

基于多注意力特征融合的SAR图像目标分类算法

许丽龙¹, 王春柳¹, 侯宇超², 王鹏¹

(1. 中北大学 数学学院, 山西 太原 030051; 2. 山西师范大学 数学与计算机科学学院, 山西 太原 030031)

摘要: 针对SAR图像目标分类问题提出了一种新的多注意力特征融合网络(Multi-Attention Feature Fusion Network, MAFNet)分类模型。该模型首先对初始图像使用多头自注意力机制来捕获全局信息;其次,通过引入协方差注意力机制进一步增强通道特征和空间特征的表达;然后引入浅层鲁棒特征下采样模块,以更高效地提取原始图像的有效信息;最后对前述获得的三个注意力特征进行融合,获取了更有表征能力的SAR图像特征。本文模型克服了传统卷积神经网络仅在局部范围进行特征提取的问题,对深层特征进行通道和空间两个维度的信息筛选和增强,同时融合包含全局信息的特征,有效提升了模型的分类准确性和鲁棒性。在SOC条件下的MSTAR数据集上的实验结果表明,MAFNet的分类准确率达99.96%,明显优于其他算法。

关键词: 雷达图像目标分类; MSTAR数据集; 注意力机制; 特征提取

中图分类号: TP751 **文献标识码:** A **doi:** 10.62756/jnuc.issn.1673-3193.2025.03.0012

引用格式: 许丽龙, 王春柳, 侯宇超, 等. 基于多注意力特征融合的SAR图像目标分类算法[J]. 中北大学学报(自然科学版), 2025, 46(5): 561-573.

XU Lilong, WANG Chunliu, HOU Yuchao, et al. Target classification algorithm for sar images based on multi-attention feature fusion[J]. Journal of North University of China(Natural Science Edition), 2025, 46(5): 561-573.

Target Classification Algorithm for SAR Images Based on Multi-Attention Feature Fusion

XU Lilong¹, WANG Chunliu¹, HOU Yuchao², WANG Peng¹

(1. School of Mathematics, North University of China, Taiyuan 030051, China;

2. School of Mathematics and Computer Science, Shanxi Normal University, Taiyuan 030031, China)

Abstract: A novel classification model for SAR image target classification, named Multi-Attention Feature Fusion Network (MAFNet) was proposed. Firstly, a multi-head self-attention mechanism was applied to the original image to capture global information. Secondly, a covariance attention mechanism was introduced to further enhance the representation of channel and spatial features. Thirdly, a shallow robust feature down-sampling module was incorporated to more efficiently extract effective information from the raw image. Finally, the three attention-based features were fused to obtain more representative SAR image features. This approach overcomes the limitation of traditional convolutional neural networks, which only extract features within a local receptive field. By enhancing deep features in both channel and spatial dimensions and integrating features containing global information, the model significantly improves classification accuracy and robustness. Experimental results on the MSTAR dataset under the SOC condition show that MAFNet achieves a clas-

收稿日期: 2025-03-24

基金项目: 山西省留学回国人员科技活动择优资助项目(20240011); 山西省基础研究计划项目(202303021212164, 202103021224195, 202103021224212); 山西省回国留学人员科研项目(2021-108)

作者简介: 许丽龙(2000-), 女, 硕士生, 主要从事深度学习和图像分类的研究。

通信作者: 王鹏(1977-), 男, 教授, 博士, 主要从事智能信息处理的研究。E-mail: wpmath@nuc.edu.cn.

sification accuracy of 99.96%, outperforming other existing algorithms.

Key words: SAR image classification; MSTAR dataset; attention mechanism; feature extraction

0 引言

合成孔径雷达(Synthetic Aperture Radar, SAR)作为一种主动微波成像传感器,能够不受光照条件的限制全天候获取目标的结构、大小以及散射信息,可拍摄高标称分辨率图像,在地形测绘、地质勘探、海洋观测等领域应用广泛,尤其在国防安全、农业估算和地质勘测中作用突出。合成孔径雷达自动目标识别(Synthetic Aperture Radar Automatic Target Recognition, SAR ATR)是SAR图像解译的关键环节,它可以对输入的目标区域进行类别判断,在民用和军事等众多领域都有着极其重要的应用价值。正因为如此,SAR ATR受到了世界各国的高度关注,取得了丰硕的研究成果。

SAR ATR方法主要分为基于模板的方法^[1]、基于模型的方法^[2]和基于机器学习^[3-12]的方法。基于模板的方法通过匹配待分类图像与预定义模板确定类别,计算匹配度(如欧几里得距离),但受成像条件的影响较大,泛化能力弱,计算复杂。基于模型的方法利用先验知识构建目标或场景的数学模型,并通过匹配分类,具有较强的解释性,但依赖先验知识,计算资源需求高。

随着支持向量机(Support Vector Machine, SVM)等稳健分类器的发展,机器学习范式逐渐被应用于SAR ATR任务^[3]。近年来,深度卷积神经网络(Deep Convolutional Neural Network, DCNN)凭借其强大的特征提取能力在图像识别领域取得显著突破,并逐渐成为SAR图像目标分类的主流架构^[13-19]。然而,SAR图像对姿态变化敏感且存在相干斑噪声,其获取成本高且标注困难(尤其在非合作目标场景下难以获取大量标注样本),导致DCNN在实际应用中仍面临数据不足、过拟合及对复杂成像条件鲁棒性不足等挑战。针对过拟合与数据不足的问题,Chen等^[4]设计了全卷积网络(A-ConvNets)架构,通过稀疏连接层替代传统卷积网络中的全连接层;Shang等^[5]提出深度记忆卷积神经网络(M-Net),为解决训练问题通过转移参数技术分两步进行训练,通过添加信息记录器来记忆样本空间特征并利用其预测未知样本标签;Deng等^[6]提出基于视觉转换器(Vision Transformer, ViT)模型的自监督对比学习

模型(Sup-ConL),在预训练阶段通过裁剪、旋转、高斯噪声和乘性噪声等SAR特异性增强策略模拟姿态变化与噪声干扰,提升特征鲁棒性,微调阶段结合对比损失与交叉熵损失,融合标签信息与语义信息,优化了分类能力;Fu等^[7]采用深度残差网络,引入dropout层缓解过拟合,将新的损失函数center loss与softmax loss结合来训练网络;Wilmanski等^[8]探讨了应用于SAR ATR的权重初始化、随即池化、自适应优化算法等现代CNN训练方法,通过对训练方法的改进,在小样本的MSTAR数据集上实现了比以往方法更高的分类准确率。Sun等^[9]设计了SOIF-CycleGAN网络,通过SAR光学图像双向翻译生成人工数据来解决特征提取难和数据集匮乏的问题,同时解决了相干斑噪声的问题。针对姿态变化敏感问题,Banas等^[10]将自然图像领域的3D图像识别技术应用于SAR,并采用了多视图卷积神经网络(MVCNN)框架,结合了A-ConvNets骨干网络及针对SAR数据的预处理和增强方法,显著提高了SAR ATR性能;Ren等^[11]提出类导向分层表示混合推理网络(HRNCHR)并将其应用于少样本SAR目标识别,通过局部-全局信息并发细化机制(LGICRM)来提取多样特征,在双曲空间建立层级表示空间以学习类别间的层级关系,采用混合推理策略来融合实例和原型推理结果。针对多视图自动目标识别(Automatic Target Recognition, ATR)方法在输入图像数量变化时面临结构复杂和计算量大的问题,Zhang等^[12]提出了基于图像注意力的自适应多视图融合网络(IA-AMF-Net),通过深度可分离卷积提取特征,利用挤压与激励操作生成图像注意力权重融合特征,增强对目标分类特征的关注,且融合特征维度与输入图像数量无关。这些算法都是针对SAR图像某些特点设计的,并没有完全解决SAR图像目标分类存在的问题。

以上算法均在SAR图像目标分类领域取得了显著成效,但仍存在一些不足。传统卷积神经网络(Convolutional Neural Network, CNN)仅在局部范围进行特征提取,无法建模远距离依赖关系。针对这个问题,不少网络引入了注意力机制来实现通道和空间维度的信息交互。例如,孙靖森等^[13]提出的ICBAM_CNN在传统CNN中融合了通道与空间注意力以增强全局信息捕捉,但它仍难以建模长距离

依赖关系。ViT 模型的自注意力机制可以直接捕捉全局信息,能建模长距离依赖关系^[20],但它不像 CNN 那样在小数据集上也具有良好的泛化能力,ViT 依赖大规模数据训练,否则容易过拟合,且缺乏 CNN 的局部特征提取能力,收敛速度较慢,需要更长的训练时间,需要大量正则化技巧才能稳定训练。因此,本文提出多注意力特征融合网络(Multi-Attention Feature Fusion Network, MAFNet),引入浅层鲁棒特征下采样模块来处理初始图像,融合协方差注意力机制和自注意力机制,在增强局部信息提取的同时捕捉全局特征,从而保证分类性能。

1 多注意力特征融合网络

尽管当前 SAR 图像的标称分辨率已显著提升,但其数值仍普遍低于高分辨率光学图像。更为关键的是, SAR 图像与光学数据存在本质特性差异: 1) SAR 成像依赖主动微波散射机制,导致图像受斑点噪声干扰且地物特征由散射特性主导(如金属目标强反射、水面弱反射); 2) 侧视成像几何引发叠掩、阴影等畸变,进一步增加了目标解译的复杂性。因此,在下采样时容易受干扰信息影响,而且传统 CNN 局部特征堆砌导致整体语义碎片化,不能很好地关注整体特征。为应对以上挑战,本文提出了一种融合多种注意力特征的分类模型——MAFNet,其结构如图 1 所示。当初始图像输入网络时,往往包含大量的信息,其中不仅有重要的目标特征信息,还存在散斑噪声以及复杂背景环境所造成的干扰信息。本文提出的网络 MAFNet 一方面基于视觉转换器(Vision Transformer, ViT)模型捕获 SAR 图像全局信息,获得包含长距离依赖关系的特征图(维度为 $256 \times 8 \times 8$,即通道数 \times 高度 \times 宽度,其中高度和宽度的单位为像素);另一方面采用浅层鲁棒特征下采样模块(Shallow Robust Feature Downsampling, SRFD)进行下采样,在抑制原始图像中冗余信息的同时有选择地保留显著特征,得到浅层特征图后再分别经过 3×3 卷积层、协方差通道注意力模块、协方差空间注意力模块获得深层特征图($256 \times 8 \times 8$)、协方差通道注意力($256 \times 1 \times 1$)、协方差空间注意力($1 \times 8 \times 8$),然后将深层特征图分别与通道注意力和空间注意力相乘,实现通道和空间两个维度的信息筛选和增强,得到通道信息增强的协方差通道注意力特征图和空间信息增

强的协方差空间注意力特征图。最后,将三种注意力特征图融合获得包含多维度注意力信息的多注意力融合特征图,通过平均池化层对该特征图进行降维后展平(flatten),再输入到全连接层,最终输出 SAR 图像目标的分类结果。

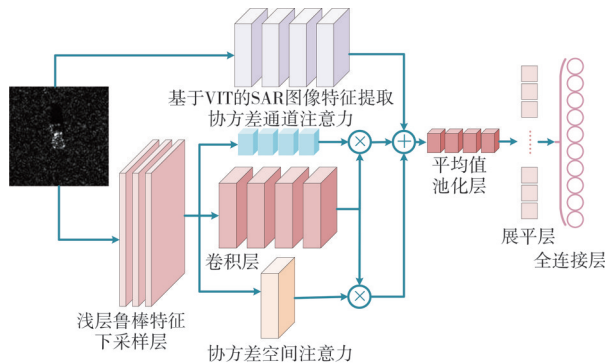


图 1 多注意力特征融合网络

Fig. 1 Multi-attention feature fusion network

SRFD 模块和 3×3 卷积层用于提取局部细节,提供高效的局部基础特征,但通道间信息仅通过卷积核隐式学习,缺乏对关键散射通道的显式筛选,且局部感受野限制了对 SAR 目标空间分布模式的全局捕捉。协方差通道注意力模块通过建模通道间的二阶统计信息来增强 SAR 图像目标关键散射通道的表征能力,弥补了 CNN 隐式通道交互的不足;协方差空间注意力模块通过建模空间位置间的二阶统计信息来增强 SAR 图像核心区域的特征表达,弥补了 CNN 局部感受野对空间全局分布建模的不足。尽管协方差注意力模块优化了通道与空间特征,但仍未完全解决长距离依赖问题,因此在引入基于 ViT 的 SAR 图像特征提取模块后,通过多头自注意力机制(Multi-Head Self-Attention, MHSA)显式捕捉任意位置间的长距离依赖,可以打破 CNN 局部感受野的限制,增强模型对 SAR 图像中多尺度目标的表征能力。最后,融合三种注意力特征图,使得 MAFNet 既兼具 CNN 的高效局部特征提取和噪声抑制能力,又融合了 ViT 对长距离依赖关系与全局信息的建模优势,能够在噪声抑制、特征表达能力和鲁棒性上实现多重优势互补,形成“局部-全局-通道-空间”多维度的特征表达,为 SAR 图像分类任务提供了更强的性能保障。

1.1 基于 ViT 的 SAR 图像特征提取

视觉转换器(Vision Transformer, ViT)模型是最早将多头自注意力机制(Multi-Head Self-Attention, MHSA)应用于计算机视觉的模型,在

对图像进行分块处理和位置编码后,通过多头自注意力机制实现全局信息的直接建模。为解决传统CNN在捕捉长距离依赖和全局信息上的不足,本文借鉴ViT的思路,通过MHSA来提取SAR图像的整体特征。基于ViT的SAR图像特征提取流程如图2所示,首先将输入的SAR图像划分为64个块(patch),并将每个块投影为一个256大小的一维特征向量,再加入位置编码保留位置信息。然后通过多头自注意力模块捕捉图像中不同子空间的上下文关联,直接建模图像全局信息。将最终得到的特征序列重塑为 $256 \times 8 \times 8$ 大小的特征图,其中既包含了多维度特征信息,也保留了长距离依赖关系,为后续的特征融合与分类提供了丰富的语义基础。

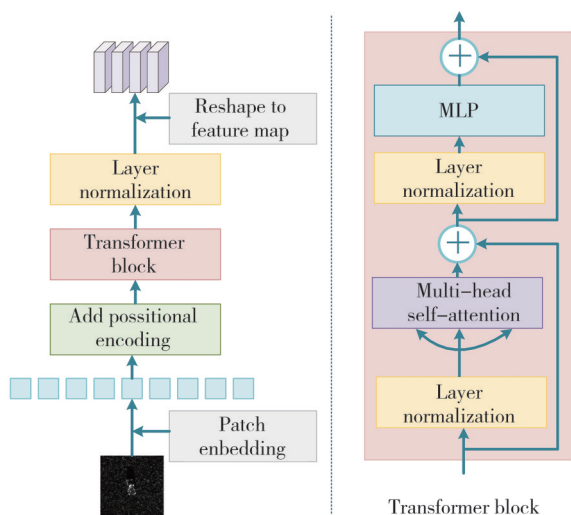


图2 基于ViT的SAR图像特征提取

Fig. 2 SAR image feature extraction based on ViT

本模型使用的多头自注意力模块计算流程如图3所示,首先通过3个可学习的线性变换矩阵 W^Q 、 W^K 、 W^V 将输入特征分别映射为查询(Query, Q)、键(Key, K)、值(Value, V),再将 Q 、 K 、 V 沿通道维

度分割为8个独立的头(Head),对每个头 i 求单头注意力

$$A_i = \text{Soft max} \left(\frac{Q_i K_i^T}{\sqrt{d_k}} \right) \cdot V_i, \quad (1)$$

式中: Q_i 、 V_i 表示第 i 个头的查询、值; K_i^T 表示第 i 个头的键的转置; d_k 为 K_i 的维度; Soft max 为激活函数。最后整合多维度信息

$$A = A_{\text{concat}} \times W^O, \quad (2)$$

式中: A_{concat} 由 A_i 沿通道维度拼接得到; W^O 是可学习的线性变换矩阵。

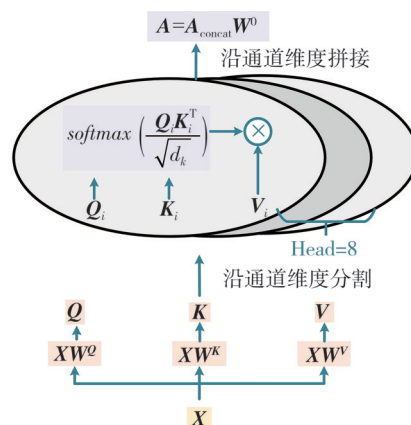


图3 多头自注意力

Fig. 3 Multi-head self-attention

1.2 鲁棒特征下采样

在图像处理领域,特征下采样是一个关键步骤,它涉及到在降低特征图空间分辨率的同时保留关键信息的问题。然而,传统的卷积下采样方法往往会丢失对于目标至关重要的细节信息,这限制了模型在SAR图像分类中的性能。针对这一问题,本文引入了一种新型的浅层鲁棒特征下采样模块(Shallow Robust Feature Downsampling, SRFD)^[21],用于处理网络输入的初始特征图。

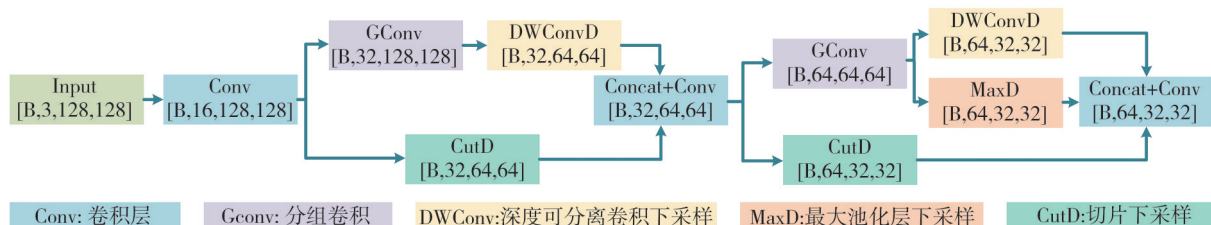


图4 浅层鲁棒特征下采样

Fig. 4 Shallow robust feature downsampling

在初始阶段,特征图包含了大量的细节信息,SRFD的目标是在降低特征图维度的同时,尽可能地保留这些细节,在MAFNet中的实现步骤

如图4所示,其中切片下采样(CutD)层将输入特征图切成不重叠的块,进行 $1 \times N$ 和 $N \times 1$ 的卷积处理,然后进行拼接。这种方法有效地减少了空

间维度,同时保留了原始数据信息。CutD的实现步骤如图 5 所示。

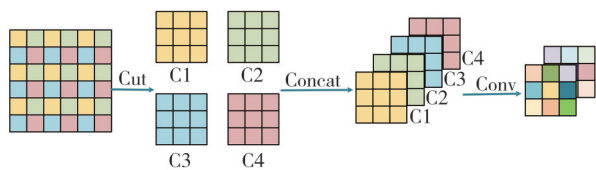
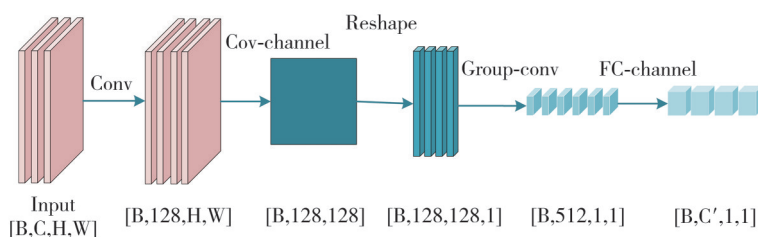


图 5 切片下采样
Fig. 5 Cut downsampling

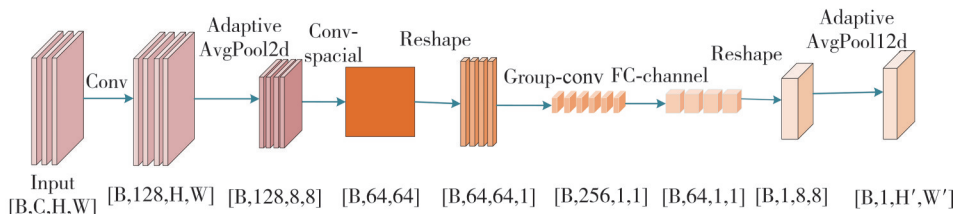
1.3 协方差注意力

协方差注意力机制 (Covariance Attention

Mechanism)通过挖掘特征张量的二阶统计特性建立注意力权重,为注意力生成提供高阶统计依据,能够充分表征特征通道和空间位置间的相关性,并编码特征分布的几何结构信息^[22]。具体而言,该机制分别计算通道维度和空间维度的协方差矩阵,以衡量通道间的相关性以及空间位置之间的依赖关系。通过轻量化矩阵运算实现特征交互建模,具有计算高效、几何可解释性强等特性。协方差注意力包含协方差通道注意力 (Covariance Channel Attention, CCA) 与协方差空间注意力 (Covariance Spatial Attention, CSA),其结构如图 6 所示。



(a) 协方差通道注意力



(b) 协方差空间注意力

图 6 协方差注意力

Fig. 6 Covariance attention

2 实验及结果分析

2.1 数据集及环境介绍

本实验采用标准操作条件 (Standard Operating Condition, SOC) 下的 MSTAR (Moving and Stationary Target Acquisition and Recognition) 数据集评估算法分类性能。该数据集由美国国防高级研究计划署 (DARPA) 和空军研究实验室 (AFRL) 提供,包含通过聚束式合成孔径雷达 (Spotlight SAR) 采集的前苏联军事车辆的 SAR 图像,分辨率 $0.3\text{ m} \times 0.3\text{ m}$,工作于 X 波段 (HH 极化)。SAR 成像方位角覆盖 $0^\circ \sim 360^\circ$ (间隔 $1^\circ \sim 2^\circ$),经预处理后提取目标图像切片。SOC 下 MSTAR 数据集包含 10 类地面目标:自行榴弹炮

2S1、步兵战车 BMP2、装甲侦察车 BRDM2、装甲运输车 BTR60/BTR70、推土机 D7、坦克 T62/T72、卡车 ZIL131、自行高炮 ZSU234 (光学与 SAR 图像对比见图 7)。训练集 (俯角 17°) 与测试集 (俯角 15°) 采用相同目标序列号,具体样本分布见表 1。

本文的实验环境配置如下: Inter 酷睿 i5-13500 处理器, 32G 内存, NVIDIA GeForce RTX3060 显卡, CUDA 版本为 11.8, 操作系统为 win10。SOC 下的 MSTAR 数据集的图像尺寸为 128×128 到 192×193 不等,因此使用中心裁剪得到图像尺寸均为 128×128 的数据集。本文采取动态学习率 (Dynamic Learning Rate), 每 10 轮进行一次学习率的迭代,模型在训练时的超参数设置如表 2 所示。

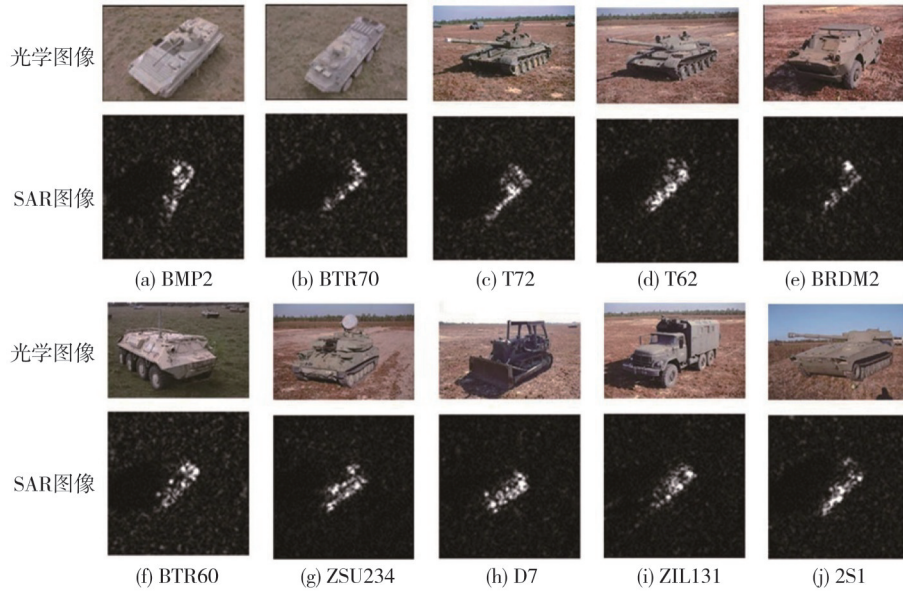


图7 MSTAR数据库中各种军事目标的光学图像和相应的SAR图像

Fig. 7 Optical images and corresponding SAR images of various military targets in the MSTAR dataset

表1 标准操作条件下的10类目标

Tab. 1 10 types of targets under standard operating conditions

类别	序列号	样本大小	训练集样本数目(俯角17°)	测试集样本数目(俯角15°)	总数量
2S1	b01	158×158	299	274	573
BMP2	9 563	128×128	233	195	428
BRDM2	E-71	128×129	298	274	572
BTR60	k10yt7 532	128×128	256	195	451
BTR70	C71	128×128	233	196	429
D7	92v13 015	177×178	299	274	573
T62	A51	172×173	299	273	572
T72	132	128×128	232	196	428
ZIL131	E12	192×193	299	274	573
ZSU234	d08	158×158	299	274	573
合计			2 747	2 425	5 172

表2 实验参数

Tab. 2 Experimental parameters

超参数名称	超参数值
输入图像的尺寸	128×128×3
损失优化器	Adam
损失函数	交叉熵损失函数
学习率	动态学习率
初始学习率	0.000 1
每10轮的学习率降低值	0.000 01
训练轮数	100
批量训练数目	16

2.2 评价指标

为充分验证MAFNet在SAR图像目标分类任务中克服散斑噪声、复杂背景干扰等问题的有效性,实验以准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-Score、ROC曲线和AUC值为评价指标来与其他网络进行对比,以验证MAFNet的优越性。

2.2.1 Accuracy

Accuracy的计算公式为

$$A_{cc} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (3)$$

式中: N_{TP} 表示模型中正确识别的正样本数量; N_{TN} 表示模型正确识别的负样本数量; N_{FN} 表示模型错误地将正样本分类为负样本的数量; N_{FP} 表示模型错误地将负样本分类为正样本的数量; A_{cc} 是模型正确预测的样本在总样本中所占的比例,反映全局判别能力。

2.2.2 Precision

Precision的计算公式为

$$P_{re} = \frac{N_{TP}}{N_{TP} + N_{FP}}。 \quad (4)$$

Precision是模型准确识别出的正样本在所有被模型判定为正样本的样本中所占的比例,用于衡量正类预测的可靠性。

2.2.3 Recall

Recall的计算公式为

$$R_{ec} = \frac{N_{TP}}{N_{TP} + N_{FN}}。 \quad (5)$$

Recall为模型正确识别出的正样本在所有实际为正样本的样本中的比例,反映对正类的捕捉能力。

2.2.4 F1-Score

F1-Score 的计算公式为

$$F1-Score = \frac{2 \times P_{re} \times R_{ec}}{P_{re} + R_{ec}} \quad (6)$$

F1-Score 是 Precision 和 Recall 的调和均值, 平衡两者的 trade-off, 用于综合评估模型的性能。

2.2.5 ROC 曲线

ROC 曲线以假正率(False Positive Rate, FPR)为横轴, 以真正率(True Positive Rate, TPR)为纵轴。

TPR 的计算公式为

$$R_{TP} = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (7)$$

FPR 的计算公式为

$$R_{FP} = \frac{N_{FP}}{N_{FP} + N_{TN}} \quad (8)$$

ROC 曲线可以直观展示模型对正负样本的区分能力。曲线下面积(Area Under Curve, AUC)是 ROC 曲线的量化指标, 取值范围为[0,1], 反映模型对正例和负例的整体区分能力。

更高的 Accuracy、Precision、Recall、F1-Score 和 AUC 值表示模型在分类任务中的性能更强, 越接近 1 说明模型的分类能力越优秀, 模型的 ROC 曲线越陡峭、越靠近左上角, 表示模型的综合判别能力越强。

2.3 实验结果分析

2.3.1 MAFNet 分类性能分析

在本实验中, 采用混淆矩阵作为评价 MAFNet 性能的工具, 结合十折交叉验证实验进行补充验证。

a) 混淆矩阵

MAFNet 在 MSTAR 数据集上的分类混淆矩阵如图 8 所示。

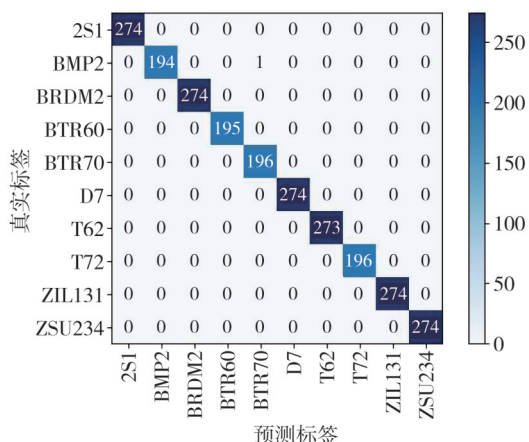


图 8 MAFNet 在 MSTAR 数据集上的分类混淆矩阵

Fig. 8 Classification confusion matrix of MAFNet on the MSTAR dataset

由图 8 可以看出, 10 类样本中, 仅 BMP2 中有一例被错误分类为 BTR70, 观察 BMP2(见图 7(a)) 和 BTR70(见图 7(b)) 的光学图像不难发现, 这两类目标较为相似, 2S1、BRDM2、BTR60、BTR70、D7、T62、T72、ZIL131、ZSU234 的所有样本在分类任务中全部分类正确, 分类准确率达 100%。这一结果验证了 MAFNet 在 SAR 图像目标分类中的优异性能, 为其在实际应用中的有效性提供了可靠支持。

b) 十折交叉验证

考虑到样本数量不多, 本实验将训练集和测试集按类别合并后打乱顺序得到新的数据集, 在此数据集上进行十折交叉验证实验, 为确保实验结果的可靠性, 进行了 5 次实验, 每次实验的准确率为 100%, 取平均值作为最终结果, 如表 3 所示, MAFNet 在每一折的准确率均达 100%。

表 3 5 次十折交叉实验的平均准确率

Tab. 3 Average accuracy of 5 repeated 10-fold cross-validation

折序	准确率/%	折序	准确率/%
第 1 折	100	第 6 折	100
第 2 折	100	第 7 折	100
第 3 折	100	第 8 折	100
第 4 折	100	第 9 折	100
第 5 折	100	第 10 折	100

MAFNet 在十折交叉验证实验中的表现优于在 SOC 下 MSTAR 数据集上的表现, 因为 MSTAR 数据集中训练集和测试集的图像存在一定区别, 如俯角不同(训练集俯角 17°, 测试集俯角 15°)、背景干扰不同等, 十折交叉验证实验中将训练集和测试集混合以后, 训练集和测试集中均同时包含俯角为 15° 和俯角为 17° 的目标图像, 在一定程度上削减了对模型泛化能力的考查, 这说明了 MAFNet 的泛化能力存在一定的进步空间。受限于 SAR 图像获取难度, 现有数据集的样本多样性不足, 网络泛化能力难以得到充分验证。在实际应用中, 目标俯角差异会更大且面临更多干扰因素, 因而研究出更具泛化能力的网络将成为该领域未来的重要研究方向。

2.3.2 MAFNet 与光学图像分类领域常用网络的对比

为验证本文提出的 MAFNet 模型在 SAR 图像目标分类任务中的性能优势, 本文将其与当前广泛使用的图像分类网络进行对比试验, 包括 Mobilenet_V2^[14]、Mobilenet_V3_large^[15]、Shufflenet_v2_×0.5^[16]、Efficientnet_b0^[17]、Resnet^[18]、Densenet121^[19]。

根据表 2 中的参数设置进行实验, 采用多维度

指标(Accuracy、Precision、Recall、F1-Score、AUC值和ROC曲线)对分类结果进行综合评估。分类精度对比数据见表4,细粒度分析结果如图9和图10所示。

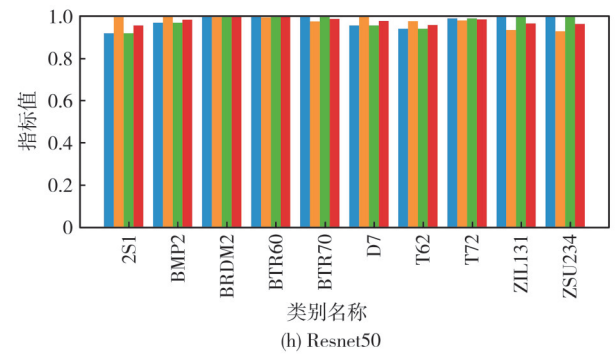
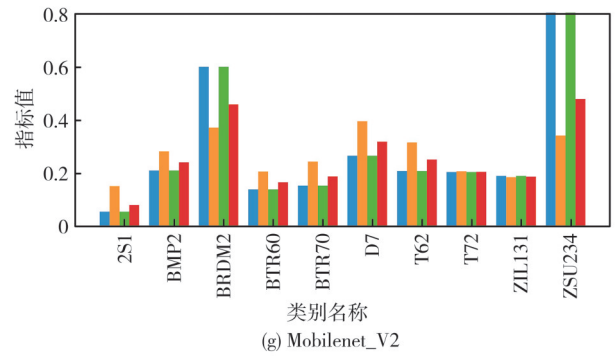
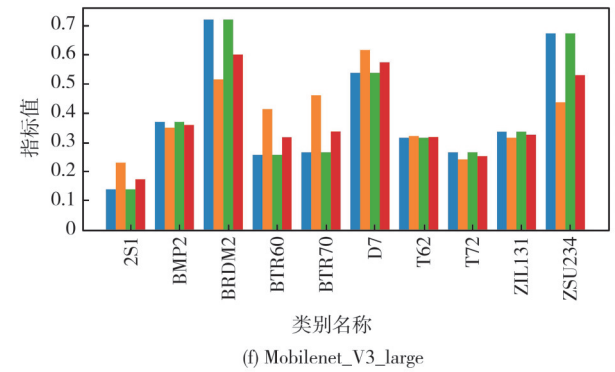
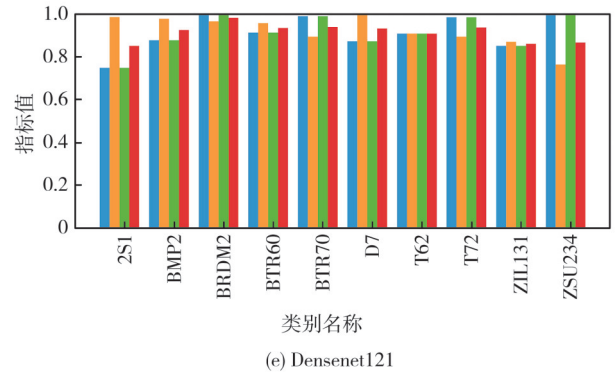
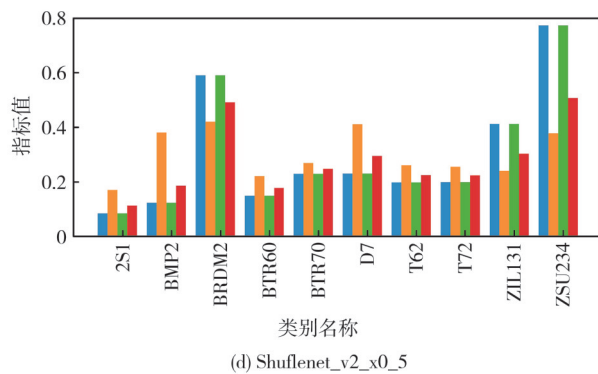
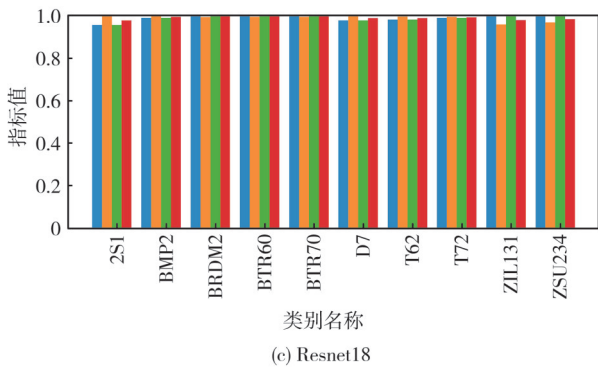
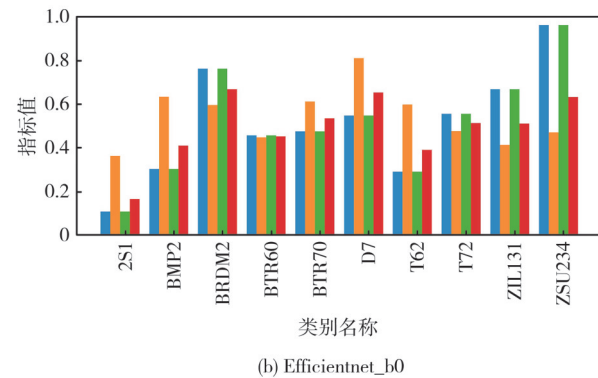
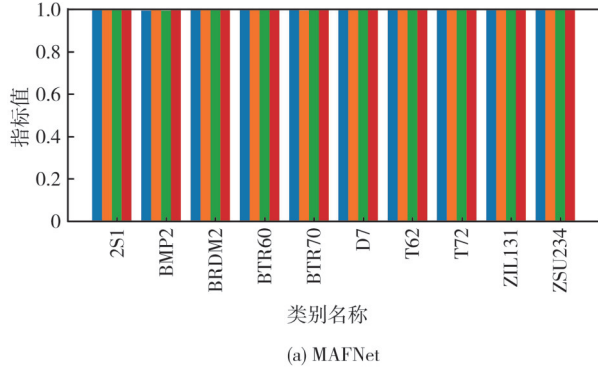


图9 MAFNet与光学图像分类领域常用DCNN相关指标的对比

Fig. 9 Comparison of performance metrics between MAFNet and common DCNNs in optical image classification

这些DCNN在光学图像分类领域取得了很好的效果,但是观察实验结果不难发现,部分DCNN并不适用于SAR图像目标分类任务,只有Resnet、

Densenet121取得了较高的准确率和AUC值。其中, DenseNet121在两类目标分类中的准确率达100%, 其ROC曲线在BRDM2($AUC=1.0000$)类别上与MAFNet完全重合, 在BTR70($AUC=0.9995$)和ZSU234($AUC=0.9994$)类别上也近乎重合。ResNet18与ResNet50在五类目标分类中准确率均达100%。具体而言, ResNet18在8个AUC值为1.0000的类别中与MAFNet的ROC曲线完全重合,

在D7($AUC=0.9996$)和T62($AUC=0.9999$)类别上也基本与MAFNet的ROC曲线重合。ResNet50在多个类别上与MAFNet高度一致, 在5个类别上AUC值为1且ROC曲线完全重合, 在2S1($AUC=0.9998$)、T72($AUC=0.9999$)和ZSU234($AUC=0.9998$)类别上与MAFNet基本重合, 仅在D7($AUC=0.9991$)和T62($AUC=0.9992$)类别上存在细微差异。

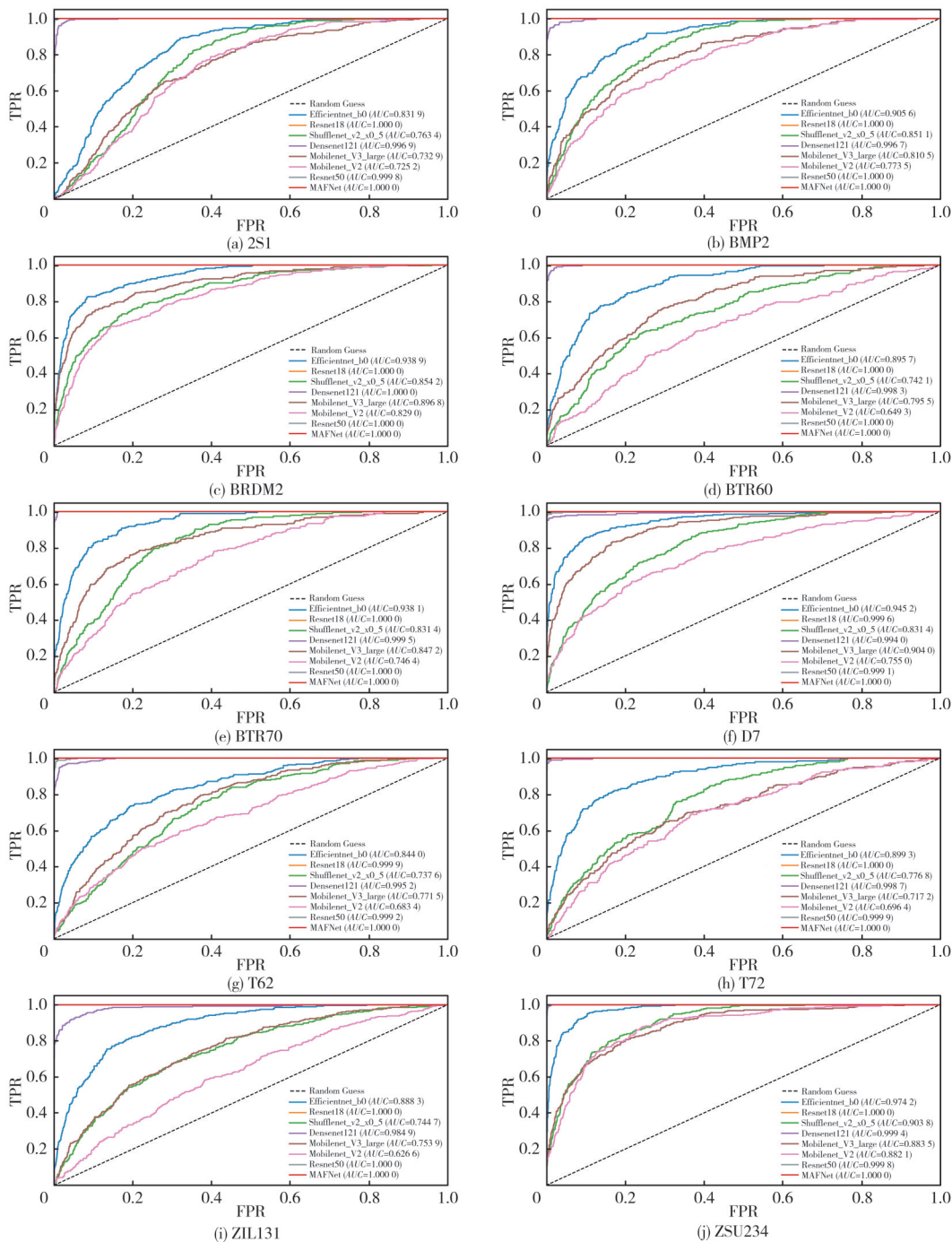


图 10 MAFNet与光学图像分类领域常用DCNN的ROC曲线对比

Fig. 10 Comparison of the ROC curves between MAFNet and the commonly used DCNNs in the field of optical image classification

上述结果表明 Resnet 对 SAR 图像目标具有较强的区分能力，可以正确识别出大部分类别，这证明了残差结构提取 SAR 图像目标特征的有效性。相比之下，MAFNet 在 SAR 图像分类任务中展现出显著优于表 3 中所列 DCNN 模型的性能，

其平均分类准确率达 99.96%，各类别目标的 AUC 值均达 1，且在 Accuracy、Precision、Recall、F1-Score 四项指标上均全面超越其他对比模型，验证了其对于 SAR 图像复杂散射特性的高效建模能力。

表 4 光学图像分类领域优秀网络对各类目标的分类精度对比
Tab. 4 Comparison of classification accuracy of outstanding networks for different targets in optical image classification

分类方法	不同目标类型的分类精度/%										
	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL131	ZSU234	Total
MAFNet	100.00	99.49	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.96
EfficientNet-B0	63.50	68.21	94.89	96.92	93.37	92.70	84.62	93.37	99.64	98.91	88.70
Resnet18	95.62	98.97	100.00	100.00	100.00	97.81	98.17	98.98	100.00	100.00	98.89
ShuffleNet	9.12	37.44	75.91	25.64	58.67	48.54	36.63	48.47	56.57	93.80	49.94
DenseNet121	96.35	98.97	100.00	99.49	100.00	93.80	97.44	99.49	99.27	99.64	98.31
MobileNetV3	30.66	38.46	70.44	45.64	43.88	70.44	40.66	39.29	62.77	70.80	52.54
MobileNetV2	25.18	42.05	84.31	52.31	57.14	54.74	35.90	57.14	68.25	96.35	58.02
Resnet50	91.97	96.92	100.00	100.00	100.00	95.62	94.14	98.98	100.00	100.00	97.61

图 11 为各模型训练损失曲线对比，Resnet18 在前几轮的损失最低，但训练损失曲线不够平滑，表明训练过程存在一定的不稳定性。MAFNet 的训练损失下降过程高效且稳定，在早期训练阶段即展现出快速收敛的特性，且最早收敛至较低损失值，下降至低损失值后，曲线保持平稳，无显著波动，说明模型在训练后期趋于稳定，这体现了 MAFNet 对任务的适应性优势，证明了其设计上的合理性与有效性。

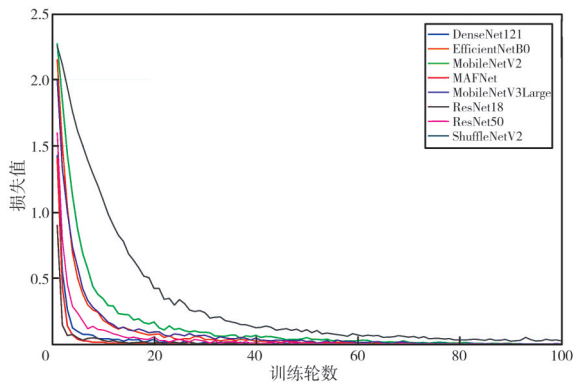


图 11 训练损失曲线的对比

Fig11 Comparison of training loss curves

2.3.3 MAFNet 与 SAR 图像目标分类网络对比

为进一步验证 MAFNet 在 SAR 图像目标分类领域的贡献，采用 Accuracy、Precision、Recall 和 F1-Score 作为评价标准，将 MAFNet 与一系列在该领域表现优异的 DCNN 进行对比，这些 DCNN 包括 A-ConvNet^[4]、M-Net^[5]、Sup-ConL^[6]、CNN-A^[8]、ICBAM_CNN^[13]。为公平对比各网络在理想条件下的最优分类能力，本文采用原论文的实验结果进行对比，实验设置以及分类结果对比如表 5 所示，各类别分类精度对比如表 6 所示，Accuracy、Precision、Recall、F1-Score 四类指标的对比如图 12 所示。

表 5 不同 SAR 图像分类网络的实验设置及分类结果对比
Tab. 5 Comparison of experimental settings and classification results for different SAR image classification networks

方法	图片大小	俯角/(°)		准确率/%
		训练集	测试集	
ICBAM_CNN	128×128			95.59
CNN-A	128×128			98.50
Sup-ConL	224×224			99.06
A-ConvNet	88×88	17	15	99.13
Resnet-D-J	64×64			99.67
M-Net	70×70			99.71
MAFNet	128×128			99.96

表 6 不同 SAR 图像分类网络在 MSTAR 数据集上的分类精度对比

Tab. 6 Comparison of classification accuracy of different SAR image classification networks on the MSTAR dataset

分类方法	不同目标类型的分类精度/%										
	2S1	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL131	ZSU234	Total
ICBAM_CNN ^[13]	91.97	88.21	95.26	93.85	98.47	99.27	93.04	100.00	99.27	95.97	
M-Net ^[5]	99.64	99.49	99.64	99.49	100.00	99.64	100.00	100.00	99.27	100.00	
A-ConvNet ^[4]	98.18	98.98	99.27	96.41	99.49	99.27	99.63	100.00	99.64	99.64	
Sup-ConL ^[6]	98.18	91.84	98.54	96.92	92.35	98.91	97.44	95.41	98.18	99.27	
CNN-A ^[8]	97.08	93.19	98.91	90.77	98.47	99.64	98.17	99.83	99.27	98.91	
MAFNet	100.00	99.49	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	

A-ConvNet^[4]、M-Net^[5] 分别将原始图像裁剪为 88×88、70×70 来进行数据增强。从具体分类情况

看，M-Net 对三类目标的分类准确率达 100%，A-ConvNet 对一类目标的分类准确率达 100%。从综

合评估指标来看, Accureacy、Precision、Recall、F1-Score这四个指标均表现不错,这两种方法总的准确率均达到了99%以上,说明它们能准确分辨部分类别,能正确分类大部分目标。Sup-ConL^[6]是一种基于ViT的自监督对比学习模型,经过数据增强后,把所有图像调整为224×224的统一尺寸以满足ViT模型的输入要求,但其对BMP2(91.84%)、BTR70(92.35%)的分类准确率较低,四项指标整体落后于A-ConvNet、M-Net,但总准确率仍达99.06%,说明该网络对SAR图像目标有较强的分类能力。CNN-A^[8]、ICBAM_CNN^[13]、MAFNet均使用128×128大小的SAR图像,CNN-A通过特定拓扑与训练方法实现98.50%的准确率,对D7、T72、ZIL131的分类准确率达99%以上。

ICBAM_CNN将方差和中心坐标注意力机制集成到卷积神经网络来提高模型的性能,对T72的分类准确率达100%,对其他目标的分类准确率较低,Accureacy、Precision、Recall、F1-Score四个指标均表现均一般,总准确率仅为95.59%,这与网络设计有一定关系,因为该论文的实验重在突出中心坐标注意力机制的作用,没有使用特征提取能力较强的卷积模块。相较之下,本文提出的MAFNet准确率为99.96%,Accureacy、Precision、Recall、F1-Score四个指标均表现优秀,进一步说明了MAFNet在SAR图像目标分类任务中的性能较为出色。

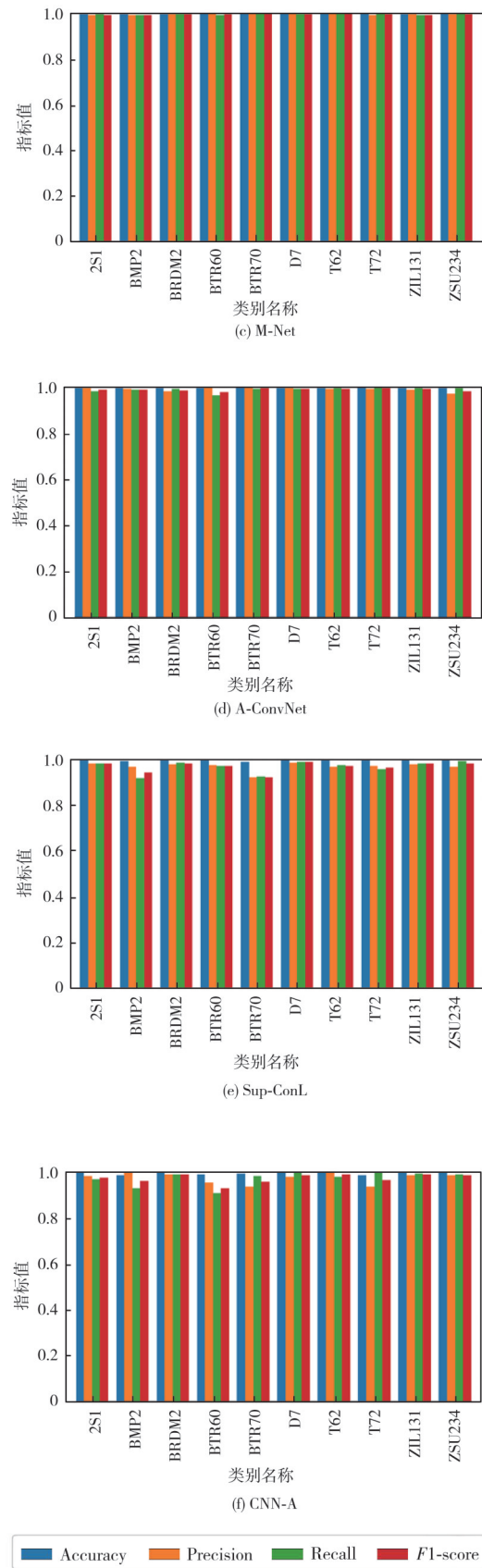
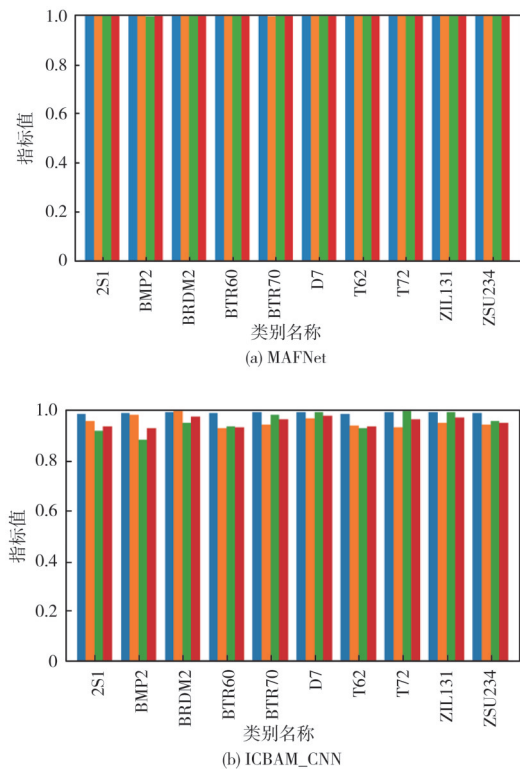


图 12 不同SAR图像分类网络的相关指标
Fig. 12 Relevant metrics for different SAR image classification networks

2.4 消融实验

为验证MAFNet中加入协方差通道注意力机制、协方差空间注意力机制和多头自注意力机制的有效性,以分类精度为评价标准进行消融实验,表7记录了在MSTAR数据集上的实验结果。

根据表7中的数据分析可知:只加入协方差通道注意力机制时,分类准确率为99.63%;只加入协方差空间注意力机制时,分类准确率为99.75%;只使用多头自注意力机制时,分类准确率为99.84%;同时使用协方差注意力模块和多头自注意力模块时,分类准确率达到99.96%,比只使用协方差通道注意力模块提高了0.33个百分点,比只使用协方差空间注意力模块提高了0.21个百分点,比只使用自注意力模块提高了0.12个百分点,进一步验证了融合多种注意力机制在提高SAR图像目标分类精度方面的有效性。

表7 消融实验结果

Tab.7 Ablation experiment results

协方差通道 注意力	协方差空间 注意力	多头自注意力	分类精度/%
✓			99.63
	✓		99.75
		✓	99.84
✓	✓	✓	99.96

3 结论

本文提出一种融合多注意力特征的网络——MAFNet,通过协方差注意力机制充分表征特征通道间和位置间的相关性,增强了特征表达,提高了鲁棒性;同时,通过自注意力机制直接捕捉全局信息,更加有效地解决了SAR图像各类别难以有效区分的问题;该架构通过跨尺度注意力机制动态整合不同分辨率的散射特征,有效解决了SAR图像中目标尺寸变化大和背景杂波干扰强的难题。在MSTAR数据集上的实验结果表明,其准确率高达99.96%,混淆矩阵进一步证实了其对各类目标的精准分类能力。与其他代表性算法相比,MAFNet在分类精度上展现出显著优势,这为SAR图像在目标侦察、监视等实际应用中的自动目标识别提供了可靠技术的支撑。探索具备更强泛化能力的网络模型将成为该领域未来的重要研究方向。

参考文献:

[1] NOVAK L M, OWIRKA G J, BROWER W S,

et al. The automatic target-recognition system in SAIP [J]. The Lincoln Laboratory Journal, 1997, 10(2): 187-202.

[2] IKEUCHI K, WHEELER M D, YAMAZAKI T, et al. Model-based SAR ATR system[J]. Algorithms for Synthetic Aperture Radar Imagery III, 1996, 2757: 376-387.

[3] ZHAO Q, Principe J C. Support vector machines for SAR automatic target recognition[J]. IEEE Transactions on Aerospace and Electronic Systems, 2001, 37(2): 643-654.

[4] CHEN S Z, WANG H P, XU F, et al. Target classification using the deep convolutional networks for SAR images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4806-4817.

[5] SHANG R H, WANG J M, JIAO L C, et al. SAR targets classification based on deep memory convolution neural networks and transfer parameters[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(8): 2834-2846.

[6] DENG J P, ZHU Y Y, ZHANG S N, et al. SAR image recognition using ViT network and contrastive learning framework with unlabeled samples[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 4000205.

[7] FU Z Z, ZHANG F, YIN Q, et al. Small sample learning optimization for resnet based sar target recognition [C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018: 2330-2333.

[8] WILMANSKI M, KREUCHER C, LAUER J. Modern approaches in deep learning for SAR ATR [C]//Proceedings of SPIE 9843, Algorithms for Synthetic Aperture Radar Imagery XXIII, 2016: 98430N.

[9] SUN Y C, YAN K J, LI W Z. CycleGAN-based SAR-optical image fusion for target recognition[J]. Remote Sensing, 2023, 15(23): 5569.

[10] BANAS K M, KREUCHER C. A multi-view CNN for SAR ATR [C]//2024 IEEE Radar Conference (RadarConf24), 2024: 1-6.

[11] REN H H, LIU S, MIAO L, et al. Hybrid reasoning network with class-oriented hierarchical representation for few-shot SAR target recognition [J]. IEEE Sensors Journal, 2024, 24(16): 26091-26103.

[12] ZHANG R L, DUAN Y Z, ZHANG J D, et al. An adaptive multiview SAR automatic target recognition network based on image attention[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17(1): 13634-13645.

- [13] 孙靖森, 李宗豫, 杨森, 等. 基于集成改进卷积注意力块的SAR图像目标分类算法[J]. 海军航空大学学报, 2024, 39(4): 445-452.
SUN Jingsen, LI Zongyu, YANG Sen, et al. SAR image target classification algorithm based on integrated improved convolutional block attention module [J]. Journal of Naval Aviation University, 2024, 39(4): 445-452. (in Chinese)
- [14] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [15] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 1314-1324.
- [16] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [17] TAN M X, LE Q. EfficientNet: Rethinking model scaling for convolutional neural networks [DB/OL]. (2019-05-28) [2025-03-24]. <https://arxiv.org/abs/1905.11946>.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [19] HUANG G, LIU Z, MAATEN L, et al. Densely connected convolutional networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261-2269.
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [DB/OL]. (2021-06-03) [2025-03-24]. <https://arxiv.org/abs/2010.11929>.
- [21] LU W, CHEN S B, TANG J, et al. A robust feature downsampling module for remote-sensing visual tasks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4404312.
- [22] GAO Z L, XIE J T, WANG Q L, et al. Global second-order pooling convolutional networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 3019-3028.