

## SPFusion: A Lightweight Multi-Modality Deep Learning Model for Parameter Estimation of Red Clump Stars

2 YUCHEN HE <sup>†</sup> , <sup>1,2,3</sup> JINGJING WU <sup>†</sup> , <sup>1,2</sup> BIN JIANG  AND <sup>3</sup> YANXIA ZHANG 

3 <sup>1</sup> Shenzhen Research Institute of Shandong University, Shenzhen, 518057, Guangdong, China

4 <sup>2</sup> School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, 264209, Shandong, China

5 <sup>3</sup> CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, 100101, Beijing, China

### 6 ABSTRACT

7 Red Clump stars, well known for their moderate luminosity and red colors, are widely utilized in  
8 studies of galactic evolution, extinction, and stellar distribution. Therefore, rapid and accurate es-  
9 timation of their physical parameters is essential. Numerous models for parameter estimation have  
10 been developed, using either photometric images or spectral data. However, relying on a single data  
11 modality often limits accuracy due to insufficient feature information. To address the issues above,  
12 we propose SPFusion, a lightweight, end-to-end, and multi-modality parameter estimation model de-  
13 signed to combine feature-sparse photometric images with feature-rich spectral data, thereby impro-  
14 ving estimation accuracy. Its efficient architecture reduces computational costs and enhances infer-  
15 ence speed, making it particularly suitable for large-scale, multi-source survey tasks. Comparative  
16 experiments demonstrate that SPFusion outperforms baseline models, achieving the Mean Absolute  
17 Errors (MAE) of 0.0640 and 0.7456 for mass and age parameters estimation, respectively. The ab-  
18 blation experiments validate the effectiveness of the structural design. Furthermore, we conduct ex-  
19 ploratory analyses to provide insights into the interpretability of SPFusion. The code is available at  
20 <https://github.com/qintianjian-lab/SPFusion>.

21 *Keywords:* Neural networks (1933) — Astronomy data analysis (1858) — Red giant clump (1370)

### 22 1. INTRODUCTION

23 Red Clump (RC) stars are low-mass, metal-rich stars  
24 with helium-burning cores (Cassisi & Salaris 1997), and  
25 they are widely distributed across the Magellanic Clouds  
26 and the Galactic disk (Girardi 2016). Due to their dis-  
27 tinctive characteristics, such as stable luminosity, consis-  
28 tent color and extensive spatial distribution (Wan et al.  
29 2015), RC stars are frequently used as standard candles  
30 in numerous studies aimed at probing galactic evolu-  
31 tion, extinctions, kinematics, distances, and spatial dis-  
32 tributions (Hatzidimitriou 1991; Stanek 1996; Wozniak  
33 & Stanek 1996; Mao & Paczyński 2002; Sumi et al. 2003;  
34 Nataf et al. 2013; De Marchi et al. 2013). To explore the  
35 tasks above in greater depth via RC stars, it is essential  
36 to investigate its accurate physical parameters, such as

37 the mass and age estimated in He et al. (2022) by using  
38 the spectra from the Large Sky Area Multi-Object Fiber  
39 Spectroscopic Telescope (LAMOST) data release.

40 Traditionally, some studies employed modeling or  
41 fitting methods based on astrophysical principles  
42 (Casamiquela et al. 2017; Li et al. 2015; Hawkins et al.  
43 2017) to analyze various physical parameters of celestial  
44 objects with high accuracy. However, due to the vast  
45 photometric and spectral data produced by large-scale  
46 sky surveys, traditional methods often need to be revised  
47 to analyze big data rapidly. Consequently, many studies  
48 shifted to machine-learning-based methods (Zhang et al.  
49 2020; Xiang et al. 2021), end-to-end estimating the phys-  
50 ical parameters of specific celestial objects from spectral  
51 data. For example, Xiang et al. (2016) applied a kernel-  
52 based principal component analysis method to estimate  
53 stellar atmospheric parameters, and Bu & Pan (2014)  
54 used Gaussian process regression for the same task. In  
55 recent years, deep learning methods based on neural  
56 networks have demonstrated their ability to achieve ex-  
57 ceptionally high accuracy in complex astronomy tasks

<sup>1</sup> <sup>†</sup> These authors contributed equally to this work.

Corresponding author: Bin Jiang, Yanxia Zhang

jiangbin@sdu.edu.cn, zyx@bao.ac.cn

while offering rapid inference capabilities. These advantages make them well-suited to vast astronomical data analysis demands (Wu et al. 2024). For instance, the autoencoder model proposed by Yang & Li (2015) and the StarGRUNet model proposed by Li & Lin (2023) were employed to estimate stellar atmospheric parameters. Wu et al. (2018) developed a classifier based on deep learning model for galaxy morphological classification. Additionally, some researches (Pasquet-Itam, J. & Pasquet, J. 2018; Pasquet, Johanna et al. 2019; Henghes et al. 2022) have used neural networks to estimate the redshifts of galaxies and quasars.

To further improve the accuracy of neural networks in astronomical tasks, multi-modality models that integrate data from diverse modalities have been introduced into relevant studies. For example, Parker et al. (2024) and Rizhko & Bloom (2024) constructed CLIP-75 based (Radford et al. 2021) multi-modality models for galaxy morphology tasks and celestial object classification, respectively. Ait Ouahmed, R. et al. (2024) employed a multi-modality Convolutional Neural Network (CNN) for photometric redshift estimation. However, these models focus on improving estimation accuracy while overlooking the inference speed. Furthermore, their architectural designs are tailored for more complex targets, such as galaxies, whose photometric images contain richer morphology information than RC. Therefore, multi-modality models based on photometric images and spectral data may encounter significant overfitting issues when extracting features of RC. This can hinder the model's ability to derive meaningful information from the photometric image modality, ultimately preventing any improvement in performance.

The paper proposes the Spectral data and Photometric images **Fusion** (SPFusion) lightweight multi-modality model for RC physical parameter estimation to solve the problems above. The subsequent sections of this paper are organized as follows: Section 2 describes the RC catalog, their spectral data and photometric images, as well as the construction of the datasets used in the experiments. Section 3 presents a detailed explanation of the structural design of each module in the SPFusion model. Section 4 discusses the experimental setup and analyzes the results, including comparative experiments with other models and ablation studies on the SPFusion architecture. Section 5 offers an interpretability analysis of SPFusion, incorporating signal-to-noise ratios (SNR) and the spectral features of RC. Finally, Section 6 concludes our work.

## 2. DATA PREPARATION AND DATASET

We use the catalog from He et al. (2022), which provides mass ( $m_{\odot}$ ) and age (Gyr) parameters for 184,318 RCs. The catalog is cross-matched with the Low-Resolution Spectra (LRS) from LAMOST DR10 (Cui et al. 2012; Zhao et al. 2012), and the resulting data is further cross-matched with photometric images from the Sloan Digital Sky Survey (SDSS; Abdurro'uf et al. 2022) DR17. Defective images and spectra are then removed. The flux calibration is performed for each LAMOST spectrum, and all spectra are cropped to the wavelength range from 4000Å to 9000Å. For the photometric images, we use the frames of  $g$ ,  $r$ , and  $i$ -bands and align the three bands by targeting the  $g$ -band with the Python-based Astropy (Astropy Collaboration et al. 2013, 2018, 2022) and Reproject (Robitaille et al. 2013) packages. Each frame is cropped to focus on a  $128 \times 128$  pixel area centered on the target RC to remove unnecessary information from the images. After data preparation, the final dataset includes 55,697 RCs with both photometric images and spectral data.

To evaluate the performance of our model across the entire dataset, we construct the three-fold cross-validation dataset. In each fold, one-third of the data is designated as the test set, while the remaining two-thirds are split into training and validation sets in a 7:3 ratio. The final data partitioning is summarized in Table 1. Following Zheng et al. (2020) and Wu et al. (2023), we apply min-max normalization to each spectral data according to the following formula, thereby improving the efficiency of model training (Miyato et al. 2018):

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

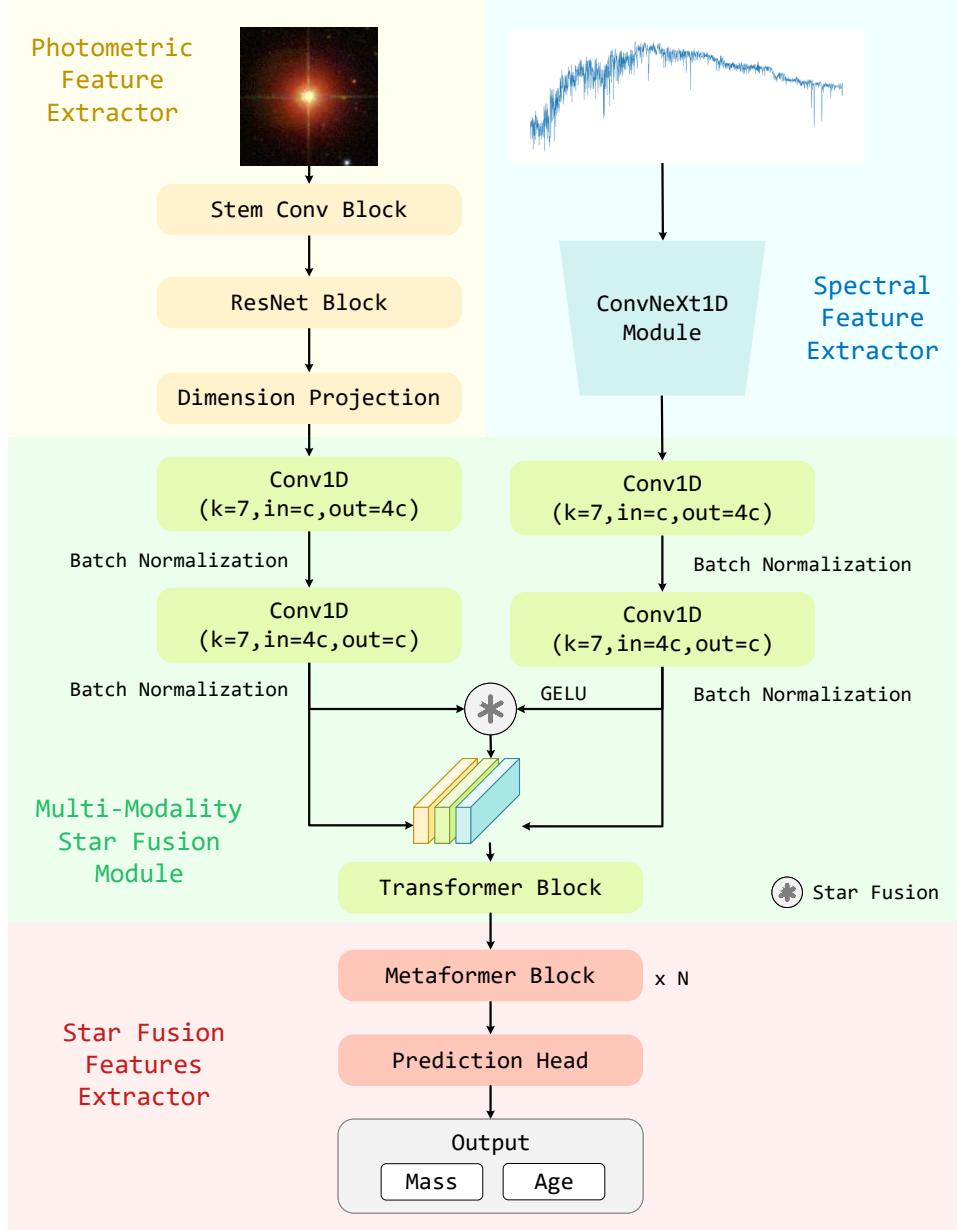
where  $x_{\min}$  and  $x_{\max}$  represent the minimum and maximum flux of the given spectral data, respectively. Similarly, the min-max normalization is applied to each band for the photometric images.

**Table 1.** The number of data in the training, validation, and test sets for each fold in the dataset.

Dataset	No. of data
Training set	25,991
Validation set	11,140
Test set	18,566
Total No.	55,697

## 3. METHOD

To solve the problems summarized in Section 1, we propose a lightweight multi-modality end-to-end model for physical parameter estimation, referred to as SPFU-



**Figure 1.** The architecture of SPFusion model.

sion. As shown in Figure 1, the SPFusion model contains four modules. The photometric feature extractor module extracts the photometric modality features from the three-band photometric images and aligns their dimensions to the spectral modality. Similarly, based on a ConvNeXt 1D model (Liu et al. 2022), the spectral feature extractor module extracts significant spectral features from the input spectral data.

Subsequently, the photometric and spectral modality features are fused by the multi-modality star fusion module, which captures critical deep semantic features correlated with both modalities. Finally, the photometric, spectral, and fused features are concatenated at the

channel level and passed through the star fusion features extractor module to refine the deep semantic features of the fused form. The model then outputs the final estimations for the mass and age of RCs.

During the training stage, we use the loss function defined in Equation 2 to calculate the losses for multiple parameter estimations simultaneously:

$$Loss = \alpha L_{MSE_{mass}} + \beta L_{MSE_{age}} \quad (2)$$

where  $\alpha$  and  $\beta$  are the coefficients for the loss functions associated with mass and age, respectively, and are set to  $1 \times 10^2$  and 1, respectively.  $L_{MSE_{mass}}$  and  $L_{MSE_{age}}$

represent the Mean Squared Error (MSE) losses for the mass and age estimation, respectively.

### 3.1. Photometric and Spectral Feature Extractor

Unlike photometric images of galaxies with noticeable morphological structures, photometric images of RCs contain limited feature information. Therefore, commonly using high-accuracy image feature extractors, such as Vision Transformer (Dosovitskiy et al. 2021), MLP-Mixer (Tolstikhin et al. 2021), or RepLKNet (Ding et al. 2022), may lead to severe overfitting in the extractor or limiting the feature extraction from the photometric modality, and failing to improve and even degrading parameter estimation accuracy.

As shown in Figure 1, we adopt a shallow, small-scale convolutional neural network consisting of a single ResNet (He et al. 2016) block as the photometric feature extractor to avoid the overfitting. There are three substructures in the extractor. For the input photometric image  $x_{in_p} \in \mathbb{R}^{b \times c \times h \times w}$ , we use a stem convolution block to obtain features map  $x_p \in \mathbb{R}^{b \times 4n \times h/2 \times w/2}$ , as defined by Equations 3 to 5:

$$x_p = \sigma(BN(Conv_{c_{out}=3n, k=3, s=2}(x_{in_p}))) \quad (3)$$

$$x_p = \sigma(BN(Conv_{c_{out}=4n, k=3, s=1}(x_p))) \quad (4)$$

$$x_p = Conv_{c_{out}=4n, k=3, s=1}(x_p) \quad (5)$$

where  $b$  means the batch size,  $c$  represents the number of input channels, and  $h$  and  $w$  correspond to the height and width of the input photometric images, respectively.  $\sigma$  represents the ReLU activation function (Glorot et al. 2011), and  $BN$  refers to the batch normalization layer.  $n$  indicates the number of feature channels in hidden layers, which is set to 8 in our model. Among the three convolutional operators with a kernel size of 3 described in Equations 3-5, only the first operator uses a stride of 2 to reduce the feature map size by half.

Afterward, the feature map output by the stem convolution block is processed via a ResNet block to obtain the shallow semantic features of the photometric modality, indicated as  $x_p \in \mathbb{R}^{b \times 8n \times h/4 \times w/4}$ . Finally, these photometric modality features are passed through a dimension projection module to align their dimensions with the spectral modality. The calculation is expressed as follows:

$$x_p = BN(Conv_{c_{out}=4n, k=3, s=1}(x_p)) \quad (6)$$

$$x_p = Conv_{c_{out}=4n, k=3, s=1}(x_p) \quad (7)$$

$$x_p = \delta(Linear(x_p)) \quad (8)$$

where  $\delta$  denotes the GELU activation function (Hendrycks & Gimpel 2023), and  $Linear$  represents a linear layer.

For the spectral data, following Wu et al. (2024), we apply a one-dimensional implementation of ConvNeXt (Liu et al. 2022), referred to as ConvNeXt 1D. The module is lightweight with a few parameters, offering efficient inference capabilities while retaining robust feature extraction performance for one-dimensional data.

### 3.2. Multi-Modality Star Fusion Module

In multimedia, multi-modality models such as CLIP utilize complex architectures to fuse features from modalities with abundant information. However, for multi-modality tasks based on photometric images and spectral data, the feature information in the data is sparse, particularly in photometric images. Thus, a more straightforward structure is required to fuse the features effectively, avoiding severe overfitting and significantly reducing parameters.

Therefore, we put forward the multi-modality star fusion module based on the star operator (Ma et al. 2024). The star operator has demonstrated excellent feature extraction capabilities in natural image classification. According to Ma et al. (2024), when combined with multiple convolution operators, the star operator can fuse feature information from two tensor streams. Leveraging this advantage, we construct a module specifically designed for feature fusion with the star operator.

As indicated in Figure 1, for the multi-modality star fusion module, the semantic feature vectors  $x_p \in \mathbb{R}^{b \times 4n \times d}$  from the photometric feature extractor and  $x_s \in \mathbb{R}^{b \times 4n \times d}$  from the spectral feature extractor are first processed separately. The processing involves a one-dimensional feature expansion convolution operator with a kernel size of 7 and output channels four times the input channels, coupled with a batch normalization layer, and a one-dimensional feature compression convolution operator with a kernel size of 7 and output channels one-fourth of the input channels, also coupled with a batch normalization layer. Subsequently, the two feature maps are fused using the star operator as defined in Equation 9 to generate the fused feature  $x_f \in \mathbb{R}^{b \times 4n \times d}$ .

$$x_f = x_p * \delta(x_s) \quad (9)$$

where  $d$  represents the dimension of the feature vector, and  $*$  denotes the star operator.

The fused features  $x_f$ , photometric modality features  $x_p$ , and spectral modality features  $x_s$  are concatenated along the channel dimension to obtain the mixed features  $x_{mix} \in \mathbb{R}^{b \times 3 \cdot 4n \times d}$ . A one-dimensional Transformer block (Vaswani et al. 2017) is then applied to model the relationships between feature vectors across channels from different modalities, achieving deep semantic feature fusion.

Given that the fused feature information is lower than that of natural images, the number of heads in the Multi-Head Self-Attention (MHSA) mechanism of the Transformer block is reduced to 1 to mitigate overfitting. To further reduce the parameters and computational complexity, we replace the linear-based Multi-Layer Perceptron (MLP) in the traditional Transformer block with a feedforward module based on a one-dimensional channel convolution with an expansion rate of 4. The computation process is defined as follows:

$$x_{\text{mix}} = \delta(\text{Conv}_{c_{\text{out}}=4 \cdot 3 \cdot 4n, k=1, s=1}(x_{\text{mix}})) \quad (10)$$

$$x_{\text{mix}} = \text{Conv}_{c_{\text{out}}=3 \cdot 4n, k=1, s=1}(x_{\text{mix}}) \quad (11)$$

In Section 4, we implement ablation experiments to evaluate the impact of the number of attention heads and the hidden channels of the feedforward unit on the model performance.

### 3.3. Star Fusion Features Extractor

After feature fusion, the multi-modality mixed features  $x_{\text{mix}}$  undergo deep semantic feature extraction through the star fusion features extractor and subsequently via a prediction head to output the parameters shown in Figure 1. In this module, we adopt a Metaformer block proposed by Yu et al. (2022) as the feature extractor. The structure retains the traditional Transformer block with residual connections, a token mixer, and an MLP unit, but replaces the MHSA layer with a global average pooling layer, which serves as a token mixer while reducing the computational complexity to accelerate the inference process without compromising performance.

The features  $x_{\text{mix}}$  output by the Metaformer block are passed through the prediction head to output the parameters of the target object. Specifically,  $x_{\text{mix}}$  first undergoes a channel attention module based on a Squeeze-and-Excitation (SE) block (Hu et al. 2018), denoted as  $CA$  in Equation 12. The module identifies effective feature channels from the stacked feature channels, increasing the weights while reducing irrelevant information weights. Subsequently, the weighted feature map is processed by a one-dimensional convolution with a kernel size of 7 and a batch normalization layer to further extract information from the effective feature channels.

$$x_{\text{mix}} = BN(\text{Conv}_{c_{\text{out}}=3 \cdot 4n, k=7}(CA(x_{\text{mix}}))) \quad (12)$$

Finally,  $x_{\text{mix}} \in \mathbb{R}^{b \times 3 \cdot 4n \times d}$  transforms the dimensions by averaging the feature dimension across each feature channel, resulting in a compressed representation  $x_{\text{mix}} \in \mathbb{R}^{b \times 3 \cdot 4n}$ . The compressed feature vector is then passed through a linear layer to output the results of  $k$  parameters, represented as  $x_{\text{out}} \in \mathbb{R}^{b \times k}$ .

## 4. EXPERIMENTS AND RESULTS

To evaluate the performance of SPFusion in the parameter estimation, we implement comparative experiments with models based on spectral data or photometric images. Additionally, we perform ablation experiments to verify the effectiveness of various structures within SPFusion. All the aforementioned experiments are executed on the RC dataset with physical parameters described in Section 2, using three-fold cross-validation to ensure that the test set encompasses the entire dataset.

For the experimental setup, our model is based on the PyTorch framework (Paszke et al. 2019; Ansel et al. 2024). During training, we employ a cosine annealing learning rate scheduler (Loshchilov & Hutter 2016) to reduce overfitting, with an initial learning rate of  $1 \times 10^{-3}$ . The annealing process starts at the 20th epoch, with an annealing factor set to 1 and a minimum learning rate of  $1 \times 10^{-8}$ . We select Adam (Kingma & Ba 2017) as the optimizer for training and adopt DropPath (Larsson et al. 2017) layers with a probability of 30% to reduce overfitting further. All experiments are completed on an NVIDIA RTX 3090 GPU.

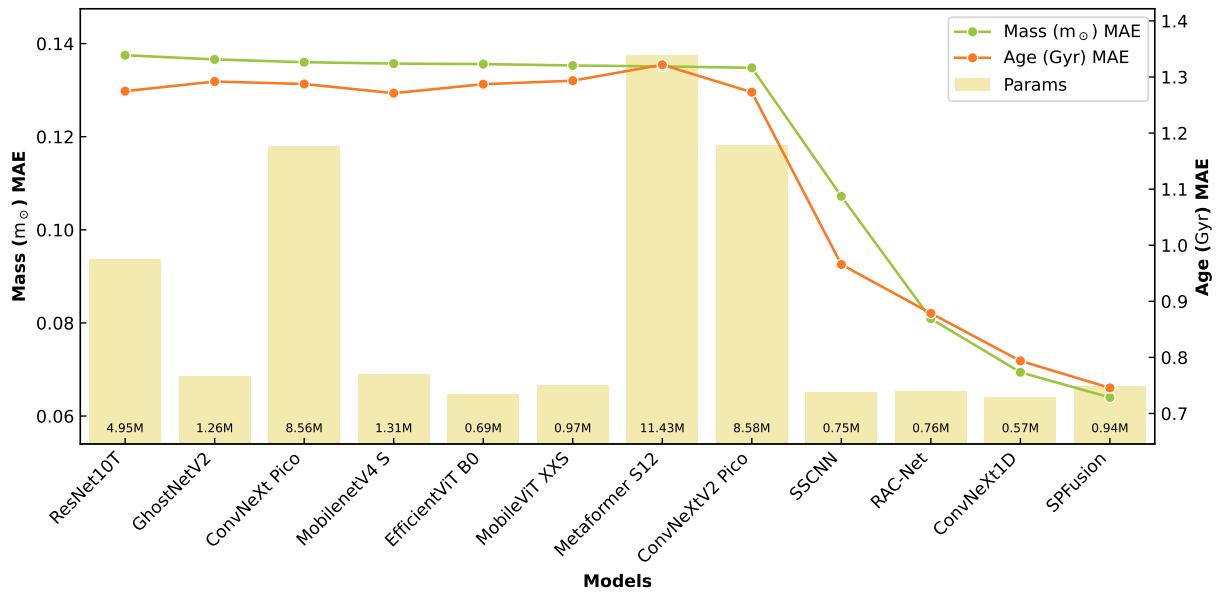
### 4.1. Comparative Experiments

In the comparative experiments, we select lightweight feature extractors that performed excellently in natural image tasks as photometric image modality models for parameter estimation. For spectral data, we use the three top-performing models in spectral feature extraction tasks based on the spectral data modality. All image feature extractors are paired with regression heads composed of two cascaded linear layers, with the hidden feature dimensions determined via parameter search. Furthermore, all models undergo the grid search to identify the best hyperparameters, ensuring optimal performance on our dataset. The details of the search are provided in Appendix A.

The results are presented in Table 2. From top to bottom, ResNet10T (He et al. 2016) to ConvNeXt V2 Pico (Woo et al. 2023) represent photometry-based parameter estimation models, while SSCNN (Liu et al. 2018), RAC-Net (Zou et al. 2020), and ConvNeXt 1D (Liu et al. 2022) correspond to spectrum-based models. As shown in Figure 2, due to the limited feature information in RC photometric images compared to spectral data, all photometry-based models exhibit higher MAE than spectrum-based models and SPFusion. Furthermore, insufficient feature information renders high-parameter models, such as Metaformer S12 (Yu et al. 2022)—the model with the most parameters among the baselines—less effective for this task. Instead, they are

**Table 2.** Comparative experimental results for different models. For each model, the table presents the number of model parameters and MAE of the mass and age estimations. In the data modality column, S represents only using spectral data as input, P corresponds to photometric images only, and S&P denotes using both. \*: ConvNeXt 1D refers to implementing the ConvNeXt model customized for spectral data.

Model	Data Modality	No. of Params (M)	MAE	
			Mass ( $m_{\odot}$ )	Age (Gyr)
ResNet10T (He et al. 2016)	P	4.95	0.1375	1.2747
GhostNet V2 (Tang et al. 2022)	P	1.26	0.1366	1.2919
ConvNeXt Pico (Liu et al. 2022)	P	8.56	0.1360	1.2875
MobilenetV4 S (Qin et al. 2025)	P	1.31	0.1357	1.2711
EfficientViT B0 (Cai et al. 2024)	P	0.69	0.1356	1.2871
MobileViT XXS (Mehta & Rastegari 2022)	P	0.97	0.1353	1.2934
Metaformer S12 (Yu et al. 2022)	P	11.43	0.1351	1.3218
ConvNeXtV2 Pico (Woo et al. 2023)	P	8.58	0.1348	1.2729
SSCNN (Liu et al. 2018)	S	0.75	0.1072	0.9656
RAC-Net (Zou et al. 2020)	S	0.76	0.0809	0.8786
ConvNeXt 1D*	S	<b>0.57</b>	0.0694	0.7938
<b>SPFusion</b>	<b>S&amp;P</b>	0.94	<b>0.0640</b>	<b>0.7456</b>

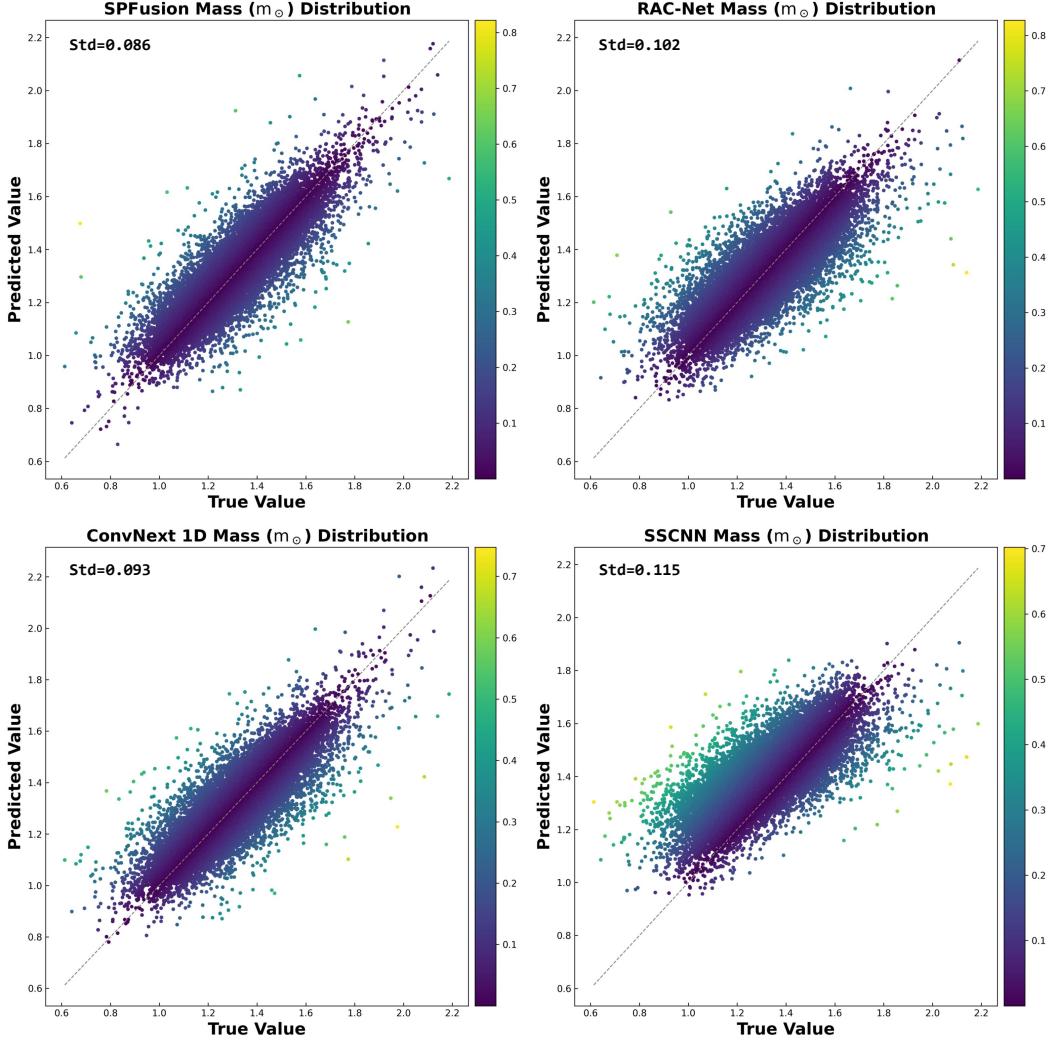


**Figure 2.** The number of model parameters vs. MAE in estimating the mass and age of RCs for different models.

more prone to overfitting, which leads to an upward trend in MAE for age estimation compared to smaller models. For spectrum-based models, the shallow architecture of SSCNN prevents it from effectively learning the complex features of spectral data, resulting in low estimation accuracy compared to RAC-Net and ConvNeXt 1D.

Although the parameters of SPFusion are slightly higher than the best-performing ConvNeXt 1D among the baselines, it remains under one million parameters, which ensures that devices equipped with inference ac-

celerators can efficiently execute inference tasks and estimate the physical parameters. Leveraging features from both modalities, SPFusion achieves superior performance in this task, reducing MAE by 7.78% for mass estimation and by 6.07% for age estimation compared to ConvNeXt 1D. Since SPFusion incorporates ConvNeXt 1D for the spectral modality, these results indicate that the photometric image modality contributes to the parameter estimation, significantly improving model performance.

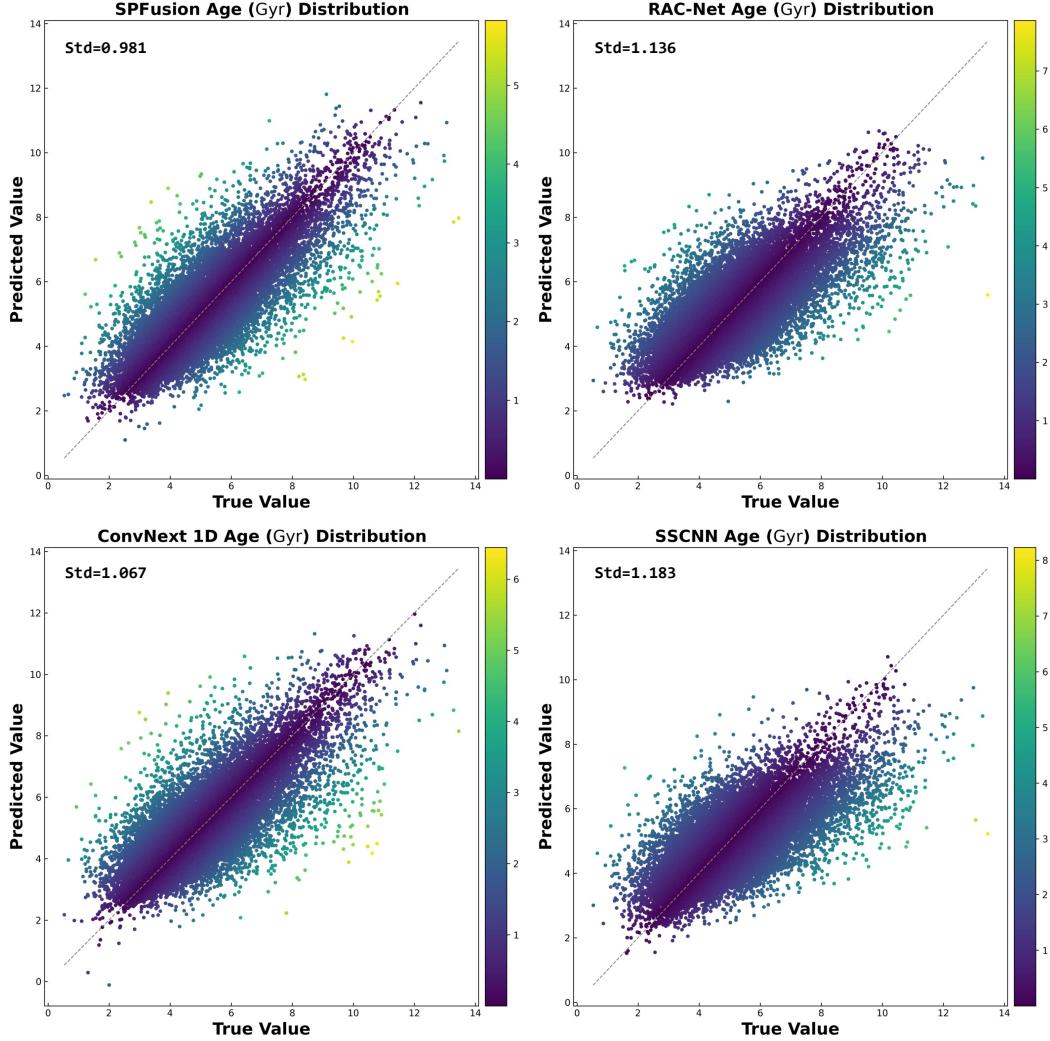


**Figure 3.** Predicted values vs. ground truth for mass estimation via SPFusion, RAC-Net, ConvNeXt 1D, and SSCNN, respectively.

From a qualitative perspective, Figures 3 and 4 show the scatter plots of the predicted values versus ground truth for mass and age via SPFusion and the three spectrum-based models, respectively. The predicted value distribution of SPFusion in mass parameter estimation is more concentrated than those of the other three models. For instance, its standard deviation is reduced by 7.53% relative to ConvNeXt 1D. Additionally, the figures reveal that among the three spectrum-based models, the poorly performing SSCNN exhibits a higher number of outlier samples in its prediction distribution. Similarly, SPFusion maintains a distribution that closely aligns with the ground truth for age estimation, whereas both RAC-Net and SSCNN show excessive outlier samples in their predictions.

Furthermore, we compare the estimation accuracy of SPFusion and three spectrum-based models based on

the *g*-band SNR of LAMOST spectra, segmented at the range of 10. As described in Figure 5, SSCNN shows significantly higher MAE than the other three models due to its limited feature extraction capability. In contrast, leveraging the feature information from photometric images, our model maintains the best performance even when the spectral SNR is low. Moreover, compared to ConvNeXt 1D, as the spectral SNR decreases, the features from the photometric modality play a more important role, leading to a growing performance gap between SPFusion and ConvNeXt 1D. Figure 6 shows the performance of the four models in the age estimation across different SNR ranges. Similarly, the results indicate that as the spectral SNR decreases, the photometric modality features effectively assist the model in achieving better accuracy.



**Figure 4.** Predicted values vs. ground truth for age estimation via SPFusion, RAC-Net, ConvNeXt 1D, and SSCNN, respectively.

Both figures clearly show the complementary relationship between features from different modalities, which helps mitigate the performance degradation caused by deficiencies in single-modality data.

#### 4.2. Ablation Experiments

We implement ablation experiments on the SPFusion model to validate the optimal model structure. Each ablation experiment uses the RC dataset constructed in Section 2 and employs three-fold cross-validation.

##### 4.2.1. Star Operator

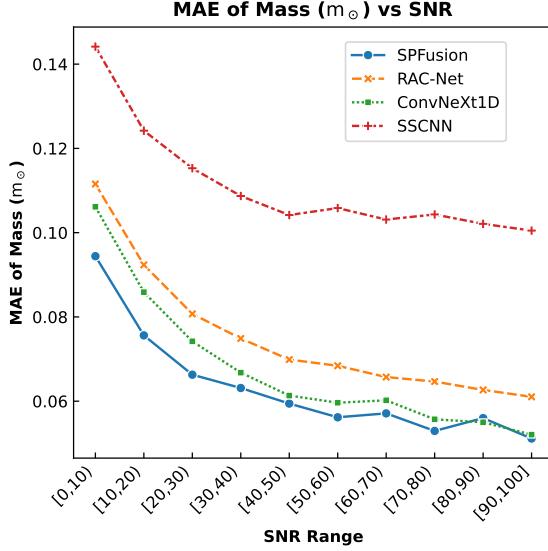
We create a feature fusion module without the star operator to replace the multi-modality star fusion module, thereby validating its effectiveness. For the new structure without the star operator, we remove the two original convolutional blocks from the two tensor streams

**Table 3.** The result of whether the star operator is applied during the star fusion process.

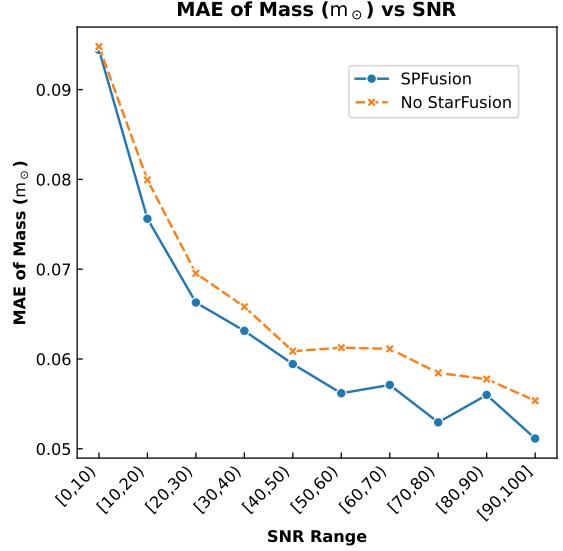
Star Operator	MAE	
	Mass ( $m_{\odot}$ )	Age (Gyr)
✗	0.0674	0.7884
✓	<b>0.0640</b>	<b>0.7456</b>

and exclude using the star operator. Instead, the photometric and spectral modality features are added along the feature dimension and stacked with the photometric and spectral features along the channel dimension. Finally, the stacked features pass through a Transformer block to capture long-range feature dependencies.

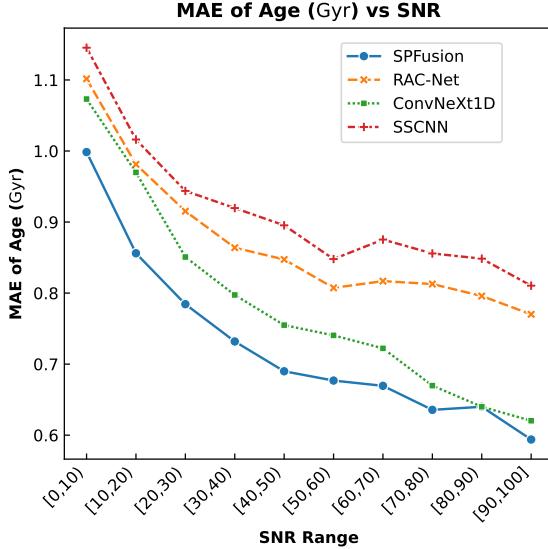
The experimental results are shown in Table 3. The model without the star operator increases MAE by 5.31% for mass estimation and by 5.74% for age esti-



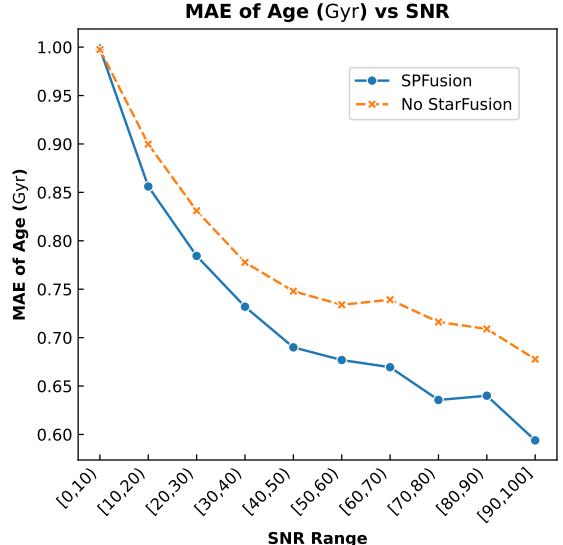
**Figure 5.** The MAE curve for mass estimation across different SNR ranges for SPFusion, RAC-Net, ConvNeXt 1D, and SSCNN, respectively.



**Figure 7.** The MAE curve for mass estimation across different SNR ranges for SPFusion and the model without the star fusion module.



**Figure 6.** The MAE curve for age estimation across different SNR ranges for SPFusion, RAC-Net, ConvNeXt 1D, and SSCNN, respectively.



**Figure 8.** The MAE curve for age estimation across different SNR ranges for SPFusion and the model without the star fusion module.

mation, respectively. Figures 7 and 8 provide a more intuitive visualization of the impact of star fusion on model performance. Without the star fusion module, the model fails to effectively learn the correlations and complementarities between the two modalities. Consequently, regardless of the variation in spectral SNR, the parameter estimation accuracy consistently lags behind that of SPFusion, underscoring the importance of the star fusion module.

#### 4.2.2. Kernel Size of the Multi-Modality Star Fusion Module

The multi-modality star fusion module relies on convolutional blocks in the two tensor streams to extract semantic features from the two modalities. The size of the convolutional receptive field impacts its extraction capability (Ding et al. 2022). To determine the opti-

**Table 4.** The impact of the kernel size in the multi-modality star fusion module’s convolution operators for the feature extraction.

Kernel Size	MAE	
	Mass ( $m_{\odot}$ )	Age (Gyr)
3	0.0662	0.7671
<b>7</b>	<b>0.0640</b>	<b>0.7456</b>
11	0.0659	0.7586

**Table 5.** The impact of the hidden channel scaling ratio in the MHSA feedforward unit within the multi-modality star fusion module.

Scaling Ratio	MAE	
	Mass ( $m_{\odot}$ )	Age (Gyr)
2	0.0683	0.7922
3	0.0662	0.7816
<b>4</b>	<b>0.0640</b>	<b>0.7456</b>
5	0.0671	0.8019

469 mal convolution kernel size, we modify the kernel size to  
470 adjust the receptive field of the convolutional operators.

471 The experimental results in Table 4 indicate that a  
472 smaller convolution kernel size (*kernelsize* = 3) limits  
473 the receptive field of the convolutional operator, render-  
474 ing it incapable of effectively capturing complete feature  
475 information. Therefore, its MAE for both mass and age  
476 estimation is higher than the structure with a kernel size  
477 of 7. Increasing the kernel size helps alleviate this issue;  
478 however, a large kernel (*kernelsize* = 11) results in an  
479 overly broad receptive field, reducing the operator’s abil-  
480 ity to separate fine-grained features and yet degrading  
481 model performance. Accordingly, we use convolutional  
482 operators with a kernel size of 7 as feature extractors for  
483 this module.

#### 484 4.2.3. Hidden Channel Scaling Ratio of the Multi-Modality 485 Star Fusion Module Feedforward Unit

486 The feedforward unit in a traditional Transformer  
487 block enhances the representational ability by scal-  
488 ing features along the feature or channel dimensions  
489 (Vaswani et al. 2017; Yu et al. 2022). Since the ra-  
490 tio affects model performance, we implement ablation  
491 experiments to determine the optimal hidden channel  
492 scale ratio.

493 As shown in Table 5, small-scale ratios, such as 2 or 3,  
494 are unfavorable for feature representation in the feedfor-  
495 ward unit, leading to an increase in MAE compared to  
496 the structure with a ratio of 4. Similar to the behavior  
497 observed with convolutional kernel sizes, large-scale ra-  
498 tios are prone to overfitting when handling limited fea-  
499 ture information, resulting in higher MAE. Hence, we

**Table 6.** The impact of MHSA heads in the multi-modality star fusion module for the feature extraction.

Heads	MAE	
	Mass ( $m_{\odot}$ )	Age (Gyr)
1	<b>0.0640</b>	0.7456
2	0.0656	0.7563
4	0.0645	<b>0.7416</b>
7	0.0665	0.7895

**Table 7.** The impact of Metaformer blocks in the star fusion features extractor for the feature extraction.

Metaformer Blocks	MAE	
	Mass ( $m_{\odot}$ )	Age (Gyr)
1	<b>0.0640</b>	<b>0.7456</b>
2	0.0652	0.7571
3	0.0679	0.7835

500 select a feedforward unit structure with a hidden chan-  
501 nel scaling ratio of 4.

#### 502 4.2.4. MHSA Heads in the Multi-Modality Star Fusion 503 Module

504 The MHSA mechanism in Transformer blocks cap-  
505 tures diverse features by employing multiple attention  
506 heads. For natural images, more heads are often used  
507 to model complex features. However, in the RC pa-  
508 rameter estimation task, where the feature is inherently  
509 limited, more attention heads do not enhance the rep-  
510 resentational ability of the model. In contrast, it leads  
511 to learning redundant features, resulting in overfitting.  
512 The experimental results shown in Table 6 verify the  
513 conclusion above. The MHSA mechanism achieves op-  
514 timal representational ability when using only an atten-  
515 tion head. On the other hand, increasing the number of  
516 attention heads causes a rise in MAE, leading to perfor-  
517 mance degradation.

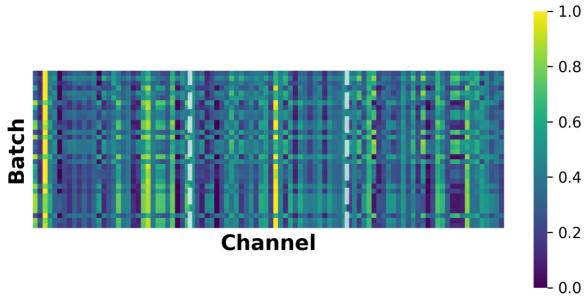
#### 518 4.2.5. Metaformer Blocks in the Star Fusion Features 519 Extractor

520 We implement the ablation experiment on the  
521 Metaformer blocks in the star fusion features extractor,  
522 as shown in Table 7. Due to the limited feature informa-  
523 tion, increasing the depth of the feature extractor does  
524 not enhance the representational ability as effectively as  
525 in natural image tasks. Instead, deeper models exhibit  
526 severe overfitting in this task. Compared to the optimal  
527 configuration, using three Metaformer blocks increases  
528 MAE by 5.08% for age estimation. Based on the exper-  
529 imental results, we select a configuration with only one  
530 Metaformer block, which achieves optimal performance,

reducing the parameters and computational cost of SP-  
Fusion.

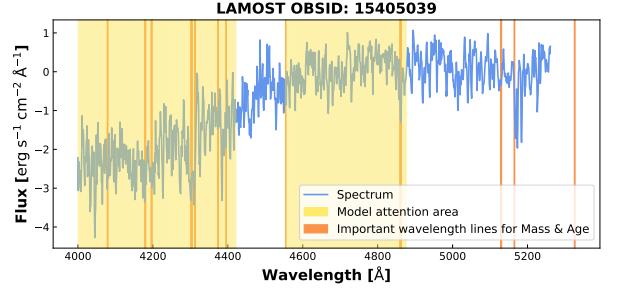
## 533 5. DISCUSSIONS

To preliminarily explore the interpretability of SPFusion, we visualize the channel attention matrix within the prediction head of the star fusion features extractor, as shown in Figure 9, which presents the channel attention heatmap for a input batch. The three sections, divided by two vertical gray dashed lines, correspond to the heatmaps for the photometric modality, fused features, and spectral modality, respectively. Notably, most feature channels effectively use the information they encapsulate. Additionally, channels with significantly higher weights are observed in the photometric modality and fused features, proving that the extracted information supplements the spectral modality.



**Figure 9.** Channel attention heatmap for a single input batch, where the x-axis represents feature channels and the y-axis corresponds to all input data within the batch. The three sections, separated by the two vertical gray dashed lines, correspond to the attention weights for the feature channels in the photometric, fused, and spectral modalities, respectively. The intensity of the color, with yellow representing the highest value, indicates the feature weight associated with each channel, suggesting that the feature is more significant. Conversely, blue indicates lower feature weights, suggesting that the corresponding feature is less important.

Furthermore, we overlay the channel attention matrix of the spectral modality onto the input spectrum. As shown in Figure 10, the model attention areas (yellow low areas) correspond to feature channels with weights greater than 0.5. The characteristic spectral lines for RC mass and age (orange lines) are derived from He et al. (2022), where the data are based on LAMOST blue-end spectra under the wavelength of 5500Å. For the spectral lines in the wavelength range from 4000Å to 4500Å, the feature weights of the spectral modality are significantly higher, which indicates that the model enhances spectral lines in this wavelength range, showing its ability to learn feature information based on these actual physical features.



**Figure 10.** An overlay of the spectral modality feature channel attention weights with the input spectrum under the wavelength of 5500Å. The model attention areas correspond to feature channels with attention weights greater than 0.5, while the important wavelength lines represent the characteristic spectral lines for the RC mass and age estimation (He et al. 2022).

## 561 6. CONCLUSION

To address the challenges of effectively fusing multi-modality features from photometric images with limited information, and spectral data with complex features, while also ensuring a fast and accurate parameter estimation model for RC stars, we propose the lightweight, multi-modality SPFusion model. This model simplifies the photometric feature extraction process by using shallow feature maps, and introduces the star operator and Transformer blocks to achieve the fusion and deep feature extraction of photometric and spectral features. We conduct a dataset for mass and age estimation of RC stars, which combines photometric images and spectral data, and implement comparative experiments with baseline feature extraction methods. The results demonstrate that our model effectively facilitates feature complementarity and fusion between the two modalities. Additionally, five ablation experiments validate the impact of various structural components in the model. Finally, we visualize the feature learning process of SPFusion and perform an exploratory analysis of its interpretability.

Due to its lightweight nature, high speed, and strong accuracy, SPFusion holds significant potential for parameter estimation of large-scale data. While SPFusion demonstrates high efficiency in this context, several limitations persist, particularly in terms of its interpretability and the need for further expansion to additional tasks for more comprehensive validation. These challenges offer promising avenues for future research.

## 591 7. ACKNOWLEDGMENTS

The study is funded by the National Natural Science Foundation of China under grants Nos.12273076 and

<sup>594</sup> 12133001, and Shenzhen Fundamental Research Program (JCYJ20230807094104009).

<sup>595</sup> The Guoshoujing Telescope (the Large Sky Area Multi-object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

<sup>596</sup> Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is [www.sdss.org](http://www.sdss.org). SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group,

<sup>597</sup> Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) /University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

<sup>598</sup> *Software:* Astropy ([Astropy Collaboration et al. 2013, 2018, 2022](#)), Reproject ([Robitaille et al. 2013](#)), PyTorch ([Paszke et al. 2019; Ansel et al. 2024](#))

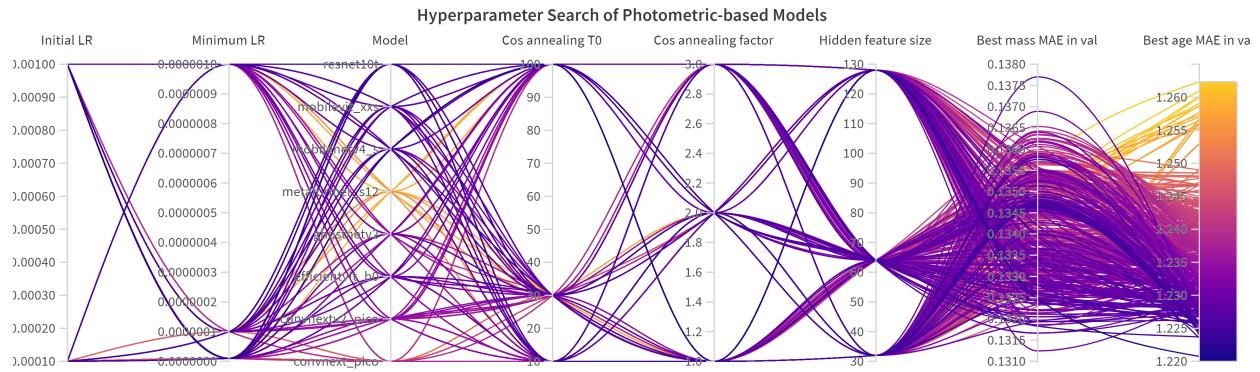
639

## APPENDIX

640

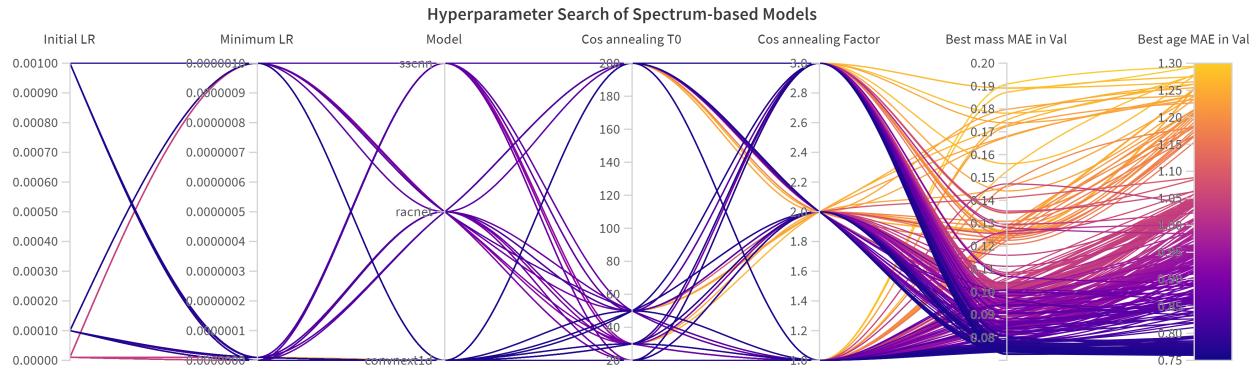
### A. HYPERPARAMETER SEARCH

<sup>641</sup> We implement the grid search to explore all combinations of hyperparameters, ensuring each comparative model <sup>642</sup> to achieve optimal performance on the RC physical parameter estimation. The hyperparameter search processes for <sup>643</sup> photometry- and spectrum-based models are shown in Figures 11 and 12, respectively. We adopt the Wandb software <sup>644</sup><sup>1</sup> to assist with automated search and find the best combinations.



**Figure 11.** The grid search process for photometry-based parameter estimation models.

<sup>1</sup> <https://wandb.ai>



**Figure 12.** The grid search process for spectrum-based parameter estimation models.

## REFERENCES

- 645 Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, The  
646 Astrophysical Journal Supplement Series, 259, 35,  
647 doi: [10.3847/1538-4365/ac4414](https://doi.org/10.3847/1538-4365/ac4414)
- 648 Ait Ouahmed, R., Arnouts, S., Pasquet, J., Treyer, M., &  
649 Bertin, E. 2024, A&A, 683, A26,  
650 doi: [10.1051/0004-6361/202347395](https://doi.org/10.1051/0004-6361/202347395)
- 651 Ansel, J., Yang, E., He, H., et al. 2024, in 29th ACM  
652 International Conference on Architectural Support for  
653 Programming Languages and Operating Systems, Volume  
654 2 (ASPLOS '24) (ACM), doi: [10.1145/3620665.3640366](https://doi.org/10.1145/3620665.3640366)
- 655 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J.,  
656 et al. 2013, A&A, 558, A33,  
657 doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- 658 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M.,  
659 et al. 2018, AJ, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- 660 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L.,  
661 et al. 2022, ApJ, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- 662 Bu, Y., & Pan, J. 2014, Monthly Notices of the Royal  
663 Astronomical Society, 447, 256,  
664 doi: [10.1093/mnras/stu2063](https://doi.org/10.1093/mnras/stu2063)
- 665 Cai, H., Li, J., Hu, M., Gan, C., & Han, S. 2024,  
666 EfficientViT: Multi-Scale Linear Attention for  
667 High-Resolution Dense Prediction.  
668 <https://arxiv.org/abs/2205.14756>
- 669 Casamiquela, L., Carrera, R., Blanco-Cuaresma, S., et al.  
670 2017, Monthly Notices of the Royal Astronomical  
671 Society, 470, 4363, doi: [10.1093/mnras/stx1481](https://doi.org/10.1093/mnras/stx1481)
- 672 Cassisi, S., & Salaris, M. 1997, Monthly Notices of the  
673 Royal Astronomical Society, 285, 593,  
674 doi: [10.1093/mnras/285.3.593](https://doi.org/10.1093/mnras/285.3.593)
- 675 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Research  
676 in Astronomy and Astrophysics, 12, 1197,  
677 doi: [10.1088/1674-4527/12/9/003](https://doi.org/10.1088/1674-4527/12/9/003)
- 678 De Marchi, G., Panagia, N., & Girardi, L. 2013, Monthly  
679 Notices of the Royal Astronomical Society, 438, 513,  
680 doi: [10.1093/mnras/stt2233](https://doi.org/10.1093/mnras/stt2233)
- 681 Ding, X., Zhang, X., Han, J., & Ding, G. 2022, in  
682 Proceedings of the IEEE/CVF Conference on Computer  
683 Vision and Pattern Recognition (CVPR), 11963–11975
- 684 Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2021, in  
685 International Conference on Learning Representations.  
686 <https://openreview.net/forum?id=YicbFdNTTy>
- 687 Girardi, L. 2016, Annual Review of Astronomy and  
688 Astrophysics, 54, 95,  
689 doi: [10.1146/annurev-astro-081915-023354](https://doi.org/10.1146/annurev-astro-081915-023354)
- 690 Glorot, X., Bordes, A., & Bengio, Y. 2011, in Proceedings  
691 of Machine Learning Research, Vol. 15, Proceedings of  
692 the Fourteenth International Conference on Artificial  
693 Intelligence and Statistics, ed. G. Gordon, D. Dunson, &  
694 M. Dudík (Fort Lauderdale, FL, USA: PMLR), 315–323.  
695 <https://proceedings.mlr.press/v15/glorot11a.html>
- 696 Hatzidimitriou, D. 1991, Monthly Notices of the Royal  
697 Astronomical Society, 251, 545,  
698 doi: [10.1093/mnras/251.4.545](https://doi.org/10.1093/mnras/251.4.545)
- 699 Hawkins, K., Leistedt, B., Bovy, J., & Hogg, D. W. 2017,  
700 Monthly Notices of the Royal Astronomical Society, 471,  
701 722, doi: [10.1093/mnras/stx1655](https://doi.org/10.1093/mnras/stx1655)
- 702 He, K., Zhang, X., Ren, S., & Sun, J. 2016, in Proceedings  
703 of the IEEE Conference on Computer Vision and Pattern  
704 Recognition (CVPR)
- 705 He, X.-J., Luo, A.-L., & Chen, Y.-Q. 2022, Monthly Notices  
706 of the Royal Astronomical Society, 512, 1710,  
707 doi: [10.1093/mnras/stac484](https://doi.org/10.1093/mnras/stac484)
- 708 Hendrycks, D., & Gimpel, K. 2023, Gaussian Error Linear  
709 Units (GELUs). <https://arxiv.org/abs/1606.08415>

- 710 Henghes, B., Thiyyagalingam, J., Pettitt, C., Hey, T., &  
 711 Lahav, O. 2022, Monthly Notices of the Royal  
 712 Astronomical Society, 512, 1696,  
 713 doi: [10.1093/mnras/stac480](https://doi.org/10.1093/mnras/stac480)
- 714 Hu, J., Shen, L., & Sun, G. 2018, in Proceedings of the  
 715 IEEE Conference on Computer Vision and Pattern  
 716 Recognition (CVPR)
- 717 Kingma, D. P., & Ba, J. 2017, Adam: A Method for  
 718 Stochastic Optimization.  
<https://arxiv.org/abs/1412.6980>
- 719 Larsson, G., Maire, M., & Shakhnarovich, G. 2017, in  
 720 International Conference on Learning Representations.  
<https://openreview.net/forum?id=S1VaB4cex>
- 721 Li, X., & Lin, B. 2023, Monthly Notices of the Royal  
 722 Astronomical Society, 521, 6354,  
 723 doi: [10.1093/mnras/stad831](https://doi.org/10.1093/mnras/stad831)
- 724 Li, X., Lu, Y., Comte, G., et al. 2015, The Astrophysical  
 725 Journal Supplement Series, 218, 3,  
 726 doi: [10.1088/0067-0049/218/1/3](https://doi.org/10.1088/0067-0049/218/1/3)
- 727 Liu, W., Zhu, M., Dai, C., et al. 2018, Monthly Notices of  
 728 the Royal Astronomical Society, 483, 4774,  
 729 doi: [10.1093/mnras/sty3020](https://doi.org/10.1093/mnras/sty3020)
- 730 Liu, Z., Mao, H., Wu, C.-Y., et al. 2022, in Proceedings of  
 731 the IEEE/CVF Conference on Computer Vision and  
 732 Pattern Recognition (CVPR), 11976–11986
- 733 Loshchilov, I., & Hutter, F. 2016, CoRR, abs/1608.03983
- 734 Ma, X., Dai, X., Bai, Y., Wang, Y., & Fu, Y. 2024, in  
 735 Proceedings of the IEEE/CVF Conference on Computer  
 736 Vision and Pattern Recognition (CVPR), 5694–5703
- 737 Mao, S., & Paczyński, B. 2002, Monthly Notices of the  
 738 Royal Astronomical Society, 337, 895,  
 739 doi: [10.1046/j.1365-8711.2002.05951.x](https://doi.org/10.1046/j.1365-8711.2002.05951.x)
- 740 Mehta, S., & Rastegari, M. 2022, MobileViT: Light-weight,  
 741 General-purpose, and Mobile-friendly Vision  
 742 Transformer. <https://arxiv.org/abs/2110.02178>
- 743 Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. 2018,  
 744 in International Conference on Learning Representations.  
<https://openreview.net/forum?id=B1QRgziT>
- 745 Nataf, D. M., Gould, A., Fouqué, P., et al. 2013, The  
 746 Astrophysical Journal, 769, 88,  
 747 doi: [10.1088/0004-637X/769/2/88](https://doi.org/10.1088/0004-637X/769/2/88)
- 748 Parker, L., Lanusse, F., Golkar, S., et al. 2024, Monthly  
 749 Notices of the Royal Astronomical Society, 531, 4990,  
 750 doi: [10.1093/mnras/stae1450](https://doi.org/10.1093/mnras/stae1450)
- 751 Pasquet-Itam, J., & Pasquet, J. 2018, A&A, 611, A97,  
 752 doi: [10.1051/0004-6361/201731106](https://doi.org/10.1051/0004-6361/201731106)
- 753 Pasquet, Johanna, Bertin, E., Treyer, M., Arnouts, S., &  
 754 Fouchez, D. 2019, A&A, 621, A26,  
 755 doi: [10.1051/0004-6361/201833617](https://doi.org/10.1051/0004-6361/201833617)
- 756 Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in  
 757 Neural Information Processing Systems 32 (Curran  
 758 Associates, Inc.), 8024–8035.  
<http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-lib.pdf>
- 759 Qin, D., Leichner, C., Delakis, M., et al. 2025, in Computer  
 760 Vision – ECCV 2024, ed. A. Leonardis, E. Ricci, S. Roth,  
 761 O. Russakovsky, T. Sattler, & G. Varol (Cham: Springer  
 762 Nature Switzerland), 78–96
- 763 Radford, A., Kim, J. W., Hallacy, C., et al. 2021, in  
 764 Proceedings of Machine Learning Research, Vol. 139,  
 765 Proceedings of the 38th International Conference on  
 766 Machine Learning, ed. M. Meila & T. Zhang (PMLR),  
 767 8748–8763.  
<https://proceedings.mlr.press/v139/radford21a.html>
- 768 Rizhko, M., & Bloom, J. S. 2024, in Neurips 2024  
 769 Workshop Foundation Models for Science: Progress,  
 770 Opportunities, and Challenges.  
<https://openreview.net/forum?id=w0iQslghXD>
- 771 Robitaille, T., Deil, C., & Ginsburg, A. 2013,  
 772 astropy/reproject. <https://github.com/astropy/reproject>
- 773 Stanek, K. Z. 1996, The Astrophysical Journal, 460, L37,  
 774 doi: [10.1086/309976](https://doi.org/10.1086/309976)
- 775 Sumi, T., Eyer, L., & Woźniak, P. R. 2003, Monthly  
 776 Notices of the Royal Astronomical Society, 340, 1346,  
 777 doi: [10.1046/j.1365-8711.2003.06411.x](https://doi.org/10.1046/j.1365-8711.2003.06411.x)
- 778 Tang, Y., Han, K., Guo, J., et al. 2022, in Advances in  
 779 Neural Information Processing Systems, ed. S. Koyejo,  
 780 S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh,  
 781 Vol. 35 (Curran Associates, Inc.), 9969–9982. [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/40b60852a4abdaa696b5a1a78da34635-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/40b60852a4abdaa696b5a1a78da34635-Paper-Conference.pdf)
- 782 Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., et al. 2021, in  
 783 Advances in Neural Information Processing Systems, ed.  
 784 M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, &  
 785 J. W. Vaughan, Vol. 34 (Curran Associates, Inc.),  
 786 24261–24272.  
[https://proceedings.neurips.cc/paper\\_files/paper/2021/file/cba0a4ee5cccd02fda0fe3f9a3e7b89fe-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/cba0a4ee5cccd02fda0fe3f9a3e7b89fe-Paper.pdf)
- 787 Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in  
 788 Advances in Neural Information Processing Systems, ed.  
 789 I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach,  
 790 R. Fergus, S. Vishwanathan, & R. Garnett, Vol. 30  
 791 (Curran Associates, Inc.).  
[https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fdbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fdbd053c1c4a845aa-Paper.pdf)

- 807 Wan, J.-C., Liu, C., Deng, L.-C., et al. 2015, Research in  
 808     Astronomy and Astrophysics, 15, 1166,  
 809     doi: [10.1088/1674-4527/15/8/006](https://doi.org/10.1088/1674-4527/15/8/006)
- 810 Woo, S., Debnath, S., Hu, R., et al. 2023, in Proceedings of  
 811     the IEEE/CVF Conference on Computer Vision and  
 812     Pattern Recognition (CVPR), 16133–16142
- 813 Wozniak, P. R., & Stanek, K. Z. 1996, The Astrophysical  
 814     Journal, 464, 233, doi: [10.1086/177315](https://doi.org/10.1086/177315)
- 815 Wu, C., Wong, O. I., Rudnick, L., et al. 2018, Monthly  
 816     Notices of the Royal Astronomical Society, 482, 1211,  
 817     doi: [10.1093/mnras/sty2646](https://doi.org/10.1093/mnras/sty2646)
- 818 Wu, J., He, Y., Wang, W., et al. 2024, The Astronomical  
 819     Journal, 167, 260, doi: [10.3847/1538-3881/ad38ae](https://doi.org/10.3847/1538-3881/ad38ae)
- 820 Wu, J., Zhang, Y., Qu, M., Jiang, B., & Wang, W. 2023,  
 821     Universe, 9, doi: [10.3390/universe9110477](https://doi.org/10.3390/universe9110477)
- 822 Xiang, G., Chen, J., Qiu, B., & Lu, Y. 2021, Publications  
 823     of the Astronomical Society of the Pacific, 133, 024504,  
 824     doi: [10.1088/1538-3873/abd997](https://doi.org/10.1088/1538-3873/abd997)
- 825 Xiang, M.-S., Liu, X.-W., Shi, J.-R., et al. 2016, Monthly  
 826     Notices of the Royal Astronomical Society, 464, 3657,  
 827     doi: [10.1093/mnras/stw2523](https://doi.org/10.1093/mnras/stw2523)
- 828 Yang, T., & Li, X. 2015, Monthly Notices of the Royal  
 829     Astronomical Society, 452, 158,  
 830     doi: [10.1093/mnras/stv1210](https://doi.org/10.1093/mnras/stv1210)
- 831 Yu, W., Luo, M., Zhou, P., et al. 2022, in Proceedings of  
 832     the IEEE/CVF Conference on Computer Vision and  
 833     Pattern Recognition (CVPR), 10819–10829
- 834 Zhang, B., Liu, C., & Deng, L.-C. 2020, The Astrophysical  
 835     Journal Supplement Series, 246, 9,  
 836     doi: [10.3847/1538-4365/ab55ef](https://doi.org/10.3847/1538-4365/ab55ef)
- 837 Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng,  
 838     L.-C. 2012, Research in Astronomy and Astrophysics, 12,  
 839     723, doi: [10.1088/1674-4527/12/7/002](https://doi.org/10.1088/1674-4527/12/7/002)
- 840 Zheng, Z.-P., Qiu, B., Luo, A.-L., & Li, Y.-B. 2020,  
 841     Publications of the Astronomical Society of the Pacific,  
 842     132, 024504, doi: [10.1088/1538-3873/ab5ed7](https://doi.org/10.1088/1538-3873/ab5ed7)
- 843 Zou, Z., Zhu, T., Xu, L., & Luo, A.-L. 2020, Publications of  
 844     the Astronomical Society of the Pacific, 132, 044503,  
 845     doi: [10.1088/1538-3873/ab7548](https://doi.org/10.1088/1538-3873/ab7548)