

# 基于小波变换平行注意力的多源遥感图像分类

王嘉毅, 高峰\*, 张天戈, 甘言海

(中国海洋大学 计算机科学与技术学院, 青岛 266100)

**摘 要:** 如何充分挖掘多源遥感图像数据特征的依赖关系, 实现不同模态图像数据间的优势互补, 已成为遥感领域的研究热点方向之一。现有的高光谱和 SAR 数据联合分类任务主要面临以下两个关键难题: 1) 图像的特征提取和特征表达不充分, 高频信息容易损失, 不利于后续的分类任务; 2) 多源图像特征的交互有限, 多模态特征关联不紧密。针对上述两个问题, 围绕图像特征的鲁棒表达和多源特征的高效关联开展了研究工作, 提出了基于小波变换和平行注意力机制的多源遥感图像分类方法。基于小波变换的特征提取器可以充分利用频域分析技术, 在可逆下采样的过程中充分捕捉粗/细粒度级别特征; 基于平行注意力机制的特征融合器充分综合多模态遥感数据的一致性和差异性, 完成强相关性特征的融合和生成, 以提升分类准确率。在两个真实多源遥感数据集上的实验表明, 所提出的分类方法具有显著优势, 在 Augsburg 和 Berlin 数据集的总体准确率分别达到 90.40% 和 76.23%, 相对 DFINet 等主流方法在两个数据集上的总体准确率分别至少提升 2.66%、12.22%。

**关 键 词:** 高光谱图像; 合成孔径雷达; 小波变换; 多源特征融合; 遥感图像

中图分类号: TP753

文献标识码: A

文章编号: 1001-5965 (XXXX) XX-XXXX-XX

近年来, 我国的卫星遥感技术发展日新月异, 多源遥感卫星数据 (高光谱、合成孔径雷达、红外、可见光等) 已经应用于国民经济、社会生活与国家安全等各个方面, 初步形成了遥感卫星应用体系。遥感图像分类技术通过对遥感图像中的地物特征进行识别、加工和分类, 达到对图像中多种地物信息进行识别的目的<sup>[1]</sup>。其中, 高光谱传感器能够获取地物丰富的光谱信息, 捕获反映物体材质的差异信息<sup>[2]</sup>。合成孔径雷达 (Synthetic Aperture Radar, SAR) 传感器能够全天时、全天候工作, 能够有效弥补恶劣天气 (云雾遮挡、雾霾等) 情况下高光谱传感器观测的局限性<sup>[3]</sup>。然而, SAR 传输图像往往受到相干斑噪声的影响, 目标细节模糊, 视觉解译难度大。高光谱图像相比于 SAR 图像具有光谱信息丰富、目标结构明显等特点, 但容易受光谱、云雾、阴影等因素的影响。融合高光谱和 SAR 图像的特征用于地物分类, 能够实现多源遥感信息优势互补, 实现地

物信息更全面、更丰富的解译。

高光谱和 SAR 图像由于成像机理上的差异导致其数据存在较大的差异, 多源图像特征的关联非常困难。针对多源遥感图像分类任务, 国内外学者提出了很多方法: Man 等人<sup>[4]</sup>采用了像素级特征融合分类器, 应用极大似然分类和支持向量机等方法, 实现了高光谱图像和激光雷达 (Light Detection And Ranging, LiDAR) 数据的融合分类。Hu 等人<sup>[5]</sup>提出了一种基于对象的融合方法, 用于联合使用极化合成孔径雷达和高光谱数据的地物分类任务。Chen 等人<sup>[6]</sup>利用卷积神经网络处理多源遥感数据, 分别提取其光谱空间特征和高度特征, 以获取深层次的特征信息并完成融合分类。Hong 等人<sup>[7]</sup>提出了深层编码器-解码器网络, 利用强制融合单元对不同模态的特征进行重构, 提升了多模态信息的利用率, 更好地完成了多源遥感图像分类任务。Zhao 等人<sup>[8]</sup>提出了 HRWN 模型以更好地解决遥感图像融合分类问

收稿日期: 2023-06-01; 录用日期: 2023-\*\*-\*\*-; 网络出版时间:

网络出版地址:

基金项目: 国家自然科学基金 (42106191); 科技创新 2030-新一代人工智能重大项目 (2022ZD0117202)

\*通信作者: E-mail: gaofeng@ouc.edu.cn

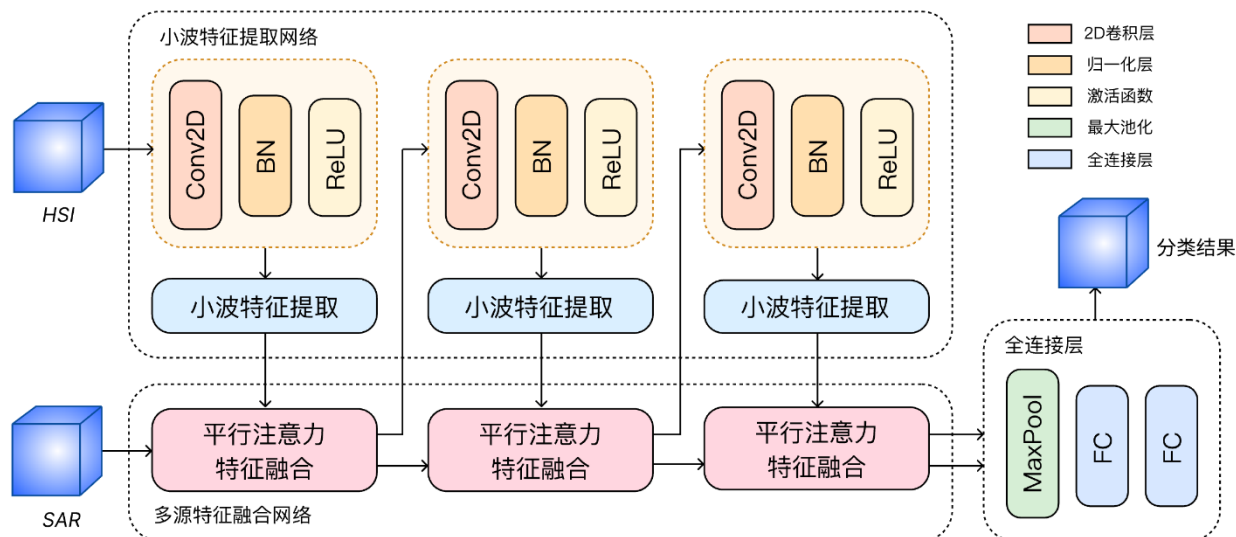


图 1 基于小波变换和平行注意力机制的多源遥感图像融合分类网络

Fig.1 Multi-source remote sensing image fusion classification network based on wavelet transform and parallel attention mechanism

题，通过双隧道卷积神经网络对空间特征和光谱特征进行全局先验，采用像素级亲和分支以增强样本特征空间的一致性，从而实现多源遥感数据的高效融合分类。Feng 等人<sup>[9]</sup>提出了一种线性自注意力融合网络模型，以实现高光谱图像和 LiDAR 图像的特征融合。该模型由特征提取单元、自注意力单元和特征融合单元构成，显著地提高了多源遥感图像融合分类的精度和准确率。Li 等人<sup>[10]</sup>提出了一种基于非对称特征融合的分类方法，利用权重共享的剩余块进行特征提取，设计了特征校准模块，利用多源特征的空间依赖性提升识别能力，提升了多源遥感图像融合分类的准确率。

尽管国内外学者近年来围绕高光谱和 SAR 融合分类技术开展了大量的研究工作，在多个国内外公开的数据集上取得了较好的效果，但仍难以满足遥感应用的实际需求，主要因为如下两个技术难点亟待攻克：（1）**图像中的高频信息容易损失**。现有特征提取方法通常会使用池化、空洞卷积等操作来减少计算成本，但是容易损失图像中的纹理细节。（2）**多模态特征的交互有限**。当前的多源图像特征融合方法大多简单拼接多源特征，或仅采用加权融合的方式，其对于不同模态特征交互信息的利用较为有限，亟需针对遥感场景的复杂性进行重构和优化。

为了解决以上2个问题，本文提出了基于小波

变换和平行注意力机制的多源遥感图像分类网络（Wavelet Transform and Parallel Attention-Based Network, WPANet）。具体而言，为了进一步充分挖掘高光谱数据中的高频特征和低频特征，本文设计了基于小波变换的特征提取器，利用频域空间分析原理，充分捕捉多源图像特征之间的特征依赖。此外，本文设计了基于平行注意力机制的特征融合器，分别提取多源数据间的一致性和差异性特征。其中，一致性分支更加关注多源特征的一致性，抑制非显著信息的干扰；差异性分析更加关注多模态特征在显著区域的差异，从而实现多模态特征的对齐。本方法在 Augsburg 和 Berlin 两个多源遥感数据集上完成实验，实验结果表明 WPANet 在多源遥感图像地物分类任务中具有较好的效果。

本文的主要贡献如下：

（1）在图像特征鲁棒表达方面，提出了基于小波变换的特征提取器，利用频域分析原理完成对粗/细粒度级别特征的提取，通过中间特征的可逆下采样，完成特征的充分挖掘。

（2）在多源特征高效关联方面，构建了基于平行注意力的特征融合器，分别关注多源特征的一致性和差异性，充分融合多源数据的特征以得到更高质量的联合特征，从而实现地物分类准确率的有效提升。

（3）在 Augsburg 和 Berlin 两个多源遥感

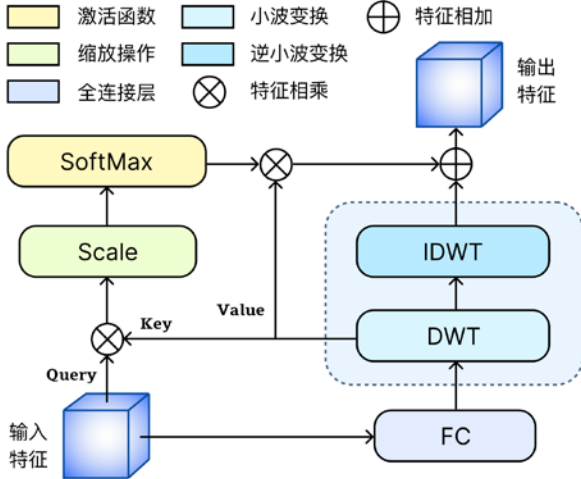


Fig.2 Wavelet transform feature extractor

数据集上进行了大量实验，从定量分析和定性分析两个角度证明了本文所提出的 WPANet 模型的有效性及其可扩展性。

## 1 本文方法

本文提出的 WPANet 模型整体架构如图1所示，由小波特征提取网络、多源特征融合网络以及全连接层组成。WPANet 主要包括两个关键模块：基于小波变换的特征提取模块和平行注意力特征融合模块。下面将分别介绍本文方法的网络整体架构、基于小波变换的特征提取模块和平行注意力特征融合模块。

### 1.1 网络整体架构

本文选择了常见的双流网络架构以完成跨模态信息特征的交互和融合。首先给定两组输入信息：高光谱图像  $I^{HSI} \in \mathbb{R}^{H \times W \times C}$  和合成孔径雷达图像  $I^{SAR} \in \mathbb{R}^{H \times W \times C}$ ，将这两种不同模态的图像作为 WPANet 的输入信息。采用主成分分析方法（Principal Component Analysis, PCA）对  $I^{HSI}$  在通道维度进行降维处理以保留其关键信息，从而减少计算量，提升网络训练效率。

在小波特征提取网络中， $I^{HSI}$  经过基于小波变换的特征提取模块，能够更好地实现图像特征在频率域的信息整合，以更好地实现高频和低频特征的提取。在多源特征融合网络中，使用基于平行注意力的特征融合器，对经过特征提取的  $I^{HSI}$  和上一级的  $I^{SAR}$  完成跨尺度和跨模态的特征融合。最后，融合的特征输入全连接层处理，将已充分融合的高维度特征映射为不同的地物特征种类，使用 Softmax 函数进行归一化，得到最终的地物分类结果。

### 1.2 基于小波变换的特征提取器

小波变换特征提取器将小波变换引入传统的 Transformer 特征编码器中，利用频域空间分析原理，充分挖掘粗/细粒度级别的特征信息，从而更好地实现中间特征的可逆下采样。如图2所示，给定输入特征图  $I \in \mathbb{R}^{H \times W \times C}$ ，通过可学习的嵌入矩阵  $W_d \in \mathbb{R}^{C \times C/4}$  将其线性变换为  $\bar{I}$ ：

$$\bar{I} = IW_d \in \mathbb{R}^{H \times W \times \frac{C}{4}} \quad (1)$$

此时通道数缩小为原来的 1/4。通过离散小波变换（DWT）将  $\bar{I}$  分解为 4 个小波子带实现下采样。

本文选择 Haar 小波函数， $2 \times 2$  Haar 变换矩阵可表示为：

$$H_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (2)$$

其中， $H_2$  的行可用于分别定义分析滤波器  $F_L = (1/\sqrt{2}, 1/\sqrt{2})$  和  $F_H = (1/\sqrt{2}, -1/\sqrt{2})$ 。沿行应用低通滤波器  $F_L$  和高通滤波器  $F_H$  将  $\bar{I}$  分解为 2 个小波子带  $I_L$  和  $I_H$ ，进一步沿列方向使用相同的低通滤波器和高通滤波器将  $I_L$  和  $I_H$  分解为 4 个小波子带： $I_{LL}$ 、 $I_{LH}$ 、 $I_{HL}$  和  $I_{HH}$ ：

$$I_{LL}, I_{LH}, I_{HL}, I_{HH} = \text{DWT}(\bar{I}) \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times \frac{C}{4}} \quad (3)$$

此时，每个小波子带可以看作  $\bar{I}$  经过无损下采样得到的特征信息，此过程中无特征信息丢失。如图3所示， $\bar{I}$  为原始的输入信息，经过 Haar 小波变换， $\bar{I}$  被分解成了 4 部分： $I_{LL}$ （原始的低频信息）， $I_{LH}$ （水平方向的高频信息）， $I_{HL}$ （垂直方向的高频信息）和  $I_{HH}$ （对角线方向的高频信息）。 $I_{LL}$  在粗粒度级别反映对象的基本结构特征，而  $I_{LH}$ 、 $I_{HL}$  和  $I_{HH}$  在细粒度级别保留对象的纹理细节信息。

沿通道维度拼接 4 个小波子带，得到频域综合特征  $\hat{I}$ ：

$$\hat{I} = \text{Concat}(I_{LL}, I_{LH}, I_{HL}, I_{HH}) \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times C} \quad (4)$$

接下来，将  $\hat{I}$  变换为 key ( $K^w$ ) 和 value ( $V^w$ )，而 query ( $Q$ ) 为原始输入特征图  $I$ 。此时基于小

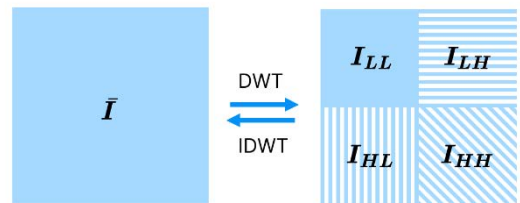


图 3 DWT 和 IDWT 过程

Fig.3 DWT and IDWT process



波变换的多头自注意力计算在 query 和已经下采样的 key 和 value 上进行:

$$\begin{aligned} \text{head}_i &= \text{Attention}(Q_i, K_i^w, V_i^w) \\ &= \text{Softmax}\left(\frac{Q_i K_i^{w\top}}{\sqrt{D_h}}\right) V_i^w \end{aligned} \quad (5)$$

其中  $K_i^w$  表示第  $i$  个头经过下采样后的 key,  $V_i^w$  表示第  $i$  个头经过下采样后的 value,  $D_h$  表示表示每个头的维度。由于多头自注意力的聚合性,  $\text{head}_i$  为输入的长距离全局依赖特征信息。对  $\hat{I}$  应用逆小波变换 (IDWT) 得到重构的特征图  $I'$ :

$$I' = \text{IDWT}(\hat{I}) \in \mathbb{R}^{H \times W \times \frac{C}{4}} \quad (6)$$

其中, IDWT 为逆小波变换。根据小波变换理论, 重构得到的特征图  $I'$  包含原始输入的细节, 从而实现局部特征信息的可逆下采样, 更好地在多尺度信息上完成自注意力学习。小波特征提取器的整体操作过程可以由式 7 表示:

$$\begin{aligned} \text{Wavelet}(I) &= \text{MultiHead}^w(Q, K^w, V^w, I') \\ &= \text{MultiHead}^w(IW^Q, \hat{IW}^K, \hat{IW}^V, I') \quad (7) \\ &= \text{Concat}(\text{head}_0, \dots, \text{head}_{N_h}, I') W^O \end{aligned}$$

其中,  $N_h$  表示头的数量,  $W^O$  为变换矩阵。与常见的注意力模块相比, 在加入小波变换之后, 小波特征提取器中的 DWT-IDWT 过程可以在不计计算成本的前提下, 将特征信息进行强相关的局部上下文处理, 其注意力模块具有更大的感受野, 能够更好地实现局部特征信息和上下文之前的交互。

### 1.3 基于平行注意力机制的特征融合器

基于平行注意力机制的特征融合器可以更好地将不同层级的两种模态图像特征在深层次完成特征信息的交互和融合。经过平行注意力融合器的计算, 类间特征的差异性得以增强, 对应局部差异的细节纹理得以保留; 类内特征的一致性可以更为充分地保证两种模态图像特征完成高效的综合。

如图 4 所示, 不同层级的两种模态图像特征  $F_{HSI}$  和  $F_{SAR}$  首先采用逐元素乘法  $\otimes$ , 得到初步的一致性特征信息  $F_{HSI} \otimes F_{SAR}$ 。紧接着将所得到的初步一致性特征经过通道注意力 (Channel Attention, CA) 的计算, 在通道维度上完成注意力的聚合, 以关注图像特征信息量较大的通道, 抑制相关性较低的通道特征信息<sup>[11]</sup>, 得到通道注意力权重  $W$ :

$$W = \text{CA}(F_{HSI} \otimes F_{SAR}) \quad (8)$$

利用通道注意力权重  $W$  来增强原始的两个模态的图像的显著性特征信息, 通过逐元素加法完成一致性联合特征  $F_C$  的计算:

$$F_C = (W \otimes F_{HSI}) \oplus (W \otimes F_{SAR}) \quad (9)$$

一致性联合特征  $F_C$  在跨层级和跨模态的图像特征信息中实现了特征一致性的显著提升。差异性联合特征  $F_D$  描述了图像特征的差异信息, 跨模态图像的类间差异得以保留和强化, 对于跨模态图像特征融合有高效地提升作用。差异性联合特征  $F_D$  关注类间差异, 其定义为:

$$F_D = |W \otimes F_{HSI} \ominus W \otimes F_{SAR}| \quad (10)$$

其中  $\ominus$  表示逐元素减法,  $||$  表示逐元素计算绝对值。通过结合一致性联合特征  $F_C$  和差异性联合特征  $F_D$  得到平行注意力融合器输出的融合信息  $F_{CD}$ :

$$F_{CD} = F_C \oplus F_D \quad (11)$$

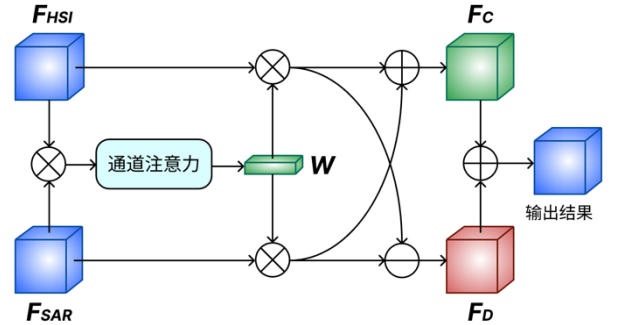


图 4 基于平行注意力机制的特征融合器

Fig.4 Feature fuser based on parallel attention mechanism

## 2 实验结果与分析

### 2.1 数据集

1) Augsburg 数据集采集于德国奥格斯堡市上空, 包括星载 HSI 图像和双 pol 偏振合成孔径雷达 (PolSAR) 图像<sup>[12][13]</sup>。Augsburg 数据集图像场景包含了  $332 \times 485$  像素和 180 个光谱波段, 其中 HSI 图像的光谱范围为  $0.4 \mu\text{m}$  至  $2.5 \mu\text{m}$ 。SAR 数据包含了 VV 强度、VH 强度以及 PolSAR 协方差矩阵非对角元素的实部和虚部 4 个特征。

2) Berlin 数据集采集于柏林城区及其周围农村地区上空, 包括 HSI 图像和 SAR 图像<sup>[14]</sup>,

Berlin 数据集中 HSI 图像场景的空间分辨率为 30 m, 包含了  $797 \times 220$  像素和 244 个光谱带, 光谱带范围为  $0.4 \mu\text{m}$  至  $2.5 \mu\text{m}$ 。相应的 SAR 图像空间分辨率为 13.89m, 图像大小为  $1723 \times 476$  像素。由于这两种模态图像大小不同, 使用最近邻插值算法来扩展 HSI 的大小与 SAR 图像保持一致。

## 2.2 实验参数分析

Augsburg 数据集和 Berlin 数据集中高光谱图像的光谱波段数分别为 180 和 244, 选择 PCA 方法对高光谱图像中的光谱特征维度进行降维处理, 仅保留关键的特征信息, 减少神经网络的输入数据量以提升其计算效率。为探究高光谱主成分数  $n$  和图像分类效果的关系, 在 Augsburg 和 Berlin 数据集上完成实验, 选择总体分类准确度 (OA) 作为评价指标, 设置  $n$  的大小从集合 {10, 20, 30, 40, 50} 中取值, 得到  $n$  与 OA 的映射关系如图 5 所示。从图中可以看出, 当  $n$  取值为 30 时 WPANet 达到最理想的实验结果。当  $n$  小于 30 时,  $n$  与 OA 呈正相关。当  $n$  大于 30 后, OA 无明显提升效果, 甚至出现退化的情况, 因此最终将高光谱主成分数设置为 30。

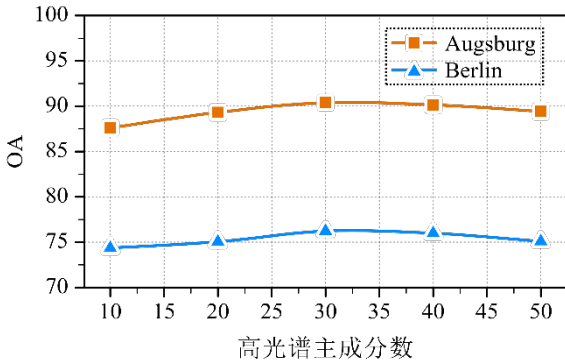


图 5 高光谱主成分数与总体准确率关系

Fig.5 Relationship between hyperspectral principal component score and overall accuracy

## 2.3 对比实验分析

为了进一步验证本文提出的 WPANet 地物分类模型的有效性, 选择支持向量机 (Support Vector Machines, SVM)、局部二进制模式-极限学习机 (Local Binary Pattern-Extreme Learning Machine, LBP-ELM)、双分支卷积神经网络 (Two-Branch CNN, TBCNN)、上下文深度卷积网络 (ContextCNN) 和深度特征交互网络 (Depthwise Feature Interaction Network, DFNet) 共 5 种不同

的遥感图像地物分类方法与 WPANet 进行对比实验分析:

1) SVM<sup>[15]</sup> 通过将数据样本特征映射到高维空间, 寻找分类超平面以解决分类的目标任务。2) LBP-ELM<sup>[16]</sup> 首先使用 LBP 充分挖掘图像局部特征信息, 如边缘特征、纹理特征、斑点特征等, 然后和全局 Gabor 特征与原始光谱特征相融合, 输入特征分类器以得到最终的分类结果。3) TBCNN<sup>[17]</sup> 可以高效地完成多源遥感图像特征提取任务。首先在特征融合阶段融合 HSI 图像的光谱和空间特征, 然后再将 HSI 特征与其他数据特征相结合, 完成多源遥感图像的高效融合。4) ContextCNN<sup>[18]</sup> 通过设计多尺度卷积滤波组模块, 利用局部的语境交互信息, 充分挖掘上下文信息的潜在交互作用, 以实现空间和光谱特征的高效分类。5) DFNet<sup>[19]</sup> 提出的深度特征交互网络可充分利用多源遥感数据的互补性, 深入挖掘多源遥感数据中的自相关和交叉相关特征信息, 增强影响分类任务的关键信息, 以提升分类准确率。

使用相同实验环境、相同的训练样本集和测试样本集以保证实验过程的科学性和实验结果的可靠性。在 Augsburg 数据集和 Berlin 数据集上完成的对比实验结果分别如表 1 和表 2 所示, 内容包括对各地物样本类别的分类准确度评价指标 OA、AA 和 Kappa 系数。对于 Augsburg 数据集, WPANet 取得了最高的 3 类评价指标结果, 并且对于低矮植物和商业区 2 种地物取得了最高的类别样本分类准确度。对于 Berlin 数据集, WPANet 的综合指标和其余分类方法相比, 同样取得了最优的地物分类效果。尤其对于住宅区、工业区、低矮植物和土壤 4 类地物, WPANet 的分类准确度已明显优于其他分类方法, 这是因为基于小波变换的特征提取器可以充分利用频域分析原理挖掘图像的深层特征信息, 同时平行注意力融合器可以完成不同模态之间的高效融合。综和 WPANet 在两种数据集的对比实验结果, 可以看出 WPANet 在多源遥感图像地物分类任务中的优势。在 Augsburg 数据集和 Berlin 数据集上完成的对比实验地物分类可视化结果分别如图 6 和图 7 所示。从图中可以对比看出 WPANet 和其余的 5 种地物分类方法相比, 对不同的地物类别均表现出了较好的地物分类效果, 表明了本文所提出的遥感图像地物分类模型 WPANet 在地物分类任务上的显著优势。

表 1 Augsburg 数据集对比实验结果

Table 1 Comparative experimental results for the Augsburg dataset

No.	类别 (训练/测试样本数)	SVM <sup>[15]</sup>	LBP- ELM <sup>[16]</sup>	TBCNN <sup>[17]</sup>	ContextCNN <sup>[18]</sup>	DFINet <sup>[19]</sup>	WPANet
1	森林 (146/13345)	90.55	93.65	94.77	94.57	<b>95.38</b>	94.81
2	住宅区 (264/30065)	89.81	86.81	95.01	<b>97.25</b>	95.84	93.66
3	工业区 (21/3830)	23.03	35.12	<b>71.17</b>	51.46	69.79	67.52
4	低矮植物 (248/26543)	83.73	83.21	85.33	86.25	86.65	<b>95.49</b>
5	配额地 (52/523)	34.23	49.33	56.41	56.02	<b>64.05</b>	50.10
6	商业区 (7/1632)	9.71	7.94	15.14	13.68	13.86	<b>19.91</b>
7	水域 (23/1502)	<b>45.92</b>	44.99	22.30	21.57	28.47	44.81
	OA (%)	81.60	81.47	87.11	87.24	88.06	<b>90.40</b>
	AA (%)	53.82	57.29	62.87	60.11	64.86	<b>66.61</b>
	Kappa (%)	73.17	73.41	81.69	81.82	82.98	<b>86.28</b>

表 2 Berlin 数据集对比实验结果

Table 2 Comparative experimental results for the Berlin dataset

No.	类别 (训练/测试样本数)	SVM <sup>[15]</sup>	LBP- ELM <sup>[16]</sup>	TBCNN <sup>[17]</sup>	ContextCNN <sup>[18]</sup>	DFINet <sup>[19]</sup>	WPANet
1	森林 (443/54484)	50.08	<b>86.17</b>	76.47	77.22	68.95	69.35
2	住宅区 (423/268219)	61.07	36.95	62.42	63.69	67.52	<b>81.06</b>
3	工业区 (499/19067)	30.68	45.46	43.22	61.44	43.42	<b>62.22</b>
4	低矮植物 (376/58906)	84.29	84.09	78.82	73.77	81.77	<b>85.17</b>
5	土壤 (331/17095)	87.30	89.72	76.33	87.22	75.58	<b>90.47</b>
6	配额地 (280/13025)	54.00	0.00	73.44	<b>82.88</b>	80.05	61.21
7	商业区 (298/24526)	26.61	0.35	<b>49.76</b>	31.13	40.94	25.08
8	水域 (170/6502)	65.40	50.17	<b>82.28</b>	74.24	79.87	80.04
	OA (%)	60.48	48.32	65.81	66.31	67.93	<b>76.23</b>
	AA (%)	57.43	49.25	67.84	68.95	67.26	<b>69.32</b>
	Kappa (%)	45.36	34.65	41.79	54.03	55.22	<b>64.36</b>

## 2.4 消融实验

为了验证本文设计的 Wavelet 提取器和并行注意力融合器的有效性, 分别在 Augsburg 和 Berlin 两个数据集上完成消融实验, 测试两个模块对整体网络架构的影响。消融实验结果如表3所示。从表中可以看出, 基础的卷积特征提取网络在 Augsburg 和 Berlin 两个数据集的OA分别为87.27%和73.86%, 加入Wavelet 提取器后提升效果分别为2.79%和2.70%; 而加入并行注意力融合器后提升效果分别为1.94%和1.77%; 同时加入二者后, OA获得了进一步提升, 相较于原始网络OA提升效果分别为3.59%和3.21%, 这是因为Wavelet 提取器充分利用了频域空间分析原理,

完成了对粗/细粒度级别特征的充分挖掘; 而并行注意力融合器高度关注中间特征信息的一致性和差异性, 可实现对跨层级和跨模态两个层面的特征信息融合补偿, 显著提升分类网络的特征融合能力。

## 3 结论

1) 本文基于小波变换分析理论, 设计并通过编程实现了 WPANet 多源遥感图像分类网络模型。本文提出了基于小波变换的特征提取器, 基于其强大的频域分析能力, 通过中间特征的可逆下采样, 以完成对粗/细粒度级别特征的充分提取, 改善了高光谱图像和 SAR 图像的特征捕捉效果,



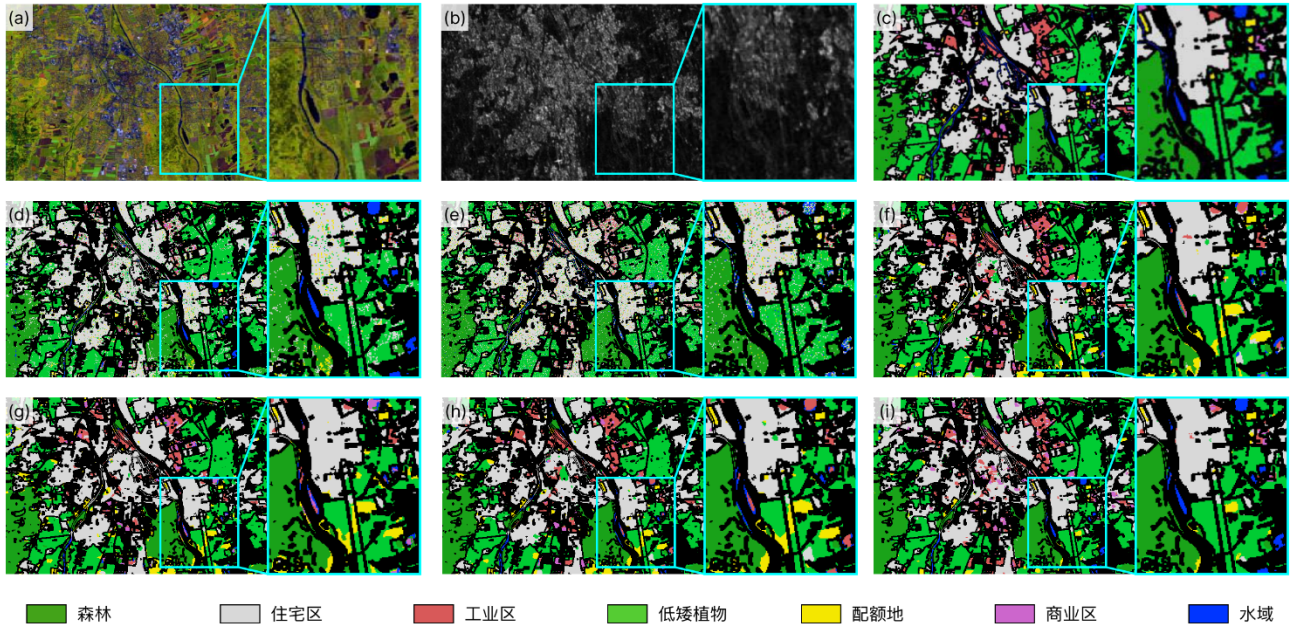


图 6 Augsburg 数据集上不同方法的分类结果。(a) HSI 图像。(a) SAR 图像。(c) 真值图。(d) SVM (81.60%). (e) LBP-ELM (81.47%). (f) TBCNN (87.11%). (g) ContextCNN (87.24%). (h) DFNet (88.06%). (i) WPANet (90.40%).

Fig.6 Classification results of different methods on Augsburg dataset. (a) HSI image. (a) SAR image. (c) True value map. (d) SVM (81.60%). (e) LBP-ELM (81.47%). (f) TBCNN (87.11%). (g) ContextCNN (87.24%). (h) DFNet (88.06%). (i) WPANet (90.40%).

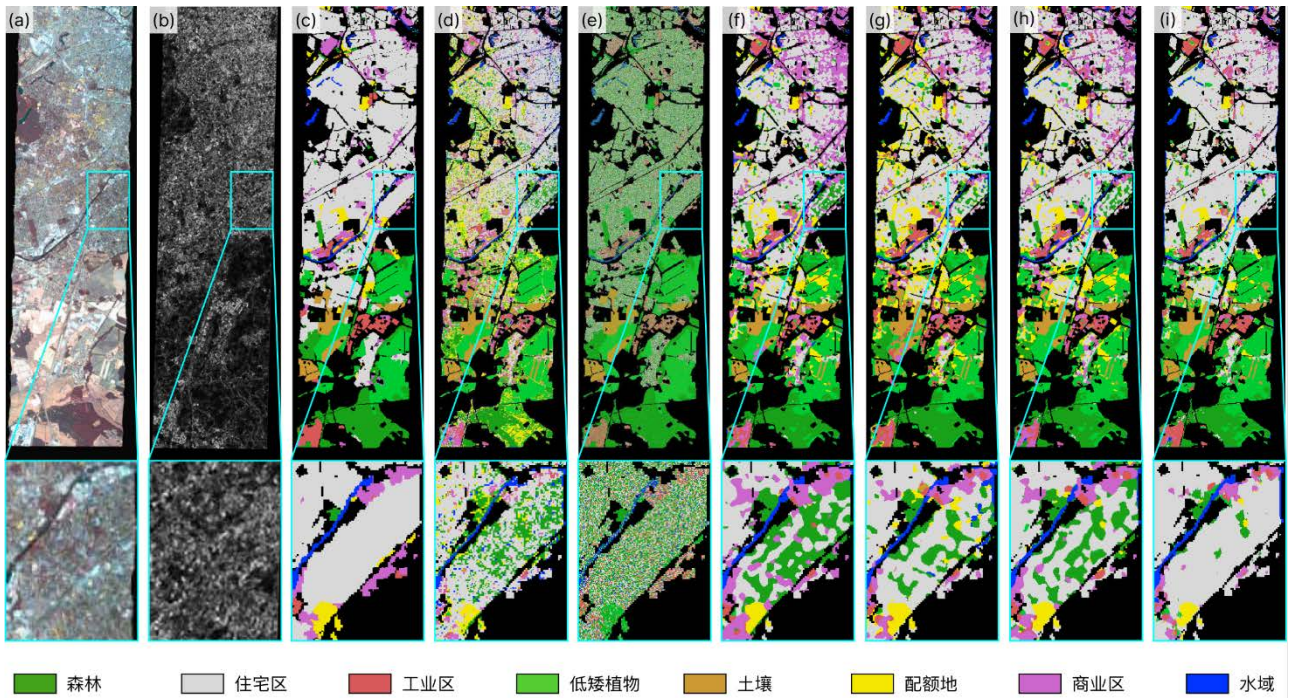


图 7 Berlin 数据集上不同方法的分类结果。(a) HSI 图像。(a) SAR 图像。(c) 真值图。(d) SVM (60.48%). (e) LBP-ELM (48.32%). (f) TBCNN (65.81%). (g) ContextCNN (66.31%). (h) DFNet (67.93%). (i) WPANet (76.23%).

Fig.7 Classification results of different methods on the Berlin dataset. (a) HSI image. (a) SAR image. (c) True value map. (d) SVM (60.48%). (e) LBP-ELM (48.32%). (f) TBCNN (65.81%). (g) ContextCNN (66.31%). (h) DFNet (67.93%). (i) WPANet (76.23%).

对多源遥感图像地物分类准确率有所提升。

2) 基于平行注意力机制的特征融合器通过结合跨层级和跨模态两个维度的融合能力，充分挖掘高光谱图像特征和 SAR 图像特征的一致性

和差异性，改善不同模态图像特征的融合效果和分类效果。

3) 本文在 Augsburg 和 Berlin 两个不同的数据集上通过 OA、AA 和 Kappa 系数3个模型评价指标对实验结果完成评价和分析，结合地物

分类可视化结果,从定量和定性分析两个角度综合表明了本文所提出的 WPANet 的有效性和可扩展能力。

4) 本文仅实现了高光谱和 SAR 两种模态图像的融合分类,可继续尝试更多模态(如多光谱图像、热红外图像等)遥感图像的融合分类,更进一步地研究其他模态图像和高光谱图像、SAR 图像之间的融合分类方法,以扩展 WPANet 的应用场景,在遥感图像处理领域为社会带来更多的实际应用价值。

**表 3 Wavelet 提取器和并行注意力融合器消融实验结果**  
**Table 3 Results of ablation experiments with Wavelet extractor and parallel attention fuser**

网络结构	OA (%)	
	Augsburg	Berlin
卷积特征提取网络	87.27	73.86
+ Wavelet提取器	89.71	75.86
+ 并行注意力融合器	88.96	75.17
Wavelet提取器+并行注意力融合器	<b>90.40</b>	<b>76.23</b>

### 参考文献 (References)

- [1] WANG C, ZHANG L, WEI W, et al. Dynamic super-pixel normalization for robust hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.
- [2] BIOUCAS-DIAS J M, PLAZA A, CAMPS-VALLS G, et al. Hyperspectral remote sensing data analysis and future challenges[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(2): 6-36.
- [3] MOREIRA A, PRATS-IRAOLA P, YOUNIS M, et al. A tutorial on synthetic aperture radar[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1 (1): 6-43.
- [4] MAN Q, DONG P, GUO H. Pixel-and feature-level fusion of hyperspectral and lidar data for urban land-use classification[J]. International Journal of Remote Sensing, 2015, 36(6): 1618-1644.
- [5] HU J, GHAMISI P, SCHMITT A, et al. Object based fusion of polarimetric sar and hyperspectral imaging for land use classification[C]//2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). 2016: 1-5.
- [6] CHEN Y, LI C, GHAMISI P, et al. Deep fusion of remote sensing data for accurate classification[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1253-1257.
- [7] HONG D, GAO L, HANG R, et al. Deep encoder-decoder networks for classification of hyperspectral and LiDAR data[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19:1-5.
- [8] ZHAO X, TAO R, LI W, et al. Joint classification of hyperspectral and lidar data using hierarchical random walk and deep cnn architecture[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10): 7355-7370.
- [9] FENG M, GAO F, FANG J, et al. Hyperspectral and lidar data classification based on linear self-attention[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. 2021: 2401-2404.
- [10] LI W, GAO Y, ZHANG M, et al. Asymmetric feature fusion network for hyperspectral and sar image classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022: 1-14.
- [11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018 : 7132-7141.
- [12] KURZ F, ROSENBAUM D, LEITLOFF J, et al. Real time camera system for disaster and traffic monitoring[C]//Proceedings of International Conference on SMPR 2011. 2011: 1-6.
- [13] BAUMGARTNER A, GEGE P, KÖHLER C, et al. Characterisation methods for the hyperspectral sensor hypspec at dlr's calibration home base[C]//Sensors, Systems, and Next-Generation Satellites XVI: SPIE, 2012, 8533: 371-378.
- [14] HONG D, HU J, YAO J, et al. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 178: 68-80.
- [15] MELGANI F, BRUZZONE L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778-1790.
- [16] LI W, CHEN C, SU H, et al. Local binary patterns and extreme learning machine for hyperspectral imagery classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(7): 3681-3693.
- [17] XU X, LI W, RAN Q, et al. Multisource remote sensing data classification based on convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(2): 937-949.
- [18] LEE H, KWON H. Going deeper with contextual CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4843-4855.
- [19] GAO Y, LI W, ZHANG M, et al. Hyperspectral and multispectral classification for coastal wetland using depthwise feature interaction network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.

### 作者简介:

**王嘉毅** 男, 硕士研究生。主要研究方向: 深度学习、计算机视觉。

**高峰** 男, 博士, 副教授, 硕士生导师。主要研究方向: 遥感数据智能处理, 多媒体计算。



# Multi-Source Remote Sensing Image Classification Based on Wavelet Transform and Parallel Attention

WANG Jiayi, GAO Feng\*, ZHANG Tiange GAN Yanhai

(School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China)

**Abstract:** Exploring the dependency relationships of multi-source remote sensing image data features to fully exploit the complementary advantages between different modalities has become a prominent research direction in the field of remote sensing. Existing joint classification tasks of hyperspectral and SAR data face two key challenges: 1) Insufficient feature extraction and representation in images, resulting in the loss of high-frequency information, which hinders subsequent classification tasks; 2) Limited interaction among multi-source image features and weak correlation between multimodal features. To address these challenges, research work has been conducted on robust feature representation and efficient correlation of multi-source features, proposing a multi-source remote sensing image classification method based on wavelet transform and parallel attention mechanism. The feature extractor based on wavelet transform can effectively utilize frequency domain analysis techniques, capturing coarse- and fine-grained level features during the process of reversible downsampling. The feature fusion module based on the parallel attention mechanism comprehensively integrates the consistency and differences of multimodal remote sensing data, accomplishing the fusion and generation of highly correlated features to enhance classification accuracy. Experimental results on two real multi-source remote sensing datasets demonstrate the significant advantages of the proposed classification method. The overall accuracy on the Augsburg and Berlin datasets reaches 90.40% and 76.23% respectively, with at least a 2.66% and 12.22% improvement in overall accuracy compared to mainstream methods like DFNet on the two datasets.

**Keywords:** Hyperspectral images(HSI); Synthetic aperture radar(SAR); Wavelet transform; Multi-source feature fusion; Remote sensing images

---

**Received:** 2023-06-01; **Accepted:** 2023-xx-xx; **Published online:** 2023-xx-xx xx:xx

**URL:**

**Foundation item:** National Natural Science Foundation of China (No. 42106191); National Key R&D Program of China (No. 2022ZD0117202)

\***Corresponding author.** Tel.: 0532-66781123 E-mail: gaofeng@ouc.edu.cn