ESNet: Estimating Stellar Parameters from LAMOST Low-Resolution Stellar Spectra

Kun Wang 1, Bo Qiu 1, *, A-li Luo 2,3, *, Fuji Ren 4 and Xia Jiang 1

1 School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; wk2021319@163.com (K.W.); xunjiang@hebut.edu.cn (X.J.)
2 CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, China
3 University of Chinese Academy of Sciences, Beijing 100049, China
4 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610056, China; ren2fuji@gmail.com
* Correspondence: qiubo@hebut.edu.cn (B.Q.); lal@bao.ac.cn (A.-l.L.)

Abstract: Stellar parameters are estimated through spectra and are crucial in studying both stellar evolution and the history of the galaxy. To extract features from the spectra efficiently, we present ESNet (encoder selection network for spectra), a novel architecture that incorporates three essential modules: a feature encoder (FE), feature selection (FS), and feature mapping (FM). FE is responsible for extracting advanced spectral features through encoding. The role of FS, on the other hand, is to acquire compressed features by reducing the spectral dimension and eliminating redundant information. FM comes into play by fusing the advanced and compressed features, establishing a nonlinear mapping between spectra and stellar parameters. The stellar spectra used for training and testing are obtained through crossing LAMOST and SDSS. The experimental results demonstrate that for low signal-to-noise spectra (0–10), ESNet achieves excellent performance on the test set, with mean absolute error (MAE) values of 82 K for $T_{\text{eff}}$ (effective temperature), 0.20 dex for $\log g$ (logarithm of the gravity), and 0.10 dex for $[\text{Fe/H}]$ (metallicity). The results indeed indicate that ESNet has an excellent ability to extract spectral features. Furthermore, this paper validates the consistency between ESNet predictions and the SDSS catalog. The experimental results prove that the model can be employed for the evaluation of stellar parameters.

Keywords: data analysis; statistical; stellar fundamental parameters

1. Introduction

With the advancement of spectroscopic surveys, a vast amount of stellar spectra have been acquired. The SEG YE plan by SDSS involved sampling stars from different distances to capture spatial substructures in the Galactic stellar halo [1]. The Anglo-Australian Observatory utilized the Schmidt telescope to determine the radial velocity and parameters of stars [2]. LAMOST launched its low-resolution spectral survey in 2012, amassing over 10 million spectra with $R=1800$ (wavelength range $3690–9100$ Å) [3]. SDSS’s APOGEE project focused on investigating Red Giants, and it provided spectra with $R = 22,500$ (wavelength range $3800–9200$ Å) [4]. Gaia-ESO obtained approximately 100,000 stellar spectra (wavelength range $3700–9000$ Å), providing valuable kinematic and elemental abundance data [5]. From the spectra, important information such as the age, temperature, and metal abundances (Fe, Mg, Al, etc.) of the objects can be extracted, which is significant for the study of celestial evolution and the understanding of the history of galaxies [6–9].

Traditionally, stellar parameters were determined by comparing observed spectra with spectral templates. For instance, Elsner introduced a weighted nearest neighbor template matching method [10]. Meanwhile, the equivalent width method is a traditional method for stellar parameters’ and chemical abundances’ determination [11]. There are also some
other methods, such as MOOG [12] and SME [13]. However, traditional methods demand high-quality spectra.

An alternative mathematical approach was employed to evaluate the stellar parameters, specifically utilizing line indices. This method established a nonlinear mapping relation between the spectral line features and the stellar parameters, such as the new Lick Balmer index [14], the Rose index [15], and the SDDD-based Lick/SDSS index [16]. However, the line indices method also necessitates high-quality spectra (SNR > 20) to produce reliable outcomes.

With the advancements in machine learning, processing vast quantities of spectra in bulk has become feasible. For instance, Cannon utilized data-driven approaches to evaluate stellar parameters like \( T_{\text{eff}} \) (effective temperature), \( \log g \) (logarithm of the gravity), and metal abundances ([Fe/H] (metallicity)) for stars [17]. Cannon2 evaluated \( T_{\text{eff}}, \log g, \) and 15 elemental abundances (C, Mg, Al, etc.) of Red Giants [18]. CNN-based neural networks were also employed to measure stellar parameters (\( T_{\text{eff}}, \log g, [\text{Fe/H}] \)) by using a synthetic training set [19]. There were some methods that used observed data for the training (as in the current study), such as [20,21]. Additionally, some parametric regression models based on traditional regression methods to evaluate \( (T_{\text{eff}}, \log g, [\text{Fe/H}]) \), such as SLAM [22] and SCDD [23], were proposed. Ref. [24] played a pioneering role in employing neural networks to estimate stellar \( T_{\text{eff}}, \log g, \) and \([M/H]\) across various SNR and spectral resolutions, encompassing even very low-resolution spectra. Similarly, ref. [25] introduced deep learning networks to predict stellar parameters using SDSS spectra, while ref. [26] concentrated on LAMOST DR5 spectra and introduced generative spectrum networks for stellar parameter estimation.

For low SNR spectra, it has been a challenge to extract effective features from them to estimate stellar parameters. Based on machine learning methods, both refs. [27,28] estimate stellar parameters for low SNR spectra, but the results are not satisfactory for SNRs between 0 and 10.

Existing methods cannot effectively remove spectral redundant information (mainly includes noise, features that are not sensitive to stellar parameters) and extract spectral features from low SNR spectra. Thus, this paper proposes ESNet (encoder selection network for spectra), as a new approach to evaluating stellar parameters.

The organization of this paper is as follows: Section 2 introduces the data and pre-processing; Section 3 presents the method; Section 4 presents the experiments and results; Section 5 presents the discussion; Section 6 presents the conclusions.

2. Data

This section provides a detailed description of the data used in this paper, which includes information about its sources, the criteria used for selection, and the pre-processing steps.

2.1. Sources

Data are from LAMOST and SDSS. LAMOST provides spectra, and SDSS provides stellar parameters.

2.1.1. LAMOST Survey

LAMOST [29] currently provides spectra for approximately one million objects, encompassing stars, galaxies, and quasars. These spectra offer crucial information concerning the physical properties, chemical composition, and motions of these objects. The wavelength range covered by the LAMOST spectra spans from 3690 to 9100 Å, and the spectra have a resolution of R 1800 [30]. In one of its data releases, LAMOST DR10, around eleven million low-resolution spectra were made available, with stellar spectra accounting for 97.1% of the total number. Among these stellar spectra, approximately 7.4 million are utilized for the evaluation of stellar parameters.
2.1.2. SDSS Survey

SDSS [31] houses an extensive collection of astronomical data, spanning a vast region of the sky. These data include essential information such as the position, brightness, and spectral details of billions of celestial objects, along with their physical properties. SDSS also offers a wide range of data, including images, galaxy redshift measurements, galaxy photometry, and a comprehensive catalog of sky survey objects. Additionally, the wavelength range covered by the SDSS DR17 APOGEE spectra with a resolution of R = 22500 [32] spans from 3800 to 9200 Å, and high-quality atmospheric parameters \([T_{\text{eff}}, \log g, [Fe/H], [\text{alpha}/Fe]]\) are provided. Estimating atmospheric parameters using high-resolution spectra is more accurate [28]. Thus, in the paper, APOGEE’s labels are used. The APOGEE Stellar Parameter and Chemical Abundances Pipeline (ASPCAP) [33] is a software pipeline for the analysis of stellar spectra, mainly used to process spectra from the Apache Point Observatory Galactic Evolution Experiment (APOGEE). ASPCAP aims to extract stellar parameters (e.g., \(T_{\text{eff}}, \log g, \) metal abundances, etc.) and chemical abundances (e.g., elemental abundances) from stellar spectra.

2.2. Selection

Spectra that have a spherical distance less than 3 arcsec are considered to belong to the same target source, and this criterion is used for the selection of spectra [34]. The data are obtained by cross-matching the LAMOST data with SDSS data using the Tool for Operations on Catalogs and Tables (TOPCAT) [35]. The spectra are then further selected based on the following limitations [23]: (1) \(4000 \leq T_{\text{eff}} \leq 8000\); (2) \(0.5 \leq \log g \leq 5.5\); (3) \(-2.5 \leq [Fe/H] \leq 1.0\); (4) \(T_{\text{eff}}_{\text{err}} < 200\); (5) \(\log g_{\text{err}} < 0.1\); (6) \([Fe/H]_{\text{err}} < 0.1\). After applying these selection criteria, there are a total of 129,108 spectra available for training and testing.

2.3. Pre-Processing

To enhance the usability of spectral information, proper processing of stellar spectra is necessary, especially considering the impact of extinction and instrument noise [28]. The specific details of the processing steps are as follows:

(1) Converting wavelength provided by LAMOST to rest frame. This will eliminate the effect of the radial velocity on spectra, as shown in Equation (1).

\[
\lambda' = \frac{\lambda}{1 + \frac{RV}{C}}
\]

where \(\lambda'\) is the wavelength in the rest frame, \(\lambda\) is the wavelength provided by LAMOST, \(RV\) is the radial velocity, and \(C\) is the speed of light.

(2) Handling bad fluxes. Using the masks provided by LAMOST, the unreliable flux values are substituted with the corresponding continuum values, instead of being deleted.

(3) Resampling. To construct the dataset, the spectra are resampled by the spline interpolation method with a step of 0.0001 in logarithmic wavelength coordinates.

(4) Normalizing spectra. Due to the uncertainty in the flux calibration and unknown extinction values in the LAMOST stellar spectra, it is necessary to normalize the spectra. We map the spectral fluxes to values between 0 and 1, so that spectra can be compared and processed on the same scale. This normalization method is very suitable for machine learning.

\[
F_{\text{new}} = \frac{F - \min(F)}{\max(F) - \min(F)}
\]

where \(F\) is spectral flux, and \(F_{\text{new}}\) is the normalized spectra.

3. Method

This section describes the network structure, including the feature encoder (FE) module, feature selection (FS) module, and feature mapping (FM) module.
3.1. Network Structure

This section provides a detailed description of the network structure, as depicted in Figure 1. ESNet is composed of three main modules: the FE module, the FS module, and the FM module. The FE module is constructed using an encoder architecture inspired by encoder–decoder. It encodes the spectra by mapping them to high-level features, thereby achieving dimension reduction. The encoder first encodes the original spectrum as a vector; to ensure that the vector can represent the original spectrum, the vector is then restored to the original spectrum by the decoder. The encoder and decoder consist of three linear layers, but their structures are symmetric. In Section 4.1, we compare the effect of whether the encoder uses convolutional layers or does not on the prediction results, and the results prove that the performance is better using only linear layers, which may be mainly due to the fact that the linear layers can pay sufficient attention to the features of the spectra. The FS module employs the LASSO method to extract spectral features and effectively remove redundant information from the spectra. The FM module is based on a Multi-Layer Perceptron (MLP) neural network. It combines the features extracted by the FE and FS modules and establishes a nonlinear mapping between the features and the stellar parameters. We compare the effects of different layers on the prediction results, and it is found that the performance is better when the FM module uses a four-layer structure. The main reason may be that the four-layer structure can sufficiently extract features while ensuring the generalization ability of the network.

![Figure 1. The structure of the ESNet. It consists of FE module, FS module, and FM module. FE and FS are trained separately. Decoder works when training FE and does not participate in testing.](image)

3.1.1. The FE Module

The FE module utilizes an encoder architecture inspired by encoder–decoder [36]. Encoder–decoder is a pair of interconnected modules commonly used in deep learning and neural networks for data transformation. The encoder takes the input data, such as spectra, and maps them to a higher-level representation space through a series of processing and transformation operations, effectively achieving dimensionality reduction. By doing so, the encoder captures essential features present in the input data and compresses them into a higher-level feature vector.

The decoder, on the other hand, performs the inverse process of the encoder. It takes the higher-level feature vector generated by the encoder and reconstructs the original data. The primary goal of the decoder is to restore the original data as accurately as possible, minimizing the difference between the reconstructed data and the original data.

During the training stage, the encoder is used to map the spectra into higher-level feature vectors, and then the decoder reconstructs the original spectra from these feature vectors. This means that the higher-level feature vectors effectively represent the essential
information present in the original spectra. During the testing stage, only the encoder works, mapping the spectra into higher-level feature vectors.

By leveraging the encoder–decoder, the FE module is capable of capturing crucial spectral features and achieving effective dimensionality reduction.

3.1.2. The FS Module

The FS module utilizes the Least Absolute Shrinkage and Selection Operator (LASSO) method to select spectral features, eliminate redundant information, and reduce the dimensionality of the spectra.

LASSO is a statistical method employed for linear regression and feature selection [37]. It accomplishes model sparsity and variable selection by imposing constraints on the regression coefficients. The penalty function of LASSO involves minimizing the residual sum of squares (RSS) with an L1-norm penalty term. The expression of the penalty function is shown in the following equation:

$$B_{\text{LASSO}} = \arg \min_B \left\{ \left| \left| Y - \sum_{j=1}^{p} X_j B_j \right| \right|^2 + a \sum_{j=1}^{p} |B_j| \right\}$$

where $Y$ represents the regression term, $X_j$ stands for the input, $B_j$ represents the coefficient, and $a$ is the adjustment parameter. By manipulating the parameter $a$, the overall regression coefficients can be compressed effectively. Through the utilization of the L1 parametric penalty term, some regression coefficients can be driven to zero, resulting in spectral dimension reduction and the elimination of redundant information.

As a result, redundant information in the spectra can be efficiently eliminated, while retaining spectral features that exhibit sensitivity to stellar parameters.

3.1.3. The FM Module

The FM, or feature mapping, is a crucial component in the ESNet, and it consists of a Multi-Layer Perceptron (MLP) [38]. The MLP is chosen for its exceptional ability to achieve high levels of parallel processing and its strong adaptive and self-learning functions, attributed to its highly nonlinear global action. As an artificial neural network model, MLP is one of the most fundamental models.

The typical architecture of MLP comprises an input layer, multiple hidden layers, and an output layer. The input layer receives the original data, while the hidden layers conduct nonlinear transformations and extract essential features from the input data. Finally, the output layer yields the ultimate result. By leveraging the power of MLP, the FM effectively maps the features extracted through FS and FE to the corresponding stellar parameters.

4. Results

This section describes the experiment and the results. Mean absolute error (MAE) is a statistical metric used to assess the average absolute difference between predicted values and actual observations. It is utilized to evaluate the accuracy of models in predicting outcomes. Thus, in this paper, MAE is employed as an evaluation metric.

4.1. Selection of Model Structure and Estimation of Model Uncertainty

In order to choose a better model structure, we conduct comparison experiments, and the results are shown in Table 1. Firstly, we compare the activation functions (ReLU, Logistic, and Tanh), and the best result is obtained by using Logistic; secondly, we compare the loss functions (L1loss and L2loss), and the experimental results show that the loss functions do not have a significant effect on the model performance. We also compare the effect of different numbers of layers of the FM, and the results show that the network is best when the number of layers is four. This may be because the network structure with three layers cannot fully extract the features, while the network with five layers can fully extract the features, but the generalization performance is not good.
Table 1. Comparative experiments for selection of model structures.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Name/Count</th>
<th>MAE ($T_{eff}$)</th>
<th>MAE (log g)</th>
<th>MAE ([Fe/H])</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReLU</td>
<td>355</td>
<td>0.09</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>Logistic</td>
<td>49</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Tanh</td>
<td>51</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>Loss</td>
<td>L1loss</td>
<td>49</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>L2loss</td>
<td>49</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>Linear</td>
<td>3 layers</td>
<td>49</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>4 layers</td>
<td>49</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>5 layers</td>
<td>49</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>FE</td>
<td>Conv(yes)</td>
<td>49</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Conv(no)</td>
<td>49</td>
<td>0.07</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note: Bold represents the best results in the comparison experiment.

For the encoder module, we compare the effect of using convolutional layers and not using them on the performance of the model, and the results show that the performance of the network without convolutional layers is better, which may be mainly due to the fact that the linear layer can focus on the global information of the spectra compared to the convolutional layer.

The uncertainty estimates for the model are shown in Table 2. First, we train eight models and then estimate the stellar parameters using each of the eight models on the test set. The variance in the predictions of the eight models is calculated as the model uncertainty [21]. The results are as follows: $\mu(T_{eff}) = 49$, $\sigma(T_{eff}) = 0.547$; $\mu(\log g) = 0.07$, $\sigma(\log g) = 0.0008$; $\mu([Fe/H]) = 0.04$, $\sigma([Fe/H]) = 0.0004$.

Table 2. The uncertainty estimates for the model.

<table>
<thead>
<tr>
<th>Structure</th>
<th>$T_{eff}$</th>
<th>log g</th>
<th>[Fe/H]</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean value $\mu$</td>
<td>49</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>standard deviation $\sigma$</td>
<td>0.547</td>
<td>0.0008</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

4.2. Evaluation of Predictions

The dataset is partitioned into three subsets: a training set, a validation set, and a test set, distributed in an 8:1:1 ratio. We sample all the spectra uniformly to obtain the training set. The distribution of the spectral SNR for the training set is shown in Figure 2. The distribution of the number of $T_{eff}$, log $g$, [Fe/H] in the training set is shown in Figure 3. Finally, the number of metal-poor stars is 181 (accounting for 0.2%). At training time, we process the $T_{eff}$ in the labels by taking logarithms.

Figure 2. Distribution of training samples along with SNR_g.
For the ESNet model, the spectra are used as inputs, while the corresponding stellar parameters serve as labels. Spectral features are extracted using the FE and FS modules. The features extracted by the FE and FS modules are fused. Subsequently, the FM maps the fused features to the stellar parameters.

The ESNet model’s performance is evaluated using the test set. The model’s predictions for the stellar parameters are compared against the labels (SDSS). The comparison between the ESNet’s predictions and the actual labels (SDSS) is shown in Figure 4. The comparison between the ESNet’s predictions and the actual labels (SDSS) using the \([Fe/H]\) color code with respect to the \(T_{\text{eff}} - \log g\) diagram is shown in Figure 5. It is evident from the figure that the distribution of the ESNet’s predictions closely resembles that of the SDSS labels. It indicates that the ESNet model achieves promising results in predicting the stellar parameters.

The consistency between the ESNet predictions and the labels is thoroughly analyzed and evaluated. Scatter density plots for the stellar parameters \(T_{\text{eff}}, \log g, [Fe/H]\) are
presented in Figure 6, providing a visual representation of the relationship between the ESNet predictions and the labels.

**Figure 5.** The comparison between the ESNet predictions and the actual labels (SDSS) using $[Fe/H]$ color code with respect to the $T_{\text{eff}}$-$\log g$ diagram. The left one is the ESNet predictions, and the right one is the labels (SDSS catalog). Color bar represents the density of data points with a change in color. Larger values on the color bar indicate greater density.

**Figure 6.** Consistency between the prediction (ESNet) and SDSS catalog. The upper left figure refers to $T_{\text{eff}}$; the upper right figure refers to $\log g$; the bottom left figure refers to $[Fe/H]$. Color bar represents the density of data points with a change in color. Larger values on the color bar indicate greater density.

Additionally, histograms depicting the differences between the ESNet predictions and the labels are shown in Figure 7. The experimental results reveal that the diagonal
distributions in the scatter density plots (Figure 6) align closely with the 45° line, indicating that there is no significant systematic deviation between the ESNet predictions and the labels, and the ESNet model exhibits good consistency in predicting the stellar parameters.

![Figure 7](image_url)

Figure 7. Distribution of residual error between predictions and SDSS catalog. The upper left figure refers to \( T_{\text{eff}} \); the upper right figure refers to \( \log g \); the bottom left figure refers to \( [Fe/H] \). To further analyze the consistency, we calculate two statistical metrics: the mean value (\( \mu \)) and the mean square error (\( \sigma \)) of the differences between the ESNet predictions and the labels for each stellar parameter (\( T_{\text{eff}}, \log g, \) and \( [Fe/H] \)). The results are as follows:

\[
\begin{align*}
\mu(T_{\text{eff}}) &= -3.12, \\
\sigma(T_{\text{eff}}) &= 83.18; \\
\mu(\log g) &= 0.002, \\
\sigma(\log g) &= 0.125; \\
\mu([Fe/H]) &= 0.00002, \\
\sigma([Fe/H]) &= 0.05800.
\end{align*}
\]

However, we also note that the ESNet model performs relatively poorly for cold and hot stars. This limitation may be attributed to the lack of spectra for such stars in the SDSS catalog [28], since the stellar parameters in this paper are from the SDSS catalog, which weakens ESNet’s ability to take spectral features of both cool and hot stars; alternatively, the divergence between the two catalogs could potentially have a significant influence.

In conclusion, the analysis and statistical results provide strong evidence for the high and stable consistency between the ESNet predictions and the labels, validating the effectiveness of the proposed approach in estimating stellar parameters from spectra.

4.3. Comparing the LAMOST Catalog with the SDSS One

In order to further validate the accuracy and stability of the ESNet predictions, the paper conducts a comparison between the stellar parameters provided by LAMOST and SDSS. For this comparison, we utilize 12,910 spectra from the test set. The validation results are presented in Figure 8, which includes scatter density plots, and Figure 9, which displays statistical differences between the stellar parameters from LAMOST and SDSS.

We present scatter plots (Figure 8) of the three parameters in the test set, along with histograms of the errors. The scatter plots reveal notable differences between the stellar parameters \( T_{\text{eff}}, \log g, \) and \( [Fe/H] \) provided by LAMOST and those from SDSS, indicating a significant systematic deviation between the two catalogs. By comparing Experiment 4.1,
we find that the ESNet predictions exhibit better agreement with the SDSS catalog than with the LAMOST catalog.

Figure 8. Consistency between the LAMOST catalog and SDSS one. The upper left figure refers to $T_{\text{eff}}$; the upper right figure refers to $\log g$; the bottom left figure refers to $[\text{Fe}/H]$. Color bar represents the density of data points with a change in color. Larger values on the color bar indicate greater density.

For the scatter plot (upper left) in Figure 8, we observe that the $T_{\text{eff}}$ provided by SDSS for cool stars are generally higher than those from LAMOST, while, for hot stars, the $T_{\text{eff}}$ from SDSS are relatively lower. The corresponding histogram reveals a systematic error of $\mu = -34.10$ and a deviation of $\sigma = 154.67$ between the two catalogs concerning $T_{\text{eff}}$.

For the scatter plot (upper right) in Figure 8, a systematic error is evident between the two catalogs, primarily observed in $\log g$ within the range of (0, 2.5), with those from the SDSS catalog being smaller than those from LAMOST. The histogram illustrates a systematic error of $\mu = -0.069$ and a deviation of $\sigma = 0.173$ for $\log g$.

Regarding $[\text{Fe}/H]$, the scatter plot (middle) in Figure 8 indicates an ambiguous systematic error. However, the dispersion between LAMOST and SDSS for $[\text{Fe}/H]$ is greater than the dispersion between the ESNet predictions and SDSS. The corresponding histogram presents a systematic error of $\mu = -0.012$ and a dispersion of $\sigma = 0.090$ between the two catalogs.

In summary, the stellar parameters provided by LAMOST ($T_{\text{eff}}, \log g, [\text{Fe}/H]$) show a significant systematic error when compared to the SDSS catalog.

Compared to the LAMOST and SDSS catalogs, the ESNet predictions exhibit higher consistency with the SDSS catalog. The systematic error and deviation of the ESNet predictions are lower than those ($T_{\text{eff}}, \log g, [\text{Fe}/H]$) provided by LAMOST. As a result, the ESNet model can offer stellar parameters that align more closely with the SDSS catalog. We should note that the more precise and accurate ESNET labels with respect to SDSS are simply due to the fact that SDSS labels are used for training.
4.4. Comparison with Other Methods

To evaluate the effectiveness of our approach, ESNet is compared with several other models, including SLAM [22], SCDD [23], StarN [27] (because its model's name is the same as that in [19], we use StarN to refer to this model), StarNet [19], Catboost [39], Lasso-MLP[28], and DenseNet [40]. The experimental results are presented in Table 3.

Based on the experimental results, the performance of the different models on the test set is as follows.

StarNet: Based on a CNN, the network structure consists of two convolutional layers and three linear layers. Data preprocessing: spectra are interpolated, resampled, and normalized. Training datasets and labels come from synthetic datasets. Limitations: for a low SNR (0–10), where the $T_{\text{eff}}$ performance is 206, $\log g$ is 0.38, and $[\text{Fe/H}]$ is 0.13. However, due to its shallow network structure, StarNet struggles to effectively extract spectral features from spectra with information redundancy, leading to poor results when using the same dataset and labels as we did.

SLAM: The network structure consists of Support Vector Regression (SVR). Data preprocessing: spectra are normalized by dividing its pseudo-continuum, smoothed by smoothing spline, and resampled to the same wavelength grid. Training datasets and labels come from LAMOST. Limitations: when using the same dataset and labels as we did, for low SNR (0–10), where the $T_{\text{eff}}$ performance is 111, $\log g$ is 0.33, and $[\text{Fe/H}]$ is 0.18. SLAM can partially extract spectral features, but its predictions for $\log g$ and $[\text{Fe/H}]$ are not satisfactory. The main reason might be SLAM’s limited ability to sufficiently extract the spectral features associated with $\log g$ and $[\text{Fe/H}]$, when using the same dataset and labels as we did.
Table 3. Comparing the ESNet with other regression methods.

<table>
<thead>
<tr>
<th>SNR</th>
<th>Methods</th>
<th>MAE ($T_{eff}$)</th>
<th>MAE ($\log g$)</th>
<th>MAE ([Fe/H])</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>StarNet</td>
<td>206</td>
<td>0.38</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>111</td>
<td>0.33</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>257</td>
<td>0.36</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>169</td>
<td>0.34</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>157</td>
<td>0.26</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>93</td>
<td>0.21</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>145</td>
<td>0.36</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td>82</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>10–20</td>
<td>StarNet</td>
<td>207</td>
<td>0.35</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>79</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>212</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>123</td>
<td>0.22</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>123</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>77</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>124</td>
<td>0.25</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td>64</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>&gt;20</td>
<td>StarNet</td>
<td>159</td>
<td>0.28</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>48</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>102</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>90</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>104</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>54</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>81</td>
<td>0.13</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td>48</td>
<td>0.06</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note: Bold represents the best results in the comparison experiment.

SCDD: Network structure consists of Catboost. Data preprocessing: spectra are divided into multiple windows and the statistical features of each window are extracted, e.g., mean, maximum, minimum, etc. Training datasets and labels come from LAMOST. Limitations: for low SNR (0–10), where the $T_{eff}$ performance is 169, $\log g$ is 0.34, and $[\text{Fe/H}]$ is 0.15. SCDD splits the spectra into windows, extracts statistical features from each window to remove redundant information, and then fuses the features to train using Catboost. However, this approach causes the significant loss of valid spectral information, resulting in poor performance, when using the same dataset and labels as we did.

StarN: Based on a CNN, the network structure consists of two convolutional layers and three linear layers. Data preprocessing: spectra are resampled and normalized. Training datasets and labels come from a synthetic dataset and LAMOST. Limitations: for low SNR (0–10), where the $T_{eff}$ performance is 257, $\log g$ is 0.36, and $[\text{Fe/H}]$ is 0.17. SCDD splits the spectra into windows, extracts statistical features from each window to remove redundant information, and then fuses the features to train using Catboost. Despite increasing the size of the convolution kernel, due to its shallow network structure, StarN struggles to effectively extract spectral features from spectra with information redundancy, leading to poor results, when using the same dataset and labels as we did.

Catboost: Network structure consists of Catboost. Data preprocessing: spectra are normalized by dividing its pseudo-continuum, smoothed by smoothing spline, and resampled to the same wavelength grid. Training datasets and labels come from LAMOST. Limitations: for a low SNR (0–10), where the $T_{eff}$ performance is 157, $\log g$ is 0.26, and $[\text{Fe/H}]$ is 0.12. Unlike SCDD, this method directly trains Catboost by using spectra without removing redundant spectral information. Despite this, it outperforms SCDD, indicating that SCDD’s removal of valid spectral information hampers its performance, when using the same dataset and labels as we did.
Lasso-MLP: Network structure consists of Lasso and MLP. Data preprocessing: spectra are interpolated, resampled, and normalized. Training datasets and labels come from LAMOST and Payne catalogs. Limitations: for a low SNR (0–10), where the $T_{\text{eff}}$ performance is 93, $\log g$ is 0.21, and $[\text{Fe}/\text{H}]$ is 0.12. For low SNR spectra, the LASSO method does not adequately extract spectral features and some features may be lost. The lost features are important in estimating the $T_{\text{eff}}$.

DenseNet: Based on a CNN, the network structure mainly consists of three dense blocks and two transition blocks. Data preprocessing: spectra are interpolated, resampled, and normalized to the range [0, 1]. Training datasets and labels come from LAMOST. Limitations: for low SNR (0–10), where the $T_{\text{eff}}$ performance is 145, $\log g$ is 0.36, and $[\text{Fe}/\text{H}]$ is 0.15 on the test set. DenseNet proposes a new network structure with increased depth to extract spectral features. It shows some effectiveness compared with StarNet, but redundant information in the spectra can interfere with its feature extraction capability, when using the same dataset and labels as we did.

ESNet: For low SNR (0–10), where the $T_{\text{eff}}$ performance is 82, $\log g$ is 0.20, and $[\text{Fe}/\text{H}]$ is 0.10. This method effectively removes redundant spectral information while retaining essential spectral features, and improves the performance compared to the other methods.

4.5. Validation of ESNet Using Other Catalogs

We cross the data with the Payne catalog [41] and obtain 248 spectra (SNR 0–10) and 968 spectra (SNR 10–20). At the same time, we estimate their stellar parameters and the results are shown in Table 4. Compared to other methods, ESNet performs best. Here, the $T_{\text{eff}}$ performance is 146, $\log g$ is 0.17, and $[\text{Fe}/\text{H}]$ is 0.10 for SNR (0–10).

Table 4. Comparison of predicted results with Payne catalog.

<table>
<thead>
<tr>
<th>SNR</th>
<th>Method</th>
<th>MAE ($T_{\text{eff}}$)</th>
<th>MAE ($\log g$)</th>
<th>MAE ($[\text{Fe}/\text{H}]$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>StarNet</td>
<td>197</td>
<td>0.42</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>161</td>
<td>0.27</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>314</td>
<td>0.36</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>212</td>
<td>0.33</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>213</td>
<td>0.24</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>158</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>202</td>
<td>0.32</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td><strong>146</strong></td>
<td><strong>0.17</strong></td>
<td><strong>0.10</strong></td>
</tr>
<tr>
<td>10–20</td>
<td>StarNet</td>
<td>191</td>
<td>0.37</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>140</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>216</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>175</td>
<td>0.23</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>177</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>138</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>182</td>
<td>0.25</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td><strong>137</strong></td>
<td><strong>0.16</strong></td>
<td><strong>0.07</strong></td>
</tr>
</tbody>
</table>

Note: Bold represents the best results in the comparison experiment.

To further validate the ESNet in estimating stellar parameters using low SNR spectra, we cross it with 2MASS [42] through SIMBAD and obtain 190 (SNR 0–10) and 872 (SNR 10–20) spectra, and the experimental results are shown in Table 5. Compared to other methods, ESNet also performs best. The $T_{\text{eff}}$ performance is 101, $\log g$ is 0.19, and $[\text{Fe}/\text{H}]$ is 0.11 for SNR (0–10).

In summary, the experimental results demonstrate that ESNet, with its ability to remove redundant spectral information while retaining crucial features, outperforms the other models, offering more accurate and reliable predictions for stellar parameters on the test set.
Table 5. Comparison of predicted results with 2MASS.

<table>
<thead>
<tr>
<th>SNR</th>
<th>Method</th>
<th>MAE (T_{\text{eff}})</th>
<th>MAE (\log g)</th>
<th>MAE ([\text{Fe/H}])</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>StarNet</td>
<td>298</td>
<td>0.48</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>SLAM</td>
<td>132</td>
<td>0.28</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>StarN</td>
<td>355</td>
<td>0.35</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>SCDD</td>
<td>163</td>
<td>0.30</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Catboost</td>
<td>113</td>
<td>0.26</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Lasso-MLP</td>
<td>106</td>
<td>0.20</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>DenseNet</td>
<td>153</td>
<td>0.36</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>This method</td>
<td><strong>101</strong></td>
<td><strong>0.19</strong></td>
<td><strong>0.11</strong></td>
</tr>
</tbody>
</table>

| 10–20| StarNet      | 232                     | 0.34           | 0.12                   |
|      | SLAM         | 95                      | 0.19           | 0.12                   |
|      | StarN        | 212                     | 0.24           | 0.12                   |
|      | SCDD         | 122                     | 0.22           | 0.12                   |
|      | Catboost     | 83                      | 0.17           | 0.10                   |
|      | Lasso-MLP    | 82                      | 0.14           | 0.09                   |
|      | DenseNet     | 126                     | 0.25           | 0.12                   |
|      | This method  | **75**                  | **0.14**       | **0.09**               |

Note: Bold represents the best results in the comparison experiment.

5. Discussion

In this paper, we conduct a further investigation into the effect of the spectral SNR on stellar parameters, as depicted in Figure 10. The experimental results clearly demonstrate that the SNR significantly impacts the predictions of stellar parameters. Specifically, higher SNR values lead to more accurate predictions with reduced dispersion. Conversely, lower SNR values result in less accurate stellar parameter predictions with increased dispersion.

Figure 10. The relationship between residual error and SNR. The upper left figure refers to \(T_{\text{eff}}\); the upper right figure refers to \(\log g\); the bottom left figure refers to \([\text{Fe/H}]\). Color bar represents the density of data points with a change in color. Larger values on the color bar indicate greater density.
We analyze the distribution of the SNR in the test set and find that the largest number of spectra fall within the SNR range of 0 to 200, as illustrated in Figure 11.

![Figure 11. Distribution of the number of spectra with respect to the SNR.](image)

We explore the impact of the SNR on stellar parameters ($T_{\text{eff}}$, $\log g$, $[\text{Fe}/\text{H}]$). Comparing the predictions to the SDSS catalog, we observe that the differences are smaller than those between LAMOST and SDSS. Specifically, for $T_{\text{eff}}$, the MAE of the former steadily decreases and converges with increasing SNR, while the latter also exhibits a tendency to converge, although with less stability.

For $\log g$, the MAE of the former consistently decreases and converges as the SNR increases. On the other hand, the MAE of the latter also converges, but at a higher value, indicating larger errors.

Regarding $[\text{Fe}/\text{H}]$, the MAE of the former continuously decreases and converges with increasing SNR. In contrast, the latter does not show a clear trend of convergence, displaying a diffused and unstable pattern, as depicted in Figure 12.

These findings highlight the significance of the SNR in influencing the accuracy and stability of stellar parameter predictions, with higher SNR values leading to more precise and reliable results.

Additionally, we explore the influence of the SNR on the dispersion $\sigma$. The results indicate that the dispersion $\sigma$ between the predictions (including $T_{\text{eff}}$, $\log g$, $[\text{Fe}/\text{H}]$) and the SDSS catalog is smaller than the dispersion between LAMOST and SDSS. The experimental findings demonstrate a higher level of consistency between the predicted stellar parameters and the SDSS catalog, indicating the better stability of ESNet, as illustrated in Figure 13.

Specifically, for $T_{\text{eff}}$, the dispersion $\sigma$ of the former steadily decreases and converges with the increase in SNR. The latter also exhibits a tendency to converge, but with less stability compared to the former. It can be found that there are some jumps observed in the ESNet red curve for $T_{\text{eff}}$. The main reason is that ESNet has a few values with large prediction errors at the jumps (in Figure 10), but even so, the dispersion $\sigma$ of ESNet is still smaller than that of the LAMOST catalog.

Concerning $\log g$, the dispersion $\sigma$ of the former steadily decreases and converges as the SNR increases. Similarly, the latter also shows convergence, but the final convergence value is higher, indicating larger errors.

Regarding $[\text{Fe}/\text{H}]$, the dispersion $\sigma$ of the former consistently decreases and converges with increasing SNR. However, the latter does not display a clear trend of convergence, and the pattern is diffused and unstable.

The experimental results reveal that the SNR significantly impacts the stability of ESNet, leading to improved consistency and accuracy between the predicted stellar parameters and the SDSS catalog compared to LAMOST. The results underscore the robustness and reliability of ESNet in estimating stellar parameters.
Figure 12. The consistencies between ESNet predictions, LAMOST catalog, and SDSS catalog. The upper left figure refers to $T_{\text{eff}}$; the upper right figure refers to $\log g$; the bottom left figure refers to $[\text{Fe}/H]$.

Figure 13. The dispersion of the difference between SDSS catalog and ESNet predictions or LAMOST catalog. The upper left figure refers to $T_{\text{eff}}$; the upper right figure refers to $\log g$; the bottom left figure refers to $[\text{Fe}/H]$.

The FE module in ESNet allows it to provide features that are compressed, as shown in Figure 14. It can be found that different parameters correspond to different spectral features.
The FE indeed compresses the spectra and extracts spectral features that are sensitive to the parameters. For example, (1) some spectral lines (Hβ, Hα, and SII) play an important role in estimating $T_{eff}$, log g, and [Fe/H]. (2) Some atomic metal lines (CaK, CaH, CaI, FeI, Mg, Ca) are also critical in estimating $T_{eff}$, log g, and [Fe/H]. (3) At the same time, we note that Na contributes more to estimating log g and [Fe/H] than to $T_{eff}$.

Figure 14. Compressed feature figure obtained by compressing spectra through FE. Spectral features sensitive to atmospheric parameters are shown in the figure.

6. Conclusions

In this paper, we propose an ESNet for the estimation of stellar parameters ($T_{eff}$, log g, and [Fe/H]). Our approach utilizes stellar spectra common to both LAMOST and SDSS as the datasets, with spectra obtained from LAMOST and labels (stellar parameters) from SDSS. The experimental results demonstrate that the ESNet predictions align well with those of the SDSS catalog. For a low SNR (0–10), ESNet achieves impressive performance, with predictions of $T_{eff}$ at 82 K, log g at 0.20 dex, and [Fe/H] at 0.10 dex in the test set, which proves that ESNet has an excellent ability to extract spectral features and remove redundant information in spectra. Additionally, we explore the consistency between the SDSS and LAMOST catalogs, revealing errors of 92K, 0.31 dex, and 0.12 dex for the three parameters, respectively. Compared to the results in Section 4, the ESNet predictions demonstrate improved consistency with the SDSS catalog compared to the LAMOST catalog.

To evaluate the proposed model, we compare it with seven methods. The results indicate that ESNet exhibits robust and better performance. At the same time, for a low SNR (0–10), compared to the Payne catalogs, ESNet achieves good performance, with predictions of $T_{eff}$ at 146 K, log g at 0.17 dex, and [Fe/H] at 0.10 dex. Compared to 2MASS, ESNet achieves good performance, with predictions of $T_{eff}$ at 101 K, log g at 0.19 dex, and [Fe/H] at 0.11 dex. The model’s stability is also analyzed, indicating that ESNet displays better robustness. Furthermore, we investigate the effect of the SNR on the predictions and find that as the SNR increases, ESNet exhibits better convergence and stability.

However, ESNet does not perform as well in extracting the spectral features of cool and hot stars, which proves that ESNet’s effectiveness in handling small data samples needs to be strengthened. Moreover, the performance in extracting low SNR spectral features is not as strong as the high SNR performance.

In conclusion, we will improve the model’s ability to extract spectral features with small data samples (hot and cool stars) and low SNR spectra in the next stage. In the future, 4MOST and SDSS-V will provide a large number of spectra that can be analyzed to reveal the properties and evolution of stars, galaxies, and the universe. It will be possible for ESNet to use these spectra to estimate the parameters.
Author Contributions: Methodology, K.W.; data curation, K.W.; writing—original draft preparation, K.W.; writing—review and editing, B.Q., A.-l.L., F.R. and X.J.; supervision, B.Q., A.-l.L., F.R. and X.J.; project administration, B.Q.; funding acquisition, B.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Tianjin, 22JCYBJC00410, and the Joint Research Fund in Astronomy, National Natural Science Foundation of China, U1931134.

Data Availability Statement: We are grateful to LAMOST for providing us with the data and SDSS for providing us with the stellar parameters. LAMOST DR10 Website: http://www.lamost.org/dr10/v1.0/ (accessed on 20 April 2023) and SDSS DR17 Website: https://www.sdss.org/dr17/ (accessed on 20 April 2023) and SIMBAD Website: http://simbad.u-strasbg.fr/simbad/ (accessed on 25 April 2023)

Acknowledgments: Software: Pytorch [43], Scipy[44], Matplotlib [45], Astropy [46].

Conflicts of Interest: The authors declare no conflict of interest.

Reference


41. Lingchang, P.; Zhenping, Y.; Xiaoming, K.; Yude, B. Atmospheric parameter measurement of Low-S/N stellar spectra based on deep learning. *Optik* 2020, 218, 165004. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.