

# 多尺度特征融合的高光谱图像分类网络<sup>①</sup>

施 伟<sup>1</sup>, 冯先伟<sup>2</sup>

<sup>1</sup>(辽宁工程技术大学 应用技术与经济管理学院, 阜新 123000)

<sup>2</sup>(辽宁工程技术大学 软件学院, 葫芦岛 125105)

通信作者: 施 伟, E-mail: [1277403835@qq.com](mailto:1277403835@qq.com)



**摘 要:** 高光谱图像每个像素波段的关联性强, 在进行识别的过程中容易出现同谱异物的问题, 并且像素波段的维度也较高, 传统的模型难以在光谱空间和二维空间之间建立有效的特征关联. 针对这些问题, 本文提出一种结合分层深度可分离卷积与图卷积和分组自注意力 (group separable self-attention, GSA) 机制多尺度特征融合 (SSG-VIT) 的高光谱图像分类模型. 首先利用分层深度可分离卷积这种分层次的结构用不同大小的卷积核多尺度的提取高光谱图像中的局部空间特征并利用 GSA 捕获全局空间特征关系, 同时并行引入图卷积模块针对高光谱图像的光谱特征进行结构化特征建模去除冗余信息增强光谱特征的表达能力, 最后采用自适应特征融合 (adaptive feature fusion, AFF) 机制进行融合. 模型在 Indian pines、Salinas、Botswana 这 3 个高光谱数据集上进行了多次实验, 结果显示本模型的总体精度 (OA), 分别达到了 99.32%、99.67%、99.69%.

**关键词:** 高光谱图像; 深度可分离卷积; 图卷积; 自适应特征融合; 分层卷积; 分组自注意力机制

引用格式: 施伟, 冯先伟. 多尺度特征融合的高光谱图像分类网络. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9887.html>

## Hyperspectral Image Classification Network Based on Multi-scale Feature Fusion

SHI Wei<sup>1</sup>, FENG Xian-Wei<sup>2</sup>

<sup>1</sup>(School of Applied Technology and Economics and Management, Liaoning Technical University, Fuxin 123000, China)

<sup>2</sup>(Software College, Liaoning Technical University, Huludao 125105, China)

**Abstract:** The spectral correlation of each pixel in hyperspectral images is strong, making it prone to the “same spectrum, different objects” problem during the recognition process. In addition, the high dimensionality of spectral bands poses challenges for traditional models in effectively associating spectral and spatial features. To address these issues, an SSG-VIT model is proposed, integrating hierarchical depthwise separable convolution, graph convolution, and group separable self-attention (GSA) for multi-scale feature fusion. Specifically, hierarchical depthwise separable convolution is employed to extract local spatial features at multiple scales using different kernel sizes, while GSA captures global spatial relationships. A graph convolution module is incorporated to model the structured spectral features, eliminate redundant information, and enhance spectral feature representation. Finally, an adaptive feature fusion (AFF) mechanism integrates the spatial and spectral features. The proposed model is evaluated on three hyperspectral datasets: Indian pines, Salinas, and Botswana. The overall accuracies (OA) achieved are 99.32%, 99.67%, and 99.69%, respectively, across multiple experiments.

**Key words:** hyperspectral image; depth-separable convolution; graph convolution; adaptive feature fusion (AFF); hierarchical convolution; group separable self-attention (GSA) mechanism

① 基金项目: 国家自然科学基金 (42071428)

收稿时间: 2024-11-28; 修改时间: 2025-01-07; 采用时间: 2025-01-21; csa 在线出版时间: 2025-04-25

## 1 引言

高光谱成像技术是一种结合了成像技术与光谱分析的先进技术,与传统多光谱成像相比高光谱成像的光谱分辨率更高,每个像素点可提供几十到数百个波段的数据可以用来区分不同材料或物质.因此在环境保护<sup>[1]</sup>、城市规划<sup>[2]</sup>、精准农业<sup>[3]</sup>、国防<sup>[4]</sup>等领域得到了广泛的应用.由于在高光谱图像(hyperspectral image, HSI)中每个像素的数百个波段之间具有很强的相关性,所以HSI在分类过程中容易出现同谱异物的问题,其次由于高光谱图像相较于普通数字图像每个像素拥有数百个光谱波段,分类过程如何避免产生维度爆炸也是一个难题.为解决这些问题,研究人员不断的进行研究和探索,在早期阶段出现了各种传统机器学习方法,如K-近邻算法(K-nearest neighbor, KNN)<sup>[5]</sup>、稀疏表示<sup>[6]</sup>、形态轮廓<sup>[7]</sup>、支持向量机(support vector machine, SVM)<sup>[8]</sup>、随机森林(RF)<sup>[9]</sup>等方法,这些传统机器学习算法主要针对高光谱图像的光谱特征进行处理.如KNN算法它通过比较样本光谱特征之间的相似性来进行分类, SVM则是一种基于超平面的分类方法,其目的是找到一个最优的超平面来最大化类间距离.而高光谱图像数据通常是高维且复杂的并且可能非线性可分,所以要选择合适的核函数将高维数据映射到低维空间使其线性可分.但这些传统机器学习算法往往忽略了对于高光谱图像空间特征的表达导致其分类效果的不理想.并且在过去针对高光谱图像的高纬度问题,主要采用的是主成分分析<sup>[10]</sup>的方法,但是高光谱数据中的光谱特征通常是非线性的,尤其是在不同地物或材料之间的光谱差异上线性降维方法可能会丢失重要的非线性特征,从而影响分类性能.后来随着深度学习领域的不断发展,卷积神经网络(convolutional neural network, CNN)被广泛应用于HSI的处理任务中.如Chhapariya等人<sup>[11]</sup>提出一种多层级的3D-CNN模型,采用3D-CNN在对高光谱数据进行光谱-空间特征提取时比传统方法具有更好的表现,通过卷积操作在能够有效降维的同时还能提取重要的特征信息.另外Hu等人<sup>[12]</sup>提出了一种2D卷积神经网络对于HSI的空间特征进行提取,解决了传统机器学习算法对于空间特征的忽略.虽然2D-CNN扩大了感受野使网络可以从相邻像素中学习到相关信息,但是无法捕捉相邻波段间的光谱信息.为解决这一问题, Roy等人<sup>[13]</sup>提出了一种基于2D-CNN和3D-CNN的HybridSN模型,利用

3D-CNN提取HSI的光谱特征2D-CNN学习HSI的空间特征,这种混合结构降低了模型的复杂度提高了分类效果. Guo等人<sup>[14]</sup>则提出了一种基于卷积神经网络的新型空间特征融合算法,将空间特征融合到光谱特征中提高了分类效果.对于高光谱图像在分类过程中出现的同谱异物问题, Zhou等人<sup>[15]</sup>提出了一种光谱-空间长短期记忆网络(spectral-spatial LSTM)通过将其不同通道的光谱值逐一输入到光谱长短期记忆网络(LSTM)中以学习其中的光谱特征. Zhong等人<sup>[16]</sup>则开发了一种三维空间光谱残差网络(SSRN),用于使用多个连接的残差块来提取空间光谱特征.

如今基于Transformer的网络在HSI分类任务中也取得了巨大的成功,如Hong等人<sup>[17]</sup>在ViT(vision Transformer)的基础上引入了SpectralFormer,使模型能够从相邻的光谱通道中学习光谱特征,同时还引入了跨层自适应融合促进了深层特征和浅层特征的融合. Zhao等人<sup>[18]</sup>开发出了一种GSC-ViT网络,采用逐点自注意(PWSA)增强了对全局空间特征的提取,并且采用具有增强跳过连接机制的简单逐点层替代ViT中的多层感知器(MLP)层以消除不必要的非线性变换提高了分类的效果.为了解决感受野受限的问题, He等人<sup>[19]</sup>提出了具有全局感受野的模型HSI-BERT.此外对于Transformer对空间信息利用不足的问题, Roy等人<sup>[20]</sup>设计了一种将CNN提取的空间特征与Transformer进行结合,利用密集连接来捕获光谱序列信息的网络.对于处理模型稳定性的问题中, 晏远翔等人<sup>[21]</sup>提出一种生成式对抗网络缓解GAN模型崩溃的问题.

针对上述问题本文则提出了一种结合分层深度可分离卷积以及图卷积和GSA多尺度特征融合的高光谱图像分类模型,本文相关研究和如下.

(1) 针对HSI数据中像素的光谱特征采用图卷积模块进行特征提取, GCN能够将高光谱图像的空间和光谱信息同时表示为图结构从而更好地对像素间的相关性和波段间的关联性进行建模,能够更好地处理高光谱图像中的同谱异物问题.

(2) 提出了一种分层深度可分离卷积和GSA注意力机制的结构.采用分层次的结构对高光谱图像的局部空间特征进行提取,分层深度可分离卷积采用不同大小的卷积核能够更全面地捕获像素间复杂的空间和光谱关系,以及采用GSA注意力机制对长距离依赖关系进行建模并提取全局空间特征逐步的聚合特征,使

模型能够在不同层次上提取高频和低频信息.

(3) 提出了图卷积与 GSA 注意力机制的双分支结构, 分别对高光谱图像的光谱特征以及空间特征进行提取, 并利用自适应加权融合机制引入可学习参数自适应调节 GCN 与 GSA 的输出权重, 以提高模型分类效果.

## 2 模型架构

基于 SSG-VIT 的高光图像分类模型主要包含 4 部分. 如图 1 所示, 该模型基于分层深度可分离卷积模块 (SSD-CBAM), GSA 模块以及图卷积模块和自适应特征融合模块.

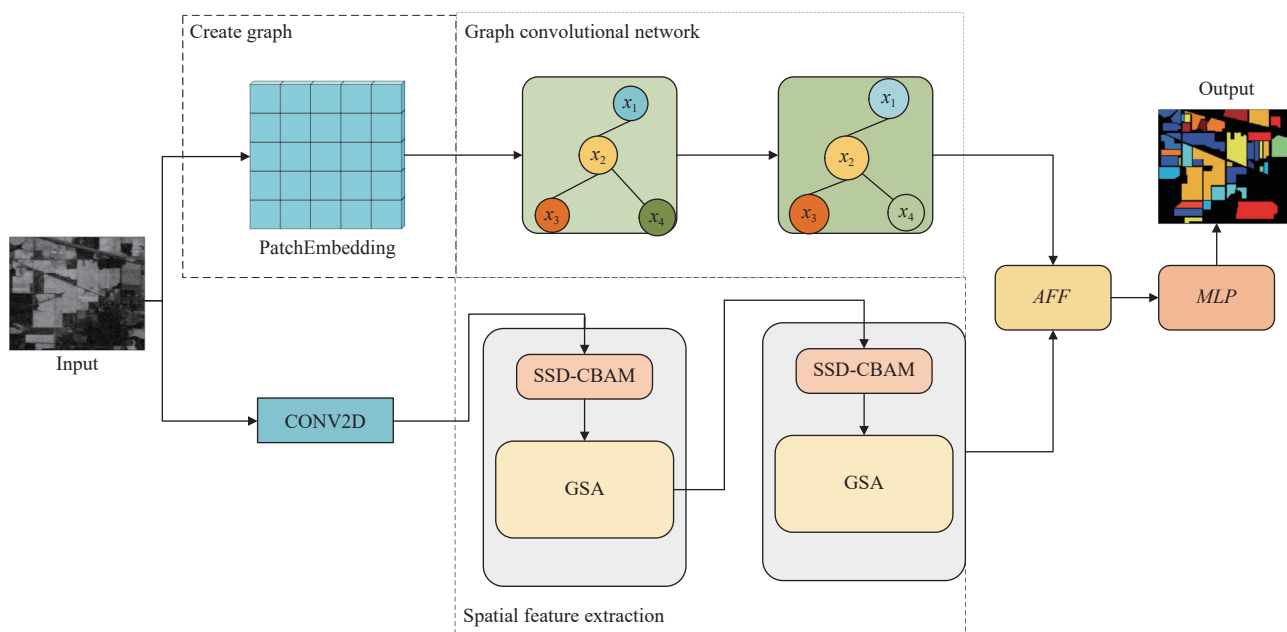


图 1 算法整体框架

首先高光谱图像将经过 PatchEmbedding 层以及卷积层的处理. 因为 HSI 中的每个像素具有波段多、数据维度高的特点, 需要经过卷积层进行特征提取与降维减少不必要的冗余信息同时保留其重要的特征. 经过卷积层进行特征提取后原先的高光谱数据维度由  $CWH$  ( $C$  为波段数,  $W$  为波段跨密度,  $H$  为波段高度) 变为  $BWH$  ( $B$  为进行特征提取之后波段数), 并且输入的图像需要经过 PatchEmbedding 层变为图结构数据才能送入图卷积网络中进行特征提取. 在构建图结构数据时之所以采用 PatchEmbedding 方法而非传统的超像素分割<sup>[22]</sup>的方法, 因为在 PatchEmbedding 模块中利用卷积操作对高光谱图像进行块划分能够通过固定的划分规则确保光谱信息在每个 Patch 内保持完整而不会出现因分割误差导致波段信息丢失或不均匀分布的问题. 高光谱图像数据在经过 PatchEmbedding 层之后会变为若干个 Patch, 然后通过对所有 Patch 之间的余弦相似度进行判别并构建邻接矩阵以此作为图卷积网络的输入. 另外高光谱数据在经过 CONV2D 的特征

提取后会进入分层深度可分离卷积模块 (SSD-CBAM), 在每一层中首先采用大小为 (1, 1) 的卷积核进行逐点卷积操作, 这种方式能够高效地提取光谱信息并实现特征降维而不受空间维度的干扰, 同时在保证特征表达能力的前提下, 逐点卷积大幅减少了计算资源的消耗. 之后每一层采用不同大小的卷积核对于高光谱数据的局部空间特征进行多尺度特征提取, 最后采用通道-空间 (channel-space attention mechanism, CBAM) 注意力机制对于融合之后的局部空间特征再次提取. 因为在高光谱数据中具有复杂的空间和光谱信息且这两种信息之间往往存在深层次的依赖关系, CBAM 注意力能够通过跨注意力机制, 动态的学习空间和光谱通道之间的相互作用, 从而更精确地捕捉到数据中隐含的多维度信息. 同时 CBAM 注意力能够更细粒度地建模不同波段之间的相互关系帮助提取出更具区分性的特征, 之后将数据送入 GSA 模块中对全局空间进行特征提取. 最后与图卷积网络输出的特征进行自适应特征融合之后送入 MLP 层进行分类预测, 总体网络计算

过程如式 (1) 所示:

$$F_{out} = AFF[(f_{gsa}(f_{ssd-cbam}(x)), f_{gcn}(x))] \quad (1)$$

其中,  $F_{out}$  表示融合之后的特征数据,  $x$  表示输入的高光谱数据,  $f_{gsa}$  表示 GSA 模块对于高光谱数据的处理结果,  $f_{ssd-cbam}$  表示 SSD-CBAM 模块的输出,  $f_{gcn}$  表示 GCN 模块对于高光谱数据的处理结果.  $AFF$  表示自适应特征融合模块.

### 2.1 基于图卷积模块的特征提取

采用图卷积网络是因为 GCN 允许节点 (像素) 之间的信息在图结构中传播并能够捕捉全局的特征关系. 这种全局信息的共享有助于模型能更好地区分光谱特征相近的类别以处理分类过程中的同谱异物问题. 同时针对高光谱图像这种具有数百个波段的高维数据, GCN 可以有效降维去除冗余特征保留关键信息减少维度灾难的问题从而提升分类效率和性能. 同时 GCN 还能够有效捕捉像素间的空间和光谱关系, 可以利用邻近像素的特征来增强分类性能. 并且 GCN 对噪声和异常值的影响也相对较小, 通过邻域信息聚合能够增强对局部噪声的鲁棒性使得模型在噪声较多的场景下依然可以获得较好的分类结果, 增强了模型的稳定性.

在图卷积网络中对于输入的图片首先经过 Patch-

Embedding 变成图节点并创建邻接矩阵, 之后送入 GCN 层中对像素节点的光谱特征进行特征提取. 之后进行 LeakyReLU 激活, 具体过程如图 2 所示. 高光谱数据通常具有高度的光谱冗余性和复杂的空间特征, 使用 LeakyReLU 激活可以让模型在传递特征时不会因为负值而完全丢失信息提高模型泛化能力. 在 LeakyReLU 中, 负值部分会以一个很小的斜率通过, 这样确保了即使输入为负神经元也能保持一定的梯度避免梯度消失, 之后采用 Dropout 减轻模型的计算复杂度. 图卷积模块的计算过程如式 (2)–式 (5) 所示:

$$F_{gcn\_out} = \partial(\sigma(F_{gcn}(F_{pe}(x)))) \quad (2)$$

$$L_{sym} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \quad (3)$$

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)}) \quad (4)$$

$$\tilde{A} = A + I \quad (5)$$

其中,  $x$  表示输入的特征,  $F_{pe}$  表示高光谱图像经过处理变成图数据,  $\tilde{A}$  表示带自连的邻接矩阵,  $I$  表示单位矩阵,  $\tilde{D}$  表示度矩阵,  $H^{(l)}$  是第  $l$  层的节点特征表示,  $W^{(l)}$  是第  $l$  层的权重矩阵,  $L_{sym}$  表示标准化的拉普拉斯矩阵, 用来保证 GCN 中信息传播的稳定性.  $\sigma$  表示 LeakyReLU 激活函数用于引入非线性,  $\partial$  表示 Dropout.

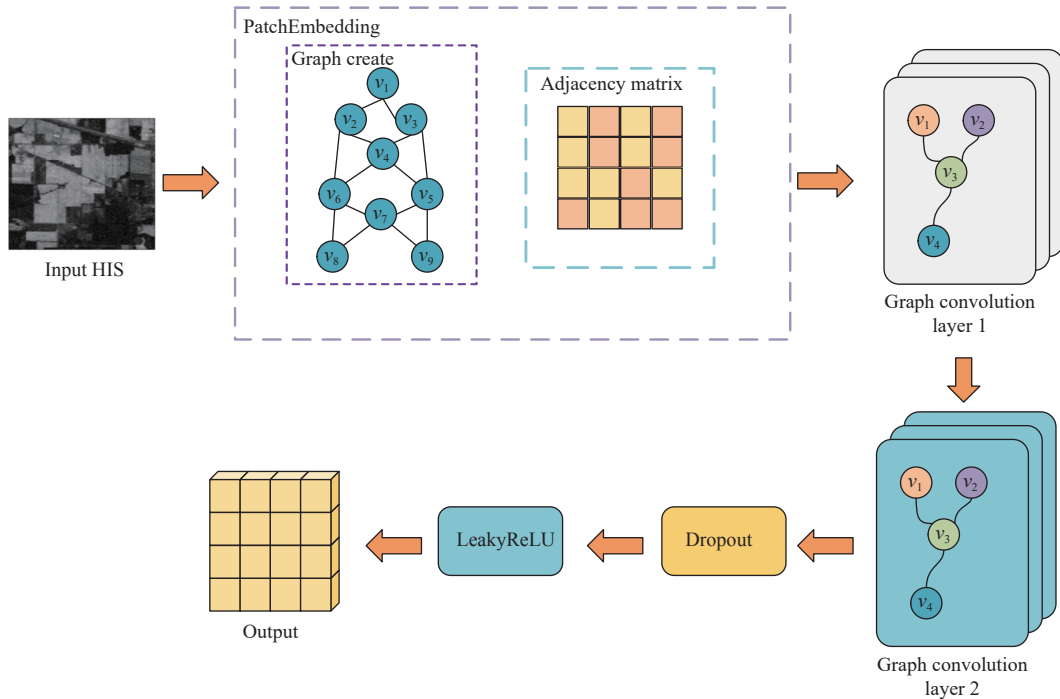


图 2 图卷积 (GCN) 模块结构示意图



### 2.2 构建邻接矩阵

在邻接矩阵中顶点集有 HIS 像素构成, 边集由任意两个顶点  $v_i$  和  $v_j$  之间相似性组成 (余弦相似度 (cosine similarity)<sup>[23]</sup>). 图邻接矩阵为相同类别之间顶点连接的集合. 一般来说类内节点的谱值比类间节点的谱值更接近, 所以图的邻接矩阵我们采用了余弦相似度的方法进行创建. 余弦相似度是一种衡量两个向量之间相似度的指标, 适用于高维空间中两个向量是否相似的判别. 通过对所有的节点的遍历, 如果某两个节点  $v_i$  和  $v_j$  的相似度超过规定的阈值, 则在邻接矩阵中将两个节点的连接关系设置为 1, 否则为 0. 计算过程如式 (6) 所示:

$$CosineSimilarity(A_i, B_j) = \frac{A_i \cdot B_j}{\|A_i\| \times \|B_j\|} \quad (6)$$

其中,  $A_i$ 、 $B_j$  表示为特征向量.

### 2.3 分层深度可分离卷积模块 (SSD-CBAM)

采用分层深度可分离卷积能有效提取空间和光谱特征, 增强了对高维数据的处理能力降低模型的过拟

合风险. 如图 3 所示, 采用这种分层次的结构在每一层使用不同大小和形状的卷积核, 构建了一个多尺度、多方向、逐层融合的特征提取网络. 在第 1 层我们使用大小为 (1, 1), (3, 3), (1, 3), (3, 1) 的卷积核捕捉不同方向的空间局部特征同时结合光谱信息. 在第 2 层我们使用大小为 (1, 1), (3, 3), (3, 3) 的卷积核进一步扩大感受野, 同时增强不同尺度的空间上下文信息提取能力. 并且采用不同卷积核尺寸的组合, 能够捕获小目标的细节特征和大目标的全局信息. 第 3 层第 4 层分别采用大小为 (1, 1), (3, 3) 和 (1, 1) 的卷积核逐层融合光谱和空间特征, 实现对高光谱图像的多维特征提取. 并且通过深度可分离卷积的方式, 将标准卷积拆分为逐通道卷积和逐点卷积. 逐通道卷积对每个输入通道单独进行卷积操作, 之后采用逐点卷积使用大小为 (1, 1) 的卷积核在逐通道卷积的输出上进行逐点卷积操作, 同时引入空间-通道注意力机制 (CBAM) 提取再次各层学习的局部空间特征.

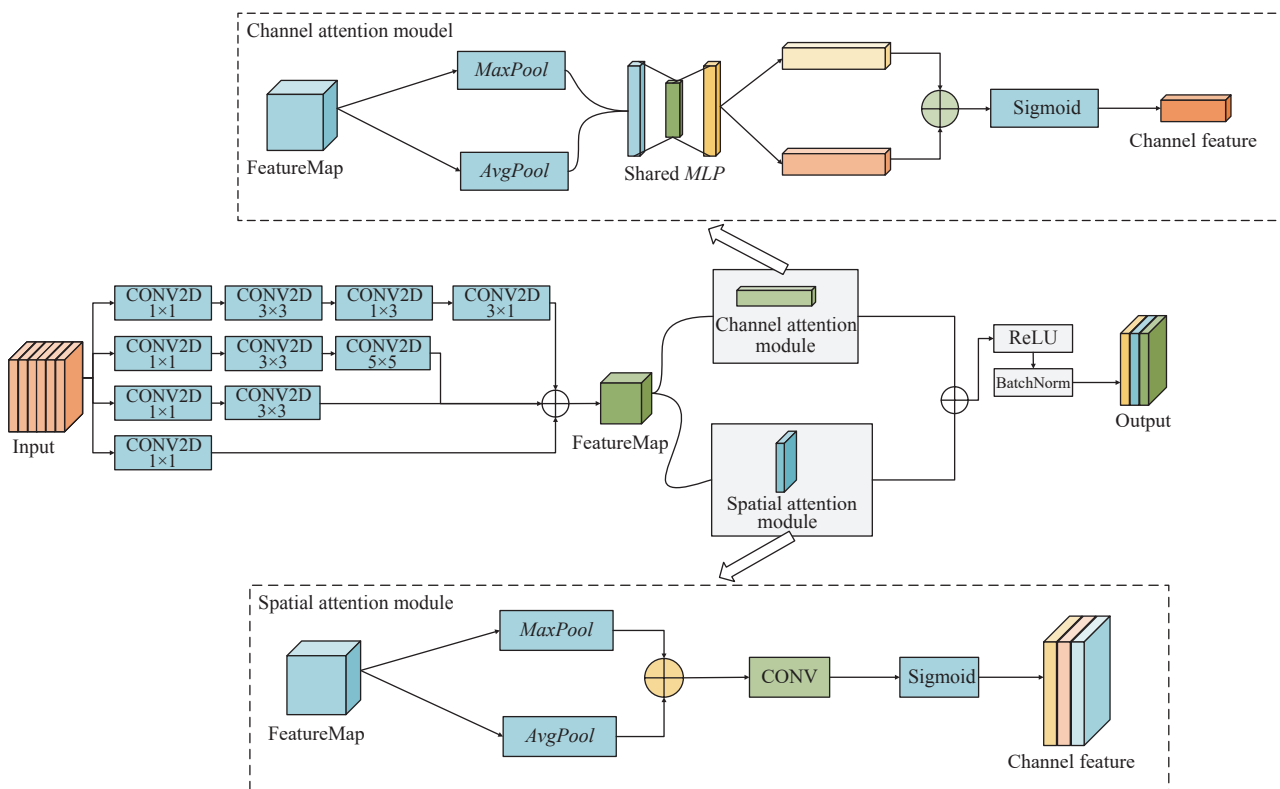


图 3 分层深度可分离卷积 (SSD-CBAM) 模块结构示意图

分层深度可分离卷积模块的每一层可以专注于学习特定的空间特征, 从而使模型能够识别更加复杂的形状和纹理. 并且在针对每个波段对应不同的物理信

息中, 逐通道卷积避免了不同通道间的权值共享, 从而更好地保留每个波段的特征. 而后通过空间-通道注意力机制 (CBAM) 对于卷积层的输出进行重塑和定向.

并且分层次的结构允许网络在不同层次上学习不同尺度的空间特征,能够更好地捕捉图像中的局部空间信息.同时分层深度可分离卷积模块可以有效融合来自不同层的局部空间特征,增强模型对复杂背景和多样性目标的适应能力.计算机过程如式(7)–式(13)所示:

$$F_1 = \sigma_{(3,1)}(\sigma_{(1,3)}(\sigma_{(3,3)}(\sigma_{(1,1)}(X)))) \quad (7)$$

$$F_2 = \sigma_{(5,5)}(\sigma_{(3,3)}(\sigma_{(1,1)}(X))) \quad (8)$$

$$F_3 = \sigma_{(3,3)}(\sigma_{(1,1)}(X)) \quad (9)$$

$$F_4 = \sigma_{(1,1)}(X) \quad (10)$$

$$F_{\text{cn-out}} = F_1 + F_2 + F_3 + F_4 \quad (11)$$

$$M_c(F_{\text{cn-out}}) = \sigma(\text{MLP}(\text{AvgPool}(F_{\text{cn-out}})) + \text{MLP}(\text{MaxPool}(F_{\text{cn-out}}))) \quad (12)$$

$$M_s(F_{\text{mc}}) = \varepsilon(f^{7 \times 7}(\text{AvgPool}(F_{\text{mc}}); \text{MaxPool}(F_{\text{mc}}))) \quad (13)$$

其中,  $X$  表示输出的特征图,  $\sigma_{(i,j)}$  表示不同大小的卷积核,  $F_1$ 、 $F_2$ 、 $F_3$ 、 $F_4$  分别表示第 1–4 层的卷积输出,  $F_{\text{cn-out}}$  表示卷积层的最终输出.  $M_c$  表示通道注意力的输出,  $\varepsilon$  表示非线性激活 ReLU,  $M_s$  表示融合空间注意力,  $\text{AvgPool}$  表示平均池化,  $\text{MaxPool}$  表示最大池化.

采用深度可分离卷积与标准卷积的计算量进行对比.对于输入的特征图的大小为  $H \times W \times C_{\text{in}}$  ( $H$  表示图像高度,  $W$  表示图像宽度,  $C_{\text{in}}$  表示特征图的光谱维度), 卷积核大小为  $K \times K$ , 输出通道为  $C_{\text{out}}$ , 输出的特征图大小为  $H_{\text{out}}$ 、 $W_{\text{out}}$  的特征图.对于采用标准卷积的计算量如式(14)所示:

$$CA_{\text{sc}} = H_{\text{out}} \times W_{\text{out}} \times C_{\text{out}} \times (K \times K \times C_{\text{in}}) \quad (14)$$

采用深度可分离卷积的计算量如式(15)所示:

$$CA_{\text{ds}} = H_{\text{out}} \times W_{\text{out}} \times [C_{\text{in}} \times (K \times K) + C_{\text{in}} \times C_{\text{out}}] \quad (15)$$

对于传统卷积相比,深度可分离卷积的所减少的计算量如式(16)所示:

$$\frac{C_{\text{in}} \times K^2 + C_{\text{in}} \times C_{\text{out}}}{C_{\text{in}} \times C_{\text{out}} \times K^2} \quad (16)$$

## 2.4 分组自注意力模块

GSA 表示为将传统的 MSA 进行分解.与在一个 Patch 内对所有 Token 同时进行 MSA 不同的是在 GSA 中为了提取全局信息和局部空间信息同时降低计算复杂度,我们采用沿空间维度对 patch 进行分组的方式.如图 4 所示,首先沿空间维度对 patch 进行分组,输入

特征图的空间维度首先被划分  $N$  个大小相等的子图,每个子图称为一个组,在这里引入了“组 Token”的概念,使用 *AdaptiveAvgPool2d* 的方式聚合每个组的特征生成组 Token.然后对每个组内的 Token 执行 MSA (包括组内空间像素 Token 和组 Token).重要的是 MSA 是针对每个组独立执行的,所以可以实现每个组内特征的全局表示,并且相较于传统的自注意力机制这种方式降低了计算的复杂度.所以在 GSA 的过程中,每个组 Token 通过与组内的所有 Token 进行自注意力,从而得到局部空间信息的特征表达.然后再使用这些组 Token 计算每个组之间的注意力关系,这种方式强调了对于全局信息的表达,最后将组间特征与组内特征进行融合之后经过 GELU 激活和 Layer Normalization 层后生成注意力权重矩阵. GSA 的过程代表一种结合局部和全局考虑的空间特征提取方式,计算过程如式(17)–式(25)所示,假设输入的特征图  $X \in \mathbb{R}^{h \times w \times c}$ , 划分  $N$  个子图:

$$X = \{X_1, X_2, \dots, X_n\}, X_i \in \mathbb{R}^{h \times w \times c} \quad (17)$$

其中,  $H \times W = N(h \times w)$ .对于每个子图  $X_i$ , 生成一个组 Token,  $T_i \in \mathbb{R}^d$ :

$$T_i = F(X_i) \quad (18)$$

对每个组  $X$  中的所有像素  $x_i$  和组 Token  $T_i$ , 进行自注意力计算,拼接组内所有 Token.

$$T_i = [x_1, x_2, \dots, x_n, t_i] \in \mathbb{R}^{(N+1) \times c} \quad (19)$$

$$Q = T_i W^Q, K = T_i W^K, V = T_i W^V \quad (20)$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (21)$$

之后将所有的组 Token  $t_i$ , 组成新的特征矩阵:

$$G = [t_1, t_2, t_3, \dots, t_n] \in \mathbb{R}^{M \times C} \quad (22)$$

对组 Token 矩阵  $G$  执行自注意力:

$$Q_g = GW^Q, K_g = GW^K, V_g = GW^V \quad (23)$$

$$\text{Attention}(Q_g, K_g, V_g) = \text{Softmax}\left(\frac{Q_g K_g^T}{\sqrt{d_k}}\right)V \quad (24)$$

GSA 模块的整体计算过程如式(25)所示:

$$T_{\text{out}} = \sigma(F_{\text{bn}}(F_{\text{ln}}(\text{GSA}(F_{\text{ln}}(x)))))) \quad (25)$$

其中,  $\sigma$  表示非线性激活 (ReLU),  $x$  表示输入高光谱特征图,  $F_{\text{ln}}$  表示 LayerNorm 层,  $F_{\text{bn}}$  表示 BatchNorm 层.

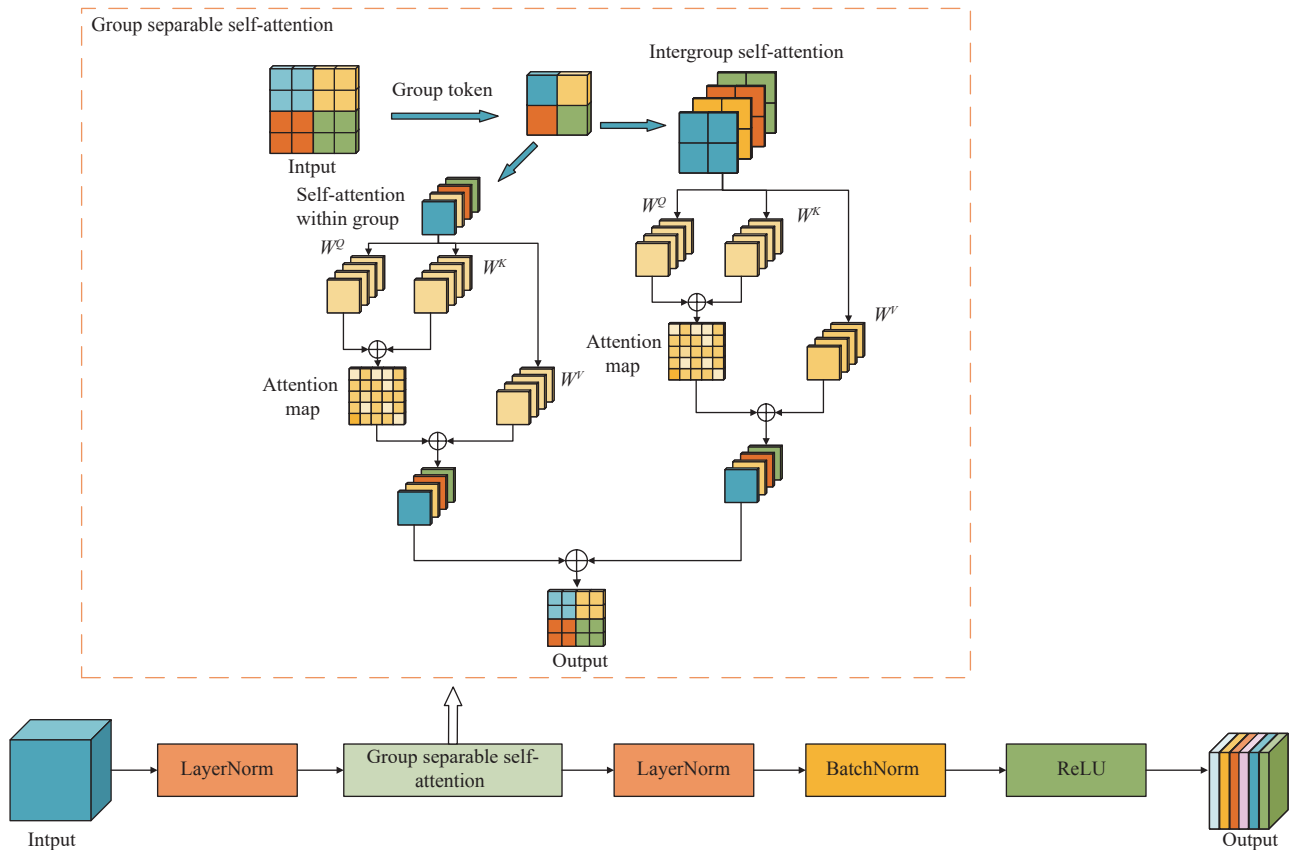


图4 GSA 模块结构示意图

### 2.5 自适应特征融合模块

GCN 能够基于邻近像素的光谱信息和空间结构进行学习,而自注意力机制可以建模数据中远距离的依赖性和复杂的全局关系.因此对于图卷积与 GSA 的输出结果我们提出了一种自适应特征融合机制.自适应特征融合模块可以将 GCN 捕捉到的光谱信息与 GSA 捕捉到的全局空间信息相结合,从而获得更加丰富和全面的特征表示.这种多视角的信息融合可以提升模型对不同结构的识别能力,从而提高预测的准确性和可靠性.

主要方式为定义两个可学习参数  $\alpha$  和  $\beta$  作为 GCN 和 GSA 特征的加权系数,采用 *Softmax* 函数对  $\alpha$  和  $\beta$  进行归一化,确保特征加权和为 1 或在特定范围内,这样避免权重过大或过小.通过引入可学习的加权系数后能够让模型在训练过程中自动调整 GCN 和 GSA 特征的权重,从而提高模型的分类效果.其计算过程如式 (26) 所示:

$$F_{\text{fusin}} = \xi(\alpha)F_{\text{gcn}} + \xi(\beta)F_{\text{gs-vit}} \quad (26)$$

其中,  $\xi$  表示 *Softmax*.

## 3 实验结果及分析

### 3.1 实验数据集

为了评估 SSG-VIT 的性能,本文采用 3 个高光谱数据 Indian pines、Salinas、Botswana 进行实验,数据集参数信息如表 1 所示.

表 1 HSI 数据集参数

数据集	Indian pines	Salinas	Botswana
采集地点	印第安纳州	萨利纳斯	博茨瓦纳
采集设备	ROSIS	AVIRIS	Hyperion
谱范围 ( $\mu\text{m}$ )	0.4–2.5	0.4–2.5	0.4–2500
大小 (像素)	145×145	512×217	1476×256
空间分辨率 (m)	20	3.7	20
波段数	224	224	242
去噪波段数	200	204	145
样本数	10249	54129	3248
类别数	16	16	14

### 3.2 实验设置

本实验所采用的开发环境以及工具为 Python 3.9, PyTorch 1.13, GPU 型号为 NVIDIA GeForce RTX3080 显存为 12 GB, RAM 为 32 GB, CPU 型号为 Intel(R) Core(R) i5-12600KF. 对于 Indian pines、Salinas、

Botswana 数据集, 随机从每个类别中选取 10%、10% 和 80% 组成训练集, 验证集与测试集. SSG-VIT 网络模型的输入窗口大小为  $8 \times 8$ , 训练的迭代次数为 200 次. 本文采用 3 个指标: 总体精度 (OA), 平均精度 (AA) 和 Kappa 系数对我们提出的模型进行有效性评估.

### 3.3 训练样本数量的影响

本节探讨了训练样本数量的不同对于不同方法的性能影响. 图 5 表示当训练集的规模为总体样本的 6%、

10% 和 20% 时, 不同方法在 3 个数据集上得到的总体正确率, 需要注意的是样本是根据每个类别的百分比随机选择的. 从图 5(a) 中可看出在 Indian pines 数据集中随着训练样本的增长, 采用 2D-CNN 模型进行训练分类结果反而出现了下降. 而我们的模型在采用较少的训练样本时在 3 个数据集中依然能够保持较高的分类效果, 这些结果证明了 SSG-VIT 的有效性和稳定性.

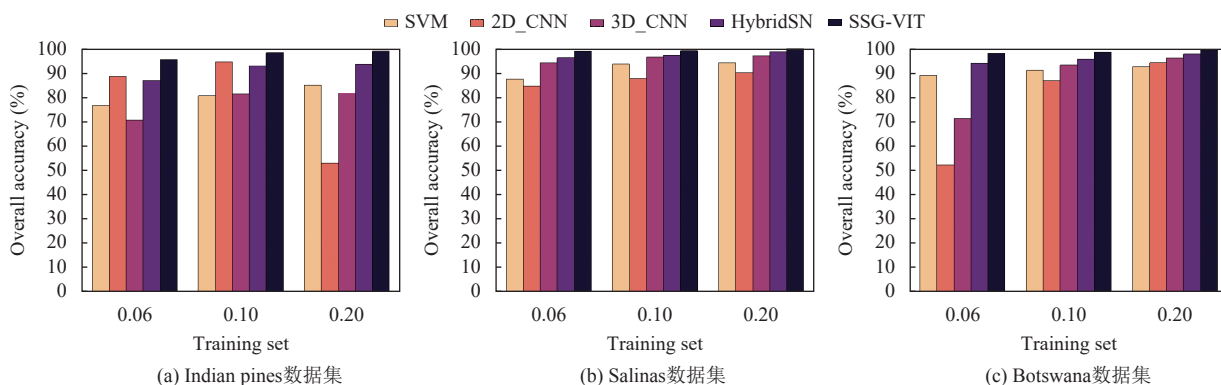


图 5 采用不同比例的训练集得到的总体准确率

### 3.4 消融实验

如前所述, 本文提出的 SSG-VIT 包含 4 个关键的模块, 图卷积模块 (GCN) 和分层深度可分离卷积模块 (SSD-CBAM), 以及 GSA 模块和自适应特征融合模块. 本节设计了消融实验来验证 4 个模块的实际效果提升, 消融实验结果如表 2 所示.

表 2 消融实验结果表 (%)

编号	GCN	SSD-CBAM	GSA	AFF	Indian pines	Salinas	Botswana
1	×	×	×	×	98.42	99.34	98.76
2	×	×	√	×	95.46	96.57	97.65
3	×	√	×	×	98.12	97.45	98.45
4	√	×	×	×	97.46	98.23	97.79
5	√	√	√	×	99.12	99.47	99.59
6	√	√	√	√	99.32	99.67	99.69

首先要以光谱-空间全局图推理 (spectral-spatial global graph reasoning) 网络<sup>[24]</sup>在 HSI 图像分类的效果为基线. 然后我们在网络中添加 GCN 模块, SSD-CBAM 模块以及 GSA 模块之后经过对比在 Indian pines、Salinas、Botswana 这 3 个数据集中分别有 0.7 个百分点、0.13 个百分点和 0.83 个百分点的提升. 但是若模型仅采用一种模块进行训练, 其训练效果都出现了不同程度的下降. 因为对于高光谱图像若只针对其中一种特征进行提取, 忽略其他特征就会导致分类效果的

不理想. 并且对于 GSA 模块的输出结果和 GCN 模块的输出结果在没有采用自适应特征融合只进行简单的拼接时, 模型的效果在 3 个数据集上的提升分别为 0.7 个百分点、0.13 个百分点、0.83 个百分点. 而在采用自适应特征融合后, 模型能够自适应的调节高光图像光谱维度, 与空间维度的重要性, 所以在 Indian pines、Salinas、Botswana 这 3 个数据集上分别有 0.9 个百分点、0.33 个百分点、0.93 个百分点提升. 实验结果表明了 SSG-VIT 模型的有效性.

### 3.5 不同模型之间的对比实验

为评估本文所提出 SSG-VIT 的模型的有效性, 我们将与 SVM、2D-CNN、3D-CNN、HybridSN 模型进行实验对比分析, 为保证其实验结果的公平性对比实验均采用相同的训练集比例 (20%)、验证集比例 (10%) 以及测试集比例 (70%) 并且实验的环境均一致. 表 3、表 4、表 5 表示不同方法在不同数据集上的实验结果, 结果显示 SSG-VIT 在 3 个不同的数据集上都表现出其出色的性能. 图 6、图 7、图 8 分别表示 3 个数据集的分类效果图, 并且在图 6 分类效果图中可以看出 SSG-VIT 模型在 Salinas 数据集的训练中表现出了特别优秀的分类效果, 其分类的边界结果十分明确.



但在 Indian pines 数据集中传统的机器学习分类模型 SVM 受到了严重的椒盐噪声的影响导致了其分类效果不是很理想. 从这些分类效果图及分类数据表中可以看出 SSG-VIT 的分类效果均优于另外的 4 种模型.

表 3 不同方法下, 在 Indian pines 数据集上获得的各类精度以及总体分类精度 OA、平均精度 AA 和 Kappa 系数 (%)

Class	SVM	2D-CNN	3D-CNN	HybridSN	SSG-VIT
1	67	58.05±26.65	38.38±10.45	77.30±16.63	91.88±5.45
2	68	84.48±3.62	61.59±4.14	86.26±3.25	97.62±1.92
3	75.00	90.52±4.49	49.77±9.96	82.95±6.48	99.04±0.62
4	67.00	88.83±2.74	24.79±5.00	88.74±4.14	96.02±3.86
5	85.00	91.95±2.53	54.94±13.21	88.05±1.50	97.46±2.05
6	86.00	96.04±1.58	95.95±1.17	96.29±2.21	99.88±0.16
7	62.00	52.80±26.70	3.20±6.40	64.00±20.71	93.00±11.66
8	86.00	97.40±1.96	97.72±1.59	96.56±1.94	100.00±0.00
9	64.00	15.56±11.86	17.50±15.00	77.50±16.11	82.86±10.69
10	69.00	80.62±4.38	51.22±2.69	82.51±3.69	97.88±1.63
11	75.00	90.62±0.71	84.05±2.51	89.06±1.79	98.07±1.24
12	63.00	73.52±5.31	28.31±7.58	70.45±5.28	96.43±0.74
13	85.00	98.92±1.49	98.59±0.73	98.92±0.59	99.58±0.56
14	88.00	96.19±1.98	95.80±1.98	96.84±1.15	99.57±0.37
15	71.00	78.10±6.81	71.97±5.14	70.32±4.14	97.48±0.72
16	86.00	97.14±2.97	92.86±3.61	97.86±2.65	98.15±0.62
OA	76.44	88.59±0.81	81.40±1.61	92.92±0.51	98.44±0.88
AA	75.00	86.98±0.92	56.32±2.91	90.62±2.02	97.50±1.09
Kappa	76.00	86.98±0.92	65.78±2.74	91.93±0.57	97.99±0.77

图 9、图 10 分别表示 SSG-VIT 模型在不同数据集的混淆矩阵图以及不同方法下的损失曲线图, 从损失曲线图中可以看出 SSG-VIT 的收敛速度是最快的, 并且其稳定性的表现也比较优异. 但是, 采用 2D-CNN 的方法在 Salinas 数据集的训练中模型没有收敛, 并且从混淆矩阵图中可以看出 SSG-VIT 模型在对于类别不均衡的处理中表现也十分优异, 对于类别数量较少的部分模型仍然能够进行精准的预测. 通过混淆矩阵对角线上的值可以了解到模型整体分类的准确率较高.

表 4 不同方法下在 Salinas 数据集上获得的各类精度以及总体分类精度 OA、平均精度 AA 和 Kappa 系数 (%)

Class	SVM	2D-CNN	3D-CNN	HybridSN	SSG-VIT
1	83.00	39.86±48.82	100.00±0.00	99.83±0.32	100.00
2	99.00	95.50±4.09	99.94±0.07	99.97±0.03	99.90
3	86.00	88.12±17.46	99.80±0.16	99.87±0.16	100.00
4	99.00	79.73±39.69	98.92±0.44	99.30±0.88	99.91
5	100.00	76.68±23.02	99.06±0.39	99.59±0.46	100.00
6	87.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00
7	99.00	66.92±39.82	99.96±0.04	99.97±0.03	100.00
8	98.00	91.39±1.99	93.22±0.39	97.10±0.10	99.89
9	78.00	99.56±0.76	99.96±0.03	99.98±0.03	100.00
10	71.00	90.50±4.45	97.80±0.79	99.37±0.38	99.81
11	100.00	58.62±45.56	99.18±1.18	99.39±0.41	99.53
12	97.00	97.67±1.55	99.55±0.53	99.64±0.31	99.94
13	87.00	96.15±6.77	100.00±0.00	99.89±0.10	99.73
14	98.00	94.93±5.52	99.67±0.27	99.53±0.26	99.65
15	99.00	64.20±14.79	89.70±0.81	95.16±1.11	99.36
16	86.00	93.54±5.42	99.45±0.61	99.14±0.40	100.00
OA	94.21	84.66±8.34	96.92±0.16	98.59±0.14	99.66±0.11
AA	92.00	83.34±11.01	98.51±0.17	99.23±0.07	99.26±0.16
Kappa	89.00	82.83±9.36	96.57±0.17	98.43±0.16	98.95±0.20

表 5 不同方法下在 Botswana 数据集上获得的各类精度以及总体分类精度 OA、平均精度 AA 和 Kappa 系数 (%)

Class	SVM	2D-CNN	3D-CNN	HybridSN	SSG-VIT
1	100.00	99.84±0.33	100.00±0.00	99.91±0.19	99.54±0.59
2	93.00	85.93±5.13	91.36±9.30	96.79±5.21	99.75±0.49
3	97.00	88.32±8.33	97.71±1.43	98.31±1.39	99.90±0.20
4	99.00	77.94±6.96	98.72±0.93	98.60±0.94	99.77±0.47
5	94.00	75.70±15.44	87.63±5.18	93.02±4.04	98.14±1.58
6	76.00	70.58±14.55	86.79±6.83	96.93±1.04	97.95±1.78
7	100.00	95.28±1.36	99.61±0.19	99.90±0.19	100.00±0.00
8	94.00	73.33±9.75	97.28±2.01	99.38±0.68	100.00±0.00
9	94.00	77.24±7.42	94.66±2.54	93.71±1.94	99.44±0.54
10	81.00	84.48±19.38	97.69±1.21	98.79±0.93	100.00±0.00
11	95.00	95.27±3.54	98.85±2.10	99.67±0.66	99.18±1.27
12	84.00	92.88±6.99	99.45±0.68	99.72±0.55	99.72±0.55
13	98.00	95.93±3.84	99.44±0.68	99.81±0.37	99.91±0.19
14	99.00	95.93±3.84	91.32±5.62	93.42±4.56	97.11±2.68
OA	92.49	94.21±1.52	96.02±0.42	97.84±0.66	99.57±0.12
AA	93.00	94.20±1.61	95.75±0.45	97.71±0.88	99.32±0.38
Kappa	93.00	93.72±1.65	95.69±0.45	97.66±0.71	99.32±0.33

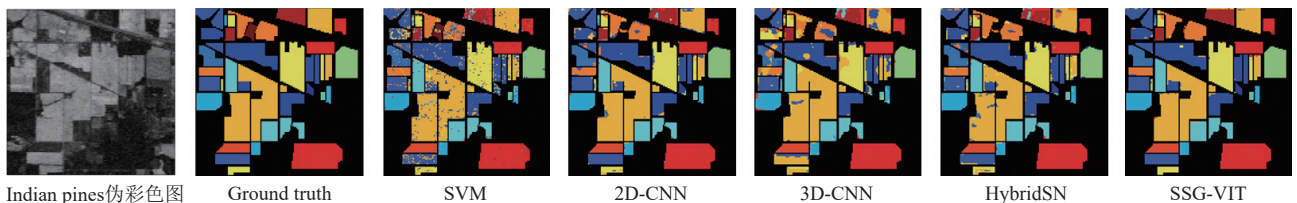


图 6 不同方法下 Indian pines 分类效果图

#### 4 结论与展望

针对高光谱数据波段多、数据冗余高, 以及同谱异物的问题, 本文提出了一种基于图卷积、分层深度

可分离卷积与 GSA 和相融合的分类网络 (SSG-VIT). 通过图卷积模块提取高光谱图像高维度的光谱特征, 以减少光谱波段的冗余信息降低计算的复杂度同时保

留光谱波段的重要特征,之后采用分层深度可分离卷积模块对高光谱数据的局部空间特征进行特征提取,然后采用 GSA 对其全局空间特征进行特征融合,最后在特征融合过程中采用自适应特征融合方式.本模型在减少参数提高鲁棒性的同时,保持了较高的分类精度.与其他模型对比在采用较少的训练样本下,依旧能够保持较高的分类精度,但因模型引入了图卷积模块其训练时间相比他模型略有增加.本文所提出的 SSG-VIT 模型在 Indian pines、Salinas、Botswana 数据集上

进行了多次实验,实验结果中总体分类精度分别达到了 99.32%、99.67%、99.69%,证明了本文提出方法的有效性.将本文提出的 SSG-VIT 模型与 SVM、2D-CNN、3D-CNN、HybridSN 模型进行实验对比分析,根据分类效果图显示模型分类的边界结果十分明确,并且模型在不同数据集的混淆矩阵图及不同方法下的损失曲线图中显示,SSG-VIT 模型的收敛速度是最快的且具有优异的稳定性.未来可对模型的复杂度做进一步的深入研究,进一步提升模型的训练时间和效率.

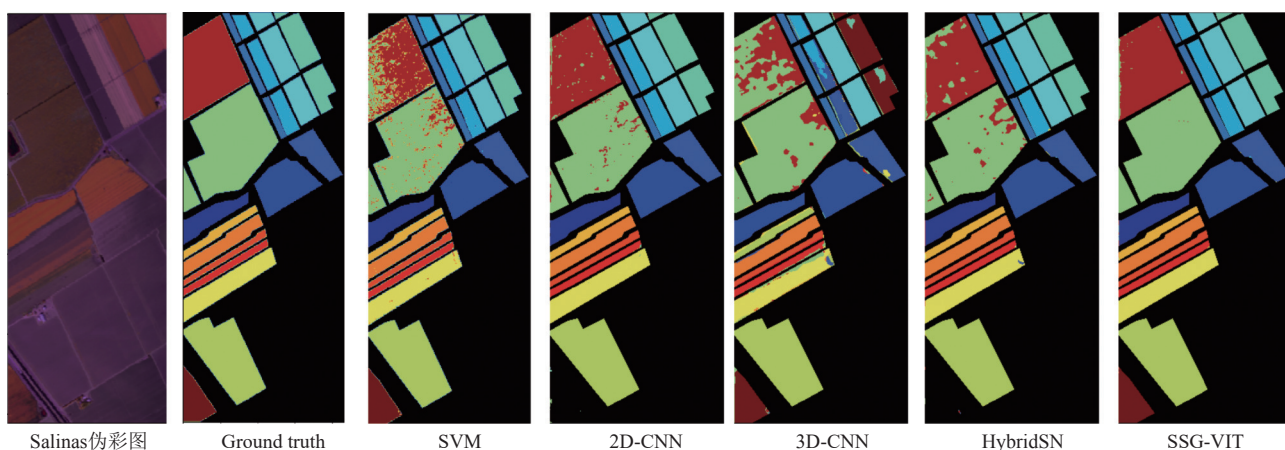


图 7 不同方法下 Salinas 分类效果图

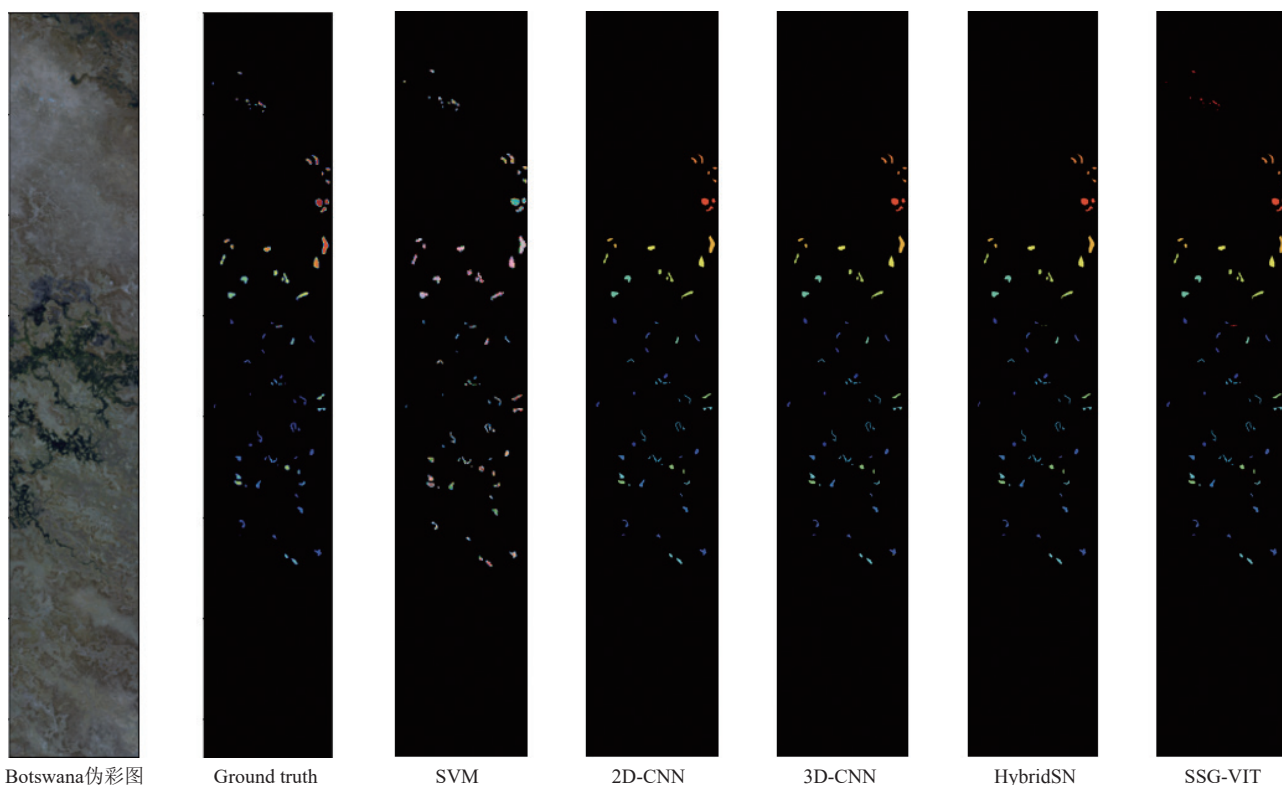


图 8 不同方法下 Botswana 分类效果图

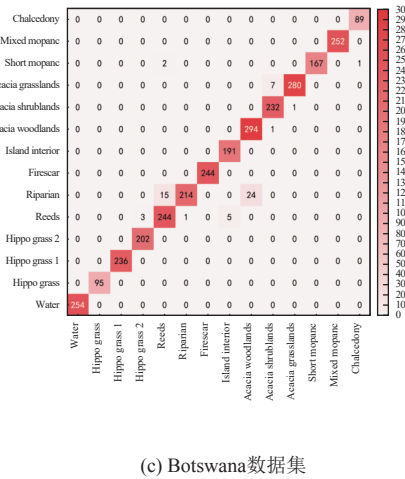
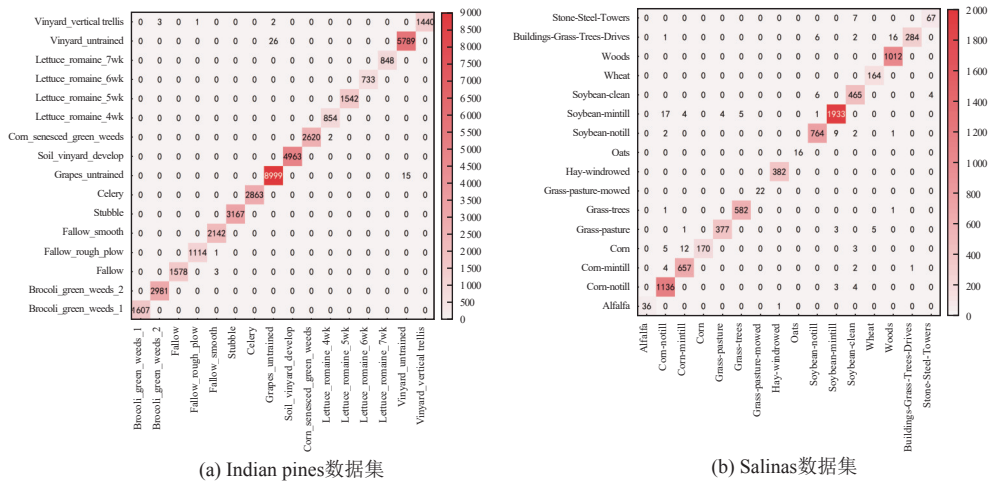


图9 不同数据集下的混淆矩阵图

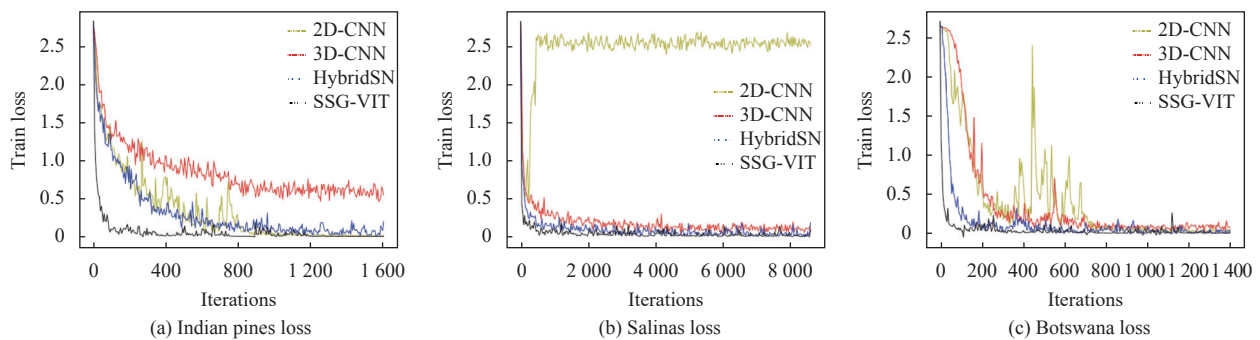


图10 不同数据集的损失曲线图

参考文献

- 蒋凯, 任宇鹏, 徐丹丹. 数字技术赋能生态环境保护——高光光谱走航监测系统助力水环境污染溯源应用. 中国安防, 2024(4): 48–51. [doi: 10.3969/j.issn.1673-7873.2024.04.011]
- 袁静文, 武辰, 杜博, 等. 高分五号高光光谱遥感影像的城市土地利用景观格局分析. 遥感学报, 2020, 24(4): 465–478. [doi: 10.11834/jrs.20209252]
- 甘甫平, 王润生. 高光光谱遥感技术在地质领域中的应用. 国土资源遥感, 2007, 74(4): 127–128.
- 史华湘. 基于机器学习的红外高光光谱云检测及其同化应用 [硕士学位论文]. 长沙: 国防科技大学, 2021.
- Mucherino A, Papajorgji PJ, Pardalos PM. K-nearest neighbor classification. In: Mucherino A, Papajorgji PJ, Pardalos PM, eds. Data Mining in Agriculture. New York:

- Springer, 2009. 83–106. [doi: [10.1007/978-0-387-88615-2\\_4](https://doi.org/10.1007/978-0-387-88615-2_4)]
- 6 Zhang Z, Xu Y, Yang J, *et al.* A survey of sparse representation: Algorithms and applications. *IEEE Access*, 2015, 3: 490–530. [doi: [10.1109/ACCESS.2015.2430359](https://doi.org/10.1109/ACCESS.2015.2430359)]
- 7 Benediktsson JA, Palmason JA, Sveinsson JR. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(3): 480–491. [doi: [10.1109/TGRS.2004.842478](https://doi.org/10.1109/TGRS.2004.842478)]
- 8 Huang SJ, Cai NG, Pacheco PP, *et al.* Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics & Proteomics*, 2018, 15(1): 41–51. [doi: [10.21873/cgp.20063](https://doi.org/10.21873/cgp.20063)]
- 9 Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5–32. [doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)]
- 10 Wang CZ, Wu P, Yan LY, *et al.* Image classification based on principal component analysis optimized generative adversarial networks. *Multimedia Tools and Applications*, 2021, 80(6): 9687–9701. [doi: [10.1007/s11042-020-10137-8](https://doi.org/10.1007/s11042-020-10137-8)]
- 11 Chhapariya K, Buddhiraju KM, Kumar A. Spectral-spatial classification of hyperspectral images with multi-level CNN. *Proceedings of the 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. Rome: IEEE, 2022. 1–5. [doi: [10.1109/WHISPERS56178.2022.9955063](https://doi.org/10.1109/WHISPERS56178.2022.9955063)]
- 12 Hu W, Huang YY, Wei L, *et al.* Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015, 2015: 258619. [doi: [10.1155/2015/258619](https://doi.org/10.1155/2015/258619)]
- 13 Roy SK, Krishna G, Dubey SR, *et al.* HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(2): 277–281. [doi: [10.1109/LGRS.2019.2918719](https://doi.org/10.1109/LGRS.2019.2918719)]
- 14 Guo AJX, Zhu F. A CNN-based spatial feature fusion algorithm for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 7170–7181. [doi: [10.1109/TGRS.2019.2911993](https://doi.org/10.1109/TGRS.2019.2911993)]
- 15 Zhou F, Hang RL, Liu QS, *et al.* Hyperspectral image classification using spectral-spatial LSTMs. *Neurocomputing*, 2019, 328: 39–47. [doi: [10.1016/j.neucom.2018.02.105](https://doi.org/10.1016/j.neucom.2018.02.105)]
- 16 Zhong ZL, Li J, Luo ZM, *et al.* Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 847–858. [doi: [10.1109/TGRS.2017.2755542](https://doi.org/10.1109/TGRS.2017.2755542)]
- 17 Hong DF, Han Z, Yao J, *et al.* SpectralFormer: Rethinking hyperspectral image classification with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5518615. [doi: [10.1109/TGRS.2021.3130716](https://doi.org/10.1109/TGRS.2021.3130716)]
- 18 Zhao ZY, Xu X, Li ST, *et al.* Hyperspectral image classification using groupwise separable convolutional vision Transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5511817. [doi: [10.1109/TGRS.2024.3377610](https://doi.org/10.1109/TGRS.2024.3377610)]
- 19 He J, Zhao LN, Yang HW, *et al.* HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(1): 165–178. [doi: [10.1109/TGRS.2019.2934760](https://doi.org/10.1109/TGRS.2019.2934760)]
- 20 Roy SK, Deria A, Shah C, *et al.* Spectral-spatial morphological attention Transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5503615. [doi: [10.1109/TGRS.2023.3242346](https://doi.org/10.1109/TGRS.2023.3242346)]
- 21 晏远翔, 曹国, 张友强. 基于 Wasserstein 距离与生成对抗网络的高光谱图像分类. *计算机系统应用*, 2024, 33(2): 13–22. [doi: [10.15888/j.cnki.csa.009388](https://doi.org/10.15888/j.cnki.csa.009388)]
- 22 Subudhi S, Patro RN, Biswal PK, *et al.* A survey on superpixel segmentation as a preprocessing step in hyperspectral image analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 5015–5035. [doi: [10.1109/JSTARS.2021.3076005](https://doi.org/10.1109/JSTARS.2021.3076005)]
- 23 Xia PP, Zhang L, Li FZ. Learning similarity with cosine similarity ensemble. *Information Sciences*, 2015, 307: 39–52. [doi: [10.1016/j.ins.2015.02.024](https://doi.org/10.1016/j.ins.2015.02.024)]
- 24 Wang D, Du B, Zhang LP. Spectral-spatial global graph reasoning for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(9): 12924–12937. [doi: [10.1109/TNNLS.2023.3265560](https://doi.org/10.1109/TNNLS.2023.3265560)]

(校对责编: 张重毅)