

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2025)03-0883-12

论文引用格式: Jin X P, Gao F, Shi X C and Dong J Y. 2025. Gated cross-modal aggregation network for multi-source remote sensing data classification. Journal of Image and Graphics, 30(3):0883-0894(金学鹏, 高峰, 石晓晨, 董军宇. 2025. 针对多源遥感图像分类的门控跨模态聚合网络. 中国图象图形学报, 30(3):0883-0894)[DOI:10.11834/jig.240359]

针对多源遥感图像分类的门控跨模态聚合网络

金学鹏, 高峰*, 石晓晨, 董军宇

中国海洋大学计算机科学与技术学院, 青岛 266100

摘要: 目的 为了突破单一传感器的技术限制并弥补单一数据源应用的局限性, 多源遥感数据融合成为了遥感应用领域的研究热点。当前的高光谱图像与激光雷达(light detection and ranging, LiDAR)/合成孔径雷达(synthetic aperture radar, SAR)数据融合分类方法未能充分利用高光谱图像的光谱特征以及 LiDAR/SAR 数据的地物结构信息。由于不同成像模态的图像在数据特性上存在本质差异, 这些差异为多源图像特征的关联带来了重大挑战。尽管采用深度学习技术的一些方法在结合高光谱与 LiDAR/SAR 数据进行分类的任务中显示出了优秀的结果, 但它们在融合过程中未能充分利用多源数据中的纹理信息和几何信息。**方法** 为了应对这一关键问题, 提出了一种基于门控注意力聚合网络的多源遥感图像分类方法, 可以更加全面地挖掘多源数据中的互补信息。首先, 设计了一个门控跨模态聚合模块, 利用交叉注意力特征融合将 LiDAR/SAR 数据中的地物精细结构信息与高光谱图像特征有机融合。然后, 使用精细化的门控模块将关键的 LiDAR/SAR 特征集成到高光谱图像特征中, 从而增强多源数据的融合效果。**结果** 在 Houston2013 和 Augsburg 数据集上与 7 种主流方法进行实验比较, 在总体精度(overall accuracy, OA)、平均精度(average accuracy, AA)和卡帕系数(Kappa coefficient, Kappa)指标上都取得了最优表现。特别是在 Augsburg 数据集中, 本文方法在大多数类别上均取得了最佳指标。在分类的可视化结果中可以明显看出, 本文所提出的分类模型在性能上具有显著优势。**结论** 实验结果表明, 本文所提出的 GCA-Net(gated cross-modal aggregation network)具有优异的性能, 显著优于 HCT(hierarchical CNN and Transformer)、MACN(mixing self-attention and convolutional network)等主流方法。该方法能够根据不同模态的特点充分融合不同模态的信息进行分类, 为多源遥感数据的融合分类提供了理论支持。

关键词: 高光谱图像(HSI); 激光雷达(LiDAR); 合成孔径雷达(SAR); 后向散射信息; 多源特征融合

Gated cross-modal aggregation network for multi-source remote sensing data classification

Jin Xuepeng, Gao Feng*, Shi Xiaochen, Dong Junyu

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

Abstract: Objective In recent years, multisource remote sensing data fusion has emerged as a research hotspot in the field of remote sensing applications. This trend aims to overcome the technical limitations of single sensors and address the constraints associated with relying on a single data source. Traditional remote sensing methods, which often depend on a single type of sensor, face considerable challenges in providing comprehensive and accurate information due to the inherent limita-

收稿日期: 2024-06-27; 修回日期: 2024-09-04; 预印本日期: 2024-09-11

* 通信作者: 高峰 gaofeng@ouc.edu.cn

基金项目: 新一代人工智能国家科技重大专项(2022ZD0117202); 国家自然科学基金项目(42106191)

Supported by: National Science and Technology Major Project(2022ZD0117202); National Natural Science Foundation of China(42106191)

tions of the sensors. For instance, hyperspectral sensors capture detailed spectral information but may lack spatial resolution, while LiDAR (light detection and ranging) and SAR (synthetic aperture radar) sensors excel in capturing structural information but fail to provide sufficient spectral details. The integration of hyperspectral images and LiDAR/SAR data holds remarkable promise for enhancing remote sensing applications. However, current methods for fusion classification have not fully utilized the rich spectral features of hyperspectral images and the structural information of ground objects provided by LiDAR/SAR data. The two types of data have fundamentally different characteristics, which pose considerable challenges for effective feature correlation. Hyperspectral images contain abundant spectral information that can identify different materials, while LiDAR provides 3D structural information, and SAR offers high-resolution imaging under various weather conditions. The differences in data characteristics among these imaging modalities create substantial challenges in correlating multisource image features. Although some deep learning-based methods have shown promising results in the fusion classification tasks of hyperspectral and LiDAR/SAR data, they often fall short in fully exploiting the texture and geometric information embedded within the multisource data during the fusion process. These methods may perform well in specific scenarios but often lack the necessary robustness and versatility for broader applications. Consequently, highly sophisticated approaches that can leverage the complementary strengths of different data sources are urgently needed to improve classification accuracy and reliability.

Method This paper proposes a novel multisource remote sensing image classification method based on a gated cross-modal aggregation network (GCA-Net) to address this critical issue. This approach aims to comprehensively exploit the complementary information available in multisource data. The core innovation of the proposed method lies in its capability to effectively integrate the unique advantages of different sensor data types through a series of advanced neural network modules. First, a gated cross-modal aggregation module is designed to facilitate the organic integration of detailed structural information from LiDAR/SAR data with the spectral features of hyperspectral images. This module leverages cross-attention mechanisms to enhance the feature fusion process, ensuring that the distinctive information from each data source is effectively utilized. The cross-attention feature fusion mechanism allows the model to focus on the most relevant features from each modality, enhancing the overall representation of the fused data. Second, a refined gating module is employed to integrate valuable LiDAR/SAR features into the hyperspectral image features. This gating mechanism selectively incorporates the most informative features, thereby enhancing the fusion effect of multisource data. The gating module acts as a filter that prioritizes the integration of features that provide the most contribution to the classification task, ensuring that the fused data retains the critical information from both sources. This approach not only improves the accuracy of the classification but also enhances the robustness of the model across different datasets and scenarios.

Result The proposed method was rigorously tested through experiments conducted on two widely recognized datasets: Houston2013 and Augsburg. These datasets provide a diverse range of scenes and conditions, making them ideal for evaluating the effectiveness of the proposed method. The performance of the GCA-Net was compared with seven mainstream methods to assess its efficacy. Experimental results demonstrated that the proposed method achieved the best performance in terms of overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). Specifically, the GCA-Net outperformed the other methods by a substantial margin, highlighting its superior capability to fuse and classify multisource remote sensing data. Particularly, on the Augsburg dataset, the proposed method achieved the best indicators across most categories. This dataset includes a variety of urban and rural scenes, challenging the classification models with diverse textures and structures. The capability of the GCA-Net to handle such diversity showcases its robustness and versatility. Moreover, the classification visualization results show that the proposed classification model has a notable performance advantage. The visualized results provide a clear and intuitive representation of the classification accuracy, with the GCA-Net producing more precise and consistent classifications compared to other methods. This visual evidence emphasizes the practical benefits of the proposed method, demonstrating its potential for real-world applications in remote sensing.

Conclusion The experimental results on the Houston2013 and Augsburg datasets strongly support the excellent performance of the proposed GCA-Net, drastically surpassing current mainstream methods such as HCT (hierarchical CNN and Transformer) and MACN (mixing self-attention and convolutional network). A key factor in the success of the GCA-Net lies in its capability to fully integrate information from different modalities based on their unique characteristics. This integration allows the model to leverage the strengths of each data source, resulting in a highly accurate and reliable classification.

The superior performance of the proposed method can be attributed to its innovative use of gated attention mechanisms and cross-modal feature fusion. These techniques enable the model to selectively focus on the most relevant features from each modality, enhancing the overall quality of the fused data. The GCA-Net sets a new benchmark for multisource remote sensing data fusion by effectively combining the spectral richness of hyperspectral images with the structural details of LiDAR/SAR data. Overall, the GCA-Net provides a robust theoretical foundation for the fusion classification of multisource remote sensing data. The capability of this network to address the limitations of single-sensor approaches and leverage the complementary strengths of different data types increases its value as a tool for advancing remote sensing applications. The proposed method not only improves classification accuracy but also enhances the practical applicability of remote sensing technologies, paving the way for highly sophisticated and effective solutions in various domains, such as environmental monitoring, urban planning, and disaster management.

Key words: hyperspectral image (HSI); light detection and ranging (LiDAR); synthetic aperture radar (SAR); back scattering; multi-source feature fusion

0 引言

高分辨率卫星遥感图像在土地覆盖分类中的应用受到国内外学者的广泛关注(Arshad和Zhang, 2024)。然而,由于传感器的性能限制和数据传输带宽的约束,它们在实现高精度的空间细节和丰富的光谱信息方面面临着挑战(高峰等,2024)。因此,通过集成高光谱、合成孔径雷达(synthetic aperture radar, SAR)和激光雷达(light detection and ranging, LiDAR)等不同传感器的信息,可以显著提高地物识别和分类的准确性(Zhao等,2023;Wang等,2021)。

高光谱传感器(胡帅等,2024)能够捕获数百个甚至上千个连续的光谱带,提供关于地表物质和植被类型的丰富信息(Bioucas-dias等,2013)。SAR传感器(肖振久等,2024)则不受天气条件和光照限制,能够提供地物详细的结构、材质信息,如地表粗糙度、地形和建筑物结构等(Moreira等,2013)。LiDAR传感器通过测量从地面反射回来的激光脉冲的时间,能够提供高精度的地表高程信息(Roy等,2024)。尽管每种传感器都有其独特的优势,但它们也存在一定的局限性。例如,高光谱图像(hyperspectral image, HSI)的光谱特征可能会因大气条件的变化受到影响,而SAR图像则容易受到相干斑噪声的干扰,影响其图像质量。同样,茂密的植被可能会阻挡LiDAR直接触及地面,从而对地面高程的精准测量造成影响。

通过多源遥感图像融合分类,可以将不同数据源的优势结合起来,弥补各自的不足。例如,SAR和LiDAR数据在恶劣天气条件下能辅助高光谱传感器

进行分类,而高光谱数据丰富的光谱信息则能增强SAR和LiDAR数据的分类能力。

考虑到不同成像技术产生的图像在数据特性上具有根本性的区别,这些区别在关联多源图像的特征时带来了显著的困难。为了应对多源遥感图像分类任务,国内外众多学者已经提出了一系列创新性的方法。Man等人(2015)的研究主要集中在高光谱与LiDAR数据的融合与分类上。在他们的研究中,通过像素级的特征整合,构建了一个分类器,该分类器融合了极大似然法和支持向量机等方法的优势。这种方法的优势在于能够有效地融合不同成像模态的数据,但可能受限于特征融合的精度和效率。Chen等人(2017)使用了一个全连接网络来融合由双分支卷积神经网络(convolutional neural network, CNN)获得的异构特征。然而,全连接网络可能存在过拟合和计算复杂度高的问题。Hong等人(2022)采用深度编码器—解码器网络架构,用于高光谱和LiDAR数据的分类,通过强制融合的特征重构多模态输入来融合多模态信息,更好地激活跨模态的神经元,更好地实现多源遥感图像分类。Zhao等人(2020)在多源遥感图像融合分类领域提出了一种高效的模型HRWN(hierarchical-random-walk-network),该模型利用双隧道卷积神经网络,实现了对空间和光谱特征的全面学习,确保了全局特征信息的充分利用。Li等人(2022)在多源遥感图像分类的研究中,创新性地设计并实现了一个图特征提取模块。该模块致力于捕捉并整合拓扑结构信息与丰富的光谱—空间信息。在此过程中,初始通过卷积操作得到的特征被映射至图特征空间,随后利用图卷积技术在节点间进行信息传播,确保了拓扑结构

信息得以保留和利用。Ma 等人(2024)提出了一个创新的跨源交互增强融合模块,旨在处理多级多源遥感特征。该模块通过强化内源特征间的相互依赖性,并实施源间的增强型融合策略,有效提炼了内源相关性特征,并显著缓解了不同源之间的语义差异问题。

尽管这些技术在结合高光谱数据和 LiDAR/SAR 数据进行分类时显示出了良好的性能,但它们在融合过程中未能充分利用多源数据中的纹理信息和几何信息,未能根据不同数据模态的特性进行动态调整,限制了多源数据融合的灵活性和有效性。此外,现有方法在挖掘高光谱与 LiDAR/SAR 数据的互补性方面存在不足,未能充分利用两者之间的协同信息,从而导致分类性能的提升有限。为解决这一问题,本文提出了基于门控注意力的聚合网络(gated cross-modal aggregation network, GCA-Net),可以更好地挖掘隐藏在光谱和 LiDAR/SAR 数据中的多模态信息。具体而言,本文设计了一个门控跨模态聚合模块(gated cross-modal aggregation module, GCAM),如图 1 所示。在每个 GCAM 中,利用交叉注意力特征融合将 LiDAR/SAR 数据中的地物精细结构信息与高光谱图像特征进行有效整合。接下来,通过细化门控模块(refinement gating)将关键的融合特征集成到光谱特征中。在 Houston2013 和 Augsburg 多源遥感数据集上的实验结果表明,本文所提出的 GCA-Net 具有优异的性能,显著优于层次卷积神经网络和 Transformer(hierarchical CNN and Transformer, HCT)、混合自注意力机制和卷积网络(mixing self-attention and convolutional network, MACN)等当前主流方法。

本文的主要贡献如下:1)提出了交叉注意力特征融合与细化门控机制。通过交叉注意力机制,实现光谱的光谱特征与 LiDAR/SAR 数据的空间特征之间的深度对齐和互补信息的挖掘。在此基础上,细化门控模块进一步优化了特征表示,通过动态调整不同特征的重要性,强化了模型对关键信息的捕捉能力。2)设计并实现了一个高效的高光谱图像特征编码器,具备同时捕捉局部细节和全局上下文信息的能力。该编码器使得模型不仅能够识别出图像中的局部模式和细微差别,还能够理解局部特征在整个场景中的上下文关联关系。3)在 Houston2013 和 Augsburg 这两个多源遥感数据集上进行

了广泛的实验。这些实验从定量和定性两个维度验证了所提出的 GCA-Net 网络的卓越性能。

1 本文方法

本文提出的 GCA-Net 架构,专为高光谱图像与 LiDAR/SAR 数据的分类而设计,其结构如图 1 所示。该框架包含 3 个门控跨模态聚合模块(GCAM),专注于跨模态特征的深度提取,并配合一个全连接层来实现地物精确分类。每个 GCAM 模块都集成了高光谱和 LiDAR/SAR 的特征编码器,即 HSI encoder 与 LiDAR/SAR encoder,通过交叉注意力机制 Cross-attention feature fusion 实现特征的紧密融合,并通过细化门控模块进一步提升特征的表达精度。

1.1 高光谱图像特征编码器

高光谱图像编码器的详细结构如图 1 所示。首先,输入数据 X 通过 $3 \times 3 \times 3$ 的 3D 卷积提取初步特征,这个初步特征表示为 $X_f \in \mathbf{R}^{C \times H \times W}$ 。然后,将特征输入到两个并行分支中:卷积分支和自注意力分支。对于卷积分支,使用 3×3 卷积和深度可分离卷积进行局部特征提取,得到的特征为 X_{conv} 。这种组合设计旨在综合利用常规卷积的特征提取能力和深度可分离卷积的高效性,从而在保证模型性能的同时,降低参数量和计算量。对于另一个自注意力分支,将初步特征 X_f 分成 3 部分:查询 Q 、键 K 和值 V ,以建模全局特征交互。上述两个分支具体为

$$X_f = \text{Conv}(X) \quad (1)$$

$$X_{\text{conv}} = \text{DWConv}(\text{Conv}(X_f)) \quad (2)$$

$$X_{\text{att}} = f_{\text{softmax}}\left(\frac{QK^T}{\tau}\right)V \quad (3)$$

式中, $\text{Conv}(\cdot)$ 表示卷积, $\text{DWConv}(\cdot)$ 表示深度可分离卷积。 Q 、 K 和 $V \in \mathbf{R}^{C \times H \times W}$ 分别表示查询、键和值。 τ 是一个可学习的超参数。 softmax 函数是用于多类分类问题的激活函数。最后,两个路径的输出 X_H 通过逐元素求和进行融合,即

$$X_H = X_{\text{conv}} + X_{\text{att}} \quad (4)$$

在对高光谱图像特征编码的过程中,通过卷积分支捕捉到细粒度的局部特征。而自注意力机制则能够建模全局依赖关系,通过查询、键和值的交互过程来捕捉长距离特征的相互作用。这种结合使得高光谱特征编码器能够同时学习局部细节和全局上下文信息。两个分支的并行处理并通过最终的相加操

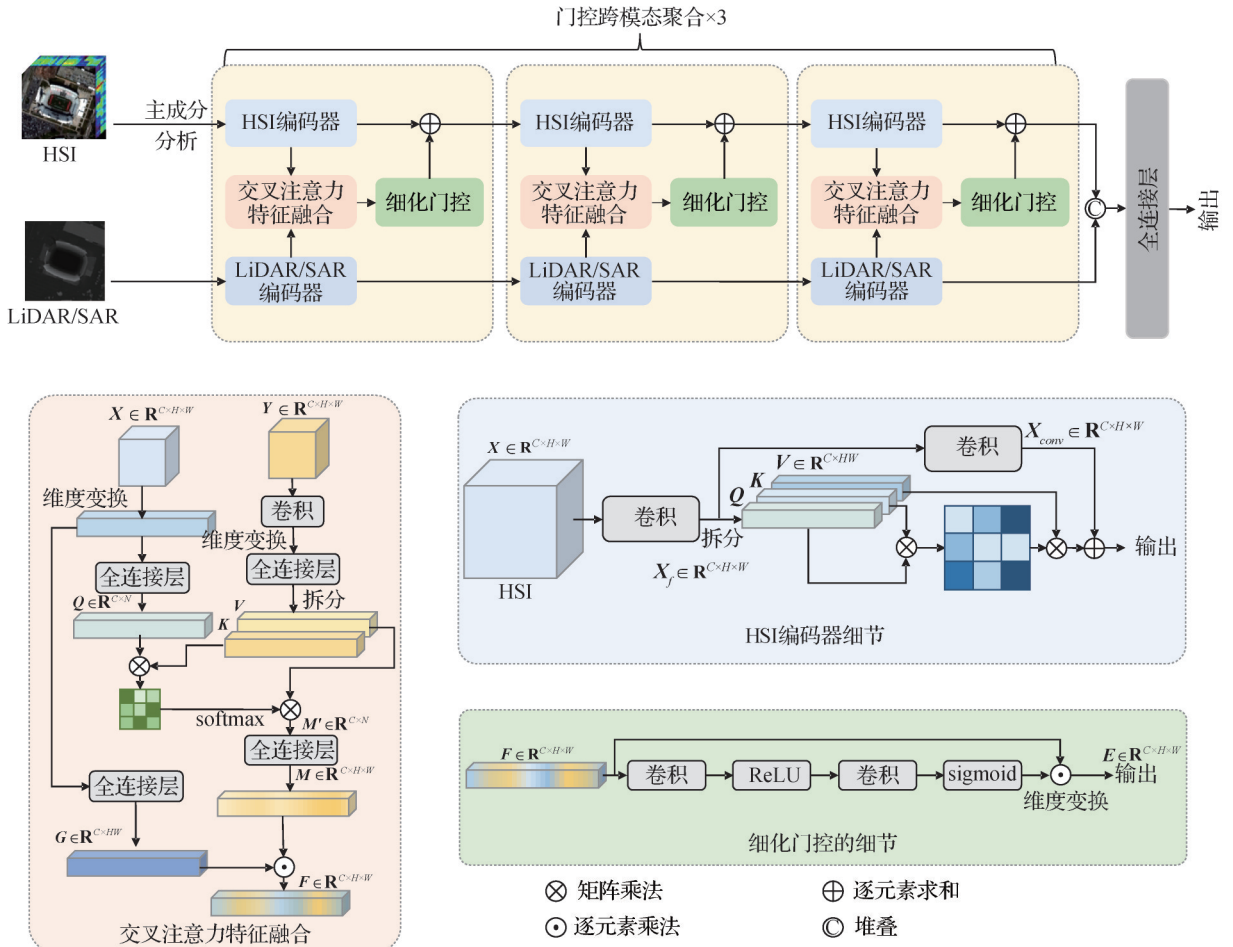


图1 基于门控注意力的聚合网络

Fig. 1 Gated cross-modal aggregation network

作实现了特征的互补,这样可以提升模型对复杂场景的识别能力,从而提高网络的分类性能。对于LiDAR/SAR特征编码器,本文简单地使用两个2D卷积进行特征提取。

1.2 交叉注意力的特征融合

交叉注意力特征融合模块 Cross-attention feature fusion 的核心目标是实现高光谱特征与 LiDAR/SAR 特征的有效对齐,并将这些特征整合到一个统一的特征空间中,以便充分挖掘和利用两种模态之间的互补信息。在这个过程中,本文特别关注从 LiDAR 数据中提取的高程特征,以及从 SAR 数据中提取的后向散射特征。通过交叉注意力机制,构建了跨模态之间的交互模型,这不仅促进了特征间的相互融合,还实现了对高程或后向散射特征的上下文编码,从而增强了最终特征的表达能力。

图1显示了交叉注意力特征融合的信息。给定输入的高光谱特征 $X \in \mathbb{R}^{C \times H \times W}$ 和 LiDAR/SAR

特征 $Y \in \mathbb{R}^{C \times H \times W}$,该模块通过两个步骤对它们进行聚合:

在第1步中,使用卷积来提取 LiDAR/SAR 的局部特征。这一步生成一系列空间特征图, $M \in \mathbb{R}^{C \times H \times W}$ 。具体为

$$Y' = \text{Reshape}(\text{Conv}(Y)) \quad (5)$$

$$Q = W_q(\text{Reshape}(X)) \quad (6)$$

$$K = W_k(Y') \quad (7)$$

$$V = W_v(Y') \quad (8)$$

$$M' = f_{\text{softmax}}\left(\frac{QK^T}{\sqrt{C}}\right)V \quad (9)$$

$$M = W_m(\text{Reshape}(M')) \quad (10)$$

式中, W_q 、 W_k 、 W_v 和 W_m 是投影层的权重, Reshape 操作用来改变特征维度, C 是 V 向量的维度。

在第2步中,在获得空间特征图 M 之后,通过逐元素乘法将 M 和 X 结合起来,得到最终输出的融合特征 F ,具体为

$$G = W_g(\text{Reshape}(X)) \quad (11)$$

$$F = W_f(G \odot M) \quad (12)$$

式中, $\odot$ 表示逐元素乘法, W_g 和 W_f 是投影层的权重。

1.3 细化门控

本文旨在充分发挥 LiDAR/SAR 在空间信息方面的强项以及高光谱图像在光谱信息方面的优势, 为此, 引入了一个细化门控模块来细化融合后的特征。该模块通过融合特征 F 学习逐元素权重参数, 对特征进行重新标定, 确保 LiDAR/SAR 的信息能够自适应地融入到后续处理过程中。通过这种门控机制, 能有效地提升对任务关键特征的识别能力, 它能够强化最有助于当前任务的特征, 并抑制那些可能引起信息冗余或不相关的特征, 从而使得模型在处理多源遥感数据时更加精确和高效。

图 1 显示了细化门控模块的详细结构, 最终的输出结果 E 的计算式为

$$W = f_{\text{sigmoid}}(\text{Conv}(\text{ReLU}(\text{Conv}(F)))) \quad (13)$$

$$E = \text{Rehsape}(W \odot F) \quad (14)$$

式中, $\odot$ 表示逐元素乘法, $\text{Conv}(\cdot)$ 表示卷积层。输入特征 F 经过两个卷积层处理, 第 1 层是一个 3×3 卷积, 后跟 ReLU (rectified linear unit) 激活函数, 第 2 层是一个 1×1 卷积, 后跟 sigmoid 激活函数。

2 实验结果与分析

2.1 数据集

1) Houston2013 数据集。作为 2013 年 IEEE 地球科学与遥感学会数据融合大赛的一部分, 提供了一种独特的视角来观察美国德克萨斯州休斯敦市及其周边地区的城市土地覆盖情况。数据集以 2.5 m 的空间分辨率, 记录覆盖 380~1 050 nm 光谱范围的 144 个波段的高光谱图像, 以及提供精确海拔高度信息的 LiDAR 数据, 涵盖了 15 种不同的土地覆盖类型。

2) Augsburg 数据集。包括了高光谱和双极化 SAR 图像(Baumgartner 等, 2012), 详细记录了该地区的多种土地覆盖类型。数据集的图像尺寸为 332 × 485 像素, 其核心是一个 180 波段的高光谱图像, 光谱覆盖为 0.4~2.5 μm 。此外, 数据集还包括了垂直极化(VV)和垂直—水平极化(VH)的 SAR 数据。

这些数据共同提供了对德国奥格斯堡市土地覆盖情况的全面描述。

2.2 实验设置

实验环境包括 32 GB 内存、Intel Xeon E5-2678 v3 @ 2.50 GHz CPU (central processing unit) 以及 NVIDIA GeForce RTX 2080Ti GPU (11 GB VRAM)。选择 Adam 优化器, 并将初始学习率设定为 0.000 1。Adam 优化器结合了动量和自适应学习率调整的优点, 能够有效应对稀疏梯度和非平稳目标。设定的初始学习率有助于稳定训练过程, 防止模型在早期阶段出现较大振荡, 确保渐进式优化。与大多数多源遥感图像分类方法类似, 本文方法对输入数据以每个像素点为中心裁剪为 11×11 大小的块进行处理。batchsize 设置为 128, 训练周期为 200 个 epoch。较大的 batchsize 能够充分利用 GPU (graphics processing unit) 的并行计算能力, 同时也有助于梯度估计的稳定性。总体而言, 本文的实验设置确保了高光谱图像分类任务的有效性和可靠性。

2.3 模型参数分析

使用 Houston2013 和 Augsburg 数据集来研究高光谱图像的光谱特征降维对图像分类效果的影响。Houston2013 数据集包含 144 个光谱波段, Augsburg 数据集包含 180 个光谱波段。为了提升神经网络的计算效率, 采用主成分分析 (principal component analysis, PCA) 方法对高光谱图像进行降维处理, 保留关键特征信息。

为了分析高光谱图像的主成分数对分类性能的影响, 在 Houston2013 和 Augsburg 数据集上进行实验, 采用总体分类准确度 (overall accuracy, OA) 作为衡量标准。实验中的主成分数 n 从集合 {20, 25, 30, 35, 40} 中取值, 从图 2 中可以看出 n 与 OA 之间的关系。这种选择基于对高光谱图像数据的经验和先验知识。根据以往的研究和经验, 发现保留 20~40 个主成分通常能够覆盖数据中大部分的变化和差异。这些主成分能够包含数据中的主要特征, 既能有效减少数据维度, 降低计算负担, 又不至于丢失关键信息。因此, 选择这些值作为主成分数的取值范围, 以确保在减少数据维度的同时, 仍能保持良好的分类效果。选择固定的几个主成分数值, 有助于设计实验并进行对比分析。这些值的间隔为 5, 可以观察到主成分数逐步增加时分类准确度的变化, 从而更直观地了解 n 与 OA 之间的关系。

从图2可以看出,当 $n = 30$ 时,GCA-Net模型的性能达到最优。当 $n < 30$ 时, n 与OA之间存在正向的相关性。然而,当 $n > 30$ 后,观察到OA并没有显著增加,反而在某些情况下,模型的准确度有所下降。基于上述分析,综合考虑模型性能与参数 n 之间的关系,本文决定将高光谱数据的主成分数设定为30。

在本实验中,选择30个主成分能够在不丢失关键信息的前提下有效降低数据维度,提高计算效率。合理选择主成分数能够显著提升模型的性能,但过多的主成分可能会引入噪声,导致性能下降。

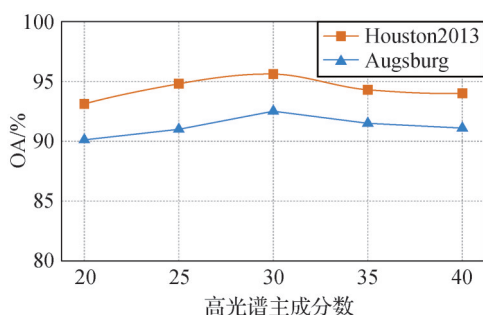


图2 高光谱图像中主成分数对总体准确率的影响

Fig. 2 The impact of the number of principal components on the overall accuracy in hyperspectral images

2.4 对比实验分析

为了深入评估本文提出的地物分类模型 GCA-Net 的性能,选择支持向量机 (support vector machines, SVM)、双分支卷积神经网络 (two-branch CNN, TBCNN)、空间光谱跨模态增强网络 (spatial-spectral cross-modal enhancement network, S²ENet)、非对称特征融合网络 (asymmetric feature fusion network, AsyFFNet)、拓展视觉 Transformer (extended vision Transformer, ExViT)、HCT 和 MACN 共 7 种不同的遥感图像地物分类方法与 GCA-Net 进行性能对比实验。这些方法的细节如下:

1) SVM (Melgani 和 Bruzzone, 2004) 通过映射数据样本至更高维度的空间,并在此空间中探寻一个最优的分类边界,以此来达到分类任务的目标。

2) TBCNN (Xu 等, 2018) 能够高效地处理多源遥感图像的特征提取工作。该方法首先在特征融合阶段融合了高光谱图像的光谱和空间特征,然后将有效地将这些特征与来自其他传感器的数据特征结合起来,以实现多源遥感数据的全面特征整合。

3) S²ENet (Fang 等, 2022) 提出了一个名为空间光谱增强模块的方法,旨在深度神经网络中实现多模态遥感数据 (如高光谱和 LiDAR 数据) 的跨模态信息交互。空间光谱增强模块包含两个子模块:空间增强模块用于通过 LiDAR 特征增强高光谱数据的空间表示;光谱增强模块用于通过高光谱特征增强 LiDAR 数据的光谱表示。

4) AsyFFNet (Li 等, 2023b) 提出了一种非对称特征融合网络 (AsyFFNet) 多源遥感数据联合分类方法。该方法通过使用权重共享的残差块进行特征提取,并保持独立的批量归一化层来处理不同数据源间的信息不平衡问题。在训练过程中,通过 BN (batch normalization) 层中的缩放因子自适应地确定并消除冗余通道,同时施加稀疏约束以提升模型的泛化能力。此外,引入了特征校准模块来增强多源特征的空间依赖性,进一步提升了分类的区分能力。

5) ExViT (Yao 等, 2023) 提出了一个多模态深度学习框架,它通过对传统的 ViT (vision Transformer) 进行修改来处理土地利用和土地覆盖分类任务。该框架通过并行分支处理多模态遥感图像数据,并引入了跨模态注意力模块来增强不同模态间的信息交换。

6) HCT (Zhao 等, 2023) 通过结合层次卷积神经网络和 Transformer 网络来捕捉和学习 HSI 数据的光谱—空间特征以及 LiDAR 数据的高程特征,然后通过交叉标记注意 (cross-token attention, CTA) 融合编码器进行数据融合。这种方法能够有效融合多源异构信息,并提高联合分类性能。

7) MACN (Li 等, 2023a) 利用两个自适应 CNN 编码器从多源遥感数据中提取浅层卷积特征。通过自注意力和卷积的巧妙融合,提出的混合自注意力和卷积 Transformer 层实现了局部和全局多尺度感知。最后,设计了一个多源交叉引导融合模块,用于深度融合多源遥感数据特征。

为了保证实验过程的科学性和实验结果的可靠性,使用了相同的实验环境、相同的训练样本集和测试样本集。对比实验的结果展示在表1和表2中,分别对应 Houston2013 和 Augsburg 数据集,详细列出了各种土地覆盖类别的分类准确率以及总体分类准确度 (OA)、平均分类准确度 (average accuracy, AA) 和 Kappa 系数。

在对 Houston2013 数据集的评估中, GCA-Net 在

3项核心评价指标上均达到了最高水平。对于Augsburg数据集,GCA-Net同样展现出卓越的性能,其综合指标在众多分类方法中名列前茅。特别是在区分森林、工业区、低矮植被和水域这4类地物时,GCA-Net的分类准确率显著超越了其他方法。这一显著优势源于GCA-Net的独特设计,它采用了交叉

表1 在Houston2013数据集上各算法的指标对比
Table 1 Comparison of metrics for various algorithms on the Houston2013 dataset

方法	健康草地	受压草地	合成草地	树木	土壤	水资源	住宅区	商业区	道路
	198/1 053	190/1 064	192/505	188/1 056	186/1 056	182/143	196/1 072	191/1 053	193/1 059
SVM	82.72	83.18	100	94.03	99.15	99.30	79.66	55.27	75.83
TBCNN	83.10	84.96	97.62	93.37	100	98.60	88.15	82.15	85.36
S ² ENet	83.10	85.53	99.80	99.34	99.91	100	91.98	91.93	92.35
AsyFFNet	83.10	82.61	100	93.37	99.72	100	92.26	91.55	86.02
ExViT	91.55	85.15	98.61	98.61	100	98.60	88.90	93.35	89.52
HCT	99.89	88.52	100	97.03	98.03	98.43	97.94	91.2	94.62
MACN	100	99.15	99.21	97.38	100	58.13	72.51	90.89	90.02
GCA-Net(本文)	99.77	89.19	100	96.42	98.69	99.31	92.97	92.73	98.32

方法	高速公路	铁路	停车场1	停车场2	网球场	跑道	OA	AA	Kappa
	191/1 036	181/1 054	192/1 041	184/285	181/247	187/473			
SVM	62.55	96.87	86.55	53.68	98.79	98.10	82.91	84.38	81.43
TBCNN	64.67	79.22	95.20	80.88	92.71	94.50	86.91	88.70	85.84
S ² ENet	93.05	97.34	98.37	90.88	100	100	93.97	94.90	93.45
AsyFFNet	66.31	95.92	94.24	98.60	100	100	90.06	92.25	89.24
ExViT	65.54	95.83	90.39	90.53	99.60	89.85	90.62	91.76	89.83
HCT	90.94	75.60	93.06	97.60	100	99.16	92.94	93.12	92.33
MACN	98.06	74.96	95.06	88.77	100	100	90.82	92.34	90.04
GCA-Net(本文)	94.17	94.95	95.77	92.11	100	100	95.61	96.14	95.23

注:加粗字体表示各列最优结果。类别下面的数字表示训练样本数/测试样本数。

表2 在Augsburg数据集上各算法的指标对比
Table 2 Comparison of metrics for various algorithms on the Augsburg dataset

方法	森林	住宅区	工业区	低矮植物	配额地	商业区	水域	OA	AA	Kappa
	146/13 361	264/30 065	21/3 830	248/26 609	52/523	7/1 638	23/1 507			
SVM	90.55	89.81	23.03	83.73	34.23	9.71	45.92	81.60	53.82	73.17
TBCNN	90.88	95.89	8.28	91.97	38.24	1.40	10.82	84.53	47.92	77.13
S ² ENet	98.10	99.08	12.19	91.78	45.12	1.22	24.09	88.22	53.08	82.61
AsyFFNet	96.01	97.96	40.81	91.74	65.39	4.58	34.62	90.15	65.51	85.74
ExViT	90.04	95.44	34.58	90.68	51.82	28.63	17.65	86.65	58.40	80.79
HCT	96.03	91.72	64.97	96.60	23.02	11.02	79.94	91.38	62.45	87.52
MACN	96.07	90.60	64.40	98.84	24.50	11.40	75.73	91.22	63.67	87.30
GCA-Net(本文)	98.12	90.08	66.86	98.91	45.88	26.21	90.55	92.51	66.30	89.12

注:加粗字体表示各列最优结果。类别下面的数字表示训练样本数/测试样本数。

注意力模块来加深不同数据模态间的特征融合,使得网络能够更深入地挖掘和利用多源遥感数据的互补性。此外,GCA-Net引入的门控机制能够根据分类任务的需要动态调整特征权重,突出那些对于区分不同地物类别最为关键的特征。通过对两个数据集的对比实验结果进行综合分析,可以清晰地看到GCA-Net在处理多源遥感图像地物分类任务时的强大优势,这一优势不仅体现在个别类别的准确率提升上,更在于其整体性能的全面提升和对复杂地物类别的精准识别能力。

Houston2013 数据集上的地物分类对比实验结

果通过图3进行了可视化呈现。从细节图可以看出,本文提出的方法在细节放大图上可以看出与真值图更加接近。特别是在商业区、树木和高速公路类别的分类结果上,本文方法明显优于其他方法,充分体现了算法的优势。Augsburg数据集上的地物分类对比实验结果通过图4进行了可视化呈现。从图3和图4的比较可以清楚地看出,GCA-Net的结果相比于其他算法与真值图更加接近。尤其在工业区、水域和低矮植物分类的结果上,GCA-Net比其他算法更加准确。综上所述,本文提出的遥感图像地物分类模型GCA-Net在地物分类领域具有明显的优越性。

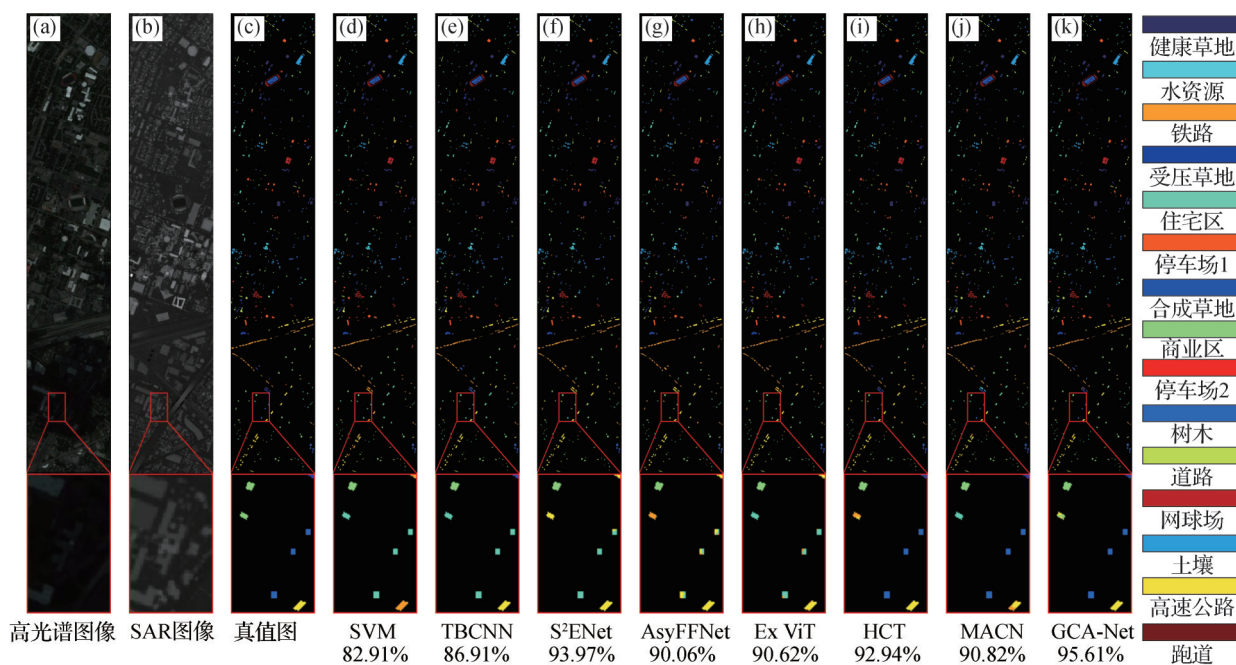


图3 各算法在 Houston2013 数据集上的可视化图以及对应的放大细节

Fig. 3 Visualization and corresponding detailed zoomed-in images of various algorithms on the Houston2013 dataset

2.5 消融实验

为了评估本文提出的交叉注意力特征融合和细化门控模块的性能,在 Houston2013 与 Augsburg 数据集上进行了深入的消融实验,结果如表3所示。可以看出,随着网络结构的逐步改进,分类精度得到了显著增强。

首先,基础的卷积特征提取网络在 Houston2013 和 Augsburg 数据集上分别取得了 92.40% 和 91.16% 的整体准确率(OA),92.90% 和 64.47% 的平均准确率(AA),以及 91.91% 和 86.60% 的 Kappa 系数。这些基准结果为后续引入模块提供了对比基础。

引入交叉注意力特征融合机制后,分类性能显

著提升。在 Houston2013 数据集上,OA、AA 和 Kappa 系数分别提高至 93.78%、94.45% 和 93.25%;在 Augsburg 数据集上,分别提升至 92.22%、65.42% 和 88.72%。这种性能提升主要归功于交叉注意力特征融合机制能够深度整合高光谱数据的光谱信息与 LiDAR/SAR 数据的空间信息,其增强了模型对局部细节的敏感性,并扩展了对全局上下文的理解。

通过加入细化门控结构,进一步优化了特征融合过程。在 Houston2013 数据集上,OA、AA 和 Kappa 系数分别达到 95.61%、96.14% 和 95.23%;在 Augsburg 数据集上,分别提升至 92.51%、66.30% 和 89.12%。细化门控模块通过动态调整特征权重,确

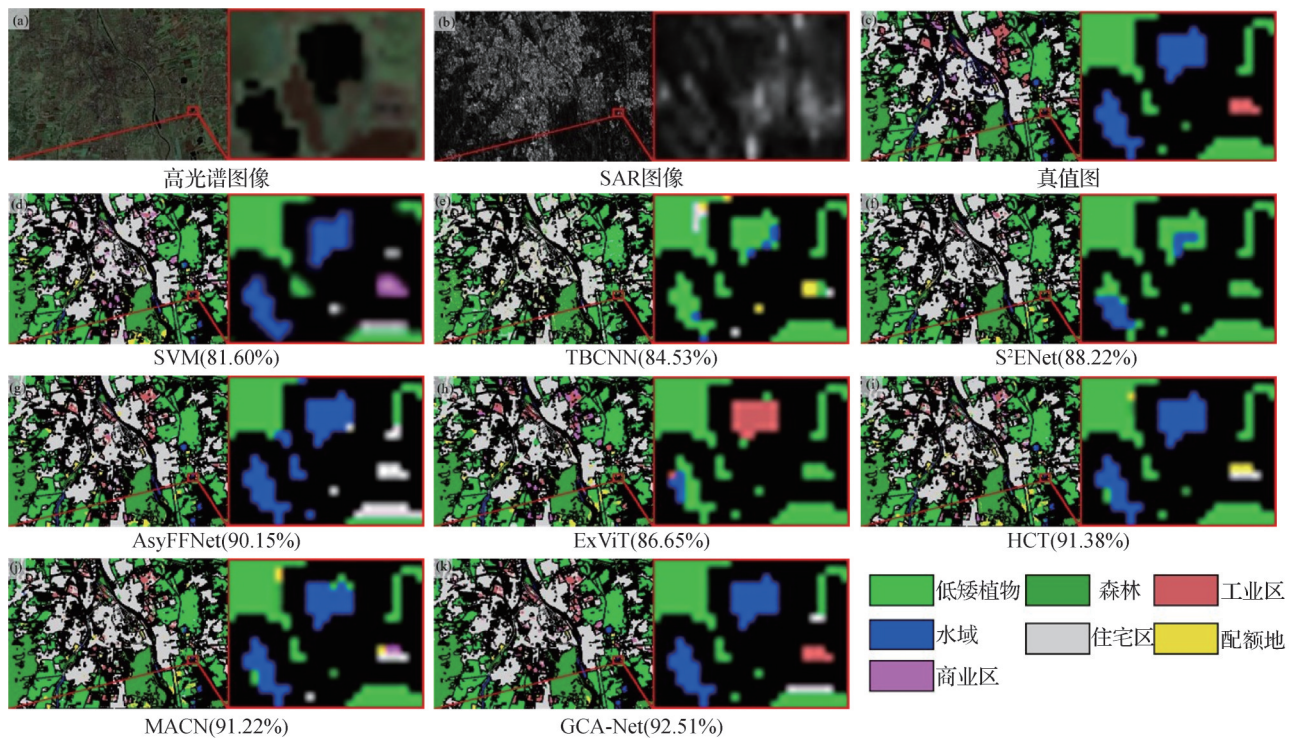


图4 各算法在 Augsburg 数据集上的可视化图以及对应的放大细节
Fig. 4 Visualization and corresponding detailed zoomed-in images of various algorithms on the Augsburg dataset

保模型聚焦于对分类任务最为关键的信息,这种精准的特征筛选与加权机制显著提升了模型在处理复杂遥感图像时的特征表示准确性。

这一系列性能提升的原因在于,交叉注意力特征融合机制深度整合了高光谱数据的光谱信息与 LiDAR/SAR 数据的空间信息,增强了模型对局部细

节的敏感性,并扩展了其全局上下文的理解,从而在分类任务中实现了更加全面的特征覆盖。此外,细化门控模块通过动态调整特征权重,确保模型聚焦于对分类任务最为关键的信息,这种精准的特征筛选与加权机制显著提升了模型在处理复杂遥感图像时的特征表示准确性。

表 3 交叉注意力特征融合和细化门控模块消融实验结果
Table 3 Results of ablation experiments on cross-attention feature fusion and refinement gating

网络结构	/%					
	OA		AA		Kappa	
	Houston2013	Augsburg	Houston2013	Augsburg	Houston2013	Augsburg
卷积特征提取网络	92.40	91.16	92.90	64.47	91.91	86.60
交叉注意力特征融合	93.78	92.22	94.45	65.42	93.25	88.72
交叉注意力特征融合 + 细化门控	95.61	92.51	96.14	66.3	95.23	89.12

3 结 论

本文提出了一种基于门控注意力的聚合网络 (GCA-Net),旨在应对多源遥感图像分类中的复杂挑战。GCA-Net 的核心优势在于其独特的门控跨模态聚合模块 (GCAM)。该模块采用交叉注意力特征

融合技术,深入整合了高光谱数据的光谱特征与 LiDAR/SAR 数据的空间特征,实现了信息的互补和深度挖掘。通过细化门控模块,GCAM 能够动态地调整特征,捕捉到更精准的关键信息,显著提升了特征表达的精度和模型的泛化性能。进一步地,本文还设计了一种高效的高光谱特征编码器,它不仅能够捕捉局部细节,还能理解全局上下文信息,从而极

大地增强了模型对复杂遥感数据的解析能力。这种全面的特征学习策略使得模型能够在识别图像中的局部模式和细微特征的同时,也能在更宽广的上下文中赋予它们意义。在实验验证方面,选取了Houston2013和Augsburg两个具有代表性的遥感数据集,通过总体精度(OA)、平均精度(AA)和Kappa系数等关键评价指标,对GCA-Net的性能进行了全面评估。通过地物分类的可视化结果,从定量和定性两个维度全面验证了GCA-Net模型的有效性。

尽管本文提出的GCA-Net在多源遥感图像分类任务中取得了较好的分类性能,未来的工作仍有很大的拓展空间。例如,可以考虑将多光谱图像、热红外图像等更多类型的遥感数据纳入模型中,这将有助于实现更为全面和细致的地物分类,丰富特征表示。

参考文献(References)

- Arshad T and Zhang J P. 2024. Hierarchical attention Transformer for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 21: #5504605 [DOI: 10.1109/LGRS.2024.3379509]
- Baumgartner A, Gege P, Köhler C, Lenhard K and Schwarzmaier T. 2012. Characterisation methods for the hyperspectral sensor HyS-pex at DLR's calibration home base//*Proceedings Volume 8533, Sensors, Systems, and Next-Generation Satellites XVI*. Edinburgh, United Kingdom: SPIE: 371-378 [DOI: 10.1117/12.974664]
- Bioucas-Dias J M, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N and Chanussot J. 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2): 6-36 [DOI: 10.1109/MGRS.2013.2244672]
- Chen Y S, Li C Y, Ghamisi P, Jia X P and Gu Y F. 2017. Deep Fusion of Remote Sensing Data for Accurate Classification. *IEEE Geoscience and Remote Sensing Letters*, 14(8): 1253-1257 [DOI: 10.1109/LGRS.2017.2704625]
- Fang S, Li K Y and Li Z. 2022. S²ENet: spatial-spectral cross-modal enhancement network for classification of hyperspectral and LiDAR data. *IEEE Geoscience and Remote Sensing Letters*, 19: #6504205 [DOI: 10.1109/LGRS.2021.3121028]
- Gao F, Meng D S, Xie Z Y, Qi L and Dong J Y. 2024. Multi-source remote sensing image classification based on Transformer and dynamic 3D-convolution. *Journal of Beijing University of Aeronautics and Astronautics*, 50(2): 606-614 (高峰, 孟德森, 解正源, 元林, 董军宇. 2024. 基于Transformer和动态3D卷积的多源遥感图像分类. *北京航空航天大学学报*, 50(2): 606-614) [DOI: 10.13700/j.bh.1001-5965.2022.0397]
- Hong D F, Gao L R, Hang R L, Zhang B and Chanussot J. 2022. Deep encoder-decoder networks for classification of hyperspectral and LiDAR data. *IEEE Geoscience and Remote Sensing Letters*, 19: #5500205 [DOI: 10.1109/LGRS.2020.3017414]
- Hu S, Gao F, Gong Z R, Tao S E, Shangguan X Y and Dong J Y. 2024. Parallel channel shuffling and Transformer-based denoising for hyperspectral images. *Journal of Image and Graphics*, 29(7): 2063-2074 (胡帅, 高峰, 龚卓然, 陶盛恩, 上官心语, 董军宇. 2024. 基于Transformer和通道混合并行卷积的高光谱图像去噪. *中国图象图形学报*, 29(7): 2063-2074) [DOI: 10.11834/jig.230381]
- Li K, Wang D, Wang X, Liu G, Wu Z L and Wang Q. 2023a. Mixing self-attention and convolution: a unified framework for multisource remote sensing data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5523216 [DOI: 10.1109/TGRS.2023.3310521]
- Li W, Gao Y H, Zhang M M, Tao R and Du Q. 2023b. Asymmetric feature fusion network for hyperspectral and SAR image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10): 8057-8070 [DOI: 10.1109/TNNLS.2022.3149394]
- Li W, Wang J J, Gao Y H, Zhang M M, Tao R and Zhang B. 2022. Graph-feature-enhanced selective assignment network for hyperspectral and multispectral data classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5526914 [DOI: 10.1109/TGRS.2022.3166252]
- Ma W P, Guo Y S, Zhu H, Yi X Y, Zhao W H, Wu Y, Hou B and Jiao L C. 2024. Intra- and intersource interactive representation learning network for remote sensing images classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5401515 [DOI: 10.1109/TGRS.2024.3352816]
- Man Q X, Dong P L and Guo H D. 2015. Pixel- and feature-level fusion of hyperspectral and LiDAR data for urban land-use classification. *International Journal of Remote Sensing*, 36(6): 1618-1644 [DOI: 10.1080/01431161.2015.1015657]
- Melgani F and Bruzzone L. 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8): 1778-1790 [DOI: 10.1109/TGRS.2004.831865]
- Moreira A, Prats-Iraola P, Younis M, Krieger G, Hajnsek I and Papathanassiou K P. 2013. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1): 6-43 [DOI: 10.1109/MGRS.2013.2248301]
- Roy S K, Sukul A, Jamali A, Haut J M and Ghamisi P. 2024. Cross hyperspectral and LiDAR attention Transformer: an extended self-attention for land use and land cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5512815 [DOI: 10.1109/TGRS.2024.3374324]
- Wang J J, Gao F, Dong J Y and Du Q. 2021. Adaptive DropBlock-

- enhanced generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6): 5040-5053 [DOI: 10.1109/TGRS.2020.3015843]
- Xiao Z J, Lin B H and Qu H C. 2024. SAR ship detection with multi-mechanism fusion. *Journal of Image and Graphics*, 29(2): 545-558 (肖振久, 林渤翰, 曲海成. 2024. 融合多重机制的SAR舰船检测. *中国图象图形学报*, 29(2): 545-558) [DOI: 10.11834/jig.230166]
- Xu X D, Li W, Ran Q, Du Q, Gao L R and Zhang B. 2018. Multi-source remote sensing data classification based on convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 937-949 [DOI: 10.1109/TGRS.2017.2756851]
- Yao J, Zhang B, Li C Y, Hong D F and Chanussot J. 2023. Extended Vision Transformer (ExViT) for land use and land cover classification: a multimodal deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5514415 [DOI: 10.1109/TGRS.2023.3284671]
- Zhao G R, Ye Q L, Sun L, Wu Z B, Pan C S and Jeon B. 2023. Joint classification of hyperspectral and LiDAR data using a hierarchical CNN and Transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5500716 [DOI: 10.1109/TGRS.2022.3232498]
- Zhao X D, Tao R, Li W, Li H C, Du Q, Liao W Z and Philips W. 2020. Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10): 7355-7370 [DOI: 10.1109/TGRS.2020.2982064]

作者简介

金学鹏,男,硕士研究生,主要研究方向为多源遥感图像分类和深度学习。E-mail: jinxuepeng@stu.ouc.edu.cn

高峰,通信作者,男,副教授,主要研究方向为多媒体计算与高光谱图像处理。E-mail: gaofeng@ouc.edu.cn

石晓晨,男,硕士研究生,主要研究方向为多源遥感图像分类和深度学习。E-mail: 624122050@qq.com

董军宇,男,教授,主要研究方向为计算机视觉与模式识别。E-mail: dongjunyu@ouc.edu.cn