

文章编号 1004-924X(2023)18-2752-13

面向高光谱医学图像分类的空-谱自 注意力 Transformer

李 远¹, 时 旭¹, 杨正春², 谭崎娟^{3*}, 黄 鸿^{1*}

(1. 重庆大学 光电技术与系统教育部重点实验室, 重庆 400044;

2. 重庆市妇幼保健院 超声科, 重庆 401147;

3. 重庆大学附属肿瘤医院 影像科, 重庆 400030)

摘要: 高光谱成像技术的飞速发展给非侵入式医学成像带来新的契机, 但高光谱医学图像具有高维度、高冗余以及“图谱合一”的特点, 亟需针对上述特点设计智能诊断算法。近年来, Transformer 已经在高光谱医学图像处理领域得到广泛应用。然而, 不同仪器设备、不同采集操作所获得的高光谱医学图像差异较大, 这给现有 Transformer 诊断模型的实际应用带来了巨大挑战。针对上述问题, 本文提出了一种空-谱自注意力 Transformer (S³AT), 自适应挖掘像素与像素间、波段与波段间的内蕴联系, 并在分类阶段融合多个视野下的预测结果。首先, 在 Transformer 编码器中, 设计一种空-谱自注意力机制, 获取不同视野下高光谱图像上的关键空间信息和重要波段, 并将不同视野下所获得的空-谱自注意力进行融合。其次, 在模型分类阶段, 将不同视野下的预测结果根据可学习权重进行加权融合, 对图像进行综合预测。在 In-vivo Human Brain 和 BloodCell HSI 两个数据集上, 本文算法总体分类精度分别达到 82.25% 和 91.74%。实验结果表明, 所提出的算法有效改善高光谱医学图像分类性能。

关 键 词: 高光谱医学图像; Transformer; 空-谱自注意力; 预测融合

中图分类号: TP394.1; TH691.9 **文献标识码:** A **doi:** 10.37188/OPE.20233118.2752

Spatial-spectral Transformer for classification of medical hyperspectral images

LI Yuan¹, SHI Xu¹, YANG Zhengchun², TAN Qijuan^{3*}, HUANG Hong^{1*}

(1. Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry of China,

Chongqing University, Chongqing 400044, China;

2. Women and Children's Hospital of Chongqing Medical University, Chongqing 401147, China;

3. Department of Radiology, Chongqing University Cancer Hospital & Chongqing Cancer Institute & Chongqing Cancer Hospital, Chongqing 400030, China)

* Corresponding author, E-mail: hhuang@cqu.edu.cn; jiangliao2000@163.com

Abstract: The development of hyperspectral imaging (HSI) technology offers new avenues for non-invasive medical imaging. However, medical hyperspectral images are characterized by high dimensionality,

收稿日期: 2023-01-23; 修订日期: 2023-02-28.

基金项目: 国家自然科学基金资助项目 (No. 42071302); 重庆市留学人员回国创业创新支持计划资助项目 (No. cx2019144); 重庆市研究生研究与创新基金资助项目 (No. CYB21060)

high redundancy, and the property of “graph-spectral uniformity,” necessitating the design of high-precision diagnostic algorithms. In recent years, transformer modes have been widely applied in medical hyperspectral image processing. However, medical hyperspectral images obtained using various instruments and acquisition methods have significant differences; this considerably hinders the practical applications of existing transformer-based diagnostic models. To address the aforementioned issues, a spatial - spectral self-attention transformer (S3AT) algorithm is proposed to adaptively mine the intrinsic relations between pixels and bands. First, in the transformer encoder, a spatial - spectral self-attention mechanism, which is designed to obtain key spatial information and important bands on hyperspectral images from different view-points, is employed. Thus, the spectral - spectral self-attention obtained from different views is fused. Second, in the classification stage, the predictions from different views are fused according to the learned weights. The experimental result on in-vivo human brain and blood cell HSI datasets indicate that the overall classification accuracies reach 82.25% and 91.74%, respectively. This demonstrates that the proposed S3AT algorithm yields enhanced classification performance on medical hyperspectral images.

Key words: medical hyperspectral images; transformer; spatial-spectral self-attention; predictions fusion

1 引 言

高光谱成像(Hyperspectral Imaging, HSI)技术是一种先进的图像空间信息与光谱信息提取技术,能同时获取拍摄对象的二维空间信息和一维光谱信息,覆盖可见光、红外和紫外等光谱范围,其已成功应用于遥感监测、艺术保护以及食品安全等^[1]。在生物医学领域,高光谱成像作为一种非侵入性的辅助诊断手段,因其可提供有关组织生理、形态和生化成分的诊断信息,为生物组织学研究提供精细的光谱特征,正逐渐受到广泛关注^[2-5],并已成功应用于非侵入性疾病的诊断和监测^[6-7]、图像引导的微创手术^[8]和药物剂量评估^[9]等。近年来,随着精准医学理论的高速发展,如何针对高光谱医学图像高维度、高冗余度以及“图谱合一”的特点,设计高效与精确的诊断算法已成为高光谱医学图像分析领域的研究热点。

传统的高光谱医学图像分类方法通常在提取手工特征之后,使用分类器对其分类。Duan^[10]等提出一种旋转不变的局部二值化模式作为纹理特征,同时结合形状特征和光谱特征,采用支持向量机(Support Vector Machine, SVM)对白细胞进行分类。Ruiz^[11]等使用SVM和随机森林(Random Forest, RF)对活体大脑高光谱图像进行分类,验证了传统分类方法的潜力。Baltus-

sen^[12]等对腹腔镜获取的结肠癌高光谱图像进行特征提取后,采用SVM对三种组织类型进行区分。然而传统的高光谱图像分类方法无法提取深层特征,其性能受到很大限制。

近年来,深度学习作为一种端到端的方法,已开始应用于高光谱医学图像处理领域。其中,卷积神经网络(Convolutional Neural Network, CNN)成为主流,其使用局部感受野,并且随着网络层数的加深,逐渐提取深层特征,在诊断任务中表现优异。Huang等^[13]提出了一种将调制Gabor小波与深度卷积神经网络核相结合的血细胞分类框架(Modulated Gabor CNN, MGCNN),将调制Gabor滤波与卷积神经网络相结合,对血细胞进行分类。Wei^[14]等设计了一种双通道CNN提取局部特征与全局特征,取得了比传统卷积神经网络更好的分类结果。Zhang^[15]等提出了一种基于卷积组合单元的三维卷积神经网络(3D-PulCNN),对肺癌三种亚型进行分类。Hu^[16]等提出一种空-谱联合卷积神经网络,对胃部病理组织进行识别。然而,高光谱图像波段数量丰富,传统的卷积神经网络无法在长距离波段之间挖掘有效的关系信息,并扭曲其原有的光谱序列关系。这限制了卷积神经网络方法在高光谱医学图像上的性能。

Vision Transformer(ViT)以其强大的全局建模能力而受到广泛关注^[17]。ViT中的自注意

力机制,可以捕获长距离光谱波段间的关系,更好地对光谱序列建模,已在高光谱医学图像领域取得一定成效。Zhou 等^[18]提出一种 Swin-spectral Transformer 用来获得有效的光谱和空间特征表示。Li^[19]等提出一种光谱纹理 Transformer,用来感知光谱上下文信息。Li^[20]等提出一种多层协同生成对抗 Transformer,用于缓解高光谱医学标记样本数量的不足,加深而受限的问题。然而高光谱医学图像在获取过程中,由于采集设备、操作手段,以及预处理方式(光谱矫正、降噪以及解混等)的不同,其光谱分辨率、空间分辨率也往往不同,所拍摄生物组织的光谱曲线差异较大。因此,每个具体的诊断任务往往需要设计不同的算法。当上述算法应用于不同的诊断任务时,其性能难以满足更进一步的精度需求。

最近,研究人员开始结合 Transformer 模型和空-谱注意力机制,以提升高光谱图像分类的精度。空-谱注意力机制可以更好地捕捉关键的空间和光谱信息,并根据不同类型的高光谱图像的特点挖掘关键的空-谱信息。Peng^[21]等设计了一种双分支结构的交叉空-谱注意力,其中空间分支用来获取细粒度的空间信息,光谱分支用来建立光谱序列之间的关系。Ouyang^[22]等提出一种空-谱注意力机制,用于依次捕获空间信息和光谱信息,使模型更加关注差异化的空间和光谱位置。Liu 等^[23]提出一种双流深度空-谱注意力机制,分别用于关注空间维度和光谱维度的特征。然而,这些空-谱注意力机制只是简单地将输入特征分别处理成光谱序列或空间序列,再依次或并行使用自注意力机制对这两种序列进行长距离关系捕获,并未对自注意力机制本身进行改进,使其具备空-谱特征提取能力。此外,这些高光谱图像分类算法中,往往只进行单一地输出预测,未能结合多个视野的信息对图像类别进行综合预测,这给模型的性能带来了瓶颈。

基于此,本文提出了一种空-谱自注意力 Transformer (Spatial-spectral Self-attention Transformer, S³AT)。首先,为了适应不同仪器设备所采集的高光谱医学图像的空-谱信息密度不同,该模型在原有自注意力机制基础上,将空间注意力和光谱注意力融入自注意力机制中,寻

找空间特征和光谱特征之间的内在关联,得到空-谱自注意力。其次,将不同空-谱 Transformer 编码器中的卷积核大小设计为不同尺寸,以获得不同视野下的空-谱自注意力,并对其进行融合。在最后分类过程中,网络在不同视野下分别进行预测,并将预测结果使用可学习的预测权重进行融合,形成最终分类结果。在 In-vivo Human Brain 和 BloodCell HSI 数据集上进行的实验表明,该方法充分挖掘了空-谱特征,有效地融合不同视野下获取的信息,在不同仪器所获得的高光谱医学图像上均具有明显的精度优势。

2 空-谱自注意力 Transformer(S³AT)

本文所提出的 S³AT 算法如图 1 所示。首先以高光谱图像的一个像素点为中心取出一个图像块作为样本,沿光谱维将图像块展开后输入 S³AT 网络。网络由三阶段空-谱 Transformer 编码器组成,其内部卷积核的空间感受野(kernel size)依次由大到小。在每个阶段的编码器中,所获得的空-谱自注意力会与之前更大视野下所获得空-谱自注意力进行融合。最后,通过可训练系数对这些不同视野下预测进行加权融合,形成最终的输出结果。下面对空-谱 Transformer 编码器和预测加权融合分别进行介绍。

2.1 空-谱 Transformer 编码器

在不同仪器、不同获取条件获得高光谱医学图像中,信息在空间像素间、光谱波段间的分布往往存在较大差异,这需要模型精细地描绘像素与像素间、波段与波段之间的关系,自适应地挖掘空间信息与光谱信息之间的内蕴关联。因此,本文设计了一种空-谱 Transformer 编码器,其如图 2(a)所示。首先,通过层归一化、线性层和 Re-shape(sequence to patch)操作,将输入特征映射为三个矩阵 $Q \in \mathbb{R}^{b \times w \times w}$, $K \in \mathbb{R}^{b \times w \times w}$ 和 $V \in \mathbb{R}^{b \times w \times w}$,其中 w 为特征的空间尺度, b 为特征所含波段数。随后,这三个矩阵输入空-谱自注意力(Spatial-spectral Self-attention, S³A)机制模块,以获取空-谱特征。在空-谱自注意力模块中,将 Q 输入空间注意力模块,以精确地挖掘高光谱医学图像中不同像素间的关系,赋予特征图不同

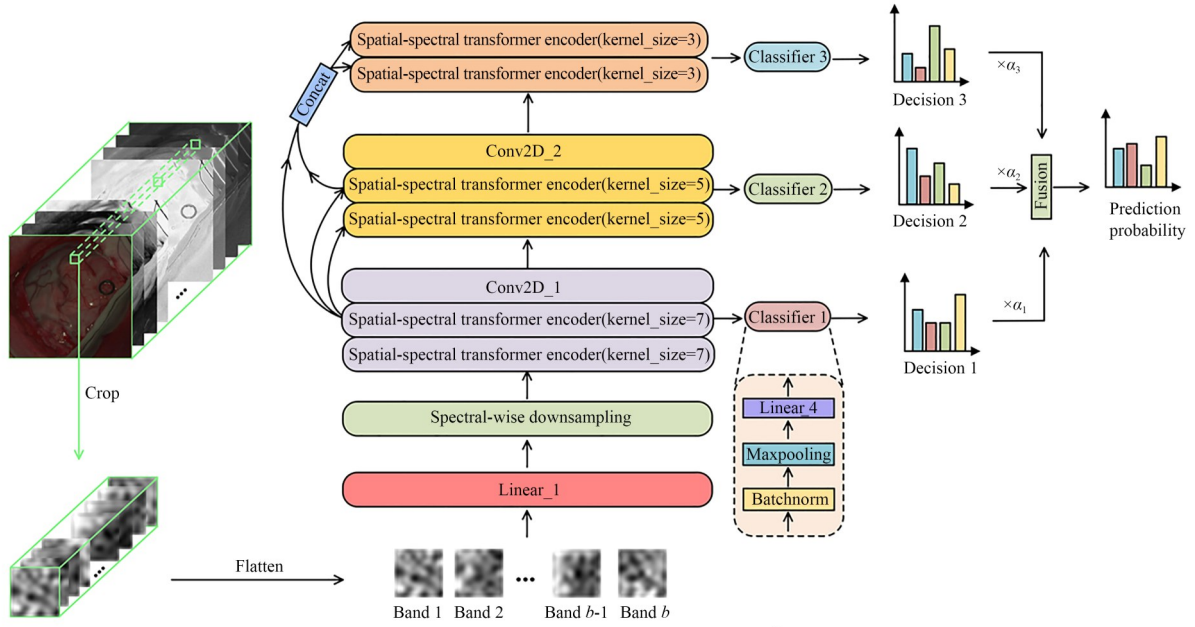


图 1 空-谱自注意力 Transformer 流程图

Fig. 1 Flowchart of spatial-spectral self-attention transformer

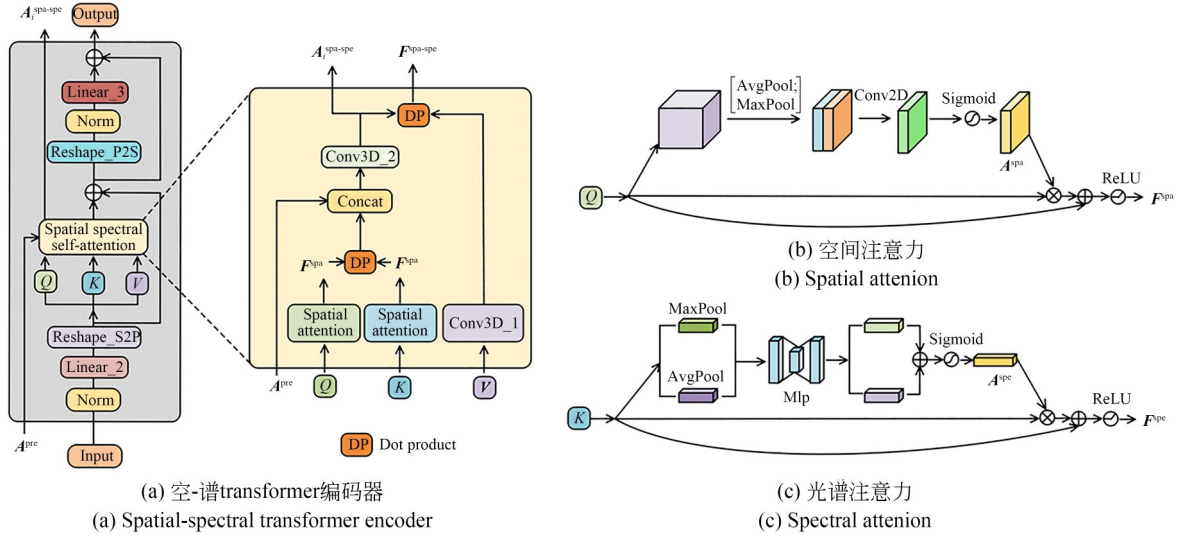


图 2 空-谱 Transformer 编码器结构图

Fig. 2 Structure of spatial-spectral transformer encoder

空间位置的以不同的重要性,提取更加具有鉴别性的空间特征,其具体结构如图 2(b)所示。首先通过全局最大池化和全局平均池化对 Q 的通道域特征进行压缩,并将所得的两个特征沿通道维进行拼接。接着,通过一个卷积层将这个二通道特征转换为单通道特征,再以一个 Sigmoid 函数对其激活得到空间注意力。此过程可表示为:

$$A^{spa} = \text{Sigmoid}(f_{n \times n}([\text{AvgPool}(Q); \text{MaxPool}(Q)])), \quad (1)$$

其中, $\text{AvgPool}(\cdot)$ 表示平均池化操作,

$[\cdot]$ 表示特征图拼接, $\text{MaxPool}(\cdot)$ 表示最大池化操作, $\text{Sigmoid}(\cdot)$ 为 Sigmoid 激活函数操作, $f_{n \times n}(\cdot)$ 表示卷积层,其中 n 表示感受野大小。将 Q 与空间注意力进行点乘并残差连接,再以 Re-

LU 函数激活,可得空间特征:

$$F^{\text{spa}} = \text{ReLU}(\mathbf{Q} + \langle \mathbf{Q}, \mathbf{A}^{\text{spa}} \rangle_{3,2}), \quad (2)$$

其中: $\langle \mathbf{X}, \mathbf{Y} \rangle_{a,b}$ 表示 a 维矩阵 $\mathbf{X}$ 与 b 维矩阵 $\mathbf{Y}$ 按广播机制进行点乘, $\text{ReLU}(\cdot)$ 为 ReLU 激活函数。

获取空间特征之后,为了精确地描绘波段与波段之间的关系,更好地赋予不同波段以不同权重,提取对诊断有帮助的波段,将 $\mathbf{K}$ 输入光谱注意力模块,其结构如图 2(c) 所示。首先,使用最大池化和均值池化,把 $\mathbf{K}$ 的每个波段内的空间特征信息进行压缩。然后采用 MLP 对压缩特征进行映射,以提高压缩信息的迁移能力。最后,在每个波段上对两种压缩方式得到的压缩信息相加融合并以 Sigmoid 函数激活,得到光谱注意力。这个过程可表示为:

$$A_i^{\text{spa-spe}} = \begin{cases} f_{3 \times 3 \times 3}(\langle F^{\text{spa}}, F^{\text{spe}} \rangle_{3,3}), & i = 1 \\ f_{3 \times 3 \times 3}(\text{Concat}(A_1^{\text{spa-spe}}, \langle F^{\text{spa}}, F^{\text{spe}} \rangle_{3,3})), & i = 2 \\ f_{3 \times 3 \times 3}(\text{Concat}(A_1^{\text{spa-spe}}, A_2^{\text{spa-spe}}, \langle F^{\text{spa}}, F^{\text{spe}} \rangle_{3,3})), & i = 3 \end{cases} \quad (5)$$

随后, $A_i^{\text{spa-spe}}$ 会分成两条支路:一路会直接输入下一个编码器进行不同视野关键信息融合,另一路会和 $f_{n \times n \times n}(\mathbf{V})$ 进行点乘,得空-谱特征为:

$$F^{\text{spa-spe}} = \langle A_i^{\text{spa-spe}}, f_{n \times n \times n}(\mathbf{V}) \rangle_{3,3}, \quad (6)$$

其中: $f_{n \times n \times n}(\cdot)$ 表 3D 卷积层, n 表示本层编码器感受野大小。最后,经过 Reshape (patch to sequence), LayerNorm 以及 Linear 层后,空-谱特征从本层编码器输出。

2.2 多视野预测融合

如图 1 所示,文本设计了一种多视野预测融合 (Multi-View Predictions Fusion, MVPF) 策略,将不同感受野下的编码器对样本分别进行预测,并对所有预测结果进行有机融合,弥补网络模型单一预测的不足。具体而言,将三个不同视野下的 Transformer 编码器所得空-谱特征分别接以一个分类器,输出三个类别预测,再对这三个预测进行加权融合。将大视野到小视野所得的三个预测分别表示为 p_1, p_2 和 p_3 , 则最终融合预测 p^{fusion} 满足:

$$A^{\text{spe}} = \text{Sigmoid}(\text{MLP}(\text{AvgPool}(\mathbf{K})) + \text{MLP}(\text{MaxPool}(\mathbf{K}))), \quad (3)$$

其中, $\text{MLP}(\cdot)$ 为多层感知机。随后可得光谱特征:

$$F^{\text{spe}} = \text{ReLU}(\mathbf{K} + \langle \mathbf{K}, A^{\text{spe}} \rangle_{3,1}). \quad (4)$$

为了获得空间特征和光谱特征的内蕴关联,详细地刻画高光谱医学图像空-谱信息分布,将 F^{spa} 和 F^{spe} 进行点积,可得本层编码器下空-谱自注意力。为了更好地利用不同视野下所获得的空-谱自注意力,对不同视野下所获的关键空-谱信息进行整合,本文设计的空-谱自注意力中,将之前更大视野下获得空-谱自注意力 A^{pre} 与本层所得空-谱自注意力进行拼接融合,并接一个卷积层对拼接后的维度进行降维。记本层编码器的序号为 $i, i \in \{1, 2, 3\}$, 则多视野融合后的空-谱自注意力可表示为:

$$\begin{cases} p^{\text{fusion}} = \sum_{i=1}^3 \alpha_i p_i \\ s. t. \sum_{i=1}^3 \alpha_i = 1 \end{cases}, \quad (7)$$

其中, $\alpha_i, i \in \{1, 2, 3\}$ 为第 i 个视野下的可训练预测权重。这些权重的训练使用标准的反向传播算法来进行。由 (7) 式可得:

$$p^{\text{fusion}} = \alpha_1 p_1 + \alpha_2 p_2 + (1 - \alpha_1 - \alpha_2) p_3. \quad (8)$$

设损失函数为 $L(p^{\text{fusion}})$, 则对于可训练参数 α_1 和 α_2 的梯度可以通过链式法则计算得到,其表达式如下:

$$\frac{\partial L}{\partial \alpha_1} = \frac{\partial L}{\partial p^{\text{fusion}}} \frac{\partial p^{\text{fusion}}}{\partial \alpha_1} = \frac{\partial L}{\partial p^{\text{fusion}}} (p_1 - p_3), \quad (9)$$

$$\frac{\partial L}{\partial \alpha_2} = \frac{\partial L}{\partial p^{\text{fusion}}} \frac{\partial p^{\text{fusion}}}{\partial \alpha_2} = \frac{\partial L}{\partial p^{\text{fusion}}} (p_2 - p_3), \quad (10)$$

其中, $\frac{\partial L}{\partial p^{\text{fusion}}}$ 表示损失函数对最终输出 p^{fusion} 的梯度,根据这些梯度,使用优化器更新可训练参数 α_1 和 α_2 。

2.3 网络模型参数

为了详细说明 S³AT 的结构,受篇幅所限,其重要可训练网络参数如表 1 所示。

表 1 S³AT 的模型参数
Tab. 1 Model parameter of S³AT

Layer name	Parameter
Linear_1	Linear, 81×81
Spectral-wise Downsampling	Conv2D, 3×3 , 80, stride 1, padding 1
Linear_2	Linear, 81×384
Conv2D (Spatial Attention)	Conv2D, kernel_size $\times$ kernel_size, 1, stride 1, padding 1
MLP (Spectral Attention)	Conv2D, 3×3 , 160, stride 1, padding 1
	Conv2D, 3×3 , 80, stride 1, padding 1
Conv3D_1	Conv3D, $3 \times 3 \times 3$, 2, stride 1, padding 1
Conv3D_2	Conv3D, $3 \times 3 \times 3$, 2, stride 1, padding 1
Conv2D_1	Conv2D, 3×3 , 80, stride 1, padding 1
Conv2D_2	Conv2D, 3×3 , 80, stride 1, padding 1
Linear_3	Linear, 128×144
Linear_4	Linear, 80×4

3 实验结果与分析

3.1 实验数据

为了验证 S³AT 在面向不同仪器、不同成像以及不同预处理方式下所获得高光谱医学数据,均具有出色的分类性能,本文采用 In-vivo Human Brain HSI Dataset 和 BloodCell HSI Dataset 来进行对比实验。下面分别对其进行介绍:

(1) In-vivo Human Brain HSI Dataset (Brain HSI Dataset): 该数据集由英国南安普顿大学医院(UHS)和西班牙拉斯帕尔马斯大学内格林医院(UHDRN)共同采集。采集系统由 Hyper-spec[®] VNIR A-Series 相机组成。相机基于推扫技术,使用硅 CCD 探测器阵列,最低帧率为 90 帧/秒,光谱范围为 400~1 000 nm,光谱分辨率为 2~3 nm,可捕获 826 个光谱波段,每行 1 004 个空间像素。采集对象为进行开颅切除脑肿瘤手术过程中的 16 名成年患者,最终获得 26 张高光谱图像,其共包含背景、正常、肿瘤以及血管四个类别。在本文实验中,选取包含全部四种类别的高光谱图像进行实验,共包含 6 个病人、9 张图像。

(2) BloodCell HSI Dataset: 该数据集是通过将显微镜和硅电荷耦合装置与 VariSpec[®] 液晶可调谐滤波器(Liquid Crystal Tunable Filter, LCT-Fs)结合起来收集。该数据集包含两张血细胞图像,其分别命名为 Bloodcell1-3 和 Bloodcell2-2。Bloodcell1-3 的大小为 973×799 pixel, Bloodcell2-2 的大小为 462×451 pixel,它们都含 33 个

波段。每张高光谱图像含有红细胞、白细胞和背景 3 个类别。

由于成像方式不同,采集设备不同,以上两个数据集中的高光谱图像的空间分辨率和光谱分辨率存在较大差异,进而空间信息和光谱信息分布有所不同。为了展示这个特性,在两个数据集中各自随机选取一个样本点,以其为中心裁剪出一个 patch,分别做出光谱曲线以及某个随机波段的二维图像,最终可视化结果如图 3 所示。

由图 3 可知,本文采用的两个数据集中图像的波段数、光谱曲线以及空间分辨率存在较大差异,因此可以验证所提出算法在不同类型高光谱医学图像上的有效性。

3.2 实验设置

为验证本文算法的有效性,选取卷积神经网络方法 HybridSN^[24], SSRN(Spectral-Spatial Residual Network)^[25] 和 DBDA(Double-Branch Dual-Attention)^[26], Transformer 深度学习方法 Spectral-wise ViT^[27], SSFTT^[28] 和 CTMixer^[29] 作为对比算法。每种算法重复进行 10 次实验,以均值±标准差(Standard Deviation, STD)的形式表征总体分类精度(Overall Accuracy, OA)、平均分类精度(Average Accuracy, AA)以及 Kappa 系数(Kappa Coefficient, KC),以便综合比较并判断各算法的分类性能。在实验中,按波段对高光谱血细胞数据进行归一化处理,而学习率以及样本 Patch 大小,均由实验确定。在两个数据上的实验设置如表 2 所示。

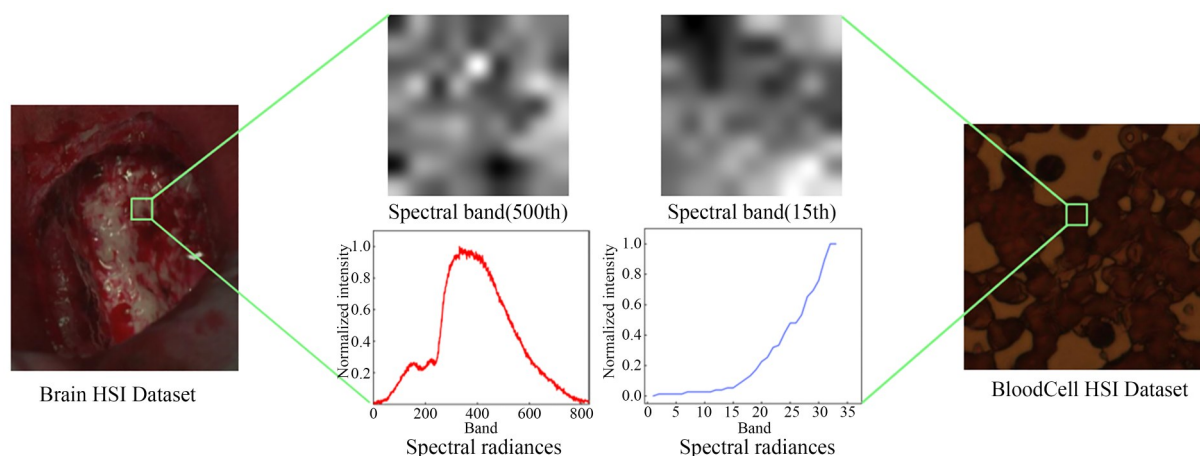


图3 Brain和BloodCell HSI数据集上的空间和光谱信息可视化

Fig. 3 Visualization of spatial and spectral information on Brain and BloodCell HSI Dataset

表2 Brain和BloodCell HSI数据集上的实验设置

Tab. 2 Experimental setup on Brain and BloodCell HSI datasets

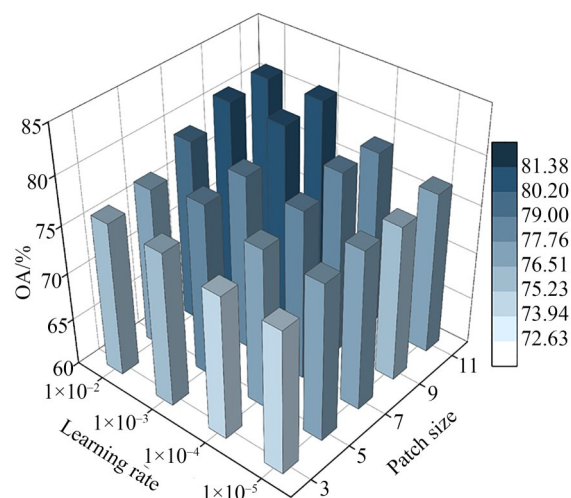
Setting Item	Dataset	
	Brain HSI	BloodCell HSI
Training Set	50% of patients BloodCell1-3	
Valid Set		
Test Set	50% of patients BloodCell2-2	
Training Samples	30 samples/class	
Valid Samples	10 samples/class	
Test Samples	All samples	
Loss Function	Cross Entropy Loss	
Epoch	300	
Batchsize	128	
Optimizer	Adamw	
Python Version	3.7	
Deep Learning Framework	Paddle	
GPU	Tesla V100 (32 G)	
Computing Platform	Baidu AI Studio	

3.3 参数实验分析

为了对所提出的 S^3AT 进行全面的研究,本文分析了样本patch大小和学习率对分类精度的影响。对于输入样本patch大小,不仅影响空间信息量,而且影响模型的复杂性。至于学习率,学习率过大会导致模型过快收敛到次优解,而学习率过小则会导致收敛过程停滞。因此,这

些参数需要进行实验以获得更好的分类精度。对于学习率,选取范围为 $\{1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}\}$ 。对于样本patch大小,选取范围为 $\{3, 5, 7, 9, 11\}$ 。实验中,采用网格搜索确定最佳参数,数据集划分与表1相同,结果如图4所示。

由图4(a)和图4(b)可知,增加样本patch大小可以明显提高分类精度。这是因为更大的patch包含了更多的空间信息,提高了样本的鉴别性。考虑到运算效率,在两个数据集上,patch大小均设为9。同时,过小的学习率会使得模型更难获得高级特征,而学习率过大会使模型发散和梯度爆炸,因此在本文两个数据集中,学习率均设置为 1×10^{-3} 。

(a) Brain HSI数据集中 S^3AT 的参数分析(a) Parameter analysis of S^3AT on brain HSI dataset

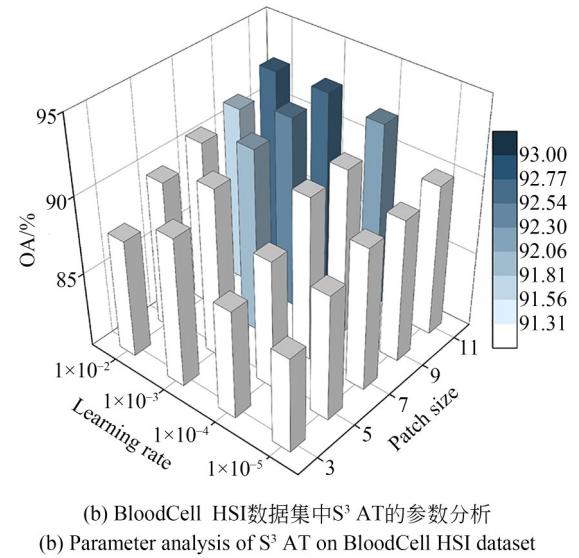


图 4 Brain 和 BloodCell HSI 数据集上的参数分析
Fig. 4 Parameter analysis on Brain and BloodCell HSI datasets

3.4 消融实验分析

在文本所提出的 S³AT 模型中,空-谱自注意力机制和多视野预测融合策略占据关键地位。为了验证它们的有效性,以单视野下的原始 Transformer 网络为 Baseline,在 Brain HSI 数据集上进行关于视野个数 (Number of Views, NV) 以及空-谱自注意力机制的消融实验。结果如表 3 所示。

由表 3 可知,将 S³A 单独添加到单视野网络后,OA,AA 和 KC 分别提升了 3.39%,4.44% 和 10.35%。这说明 S³A 模块成功关注到了关键的空-谱特征区域,提取出了更具鉴别性空-谱特征。在添加 S³A 的状态下,随着视野数的增多,模型的预测能力得到进一步的提升。这是因为多视野预测融合策略成功融合不同视野下的决策。

表 3 S³AT 关于不同模块的消融实验分析
Tab. 3 Ablation analysis of the proposed S³AT with a combination of different components

NV	S ³ A	Metrics		
		OA/%	AA/%	KC/%
1	×	72.83	72.49	62.88
1	✓	76.22	76.93	73.23
2	✓	79.09	77.64	69.92
3	✓	80.62	78.37	74.96
4	✓	79.33	78.00	71.61
5	✓	76.43	76.42	67.80

当视野个数为 3 时,模型分类表现达到最优。其 OA,AA 和 KC 分别提升了 7.79%,5.88% 和 12.08%。当视野个数进一步增大时,模型预测性能下降,这是因为过多的视野使得模型过于复杂,引入过多可训练参数,陷入过拟合状态。

3.5 多视野预测融合分析

在本文所提出的 S³AT 模型中,多视野预测融合策略占据了重要地位。为了展示不同视野下的预测效果,在 Brain HSI 数据集上,分别使用三个视野下对应的分类器单独对整张测试图像进行预测,同时使用预测融合后的预测图作为对照,实验结果如图 5 所示。

图 5 可以看出,在原有各个视野下的预测的基础上,多视野预测融合策略取得了更好的分类效果。这是由于不同视野下的空-谱 Transformer 编码器获得了不同的关键信息,所设计的多视野融合策略有效地对这些信息赋予不同的权重,有效地对其进行整合,从而更好地利用不同视野下所获的信息,取得更高的分类精度。值得注意的是,从大视野至小视野,模型所获的预测权重分别为 0.41,0.36 和

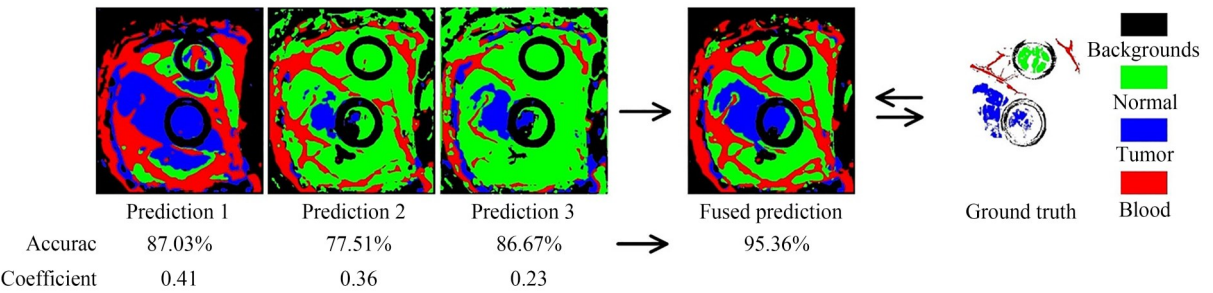


图 5 多视野预测融合分析
Fig. 5 Analysis of multi-view predictions fusion

0.23。这是因为模型在大视野下获得了更多的整体信息,而在小视野下,模型获得更多的细节信息作为补充。

3.6 对比实验结果

3.6.1 Brain HSI数据集结果

在 Brain HSI数据集上,实验结果由表4所示。由表4可知,ViT取得了最差的分

类精度。这是由于ViT偏重于光谱特征提取,未能充分提取到充足的空间特征。其余对比方法均取得了稍好的分类结果,这是因为它们均包含空-谱特征提取模块,提升了模型的鉴别能力。在对比方法中,SSRN取得了更好的分类结果,这是因为SSRN中的空间注意力和光谱注意力模块关注到了重点空间区域和重要光谱波段,消除了特征冗余。然而,其单一的预测使得模型精度受限。在所有方法中,本文提出的方法取得了更好的分类结果。这是因为S³AT中的空-谱自注意力机制赋予空-谱特征以不同权重,并且其将不同视野下的诊断预测按不同的权重进行融合,提升了模型的预测能力。

表4 Brain HSI数据集上不同算法的分类结果

Tab. 4 Classification results of different algorithms on Brain HSI Dataset (%)

Class	CNN-based Methods			Transformer-based Methods			
	HybridSN	SSRN	DBDA	ViT	CTMixer	SSFTT	S ³ AT
Background	61.64±5.26	72.76±2.01	67.29±6.76	62.63±9.95	71.42±3.85	60.53±4.50	77.27±9.97
Normal	31.73±9.75	37.32±9.61	35.28±9.04	15.90±4.59	35.25±9.32	23.08±5.26	73.00±9.99
Tumor	84.02±4.70	84.25±3.59	82.51±1.46	67.45±12.37	83.09±1.83	83.84±0.87	83.48±2.15
Blood	98.95±2.12	100.00±0.00	95.12±10.22	69.22±16.45	99.99±0.03	99.55±0.93	95.35±7.88
OA	68.32±3.75	72.25±2.67	68.99±3.98	51.75±5.93	71.07±2.77	65.66±1.35	82.25±2.88
AA	69.08±3.29	73.58±2.15	70.05±3.79	53.80±5.07	72.44±2.28	66.75±1.20	82.27±3.03
KC	58.15±4.79	63.13±3.47	58.87±5.30	36.72±7.42	61.76±3.52	54.65±1.74	76.17±3.86

为了直观对比不同方法的预测,在上述实验中,取其中一次实验中一张预测图作为展示,结果如图6所示。可以看出,本文算法相比其他方法,分类图错分点较少,更为平滑。这是因为S³AT在面向基于反射光成像的高光谱图像时,

可以自适应地获取空-谱自注意力,详细地描绘出图像的空-谱信息分布,挖掘更具鉴别性的空-谱特征,并将多视野所得到的空-谱自注意力融合。此外,在预测阶段,多视野预测的有机融合,使得模型的预测更加精确。

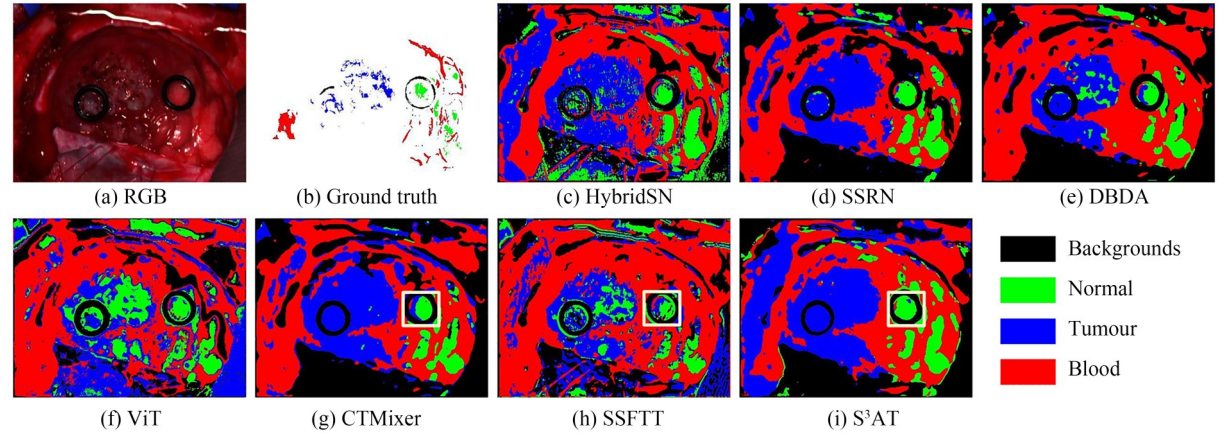


图6 各算法在 Brain HSI数据集上的分类结果图

Fig. 6 Classification maps of different methods on Brain HSI Dataset

3. 6. 2 BloodCell HSI数据集结果

在 BloodCell HSI数据集上,实验结果如表 5 所示。由表 5 可知,S³AT 在大多数指标上,依旧取得更好的分类结果。这是因为在面向基于显微镜透射光成像的高光谱图像时,S³AT 的空-谱自注意力模块仍然能够捕获关键的空-谱信息,提高模型的分类能力,并且融合了不同视野下的诊断信息。这说明 S³AT 可以适用于不同仪器、

不同成像方式所获取的高光谱医学图像,具有较好的泛化性,节约了模型开发成本。为了直观对比不同方法的预测,在上述实验中,取其中一次实验的预测图作为展示,结果如图 7 所示。由图 7 可知,S³AT 所得分类图更为光滑,误分点较少。这说明基于空-谱自注意力机制和多视野预测融合的 S³AT 算法的分类性能有明显提升,具有强的鲁棒性,更适合实际应用场景。

表 5 BloodCell HSI数据集上不同算法的分类结果
Tab. 5 Classification results of different algorithms on BloodCell HSI Dataset (%)

Class	CNN-based Methods			Transformer-based Methods			
	HybridSN	SSRN	DBDA	ViT	CTMixer	SSFTT	S ³ AT
RedBloodCell	87.16±1.21	85.35±3.07	85.60±1.21	70.79±3.08	88.52±1.48	88.39±3.59	89.22±2.69
Background	96.71±1.46	96.82±3.79	95.54±1.61	93.77±4.19	96.59±3.93	95.67±1.29	98.64±3.79
WhiteBloodCell	61.67±3.68	84.60±2.60	83.44±1.83	76.50±3.39	55.35±3.26	47.32±2.55	80.90±2.93
OA	89.21±3.88	88.71±2.40	88.25±3.45	77.13±3.19	90.45±1.61	89.50±1.39	91.74±2.79
AA	81.85±1.32	89.26±1.75	88.19±0.85	80.35±2.01	80.37±1.64	77.13±1.48	88.97±1.75
KC	77.06±2.89	76.66±3.02	75.53±1.44	57.78±3.36	78.34±0.77	77.06±2.32	81.86±2.19

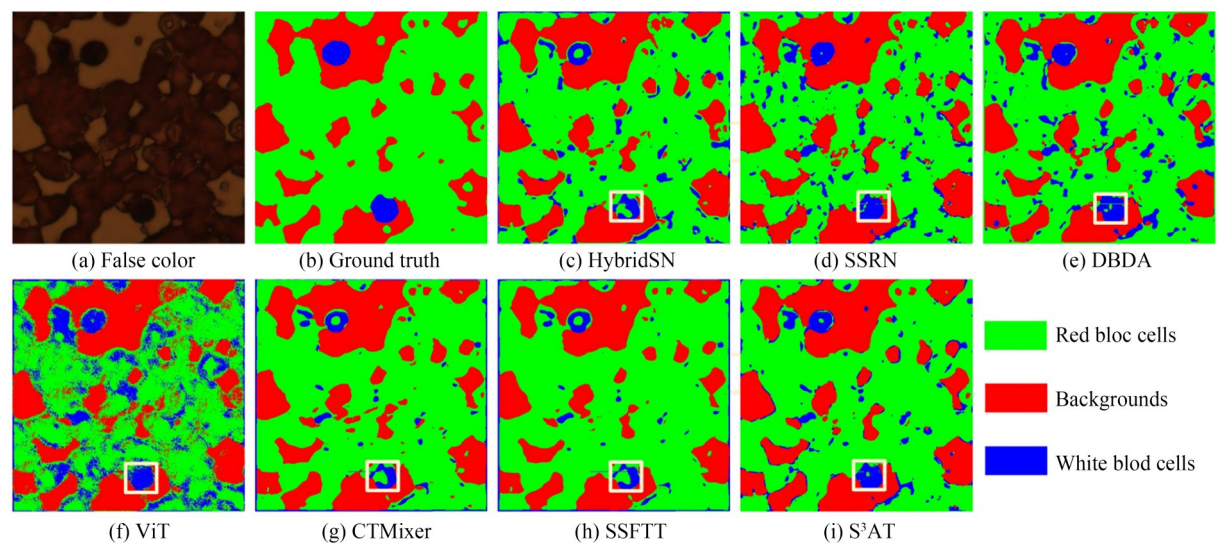


图 7 各算法在 BloodCell HSI数据集上的分类结果图
Fig. 7 Classification maps of different methods on BloodCell HSI Dataset

3. 7 模型时空代价分析

所有算法在执行时都需要消耗时间和空间资源,因此,对算法的时空代价进行分析非常必要。在神经网络中,参数数量可用于表示网络的空间复杂度和大小,也对应计算机内存资源的消耗。该指标越小,则表示网络的空间复杂度越小。浮点运算次(Floating Point Operations,

FLOPs)表示每秒完成预测所需的计算量,用来衡量网络的运算速度。该指标越小,则表示网络的时间复杂度越小。而推理时间则直接反映了一个算法在某一个设备上运算效率。使用 Brain HSI Dataset 上对 32 个测试样本在对本文使用的所有算法进行时空代价分析,实验平台如表 1 所示,实验结果如表 6 所示。

表 6 不同算法的参数量、FLOPs 以及推理时间比较

Tab. 6 Parameters, FLOPs and inference time comparison of different algorithms

	Parameters/ ($\times 10^6$)	FLOPs/ ($\times 10^9$)	Inference Time/s
HybridSN	14.33	8.15	0.055 6
SSRN	1.26	20.78	0.096 4
DBDA	2.31	32.20	0.169 1
ViT	18.82	54.12	0.179 3
CTMixer	0.61	1.63	0.018 8
SSFTT	3.63	3.32	0.027 6
S ³ AT	2.00	7.00	0.045 6

如表 6 可知, S³AT 相比 HybridSN, SSRN, DBDA, ViT 取得了更少的运算次数、更少的推理时间以及更少的模型参数(DBDA 除外)。虽然 SSFTT 和 CTMixer 在时间复杂度上取得更低的结果, 但 S³AT 分类性能比上述两个算法有显著性提升。这说明本文所提出算法具有较高的性能和效率, 在高光谱医学图像分类任务中得到更好的表现, 进而可以在计算资源受限的场景下得到更好地应用, 具有较高的实用价值和推广前景。

参考文献:

- [1] 李伟, 吕蒙, 陈天虹, 等. 高光谱图像在生物医学中的应用[J]. 中国图象图形学报, 2021, 26(8): 1764-1785.
LI W, LYU M, CHEN T H, *et al.* Application of a hyperspectral image in medical field: a review[J]. *Journal of Image and Graphics*, 2021, 26(8): 1764-1785. (in Chinese)
- [2] CHEN Y R, WANG Y N, ZHANG H, *et al.* 医学高光谱显微成像与智能分析关键技术研究及应用[J]. *AI-View*, 2022(3): 22-32. (in Chinese)
陈煜嵘, 王耀南, 张辉, 等. 医学高光谱显微成像与智能分析关键技术研究及应用[J]. *人工智能*, 2022(3): 22-32.
- [3] UL A, REHMAN. A review of the medical hyperspectral imaging systems and unmixing algorithms' in biological tissues[J]. *Photodiagnosis and Photodynamic Therapy*, 2021, 33: 102165.
- [4] 闫敬文, 陈宏达, 刘蕾. 高光谱图像分类的研究进展[J]. 光学精密工程, 2019, 27(3): 680-693.
YAN J W, CHEN H, LIU L. Overview of hyper-

4 结 论

在高光谱医学图像分类任务中, 为了克服 Transformer 网络难以适应不同类型的高光谱图像而导致的性能表现差异较大, 以及未能使用多个感受野的空-谱信息的问题, 本文基于空-谱自注意力机制以及多视野预测融合策略, 提出一种空-谱自注意力 Transformer (S³AT)。该方法能根据高光谱医学图像中的空-谱信息分布, 自适应挖掘重点空-谱信息, 并将不同感受野下所得空-谱自注意力进行融合, 且将不同感受野下的预测进行加权融合。在 Brain 和 Bloodcell HSI 高光谱数据集上, OA, AA 和 KC 分别获得了 82.25%, 82.27% 和 76.17% 以及 91.74%, 88.97% 和 81.86%。实验结果表明, S³AT 对不同类型的高光谱医学图像, 均具有高精度的分类效果。然而, S³AT 中的各个感受野大小为手工设定, 未能根据图像自适应进行尺寸调整。因此下一步研究工作将关注如何设计一种自适应动态感受野, 从而使模型更加有效地获取不同视野下的空-谱信息。

- spectral image classification [J]. *Opt. Precision Eng.*, 2019, 27(3): 680-693. (in Chinese)
- [5] 宋蓓蓓, 马穗娜, 何帆, 等. Res2-Unet 深度神经网络的 RGB-高光谱图像重建[J]. 光学精密工程, 2022, 30(13): 1606-1619.
SONG B B, MA S N, HE F, *et al.* Hyperspectral reconstruction from RGB images based on Res2-Unet deep learning network[J]. *Opt. Precision Eng.*, 2022, 30(13): 1606-1619. (in Chinese)
- [6] 时旭, 李远, 黄鸿. 面向高光谱显微图像血细胞分类的空-谱可分离卷积神经网络[J]. 光学精密工程, 2022, 30(8): 960-969.
SHI X, LI Y, HUANG H. Spatial-spectral separable convolutional neural network for cell classification of hyperspectral microscopic images[J]. *Opt. Precision Eng.*, 2022, 30(8): 960-969. (in Chinese)
- [7] 郑少佳, 邱崧, 李庆利, 等. 傅里叶变换通道注意力网络的胆管癌高光谱图像分割[J]. 中国图象图形学报, 2021, 26(8): 1836-1846.
ZHENG S J, QIU S, LI Q L, *et al.* Fourier transform channel attention network for cholangiocarcino-

- ma hyperspectral image segmentation [J]. *Journal of Image and Graphics*, 2021, 26(8): 1836-1846. (in Chinese)
- [8] FABELLO H, ORTEGA S, SZOLNA A, *et al.* *In-vivo* hyperspectral human brain image database for brain cancer detection [J]. *IEEE Access*, 2019, 7: 39098-39116.
- [9] BØTKER J, WU J X, RANTANEN J. *Data Handling in Science and Technology* [M]. Elsevier, 2020.
- [10] YIFAN, DUAN, . Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images [J]. *Optics & Laser Technology*, 2019, 112: 530-538.
- [11] RUIZ L, MARTÍN A, URBANOS G, *et al.* Multiclass brain tumor classification using hyperspectral imaging and supervised machine learning [C]. 2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS). November 18-20, 2020, Segovia, Spain. IEEE, 2020: 1-6.
- [12] BALTUSSEN E J M, KOK E N D, DE KONING S G B, *et al.* Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery [J]. *Journal of Biomedical Optics*, 2019, 24(1): 016002.
- [13] HUANG Q, LI W, ZHANG B C, *et al.* Blood cell classification based on hyperspectral imaging with modulated Gabor and CNN [J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(1): 160-170.
- [14] WEI X L, LI W, ZHANG M M, *et al.* Medical hyperspectral image classification based on end-to-end fusion deep neural network [J]. *IEEE Transactions on Instrumentation and Measurement*, 2019, 68(11): 4481-4492.
- [15] ZHANG Q, WANG Y, QIU S, *et al.* 3D-Pul-CNN: pulmonary cancer classification from hyperspectral images using convolution combination unit based CNN [J]. *Journal of Biophotonics*, 2021, 14(12): e202100142.
- [16] HU B L, DU J, ZHANG Z F, *et al.* Tumor tissue classification based on micro-hyperspectral technology and deep learning [J]. *Biomedical Optics Express*, 2019, 10(12): 6370.
- [17] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [EB/OL]. 2020; *arXiv*: 2010.11929. <https://arxiv.org/abs/2010.11929.pdf>
- [18] ZHOU Z H, QIU S, WANG Y, *et al.* Swin-spectral transformer for cholangiocarcinoma hyperspectral image segmentation [C]. 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 23-25, 2021, Shanghai, China. IEEE, 2021: 1-6.
- [19] LI N Y, XUE J Q, JIA S. Spectral context-aware transformer for cholangiocarcinoma hyperspectral image segmentation [C]. *Proceedings of the 2022 5th International Conference on Image and Graphics Processing*. January 7 - 9, 2022, Beijing, China. New York: ACM, 2022: 209-213.
- [20] LI Y, SHI X, YANG L P, *et al.* MC-GAT: multi-layer collaborative generative adversarial transformer for cholangiocarcinoma classification from hyperspectral pathological images [J]. *Bio-medical Optics Express*, 2022, 13(11): 5794-5812.
- [21] PENG Y S, ZHANG Y W, TU B, *et al.* Spatial-spectral transformer with cross-attention for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-15.
- [22] OUYANG E, LI B, HU W J, *et al.* When multi-granularity meets spatial - spectral attention: a hybrid transformer for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-18.
- [23] LIU B, YU A Z, GAO K L, *et al.* DSS-TRM: deep spatial-spectral transformer for hyperspectral image classification [J]. *European Journal of Remote Sensing*, 2022, 55(1): 103-114.
- [24] ROY S K, KRISHNA G, DUBEY S R, *et al.* HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification [J]. *IEEE Geoscience and Remote Sensing Letters*, 2020, 17(2): 277-281.
- [25] ZHONG Z L, LI J, LUO Z M, *et al.* Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 847-858.
- [26] LI R, ZHENG S Y, DUAN C X, *et al.* Classifi-

- cation of hyperspectral image based on double-branch dual-attention mechanism network[J]. *Remote Sensing*, 2020, 12(3): 582.
- [27] HONG D F, HAN Z, YAO J, *et al.* SpectralFormer: rethinking hyperspectral image classification with transformers[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-15.
- [28] SUN L, ZHAO G R, ZHENG Y H, *et al.* Spectral - spatial feature tokenization transformer for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-14.
- [29] ZHANG J J, MENG Z, ZHAO F, *et al.* Convolution transformer mixer for hyperspectral image classification [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.

作者简介:



李 远(1993—),男,河南三门峡人,博士研究生,2015年于西北工业大学获得学士学位。主要从事高光谱医学图像处理及深度学习等方面的研究。
E-mail:yuan_li@cqu.edu.cn

通讯作者:



黄 鸿(1980—),男,湖南新宁人,博士生导师,2003、2005、2008年于重庆大学分别获得学士、硕士和博士学位。主要从事医学影像智能化处理、模式识别、深度学习等方面的研究。E-mail:hhuang@cqu.edu.cn