

# Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification

Yanni Dong, Quanwei Liu, Bo Du and Liangpei Zhang

## 摘要

卷积神经网络 (CNN) 和图神经网络 (GNN), 例如图注意力网络 (GAT), 是两种经典的神经网络模型, 分别应用于网格数据和图数据的处理。他们在高光谱图像 (HSI) 分类领域取得了出色的表现, 引起了人们的极大兴趣。然而 CNN 一直面临着小样本的问题, 而 GNN 则要付出巨大的计算成本, 这都制约了两个模型的性能。在本文中, 我们利用基于超像素的 GAT 和基于像素的 CNN 的特点, 提出了用于 HSI 分类的卷积神经网络和图注意力网络 (WFCG) 的加权特征融合, 这被证明是互补的。我们首先在基于超像素的编码器和解码器模块的帮助下建立 GAT。然后我们结合注意力机制来构建 CNN。最后, 将特征与两个神经网络模型的特征进行加权融合。对三个真实 HSI 数据集的严格实验表明, WFCG 可以充分探索 HSI 的高维特征, 并与其他最先进的方法相比获得有竞争力的结果。

**关键词:** 高光谱图像分类; 卷积神经网络; 图注意力网络; 加权特征融合; 注意力机制

## 1 引言

高光谱传感器可以同时捕获被观察物质的光谱信息和空间信息, 形成高光谱图像 (HSI) [1]。通常, HSI 由数百个连续光谱带组成, 可以将它们组合成数据立方体以提供非常详细的光谱信息。此外, 随着高光谱技术和传感器平台的发展, 高光谱传感器还可以获得更详细的空间信息 [2]。丰富的光谱和空间信息使得 HSI 具有广泛的用途, 如环境监测矿产调查、农业评价、军事侦察等 [3–7]。HSI 分类因其重要性而引起了相当大的关注, 它是指确定 HSI 中每个像素所代表的地物类别 [8]。早期, 基于机器学习的经典方法大多是 k 近邻 [9]、逻辑回归 [10] 和随机森林 [11] 等监督学习方法, 在理想情况下可以取得令人满意的结果。然而, 传统的分类方法依赖专家手动设计特征, 这些特征通常很浅, 导致准确性进一步受到限制 [12]。相比之下, 使用神经网络的深度学习 (DL) [13] 可以通过端到端学习自动提取层次和非线性特征。它在无人驾驶汽车、目标识别、机器翻译等许多领域取得了巨大成功 [14–17]。

## 2 相关工作

### 2.1 卷积神经网络应用于 HSI 分类

在 HSI 分类中, 最广泛使用的深度神经网络包括深度信念神经网络 (DBN) [18]、自动编码器 (AE) [19]、循环神经网络 (RNN) [20] 和卷积神经网络 (CNN) [21], [22]。其中, CNN 具有局部连接和权重共享的特点, 可以显著减少参数量。此外, 通过划分 Patch, 还可以同时捕获光谱信息和空间信息 [23]。基于 CNN 的方法在 HSI 分类领域发挥着越来越重要的作用。第一个谱空间 CNN 方法在 [24] 中提出, 在 HSI 分类领域比传统机器学习方法取得了更高的精度。端到端频谱空间残差网络 (SSRN) [25] 可以将原始 3D 立方体作为输入数据, 从而避免 HSI 上的复杂特征工程。一种混合光谱 CNN (HybridSN) 被 [26] 提出, 它是光谱空间 3D-CNN, 然后是空间 2D-CNN, 以进一步学习更抽象级别的空间表示。此外, 越来越多的深度学习方法被应用于 HSI 分类, 例如残差网络和密集连接的 CNN [27]、[28]。基于 CNN 的模型经过不断探索, 在 HSI 分类领域取得了令人瞩目的成就, 但仍存在以下两大具有挑战性的问题需要解决: 1) 众所周知, 获取标记样本非常耗费资源 [29], [30]。对于基于卷积神经网络 (CNN) 的监督方法, 样本数量不足的问题是突出的, 容易出现过拟合现象。2) CNN 被设计用于处理欧几里得数据, 用于处理具有规则空间结构的数据, 因此它不能轻松捕捉高光谱影像中不同地物覆盖之间的内部关联 [31]。

### 2.2 图神经网络应用于 HSI 分类

图神经网络 (GNN) [32] 因其广泛的适用性和强大的性能而迅速引起人们的关注。GNN 能够进行非欧几里得数据的深度学习, 即它能够同时进行节点特征信息和结构信息的端到端表示学习。HSI 数据可以通过基于超像素的方法转换为图数据, 然后可以使用 GNN 方法对光谱空间上下文信息进行有效建模。这样就隐式扩大了标签数量, 一定程度上缓解了小样本的问题。在 [33] 中, 基于超像素分割的图卷积网络 (GCN) 首次应用于 HSI。它使用多阶邻居节点构造邻接矩阵, 使 GCN 能够捕获多尺度的空间信息。然后, 提出了一种在训练过程中自动学习图结构的方法 [34], 该方法可以促进节点特征学习并使图更适应 HSI 内容。在 GCN 中, 节点之间的权重是固定的, 无法改变, 这限制了网络的表达能力。为了动态改变节点之间的权重, 提出了一种名为图注意力网络 (GAT) [35] 的新 GNN 模型, 该模型能够为邻域中的不同节点指定不同的权重, 而不依赖于预先了解图结构或需要任何一种计算成本高昂的矩阵操作。GAT 利用 k 近邻选择邻居节点构建邻接矩阵, 可以计算每个不同节点的权重, 用于 HSI 分类 [36]。为了减轻巨大的计算成本, 类似于基于 CNN 的 HSI 分类, 使用划分子图的策略来构造邻接矩阵 [37]。金字塔结构通常可用于特征提取 [38]。光谱金字塔 GAT [39] 使用 3D-CNN 提取多尺度光谱信息, 然后应用图注意力机制显式提取高阶空间特征, 显著提高了分类精度。

### 3 本文方法

#### 3.1 本文方法概述

本文基于 CNN 和 GAT 这两个网络，设计了两个简单的分类器来探索它们的特征，然后提出 CNN 和 GAT 网络的加权特征融合 (WFCG) 来结合它们的特征。从技术上来说，首先，我们使用两个  $1 \times 1$  卷积层来处理原始 HSI 以进行光谱特征压缩和去噪。然后，输出光谱特征分别传递到两个分支，即分支一和分支二。在分支一中，为了获得稳定的超像素级特征，我们使用多头 GAT 和归一化层来提取超像素级特征，并且它们之间的连接通过图编码器和解码器实现。在第二个分支中，处理后的像素级光谱特征被传递到 CNN 中，并与注意力模块结合以捕获远程深层特征。最后将两个分支得到的特征进行加权融合。使用 softmax 函数分类后可以得到分类标签  $Y$ 。图 1 中的流程图总结了所提出的 HSI 分类 WFCG 框架。

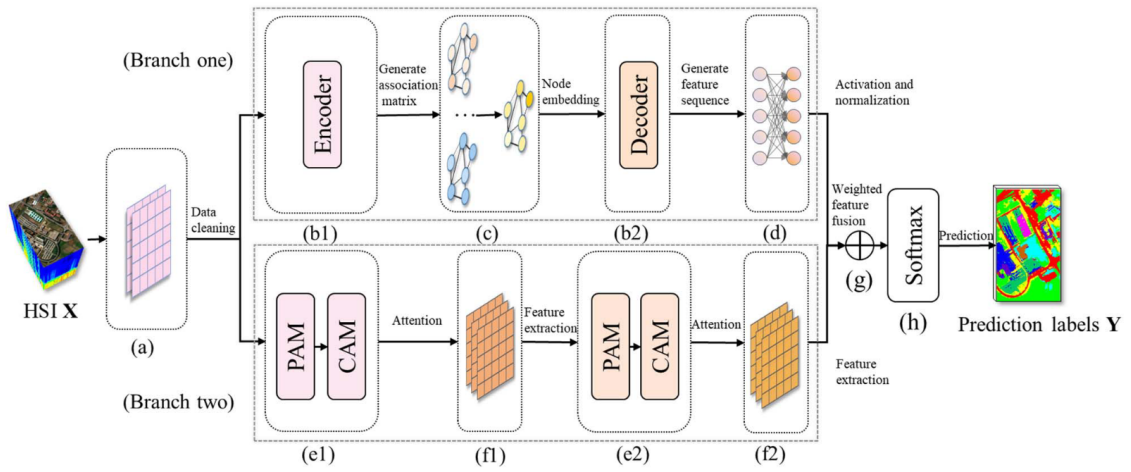


图 1. WFCG 模型框架。该模型主要分为八个部分。(a) 光谱卷积层。(b1) 和 (b2) 数据转换模块。(c) 图注意力模块。(d) 非线性特征变换模块。(e1) 和 (e2) 位置注意模块和通道注意模块。(f1) 和 (f2) 深度可分离卷积层。(g) 加权特征融合模块。(h) Softmax 层。

#### 3.2 特征提取模块

##### 3.2.1 卷积层

受到 Network-in-Network [40] 的启发，在网络的初始阶段，我们使用两个  $1 \times 1$  卷积层。由于原始 HSI 含有噪声和冗余信息，它们主要用作跨通道信息交换，去除无用的光谱信息，以提高辨别能力。然后，它们被用作降维模块以消除计算成本。这不仅可以增加网络的深度，还可以增加网络的宽度，而不会造成明显的性能损失。 $1 \times 1$  卷积只能提取谱域的特征，而 HSI 包含丰富的空间信息。2D-CNN 可用于提取上下文空间特征。然而，由于卷积核中的参数众多以及任务中的训练样本有限，2DCNN 可能很容易过拟合。因此，我们使用轻量级的深度可分离卷积 [41] 来构建 2D-CNN，可以大大减少参数并增强鲁棒性。

### 3.2.2 特征转换和超像素级特征提取

GNN 只接受图数据作为输入数据，但模块 (a) 生成的特征图排列在标准矩形网格中。虽然我们可以将图像像素视为图中的节点，但 GNN 需要输入邻接矩阵，这将使计算成本难以承受。为了解决这个问题，我们采用简单线性迭代聚类 (SLIC) [42] 将像素分组为感知上有意义的超像素。SLIC 分割采用 k-means 聚类方法生成超像素作为图节点。图注意力使神经网络能够通过选择性地关注不同的节点来学习有用的图表示。由于超像素中包含的像素数量不同，我们无法将上述分割方法直接集成到 WFCG 框架中。受 [34] 的启发，我们应用数据转换来允许特征在像素和超像素之间传播，如图 2 所示。

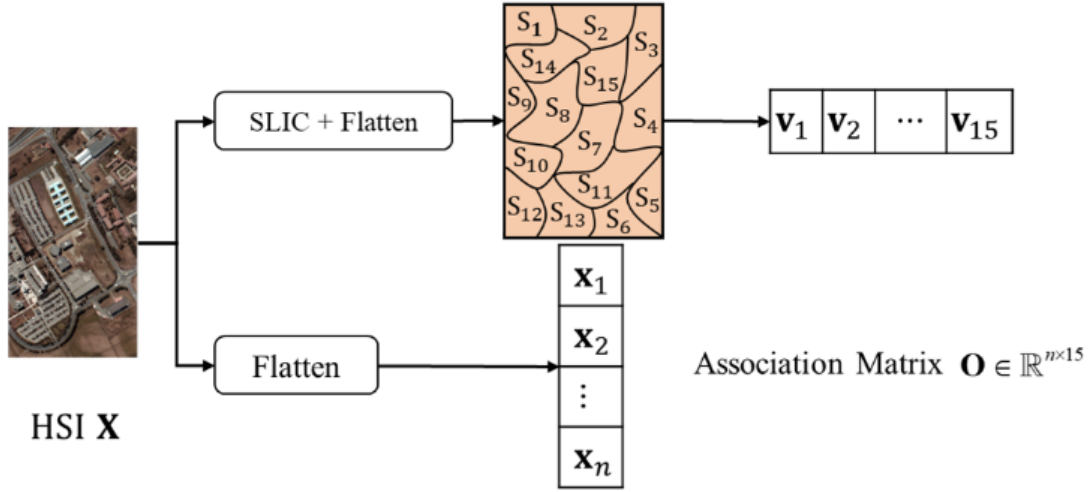


图 2. 像素和超像素数据转换模块的流程图，其中  $X_i$  表示扁平化 HSI 中的第  $i$  个像素， $V_j$  表示超像素  $S_j$  中包含的像素的平均亮度。

具体来说，令  $O \in \mathbb{R}^{HW \times Z}$  为像素和超像素之间的关联矩阵，其中  $Z$  表示超像素的数量，则我们有

$$\mathbf{O}_{i,j} = \begin{cases} 1, & \text{if } \bar{\mathbf{X}}_i \in S_j \\ 0, & \text{if } \bar{\mathbf{X}}_i \notin S_j \end{cases} \quad \bar{\mathbf{X}} = \text{Flatten}(\mathbf{X}), \quad (1)$$

其中  $\text{Flatten}(\cdot)$  表示 HSI 数据扁平化通过空间维度。 $O_{i,j}$  表示位置  $(i, j)$  处  $O$  的值。 $\bar{\mathbf{X}}_i$  表示  $\bar{\mathbf{X}}$  中的第  $i$  个像素。那么，我们可以使用以下公式来实现特征转换：

$$\mathbf{V} = \text{Encoder}(\mathbf{X}; \mathbf{O}) = \mathbf{O}^T \text{Flatten}(\mathbf{X}), \quad (2)$$

$$\tilde{\mathbf{X}} = \text{Decoder}(\mathbf{V}; \mathbf{O}) = \text{Reshape}(\mathbf{OV}), \quad (3)$$

其中  $\bar{O}$  表示按列归一化的  $O$ 。 $V$  表示由超像素组成的节点， $\text{Reshape}(\cdot)$  表示恢复扁平化数据的空间维度。 $\tilde{\mathbf{X}}$  表示转换回网格的特征。经过 SLIC 分割后，特征可以看作是一个无向图  $G = (V, E)$ ，其中  $V$  和  $E$  分别表示节点和边。这里节点的特征是超像素中像素特征的平均值。将像素的潜在表示编码为超像素后，我们应用注意力模块来获得节点嵌入，通过聚合邻居信息来表示当前节点，因此它可以学习空间特征。



### 3.2.3 位置注意力模块和通道注意力模块

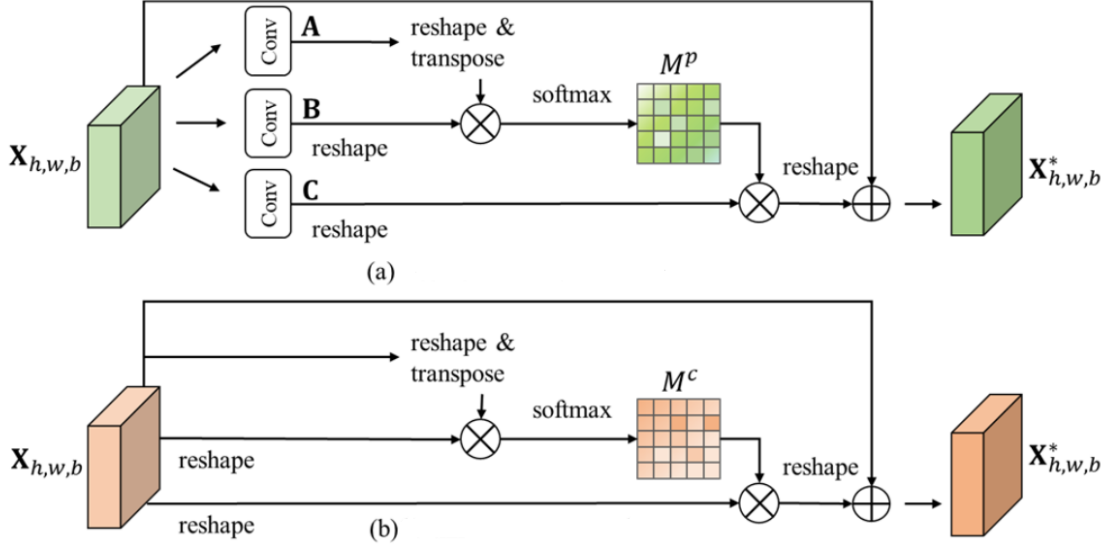


图 3. (a) 位置注意模块和 (b) 通道注意模块的详细信息。Conv 表示卷积层。\$M^P\$ 和 \$M^C\$ 表示注意力权重矩阵 \$\otimes\$ 是矩阵乘法, \$\oplus\$ 是逐元素加法。

位置注意力模块如图 3(a) 所示, 给定一个局部特征 \$X\_{h,w,b}\$, 我们首先使用卷积实现非线性变换, 生成两个新的特征图, 即 \$A\$ 和 \$B\$, 其中 \$\{A, B\} \in R^{H \times W \times C}\$。它们的维度被转换成 \$R^{N \times C}\$, 其中 \$N\$ 是输入局部特征像素的总数。然后, 我们在 \$A\$ 和 \$B\$ 的转置之间执行矩阵乘法, 应用 \$softmax\$ 层来计算位置注意力图 \$M^P \in R^{N \times N}\$ 为

$$m_{ji}^p = \frac{\exp(A_i \cdot B_j)}{\sum_{i=1}^N \exp(A_i \cdot B_j)}, \quad (4)$$

其中 \$m\_{ji}^p\$ 衡量了第 \$i\$ 个位置对第 \$j\$ 个位置的影响。位置注意力图是一个矩阵, 其形状与输入特征图的数量相同, 值在 \$(0, 1)\$ 范围内。这个值的大小代表了这个位置的重要性。以同样的方式, 我们得到特征图 \$C\$ 并将其与空间注意力图进行矩阵乘法, 将结果重塑为 \$R^{H \times W \times C}\$。最后, 为了更好地提取注意力信息而不丢失原始信息, 我们将其乘以可学习的尺度参数 \$\alpha\$ [43], 并与输入特征 \$X\_{h,w,b}\$ 进行逐元素求和运算以获得最终输出 \$X\_{h,w,b}^\*\$。

与位置注意相反, 通道注意细化了特征图的权重。通道注意力模块的结构如图 3(b) 所示。我们直接根据原始特征 \$X\_{h,w,b}\$ 计算通道注意力图 \$M^C \in R^{C \times C}\$, 如下所示

$$m_{ji}^c = \frac{\exp(X_{h,w,b,i} \cdot X_{h,w,b,i})}{\sum_{i=1}^N \exp(X_{h,w,b,i} \cdot X_{h,w,b,i})}. \quad (5)$$

此外, 与位置注意模块类似, 我们在空间注意力图和输入特征 \$X^\*\$ 之间执行矩阵乘法, 将结果重塑为 \$R^{H \times W \times C}\$。最后, 我们将其乘以可学习的尺度参数 \$\beta\$, 并与输入特征 \$X^\*\$ 执行逐元素求和运算, 以获得最终输出 \$Z \in R^{H \times W \times C}\$。

## 4 复现细节

### 4.1 与已有开源代码对比

高光谱图像是通过分析目标在不同波段上的光谱反射或辐射数据而获得的，通常具有数百个波段。这种多光谱信息为地物的识别提供了更丰富的光谱特征，使其在遥感图像分类中得到广泛应用。然而，高光谱数据的维度较高，这可能导致维度灾难问题和计算负担，因此在源代码的基础上采用主成分分析 (*PCA*) 降维技术来提取最重要的信息。*PCA* 是一种常用的降维技术，其主要目标是通过线性变换将原始数据映射到一个新的坐标系，使得数据在新坐标系中的方差最大。在高光谱图像分类中，*PCA* 可以用来减少冗余信息，提取最显著的特征，同时保留尽可能多的信息。通过 *PCA*，可以将高维的光谱数据转换为较低维度的特征，有助于提高分类算法的效率、减小计算负担，并防止维度灾难。

同时，在高光谱图像分类领域，通道注意力模块已被广泛用于提高模型对关键光谱信息的关注。然而，我们注意到原论文中采用的通道注意力模块并没有引入可学习参数，这限制了其在不同数据集和任务上的适应性。为了克服这一限制，我们决定对通道注意力模块进行改进，将其转变为带有可学习参数的版本，以增强其表达能力和灵活性。

在本次研究中，我们致力于复现先前的工作，并在此过程中，我们参考了本论文的开源代码作为我们复现的基础，链接如下<https://github.com/raglandman/WFCG>。

### 4.2 实验环境搭建

操作系统	Windows 10 家庭中文版
CPU	Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz
机带 RAM	16GB
内存	48GB
存储器容量	1TB

CPU	12 x Xeon Gold 6271
GPU	NVIDIA Tesla P100-16GB
内存	16GB
内存	48GB
存储	1.7TB

### 4.3 创新点

通过引入主成分分析 (*PCA*) 降维技术，对高维光谱数据进行处理。这旨在减少维度灾难问题和计算负担，同时提取最重要的信息，以便在分类算法中更有效地应用。此外，对原有通道注意力模块进行改进，引入可学习参数，以增强其适应性和灵活性。这一改进有助于模型在不同数据集和任务上更好地适应并提取关键信息。

## 5 实验结果分析

我们引入了主成分分析 (PCA) 降维技术, 有效地降低了数据的维度, 使得模型更关注数据的本质特性。另外, 我们对原有通道注意力模块进行了改进, 引入了可学习参数, 以增强其适应性和灵活性。这一优化使得模型能更好地捕捉输入数据中的关键信息, 呈现更强的鲁棒性。实验结果表明, 引入 PCA 降维技术和对通道注意力模块改进后, 模型性能略微提升, 减少了过拟合并提高了泛化能力, 使其在不同数据集和任务上更为稳健, 如图 4、图 5 所示。

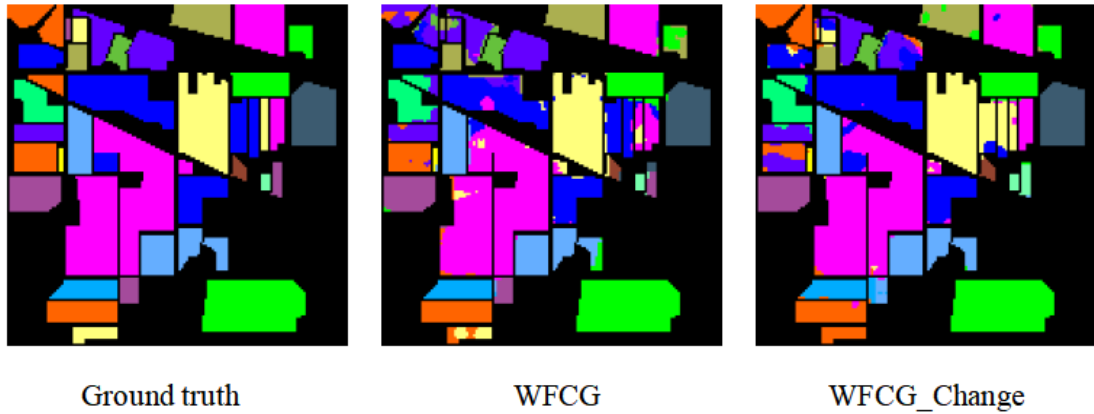


图 4. *Indian Pines* 实验结果

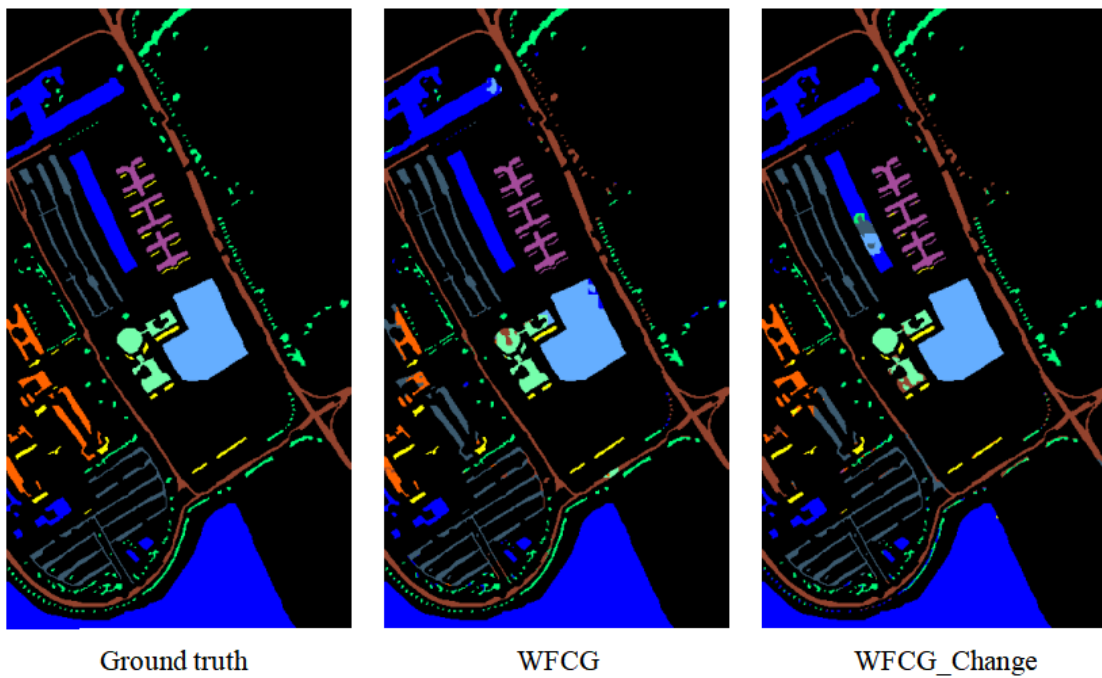


图 5. *Pavia University* 实验结果

在实验中, 我们还观察到引入 PCA 降维技术后, 模型的运行时间显著缩短。通过降低特征数量, 我们成功地减轻了计算负担, 使得模型在训练和推断阶段更为高效。这一时间优化进一步强化了方法的实用性, 使其更适用于对计算效率有更高要求的实际应用场景。

## 6 总结与展望

在本文中，我们全面探讨了基于 *WFCG* 的模型框架，并在此基础上进行了有益的扩展。主要的创新点包括引入主成分分析 *PCA* 降维技术和改进通道注意力模块，以提升模型在 *HSI* 分类任务中的性能。

然而，在当前研究中仍存在一些不足之处。例如，我们可能需要进一步调整和优化 *PCA* 降维技术的参数，以达到更好的性能。此外，通道注意力模块的改进可能还有进一步的空间，可以考虑不同的参数设置或结构调整。

未来的研究方向包括但不限于以下几点：首先，可以探索其他降维技术，与 *PCA* 进行比较，以找到更适合特定任务的方法。其次，可以考虑在模型中引入更多先进的注意力机制或者深度学习技术，以进一步提升模型的性能。此外，跨领域的迁移学习和模型的可解释性也是未来研究的潜在方向。

## 参考文献

- [1] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.*, 158:279–317, Dec 2019.
- [2] S. Liu, Q. Shi, and L. Zhang. Few-shot hyperspectral image classification with unknown classes using multitask deep learning. *IEEE Trans. Geosci. Remote Sens.*, 59(6):5085–5102, Jun 2021.
- [3] S. Sabbah. Remote sensing of gases by hyperspectral imaging: System performance and measurements. *Opt. Eng.*, 51(11), Jul 2012. Art. no. 111717.
- [4] D. Krupnik and S. Khan. Close-range, ground-based hyperspectral imaging for mining applications at various scales: Review and case studies. *Earth-Sci. Rev.*, 198, Nov 2019. Art. no. 102952.
- [5] J. Fan, N. Zhou, J. Peng, and Y. Gao. Hierarchical learning of tree classifiers for large-scale plant species identification. *IEEE Trans. Image Process.*, 24(11):4172–4184, Nov 2015.
- [6] M. Shimoni, R. Haelterman, and C. Perneel. Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geosci. Remote Sens. Mag.*, 7(2):101–117, Jun 2019.
- [7] J. Xie, L. Fang, B. Zhang, J. Chanussot, and S. Li. Super resolution guided deep network for land cover classification from remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, Oct 2021. Early access.
- [8] Y. Dong, T. Liang, Y. Zhang, and B. Du. Spectral-spatial weighted kernel manifold embedded distribution alignment for remote sensing image classification. *IEEE Trans. Cybern.*, 51(6):3185–3197, Jun 2021.



- [9] M.-L. Zhang and Z.-H. Zhou. MI-knn: A lazy learning approach to multi-label learning. *Pattern Recognit.*, 40(7):2038–2048, Jul 2007.
- [10] W. Xue, L. Zhang, X. Mou, and A. C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.*, 23(2):684–695, Feb 2014.
- [11] M. Belgiu and L. Dr̃agu t. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, 114:24–31, Apr 2016.
- [12] Z. Wang, B. Du, and Y. Guo. Domain adaptation with neural embedding matching. *IEEE Trans. Neural Netw. Learn. Syst.*, 31(7):2387–2397, Jul 2019.
- [13] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [14] Y. Dong, W. Shi, B. Du, X. Hu, and L. Zhang. Asymmetric weighted logistic metric learning for hyperspectral target detection. *IEEE Trans. Cybern.*, May 2021. Early access.
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. pages 779–788, Jun 2016.
- [16] Y. Dong, B. Du, L. Zhang, and L. Zhang. Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning. *IEEE Trans. Geosci. Remote Sens.*, 55(5):2509–2524, May 2017.
- [17] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi. Self-supervised learning with adaptive distillation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 60:1–13, 2021.
- [18] A. Mughees and L. Tao. Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images. *Tsinghua Sci. Technol.*, 24(2):183–194, Apr 2019.
- [19] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 7(6):2094–2107, Jun 2014.
- [20] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.*, 45(11):2673–2681, Nov 1997.
- [21] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.*, 145:120–147, Nov 2018.
- [22] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo. Exploring hierarchical convolutional features for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 56(11):6712–6722, Nov 2018.

- [23] N. Wang, X. Gao, and J. Li. Random sampling for fast face sketch synthesis. *Pattern Recognit.*, 76:215–227, Apr 2018.
- [24] J. Yue, W. Zhao, S. Mao, and H. Liu. Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sens. Lett.*, 6(6):468–477, May 2015.
- [25] Z. Zhong, J. Li, Z. Luo, and M. Chapman. Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Trans. Geosci. Remote Sens.*, 56(2):847–858, Feb 2018.
- [26] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri. Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.*, 17(2):277–281, Feb 2020.
- [27] L. Mou, P. Ghamisi, and X. X. Zhu. Unsupervised spectral-spatial feature learning via deep residual conv-deconv network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 56(1):391–406, Jan 2018.
- [28] M. Paoletti, J. Haut, J. Plaza, and A. Plaza. Deepdense convolutional neural network for hyperspectral image classification. *Remote Sens.*, 10(9):1454, Sep 2018.
- [29] Y. Gu, Q. Wang, H. Wang, D. You, and Y. Zhang. Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 8(6):2739–2751, Jun 2015.
- [30] Q. S. U. Haq, L. Tao, F. Sun, and S. Yang. A fast and robust sparse approach for hyperspectral data classification using a few labeled samples. *IEEE Trans. Geosci. Remote Sens.*, 50(6):2287–2302, Jun 2012.
- [31] J. B. Estrach, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and deep locally connected networks on graphs. pages 1–14, Apr 2014.
- [32] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. 2016.
- [33] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang. Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 58(5):3162–3177, May 2019.
- [34] Q. Liu, L. Xiao, J. Yang, and Z. Wei. Cnn-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 59(10):8657–8671, Oct 2021.
- [35] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. page 12, 2018.

- [36] A. Sha, B. Wang, X. Wu, and L. Zhang. Semisupervised classification for hyperspectral images using graph attention networks. *IEEE Geosci. Remote Sens. Lett.*, 18(1):157–161, Jan 2021.
- [37] Z. Zhao, H. Wang, and X. Yu. Spectral–spatial graph attention network for semisupervised hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.*, 19:1–5, 2021.
- [38] D. Han, J. Kim, and J. Kim. Deep pyramidal residual networks. pages 5927–5935, Jul 2017.
- [39] T. Wang, G. Wang, K. E. Tan, and D. Tan. Spectral pyramid graph attention network for hyperspectral image classification. 2020.
- [40] M. Lin, Q. Chen, and S. Yan. Network in network. Mar 2014.
- [41] A. G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. Apr 2017.
- [42] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282, Nov 2012.
- [43] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena. Self-attention generative adversarial networks. pages 7354–7363, 2019.