



Estimating Stellar Atmospheric Parameters and Elemental Abundances Using Fully Connected Residual Network

Shuo Li^{1,2} , Yin-Bi Li^{1,2} , A-Li Luo^{1,2} , Jun-Chao Liang^{1,2} , Hai-Ling Lu^{1,2} , and Hugh R. A. Jones³

¹ CAS Key Laboratory of Optical Astronomy, ² National Astronomical Observatories, Beijing 100101, China; ybli@bao.ac.cn, lal@nao.cas.cn, liangjc@nao.cas.cn

³ University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Physics, Astronomy and Mathematics, University of Hertfordshire, UK

Received 2025 April 30; revised 2025 July 12; accepted 2025 August 5; published 2025 October 17

Abstract

Stellar atmospheric parameters and elemental abundances are traditionally determined using template matching techniques based on high-resolution spectra. However, these methods are sensitive to noise and unsuitable for ultra-low-resolution data. Given that the Chinese Space Station Telescope (CSST) will acquire large volumes of ultra-low-resolution spectra, developing effective methods for ultra-low-resolution spectral analysis is crucial. In this work, we investigated the Fully Connected Residual Network (FCResNet) for simultaneously estimating atmospheric parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$) and elemental abundances ($[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, $[\text{Mg}/\text{Fe}]$). We trained and evaluated FCResNet using CSST-like spectra ($R \sim 200$) generated by degrading LAMOST spectra ($R \sim 1800$), with reference labels from APOGEE. FCResNet significantly outperforms traditional machine learning methods (KNN, XGBoost, SVR) and CNN in prediction precision. For spectra with the g -band signal-to-noise ratio greater than 20, FCResNet achieves precisions of 78 K, 0.15 dex, 0.08 dex, 0.05 dex, 0.10 dex, and 0.05 dex for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, respectively, on the test set. FCResNet processes one million spectra in only 42 s while maintaining a simple architecture with only 348 KB model size. These results suggest that FCResNet is a practical and promising tool for processing the large volume of ultra-low-resolution spectra that will be obtained by CSST in the future.

Key words: methods: data analysis – stars: abundances – stars: fundamental parameters – methods: analytical

1. Introduction

Stellar atmospheric parameters—including effective temperature (T_{eff}), surface gravity ($\log g$), and metallicity ($[\text{Fe}/\text{H}]$)—together with element-to-iron abundance ratio ($[\text{X}/\text{Fe}]$) such as $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$, are fundamental parameters for characterizing stars and understanding the structure and chemical evolution of the Milky Way. Large-scale spectroscopic surveys such as the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012) Experiment for Galactic Understanding and Exploration (LEGUE; Deng et al. 2012), the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009), the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), the Galactic Archaeology with HERMES (GALAH; De Silva et al. 2015), the RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006) and the Dark Energy Spectroscopic Instrument (DESI) Milky Way Survey (Cooper et al. 2023) have provided vast quantities of stellar spectra and parameters, which have significantly advanced our understanding of the evolution and chemical formation of the Milky Way.

Among stellar parameters, atmospheric parameters are relatively easier to estimate than elemental abundances ($[\text{X}/\text{H}]$) using traditional methods. Gray & Johanson (1991) and

Kovtuykh et al. (2003) used line-depth ratios to estimate T_{eff} from high-resolution spectra, achieving measurement uncertainties as low as ~ 10 K. Sousa et al. (2012) used a large sample of 451 solar-type stars to develop a calibration based on equivalent widths, enabling efficient estimation of $[\text{Fe}/\text{H}]$. Asteroseismology is particularly well-suited for estimating $\log g$ of stars and has been used to calibrate $\log g$ in both LAMOST and RAVE surveys for dwarfs and giants (Wang et al. 2016; Valentini et al. 2017). The accurate measurement of $[\text{X}/\text{H}]$ typically requires high-resolution spectroscopy, where absorption lines of specific elements can be clearly resolved and reliably identified. At medium- and low-resolution, most absorption lines are weak, blended, or both, which makes precise abundance determination difficult. However, spectral fitting across wide wavelength ranges can still enable reliable estimates of α -element-to-iron abundance ratio ($[\alpha/\text{Fe}]$). For example, $[\alpha/\text{Fe}]$ has been measured from medium-resolution spectra with a precision of approximately 0.05 dex (Kirby et al. 2008), and from low-resolution spectra with a precision better than 0.10 dex (Lee et al. 2011). Despite these successes, accurate determination of atmospheric parameters and elemental abundances from low- and ultra-low-resolution spectra remains challenging for traditional methods such as template matching and line-depth ratios. Template matching is sensitive to noise and

heavily depends on the coverage and quality of theoretical grids. Similarly, line-depth ratio techniques are limited by line blending and resolution constraints, making $[X/H]$ estimation unreliable. Recently, with the rise of machine learning and deep learning techniques, it has become possible to overcome these limitations by learning complex mappings between spectra and stellar parameters, enabling the transfer of high-precision labels to low- and ultra-low-resolution spectra.

Label transfer methods have already been successfully applied to low-resolution spectra, such as those from LAMOST. For example, Liu et al. (2015) applied a Support Vector Regression (SVR) model trained on LAMOST giant spectra with Kepler seismic $\log g$, achieving an uncertainty of about 0.10 dex for spectra with signal-to-noise ratio (S/N) in the g band (S/N_g) higher than 20. Zhang et al. (2019) and Zhang et al. (2020) transferred stellar parameters from APOGEE to LAMOST spectra using a neural network (StarNet; Fabbro et al. 2018) and an SVR-based model (SLAM), achieving uncertainties of 45/49 K for T_{eff} , 0.10/0.10 dex for $\log g$, 0.05/0.04 dex for $[M/H]$, 0.03/0.03 dex for $[\alpha/M]$, 0.06/0.06 dex for $[C/M]$, and 0.07/0.11 dex for $[N/M]$ for spectra with S/N_g larger than 10/100, respectively. Xiang et al. (2019) developed the DD-Payne, a data-driven model that inherits essential ingredients from *The Payne* (Ting et al. 2019) and *The Cannon* (Ness et al. 2015), incorporating constraints from theoretical models and trained on LAMOST spectra with stellar parameters from GALAH and APOGEE, to derive T_{eff} , $\log g$, and 16 elemental abundances in $[X/Fe]$, achieving typical precisions of better than 30 K and 0.07 dex for T_{eff} and $\log g$, 0.03–0.10 dex for the majority of elements, and 0.20–0.30 dex for $[Cu/Fe]$ and $[Ba/Fe]$ for spectra with $S/N_g > 50$. Li & Lin (2023) proposed StarGRUNet, a deep learning model trained on LAMOST spectra with stellar labels from APOGEE, to estimate T_{eff} , $\log g$, and 13 elemental abundances in $[X/H]$, achieving typical precisions of 94 K for T_{eff} , 0.16 dex for $\log g$, 0.07–0.16 dex for most abundances, 0.18 dex for $[N/H]$ and 0.22 dex for $[Cr/H]$ for spectra with $S/N_g > 5$.

Although previous studies have demonstrated the feasibility of label transfer methods on low-resolution spectra, related work on ultra-low-resolution spectra, such as those from the Chinese Space Station Telescope (CSST; $R \sim 200$), remains scarce. For example, Wu et al. (2024) developed a two-dimensional Convolutional Neural Network (CNN) trained on LAMOST spectra degraded to $R \sim 200$ to estimate atmospheric parameters and $[C/Fe]$, achieving mean absolute errors of 99 K for T_{eff} , 0.22 dex for $\log g$, 0.14 dex for $[Fe/H]$, and 0.26 dex for $[C/Fe]$ for spectra with $S/N_g > 10$. To support future stellar parameter estimation from ultra-low-resolution CSST spectra, this work focused on LAMOST spectra degraded to $R \sim 200$. Given that CSST is expected to collect an enormous amount of spectroscopic data, it is crucial to develop a precise and computationally efficient method capable of estimating atmospheric parameters and elemental abundances. In this work, we proposed a neural network

architecture, the Fully Connected Residual Network (FCResNet), which demonstrates superior performance in estimating T_{eff} , $\log g$, $[Fe/H]$, $[C/Fe]$, $[N/Fe]$, and $[Mg/Fe]$.

The structure of this paper is organized as follows. Section 2 describes the data set and preprocessing procedures. Section 3 introduces the FCResNet architecture. In Section 4, we present and analyze the experimental results in detail. Finally, Section 5 summarizes the main results of this study.

2. Data

In this section, we describe the data sources and the preprocessing procedures. To simulate the observational characteristics of the future CSST and obtain data with similar properties, we selected low-resolution spectra from LAMOST DR8, degraded to a resolution of $R \sim 200$. The corresponding stellar parameters— T_{eff} , $\log g$, $[Fe/H]$, $[C/Fe]$, $[N/Fe]$, and $[Mg/Fe]$ —were adopted from APOGEE DR16.

LAMOST is located at the Xinglong Observatory, northeast of Beijing, China. It is a national scientific research facility operated by the National Astronomical Observatories of the Chinese Academy of Sciences (Su & Cui 2004). LAMOST is a special reflecting Schmidt telescope that combines a wide field of view (5°) with a large effective aperture (ranging from 3.6 to 4.9 m, depending on the pointing altitude and hour angle). This configuration allows LAMOST to obtain 4000 spectra simultaneously in a single exposure, with a spectral resolution of $R \sim 1800$ (Cui et al. 2012). The primary scientific goal of LAMOST is to collect millions of stellar spectra to enable studies of stellar astrophysics, the structure of the Milky Way, as well as extragalactic astronomy and cosmology (Zhao et al. 2012). The spectroscopic survey officially began in 2012 September and has since obtained over 10 million spectra (Luo et al. 2015). LAMOST DR8 includes spectra from the pilot survey conducted between 2011 and 2012, as well as from the regular survey conducted from 2012 to 2020, comprising a total of 10,388,423 stellar spectra.

APOGEE is one of the major programs within SDSS III (Majewski et al. 2017) and IV (Blanton et al. 2017), and represents a large-scale stellar spectroscopic survey conducted in the near-infrared (H -band) portion of the spectrum. The survey targets over 700,000 stars, primarily Galactic red giants spanning all stellar populations. It also includes red giants from the Magellanic Clouds and other nearby dwarf galaxies, as well as a significant number of FGKM-type dwarf stars (Zasowski et al. 2013; Smith et al. 2021). In this work, the stellar atmospheric parameters and elemental abundances were adopted from APOGEE DR16, which provides high-resolution ($R = 22,500$), near-infrared spectra covering the wavelength range 15140–16940 Å for approximately 430,000 stars across both the northern and southern skies (Jönsson et al. 2020). These spectra are used to derive precise stellar atmospheric parameters, radial velocities, and chemical abundances of up to

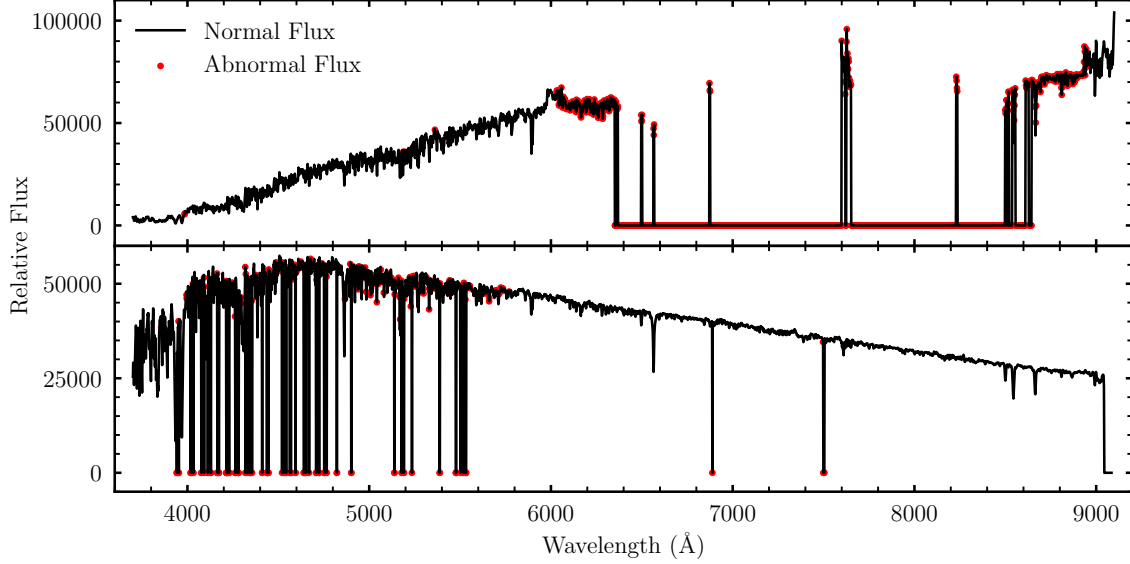


Figure 1. Examples of abnormal LAMOST spectra. The top panel shows a spectrum with missing flux, while the bottom panel shows a spectrum with abnormal flux.

26 species using the APOGEE Stellar Parameter and Chemical Abundance Pipeline (ASPCAP; García Pérez et al. 2016). In DR16, several spectroscopically derived parameters have been calibrated, improving their accuracy. A detailed description of each parameter and its calibration process is provided in Jönsson et al. (2020).

We first obtained 5,211,501 low-resolution stellar spectra from LAMOST DR8 with S/N_g greater than 20. These were then cross-matched with APOGEE DR16, resulting in 104,867 spectra with corresponding stellar parameters. We adopted stellar parameters from APOGEE rather than LAMOST due to APOGEE’s higher spectral resolution, which provides more accurate atmospheric parameters and its capability to determine elemental abundances. Subsequently, we applied the following selection criteria to filter reliable parameters and spectra, and performed spectral preprocessing to obtain CSST-like spectra:

1. For stellar parameters, Jönsson et al. (2020) noted that stars with $STARFLAG \neq 0$, $ASPCAPFLAG \neq 0$, or parameters set to -9999.99 may be affected by spectral quality issues in APOGEE DR16, potentially resulting in unreliable parameter estimates. Therefore, such entries were excluded from our sample. To eliminate scale differences among different stellar parameters, all data sets (including training, validation, and test sets) undergo Z-score normalization based on the mean and standard deviation computed from the training set, as defined in Equation (1), with the normalized parameters serving as target outputs for FCResNet training. After training, the model’s predictions require denormalization to convert back to obtain the true stellar parameter values.
2. For spectra, we first extracted the wavelength range 4000–8122 Å. Following Luo et al. (2012), spectra with

$fibermask \neq 0$ or $ormask \neq 0$ were removed to eliminate abnormalities (see Figure 1). To match the CSST resolution, we degraded the spectra to $R = 200$ using the method in Lustig-Yaeger et al. (2019), and used the resulting 143-dimensional spectra as CSST-like spectra (see Figure 2). As input to the FCResNet, each spectrum was individually normalized using Equation (1).

$$\hat{x}_i = \frac{x_i - \mu(x_i)}{\sigma(x_i)}, \quad (1)$$

where x_i denotes the original data, $\hat{x}_i$ is the normalized data, and $\mu(x_i)$ and $\sigma(x_i)$ represent the mean and standard deviation of x_i , respectively.

After the above steps, we obtained 22,632 stars containing all six stellar parameters: T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$. We randomly divided the data set into a training set (60%), a validation set (20%), and a test set (20%). The distributions of all these parameters are shown in Figure 3. It is evident that the data set includes a substantial number of both dwarf and giant stars, allowing the test set to effectively evaluate the generalization capability of our model.

3. Method

In the field of deep learning, Fully connected Neural Networks (FNN), as universal approximators (Hornik et al. 1989), possess the capability to approximate arbitrary functions (Liang & Srikant 2016) and demonstrate exceptional performance in classification and regression tasks (Arulampalam & Bouzerdoum 2003). Recent studies (Simonyan & Zisserman 2014; Szegedy et al. 2014; Srivastava et al. 2015a, 2015b) have revealed the decisive influence of network depth on model accuracy, with significant performance improvements achievable through

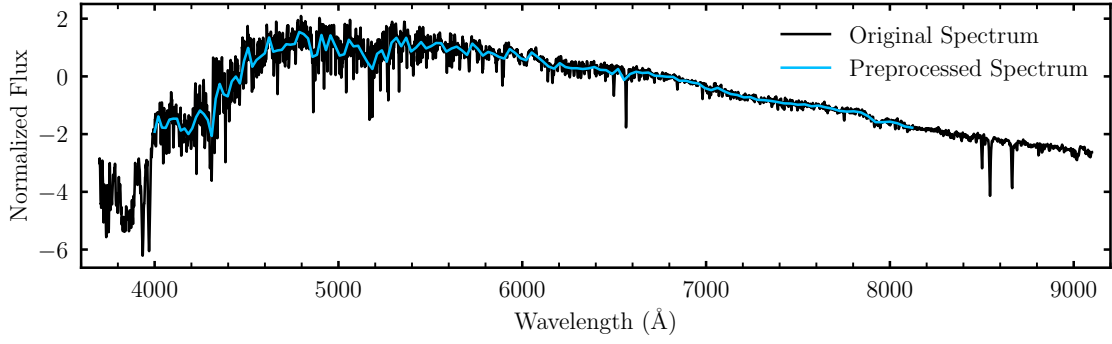


Figure 2. Comparison between the original LAMOST spectrum (black), covering the range 3690–9100 Å, and the CSST-like spectrum (blue), covering 4000–8122 Å. The CSST-like spectrum was obtained by degrading the original resolution from $R = 1800$ to $R = 200$, following the method described in Lustig-Yaeger et al. (2019).

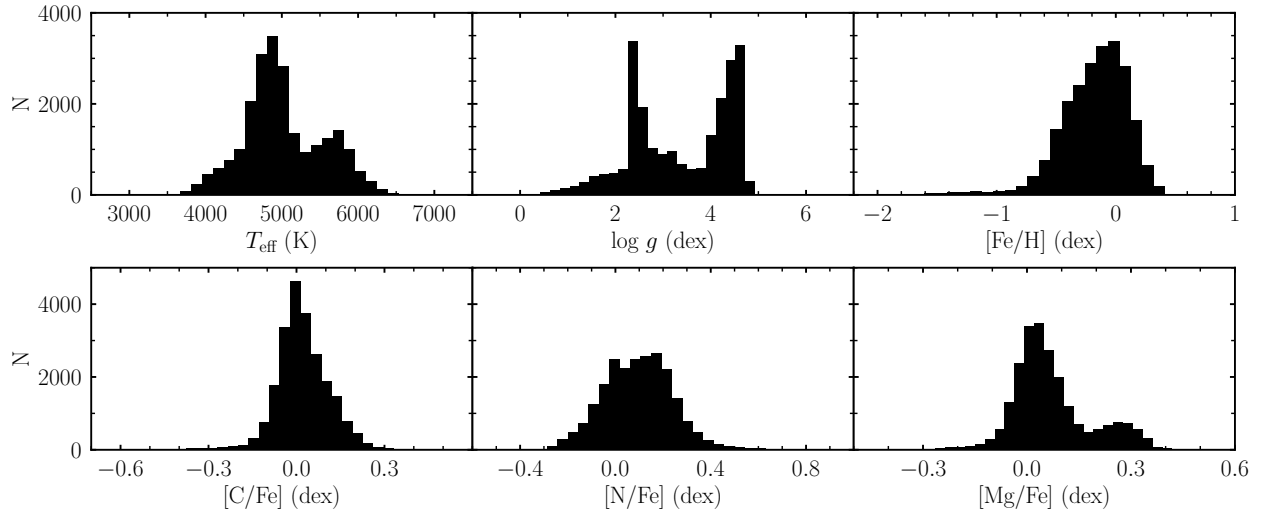


Figure 3. Distributions of stellar parameters in our data set. From left to right and top to bottom, the panels correspond to T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, respectively. These parameters were derived from APOGEE DR16 and correspond to 22,632 stars cross-matched with LAMOST DR8. All parameters were selected based on rigorous quality criteria to ensure the reliability of the sample for model training and evaluation.

increasing the number of network layers. However, as network depth increases, issues such as model degradation (He & Sun 2014; Srivastava et al. 2015a), and vanishing/exploding gradients (Bengio et al. 1994) become increasingly prominent. To address these challenges, residual block techniques have emerged as an effective solution. In this work, we combined FNN with residual blocks to construct the FCResNet, aiming to resolve the aforementioned problems encountered by deep FNN in regression tasks. We first outline the fundamental principles of FNN in Section 3.1, then elaborate on the mechanisms of residual blocks in Section 3.2, and finally describe the construction of FCResNet in Section 3.3.

3.1. Fully Connected Neural Network

In this section, we briefly outline the calculation process of FNN. For more details, we refer the reader to Svozil et al. (1997).

Suppose that (x_i, y_i) , $i = 1, 2, \dots, N$, represents the data set, where $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$ is the stellar spectrum corresponding to n dimensions, y_i is the stellar parameter associated with x_i , and $\hat{y}_i$ denotes the value estimated by the FNN. Let L denote the number of layers in the FNN (with the first layer being the input layer and the L th layer being the output layer). We use w_{jk}^l to denote the weight of the connection from the k th neuron (or node) in the $(l-1)$ th layer to the j th neuron in the l th layer, and use b_j^l to denote the bias of the j th neuron in the l th layer. The FNN fits the model through the following process:

1. *Input x_i .* Set the activation of the input layer as $a^1 = x_i$.
2. *Forward propagation.* For each layer $l = 2, \dots, L$, initialize the weights and biases. The activation a_j^l of the j th neuron in the l th layer is related to the activations in

the $(l - 1)$ th layer by:

$$a_j^l = \sigma \left(\sum_{k=1}^{n_{l-1}} w_{jk}^l a_k^{l-1} + b_j^l \right) = \sigma(z_j^l), \quad (l = 2, \dots, L - 1), \quad (2)$$

where the summation is over all neurons k in the $(l - 1)$ th layer. For the output layer ($l = L$), corresponding to the regression prediction, the output is:

$$\hat{y}_i = a_i^L = \sum_{k=1}^{n_{L-1}} w_{ik}^L a_k^{L-1} + b_i^L, \quad (3)$$

The mean square error (MSE) is used as the loss function to measure the difference between the estimated value $\hat{y}_i$ and the true value y_i :

$$\text{loss} = \sum_{i=1}^N (\hat{y}_i - y_i)^2. \quad (4)$$

3. *Back propagation.* The goal of back propagation is to update the weights and biases using the gradient descent method, thereby minimizing the loss defined in Equation (4). For each layer $l = L - 1, L - 2, \dots, 2$, substituting Equations (2) and (3) into Equation (4), the updates for w_{jk}^l and b_j^l are:

$$\begin{cases} w_{jk}^l \leftarrow w_{jk}^l - \eta \frac{\partial \text{loss}}{\partial w_{jk}^l} = w_{jk}^l - \eta a_k^{l-1} \delta_j^l, \\ b_j^l \leftarrow b_j^l - \eta \frac{\partial \text{loss}}{\partial b_j^l} = b_j^l - \eta \delta_j^l, \end{cases} \quad (5)$$

where η is the learning rate, $\delta_j^L = \frac{\partial \text{loss}}{\partial \hat{y}_i} = \hat{y}_i - y_i$, and

$$\delta_j^l = \left(\sum_{k=1}^{n_{l+1}} w_{kj}^{l+1} \delta_k^{l+1} \right) \sigma'(z_j^l).$$

Repeating steps 2 and 3, the weights w_{jk}^l and biases b_j^l are iteratively updated until the loss defined in Equation (4) approaches zero. Once the network is trained, Equation (3) can be used for making predictions.

The performance of the FNN largely depends on the depth of its hidden layers: deeper architectures generally offer greater representational capacity. However, increasing the number of hidden layers introduces several challenges that must be addressed, such as:

1. Deep architectures may lead to the problems of vanishing and exploding gradients during training (Hochreiter 1998; Pascanu et al. 2012). Specifically, in Equation (5), when l is small (i.e., layers close to the input and far from the output), the backpropagated error term δ_j^l may approach zero or infinity due to repeated gradient multiplication. As a result, the updates to w_{jk}^l and b_j^l either vanish or

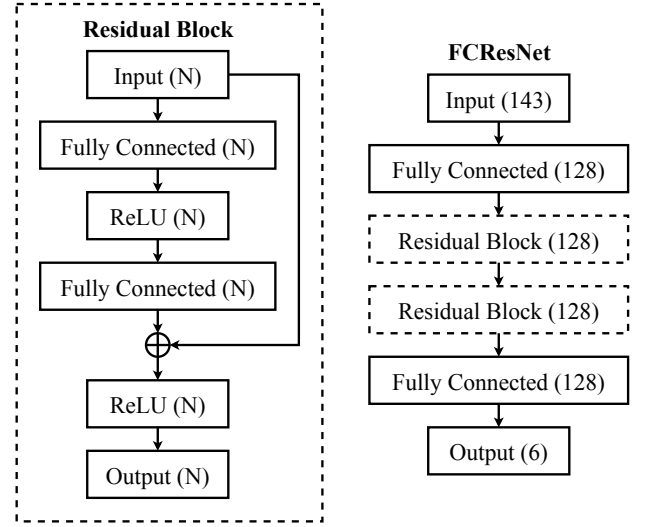


Figure 4. Left: The structure of the residual block used in this work. Right: The architecture of the proposed Fully Connected Residual Network (FCResNet) model.

diverge, hindering the learning of lower layers and degrading the model's predictive capability.

2. He et al. (2016a) observed that increasing the depth of hidden layers may lead to model degradation: as the network becomes deeper, the model's accuracy saturates and then deteriorates rapidly. This phenomenon is characterized by increasing errors on both the training and test sets, and notably, it is not caused by overfitting, but rather by optimization difficulties inherent in deep architectures.

To address the aforementioned issues, He et al. (2016a) proposed residual blocks, which effectively alleviate the problems of vanishing/exploding gradients and model degradation, particularly in deep CNNs for image classification tasks.

3.2. Residual Blocks

A residual block is a combination of several hidden layers that introduces identity mapping to ease the training of deep networks (as shown in the left panel of Figure 4). Mathematically, the residual block modifies Equation (2) as follows:

$$a_j^l = \sigma \left(\sum_{k=1}^{n_{l-1}} w_{jk}^l a_k^{l-1} + b_j^l + a_j^{l-2} \right), \quad (6)$$

where a_j^{l-2} represents the identity mapping from the $(l - 2)$ th layer to the l th layer on the j th neuron. Based on Equation (6), a recursive connection between deep and shallow hidden

layers can be established:

$$\begin{aligned} a_j^l &= \sigma(z_j^l + a_j^{l-2}) = \sigma(z_j^l + z_j^{l-2} + a_j^{l-4}) = \dots \\ &= \sigma(z_j^l + z_j^{l-2} + \dots + z_j^{l-2m} + a_j^{l-2m-2}), \end{aligned} \quad (7)$$

where m denotes the number of residual blocks, and a_j^{l-2m-2} is the identity mapping from the $(l - 2m - 2)$ th layer to the $(l - 2m)$ th layer on the j th neuron. Equation (7) demonstrates the establishment of a direct link between the l th (deep) and $(l - 2m - 2)$ th (shallow) layers.

He et al. (2016b) proved that such skip connections in residual blocks can effectively mitigate the problems of gradient vanishing and exploding, enabling the parameters w_{jk}^l and b_j^l in Equation (5) to converge more reliably toward the minimum of the loss function defined in Equation (4). In addition, these connections help address the degradation problem in deep neural networks, which has been empirically validated. The residual networks achieved top performance on several benchmarks, including ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation (He et al. 2016a).

3.3. FCResNet

We incorporated residual blocks (as shown in the left panel of Figure 4) into a Fully Connected (FC) architecture to construct our model, referred to as FCResNet. The architecture of FCResNet consists of an input layer, two residual blocks, two FC layers, and an output layer (as shown in the right panel of Figure 4).

1. *Input layer.* The input to the network is a normalized CSST-like spectrum represented by a 143-dimensional vector.
2. *Residual blocks.* Each residual block consists of a sequence of layers: the first FC layer, a ReLU activation, the second FC layer, and another ReLU activation. In addition, an identity shortcut connects the input of the block directly to the output of the second FC layer through element-wise addition.
3. *FC layers.* Each FC layer contains 128 neurons. The first FC layer projects the input spectrum into a 128-dimensional feature space, serving as the input to the residual blocks. After processing through the residual blocks, the final FC layer aggregates the feature representations learned by the residual blocks. Following common neural network design principles, inserting an FC layer before the output layer helps stabilize the model's predictions.
4. *Output layer.* The output layer contains six neurons corresponding to the normalized stellar parameters: T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$. The actual physical values are obtained by applying the inverse normalization to these outputs.

FCResNet is implemented using PyTorch. During training, 128 spectra are input in each batch (i.e., `batch_size = 128`), and the initial learning rate is set to 0.001. If the validation loss does not decrease for 20 consecutive epochs, the learning rate is reduced by a factor of 10, down to a minimum of 10^{-8} . The loss function is the mean MSE of six normalized stellar parameters, and the optimization algorithm is Adam (Kingma & Ba 2014). Training is terminated once the validation loss reaches its minimum.

4. Performance Evaluation

4.1. Evaluation Criteria

Let n denote the number of spectra in the data set, y_i the true label corresponding to the input x_i , and $\hat{y}_i$ the predicted value produced by FCResNet. The following metrics were used to evaluate the performance of our model:

1. The mean of the residuals (μ):

$$\mu = \frac{1}{n} \sum_{i=1}^n e_i, \quad (8)$$

where $e_i = \hat{y}_i - y_i$ represents the residuals of the parameters. The mean of the residuals reflects the systematic bias of a model's predictions. Specifically, it indicates whether the model tends to consistently overestimate or underestimate the true values across the data set.

2. The standard deviation of the residuals (σ):

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \mu)^2}, \quad (9)$$

The standard deviation of the residuals reflects the dispersion or precision of a model's predictions. It quantifies how much the predicted values deviate from the true values on average, regardless of the direction of the error.

4.2. Results

The trained FCResNet was applied to simultaneously predict six stellar parameters from CSST-like spectra. The first two rows of Figure 5 present the comparison between FCResNet predictions and APOGEE labels on the test set for the directly predicted parameters, while the last row of Figure 5 compares the calculated elemental abundances $[\text{X}/\text{H}] = [\text{X}/\text{Fe}] + [\text{Fe}/\text{H}]$. As shown, the precisions for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, $[\text{Mg}/\text{Fe}]$, $[\text{C}/\text{H}]$, $[\text{N}/\text{H}]$ and $[\text{Mg}/\text{H}]$ reach approximately 78 K, 0.15 dex, 0.08 dex, 0.05 dex, 0.10 dex, 0.05 dex, 0.09 dex, 0.13 dex and 0.08 dex,

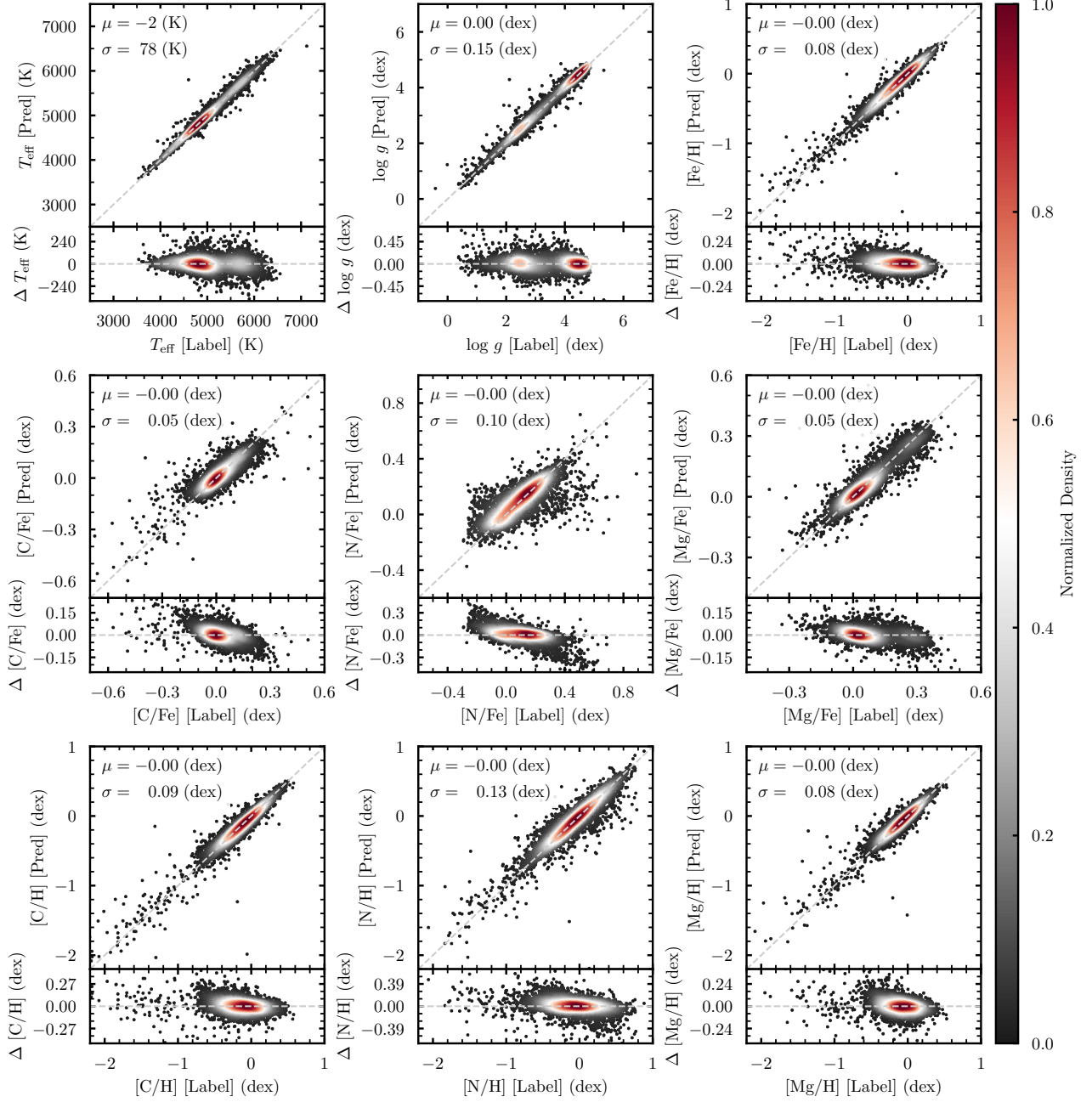


Figure 5. Comparison between FCResNet predictions and APOGEE labels on the test set. From left to right and top to bottom, the panels correspond to T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, $[\text{Mg}/\text{Fe}]$, $[\text{C}/\text{H}]$, $[\text{N}/\text{H}]$ and $[\text{Mg}/\text{H}]$, respectively. The first six parameters are directly predicted by the model, while the last three are calculated using $[\text{X}/\text{H}] = [\text{X}/\text{Fe}] + [\text{Fe}/\text{H}]$. In each panel, the upper subplot shows a density scatter plot of FCResNet predictions vs. APOGEE labels, while the lower subplot displays the residuals (predicted value minus label) vs. the labels. The μ and σ represent the systematic bias and dispersion, respectively.

respectively, with systematic biases close to zero. These results demonstrate that FCResNet achieves excellent performance in simultaneously predicting multiple stellar parameters from spectra with a resolution as low as $R = 200$. The last column of Table 1 summarizes the performance of FCResNet on both the validation and test sets. Although the selected model during training was based on its best performance on the validation

set, the performances on the validation and test sets are nearly identical. This indicates that the model does not suffer from overfitting and exhibits strong generalization capability. Such consistency across data sets demonstrates the robustness of FCResNet and its ability to generalize well to unseen spectra. To validate the stability of the model, we investigated how the prediction precision of FCResNet for the six stellar parameters

Table 1

Comparison of Model Size, Execution Time for One Million Spectra, and Prediction Precision of Six Stellar Parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$) across KNN, XGBoost, SVR, CNN, and FCResNet on the Validation and Test Sets

Method	KNN	XGBoost	SVR	CNN	FCResNet
Model Size	33 MB	405 MB	54 MB	943 KB	348 KB
Execution Time	30 s	6 minutes	27 minutes	46 s	42 s
Validation Set					
ΔT_{eff} (K)	5 ± 140	3 ± 88	3 ± 82	0 ± 79	-2 ± 76
$\Delta \log g$ (dex)	0.02 ± 0.32	0.01 ± 0.19	0.00 ± 0.17	-0.00 ± 0.15	-0.00 ± 0.15
$\Delta [\text{Fe}/\text{H}]$ (dex)	0.03 ± 0.18	0.00 ± 0.09	0.00 ± 0.08	-0.00 ± 0.07	0.00 ± 0.07
$\Delta [\text{C}/\text{Fe}]$ (dex)	-0.00 ± 0.08	-0.00 ± 0.06	0.00 ± 0.06	-0.00 ± 0.05	-0.00 ± 0.05
$\Delta [\text{N}/\text{Fe}]$ (dex)	0.00 ± 0.12	0.00 ± 0.11	-0.00 ± 0.10	0.00 ± 0.10	0.00 ± 0.10
$\Delta [\text{Mg}/\text{Fe}]$ (dex)	-0.01 ± 0.08	-0.00 ± 0.06	-0.00 ± 0.05	0.00 ± 0.05	-0.00 ± 0.04
Loss	0.3972	0.2189	0.1932	0.1703	0.1608
Test Set					
ΔT_{eff} (K)	6 ± 144	1 ± 92	1 ± 86	-2 ± 80	-2 ± 78
$\Delta \log g$ (dex)	0.03 ± 0.33	0.01 ± 0.20	0.00 ± 0.17	-0.00 ± 0.15	0.00 ± 0.15
$\Delta [\text{Fe}/\text{H}]$ (dex)	0.03 ± 0.18	0.00 ± 0.10	0.00 ± 0.08	-0.00 ± 0.08	-0.00 ± 0.08
$\Delta [\text{C}/\text{Fe}]$ (dex)	-0.00 ± 0.08	0.00 ± 0.06	0.00 ± 0.06	0.00 ± 0.05	-0.00 ± 0.05
$\Delta [\text{N}/\text{Fe}]$ (dex)	-0.00 ± 0.12	-0.00 ± 0.11	-0.01 ± 0.11	-0.00 ± 0.10	-0.00 ± 0.10
$\Delta [\text{Mg}/\text{Fe}]$ (dex)	-0.01 ± 0.08	0.00 ± 0.06	-0.00 ± 0.05	0.00 ± 0.05	-0.00 ± 0.05
Loss	0.4150	0.2348	0.2104	0.1872	0.1792

varies with S/N_g on the test set, as shown in Figure 6. It can be seen that the prediction precision remains relatively stable across different S/N_g levels, indicating that the model exhibits strong robustness when applied to spectra with varying quality.

To further validate the reliability and reasonableness of the model-predicted parameters, we compare the Kiel diagrams of FCResNet and APOGEE on the test set in Figure 7. The gray lines represent 7 Gyr isochrones generated using the theoretical models from the PAdova and TRIeste Stellar Evolution Code (PARSEC; Bressan et al. 2012), with metallicities of -0.6 , -0.3 , 0 , and 0.3 dex. The results demonstrate that FCResNet not only successfully reproduces the number density distribution and metallicity gradient patterns of APOGEE in the Kiel diagram, but also shows excellent agreement with the PARSEC isochrones. Additionally, Figure 8 presents the density distributions of $[\text{X}/\text{Fe}]$ versus $[\text{Fe}/\text{H}]$ for both APOGEE labels and FCResNet predictions on the test set. The distribution patterns of $[\text{X}/\text{Fe}]$ versus $[\text{Fe}/\text{H}]$ between FCResNet and APOGEE show remarkable similarity, further confirming the reliability of the model's elemental abundance predictions.

As shown in Figure 9, we present the pairwise density distributions of elemental abundances $[\text{Fe}/\text{H}]$, $[\text{C}/\text{H}]$, $[\text{N}/\text{H}]$, and $[\text{Mg}/\text{H}]$ for both APOGEE and FCResNet on the test set, with Pearson correlation coefficients (r) annotated in the upper left corner of each subplot. The APOGEE results reveal strong correlations among $[\text{X}/\text{H}]$ abundances, with r values ranging from 0.85 to 0.95. FCResNet exhibits highly similar

distribution patterns to APOGEE, with enhanced correlations reflected in increased r values ranging from 0.89 to 0.97, demonstrating that FCResNet not only captures but also strengthens the intrinsic relationships among elemental abundances. To investigate whether the multi-parameter prediction of FCResNet only relies on statistical correlations between parameter labels, we trained six individual FCResNet models for single-parameter prediction using the same data set. The test set results show precisions of 76 K, 0.15 dex, 0.09 dex, 0.06 dex, 0.10 dex, and 0.05 dex for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$, respectively. The experimental results demonstrate that the precision of multi-parameter model is very similar to that of single-parameter models, indicating that the network does not simply make mutual inferences based on statistical correlations between stellar parameter labels, but rather learns underlying patterns from spectral features. For ultra-low-resolution CSST-like spectra, the model may not directly identify specific absorption line features when predicting elemental abundances, but neural networks, as powerful feature extractors, can capture complex spectral patterns correlated with elemental abundances, even when the physical interpretation of these features is not always apparent.

To further investigate whether FCResNet can identify prominent absorption features (such as the MgI b triplet) when predicting $[\text{Mg}/\text{Fe}]$, we present the SHapley Additive exPlanations (SHAP) analysis for $[\text{Mg}/\text{Fe}]$ predictions by FCResNet on the test set in Figure 10, highlighting the top three wavelength regions that the model focuses on most. The

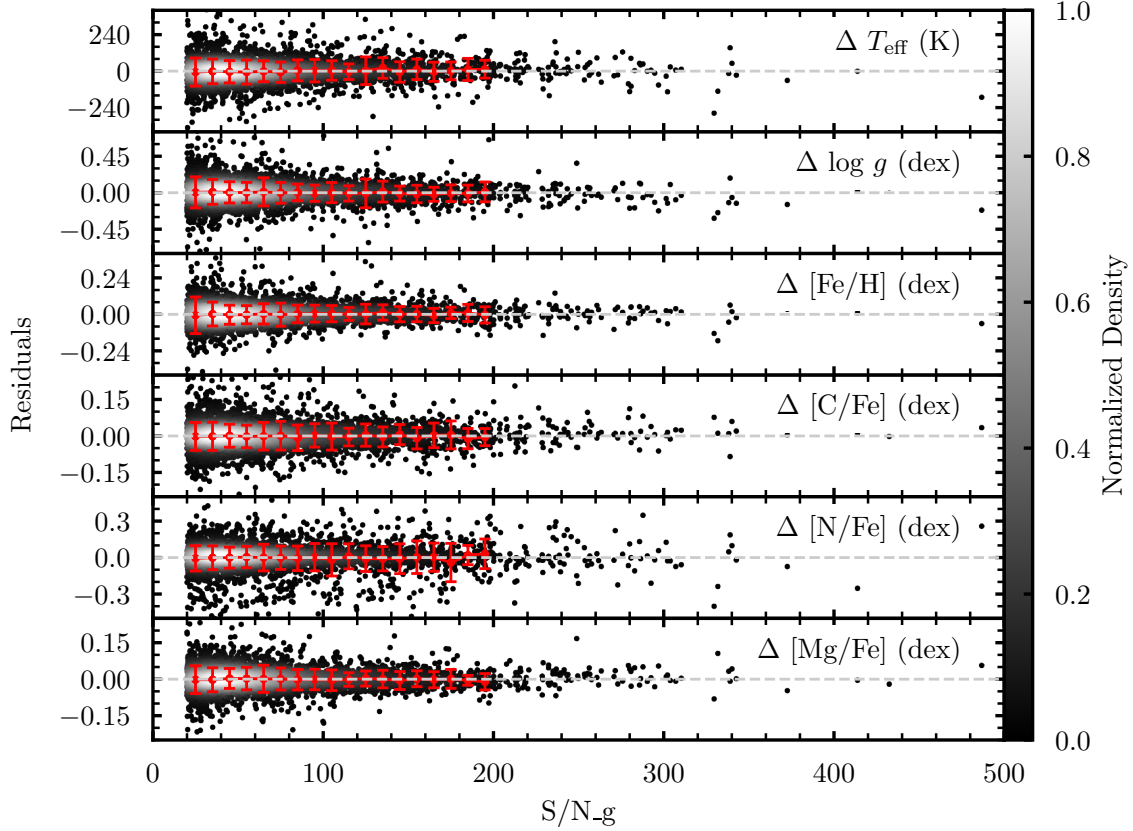


Figure 6. Density scatter plots of residuals vs. g -band signal-to-noise ratio (S/N_g) for FCResNet predictions compared to APOGEE labels on the test set. From top to bottom, the panels correspond to T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, respectively. The red error bars represent the mean and standard deviation of the residuals within each S/N_g bin. The bin width is 10, and only bins containing more than 50 data points are shown.

results confirm that the model successfully identifies the Mg I b triplet near 5184 Å, where lower flux values (deeper absorption) correspond to higher SHAP values, indicating the tendency of the model to predict higher $[\text{Mg}/\text{Fe}]$ values. This behavior is consistent with astrophysical principles. Additionally, the model recognizes other spectral features that may exhibit strong statistical correlations with $[\text{Mg}/\text{Fe}]$.

4.3. Comparison of Different Algorithms

To further evaluate the effectiveness of FCResNet, we compared it with several representative machine learning algorithms—K-Nearest Neighbors (KNN), eXtreme Gradient Boosting (XGBoost), SVR and CNN—each with distinct strengths. KNN offers simplicity and interpretability; XGBoost provides powerful structured-data performance via ensemble learning; SVR emphasizes robustness and margin-based regression; and CNN is well-suited for extracting local features in high-dimensional sequential data.

The above algorithms were implemented using Python libraries. KNN and SVR were built using the `Scikit-learn` library, XGBoost was implemented with the `xgboost` library, and CNN was constructed using `PyTorch`.

For KNN, XGBoost and SVR, we employed Bayesian optimization to search for the optimal hyperparameters based on validation loss. To evaluate the models' ability to predict multiple parameters simultaneously, the loss function was defined as the mean MSE of six normalized stellar parameters, consistent with the setting of FCResNet. The hyperparameter exploration spaces and the optimal configurations are summarized in Table 2. For the CNN model, the network architecture is summarized in Table 3, which includes four convolutional layers, four max-pooling layers, and one FC layer. All activation functions are ReLU. Other settings, including `batch_size`, learning rate, and optimizer, are kept consistent with those used in FCResNet. The best-performing model is selected based on the validation loss.

The comparison results of five trained models—KNN, XGBoost, SVR, CNN, and FCResNet—on both the validation and test sets are summarized in Table 1. Among them, FCResNet and CNN significantly outperform traditional machine learning methods in terms of prediction precision. Specifically, FCResNet achieves the lowest overall loss (0.1608 on the validation set and 0.1792 on the test set), and attains the highest precision in all six stellar parameters,

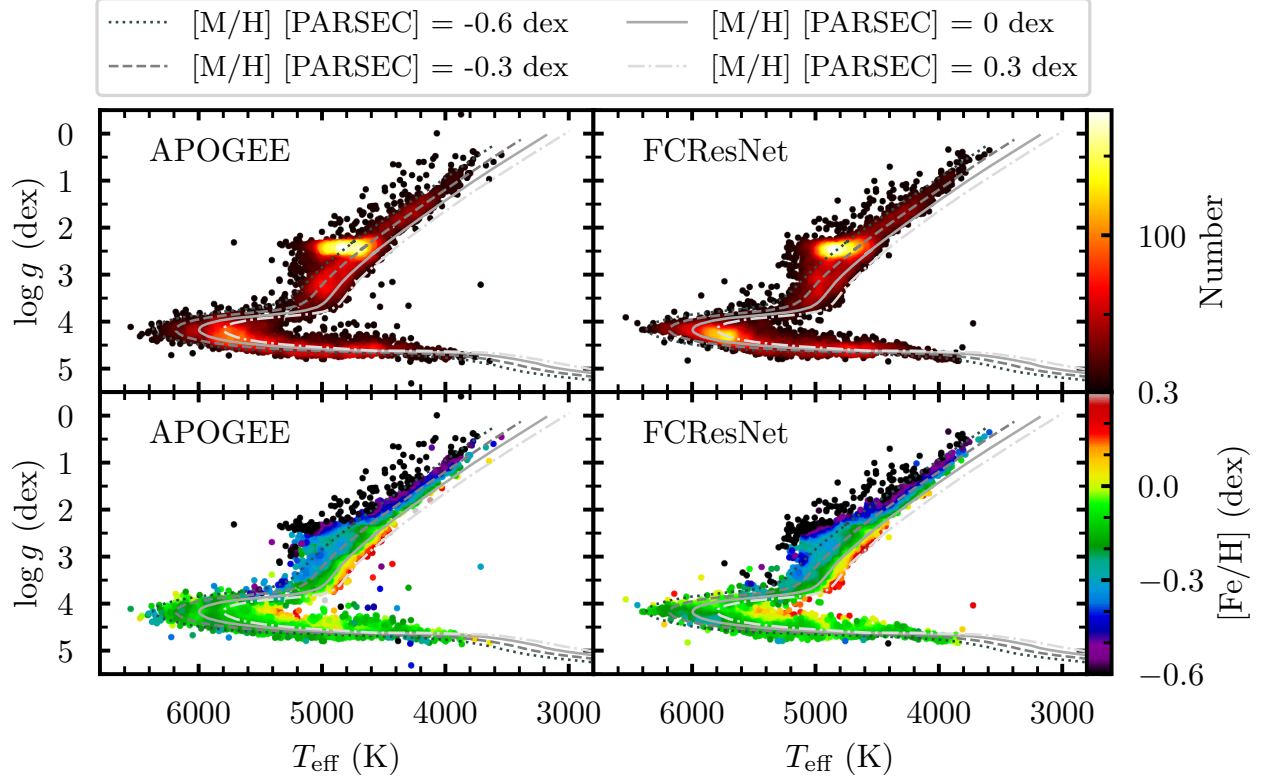


Figure 7. Kiel diagrams of APOGEE labels (left) and FCResNet predictions (right) on the test set. The color bar in the upper panels represents the number of stars within each grid cell, while that in the lower panels represents the mean $[\text{Fe}/\text{H}]$ within each grid cell. The grid widths are 100 K for T_{eff} and 0.2 dex for $\log g$, respectively. The gray lines represent 7 Gyr isochrones generated using the theoretical models from the PAdova and TRieste Stellar Evolution Code (PARSEC; Bressan et al. 2012), with metallicities of -0.6 , -0.3 , 0 , and 0.3 dex.

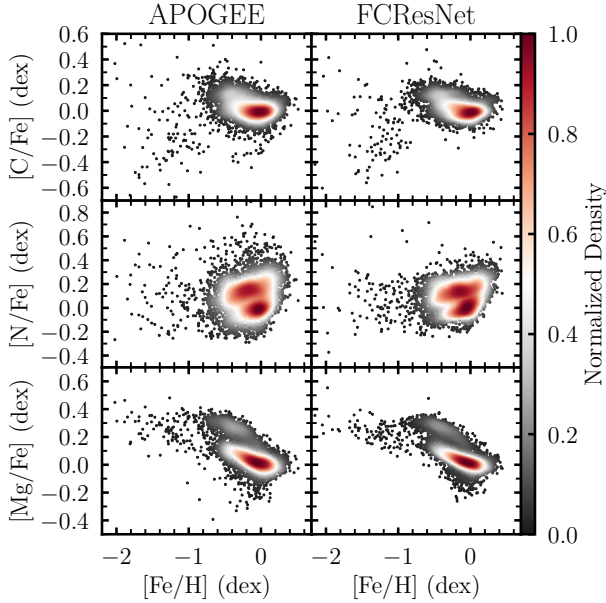


Figure 8. Density distributions of $[\text{X}/\text{Fe}]$ (where $\text{X} = \text{C}, \text{N}, \text{Mg}$) vs. $[\text{Fe}/\text{H}]$ for APOGEE labels (left) and FCResNet predictions (right) on the test set.

slightly outperforming CNN in T_{eff} . In contrast, KNN, XGBoost, and SVR exhibit noticeably higher dispersions, particularly in T_{eff} and $\log g$, with dispersions exceeding 82–140 K and 0.17–0.32 dex, respectively. This degradation in performance is likely due to the high dimensionality (143) of the input spectrum, which challenges traditional models that rely on feature engineering. These methods require manual selection or transformation of features and often struggle to extract complex patterns from high-dimensional data, leading to reduced precision. CNN alleviates the need for manual feature construction and is particularly effective at handling high-dimensional data by automatically extracting hierarchical features. However, when the spectral dimension is relatively low (143), the advantages of CNN become less pronounced, and its complex architecture may not yield significant performance gains. In terms of simplicity, FCResNet is the most lightweight model, with a size of only 348 KB, compared to 943 KB for CNN and over 400 MB for XGBoost. It is worth noting that hyperparameter tuning for XGBoost and SVR was extremely time-consuming, yet did not yield significant improvements in precision. In terms of efficiency, we also

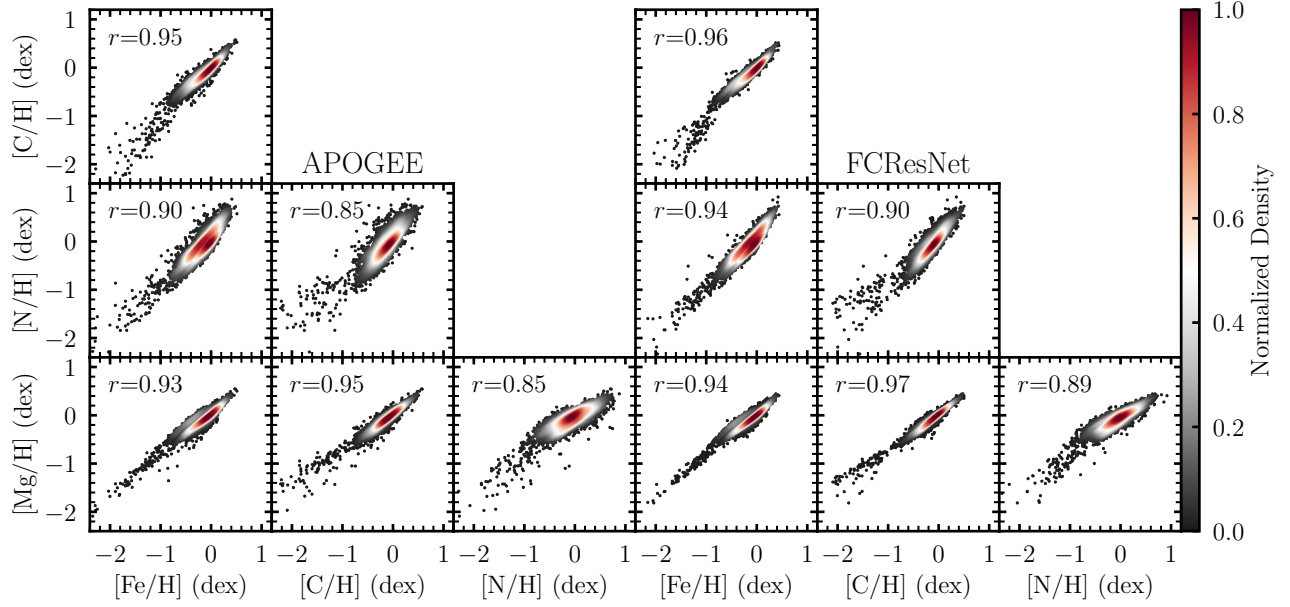


Figure 9. Pairwise density distributions of $[X/H]$ (where $X = \text{Fe, C, N, Mg}$) for APOGEE labels (left) and FCResNet predictions (right) on the test set. Note that for FCResNet, $[C/H]$, $[N/H]$, and $[Mg/H]$ are derived from the directly predicted $[X/Fe]$ and $[Fe/H]$ values using the relation $[X/H] = [X/Fe] + [Fe/H]$. The Pearson correlation coefficient (r) is displayed in the upper left corner of each panel.

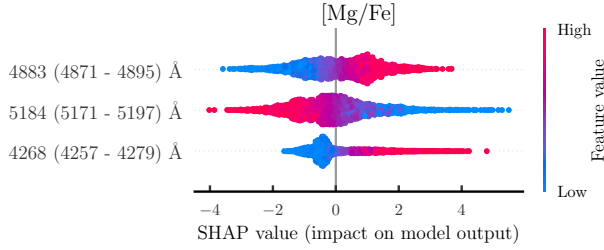


Figure 10. SHapley Additive exPlanations (SHAP) summary plot for $[Mg/Fe]$ predictions by FCResNet on the test set. Each row represents a specific wavelength region, with the horizontal position indicating the SHAP value (impact on model output) and colors representing the feature values (blue for low flux, red for high flux).

evaluated the execution time required to simultaneously predict six stellar parameters for one million spectra. FCResNet and CNN were tested on an NVIDIA RTX 3070 GPU, while KNN, XGBoost and SVR were run with multi-core parallelism on an Intel Core i7-10700K CPU. KNN is the fastest (30 s), but at the cost of the lowest precision. FCResNet, in contrast, achieves the best balance between precision and speed, completing the task in only 42 s.

Overall, FCResNet demonstrates outstanding performance on precision, simplicity and efficiency, making it a highly reliable model for simultaneously estimating multiple stellar parameters from large-scale ultra-low-resolution CSST spectra.

Table 2
The Hyperparameter Exploration Spaces and the Corresponding Optimal Configurations for KNN, XGBoost and SVR, Determined through Bayesian Optimization

Hyperparameter	Exploration Space	Optimal Configuration
KNN		
n_neighbors	[1, 20]	9
weights	{uniform, distance}	distance
algorithm	{ball_tree, kd_tree, brute}	kd_tree
metric	{euclidean, manhattan, chebyshev}	chebyshev
XGBoost		
n_estimators	[1000, 50,000]	20,000
learning_rate	(0, 1]	0.01
max_depth	[1, 20]	7
min_child_weight	[1, 20]	16
gamma	[0, 10]	0
reg_alpha	[0, 10]	2
reg_lambda	[0, 10]	0.5
subsample	(0, 1]	0.3
colsample_bytree	(0, 1]	0.7
SVR		
C	[1, 100]	10
epsilon	[0.01, 0.20]	0.12
gamma	[0.01, 0.10]	0.07

Table 3
The CNN Model Architecture Adopted for Comparison Experiments

Layer	Feature Map	Kernel Size	Stride	Padding	Shape	Activation
Input	1	...	...	...	1×143	...
Convolution	10	1×3	1	1	10×143	ReLU
Max Pooling	10	1×2	1	0	10×142	...
Convolution	20	1×3	1	1	20×142	ReLU
Max Pooling	20	1×2	2	0	20×71	...
Convolution	40	1×3	1	1	40×71	ReLU
Max Pooling	40	1×2	1	0	40×70	...
Convolution	50	1×3	1	1	50×70	ReLU
Max Pooling	50	1×2	2	0	50×35	...
Flatten	1	...	...	...	1×1750	...
Fully Connected	1	...	...	...	1×128	ReLU
Output	1	...	...	...	1×6	...

5. Summary

In this work, we explored a neural network-based model, FCResNet, for simultaneously estimating multiple stellar parameters (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$) from CSST-like spectra with a resolution of $R = 200$. The main results are as follows:

1. Data set construction and model architecture: We constructed a high-quality data set by cross-matching 22,632 low-resolution LAMOST spectra (degraded to $R \sim 200$ to simulate CSST observations) with corresponding stellar parameters from the high-resolution APOGEE survey. The data set was randomly divided into training (60%), validation (20%), and test (20%) sets. FCResNet combines fully connected neural networks with residual blocks to address gradient vanishing/exploding problems and model degradation issues encountered in deep architectures. The model consists of an input layer, two residual blocks, two fully connected layers, and an output layer.
2. Superior performance and efficiency: We implemented and compared FCResNet against baseline models including KNN, XGBoost, SVR, and CNN on the test set for spectra with $S/N_g > 20$. FCResNet achieves the highest prediction precision, with precisions of 78 K, 0.15 dex, 0.08 dex, 0.05 dex, 0.10 dex, and 0.05 dex for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, and $[\text{Mg}/\text{Fe}]$, respectively. FCResNet demonstrates significant improvements in overall loss: 4.3% better than CNN, 14.8% better than SVR, 23.7% better than XGBoost, and

56.8% better than KNN. In computational efficiency, FCResNet requires only 42 s to process one million spectra, representing speed improvements of 1.1 times faster than CNN, 38.6 times faster than SVR, and 8.6 times faster than XGBoost, while being slightly slower than KNN but with dramatically higher precision. Moreover, FCResNet is the most lightweight and easy-to-train model with only 348 KB size, compared to 943 KB for CNN, 54 MB for SVR, 405 MB for XGBoost and 33 MB for KNN. While traditional machine learning algorithms require time-consuming hyperparameter tuning with limited precision gains, and CNN achieves similar precision but relies on complex architecture and extensive tuning, FCResNet attains a strong balance between simplicity, speed, and precision, making it highly suitable for large-scale deployment in future CSST operations.

3. Physical interpretability and validation: Kiel diagram comparisons and abundance pattern analyses show excellent agreement between FCResNet predictions and APOGEE labels, as well as consistency with PARSEC theoretical isochrones. SHAP analysis confirms that FCResNet successfully identifies physically meaningful spectral features, such as the MgI b triplet near 5184 Å for $[\text{Mg}/\text{Fe}]$ predictions, demonstrating that the model learns underlying astrophysical patterns rather than merely exploiting statistical correlations between parameters.

Acknowledgments

This work is supported by the National Astronomical Observatories of Chinese Academy of Sciences (No. E4ZR0516)

and the National Natural Science Foundation of China (12273078, 12273075 and 12411530071). We also acknowledge support from Royal Society IEC\NSFC\233140 exchange grant. Guo Shou Jing 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. 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 acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org.

This research also makes use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013) and the TOPCAT tool (Taylor 2005).

Facilities: LAMOST, Sloan.

ORCID iDs

Shuo Li  <https://orcid.org/0000-0002-8913-3605>
 Yin-Bi Li  <https://orcid.org/0000-0001-7607-2666>
 A-Li Luo  <https://orcid.org/0000-0001-7865-2648>
 Jun-Chao Liang  <https://orcid.org/0009-0001-9085-8718>
 Hai-Ling Lu  <https://orcid.org/0000-0002-8252-8743>
 Hugh R. A. Jones  <https://orcid.org/0000-0003-0433-3665>

References

- Arulampalam, G., & Bouzerdoun, A. 2003, *NN*, 16, 561
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Bengio, Y., Simard, P., & Frasconi, P. 1994, *ITNN*, 5, 157
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, 154, 28
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Cooper, A. P., Kposov, S. E., Allende Prieto, C., et al. 2023, *ApJ*, 947, 37
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, 12, 1197
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, 12, 735
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, 449, 2604
- Fabbro, S., Venn, K. A., O'Briain, T., et al. 2018, *MNRAS*, 475, 2978
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, 151, 144
- Gray, D. F., & Johanson, H. L. 1991, *PASP*, 103, 439
- He, K., & Sun, J. 2015, in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2015 (Boston, MA, USA: IEEE), 5353
- He, K., Zhang, X., Ren, S., & Sun, J. 2016a, in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2016 (Las Vegas, NV, USA: IEEE), 770
- He, K., Zhang, X., Ren, S., & Sun, J. 2016b, in Computer Vision – ECCV 2016, 9908 (Cham: Springer), 630
- Hochreiter, S. 1998, *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 6, 107
- Hornik, K., Stinchcombe, M., & White, H. 1989, *NN*, 2, 359
- Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, *AJ*, 160, 120
- Kingma, D. P., & Ba, J. 2014, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Kirby, E. N., Guhathakurta, P., & Sneden, C. 2008, *ApJ*, 682, 1217
- Kovtyukh, V. V., Soubiran, C., Belik, S. I., & Gorlova, N. I. 2003, *A&A*, 411, 559
- Lee, Y. S., Beers, T. C., Allende Prieto, C., et al. 2011, *AJ*, 141, 90
- Li, X., & Lin, B. 2023, *MNRAS*, 521, 6354
- Liang, S., & Srikant, R. 2016, [arXiv:1610.04161](https://arxiv.org/abs/1610.04161)
- Liu, C., Fang, M., Wu, Y., et al. 2015, *ApJ*, 807, 4
- Luo, A. L., Zhang, H.-T., Zhao, Y.-H., et al. 2012, *RAA*, 12, 1243
- Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, *RAA*, 15, 1095
- Lustig-Yaeger, J., Robinson, T., & Arney, G. 2019, *JOSS*, 4, 1387
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Pascanu, R., Mikolov, T., & Bengio, Y. 2012, [arXiv:1211.5063](https://arxiv.org/abs/1211.5063)
- Simonyan, K., & Zisserman, A. 2014, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
- Smith, V. V., Bizyaev, D., Cunha, K., et al. 2021, *AJ*, 161, 254
- Sousa, S. G., Santos, N. C., & Israelian, G. 2012, *A&A*, 544, A122
- Srivastava, R. K., Greff, K., & Schmidhuber, J. 2015a, [arXiv:1505.00387](https://arxiv.org/abs/1505.00387)
- Srivastava, R. K., Greff, K., & Schmidhuber, J. 2015b, [arXiv:1507.06228](https://arxiv.org/abs/1507.06228)
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, 132, 1645
- Su, D.-Q., & Cui, X.-Q. 2004, *ChJAA*, 4, 1
- Svozil, D., Kvasnicka, V., & Pospichal, J. 1997, *Chemometrics and Intelligent Laboratory Systems*, 39, 43
- Szegedy, C., Liu, W., Jia, Y., et al. 2014, [arXiv:1409.4842](https://arxiv.org/abs/1409.4842)
- Taylor, M. B. 2005, in APS Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, 879, 69
- Valentini, M., Chiappini, C., Davies, G. R., et al. 2017, *A&A*, 600, A66
- Wang, L., Wang, W., Wu, Y., et al. 2016, *AJ*, 152, 6
- Wu, T., Bu, Y., Xie, J., et al. 2024, *PASA*, 41, e002
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, *ApJS*, 245, 34
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, *AJ*, 146, 81
- Zhang, B., Liu, C., & Deng, L.-C. 2020, *ApJS*, 246, 9
- Zhang, X., Zhao, G., Yang, C. Q., Wang, Q. X., & Zuo, W. B. 2019, *PASP*, 131, 094202
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *RAA*, 12, 723