MambaHSI: Spatial–Spectral Mamba for Hyperspectral Image Classification

摘要

Transformer 已被广泛探索用于高光谱图像 (HSI) 分类。然而，由于其二次计算复杂性，Transformer 在速度和内存使用方面提出了挑战。最近，Mamba 模型成为一种有前途的方法，它具有强大的远距离建模能力，同时保持线性计算复杂性。然而，由于需要集成的空间和光谱信息，表示 HSI 对于曼巴来说是一个挑战。为了弥补这些缺陷，本文提出了一种基于 Mamba 模型的 HSI 分类模型，命名为 MambaHSI，该模型可以同时建模整幅图像的长程相互作用，并以自适应的方式集成空间信息和光谱信息。具体来说，本文设计了一个空间 Mamba 块 (SpaMB) 来建模整幅图像在像素级别上的长程交互。然后，提出了一个光谱 Mamba 块 (SpeMB)，将光谱向量分成多个组，挖掘不同光谱组之间的关系，并提取光谱特征。最后，提出了一种空-谱融合模块 (SSFM) 来自适应地整合 HSI 的空间和光谱特征。这是第一个基于 Mamba 的图像级 HSI 分类模型。我本文对四个不同的 HSI 数据集进行了广泛的实验。结果证明了所提出的 HSI 分类模型的有效性和优越性。这揭示了 Mamba 作为 HSI 模型的下一代骨干的巨大潜力。

关键词：高光谱图像（HSI）分类；Mamba；状态空间模型（SSMs）；transformer

1 引言

随着地球观测技术的快速发展，获取高光谱图像（Hyperspectral Imaging，HSI）的能力显著提高。与传统的 RGB 图像系统不同，HSI 覆盖了从可见光、近红外、中红外到远红外的大量连续波段，能够提供丰富的光谱信息。这使得高光谱成像技术在诸如城市制图、资源探测、环境监测等多个领域得到了广泛应用。然而，高光谱图像分类作为这些应用的核心任务，其精度受到空间和光谱特征提取方法的限制。

目前的高光谱图像分类方法可以大致分为基于传统机器学习（如支持向量机和随机森林）和深度学习的方法（如卷积神经网络和 Transformer）。尽管深度学习方法已经显著提高了分类性能，但其在长距离依赖建模和计算复杂性方面仍存在挑战。传统的卷积神经网络由于感受野的局限性，难以捕捉全局信息，而 Transformer 尽管具有出色的全局建模能力，但其自注意力机制计算复杂度为二次方，限制了在像素级的应用。

通过将 Mamba 模型引入高光谱图像分类，该研究设计了 MambaHSI 模型，结合了空间–光谱特征提取模块和自适应融合模块，克服了现有方法在长距离依赖建模和复杂度上的缺陷。Mamba 模型的线性复杂度和强大的建模能力，为提高高光谱图像分类性能提供了理论依据。

MambaHSI 是首个基于状态空间模型（SSM）的高光谱图像分类框架，能够同时建模全图的长距离交互并集成空间与光谱信息。其成功展示了 SSM 在图像分类领域的潜力，为下一代高光谱分类模型提供了理论指导。通过设计创新性的空间和光谱模块，以及空间–光谱融合模块，该模型显著提升了高光谱图像分类精度和效率，在多个真实数据集上的实验结果证明了其优越性。模型的线性复杂度还使其在实际应用中更具可扩展性。

2 相关工作

2.1 高光谱图像分类

一般来说，现有的 HSI 分类方法由于其固有特性（例如 CNN 的局部性和二次复杂度），对长程依赖性进行建模的能力有限。与现有的 HSI 分类方法不同，MambaHSI 是第一个基于 Mamba 的 HSI 分类方法，将整个 HSI 图像作为模型的输入，可以在保持线性计算复杂性的同时对长程依赖性进行建模。一般来说，现有的 HSI 分类方法由于其固有特性（例如 CNN 的局部性和二次复杂度），对长程依赖性进行建模的能力有限。与现有的 HSI 分类方法不同，MambaHSI 是第一个基于 Mamba 的 HSI 分类方法，将整个 HSI 图像作为模型的输入，可以在保持线性计算复杂性的同时对长程依赖性进行建模。

2.2 状态空间模型

最近的研究进展激起了人们对 SSM 的兴趣。SSM 起源于经典的卡尔曼滤波器模型，现代 SSM 擅长捕获长程依赖关系，并且可以有效地并行计算 [1]。SSM 是 CNN 或 Transformer 的新颖替代品。

长序列的 SSM：结构化状态空间序列（S4）[2] 模型可以对长程依赖性进行建模。序列长度的线性复杂性这一有前景的特性吸引了进一步的探索。史密斯等人 [3] 通过将 MIMO SSM 和高效并行扫描引入 S4 层，设计了 S5 层。傅等人 [4] 填补了 SSM 和 Transformers 在语言建模方面的性能差距。梅塔等人 [5] 引入了更多的门控单元来提高 S4 的表达能力。最近，Mamba 在各种大规模真实数据上表现优于 Transformers，并在序列长度上保持线性缩放。

用于视觉应用的 SSM：Nguyen 等人 [6] 扩展了 1-D S4 以处理 2-D 图像和 3-D 视频。TranS4mer [7] 结合了 S4 和 selfattention 的优点，在电影场景检测方面实现了 SOTA 性能。王等人 [8] 向 S4 引入了选择性机制，显着提高了 S4 在长格式视频理解上的性能，同时内存占用更低。U-Mamba [9] 将 Mamba 与 U 形架构相结合，用于生物医学图像分割。

用于 HSI 分类的 SSM：最近，[10]、[11]、[12]、[13] 采用 Mamba 进行高光谱图像分类。SpectralMamba [10] 引入了 PSS 和 GSSM 模块来分别简化状态域中的顺序学习并校正频谱。黄等人 [12] 提出了一种用于 HSI 分类的光谱空间曼巴（3DSS-Mamba）。何等人 [11] 从 3D 角度探讨了 Mamba 在 3D 高光谱块中的应用。然而，上述方法利用高光谱补丁作为输入。这导致在推断整个图像时产生大量的冗余计算，阻碍了实际部署和应用。此外，补丁输入的固定大小限制了模型充分利用图像信息的能力，从而限制了其特征提取能力。

为此，本文首先设计一个空谱曼巴块来分别捕获空间和光谱信息。此外，还提出了一种空谱融合模块（SSFM）来自适应地融合空间和光谱信息。这使得所提出的 MambaHSI 能够完全集成空间和光谱信息，同时以线性复杂性方式对远程依赖性进行建模。

3 本文方法

3.1 本文方法概述

图 1显示了所提出的 MambaHSI 的主要框架。该框架包含三个主要组件：嵌入层、编码器主干和分割头。嵌入层将光谱向量投影到嵌入空间中。值得注意的是，与现有基于补丁的嵌入方法不同，嵌入层提取每个像素的嵌入。这使得本文的模型能够获得更细粒度的像素嵌入，更适合高光谱分类等密集预测任务。具体来说，细粒度像素嵌入 $E \in R^{H\times W\times D}$ 可以由 HSI $I \in R^{H\times W\times D}$ 获得，如下所示：

$$\begin{aligned} E &= Embedding(I) \\ &= SiLU(GN(Conv(I))) \end{aligned} \tag{1}$$

其中 Conv、GN 和 SiLU 分别表示内核大小为 1 × 1 的卷积层、GN 层和 SiLU 激活函数。H、W 和 C 分别表示输入 HSI 的高度、宽度和光谱通道数。D 是嵌入维度。I 和 E 表示输入 HSI 和提取的嵌入。

编码器主干用于提取用于分类的判别性空间光谱特征。具体来说，编码器主干主要包含三个组件：用于提取空间特征的空间 Mamba 块（SpaMB）、用于捕获光谱特征的光谱 Mamba 块以及用于集成空间和光谱特征的 SSFM。该过程可以定义如下：

$$H = Encoder(E) \tag{2}$$

其中 Encoder 和 H 表示编码器主干和提取的隐藏特征。

分割头采用 1×1 核大小的卷积层来获得最终的 logits l，即 l = SegHead(H)。

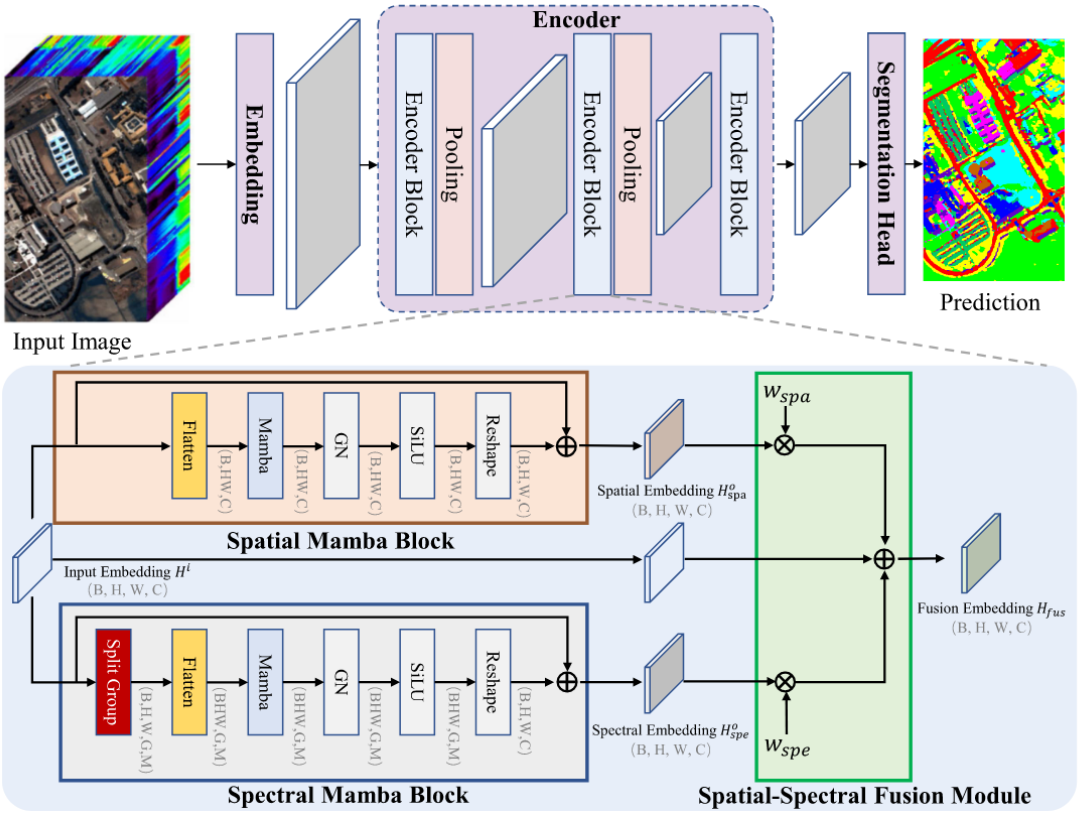

图 1. MambaHSI 框架

3.2 空间 Mamba 块

高光谱分类是像素级的分类任务。这意味着用于分类的表示需要满足两个条件。首先，表示应该被细化并且应该反映像素之间的差异。因此，与现有的补丁方法不同，本文以像素级的方式提取嵌入。其次，这些表示对于分类应该是有区别的。因此，所提出的模块具有强大的远程建模能力。值得注意的是，transformer 具有二次计算复杂度，并且不能用于在像素级建立长距离依赖关系。这迫使本文设计了一种新颖的空间特征提取器，以在线性计算复杂性中构建远程依赖关系。在本文中，采用 Mamba 层作为构建空间特征提取器的基本单元，它可以在线性计算复杂度下对远程依赖关系进行建模。

SpaMB 的详细结构如图 1所示。前向过程可以表述如下：

$$\begin{aligned} HF_{spa} &= Flatten(H^{i}), \\ HR_{spa} &= SiLU(GN(Mamba(HF_{spa}))), \\ H^{o}_{spa} &= Reshape(HR_{spa}) + H^{i} \end{aligned} \tag{3}$$

其中 $H^{i} \in R^{B\times H\times W\times D}$ 和 $H^{o}_{spa} \in R^{B\times H\times W\times D}$ 表示像素级的输入嵌入和 SpaMB 的输出特征。B、H、W 和 D 分别表示批量大小、图像高度、图像宽度和嵌入尺寸。实验中嵌入维数 D 设置为 128。$HF_{spa} \in R^{B\times L1\times D}$ 和 $HR_{spa} \in R^{B\times L1\times D}$ 表示展平输入和学习的残差空间特征。L1 等于 H×W 。Mamba 表示 [14] 中提出的标准 Mamba 块。GN 和残差连接的设计有助于 SpaMB 的学习。

3.3 光谱 Mamba 块

与通过 RGB 通道捕获图像的传统视觉系统不同，HSI 可以覆盖更大的光谱范围和更高的光谱分辨率。如何对光谱之间的关系进行建模并提取判别特征仍然是一个开放的研究问题。在本文中，设计了一个 SpeMB 来实现上述目标。整体结构如图 1所示。具体来说，论文中将光谱特征分为 G 组。然后，对不同光谱组之间的关系进行建模，然后根据光谱组之间挖掘的关系来更新光谱特征。提取的光谱特征 H^{o} 可以如下得到：

$$\begin{aligned} HG_{spe} &= SplitSpetralGroup(H^{i}), \\ HF_{spe} &= Flatten(H^{spe}), \\ HR_{spe} &= SiLU(GN(Mamba(HF_{spe}))), \\ H^{o}_{spe} &= Reshape(HR_{spe}) + H^{i} \end{aligned} \tag{4}$$

其中 $HG_{spe} \in R^{B\times H\times W\times G\times M}$，$HF_{spe} \in R^{N\times G\times M}$ ，$Hr_{spe} \in R^{N\times G\times M}$，$HF^{o}_{spe} \in R^{B\times H\times W\times D}$ 表示划分的光谱组特征，展平特征、残差特征和输出光谱特征。G 表示语义向量被分割成的组的数量。M 等于像素嵌入尺寸 D 除以组数 G。N 等于 B×H×W。Mamba 表示 [14] 中提出的标准曼巴块。

3.4 空谱 Mamba 块

空间和光谱特征对于 HSI 分类至关重要，整合空间和光谱信息有利于分类 [15]。这促使本文设计了 SSFM，SSFM 的架构如图 1所示。考虑到高光谱分类通常标记样本较少，论文中

引入了残差学习的思想来缓解训练过程中可能出现的过拟合。此外，SSFM 自适应地估计空间和光谱的重要性以指导融合。融合过程可以表述如下：

$$H_{fus} = H_i + \omega_{spa} \times H^o_{spa} + \omega_{spe} \times H^o_{spe} \tag{5}$$

其中 ω_{spa} 和 ω_{spe} 分别表示空间和光谱的融合权重。ω_{spa} 和 ω_{spe} 是随机初始化的，并且通过反向传播更新这些权重以确定最终的融合权重。

4 复现细节

4.1 与已有开源代码对比

在复现过程中，首先是跑了原论文中使用的四个数据集，分别是：Pavia University、Houston、HanChuan、HongHu，得出的精度均很低，最高才达到 40，经过不断的调试，尝试不同的环境等，最终得到的精度与原论文中所得结果差值为 0.01，在原论文使用的数据集的基础上，我还使用了 Salinas 和 Botswana 这两个数据集，得到的结果均比原论文中所得结果高，误分类也比原论文少许多，所得的分类图具有很好的对象完整性和很清晰的边界。

4.2 实验环境搭建

在复现该论文时需要配置好如图 2所示的所有环境。

```
conda create -n MambaHSI_env python=3.9
conda activate MambaHSI_env
conda install pytorch==1.13.1 torchvision==0.14.1 torchaudio==0.13.1 pytorch-cuda=11.7 -c pytorch -c nv
pip install packaging==24.0
pip install triton==2.2.0
pip install mamba-ssm==1.2.0
pip install spectral
pip install scikit-learn==1.4.1.post1
pip install calflops
```

图 2. 实验所需环境

4.3 创新点

(1) 本文提出了 MambaHSI，是第一个基于 Mamba 架构的图像级高光谱图像（HSI）分类模型。它模拟了整个图像的远程相互作用，同时集成了空间和光谱信息，克服了 cnn 和 transformer 在局部依赖关系和计算复杂性方面的局限性

(2) 本文设计了空间和光谱曼巴块来分别提取空间和光谱信息。受益于曼巴强大的远程建模能力，所提出的空间和光谱曼巴块可以对整个图像的远程交互进行建模。

(3) 本文提出了一种 SSFM，它可以自适应地估计空间和光谱信息的重要性以指导它们的融合。此外，还引入残差学习思想来帮助模块训练。

5 实验结果分析

如表1是论文中使用 Pavia University 数据集跑出来的定量的结果，表2是我使用 Salinas、Pavia University 以及 Botswana 这三个数据集跑出来的定量的结果，可以看出与我复

现出来的结果相差为 0.01，差值很小。而使用 Salinas 数据集跑出来的结果其 OA 值达到了 98.54±0.54,AA 值达到了 99.28±0.23,Kappa 值达到了 98.93±1.12, 使用 Botswana 数据集跑出来的结果其 OA 值达到了 99.64±0.27,AA 值达到了 99.69±0.24, Kappa 值达到了 99.85±0.31, 可见使用 Salinas 和 Botswana 这两个数据集得到的精度等值会比论文中提升许多。

表 1. 原论文结果

	OA	**AA**	**Kappa**
Pavia University	95.74 ± 0.90	95.86 ± 1.11	95.00 ± 2.24

表 2. 最终实验结果

	OA	**AA**	**Kappa**
Salinas	98.54 ± 0.54	99.28 ± 0.23	98.93 ± 1.12
Pavia University	95.73 ± 0.91	95.85 ± 1.11	95.00 ± 2.24
Botswana	99.64 ± 0.27	99.69 ± 0.24	99.85 ± 0.31

如图 3展示的是使用以上三个数据集跑出来的定性的结果，左边代表真值图，右边是跑出来的结果图。可以看出，与真值图相比，使用 Salinas 数据集跑出来的结果误分类的地方比较少，使用 Pavia University 数据集跑出来的结果误分类的地方相对会多一点，而使用 Botswana 数据集跑出来的结果几乎没有误分类的地方。

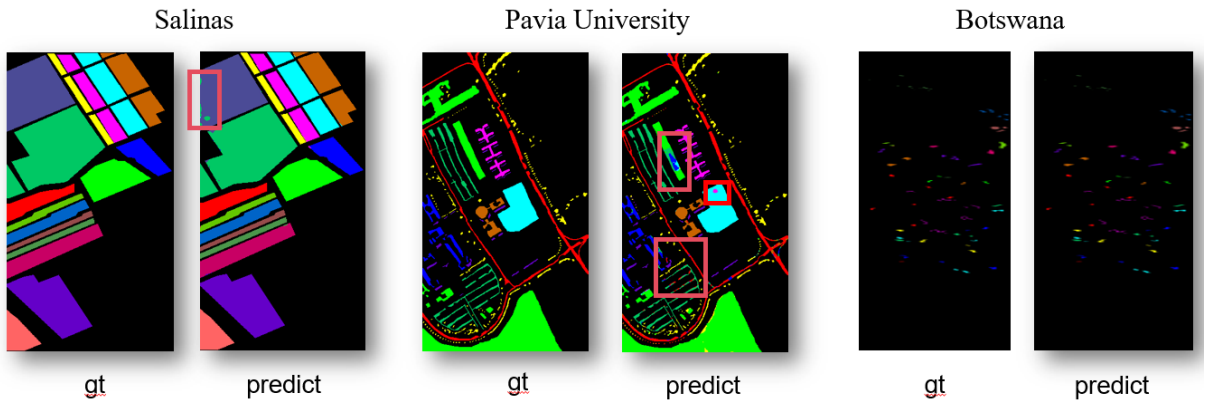

图 3. 定性结果

6 总结与展望

在本文中，提出了 MambaHSI，这是第一个基于 SSM 的图像级 HSI 分类方法，该方法具有强大的远程依赖性建模和集成光谱空间信息的能力。MambaHSI 引入 Mamba 作为基本单元来建模整个图像的长程交互，这使得模型能够捕获整个 HSI 图像的长程依赖性，同时保持线性计算复杂度。所提出的 SpaMB、SpeMB 和 SSFM 使模型能够挖掘有区别的空间和光谱特征，然后自适应地融合它们以进行 HSI 分类。从多个数据集上的广泛实验结果来看，本文主要得出以下结论：1）SSM 具有成为 HSI 分类的下一代骨干的巨大潜力，这得益于对远

程依赖性建模的强大能力，同时保持线性计算复杂性；2）集成空间和光谱信息对于 HSI 分类至关重要。未来，应将所提出的想法扩展到更多的 HSI 任务，例如弱监督 HSI 分类和 HSI 聚类。

参考文献

[1] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024.

[2] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

[3] Jimmy TH Smith, Andrew Warrington, and Scott W Linderman. Simplified state space layers for sequence modeling. *arXiv preprint arXiv:2208.04933*, 2022.

[4] Daniel Y Fu, Tri Dao, Khaled K Saab, Armin W Thomas, Atri Rudra, and Christopher Ré. Hungry hungry hippos: Towards language modeling with state space models. *arXiv preprint arXiv:2212.14052*, 2022.

[5] Harsh Mehta, Ankit Gupta, Ashok Cutkosky, and Behnam Neyshabur. Long range language modeling via gated state spaces. *arXiv preprint arXiv:2206.13947*, 2022.

[6] Eric Nguyen, Karan Goel, Albert Gu, Gordon Downs, Preey Shah, Tri Dao, Stephen Baccus, and Christopher Ré. S4nd: Modeling images and videos as multidimensional signals with state spaces. *Advances in neural information processing systems*, 35:2846–2861, 2022.

[7] Md Mohaiminul Islam, Mahmudul Hasan, Kishan Shamsundar Athrey, Tony Braskich, and Gedas Bertasius. Efficient movie scene detection using state-space transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18749–18758, 2023.

[8] Jue Wang, Wentao Zhu, Pichao Wang, Xiang Yu, Linda Liu, Mohamed Omar, and Raffay Hamid. Selective structured state-spaces for long-form video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6387–6397, 2023.

[9] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[10] Jing Yao, Danfeng Hong, Chenyu Li, and Jocelyn Chanussot. Spectralmamba: Efficient mamba for hyperspectral image classification. *arXiv preprint arXiv:2404.08489*, 2024.

[11] Yan He, Bing Tu, Bo Liu, Jun Li, and Antonio Plaza. 3dss-mamba: 3d-spectral-spatial mamba for hyperspectral image classification. *arXiv preprint arXiv:2405.12487*, 2024.

[12] Lingbo Huang, Yushi Chen, and Xin He. Spectral-spatial mamba for hyperspectral image classification. *arXiv preprint arXiv:2404.18401*, 2024.

[13] Jiamu Sheng, Jingyi Zhou, Jiong Wang, Peng Ye, and Jiayuan Fan. Dualmamba: A lightweight spectral-spatial mamba-convolution network for hyperspectral image classification. *arXiv preprint arXiv:2406.07050*, 2024.

[14] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

[15] Yonghao Xu, Bo Du, and Liangpei Zhang. Beyond the patchwise classification: Spectral-spatial fully convolutional networks for hyperspectral image classification. *IEEE Transactions on Big Data*, 6(3):492–506, 2019.