

基于小波变换和多尺度注意力机制的高光谱图像分类模型

易守恒¹, 焦 龙^{2*}

¹西安石油大学电子工程学院, 陕西 西安

²西安石油大学化学化工学院, 陕西 西安

收稿日期: 2025年2月27日; 录用日期: 2025年3月18日; 发布日期: 2025年3月31日

摘要

高光谱图像(Hyperspectral Image, HSI)分类旨在利用高光谱数据丰富的光谱信息, 实现细粒度物质识别。传统的卷积神经网络(Convolutional Neural Network, CNN)在提取空间特征方面表现出了卓越的能力, 但往往难以有效捕捉光谱特征中固有的顺序和层次关系, 从而限制了其在HSI分类任务中的性能。为了解决这一挑战, 我们提出了一种新的模型, 该模型将小波变换(Wavelet Transform, WT)与多尺度注意力机制(Multi-Scale Attention, MSA)相结合, 以提升分类性能。具体来说, WT能够有效地将光谱空间信息分解为多个频率子带, 便于细粒度特征的提取和分层分析; MSA会结合不同大小的卷积核来提取图像的不同尺度信息, 从而更好地捕捉局部特征和全局特征的关系。通过学习不同尺度的重要特征, 实现多尺度信息的高效融合。在公开的高光谱数据集上的实验结果表明, 该模型在分类精度方面显著优于传统的基于CNN模型和Transformer模型。这些结果充分表明, 将WT与MSA相结合, 能够在高光谱图像分类的任务中展现出巨大的潜力。

关键词

小波变换, 多尺度注意力机制, 高光谱图像分类

Hyperspectral Image Classification Model Based on Wavelet Transform and Multi-Scale Attention

Shouheng Yi¹, Long Jiao^{2*}

¹College of Electronic Engineering, Xi'an Shiyou University, Xi'an Shaanxi

²College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an Shaanxi

Received: Feb. 27th, 2025; accepted: Mar. 18th, 2025; published: Mar. 31st, 2025

*通讯作者。

Abstract

Hyperspectral Image (HSI) classification aims to leverage the rich spectral information in hyperspectral data for fine-grained material identification. Traditional Convolutional Neural Networks (CNNs) have shown exceptional ability in extracting spatial features but often struggle to effectively capture the inherent order and hierarchical relationships in spectral features, thus limiting their performance in HSI classification tasks. To address this challenge, we propose a novel model that combines Wavelet Transform (WT) with Multi-Scale Attention (MSA) to enhance classification performance. Specifically, WT can effectively decompose spectral-spatial information into multiple frequency subbands, facilitating the extraction of fine-grained features and hierarchical analysis. MSA, on the other hand, combines convolution kernels of different sizes to extract multi-scale information from the image, thereby better capturing the relationships between local and global features. By learning the importance of features at different scales, MSA enables efficient fusion of multi-scale information. Experimental results on publicly available hyperspectral datasets show that this model significantly outperforms traditional CNN-based models and Transformer models in classification accuracy. These results fully demonstrate that the combination of WT and MSA holds great potential in hyperspectral image classification tasks.

Keywords

Wavelet Transform, Multi-Scale Attention, Hyperspectral Image Classification

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景与现状

高光谱遥感技术作为遥感领域的前沿技术,因其同时获取空间、光谱和辐射信息的能力而备受关注。与传统多光谱遥感不同,高光谱遥感利用数百甚至上千个窄波段,能够捕捉地物的连续光谱特征曲线,为地物的精细分类和定量化分析提供了坚实基础。

近年来,高光谱遥感图像(Hyperspectral Image, HSI)分类技术已在环境监测[1]、城市发展与规划[2]、精准农业[3]、矿产资源勘探[4]等领域得到了广泛应用。传统的 HSI 分类方法主要集中在光谱特征的提取上,例如,k 近邻算法(k-Nearest Neighbor, k-NN) [5]、支持向量机(Support Vector Machine, SVM) [6]和随机森林(Random Forest, RF) [7]等经典方法被广泛应用。然而,这些方法在高维光谱数据中难以提取更深层次的特征。此外,大多数模型仅以光谱向量为输入,忽略了空间上下文信息的充分挖掘,容易导致分类结果中出现大量噪声[8],影响分类精度和视觉效果。

与传统方法相比,基于深度学习的分类方法通过多层非线性变换,能够自动提取低级特征(如形状、几何、纹理)以及高级语义信息[9],因而在 HSI 分类任务中展现出很好的表现。这些研究主要集中于卷积神经网络(Convolutional Neural Network, CNN) [10]、递归神经网络(Recurrent Neural Network, RNN) [11]和生成对抗网络(Generative Adversarial Network, GAN) [12]等技术。Chen 等人[13]首次将深度学习引入 HSI 分类领域,将稀疏自编码器(Sparse Autoencoder, SAE)与主成分分析(Principal Component Analysis, PCA)相结合,大幅提升了分类性能。随后,受 RGB 图像分类的启发, Haut 等人[14]使用二维卷积神经网络(2D-

Convolutional Neural Network, 2D-CNN)对 HSI 进行分类。此外, Zhou 等人[15]将多通道像素的时间序列和斑块的行向量输入长短时记忆网络(Long Short-Term Memory, LSTM)模型, 提取光谱和空间特征, 在 HSI 分类中取得了优异表现。

尽管基于 CNN 的分类方法取得了突破性进展, 但仍然面临着诸多挑战。例如, CNN 在捕获光谱维度的长距离依赖关系和细微光谱差异方面能力不足。此外, CNN 依赖固定大小的卷积核, 主要关注局部特征, 难以全面捕捉光谱和空间数据的全局信息。

近年来, Transformer [16]在自然语言处理(Natural Language Processing, NLP)领域取得了显著成果, 特别是在处理序列到序列任务(如机器翻译)方面表现出卓越的性能。这得益于其摒弃了 RNN 的序列依赖特性, 引入了自注意力机制(Self-Attention, SA)。这种机制能够全局捕捉长距离依赖关系, 大幅提升了时间序列数据处理能力。Transformer 的创新架构逐步扩展至图像处理与计算机视觉领域。视觉 Transformer (Vision Transformer, ViT) [17]在多种视觉任务中表现出接近或超过 CNN 的效果, 为视觉任务提供了新的思路和空间。在高光谱分类领域, Transformer 的应用也日益增多。例如, Hong 等人[18]从序列建模的角度提出了一种名为光谱 Transformer(Spectral Former, SF)的骨干网络; He 等人[19]基于 Transformer 优势, 设计了空间光谱转换器(Spatial-Spectral Transformer, SST)分类框架; Mou 等人[20]则提出了一种结合 HSI 波段注意力机制的卷积神经网络。这些方法在进一步提升 HSI 分类性能方面取得了显著成效, 为该领域提供了重要的理论支持和实践参考。然而, Transformer 模型对局部细节特征的提取不够精细, 尤其在处理高维光谱数据时, 可能导致重要局部信息的丢失。

针对上述挑战, 我们提出了一种新型模型, 将小波变换与多尺度注意机制相结合, 以进一步提升 HSI 分类性能。

2. 研究方法

本文提出了一种基于小波变换和多尺度注意力机制的高光谱图像分类模型(Wavelet Transform and Multi-Scale Attention, WTMSA), 其流程如图 1 所示包括高光谱图像预处理、小波变换模块(Wavelet Transform, WT)、逆小波变换模块(Inverse Wavelet Transform, IWT)和多尺度注意力模块(Multi-Scale Attention, MSA)的构建。通过镜像填充、裁剪子区域和最小最大归一化对输入数据进行预处理后, 将其输入 WTMSA 模型进行特征提取并完成分类。

2.1. 高光谱图像预处理

高光谱图像通常包含大量像素和光谱信息, 直接处理整幅图像计算量巨大。将其裁剪为较小的数据立方体, 有助于降低计算复杂度, 同时保留每个像素的局部上下文信息。

首先, 采用镜像填充避免边缘信息丢失, 在裁剪过程中, 图像边缘的像素可能因邻域缺失而无法得到完整的局部上下文信息。对图像上下左右分别进行镜像填充, 确保填充后的图像与原图像在边界附近的像素连续操作, 补充边缘像素的邻域, 避免边界区域分类性能下降。此部分可以表示为, 输入图像为 $C \times H \times W$, 输出填充图像为 $C \times (H + 2p) \times (W + 2p)$, 其中 $p = \text{patch}/2$ 。

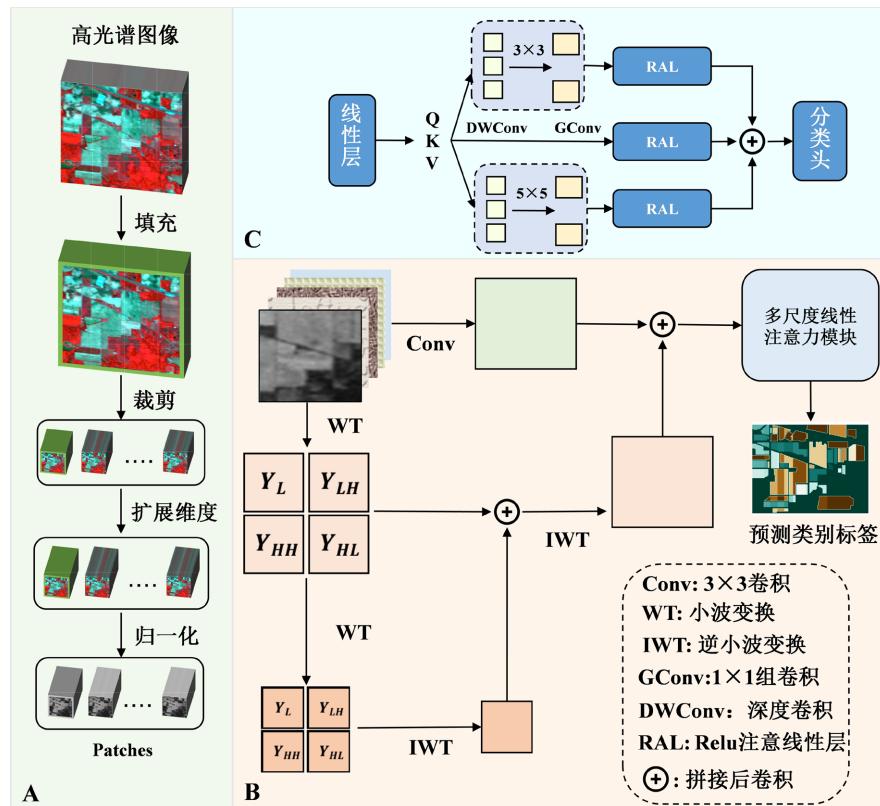
其次, 对于给定的每个中心点, 裁剪其邻域像素形成大小为 $P \times P \times C$ 的 patch。通过裁剪包含邻域信息的子区域, 可以更好地捕获局部空间特征, 并通过遍历输入点确保所有像素都能参与训练和测试。

然后, 对每个 patch 中的光谱维度 band 进行左右镜像填充, 从而更好地捕获连续波段的特征, 填充后的维度变为 $C + 2b$, 其中 $b = \text{patch}/2$ 。

最后, 通过归一化处理加快模型收敛速度, 采用最小最大归一化将数据值归一到统一范围, 增强训练的稳定性。其公式表达为:

$$x_{\text{normalized}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

其中, x 为输入的像素值; $\max(x)$ 和 $\min(x)$ 分别为该通道的最大值和最小值; $x_{\text{normalized}}$ 为归一化后的值。



2.2. WTMSA 模型

WTMSA 模型结合了小波变换和多尺度注意力机制, 通过多尺度特征提取与重加权, 充分利用了高光谱图像中丰富的空间和光谱信息。与传统的图像分类方法相比, WTMSA 能够更好地处理高光谱图像的高维和复杂特征, 从而提高分类性能。

2.2.1. WT 模块

在高光谱图像中, 像素的空间特征和光谱特征在不同尺度上存在显著差异。通过小波变换, 模型能够有效分解图像的低频和高频信息, 捕捉全局与局部的特征。低频分量保留了全局信息, 而高频分量则包含了局部细节信息, 包括水平、垂直和对角方向的细节信息。采用这种方法, 模型能够同时关注全局信息和局部细节, 确保分类时对不同尺度的信息都能进行合理的处理。具体步骤如下:

首先, 使用 Haar 小波对输入张量 X 进行单层小波分解, 得到低频子带 Y_L 和高频子带 Y_H :

$$Y_L, Y_H = \text{WT}(X) \quad (2)$$

其中, 输入张量 $X \in R^{C \times H \times W}$, C, H, W 分别表示通道数、高度和宽度。 $Y_L \in R^{C \times \frac{H}{2} \times \frac{W}{2}}$ 低频分量, Y_H 是高频

分量, 包括水平、垂直和对角细节信息分别为 $Y_{HL} \in R^{C \times \frac{H}{2} \times \frac{W}{2}}$, $Y_{LH} \in R^{C \times \frac{H}{2} \times \frac{W}{2}}$, $Y_{HH} \in R^{C \times \frac{H}{2} \times \frac{W}{2}}$ 。

其次, 对分解后的特征进行特征融合: Y_L , Y_{HL} , Y_{LH} , Y_{HH} 沿通道维度拼接:

$$X = \text{Concat}([Y_L, Y_{HL}, Y_{LH}, Y_{HH}], \text{dim}=1) \quad (3)$$

拼接后张量的形状为 $X \in R^{4 \times C \times \frac{H}{2} \times \frac{W}{2}}$

最后, 采用卷积归一化 ReLU, 对拼接后的张量 X 进行 1×1 卷积操作, 然后应用批归一化和 ReLU 激活:

$$X = \text{Relu}(\text{BatchNorm}(\text{Conv}(X))) \quad (4)$$

2.2.2. IWT 模块

IWT 模块基于逆小波变换, 旨在恢复输入图像的细节特征。通过逆小波变换, 模型能够从低分辨率的表示中恢复高分辨率的细节, 使得图像的细节特征得到充分利用, 为后续的分类任务提供更为丰富的信息。具体步骤如下:

首先, 分解输入特征, 根据离散小波变换的高频和低频特性, 将 X 解码为低频和高频部分:

$$\begin{aligned} Y_L &= X\left[:, : \frac{C}{4}, :, :\right] \\ Y_{HL} &= X\left[:, \frac{C}{4} : \frac{C}{2}, :, :\right], \quad Y_{LH} = X\left[:, \frac{C}{2} : \frac{3C}{4}, :, :\right], \quad Y_{HH} = X\left[:, \frac{3C}{4} :, :, :\right] \end{aligned} \quad (5)$$

其次, 使用逆小波变换恢复高分辨率特征:

$$X = \text{IWT}(Y_L, Y_H) \quad (6)$$

最后, 对输出张量 X 进行 1×1 卷积操作, 然后应用批归一化和 ReLU 激活:

$$X = \text{Relu}(\text{BatchNorm}(\text{Conv2D}(X))) \quad (7)$$

2.2.3. MSA 模块

MSA 模块利用多尺度特征增强和注意力机制来细化特征的表示。高光谱图像中的不同尺度特征对分类有着不同的影响, 局部的细节信息和全局的依赖关系在不同尺度下表现出不同的特征。因此, 本模块采用了多尺度卷积核进行特征增强, 并通过注意力机制自动分配不同尺度特征的重要性。通过动态调整各尺度特征的权重, MSA 模块能够有效地融合局部细节与全局上下文, 增强模型对复杂模式的识别能力。每个尺度的特征在经过 ReLU 激活后, 进入线性注意力模块进行加权, 从而生成更精确的分类特征。具体步骤如下:

首先, 给定输入特征 $X \in R^{C \times H \times W}$ 通过一个 1×1 卷积映射到查询(Query, Q)、键(Key, K)和值(Value, V)的特征空间:

$$Q, K, V = \text{Conv}_{1 \times 1}(X) \quad (8)$$

多尺度卷积操作的过程使用了不同尺度的卷积核(3×3 和 5×5)对 Q, K, V 进行特征增强。公式为:

$$\text{multi_scale_qkv} = \text{Concat}([QKV] + \text{Op}[QKV \text{ for Op in aggregations}]) \quad (9)$$

这里的每个 Op 对应于一组多尺度卷积层:

$$\text{Op}(qkv) = \text{Conv}_{k \times k}(QKV) \rightarrow \text{DepthwiseConv}_{1 \times 1}(QKV) \quad (10)$$

其中, k 是不同的卷积核大小(3, 5)。所有尺度的输出通过拼接 Concat 得到一个新的多尺度特征张量 multi_scale_qkv。

其次, 多尺度特征张量 multi_scale_qkv 被送入线性注意力模块。此过程包括对 Q 和 K 应用 ReLU 激活函数, 得到:

$$Q' = \text{ReLU}(Q), \quad K' = \text{ReLU}(K) \quad (11)$$

计算键值特征的加权表示:

$$KV = K'^T \times V \quad (12)$$

最后, 计算最终的注意力输出:

$$O = \frac{Q' \times KV}{Q' \times K'^T + \epsilon} \quad (13)$$

这里的分母用于归一化, ϵ 是一个小的数值防止分母为零,

将注意力结果 O 通过 1×1 卷积进行整合, 得到最终输出:

$$\text{Output} = \text{Conv}_{1 \times 1}(O) \quad (14)$$

所有模块搭建完成后进行特征提取, 所得到的高级特征输入到线性分类头中进行分类。

3. 实验

3.1. 数据集描述与划分

为了验证所提出的 WTMSA 模型的有效性, 本文选择了 3 种公开高光谱遥感数据集进行一系列的实验, 分别为印度松(Indian Pines)数据集、萨利纳斯(Salinas)数据集和帕维亚大学(Pavia University)数据集。

印度松数据集由 AVIRIS 传感器采集, 覆盖了美国印第安纳州西北部的一处测试区域。数据集尺寸为 145×145 像素, 最初包含 224 个波段, 波长范围为 0.4~2.5 微米。数据经过预处理后, 去除了水吸收区域的波段, 最终保留 200 个波段。该场景以农业为主, 占比约三分之二, 其余为森林和其他自然植被。此外, 还包含一些人造结构, 如双车道公路、铁路和低密度住宅等。数据划分为 16 个类别。

萨利纳斯数据集由 AVIRIS 传感器采集, 覆盖了美国加利福尼亚州萨利纳斯谷地区, 具有较高的空间分辨率。数据集尺寸为 512×217 像素, 原始数据包含 224 个波段, 去除了水吸收区域的波段, 最终保留 204 个波段。该场景主要包括蔬菜地、裸土和葡萄园等 16 个地物类别。

帕维亚大学数据集由 ROSIS 传感器采集, 覆盖了意大利帕维亚市帕维亚大学区域。数据集尺寸为 610×610 像素, 空间分辨率为 1.3 米/像素, 原始数据包含 103 个光谱波段。主要包括建筑物、道路、植被等 9 个类别。

为了模拟实际应用场景, 本文对上述数据集进行了合理的训练集与测试集划分。具体来说, 每个数据集中的每一类别样本随机按比例 15%作为训练集, 85%作为测试集。这种划分方式既保证了模型训练的多样性和代表性, 又在测试阶段提供了充足的数据量以全面评估模型的性能。[表 1~3](#) 分别展示了 Indian Pines、Salinas 和 Pavia University 数据集的类别及其划分。

3.2. 实验设置

本文提出的方法是在 Python 3.8, Pytorch1.7.8 框架上实现的, 设备配备有 Intel(R) Core (TM) i5-7300HQ CPU、NVIDIA GeForce RTX 1050 GPU 和 16G 随机存取内存。采用交叉熵损失函数(CrossEntropyLoss, CELoss)作为优化目标, 使用 Adam 优化器, 学习率为 5×10^{-4} , 并结合权重衰减参数

Table 1. Indian Pines dataset class distribution table**表 1.** Indian Pines 数据集类别划分统计表

| Class | Class Name | Train | Test | Total |
|-------|------------------------------|-------|------|--------|
| 1 | Alfalfa | 6 | 40 | 46 |
| 2 | Corn-notill | 214 | 1214 | 1428 |
| 3 | Corn-mintill | 124 | 706 | 830 |
| 4 | Corn | 35 | 202 | 237 |
| 5 | Grass-pasture | 72 | 411 | 483 |
| 6 | Grass-trees | 109 | 621 | 730 |
| 7 | Grass-pasture-mowed | 4 | 24 | 28 |
| 8 | Hay-windrowed | 71 | 407 | 478 |
| 9 | Oats | 3 | 17 | 20 |
| 10 | Soybean-notill | 145 | 827 | 972 |
| 11 | Soybean-mintill | 368 | 2087 | 2455 |
| 12 | Soybean-clean | 88 | 505 | 593 |
| 13 | Wheat | 30 | 175 | 205 |
| 14 | Woods | 189 | 1076 | 1265 |
| 15 | Buildings-Grass-Trees-Drives | 57 | 329 | 386 |
| 16 | Stone-Steel-Towers | 13 | 80 | 93 |
| Total | | 1528 | 8721 | 10,249 |

Table 2. Salinas dataset class distribution table**表 2.** Salinas 数据集类别划分统计表

| Class | Class Name | Train | Test | Total |
|-------|---------------------------|-------|--------|--------|
| 1 | Brocoli_green_weeds_1 | 301 | 1708 | 2009 |
| 2 | Brocoli_green_weeds_2 | 558 | 3168 | 3726 |
| 3 | Fallow | 296 | 1680 | 1976 |
| 4 | Fallow_rough_plow | 209 | 1185 | 1394 |
| 5 | Fallow_smooth | 401 | 2277 | 2678 |
| 6 | Stubble | 593 | 3366 | 3959 |
| 7 | Celery | 536 | 3043 | 3579 |
| 8 | Grapes_untrained | 1690 | 9581 | 11,271 |
| 9 | Soil_vinyard Develop | 930 | 5273 | 6203 |
| 10 | Corn_senesced_green_weeds | 491 | 2787 | 3278 |
| 11 | Lettuce_romaine_4wk | 160 | 908 | 1068 |
| 12 | Lettuce_romaine_5wk | 289 | 1638 | 1927 |
| 13 | Lettuce_romaine_6wk | 137 | 779 | 916 |
| 14 | Lettuce_romaine_7wk | 160 | 910 | 1070 |
| 15 | Vinyard_untrained | 1090 | 6178 | 7268 |
| 16 | Vinyard_vertical_trellis | 271 | 1536 | 1807 |
| Total | | 8112 | 46,017 | 54,129 |

Table 3. Pavia University dataset class distribution table
表 3. Pavia University 数据集类别划分统计表

| Class | Class Name | Train | Test | Total |
|-------|----------------------|-------|--------|--------|
| 1 | Asphalt | 994 | 5637 | 6631 |
| 2 | Meadows | 2797 | 15,852 | 18,649 |
| 3 | Gravel | 314 | 1785 | 2099 |
| 4 | Trees | 459 | 2605 | 3064 |
| 5 | Painted metal sheets | 201 | 1144 | 1345 |
| 6 | Bare Soil | 754 | 4275 | 5029 |
| 7 | Bitumen | 199 | 1131 | 1330 |
| 8 | Self-Blocking Bricks | 552 | 3130 | 3682 |
| 9 | Shadows | 142 | 805 | 947 |
| Total | | 8112 | 46,017 | 54,129 |

5×10^{-3} 以提升模型泛化性能, 训练轮次为 150 轮。

3.3. 损失函数

CELoss 是深度学习中常用的损失函数, 尤其适用于分类问题。它衡量了模型预测结果和真实标签之间的差异, 通过最小化交叉熵损失来优化模型的参数, 使模型的预测结果尽可能地接近真实标签。对于多分类问题, 它计算了预测概率分布与真实概率分布之间的距离。损失函数可以表示为:

$$L = -\sum_{c=1}^C y_{i,c} \log(y'_{i,c}) \quad (15)$$

其中, $y_{i,c}$ 是真实标签向量 y_i 的第 c 个元素, $y'_{i,c}$ 是预测概率向量 y'_i 的第 c 个元素。

3.4. 评价指标

为了评价模型的分类性能, 我们使用以下三项指标:

1. 总体精度(Overall Accuracy, OA): 表示分类正确的样本数占总样本数的比例, 公式如下:

$$OA = \frac{\sum_{i=1}^n M_{ii}}{\sum_{i=1}^n \sum_{j=1}^n M_{ij}} \quad (16)$$

其中, M_{ii} 表示混淆矩阵 M 中第 i 类正确分类的样本数(对角线元素)。 $\sum_{i=1}^n \sum_{j=1}^n M_{ij}$ 表示混淆矩阵中所有样本的总数。

2. 平均类别精度(Average Accuracy, AA)

平均类别精度是所有类别的分类精度的平均值, 计算公式为:

$$AA = \frac{1}{n} \sum_{i=1}^n \frac{M_{ii}}{\sum_{j=1}^n M_{ij}} \quad (17)$$

其中, n 是类别总数。 $\sum_{j=1}^n M_{ij}$ 表示第 i 类的分类精度, 即正确分类的样本数占该类别总样本数的比例。

3. Kappa 系数 (Kappa)

Kappa 系数用于衡量分类结果与随机分类的偏离程度, 其计算公式为:

$$\text{Kappa} = \frac{\text{OA} - P_e}{1 - P_e} \quad (18)$$

其中, OA 是总体精度。

P_e 是随机分类的期望精度, 计算公式为:

$$P_e = \frac{\sum_{i=1}^n \left(\sum_{j=1}^n M_{ij} \cdot \sum_{j=1}^n M_{ji} \right)}{\sum_{i=1}^n \sum_{j=1}^n M_{ij}^2} \quad (19)$$

$\sum_{j=1}^n M_{ij}$ 表示第 i 类的总样本数(按行求和)。 $\sum_{j=1}^n M_{ji}$ 表示被预测为第 i 类的样本总数(按列求和)。

以上三项指标能够全面评价分类模型的性能, 分别反映了整体分类精度、类别分类均衡性及与随机分类的偏离程度。

3.5. 实验结果定量分析

根据表 4 中的分类定量指标, 本文方法在 OA、AA 和 Kappa 三项评价指标上均显著优于其他方法。

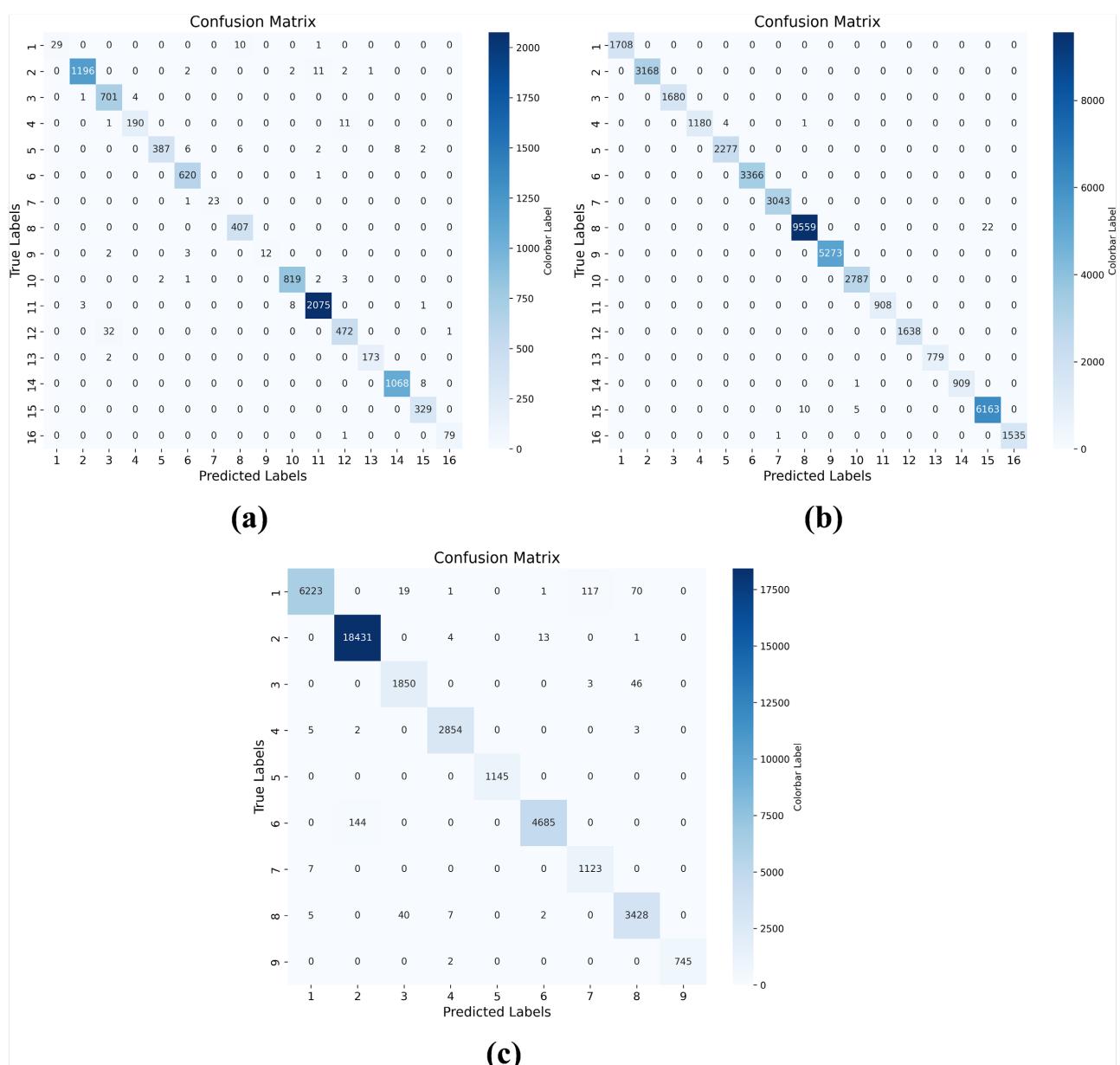
在三个高光谱数据集上的详细结果分析如下: 在 Indian Pines、Salinas 和 Pavia University 这三个数据集上, 对比了 CNN2D、LSTM、RNN、MSCNN、ViT、SF 以及 Ours 等多种方法的性能。在 Indian Pines 数据集上, 我们的 WTMSA 方法表现出色, Overall Accuracy 达到 98.38%, 比 CNN2D 高出 55.25%、比 LSTM 高出 23.08%、比 RNN 高出 23.50、比 MSCNN 高出 8.68%、比 ViT 高出 18.94%、比 SF 高出 8.89%; Kappa 指标为 98.16%, 比 MSCNN 提高 9.93%、比 SF 提高 10.07%; AA Mean 指标为 94.60, 比 SF 高出 13.83%。在 Salinas 数据集上, WTMSA 方法的 Overall Accuracy 高达 99.90%, Kappa 为 99.89%, AA Mean 为 99.93%同样远超其他方法, 在 Pavia University 数据集上, WTMSA 方法的 Overall Accuracy 为 98.80%, Kappa 为 98.38%, AA Mean 为 98.70%, 优势显著。总体而言, 我们的方法在多个数据集上的各项指标均优于其他对比方法, 展现出卓越的性能。

图 2 中的(a)、(b)和(c)分别对应 Indian Pines、Salinas 和 Pavia University 三个数据集的混淆矩阵。主对角线上的值均占据绝对主导地位, 表明模型在大多数类别上的分类性能较高。例如, 在 Indian Pines 数据集中, 第 6 类和第 8 类的准确率接近完美, 分别正确分类了 620 个和 407 个样本; 第 14 类正确分类了 1068 个样本, 误分类仅 8 个, 表现稳定。在 Salinas 数据集中, 主对角线分布清晰, 大多数类别几乎

Table 4. System resulting data of standard experiment

表 4. 标准试验系统结果数据

| 数据集 | Metric | CNN2D | LSTM | RNN | MSCNN | ViT | SF | WTMSA |
|-------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|
| Indian Pines | Overall Accuracy (%) | 43.13 | 75.30 | 74.88 | 89.70 | 79.44 | 89.49 | 98.38 |
| | Kappa (%) | 38.18 | 71.66 | 71.18 | 88.23 | 76.49 | 88.09 | 98.16 |
| | AA Mean (%) | 51.03 | 58.69 | 60.00 | 84.93 | 73.80 | 80.77 | 94.60 |
| Salinas | Overall Accuracy (%) | 63.82 | 86.92 | 74.55 | 93.02 | 91.14 | 93.9 | 99.90 |
| | Kappa (%) | 61.83 | 85.37 | 73.95 | 92.26 | 90.19 | 93.22 | 99.89 |
| | AA Mean (%) | 60.97 | 91.37 | 75.61 | 96.58 | 95.86 | 95.63 | 99.93 |
| Pavia University | Overall Accuracy (%) | 50.16 | 62.95 | 78.07 | 94.03 | 80.18 | 81.48 | 98.80 |
| | Kappa (%) | 52.56 | 53.87 | 71.28 | 91.93 | 74.26 | 76.03 | 98.38 |
| | AA Mean (%) | 57.22 | 74.18 | 80.63 | 93.98 | 87.09 | 87.26 | 98.70 |

**Figure 2.** Confusion Matrices of Different Datasets. (a) Indian Pines, (b) Salinas, (c) Pavia University**图 2.** 不同数据集的混淆矩阵。(a) Indian Pines, (b) Salinas, (c) Pavia University

没有混淆，说明模型在分类这些类别时表现非常出色。在 Pavia University 数据集中，总体分类准确率较高，主对角线上的值占绝大多数，说明模型能够很好地对大部分类别进行正确分类。

然而，非对角线上的值反映了模型在某些类别上的分类错误。例如，在 Indian Pines 数据集中，第 1 类总样本数较少(40 个)，正确分类 29 个，部分样本被误分类为第 8 类和第 11 类；第 4 类和第 12 类之间存在显著混淆，第 12 类部分样本被错误分为第 3 类。在 Salinas 数据集中，少量非对角线值表明个别类别之间可能存在边界模糊或特征相似。在 Pavia University 数据集中，类别 1 中有 117 个样本被错误分类为类别 7，70 个样本被分类为类别 8；类别 2 中有 13 个样本被误分类为类别 6，这些错误可能反映了类别特征的相似性或模型对部分类别特征提取的不足。

综上所述, 三个数据集的分类结果整体表现优异, 主对角线占据主导地位, 说明模型对大部分类别的特征具有较强的区分能力。然而, 个别类别之间仍存在一定程度的混淆, 可能由光谱特征相似、空间邻近性或样本数量不平衡引起。接下来会进一步优化模型可以通过改进特征提取和数据增强策略来减少混淆, 从而提升分类性能。

3.6. 可视化分析

在图 3~5 中, 分别展示了 WTMSA 模型和其他模型在 Indian Pines、Salinas 和 Pavia University 数据集上的分类可视化结果。从图中可以清晰地看出, 与其他对比方法相比, 所提出的 WTMSA 模型生成的分类图具有更好的类别边界清晰度和区域一致性。

具体来说, 在图 3 中, WTMSA 模型对 Indian Pines 的分类具有较高的准确性, 特别是在一些复杂类别如未耕种的玉米地和浅耕的玉米种植区的边界区域表现出显著的优势, 减少了噪声点的出现, 形成了更加平滑的分类结果。在图 4 中, WTMSA 模型能够有效地区分 Salinas 蔬菜地、裸土以及葡萄园等多种

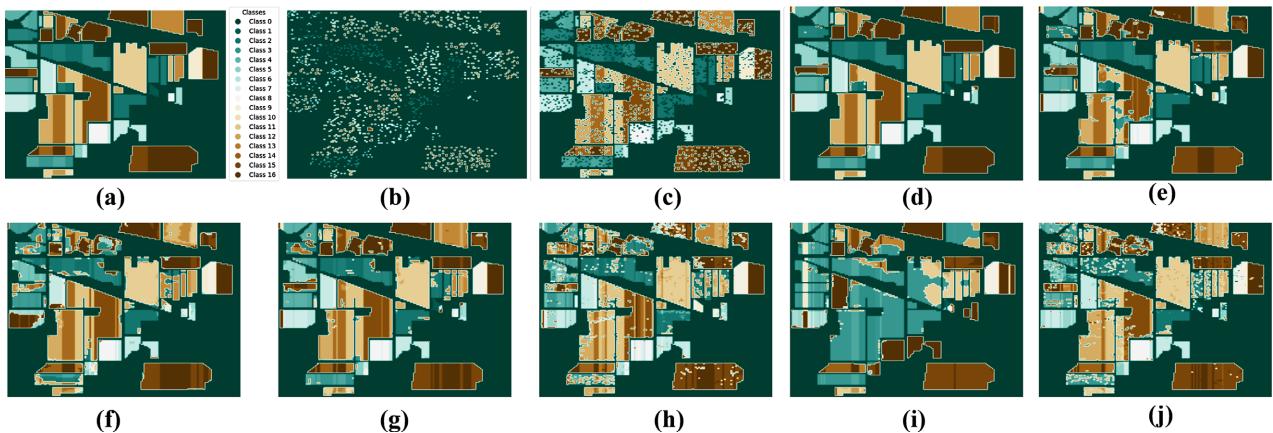


Figure 3. Visualization of the dataset division and classification visualizations of all methods on the Indian Pines dataset. (a) Ground truth, (b) Train set, (c) Test set, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

图 3. 数据集划分可视化以及所有方法在 Indian Pines 数据集上的分类可视化图。(a) 真实值, (b) 训练集, (c) 测试集, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

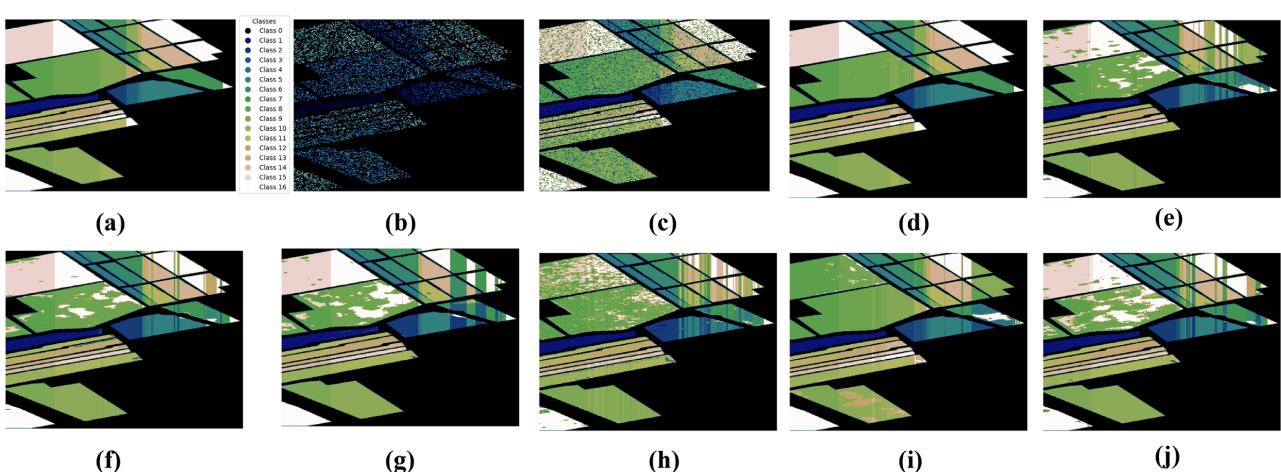


Figure 4. Visualization of the dataset division and classification visualizations of all methods on the Salinas dataset. (a) Ground truth, (b) Train set, (c) Test set, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

图 4. 数据集划分可视化以及所有方法在 Salinas 数据集上的分类可视化图。(a) 真实值, (b) 训练集, (c) 测试集, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

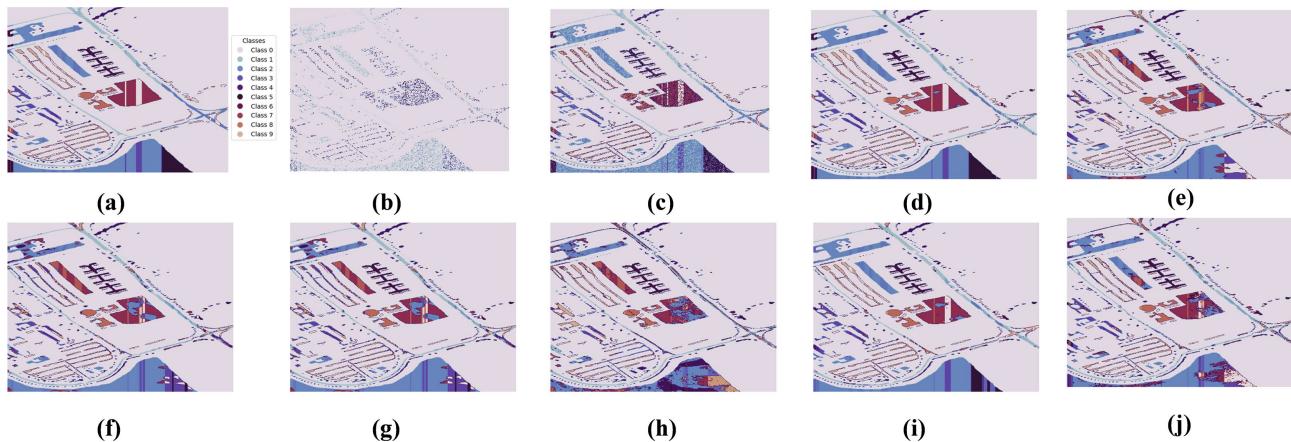


Figure 5. Visualization of the dataset division and classification visualizations of all methods on the Pavia University dataset. (a) Ground truth, (b) Train set, (c) Test set, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

图 5. 数据集划分可视化以及所有方法在 Pavia University 数据集上的分类可视化图。(a) 真实值, (b) 训练集, (c) 测试集, (d) WTMSA, (e) SF, (f) RNN, (g) MSCNN, (h) LSTM, (i) CNN2D, (j) ViT

复杂地物类别。该模型的结果能够很好地保留大型地物结构的完整性，同时显著减少了类别间的混淆。在图 5 中，在 Pavia University 数据集的分类可视化分析中，WTMSA 模型表现出对建筑物、道路和植被等主要地物类别的优异分类能力，特别是在沥青和裸露的土壤等容易混淆的类别中，分类边界清晰且过渡自然。

综上所述，通过对三组高光谱遥感数据集的分类可视化分析可以得出，所提出的 WTMSA 模型不仅能够在大规模地物场景中显著提升分类精度，还能在边界处理和小类别上取得显著优势，全面验证了模型的有效性和鲁棒性。

4. 总结

本研究提出了一种基于小波变换卷积和多尺度线性注意力机制的高光谱图像分类方法。与传统方法相比，该方法在多尺度特征提取、计算复杂度和分类精度方面实现了显著突破。具体而言，通过小波变换卷积有效捕获高光谱图像的多尺度光谱特征，克服了传统卷积感受野有限的问题；引入多尺度线性注意力机制提升了全局特征建模能力这些技术创新使得本方法在处理高维高光谱数据时展现出高效性与适应性，为实际应用提供了强有力的支持。

随着遥感技术与高光谱传感器的不断发展，高光谱图像数据的复杂性和规模将持续增加。本文所采用的多尺度特征提取与设计理念为高光谱图像分类领域提供了重要参考，但仍存在进一步改进的空间。可以探索更具创新性和有效性的多尺度注意力机制，以充分挖掘高光谱图像的空间和光谱关联信息。结合未来的硬件加速技术，本文的算法框架有望进一步优化并实现实时处理能力，从而在农业监测、环境保护和资源勘测等领域发挥更大价值。

基金项目

国家自然科学基金(No.22373075)。

参考文献

- [1] Freitas, S., Silva, H. and Silva, E. (2022) Hyperspectral Imaging Zero-Shot Learning for Remote Marine Litter Detection and Classification. *Remote Sensing*, **14**, Article 5516. <https://doi.org/10.3390/rs14215516>
- [2] Yuan, J., Wang, S., Wu, C. and Xu, Y. (2022) Fine-Grained Classification of Urban Functional Zones and Landscape

- Pattern Analysis Using Hyperspectral Satellite Imagery: A Case Study of Wuhan. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **15**, 3972-3991. <https://doi.org/10.1109/jstars.2022.3174412>
- [3] Wang, C., Liu, B., Liu, L., Zhu, Y., Hou, J., Liu, P., et al. (2021) A Review of Deep Learning Used in the Hyperspectral Image Analysis for Agriculture. *Artificial Intelligence Review*, **54**, 5205-5253. <https://doi.org/10.1007/s10462-021-10018-y>
- [4] Kumar, D., Shankar, P.S. and Jha, R. (2024) Identifying Rocks and Mineral Resources Using Hyper Spectral Analysis. 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, 15-16 March 2024, 1-6. <https://doi.org/10.1109/aimla59606.2024.10531314>
- [5] Ma, L., Crawford, M.M. and Tian, J. (2010) Local Manifold Learning-Based k -Nearest-Neighbor for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 4099-4109. <https://doi.org/10.1109/tgrs.2010.2055876>
- [6] Fauvel, M., Benediktsson, J.A., Chanussot, J. and Sveinsson, J.R. (2008) Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, 3804-3814. <https://doi.org/10.1109/tgrs.2008.922034>
- [7] Ham, J., Chen, Y.C., Crawford, M.M. and Ghosh, J. (2005) Investigation of the Random Forest Framework for Classification of Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 492-501. <https://doi.org/10.1109/tgrs.2004.842481>
- [8] Huang, X. and Zhang, L. (2013) An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **51**, 257-272. <https://doi.org/10.1109/tgrs.2012.2202912>
- [9] Fang, L., Yan, Y., Yue, J. and Deng, Y. (2023) Toward the Vectorization of Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-14. <https://doi.org/10.1109/tgrs.2023.3299154>
- [10] Ge, Z., Cao, G., Li, X. and Fu, P. (2020) Hyperspectral Image Classification Method Based on 2D-3D CNN and Multi-branch Feature Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 5776-5788. <https://doi.org/10.1109/jstars.2020.3024841>
- [11] Zhou, W., Kamata, S., Luo, Z. and Wang, H. (2022) Multiscanning Strategy-Based Recurrent Neural Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-18. <https://doi.org/10.1109/tgrs.2021.3138742>
- [12] Liang, H., Bao, W., Shen, X. and Zhang, X. (2021) Spectral-Spatial Attention Feature Extraction for Hyperspectral Image Classification Based on Generative Adversarial Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 10017-10032. <https://doi.org/10.1109/jstars.2021.3115971>
- [13] Chen, Y., Jiang, H., Li, C., Jia, X. and Ghamisi, P. (2016) Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 6232-6251. <https://doi.org/10.1109/tgrs.2016.2584107>
- [14] Haut, J.M., Paoletti, M.E., Plaza, J., Plaza, A. and Li, J. (2019) Visual Attention-Driven Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **57**, 8065-8080. <https://doi.org/10.1109/tgrs.2019.2918080>
- [15] Zhou, F., Hang, R., Liu, Q. and Yuan, X. (2019) Hyperspectral Image Classification Using Spectral-Spatial LSTMs. *Neurocomputing*, **328**, 39-47. <https://doi.org/10.1016/j.neucom.2018.02.105>
- [16] Vaswani, A., et al. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [17] Alexey, D. (2020) An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [18] Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., et al. (2022) SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-15. <https://doi.org/10.1109/tgrs.2021.3130716>
- [19] He, X., Chen, Y. and Lin, Z. (2021) Spatial-Spectral Transformer for Hyperspectral Image Classification. *Remote Sensing*, **13**, Article 498. <https://doi.org/10.3390/rs13030498>
- [20] Mou, L. and Zhu, X.X. (2020) Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 110-122. <https://doi.org/10.1109/tgrs.2019.2933609>