

# SCA-Net: 基于高效卷积和注意力机制的恒星 大气物理参数自动估计\*

张皓凯 王洪丰<sup>†</sup> 陶振钊 马 硕 张厚水

(德州学院计算机与信息学院 德州 253023)

**摘要** 对恒星大气物理参数进行估计(表面有效温度 $T_{\text{eff}}$ 、表面重力加速度 $\lg g$ 、金属丰度 $[\text{Fe}/\text{H}]$ )是研究恒星时所面临的首要任务. 为满足对海量恒星光谱数据的分析需求, 该研究基于LAMOST DR10 (Large Sky Area Multi-Object Fiber Spectroscopy Telescope Data Release 10)低分辨率实测光谱公开数据, 提出SCA-Net (Spectral Convolution-Attention Network)模型. SCA-Net融合了移动翻转瓶颈卷积, 自注意力机制与多尺度特征融合, 实现了对恒星大气物理参数的精确估计. 该研究在28913条低分辨率实测光谱上做了实验研究, 其中8913条用于训练模型, 20000条用于测试模型. 测试评价标准为平均绝对误差(Mean Absolute Error, MAE), 3个恒星大气物理参数的平均精度分别为: 71.07 ( $T_{\text{eff}}/\text{K}$ )、0.042 ( $\lg [g/(\text{cm} \cdot \text{s}^{-2})]$ )、0.040 dex ( $[\text{Fe}/\text{H}]$ ). 该研究的相关代码在以下地址提供: <https://github.com/HelKai528/SCA-Net>.

**关键词** 恒星: 基本参数, 恒星: 大气, 方法: 数据分析, 方法: 统计, 技术: 光谱

**中图分类号**: P144; **文献标识码**: A

## 1 引言

随着LAMOST<sup>[1-3]</sup> (Large Sky Area Multi-Object Fiber Spectroscopy Telescope)等大规模巡天项目的开展, 恒星光谱数据已快速增长至TB量级, 并有持续增长的趋势. 面对海量光谱数据, 传统人工处理方法缺点日益显露: 其一, 由于人类认知的局限性, 人工构造的特征与模型假设可能引入系统性偏差; 其二, 面对每时每刻都在增长的海量数据, 传统方法的处理效率无法满足实时处理数据的需求. 如何准确、高效地对海量恒星大气物理参数进行自动估计成为了一个重要课题.

早期的恒星参数估计方法主要基于模板匹配技术. Xiang等<sup>[4]</sup>开发了LAMOST恒星参数管道

LSP3 (LAMOST Stellar Parameter Pipeline at Peking University), 通过结合MILES (Medium-resolution INT Library of Empirical Spectra)和ELODIE光谱模板库, 实现了对恒星参数的精准估计. Si等<sup>[5]</sup>使用K近邻算法对SDSS DR8 (Sloan Digital Sky Survey Data Release 8)中的天体进行了估计, 该方法在本质上也可看作是一种模板匹配的思想. 此类方法受光谱模板的限制较为明显, 而光谱模板多来源于实测光谱, 其参数空间在多数情况下并不均匀, 因此又需进行插值, 导致对参数的估计具有较大的不确定性.

随着光谱数据的积累与计算能力的普及, 越来越多的学者开始用数据驱动的方法, 尤其是机器学习技术来挖掘光谱信息. 针对APOGEE

2025-08-20收到原稿, 2025-12-08收到修改稿

\* 山东省自然科学基金项目(ZR2024QA180)、德州市课题(2022dzkj097)资助

<sup>†</sup> [dzuwhf@163.com](mailto:dzuwhf@163.com)

(Apache Point Observatory Galactic Evolution Experiment)、Gaia-ESO (Gaia-European Southern Observatory)等高分辨率光谱, Ness等<sup>[6]</sup>较早地提出了名为The Cannon的数据驱动方法; Leung等<sup>[7]</sup>提出了一种结合贝叶斯神经网络与dropout变分推断的深度学习方法; Fabbro等<sup>[8]</sup>提出了基于一维卷积神经网络的StarNet模型; Candebat等<sup>[9]</sup>则提出了一种基于条件可逆神经网络的OssicoNN (Ossico Neural Network)模型以推断恒星参数和元素丰度. 在LAMOST光谱分析方面, 相关工作同样进展迅速. Wang等<sup>[10]</sup>结合生成光谱网络(Generative Spectrum Network, GNN)与贝叶斯框架实现了参数估计; Wang等<sup>[11]</sup>针对LAMOST中分辨率光谱设计了SPCANet (Stellar Parameters and Chemical Abundances Network), 利用卷积神经网络与双分支结构同时推导大气参数与13种元素丰度; 邓诗宇等<sup>[12]</sup>则尝试引入了可变形卷积网络(Deformable Convolutional Network, DCN)进行参数估计.

近年来, 为克服纯数据驱动模型可能引入的偏差, 已有学者尝试开发数据物理混合驱动模型. DD-Payne<sup>[13-14]</sup>(Data-Driven Payne)是一种代表性方法, 它结合物理驱动模型The Payne和数据驱动模型The Cannon的优点, 确保参数估计的合理性, 有效克服了传统数据驱动方法的局限.

尽管上述研究在恒星大气物理参数自动估计方面取得了显著进展, 但在LAMOST海量低分辨率光谱的特性下, 现有方法仍面临严峻挑战. 现有方法多依赖于卷积神经网络(Convolutional Neural Networks, CNN), 由于卷积核的尺寸限制, 单纯的CNN若想捕捉光谱的长程依赖关系(如连续谱整体形态), 只能通过扩大卷积核或不断堆叠网络层数来增大感受野. 然而, 这种策略不仅会让网络变得极为复杂以致计算成本暴增, 更致命的是对于信息密度本就不高的低分辨率光谱而言, 盲目增加网络深度容易引入大量冗余参数, 从而使网络过拟合的风险大幅增加. 此外, 现有模型架构均面临模型容量与泛化性的权衡难题, 且随着网络层数的加深, 浅层光谱细节(如弱金属吸收线)容易在

下采样过程中丢失, 导致最终参数估计的精度受限. 目前尚缺乏一套能同时兼顾全局与局部、深层与浅层特征, 且计算高效的端到端参数估计方法.

为有效解决上述问题, 本文提出一种名为SCA-Net (Spectral Convolution-Attention Network)的深度学习模型. SCA-Net并未采用单一的卷积堆叠策略, 而是构建了一种高效的混合架构: 首先, 融合移动翻转瓶颈卷积模块, 利用深度可分离卷积在大幅降低参数数量的同时提取光谱局部特征; 其次, 引入Transformer模块, 利用自注意力机制直接对全局上下文进行建模, 无需通过堆叠深度即可捕捉长程依赖; 最后, 本文特别设计了多尺度特征融合模块(Multi-scale Feature Fusion, MFF), 将不同阶段的特征图进行聚合. MFF能确保模型在做出最终预测时, 既拥有深层网络提取的高级语义信息, 又保留了浅层光谱的细节信息. 通过这种渐进式、多尺度的特征提取策略, SCA-Net实现了从LAMOST原始低分辨率光谱到恒星大气物理参数的高精度、鲁棒且高效的自动估计.

## 2 数据

本文实验数据来自LAMOST-DR10中A、F、G、K恒星的低分辨率( $R \sim 1800$ )实测光谱, 随机抽取信噪比高于25的28913条光谱作为实验数据集, 其中8913条光谱划分为训练集 $D_{\text{train}}$ , 剩余20000条光谱作为测试集 $D_{\text{test}}$ . 数据标签来自LAMOST参数星表, 其值由LASP<sup>[3, 15]</sup>处理得到. 作为标签的3个恒星大气物理参数分别是表面有效温度( $T_{\text{eff}}$ ), 表面重力加速度( $\lg g$ )和金属丰度( $[\text{Fe}/\text{H}]$ ). 训练集 $D_{\text{train}}$ 中的参数覆盖范围分别是:  $[3785.9, 8399.7]$  ( $T_{\text{eff}}/\text{K}$ )、 $[1.01, 4.82]$  ( $\lg [g/(\text{cm} \cdot \text{s}^{-2})]$ )、 $[-2.48, 0.73]$  ( $[\text{Fe}/\text{H}]/\text{dex}$ ), 参数分布如图1所示. 通过图1可观察到, 训练集 $D_{\text{train}}$ 中各参数覆盖范围十分广泛, 这是训练一个可靠且泛化能力强的深度学习模型所必须具备的前置条件. 值得注意的是, 模型可能无法对于那些处于训练集中样本数较少的参数空间区域的恒星做出合理的预测.

本文对实验数据集中的光谱数据进行了一系列预处理, 以保证后续谱线特征提取与分析的准确性. 首先, 由于天体光谱会因宇宙膨胀或自身径向速度而产生整体波长漂移, 因此本文从数据文件中提取红移参数用于校正光谱. 其次, 由于不同观测光谱在波长方向上存在不均匀采样, 且数据点长度不一致, 为构造定长输入, 本文在静止参

考系内构造等间距波长网格(4000–8095 Å, 共 4096 个采样点), 部分样本在此区间存在缺失采样点, 使用 3 次样条法<sup>[16]</sup>(Cubic Spline)进行插值, 对于两端溢出点则进行截断处理. 最后, 本文对光谱进行连续谱归一化处理, 方法与 Lee 等<sup>[17]</sup>提到的相同. 这一预处理流程的效果如图 2 所示.

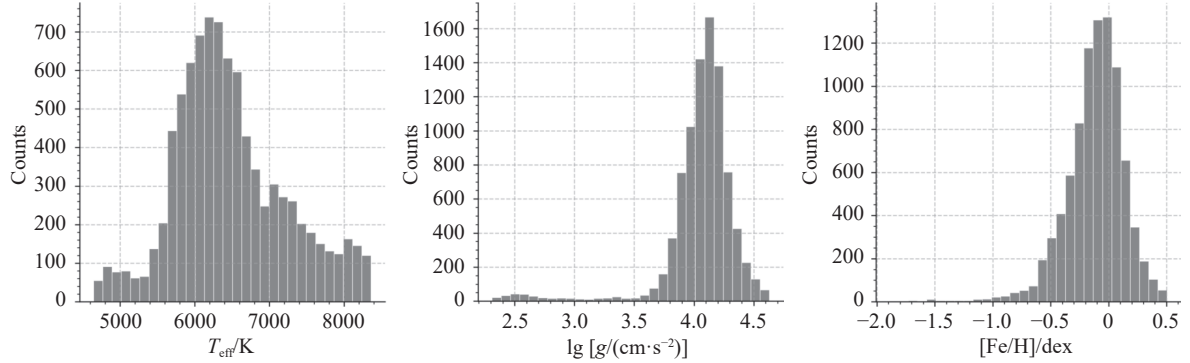


图 1 训练样本的  $T_{\text{eff}}$  (左),  $\lg g$  (中) 与  $[\text{Fe}/\text{H}]$  (右) 分布图

Fig. 1 Distribution of  $T_{\text{eff}}$  (left),  $\lg g$  (middle), and  $[\text{Fe}/\text{H}]$  (right) for the training samples

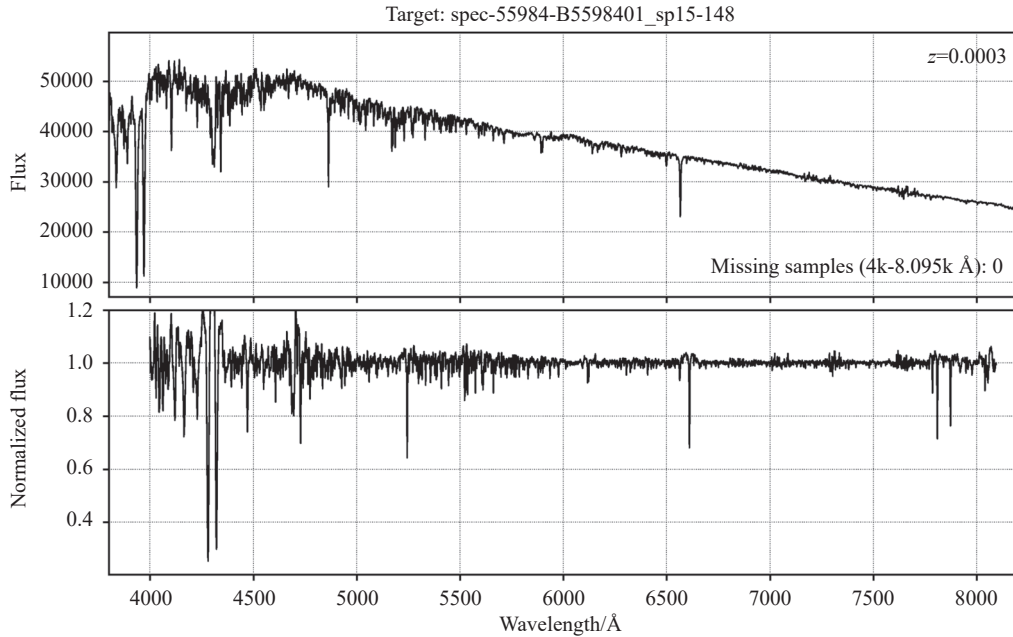


图 2 某一光谱的预处理效果图. 上图为预处理前(红移  $z$ ), 下图为预处理后.

Fig. 2 An example of spectral preprocessing results. The upper panel shows the spectrum before preprocessing (redshift  $z$ ), and the lower panel shows the spectrum after preprocessing.

由于传统1D CNN在处理光谱序列时,感受野和全局依赖之间的平衡较难把握,因此本文将长度为4096的一维光谱按S形折叠的方式转换成二维光谱(64×64). 光谱折叠后,原本在一维光谱中相距较远的波段在二维光谱中能变得相邻,这能很好地给卷积核提供新的局部模式组合. 例如,某些化学成分的多条吸收线虽然在一维上零散分布,但经折叠后能汇聚到同一局部区域, CNN的小尺寸卷积核更容易一次性捕捉到它们之间的关联. 这种维度转换,既保留了光谱连续性(因为折叠是逐行连接的),又强化了行与行之间的谱线共现信息. 二维光谱的示例见图3.

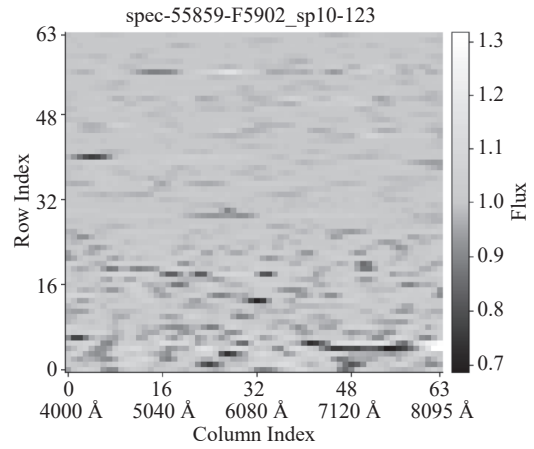


图 3 二维光谱示例图

Fig. 3 Two-dimensional spectrum example diagram

### 3 方法

深度学习模型的性能可概括为模型容量与泛化能力两大核心要素,其中模型容量是指模型的学习能力,而泛化能力则代表模型在未训练数据中的性能. Dai等<sup>[18]</sup>得出结论,卷积模块倾向于具有更好的泛化能力,而注意力模块则具有更高的模型容量. 他们将卷积和注意力机制结合,提出了CoAtNet (Combining Convolution and Attention Networks). 本文受此思想与Sandler等<sup>[19]</sup>提出的轻量化设计理念的启发,提出了SCA-Net (Spectral Convolution-Attention Network)模型.

#### 3.1 SCA-Net结构

本文利用深度学习技术构建的SCA-Net如图4所示. 二维光谱(64×64)在构建完成后会经历从S0到S5的6个处理阶段,空间分辨率逐渐降低,在每个阶段开始时总是将空间尺寸缩小2倍,并增加通道数. 此外,SCA-Net主要包含移动翻转瓶颈卷积(Mobile Inverted Bottleneck Convolution, MBConv)、Transformer和多尺度特征融合3种子模块,详见3.2节论述.

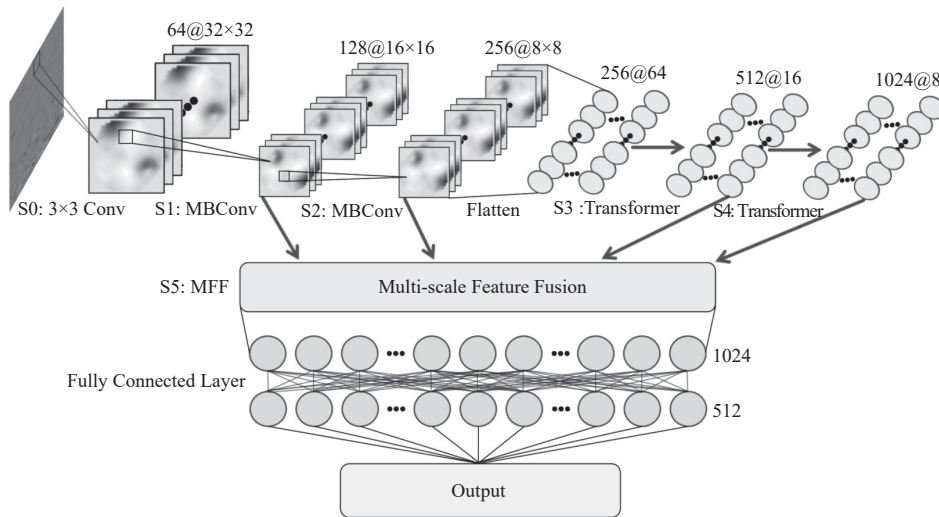


图 4 SCA-Net模型结构图

Fig. 4 SCA-Net model structure diagram

由于Transformer模块的计算复杂度随着空间尺寸呈二次方的增长, 所以S0固定是一个简单的2层卷积, 这样做能保证无论后续是何种变体(见3.3节论述)也不会导致传入Transformer模块的空间尺寸太大. S5固定是一个MFF模块以聚合来自不同网络深度的特征信息. 从S1到S4, 考虑使用MBConv模块或Transformer模块.

二维光谱在经6个阶段处理后会变成长度为1024的特征向量, 而后经两个全连接层输出参数( $T_{\text{eff}}$ 为对数尺寸)预测值.

### 3.2 SCA-Net子模块

MBConv是Sandler等<sup>[19]</sup>在MobileNetV2网络中提出的高效卷积模块, 仲峥迪等<sup>[20]</sup>已将其应用在星系形态分类任务中并取得了较好效果. MBConv通过深度可分离卷积大幅降低参数量, 利用倒置瓶颈和SE通道注意力增强模型表达能力. 其中SE通道注意力是Hu等<sup>[21]</sup>提出的注意力机制, 可帮助模型自适应地校准不同通道的重要性, 进而关注信息量更大的特征并抑制不重要的特征, 冯雪琦等<sup>[22]</sup>已用类似的思路取得了一定的成果. MBConv模块主要包括4个处理步骤: (1)使用 $1\times 1$ 卷积核扩展输入通道; (2)使用深度卷积对每个通道进行空间滤波; (3)使用SE通道注意力获取通道权重; (4)压缩通道至输出尺寸, 并进行残差连接(下采样

时残差路径匹配主路径的尺寸和通道).

Transformer是Vaswani等<sup>[23]</sup>提出的一种深度学习模型, 它摒弃了传统的循环与卷积结构, 完全依赖自注意力机制来捕捉全局上下文间的依赖关系. 值得注意的是, 此处所指的Transformer模块并非是一个独立的网络, 而是被设计的一个计算模块. 该模块主要包括3个处理步骤: (1)使用 $1\times 1$ 卷积核调整输入特征图通道与输出通道相等; (2)被调整后的特征图经多头自注意力层计算; (3)利用全连接层对注意力层的输出进行非线性变换, 并对输出特征图与被调整后的输入特征图进行残差连接.

MFF是本文受Liu等<sup>[24]</sup>SSD网络研究的启发而设计的一个多尺度特征融合模块, 结构如图5所示. 由于网络在S0到S4阶段经历了多次下采样, 深层特征图(如S4)虽然具有丰富的语义信息, 但丢失了大量空间细节. 浅层特征图(如S1)则相反. 本文通过设计MFF模块融合不同层次的特征, 令模型在做出最终参数预测前可以同时获得细节信息和语义信息. 该模块主要包括3个处理步骤: (1)使用 $1\times 1$ 卷积核与上采样技巧调整各阶段特征图尺寸大小与通道数皆一致; (2)沿通道维度拼接4组特征图并使用 $3\times 3$ 卷积核融合特征; (3)使用全局平均池化将空间尺寸压缩为 $1\times 1$ , 保留通道维度, 去除空间维度, 展平为一个向量.

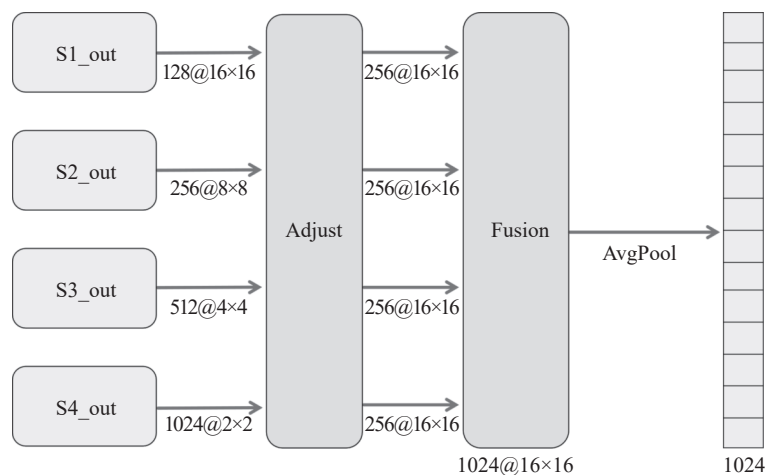


图5 MFF模块结构图

Fig. 5 MFF module structure diagram

### 3.3 SCA-Net变体

通过图4不难看出, SCA-Net中6个阶段按顺序依次工作. 由于S1到S4阶段中, 每个阶段都可以选择MBCConv或Transformer模块, 因此网络具有了不同的变体形式. 为了充分发挥卷积在处理早期特征图时的优势, 本文在网络结构层面施加一条基于先验知识的约束: MBCConv模块必须出现在Transformer模块之前. 这意味着一旦某个阶段(例如S<sub>i</sub>)使用了Transformer模块, 则其之后的所有阶段(S<sub>i+1</sub>, ..., S<sub>4</sub>)也必须是Transformer模块. 本文将MBCConv模块记作M, Transformer模块记作T, 结合约束条件, 自然而然地引出了以下5种网络变体, 其中T模块逐渐增多:

(1) M-M-M-M, 一种纯卷积结构, 其完全依赖MBCConv模块进行特征提取, 理论上具有最强的泛化能力与最低的模型容量.

(2) M-M-M-T, 一种在网络后期引入全局注意力的结构, 利用Transformer模块捕捉深层语义信息, 整体泛化能力较强, 模型容量较M-M-M-M而言有所提升.

(3) M-M-T-T, 一种均衡混合卷积和全局注意力的结构, 前半部分使用MBCConv模块提取局部特征, 后半部分使用Transformer模块整合全局信息, 具有折中的泛化能力与模型容量.

(4) M-T-T-T, 一种在网络早期便引入全局注意力的结构, 仅在S0和S1阶段使用卷积进行初步的特征提取, 之后完全依赖Transformer模块进行特征学习, 整体泛化能力较弱, 模型容量较高.

(5) T-T-T-T, 一种纯注意力结构, 除S0阶段外完全依赖Transformer模块进行特征学习, 理论上具有最弱的泛化能力与最高的模型容量.

本文在4.3节中对这5种变体做了对比实验, 最终选择M-M-T-T作为SCA-Net最终结构. 这一结果恰好与Dai等<sup>[18]</sup>在CoAtNet网络中得出的结论相似: 卷积模块越多, 泛化能力越强; 注意力模块越多, 模型容量越高. 对于大多数任务而言, 采取折中方案的模型表现会更优.

## 4 实验

### 4.1 实验环境

本文所提出的SCA-Net基于python语言和

pytorch深度学习框架实现. 硬件设备为: 2个型号为Intel (R) Xeon (R) Gold 6226R CPU @2.90 GHz的CPU, 2个型号为NVIDIA GeForce RTX 2080 Ti的GPU和4个32 GB的DDR4内存条.

### 4.2 评价指标

为准确评估模型性能, 引入平均绝对误差(Mean Absolute Error, MAE)和均方根误差(Root Mean Square Error, RMSE)作为模型预测性能的评价指标. MAE表示预测值和实际值之间绝对残差的平均值, 直观反映模型预测偏差; MSE表示预测值与实际值之间残差的标准差, 在一定程度上反映模型预测偏差的离散程度, 对较大的预测残差更敏感. 上述评价指标的数学表达式为:

$$MAE(X, h) = \frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i|,$$

$$RMSE(X, h) = \sqrt{\frac{1}{n} \sum_{i=1}^n [h(x_i) - y_i]^2}.$$

式中,  $n$ 为样本个数,  $h(x_i)$ 为模型预测值,  $y_i$ 为样本实际观测值.

### 4.3 实验结果

本文选用AdamW优化器, 设定初始学习率为0.0001, 批处理规模batch大小为32, 损失函数为MAE损失函数, 在训练集 $D_{\text{train}}$ 上对SCA-Net不同变体分别进行了80个周期的训练, 训练时按10%的比例划分验证集, 模型训练与验证损失下降情况如图6所示. 为直观显示变化趋势, 图6中从第1个epoch开始, 每隔10个epoch添加间隔标记点. 可观察到: (1)纯Transformer结构(T-T-T-T)验证损失波动最大, 说明其对训练数据最敏感, 泛化性较差; (2) M-T-T-T结构在训练前期收敛最快, 但验证损失仍有大型波动现象; (3) M-M-T-T结构前期收敛速度并未达到最快, 但在后期验证损失最为稳定; (4)在第50周期后各网络变体均趋于稳定, 未见明显过拟合现象(验证损失持续上升).

本文对每种变体重复训练了5次, 利用测试集 $D_{\text{test}}$ 中的20000条实测光谱对5种变体模型进行了评估, 将所得结果以均值±标准差形式汇总至表1.

表中结果显示, 纯Transformer架构(T-T-T-T)在各项指标上均明显劣于其余变体, 且标准差波动较大, 反映出该架构在本文数据集上泛化能力较弱、稳定性不足; 相对而言, 混合架构变体与纯卷积架构变体则表现出更好的平均性能与更小的结果波动. 在 $T_{\text{eff}}$ 的MAE统计均值层面, M-M-T-T取得了最佳均值, 较第2名M-M-M-M变体优约4.55 K; 而在 $\lg g$ 与 $[\text{Fe}/\text{H}]$ 指标上, 多数变体的指标均值相互接近, 差异主要体现在标准差层面. 本文还评估了测试集中数据标签本身的不确定度(为排除异常值影响, 对标签误差做了Z-Score过滤), 所得标签误差的均值为: 21.439 K ( $T_{\text{eff}}$ )、0.0285 dex ( $\lg g$ )、0.0156 dex ( $[\text{Fe}/\text{H}]$ ). 分析可知, 混合结构与纯卷积结构变体间的指标差异小于测试标签所固有的不确定度. 因此, 尽管M-M-T-T变体在指标统计上具有可测的优势, 但该提升幅度小于数据

本身的不确定度, 改进效果在数据层面受限.

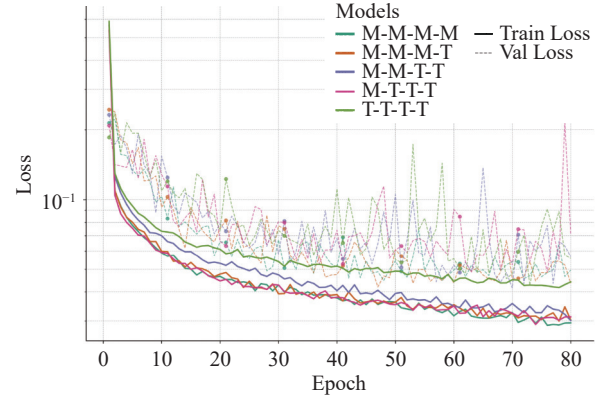


图 6 SCA-Net不同变体的训练与验证损失比较图(每隔10个epoch添加间隔标记点), 所有变体均拥有相似的数量.

Fig. 6 Comparison of training and validation (val) loss for different variants of SCA-Net (data points at 10-epoch intervals), all variants have a similar parameter count.

表 1 SCA-Net不同变体的MAE和RMSE统计表  
Table 1 Statistical table of MAE and RMSE for different variants of SCA-Net

Method	$T_{\text{eff}}/\text{K}$		$\lg [g/(\text{cm} \cdot \text{s}^{-2})]$		$[\text{Fe}/\text{H}]/\text{dex}$	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
M-M-M-M	75.62 $\pm$ 4.66	112.90 $\pm$ 3.86	0.042 $\pm$ 0.005	0.075 $\pm$ 0.004	0.041 $\pm$ 0.007	0.067 $\pm$ 0.006
M-M-M-T	81.04 $\pm$ 17.12	126.77 $\pm$ 17.08	0.042 $\pm$ 0.003	0.074 $\pm$ 0.006	<b>0.039 <math>\pm</math> 0.004</b>	0.068 $\pm$ 0.004
M-M-T-T	<b>71.07 <math>\pm</math> 9.43</b>	<b>111.84 <math>\pm</math> 11.98</b>	<b>0.042 <math>\pm</math> 0.001</b>	<b>0.074 <math>\pm</math> 0.002</b>	0.040 $\pm$ 0.006	<b>0.067 <math>\pm</math> 0.006</b>
M-T-T-T	78.07 $\pm$ 10.43	119.55 $\pm$ 15.95	0.045 $\pm$ 0.004	0.077 $\pm$ 0.003	0.045 $\pm$ 0.013	0.072 $\pm$ 0.012
T-T-T-T	127.48 $\pm$ 70.42	174.52 $\pm$ 91.26	0.069 $\pm$ 0.035	0.107 $\pm$ 0.047	0.065 $\pm$ 0.036	0.096 $\pm$ 0.037

为研究模型效率问题, 本文将SCA-Net模型中的MBConv模块替换成标准残差瓶颈块(使用标准卷积), 并对比替换前后模型的参数量、浮点运算数(FLOPs)及对单条光谱的实际推理耗时(包含数据加载与归一化过程)等指标, 论证MBConv模块对模型效率的提升, 所得结果见表2. 实验表明, MBConv模块的引入最高可将模型参数量降低至标准卷积版本的7.8%, 并将模型的实际推理耗时降低2.5倍以上.

MBConv的核心优势在于利用深度可分离卷积实现参数的线性缩放, 这一优势在模型的高维通道阶段尤为显著. 如表2所示, 在M-M-M-M变体

中, 使用标准卷积的模型参数量高达789.196M, 相比之下, 使用MBConv的SCA-Net相同变体的参数量仅为61.867M. 在将S4阶段替换为Transformer模块时, 标准卷积模型(Standard)的参数量由789M减少至628M, 总FLOPs由57.826G降至57.085G; 而SCA-Net的参数量由61M增加至83M, 总FLOPs则由4G增至7G. 这表明在高通道(1024)配置下, Transformer模块的参数复杂度与矩阵乘法开销介于标准残差瓶颈块和MBConv模块之间. 在对SCA-Net的S3、S2及S1阶段依次进行T模块的替换时, 出现了与预期不符的现象: 随着Transformer模块的增多, 模型参数量持续上升,

但总FLOPs反呈下降趋势. 与传统高分辨率图像任务(如 $224 \times 224$ )不同, 本文输入尺寸为 $64 \times 64$ , 经S0阶段下采样, 首个可含Transformer模块的S1阶段特征图分辨率仅为 $32 \times 32$ . 这意味着在浅层阶段, 序列长度未达到自注意力机制的计算瓶颈. 相反,

MBCConv模块在浅层需要对较大的空间特征图进行密集的滑窗卷积运算, 因而产生较高的FLOPs. 这一发现表明, 对于低分辨率光谱处理任务而言, 全Transformer架构或前置Transformer架构在推理速度上具有额外优势.

表 2 不同模型的参数和延迟比较  
Table 2 Comparison of parameters and latency for different models

Model	Params	FLOPs	Avg latency per sample
M-M-M-M (MBCConv)	61.867M	4.738G	1.779 ms
M-M-M-M (Standard)	789.196M	57.826G	4.660 ms
M-M-M-T (MBCConv)	83.319M	7.013G	2.231 ms
M-M-M-T (Standard)	628.581M	57.085G	4.769 ms
M-M-T-T (MBCConv)	114.239M	4.523G	1.840 ms
M-M-T-T (Standard)	166.526M	22.609G	2.664 ms
M-T-T-T (MBCConv)	118.738M	1.876G	1.418 ms
M-T-T-T (Standard)	122.218M	7.294G	1.648 ms
T-T-T-T	119.555M	0.967G	1.334 ms

经实验对比, 本文权衡模型精度与推理速度后选用M-M-T-T作为SCA-Net最终结构, 其在 $D_{\text{test}}$ 中预测值与实际值的对比情况与残差分布情况见图7. 从总体上看, 左侧散点图高密度区沿对角线分布, 表明模型预测参数与LAMOST官方参数之间存在强线性关系, 模型已能准确建模3个大气物理参数分布的主趋势. 右侧KDE曲线在零点附近对称聚集, 表明模型无严重的系统性偏差, 但非零的正偏说明模型依旧存在细微的高估现象. 虽然KDE曲线总体尖锐且在两侧迅速衰减, 但仍有一定厚尾现象(对应左侧散点图中离群点), 这意味着虽然绝大多数预测误差集中在小幅度范围, 但仍有少量显著偏离, 不过在低分辨率的光谱下, 这种量级的不确定性通常是可接受的.

为评估光谱不同波长区间对模型参数预测的贡献, 本研究采用一种基于局部信息扰动的分析方法. 具体而言, 首先将完整光谱输入已训练模型, 获取基准预测结果. 随后, 设置一个宽度为 $15 \text{ \AA}$ 的滑动窗口, 以 $5 \text{ \AA}$ 为步长, 在 $4000\text{--}8095 \text{ \AA}$ 范围内依次移动. 在每一窗口位置, 将区间内的通量替换为

连续谱对应值, 以模拟该波长范围内吸收线信息的缺失. 每次扰动后, 重新输入模型得到参数预测值, 并计算其与基准值的绝对偏差, 作为该光谱区间信息贡献的误差度量. 为减小单一样本随机性的影响, 本方法在测试集全部光谱上重复上述过程, 并取各窗口位置偏差的平均值. 最终结果如图8所示.

在有效温度( $T_{\text{eff}}$ )预测中, 模型主要依赖于 $\text{H}\beta$ 、 $\text{H}\gamma$ 以及 $4307.9 \text{ \AA}$ 处的Fe I线. 氢的巴耳末线对温度变化高度敏感, 是测定 $T_{\text{eff}}$ 的关键指标<sup>[25]</sup>. 有效温度直接决定恒星内部辐射场的强度, 对于一条来自低激发态的Fe I吸收线(如 $4307.9 \text{ \AA}$ ), 当 $T_{\text{eff}}$ 升高时, 其下能级的粒子数会相对减少, 导致吸收线强度变弱, 这种相关性可辅助约束有效温度<sup>[26]</sup>. 此外,  $4300\text{--}4310 \text{ \AA}$ 区域亦受CH分子解离影响, 进一步增强了其对温度变化的敏感性<sup>[17, 27]</sup>. 模型对 $\text{H}\beta$ 与 $\text{H}\gamma$ 的依赖程度明显高于 $\text{H}\alpha$ 与 $\text{H}\delta$ , 这一现象反映了模型在面对非优化目标干扰时所采取的一种鲁棒学习策略.  $\text{H}\alpha$ 线心易受色球活动填充, 其作为色球活动敏感指标的特性会为光球层参数估

计引入噪声<sup>[28]</sup>. H $\delta$ 则受周边金属线(尤其是CN带)污染显著,在晚型星中受金属丰度与元素比例

(如[N/Fe])影响较大,导致其作为氢线指标的稳定性降低<sup>[29]</sup>.

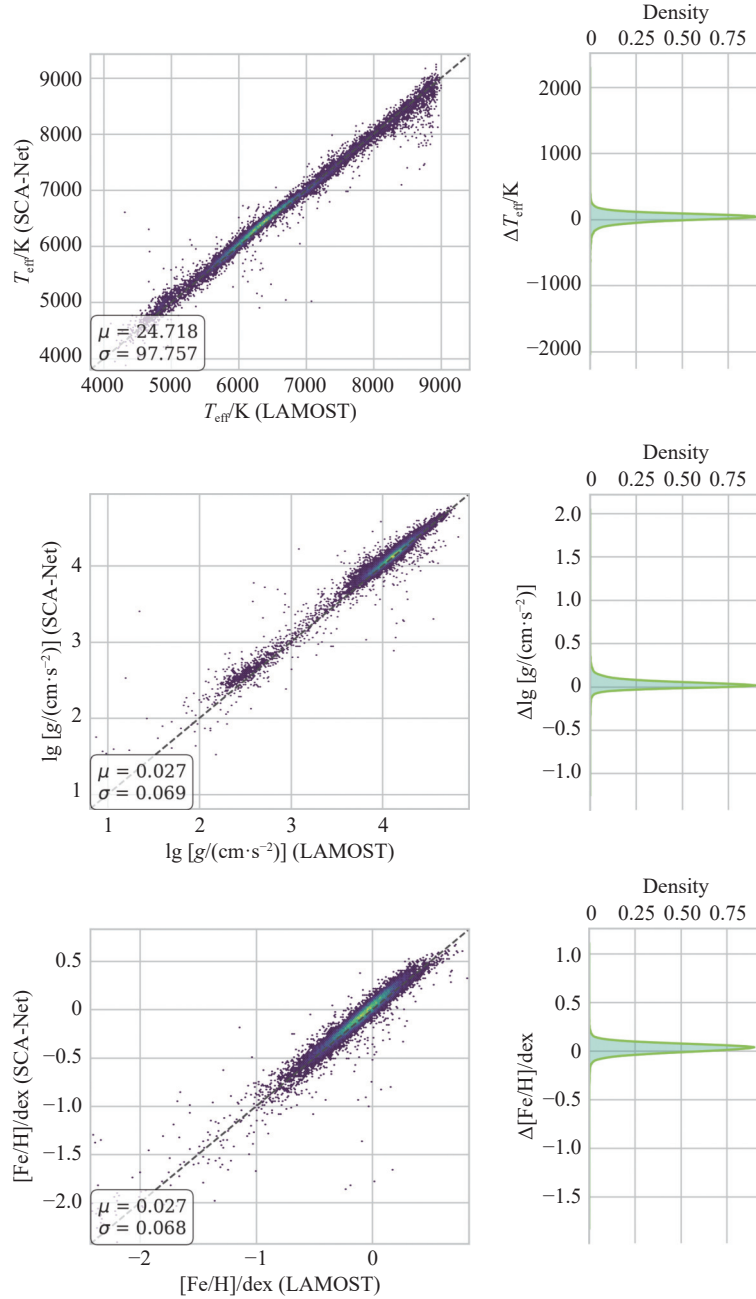


图 7 SCA-Net预测值与实际值对比和残差分布: (上)  $T_{\text{eff}}$ ; (中)  $\lg g$ ; (下)  $[\text{Fe}/\text{H}]$ . 左侧散点表示数据分布, 横轴为LAMOST官方参数, 纵轴为SCA-Net预测的参数,  $\mu$ 和 $\sigma$ 分别表示残差分布的均值和标准差. 右侧KDE曲线表示残差分布形态.

Fig. 7 Comparison of SCA-Net predicted values with actual values and residual distribution: (Top)  $T_{\text{eff}}$ ; (Middle)  $\lg g$ ; (Bottom)  $[\text{Fe}/\text{H}]$ . The scatter plots on the left represent the data distribution, with the horizontal axis showing the official LAMOST parameters and the vertical axis showing the parameters predicted by SCA-Net, where  $\mu$  and  $\sigma$  denote the mean and standard deviation of the residual distribution. The KDE curves on the right illustrate the shape of the residual distribution.

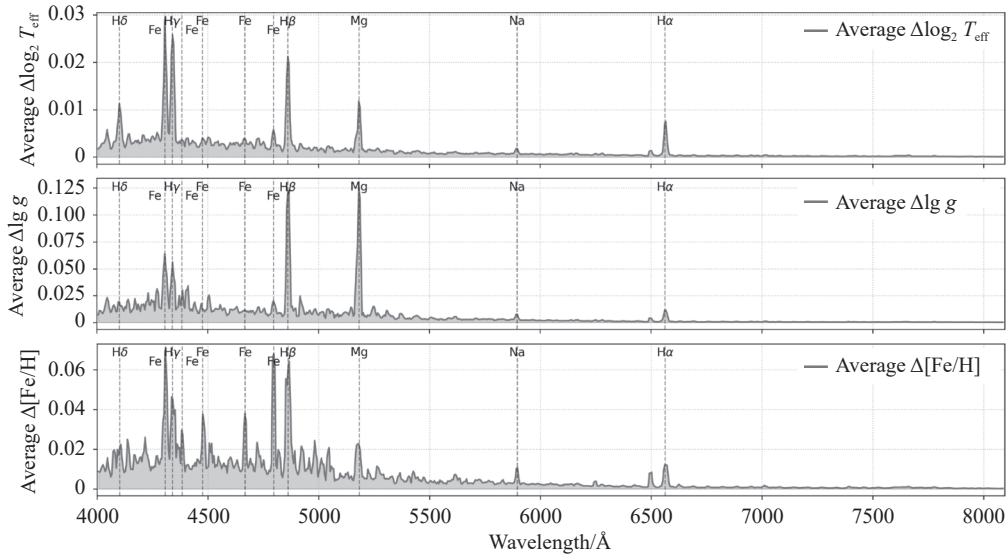


图 8 利用  $D_{\text{train}}$  训练的 SCA-Net 模型预测参数时对不同波长区域的依赖分布。横轴为窗口中心波长，纵轴为损失窗口内吸收线信息后在测试集上的平均参数偏差。

Fig. 8 Dependence distribution of the SCA-Net model trained on  $D_{\text{train}}$  in different spectral regions for parameter prediction. The horizontal axis represents the window center wavelength, and the vertical axis shows the mean parameter deviation on the test set after excluding spectral line information within the loss window.

在表面重力( $\lg g$ )预测中,模型主要依赖Mg I b线及H $\beta$ 线,其次是H $\gamma$ 线附近区域. Mg I b线的压力致宽翼是测定恒星表面重力的灵敏指示器<sup>[30]</sup>,而巴耳末线的轮廓对恒星大气中的压力敏感,而压力直接与 $\lg g$ 相关<sup>[31]</sup>.

在金属丰度( $[\text{Fe}/\text{H}]$ )预测中,模型对多个Fe I线、H $\beta$ 与H $\gamma$ 线表现出强依赖,并对Mg I b、Na以及H $\delta$ 等线表现出次要依赖. 对Fe I线而言,其强度直接反映了铁丰度,因此模型在预测参数时对其依赖显著. 氢的能级跃迁并不直接依赖金属丰度,因此H $\beta$ 、H $\gamma$ 等氢线本身不直接包含金属丰度信息,但 $[\text{Fe}/\text{H}]$ 的推导一般基于Fe I和Fe II谱线的等值宽度,并通过强制满足激发平衡和电离平衡条件来确定<sup>[32-34]</sup>. 而激发平衡条件对 $T_{\text{eff}}$ 极其敏感,若 $T_{\text{eff}}$ 不准确,不同激发能的谱线会给出系统性的、有斜率的不同丰度结果<sup>[35]</sup>. 因此,模型需借助这些氢线得到有关 $T_{\text{eff}}$ 的信息,以避免 $[\text{Fe}/\text{H}]$ 预测产生系统误差. 此外,Na与Mg元素丰度通常与Fe丰度正相关,模型可借助此相关性利用相应谱线辅助推断 $[\text{Fe}/\text{H}]$ .

本文将SCA-Net与GSN<sup>[10]</sup>和DCN<sup>[12]</sup>进行了对

比,所得结果见表3. 对比实验表明,本文所提出的SCA-Net在3种恒星大气物理参数的估计上相比其他方法均有所改善.

表 3 3种方法的MAE和RMSE统计表  
Table 3 Statistics table of MAE and RMSE for three methods

Method	$T_{\text{eff}}/\text{K}$	$\lg [g/(\text{cm} \cdot \text{s}^{-2})]$	$[\text{Fe}/\text{H}]/\text{dex}$
GSN	80.00	0.14	0.07
DCN	97.21	0.28	0.13
Ours	<b>71.07</b>	<b>0.04</b>	<b>0.04</b>

## 5 讨论

机器学习和深度学习技术为天文大数据的处理提供了强有力的工具. 在本文所使用的算力设备下(见4.1节), SCA-Net在单日内可处理约6500万条光谱,这是传统方法未曾有过的速度. 即使与数据驱动方法The Cannon 2<sup>[36]</sup>相比,后者处理相同数量的光谱,也需在小型计算集群上运行约9 d. 这种快速性能在大规模光谱巡天时代具有显著优

势. 在算法层面, 本文抛弃传统的卷积堆叠策略, 设计了卷积与注意力机制的混合架构, 以增强模型对光谱长程依赖(如连续谱形状)的捕捉能力. 此外, 本文所设计的MFF模块能够确保模型在最终预测前拥有浅层细节信息(如弱金属吸收线)以供参考.

然而, SCA-Net的预测能力受限于训练集的参数空间覆盖范围, 对超出该范围的参数预测通常不可靠. 一种解决方案是生成理论光谱, 在理论光谱中进行训练, 以此来使训练集覆盖所有可能出现的参数组合<sup>[8, 37]</sup>. 然而, 理论光谱与实测光谱在特征空间中未必属于同一分布, 这源于观测过程引入的各种效应以及理论建模中未完全涵盖的物理过程. 因此, 未来需发展能够有效融合理论光谱与实测光谱、并为数据驱动模型所用的方法.

SCA-Net不同变体的性能虽存在统计上可测的差异, 但多数变体之间的性能差异小于标签本身的不确定性, 因此该层面的改进在数据层面受限. 在具体变体方面, M-M-M-M表现出较好的稳定性, M-M-T-T在统计指标上更优, M-T-T-T具备较快的推理速度与居中的性能, 而T-T-T-T虽在本研究中性能最低, 但其推理速度最快. 值得注意的是, 尽管在本研究中T-T-T-T变体性能表现最差, 但这种现象可能与数据体量有关, 已有研究证明基于Transformer的模型在足够大规模的数据与预训练条件下具有超越卷积网络的潜力<sup>[38]</sup>. 因此, 尽管本文将M-M-T-T定为SCA-Net的最终架构, 但实际部署时仍建议根据具体任务需求与数据条件进行选择.

此外, 现阶段SCA-Net的性能受限于LASP的系统性偏差, 但本文所构建并验证的这套面向海量LAMOST低分辨率光谱的端到端参数估计流程, 为后续优化奠定了基础. LASP作为一个成熟的物理模型驱动管道, 其产出的参数星表为本文在方法开发阶段提供了一个稳定、易获取的基准标签源. 然而, LASP流程涉及多阶段、多算法的复杂操作, 包括模板匹配、 $\chi^2$ 最小化、多项式拟合、高斯卷积等. 这是一个迭代式、搜索式的优化过程, 计算成本高昂, 通常难以满足对未来数千万光谱的实时处理需求. 尽管当前阶段SCA-Net受

限于LASP, 但在实际应用部署中, 可结合利用高分辨率巡天数据(如APOGEE)对LAMOST共同观测的恒星进行定标, 生成更高质量的训练样本, 从而突破当前LASP系统偏差的限制, 这也是本文下一步的研究工作之一.

## 6 结论

本文针对LAMOST海量低分辨率巡天光谱在恒星大气参数估计上的实际需求, 提出并验证了SCA-Net. 本文在方法上的贡献包括: (1)设计了由高效卷积(MBConv)与自注意力模块结合的混合架构, 通过先卷积后注意力的结构约束(最终选择M-M-T-T变体), 在控制模型复杂度的同时对光谱中的长程依赖与局部细节进行了有效建模; (2)设计了MFF模块, 以缓解网络下采样过程中浅层细节信息的丢失, 确保模型在预测参数前拥有浅层细节信息参考; (3)将一维光谱折叠为二维表示, 增强卷积操作对谱线共现模式的捕获; (4)网络采用轻量化的MBConv模块, 显著降低了参数量与计算成本, 使其具备处理海量光谱数据的工程能力(单日可处理约6500万条光谱).

在实证评估中, SCA-Net在信噪比高于25的LAMOST-DR10测试集上取得了较高的性能, 其(M-M-T-T)在表面有效温度、表面重力加速度和金属丰度上的平均MAE精度分别为: 71.07 ( $T_{\text{eff}}/\text{K}$ )、0.042 ( $\lg [g/(\text{cm} \cdot \text{s}^{-2})]$ )、0.040 dex ( $[\text{Fe}/\text{H}]$ ), 平均RMSE精度分别为: 111.84 ( $T_{\text{eff}}/\text{K}$ )、0.074 ( $\lg [g/(\text{cm} \cdot \text{s}^{-2})]$ )、0.067 dex ( $[\text{Fe}/\text{H}]$ ). 此外, 基于局部扰动的波长重要性分析表明, 模型显著依赖的谱线区域(如巴耳末线、Mg I b线、Fe I线等)与天体物理先验一致.

综上, SCA-Net通过高效卷积、注意力机制与多尺度特征融合的协同设计, 为大规模光谱巡天的快速、精准参数估计提供了有效的解决方案.

**致谢** 感谢审稿人对文章提出的宝贵建议, 使得文章变得更加严谨. 感谢德州学院FAST人工智能实验室提供的实验条件. 感谢中国科学院国家天文台王瑞高级工程师、德州学院王丽丽教授和

秦丽博士在稿件修改阶段参与的讨论.

### 参 考 文 献

- [1] Zhao G, Zhao Y H, Chu Y Q, et al. *RAA*, 2012, 12: 723
- [2] Cui X Q, Zhao Y H, Chu Y Q, et al. *RAA*, 2012, 12: 1197
- [3] Luo A L, Zhao Y H, Zhao G, et al. *RAA*, 2015, 15: 1095
- [4] Xiang M S, Liu X W, Yuan H B, et al. *MNRAS*, 2015, 448: 822
- [5] Si J M, Luo A L, Wu F Z, et al. *Spectroscopy and Spectral Analysis*, 2015, 35: 834
- [6] Ness M, Hogg D W, Rix H W, et al. *ApJ*, 2015, 808: 16
- [7] Leung H W, Bovy J. *MNRAS*, 2018, 483: 3255
- [8] Fabbro S, Venn K A, O'Briain T, et al. *MNRAS*, 2018, 475: 2978
- [9] Candebat N, Sacco G G, Magrini L, et al. *A&A*, 2024, 692: A228
- [10] Wang R, Luo A L, Zhang S, et al. *PASP*, 2019, 131: 024505
- [11] Wang R, Luo A L, Chen J J, et al. *ApJ*, 2020, 891: 23
- [12] 邓诗宇, 刘承志, 康喆, 等. *科学技术与工程*, 2021, 21: 5223
- [13] Xiang M S, Ting Y S, Rix H W, et al. *ApJS*, 2019, 245: 34
- [14] Zhang M, Xiang M S, Ting Y S, et al. *ApJS*, 2025, 279: 5
- [15] Wu Y, Du B, Luo A, et al. *Proceedings of the International Astronomical Union*. Cambridge: Cambridge University Press, 2014, 306: 340–342
- [16] Reinsch C. *Numerische Mathematik*, 1967, 10: 177
- [17] Lee Y S, Beers T C, Sivarani T, et al. *AJ*, 2008, 136: 2022
- [18] Dai Z, Liu H X, Le Q V, et al. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc, 2021: 303
- [19] Sandler M, Howard A, Zhu M, et al. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2018: 4510
- [20] 仲峥迪, 屠良平, 冯雪琦, 等. *天文学报*, 2024, 65: 16
- [21] Hu J, Shen L, Sun G. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 7132–7141
- [22] 冯雪琦, 屠良平, 仲峥迪, 等. *天文学报*, 2024, 65: 17
- [23] Vaswani A, Shazeer N, Parmar N, et al. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc, 2017: 6000–6010
- [24] Liu W, Anguelov D, Erhan D, et al. *SSD: Single Shot MultiBox Detector* // Leibe B, Matas J, Sebe N, et al. *Computer Vision - ECCV 2016*. Cham: Springer, 2016: 21–37
- [25] Barklem P S, Piskunov N, O'Mara B J. *A&A*, 2000, 363: 1091
- [26] Mashonkina L, Sitnova T, Yakovleva S A, et al. *A&A*, 2019, 631: A43
- [27] Schuler S C, Margheim S J, Sivarani T, et al. *BAAS*, 2009, 41: 199
- [28] He H. *NatSR*, 2024, 14: 17962
- [29] Worthey G, Ottaviani D L. *ApJS*, 1997, 111: 377
- [30] Brewer J M, Fischer D A, Valenti J A, et al. *ApJS*, 2016, 225: 32
- [31] Catanzaro G, Leone F, Dall T H. *A&A*, 2004, 425: 641
- [32] Santos N C, Israelian G, Mayor M. *A&A*, 2004, 415: 1153
- [33] Sousa S G, Santos N C, Israelian G, et al. *A&A*, 2011, 533: A141
- [34] Sousa S G, Santos N C, Israelian G, et al. *A&A*, 2006, 458: 873
- [35] Sousa S G, Santos N C, Mayor M, et al. *A&A*, 2008, 487: 373
- [36] Casey A R, Hogg D W, Ness M, et al. 2016, arXiv: 1603.03040
- [37] Ting Y S, Conroy C, Rix H W, et al. *ApJ*, 2019, 879: 69
- [38] Dosovitskiy A, Beyer L, Kolesnikov A, et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. *International Conference on Learning Representations*, 2020

# SCA-Net: Automatic Estimation of Stellar Atmospheric Physical Parameters Based on Efficient Convolution and Attention Mechanisms

ZHANG Hao-kai    WANG Hong-feng    TAO Zhen-zhao    MA Shuo    ZHANG Hou-shui

*(School of Computer and Information, Dezhou University, Dezhou 253023)*

**ABSTRACT** Estimating stellar atmospheric physical parameters (effective temperature  $T_{\text{eff}}$ , surface gravity  $\lg g$ , and metallicity  $[\text{Fe}/\text{H}]$ ) is a primary task in stellar studies. To meet the demand for analyzing massive stellar spectral data, this study proposes the SCA-Net (Spectral Convolution-Attention Network) model based on publicly available low-resolution observed spectral data from LAMOST DR10 (Large Sky Area Multi-Object Fiber Spectroscopy Telescope Data Release 10). SCA-Net integrates mobile inverted bottleneck convolution, self-attention mechanisms, and multi-scale feature fusion to achieve precise estimation of stellar atmospheric physical parameters. Experiments were conducted on 28913 low-resolution observed spectra, with 8913 used for model training and 20000 for testing. The evaluation metric, Mean Absolute Error (MAE), yielded the following accuracies for the three stellar atmospheric physical parameters: 71.07 K ( $T_{\text{eff}}$ ), 0.042 ( $\lg [g/(\text{cm} \cdot \text{s}^{-2})]$ ), and 0.040 dex ( $[\text{Fe}/\text{H}]$ ). The code of this study is available at: <https://github.com/HelKai528/SCA-Net>.

**Key words** stars: fundamental parameters, stars: atmospheres, methods: data analysis, methods: statistical, techniques: spectroscopic