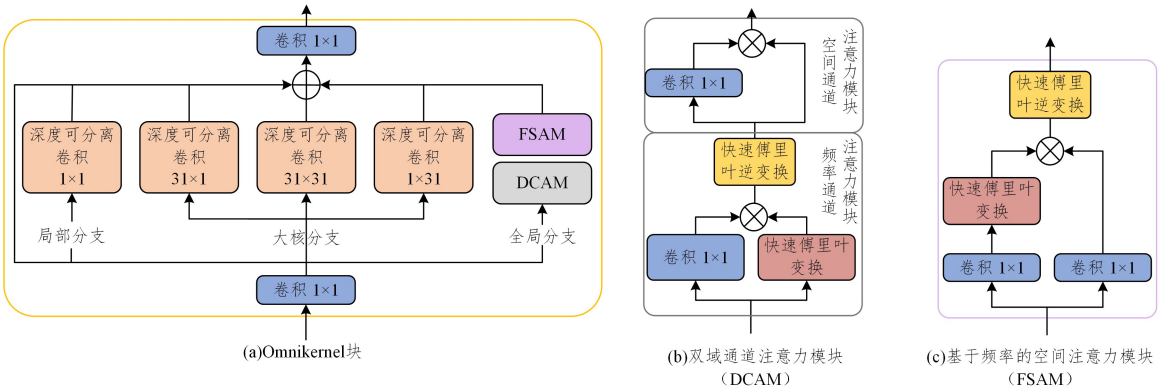

图 5　OmniKernel 模块的网络结构

Fig. 5　Network structure of OmniKernel module

在大核分支中，输入 X_{Large} 通过 3 个通道分别进行 3 次卷积操作：一次 31×31 深度可分离卷积，用于获取较大的感受野；31×1 和 1×31 的条形深度可分离卷积各一次，用于获取条形的上下文信息。

在全局分支的推理阶段，输入图像比训练图像大得多，因此 31×31 的内核无法覆盖全局域。为了缓解这个问题，本方法采用了双域处理，为全局分支增加了全局建模能力。具体来说，全局分支包括双域通道注意力模块（Dual-Domain Channel Attention Module，DCAM）和基于频率的空间注意力模块（Frequency-Based Spatial Attention Module，FSAM），通过双域处理增强了全局建模能力。

给定输入特征 X_{Global}，双域通道注意力模块首先对 X_{Global} 应用频率通道注意力（Frequency Channel Attention，FCA）：

$$X_{\text{FCA}}=\mathcal{IF}(\mathcal{F}(X_{\text{Global}})\otimes W_{1\times1}(GAP(X_{\text{Global}})) \tag{8}$$

其中，$\mathcal{F}$ 和 $\mathcal{IF}$ 分别是快速傅里叶变换及其逆变换；X_{FCA}，$W_{1\times1}$ 和 GAP 分别表示频率通道注意力模块的输出、1×1 卷积层和全局平均池化；$\otimes$ 表示逐元素乘法。

通过傅里叶变换处理后，全局特征在频谱域中经过全局调制后的特征进一步输入空间通道注意力模块（Spatial Channel Attention，SCA）：

$$X_{\text{DCAM}}=X_{\text{FCA}}\otimes W_{1\times1}(GAP(X_{\text{FCA}})) \tag{9}$$

其中，X_{DCAM} 是双域通道注意力模块的输出。双域通道注意力模块仅在通道粗粒度上增强双域特征。

随后，X_{DCAM} 输入至基于频率的空间注意力模块，在空间维度上进一步细化频谱：

$$\begin{aligned} X_{\text{FSAM}}&=\mathcal{IF}(\mathcal{F}((W_{1\times1}(X_{\text{DCAM}}))\otimes W\ 1\times1(X_{\text{DCAM}}) \\ X_{\text{FSAM}}&=\mathcal{IF}(((\mathcal{F}((W_{1\times1}(X_{\text{DCAM}}))\otimes W_{1\times1}(X_{\text{DCAM}}) \end{aligned} \tag{10}$$

其中，X_{FSAM} 是基于频率的空间注意力模块的输出。

局部分支仅将输入 X_{Local} 通过一个 1×1 的深度可分离卷积，用于局部特征的调制。

4　实验及结果分析

4.1　数据集

本文分别使用了 3 个广泛使用的公开 SAR 舰船图像数据集，即 SSDD[13]，HRSID[14] 和 SAR-Ship-Dataset[15] 进行实验。

SSDD 数据集：SSDD 数据集是第一个用于舰船检测的公开 SAR 遥感图像数据集，涵盖了多种不同环境下的舰船图

像，包括不同分辨率、尺寸大小、海况以及传感器类型。该数据集总共包含1160张图像和2456个舰船实例。

HRSID数据集：HRSID数据集由136张全景SAR图像的切片构成，包含5604张切片图像和16951个舰船实例。图像尺寸为800×800像素，分辨率为0.5～3m。该数据集涵盖了不同极化方式、海况、海域和沿海港口的SAR图像。

SAR-Ship-Dataset数据集：SAR-Ship-Dataset数据集由102张“高分三号”SAR图像和108张“哨兵一号”SAR图像的切片构成，包含43819张切片图像和59535个舰船实例。图像尺寸为256×256像素，分辨率为3～25m。

4.2 实验设置

本文基于PyTorch深度学习框架，所有实验均在64位Linux操作系统Ubuntu 20.04上进行，CPU型号为Intel(R) Xeon(R) Platinum 8368，GPU型号为4×NVIDIA A800 80GB。在实验中，使用AdamW优化器执行优化，参数设置如下：初始学习率为0.000 1，权重衰减为0.000 1，批量大小为32。输入图像大小为640×640像素，模型训练300个epoch。按照官方设定，将SSDD和SAR-Ship-Dataset数据集按照8∶2划分为训练集和测试集，将HRSID数据集按照6.5∶3.5划分为训练集和测试集。

4.3 对比实验与实验结果分析

为了验证所提方法的有效性，将其与该领域的常见目标检测算法在SSDD数据集和HRSID数据集上进行对比实验，并对结果进行分析。各检测算法在SSDD数据集和HRSID数据集上的检测结果如表1所列。

表1 所提算法和其他算法在SSDD数据集和HRSID数据集上的实验结果

Table 1 Experimental results of proposed algorithm and other algorithms on SSDD and HRSID

(%)

数据集	对比方法	主干网络	AP_{50}	$AP_{50\text{-}95}$
SSDD	Faster R-CNN[3]	ResNet50	95.7	61.1
	GLIP[19]	SwinTransformer-T	96.5	62.9
	DINO[24]	ResNet50	95.0	60.3
	Deformable DETR[25]	ResNet50	93.0	52.5
	CO-DETR[26]	ResNet50	96.5	62.0
	YOLOv8-L[27]	ResNet50	96.7	64.1
	ELLK-Net[28]	ALKS-Net	95.6	63.9
	RT-DETR[6]	ResNet50	96.8	64.3
	STFE-RTDETR	ResNet50	**97.2**	**64.4**
HRSID	Faster R-CNN[3]	ResNet50	82.3	57.6
	GLIP[19]	SwinTransformer-T	82.1	62.6
	DINO[24]	ResNet50	90.3	64.4
	Deformable DETR[25]	ResNet50	82.3	53.9
	CO-DETR[26]	ResNet50	92.1	70.2
	YOLOv8-L[27]	ResNet50	91.5	70.3
	ELLK-Net[28]	ALKS-Net	90.6	66.8
	RT-DETR[6]	ResNet50	92.7	69.5
	STFE-RTDETR	ResNet50	**93.7**	**70.3**

本文采用平均精确率(Average Precision，AP)全面评估算法的检测性能。

$$P=\frac{TP}{TP+FP}\times 100\% \tag{11}$$

$$R=\frac{TP}{TP+FN}\times 100\% \tag{12}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{13}$$

其中，P是精确率(Precision)、R是召回率(Recall)，TP是正确检测出的舰船目标数量，FP是被错误检测成舰船目标的数量，FN是舰船目标被漏检的数量。计算$IoU=0.50$时的AP，在表1中记为AP_{50}；计算在不同IoU阈值(从0.50到0.95，步长0.05)上的平均AP，在表1中记为$AP_{50\text{-}95}$。

如表1所列，作为基线模型的RT-DETR表现出色，在SSDD数据集上的AP_{50}为96.8%，$AP_{50\text{-}95}$为64.3%，在HRSID数据集上的AP_{50}为92.7%，$AP_{50\text{-}95}$为69.5%。相比之下，Faster R-CNN和DINO等经典方法在这些指标上的结果较差。YOLOv8表现出很高的性能，但仍低于基线模型RT-DETR。另一种DETR系列的新兴方法CO-DETR也取得了较高的检测精度，在SSDD数据集和HRSID数据集上的AP_{50}分别为96.5%和92.1%，$AP_{50\text{-}95}$分别为62.0%和70.2%。本文方法STFE-RTDETR在所有关键指标上都表现出卓越的整体性能，优于所有其他模型：在SSDD数据集上的AP_{50}达到97.2%，$AP_{50\text{-}95}$达到64.4%；在HRSID数据集上的AP_{50}达到93.7%，$AP_{50\text{-}95}$达到70.3%。这凸显了本文方法在舰船目标检测任务中的显著优势。

此外，还将本文算法与先进算法在大规模SAR舰船数据集SAR-Ship-Dataset上进行了对比实验，结果如表2所列。实验表明，本文算法STFE-RTDETR在AP_{50}和$AP_{50\text{-}95}$两个性能指标上均高于基线方法RT-DETR和YOLO系列算法YOLOv8-L。

表2 所提算法和其他算法在SAR-Ship-Dataset上的实验结果

Table 2 Experimental results of proposed algorithm and other algorithms on SAR-Ship-Dataset

(%)

对比方法	AP_{50}	$AP_{50\text{-}95}$
YOLOv8-L[27]	96.5	65.6
RT-DETR[6]	96.9	65.1
STFE-RTDETR	**97.6**	**67.1**

表3对各个算法的模型参数量、计算算法的浮点运算次数(FLOPs)、图像处理帧率(FPS)进行了对比。其中，模型参数量表示每个模型的空间复杂度，FLOPs衡量算法的计算时间复杂度，FPS衡量算法的实际运行速度。由表3可见，STFE-RTDETR在所有对比方法中的模型参数量最小，在同一批次大小下的FLOPs达到最小，FPS仅低于基线模型RT-DETR，因此STFE-RTDETR具有较低的时间成本和内存成本，这充分证明了所提出网络的快速性和轻便性。

表3 各方法的模型参数量、浮点运算次数和图像处理帧率

Table 3 Model parameters, FLOPs and FPS of each methods

对比方法	参数量	FLOPs	FPS
Faster R-CNN[3]	4.135×10^7	1.935×10^{12}	17
DINO[24]	4.739×10^7	2.478×10^{12}	23
Deformable DETR[25]	4.010×10^7	1.812×10^{12}	15
CO-DETR[26]	6.418×10^7	1.950×10^{12}	13
YOLOv8-L[27]	4.361×10^7	1.648×10^{12}	71
RT-DETR[6]	4.196×10^7	1.360×10^{12}	**108**
STFE-RTDETR	$\mathbf{3.551\times10^7}$	$\mathbf{1.298\times10^{12}}$	103

4.4 消融实验

本文通过在 HRSID 数据集上进行消融实验，研究了各种改进对舰船目标检测性能的影响。对基线模型 RT-DETR 的改进增强进行了评估，包括大模型提示生成网络（LMPGNet）、轻量化小目标特征增强融合网络（LSFFNet）和 AIFI-EAA 模块。

表 4 列出了不同改进模块在 HRSID 数据集上的消融实验结果。基线模型 RT-DETR 的 AP_{50} 达到 92.7%，$AP_{50\text{-}95}$ 达到 69.5%；本文方法 STFE-RTDETR 的 AP_{50} 达到 93.7%，$AP_{50\text{-}95}$ 达到 70.3%。其中，AP_{50} 提高 1 个百分点，$AP_{50\text{-}95}$ 提高 0.8 个百分点。实验结果证明，本文设计的大模型提示生成网络、轻量化小目标特征增强融合网络和 AIFI-EAA 模块均对精度提升有一定作用。

表 4 HRSID 数据集上的消融实验结果

Table 4 Ablation experiment results on HRSID

基线模型	＋LMPGNet	＋LSFFNet	＋AIFI-EAA	AP_{50}/%	$AP_{50\text{-}95}$/%
√				92.7	69.5
√	√			93.2	69.7
√	√	√		93.4	70.2
√	√	√	√	**93.7**	**70.3**

结束语 本文提出了一种基于小目标特征增强 RT-DETR 的 SAR 图像舰船目标检测方法。该方法主要由以下 3 部分组成。1）大模型提示生成网络：借助多模态大模型 Paligemma 的零样本学习能力，生成提示，以提取图像模态中更具判别性的信息。2）AIFI-EAA 模块：以 RT-DETR 为基线，改进尺度内特征交互模块，引入高效加性注意力机制来替代传统自注意力机制的二次方矩阵运算，降低算法计算复杂度。3）轻量化小目标特征增强融合网络：在多尺度特征融合网络中加入小目标检测层，设计 CSP-OmniKernel 模块进行特征融合，提升小目标的检测性能。在 3 个广泛应用的数据集上进行的综合实验，证明了所提方法在准确性上具有优势。

参 考 文 献

[1] CUI Z，QUAN H，CAO Z，et al. Sar target cfar detection via gpu oarallel operation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2018，11(12)：4884-4894.

[2] LIN H，LIU J，LI X，et al. Dcea：Detr with concentrated deformable attention for end-to-end ship detection in sar images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，2024，17：17292-17307.

[3] REN S，HE K，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，2016，39(6)：1137-1149.

[4] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-time object detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016：779-788.

[5] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end object detection with transformers[C]// European Conference on Computer Vision. 2020：213-229.

[6] ZHAO Y，LV W，XU S，et al. Detrs beat yolos on real-time object detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024：16965-16974.

[7] LIU L，FU L，ZHANG Y，et al. Clfr-det：Cross-level feature refinement detector for tiny-ship detection in sar images[J]. Knowledge-Based Systems，2024，284：111284.

[8] CAI X，LAI Q，WANG Y，et al. Poly kernel inception network for remote sensing detection[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024：27706-27716.

[9] QIN C，WANG X，LIU Y，et al. A novel end-to-end transformer network for small scale ship detection in sar images[C]// International Geoscience and Remote Sensing Symposium. 2024：8158-8162.

[10] BEYER L，STEINER A，PINTO A S，et al. Paligemma：A versatile 3b vlm for transfer[J]. arXiv：2407.07726，2024.

[11] SHAKER A，MAAZ M，RASHEED H，et al. Swiftformer：Efficient additive attention for transformer-based real-time mobile vision applications[C]// IEEE/CVF International Conference on Computer Vision. 2023：17425-17436.

[12] CUI Y，REN W，KNOLL A. Omni-kernel network for image restoration[C]// AAAI Conference on Artificial Intelligence. 2024：1426-1434.

[13] LI J，QU C，SHAO J. Ship detection in sar images based on an improved faster r-cnn[C]// SAR in Big Data Era：Models，Methods and Applications. 2017：1-6.

[14] WEI S，ZENG X，QU Q，et al. HRSID：A high-resolution sar images dataset for ship detection and instance segmentation[J]. IEEE Access，2020，8：120234-120254.

[15] WANG Y，WANG C，ZHANG H，et al. A sar dataset of ship detection for deep learning under complex backgrounds[J]. Remote Sensing，2019，11(7)：765.

[16] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016：770-778.

[17] CHEN J，LEI B，SONG Q，et al. A hierarchical graph network for 3d object detection on point clouds[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020：392-401.

[18] DING X，ZHANG X，MA N，et al. Repvgg：Making vgg-style convnets great again[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021：13733-13742.

[19] YAN H，LIU Y L，JIN L W，et al. The development，application，and future of llm similar to chatgpt[J]. Journal of Image and Graphics，2023，28(9)：2749-2762.

[20] LI L H，ZHANG P，ZHANG H，et al. Grounded language-image pre-training[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022：10965-10975.

[21] XU Y，ZHANG M，FU C，et al. Multi-modal queried object detection in the wild[C]// Advances in Neural Information Processing Systems. 2024：1-18.

[22] ZHOU K，YANG J，LOY C C，et al. Learning to prompt for vi-

sion-language models[J]. International Journal of Computer Vision, 2022, 130(9): 2337-2348.
[23] KHATTAK M U, RASHEED H, MAAZ M, et al. Maple: Multi-modal prompt learning[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 19113-19122.
[24] ZHANG H, LI F, LIU S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[J]. arXiv: 2203.03605, 2022.
[25] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection[C]// International Conference on Learning Representations. 2021: 1-16.
[26] ZONG Z, SONG G, LIU Y. Detrs with collaborative hybrid assignments training[C]// IEEE/CVF International Conference on Computer Vision. 2023: 6748-6758.
[27] JOCHER G, NISHIMURA K, MINEEVA T, et al. Yolov8 by ultralytics [EB/OL]. https://github.com/ultralytics/ultralytics.
[28] SHEN J, BAI L, ZHANG Y, et al. Ellk-net: An efficient lightweight large kernel network for sar ship detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5221514.

ZHANG Hongsen, born in 2001, postgraduate. His main research interest is pattern recognition.

WU Fei, born in 1989, Ph.D, professor, is a member of CCF(No. 86938S). His main research interests include pattern recognition and machine learning.

（责任编辑：柯颖）

CCF 访问英国 IET 总部

近日，CCF 副秘书长王新霞和 CCF 理事、北京赛博英杰科技有限公司董事长 &CEO 谭晓生一行 2 人，访问了英国 IET 总部，与 IET 首席增长官（Chief Engagement & Growth Officer）Toni Allen 和国际商务经理（International Business Manager）James Howe 举行了会谈。

CCF 就组织历史、发展历程和现状进行了简单介绍，IET 带领 CCF 访问人员参观了历史悠久的 IET 大楼。会谈期间，双方就 CNCC 合作、工程师能力培养和认证以及联合举办研讨会等议题深入交换了意见，并就可能的合作方向和模式进行了深入探讨。按照去年 6 月双方首次会谈后签订的备忘录条款，CCF 特别邀请 IET 负责人参加将于 10 月 23 日至 25 日在哈尔滨举办的中国计算机大会（CNCC2025）。IET 表示他们很期待，并将尽快协调确认相关信息。

从左到右：James Howe　Toni Allen　王新霞　谭晓生

据 CCF 微信公众号