Research in Astronomy and Astrophysics, 24:095023 (11pp), 2024 September https://doi.org/10.1088/1674-4527/ad7364

SFNet: Stellar Feature Network with CWT for Stellar Spectra Recognition

Hao Fu, Peng Liu, Xuan Qi, and Xue Mei
College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing 211816, China; liupeng@njtech.edu.cn

Received 2024 June 17; revised 2024 August 5; accepted 2024 August 15; published 2024 September 19

Abstract

Stellar spectral classification is crucial in astronomical data analysis. However, existing studies are often limited by the uneven distribution of stellar samples, posing challenges in practical applications. Even when balancing stellar categories and their numbers, there is room for improvement in classification accuracy. This study introduces a Continuous Wavelet Transform using the Super Morlet wavelet to convert stellar spectra into wavelet images. A novel neural network, the Stellar Feature Network, is proposed for classifying these images. Stellar spectra from Large Sky Area Multi-Object Fiber Spectroscopic Telescope DR9, encompassing five equal categories (B, A, F, G, K), were used. Comparative experiments validate the effectiveness of the proposed methods and network, achieving significant improvements in classification accuracy.

Key words: methods: data analysis – techniques: image processing – stars: imaging – galaxies: stellar content

1. Introduction

Stellar spectral classification is an important task in astronomical research. The classification results directly reflect the differences in stellar surface temperatures, which provide a basis for studying stellar energy radiation, internal structures, and their roles within galaxies. Stellar spectral classification also helps astronomers understand the evolution of galactic structures and the diverse effects of the interstellar medium, and supports the verification of astrophysical theories.

The Morgan–Keenan classification system categorizes stars into seven main spectral types based on surface temperature (Morgan et al. 1942). These types are denoted by the letters O, B, A, F, G, K, and M. Each type is further subdivided into 10 subclasses, represented by the numbers 0 through 9. This classification method was soon recognized by various scientific research teams upon its introduction. Over time, these teams have developed many classification methods based on this standard, which can be broadly categorized into two types: template matching and machine learning, as shown in Figure 1.

Traditional methods such as template matching held a significant position. Burbidge et al. (1957) analyzed the elements inside stars in their pioneering work, marking the beginning of using elemental analysis to explore further astronomical questions. Brown et al. (2003) further expanded the analysis of stellar spectra by adding methods such as Cross-Correlation and "Line-by-line." Smiljanic et al. (2014) released the Gaia-ESO Survey, which detailed the complete spectral analysis process, including Quality Control and Homogenization. Kesseli et al. (2017) used the Sloan Digital Sky Survey (SDSS) to establish an empirical template library, enhancing the spectral analysis task.

Compared to the template matching method, machine learning methods offer higher efficiency and are often more favored. LaSala (1994) classified 350 type B stars using the nearest neighbor method. Daniel et al. (2011) expanded the range of astronomical data objects processed, using local linear embedding to classify spectral data from the SDSS. Díaz-Hernández et al. (2014) employed sparse representation and dictionary learning methods to automatically classify stellar spectral types, achieving higher accuracy at a lower computational cost. Li et al. (2019) focused on spectra without flux standards, using the random forest algorithm to classify such spectra and demonstrating strong superiority in real-data applications.

It is noteworthy that deep learning, as an important branch of machine learning, now occupies an increasingly significant position in astronomical research. Compared with traditional machine learning methods, deep learning can automatically discover and represent high-level features in complex spectral data through end-to-end learning. Von Hippel et al. (1994) was the first to classify 575 spectra using artificial neural networks (ANN), marking the first application of deep learning in stellar classification tasks. Navarro et al. (2012) classified stellar spectra with low signal-to-noise ratios (SNRs), demonstrating the high robustness of ANN. In the same year, Mahdi studied classification of spectra from the ELODIE spectral library using probabilistic neural networks to further increase training speed (Mahdi 2008). In the face of the burgeoning volumes of astronomical data, simple ANN have become increasingly inadequate. Dieleman et al. (2015) used rotation-invariant convolutional neural networks (CNNs) in the Galaxy Zoo project to classify galaxy morphology. The success of this network provided a great reference for subsequent researches.

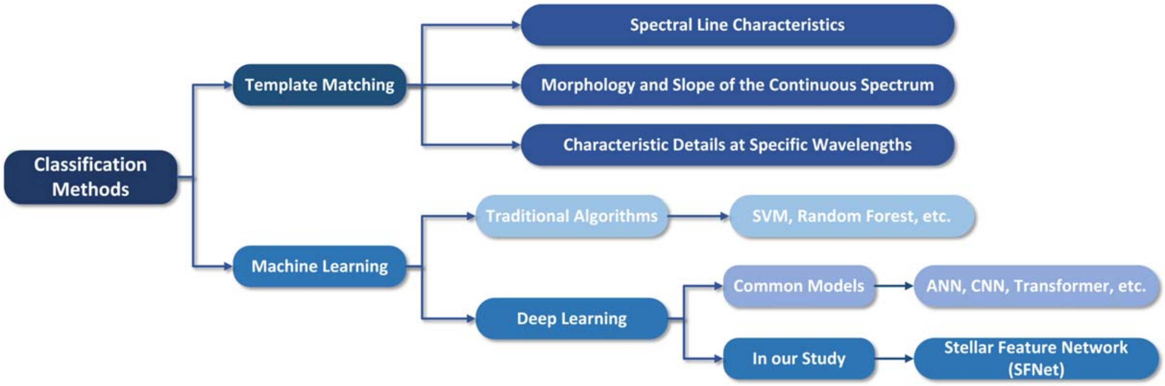


Figure 1. Common methods for stellar classification.

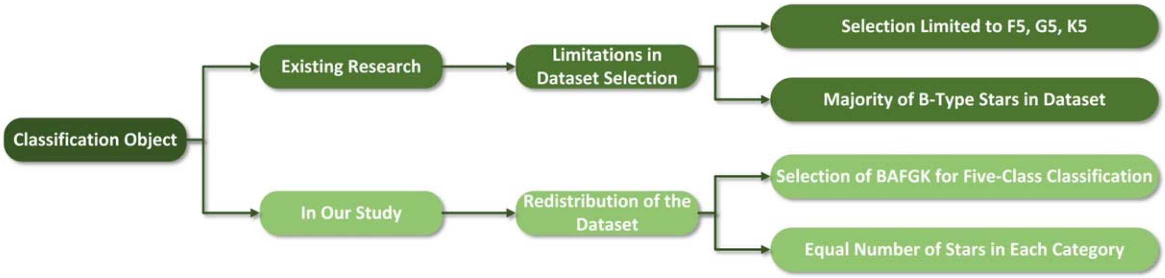


Figure 2. Comparison of research objects.

Kim & Brunner (2016) proposed a stellar-galaxy classification framework based on deep CNNs (ConvNets),further proving the feasibility of deep neural networks in various astronomical tasks. Wang et al. (2017) used deep neural networks to extract spectral features at different abstraction levels in various layers, applying this method to spectral classification and defective spectral restoration. Sharma et al. (2020) used CNNs for stellar spectral classification, comparing them with ANN and traditional machine learning methods to demonstrate the advantages of CNNs in this task. Shi et al. (2023) chose not to classify based on stellar spectra but classified stellar luminosity images using a designed CNN (SCNet), providing another promising approach.

However, due to the differences in the timing of each team's research and the specific classification tasks, it is difficult to use a single absolute standard to uniformly measure the performance. Figure 2 illustrates the limitations in the selection of research subjects in existing studies and the reason behind the data set chosen for this study.

In the early days, limited by observational capabilities, research teams often used only a small amount of data for their studies. For example, LaSala (1994) used the minimum mean method to classify only 350 type B stars, and in the same year, Von Hippel et al. (1994) first applied neural networks in this field but classified only 575 spectra. Although later, with the rise and improvement of astronomical observation methods such as SDSS (Almeida et al. 2023) and Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012), the ability to acquire stellar spectra has increased significantly. Many teams still exhibit significant bias when selecting their research subjects. For instance, Zhongbao Liu's team (Liu et al. 2016) only studied the classification of four types of K stars, three types of F stars, and three types of G stars. Liting Du's team (Li-Ting et al. 2021) only classified three subtypes of stars: F5, G5, and K5, within certain stellar categories. Although such classification results are very high, their application value still needs further exploration.Similarly, there are significant disparities in the number of instances among different stellar categories. Sharma et al. (2020) uniformly classified seven types (O, B, A, F, G, K, M) within the MK system, but spectra from F and G types alone constituted over half of the total samples in the data set.

In our works, to mitigate the impact of data set imbalance on experimental outcomes, we opted to categorize stars from LAMOST Data Release 9 (DR9) into five abundant classes: B, A, F, G, and K, ensuring an equal number of spectra for each category. We introduced the Super Morlet (SMor) wavelet function for continuous wavelet transformation (CWT) of spectral data, converting spectra into images to facilitate feature extraction by deep learning network models. Additionally, we developed the Stellar Feature Network (SFNet) specifically designed for wavelet images of these stellar classes, which enhanced the accuracy of the stellar classification tasks. Finally, a series of comparative experiments were designed to demonstrate the effectiveness of these methods.

The structure of this paper is as follows: Section 2 presents the preprocessing methods including one-dimensional convolution processing, CWT, and normalization, as well as the architecture of the SFNet. Section 3 describes the source of the data set, parameter settings for each preprocessing method and model training strategies. Section 4 presents the analysis and discussion of comparative experiments. Section 5 provides a comprehensive summary of the research and outlines future prospects.

2. Methods

In stellar classification, template matching methods typically rely on the elemental composition of stars, using empirical template libraries for reference. In contrast, machine learning approaches often train directly on one-dimensional stellar spectral data.

Many neural network models commonly used today, such as CNNs, excel at extracting local features from data structured as images for classification tasks. Consequently, using two-dimensional images as inputs to neural networks often results in superior performance compared to one-dimensional spectral data.

This study proposes using CWT to convert preprocessed one-dimensional stellar spectral data into two-dimensional wavelet images. These images are classified using the specifically designed SFNet. This approach preserves spectral information and enhances the neural network model's feature extraction capabilities for classification tasks.

2.1. One-dimensional Convolution

One-dimensional convolution is a fundamental operation in both signal processing and deep learning, typically used to smooth input signals or reduce the impact of noise. This technique is also prevalent in the analysis of spectral data (LeCun et al. 1995). The mathematical representation is as follows

$$y(t) = \sum_{a=-\infty}^{\infty} x(a) \cdot w(t-a), \tag{1}$$

where $y(t)$ represents the output of the convolution operation, $x(a)$ denotes the value of the input signal at a position, and $w(t-a)$ is the weight of the convolution kernel at position $t-a$.

In our experiment, a moving average convolution kernel is employed, where the convolution kernel is defined as [1, 1, 1]. This approach smooths the spectral data and reduces noise impacts while ensuring the retention of key features.

2.2. Continuous Wavelet Transform (CWT)

The CWT is a commonly used time–frequency analysis tool in signal processing (Torrence & Compo 1998). It employs wavelet functions to analyze the local frequency characteristics of a signal at various scales. Compared to other transformation methods such as the Fourier transform, the CWT offers finer resolution in the time–frequency domain, allowing for the simultaneous acquisition of information in both time and frequency aspects of the signal (Kaya et al. 2022).

For the task of stellar classification, converting one-dimensional stellar spectral data into two-dimensional wavelet images through CWT allows for more refined local information transformation in the time–frequency domain. This process provides a richer extraction of stellar features for subsequent neural network training and enhances the efficiency of the network. Such an approach demonstrates significant advantages when processing spectral data on a large scale.

The fundamental concept of CWT involves convolving a signal with a wavelet function that varies in scale. The scale parameter adjusts the frequency of the wavelet function, allowing the convolution results at different scales to provide local information about the signal in the frequency domain. The specific formula is as follows

$$\mathrm{CWT}(a, b) = \int_{-\infty}^{\infty} x(t) \cdot \psi^{*}\left(\frac{t-b}{a}\right) dt, \tag{2}$$

where $x(t)$ represents the input signal, $\psi(t)$ is the wavelet function, a is the scale parameter, b is the translation parameter, which is automatically generated based on the input data and covers all index positions of the data, and ψ^{*} denotes the complex conjugate of the wavelet function.

For CWT, the choice of the specific wavelet function is crucial. Commonly used wavelet functions include the Morlet wavelet, the Ricker wavelet (also known as the Mexican Hat wavelet), the Haar wavelet, and the Daubechies wavelet.

The Morlet wavelet function exhibits superior frequency smoothness, enabling the simultaneous preservation of frequency and phase information (Shao et al. 2021). The Ricker wavelet features excellent localization in the frequency domain, making it particularly effective in detecting signal edges (Alpar et al. 2022). The Haar wavelet is simple and rapid, and is better suited for analyzing abrupt changes in signals (Guo et al. 2022). The Daubechies wavelet is compactly supported in the time domain and exhibits good temporal localization characteristics, commonly used in fields such as signal denoising and compression (Mandala et al. 2023).

Based on the characteristics of various wavelet functions, this study initially selects the Morlet wavelet function. The Morlet wavelet is composed of the product of a Gaussian function and a complex exponential function. This construction endows the Morlet wavelet with localized features in both the time and frequency domains, enabling the simultaneous

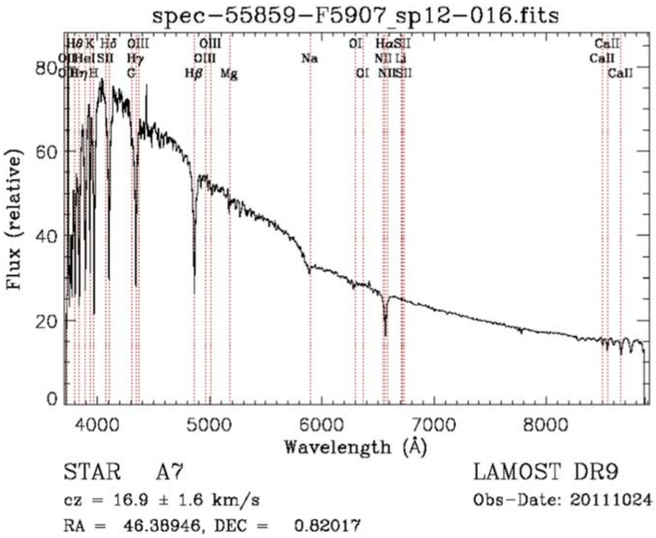


Figure 3. Wavelength–Flux curve before CWT.

analysis of temporal and frequency information in stellar spectra.

The definition of the Morlet wavelet is as follows

$$\psi(t) = \pi^{-1/4} e^{-\frac{t^2}{2}} e^{i\omega_0 t}, \tag{3}$$

where ω_0 represents the frequency parameter.

Figures 3 and 4 depict images of one-dimensional stellar spectral data before and after the application of CWT using the Morlet wavelet function respectively. In the wavelet image, the index represents the position in the data and corresponds to the length of the original spectral data (i.e., the number of sample points). The scale denotes the range of the scale parameters, which is set from 0 to 300 in this study. Magnitude represents the absolute value of the transform results, indicating the signal strength at specific scales and positions.

As shown in Figure 4, the wavelet images of stellar spectra transformed using the standard Morlet wavelet function exhibit dense information locally, but there are limitations in its ability to distinctly express features where the data are most concentrated. Consequently, there is a need to focus more on local details in the actual transformation process and to adjust the frequency bandwidth more flexibly to better adapt to areas in the spectrum where frequency changes are relatively significant, thus enriching the information displayed in the wavelet images.

Therefore, this study introduces the SMor wavelet function, an enhanced version of the Morlet wavelet. Building upon the Morlet wavelet, the SMor wavelet incorporates a bandwidth parameter, enabling the adjustment of the Gaussian window's parameters to better control localization characteristics in the time–frequency domain. The definition of the SMor wavelet function is as follows

$$\psi(t) = \pi^{-\frac{1}{4}} \sqrt{\sigma}\, e^{-\frac{t^2}{2\sigma^2}} e^{i(\omega_0 t)}, \tag{4}$$

where ω_0 represents the frequency parameter, and σ denotes the standard deviation of the Gaussian window.

Compared to the Morlet wavelet, the SMor wavelet allows for adjustment of the wavelet's frequency bandwidth, providing greater flexibility in the frequency domain. When σ is smaller, the SMor wavelet is more compact in the time domain while exhibiting a broader presence in the frequency domain. Conversely, when σ is larger, the SMor wavelet appears wider in the time domain and more compact in the frequency domain.

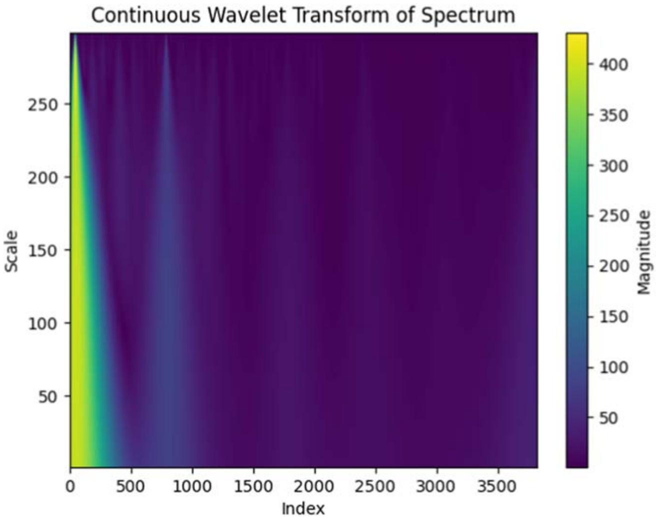


Figure 4. Wavelet image generated by CWT using Morlet wavelet function.

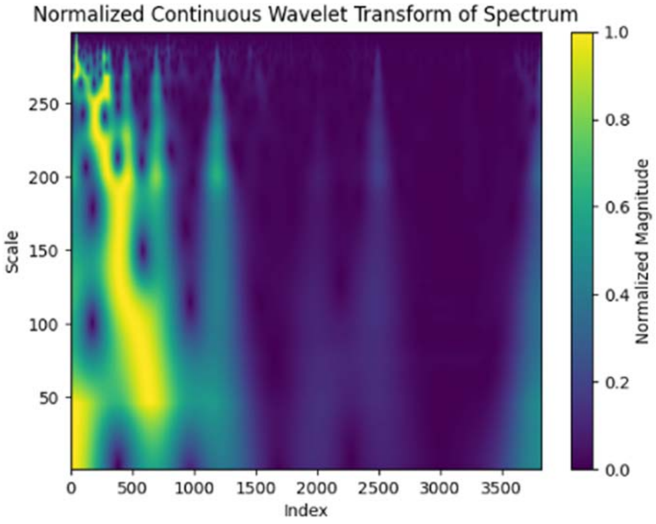


Figure 5. Wavelet image generated by CWT using SMor as the wavelet function.

Figure 5 displays the wavelet image of the same spectral data after transformation using the SMor wavelet. Compared to Figure 4, it is evident that the information in Figure 5 is more abundant, providing improved conditions for subsequent neural network training.

2.3. Normalization

In the conversion of one-dimensional stellar spectral data into two-dimensional wavelet images via CWT, normalization is an important operation. Different spectral data possess varying amplitude ranges, and it would be impractical to simply crop the images and input them into neural network training without considering the differences in amplitude values displayed by each wavelet image.

Therefore, we incorporate Min–Max Scaling in the CWT process. Min–Max Scaling is a commonly used data normalization method that linearly maps the stellar spectral data to a specified range between minimum and maximum values. This standardizes the magnitude of spectral data, thereby facilitating the training and classification processes of deep learning models (Ambarwari et al. 2020). The calculation formula for Min–Max Scaling is as follows

$$X_{\rm norm} = \frac{X - X_{\rm min}}{X_{\rm max} - X_{\rm min}}, \tag{5}$$

where $X_{\rm min}$ represents the minimum value in the data set, $X_{\rm max}$ denotes the maximum value in the data set, and $X_{\rm norm}$ is the normalized data.

2.4. Stellar Feature Network (SFNet)

To enhance the classification of stellar spectral images generated using the aforementioned methods, this study adopts core components from the ConvNeXt-V2-Base model (Woo et al. 2023) and integrates modules such as the Feature Pyramid Network (FPN; Lin et al. 2017) to better handle multi-scale features. This integration has led to the development of the SFNet. Experimental results demonstrate that SFNet exhibits superior performance in stellar spectral classification tasks, particularly when dealing with data sets produced by CWT using the SMor wavelet function.

The development of the Vision Transformer (ViT; Dosovitskiy et al. 2020) successfully adapted the Transformer (Vaswani et al. 2017) architecture for image processing, where it has consistently gained prominence, gradually achieving parity with CNNs. To demonstrate that pure CNNs still possess potential for research and development, and also because ViT does not show clear advantages in tasks such as object detection, Z. Liu et al. decided to adopt the design strategy of the Swin-Transformer (Liu et al. 2021) and introduced the ConvNeXt (Liu et al. 2022) network based on ResNet50 (He et al. 2016). This was aimed at proving that, under equivalent design and training strategies, pure CNNs still hold substantial advantages in image processing. This was confirmed by superior performance on multiple public data sets. Building on this foundation, Sanghyun Woo further integrated the Fully Convolutional Masked Autoencoder and Global Response Normalization (GRN), refining the network and naming it ConvNeXt-V2.

In the task of stellar spectral recognition, we observed that wavelet images generated from different stellar categories often exhibit distinct features at various scales. For instance, some are more prominent in local details, while others are apparent in broader global trends. Based on such a distribution of features, this study chose to further incorporate the FPN. FPN primarily establishes a top-down architecture, where the network can utilize deep, high-level features and propagate these features to shallower layers through upsampling.

Although FPN is more commonly applied in the field of object detection, it is equally capable of capturing features at various scales, from detailed to global, in wavelet images for stellar spectral classification tasks. By aggregating wavelet feature maps at different resolutions, FPN integrates features from various depth levels, enabling simultaneous capture of high-level semantic information and low-level detail. This offers a significant advantage in understanding the multi-scale structures present in wavelet images. Furthermore, by merging information from different layers, FPN also enhances the model's robustness to variations in input image sizes, which is particularly important when dealing with the diverse stellar spectral data encountered in practical astronomical challenges.

The final design of SFNet is depicted in Figure 6. The network inputs are images generated by CWT, which have been cropped to remove axes. Initially, these images are processed through a convolutional layer followed by Layer Normalization (LN; Ba et al. 2016). Within the convolution layer, a 4 × 4 kernel and a stride of 4 are used, transforming the input image from $384 \times 384 \times 3$ to a $96 \times 96 \times 256$ output feature map. This operation reduces the image size while increasing the channel count, thereby elevating the level of feature abstraction necessary for initial feature capture.

Subsequently, the process flows through four SFNet Blocks, with the detailed internal structure of each block shown on the right side of the figure. Each block comprises a submodule and a convolution layer. The submodule contains N operations, each consisting of a convolution layer, GRN, and Drop Path (Huang et al. 2016), connected in a residual structure. GRN standardizes and applies nonlinear transformations to features, significantly enhancing the model's capability to process global information, thus improving the network's expressiveness and stability. Drop Path is a regularization technique that randomly omits certain paths within the neural network, reducing overfitting and enhancing the network's generalization

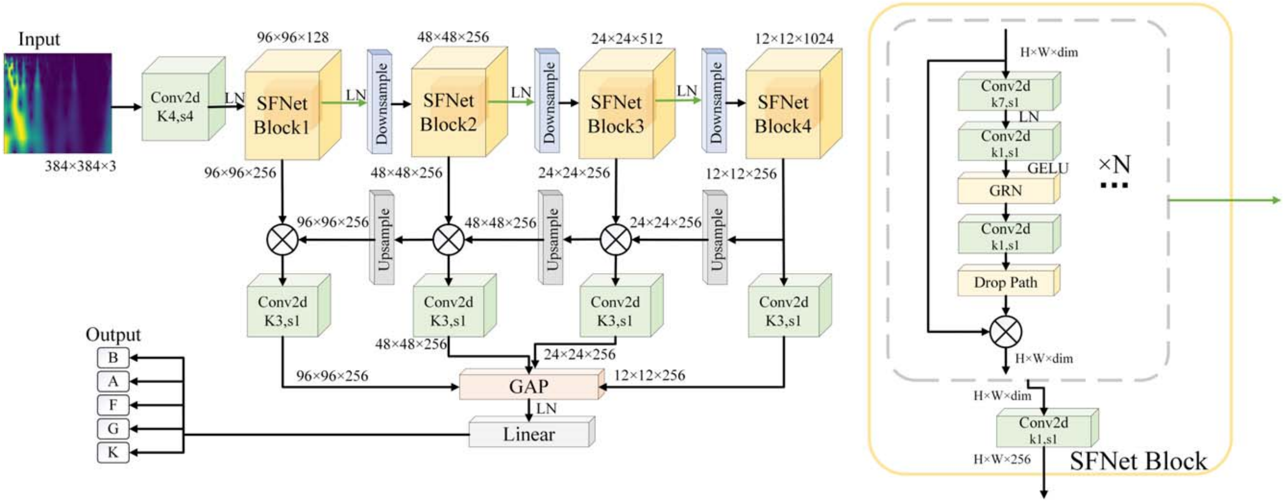


Figure 6. SFNet architecture.

capacity. Residual connections help alleviate the vanishing gradient problem by introducing skip connections that directly transmit information from previous layers, speeding up training and enhancing the performance of deeper network layers (Qi et al. 2023). In this experiment, N is set to 3 in SFNet Blocks 1, 2, and 4, and N is set to 27 in SFNet Block 3.

The feature map's input into each SFNet Block has lateral outputs in Blocks 1, 2, and 3, and vertical outputs in Blocks 1, 2, 3, and 4. Each lateral output, derived from the submodule within the block, undergoes LN and downsampling to reduce the dimensions of the feature map before being fed into the next block for further feature extraction. These lateral outputs are marked with green arrows in the diagram. After processing through the submodule, each block passes the data through a convolution layer with a kernel size of 1×1 and a stride of 1, aimed at maintaining each vertical output feature maps channel count at 256.

Subsequently, the vertical outputs from Blocks 2, 3, and 4 are upsampled and laterally connected. Using the vertical output from Block 1 as the reference size, bilinear interpolation is employed to upsample these outputs to this size and merge them. This procedure enables effective fusion of features from different levels, leveraging diverse feature information provided at various stages to better capture comprehensive features from fine-grained details to high-level semantics. Following fusion, each output undergoes Global Average Pooling (Lin et al. 2013), where an averaging operation is performed across all spatial positions of each convolutional feature map, compressing the two-dimensional spatial information into a single scalar. This application not only reduces dimensionality but also helps in mitigating overfitting. Finally, the classification of stellar types is output through a fully connected layer.

3. Experiment and Analysis

3.1. Data Acquisition

In this study, we utilized data from the ninth data release (DR9 v2.0) of LAMOST as the source for our stellar spectral classification data set. LAMOST, also known as the Guo Shoujing Telescope, is located in Xinglong, Hebei Province, China, and is operated by the National Astronomical Observatories of the Chinese Academy of Sciences. It combines the advantages of a large aperture and a wide field of view, featuring a primary mirror with a diameter of 4 meters and a secondary mirror with a diameter of 1.75 m, covering a sky area of 20 square degrees. Capable of simultaneously acquiring spectra from up to 4000 celestial bodies, its design makes it exceedingly efficient in conducting large-scale astronomical surveys, such as those measuring galaxies, quasars, and stellar spectra.

In the LAMOST DR9 v2.0 data set, there are as many as 10,809,336 spectral records covering a variety of celestial objects including galaxies, stars, and quasars. This study focuses on the classification of stellar types. For the spectral data, we required that the selected spectra have an SNR greater than 20. The number of spectra meeting the SNR criterion are as follows: 56 for O-type stars, 5,823 for B-type, 128,121 for A-type, 476,112 for F-type, 341,477 for G-type, 17,217 for K-type, and 244 for M-type.

Due to the limited number of spectral data for O and M-type stars, this study focuses primarily on the classification of five stellar types: B, A, F, G, and K. To mitigate the effects of uneven data distribution, the data set was configured to match the number of the least represented of these categories, B-type stars. Consequently, 5000 spectra were selected for each of the B, A, F, G, and K categories to form the data set.

3.2. Preprocessing Details

The preprocessing component of this study primarily consists of one-dimensional convolution processing, CWT, and Min–Max Normalization.

Initially, the spectral data in FITS format undergo one-dimensional convolution processing, where the convolution kernel is set to [1, 1, 1], the stride is set to 1, and padding is set to 1.

Subsequently, CWT is performed. Each spectral sample is first converted into a one-dimensional array, followed by scale mapping. In our study, the scale parameter a in Equation (2) is set to range from 1 to 300. This selection is based on a comprehensive consideration of the effects of different scales. Through a series of experimental comparisons, we found that transformations at smaller scales (below 200) often fail to capture the global features of the signal, resulting in the loss of important trends that could reflect stellar categories. Additionally, experiments have shown that excessively large scales (above 400) can capture low-frequency information but result in poor temporal locality, making it difficult to finely resolve transient features and subtle differences in the signal, such as spikes and abrupt changes. By choosing a scale mapping range of 1–300 for the CWT, a balance is achieved between the temporal and frequency locality of the transformation, ensuring that while global trends of the signal are captured, key local details are not lost, thereby effectively enhancing the model's predictive and feature extraction capabilities.

In addition to scale mapping, the key parameters for the Smor wavelet are set as follows: the frequency parameter is set at 6.0 to optimize frequency resolution; the scale parameter is set at 0.8, adjusting temporal resolution to capture key frequency information while maintaining sufficient temporal locality. For color mapping, viridis is chosen over options like jet, as viridis provides a consistent and continuous color gradient, avoiding misleading color jumps and enhancing the continuity of visual interpretation. The interpolation method selected is bilinear interpolation, which, compared to nearest neighbor interpolation, reduces the staircasing effect and provides a smoother image texture without introducing excessive computational complexity.

Finally, the results of the CWT are subjected to Min–Max Scaling for amplitude normalization, standardizing the magnitudes of transformations across all scales to a range between 0 and 1. This process outputs the images generated by the CWT.

3.3. SFNet Training Results

This study was conducted on a computing platform equipped with an Intel(R) Xeon(R) Platinum 8352V CPU and an NVIDIA RTX 4090 (24GB) GPU, utilizing Ubuntu 20.04 as the operating system. All experiments were performed within a Python 3.8 and PyTorch 1.11.0 environment. The network model employed was the SFNet, with the final model having a total parameter count of 90.363 million and a total number of floating-point operations of 52.907 billion. This computationally complex setup contributed to achieving high accuracy in the classification results.

The experimental data set encompasses five categories, with images from each category processed through the following standardization steps: First, image dimensions are resized to 384 × 384 pixels. Subsequently, normalization is applied using a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. This configuration is based on the statistical mean and standard deviation of RGB images from the publicly available ImageNet data set in the field of computer vision. Such preprocessing helps to standardize the input wavelet images, thereby improving the training stability and transfer learning performance of the model.

In terms of model configuration, the model did not utilize pre-trained weights and training commenced from random initialization. The normalization layer utilizes LN technology, with the epsilon value set at 1e-06; the output layer is a fully connected layer configured with 256 input features and 5 output categories, including bias terms. Further details of the remaining layer structures are presented in Section 2.4.

For training, we set a Drop Path Rate of 0.35, effectively reducing the risk of model overfitting. The training process employed the AdamW optimizer (Loshchilov & Hutter 2017), with an initial learning rate set at 3e-5 and a weight decay coefficient of 0.08, which aids in maintaining model stability. A warm-up strategy for the learning rate was applied during the first training epoch and was dynamically updated based on training progress. The overall training plan included 100 epochs, with each batch processing 8 images, ensuring sufficient training depth and data coverage.

The performance of the model is evaluated through accuracy, precision, and recall, with the calculation formulas as follows:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{6}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{7}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{8}$$

Performance data were recorded and visualized using TensorBoard, with the accuracy and loss variation curves displayed in Figures 7 and 8 respectively. These curves illustrate the performance changes of the model during the training and validation processes. As training progressed, the Train Accuracy eventually converged to 1, while the Validation Accuracy reached 95.22%.

The confusion matrix, along with the precision and recall for each category, is presented in Figure 9, further elucidating the performance of the SFNet in stellar classification tasks. It is

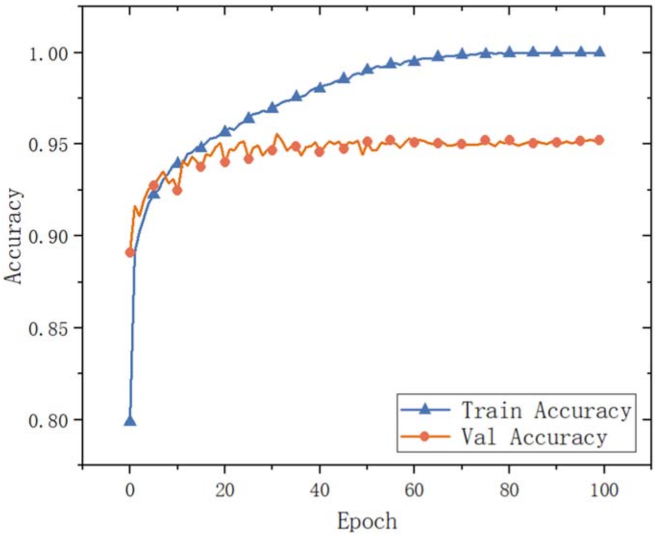


Figure 7. Accuracy curve.

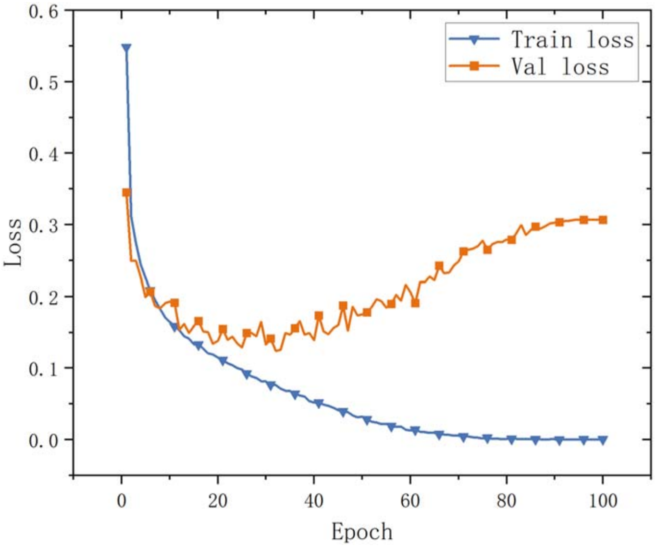


Figure 8. Loss curve.

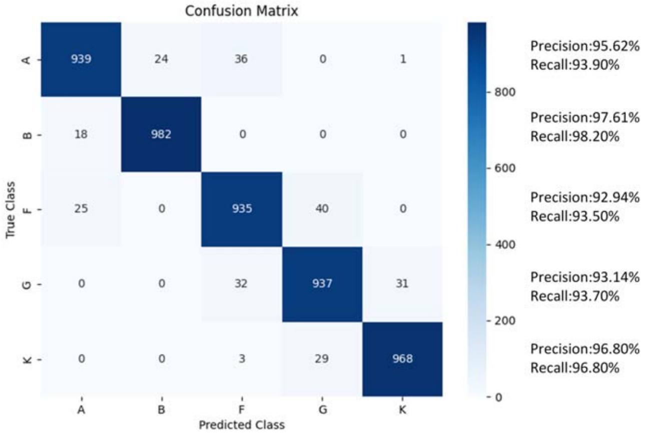


Figure 9. Confusion matrix.

evident that the recall for B-type stars is the highest, reaching 98.20%, while the recall for F-type stars is the lowest, still achieving 93.50%. Overall, the experiments achieved a final accuracy of 95.22% on the test set, demonstrating the significant effectiveness of the preprocessing techniques, such as CWT using SMor as the wavelet function, and the SFNet model in feature extraction and classification generalization. The balanced accuracy displayed across different categories further corroborates the model's capability and stability in handling diverse data sets. These results validate the potential of our methodology for practical applications and highlight the prospects of deep learning approaches in stellar classification tasks.

4. Discussions

We conducted a series of comparative experiments involving several classifiers, including the SGDClassifier, Random Forest, VGG19, and ViT. These methods include traditional machine learning techniques and two representative types of networks from deep learning approaches, enabling the evaluation and comparison of their effectiveness and efficiency in stellar image classification tasks.

The Stochastic Gradient Descent (SGD; Bottou 2010) classifier is an incremental learning method based on stochastic gradient descent. The classifier is configured with a hinge loss function combined with L2 regularization, overall representing an approximate implementation of Support Vector Machines (Cortes & Vapnik 1995).

The Random Forest classifier (Breiman 2001) is a powerful method based on ensemble learning, which enhances prediction accuracy and stability by constructing multiple decision trees and aggregating their outputs at the ensemble layer. The classifier is configured with 100 decision trees and handles large-scale image data sets through batch processing (1024 images per batch), ensuring efficient data management and computation.

VGG19 (Simonyan & Zisserman 2014) is a widely used deep CNN. In the experiments, the image size was adjusted to 384 × 384 pixels, and the data were augmented by random horizontal flipping with a probability of 0.5, followed by normalization. Model training was conducted on a CUDA-accelerated device using the AdamW optimizer with an initial learning rate set at 2e-5 and a weight decay of 5e-2. Throughout the training process, a total of 100 epochs were set, processing 16 images per batch.

Table 1
Results of Various Classifiers for Images Generated by Short-Time Fourier Transform (STFT)

Results of Various Classifiers for Images Generated by STFT					
Classifier	SGDClassifier	Random Forest	VGG19	Vision Transformer	SFNet
Objects	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K
Other Methods	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm
Accuracy	90.72%	91.52%	94.26%	93.42%	93.88%

Table 2
Results of Various Classifiers for Images Generated by Continuous Wavelet Transform (CWT) Using the Morlet Wavelet Function

Results of Various Classifiers for Images Generated by CWT-Morlet					
Classifier	SGDClassifier	Random Forest	VGG19	Vision Transformer	SFNet
Objects	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K
Other methods	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm
Accuracy	82.98%	84.88%	91.58%	90.12%	91.76%

Table 3
Results of Various Classifiers for Images Generated by Continuous Wavelet Transform (CWT) Using the Super Morlet (SMor) Wavelet Function

Results of Various Classifiers for Images Generated by CWT-SMor					
Classifier	SGDClassifier	Random Forest	VGG19	Vision Transformer	SFNet
Objects	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K	B, A, F, G, K
Other methods	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm	1D-Conv, Norm
Accuracy	91.69%	93.94%	94.70%	93.72%	95.22%

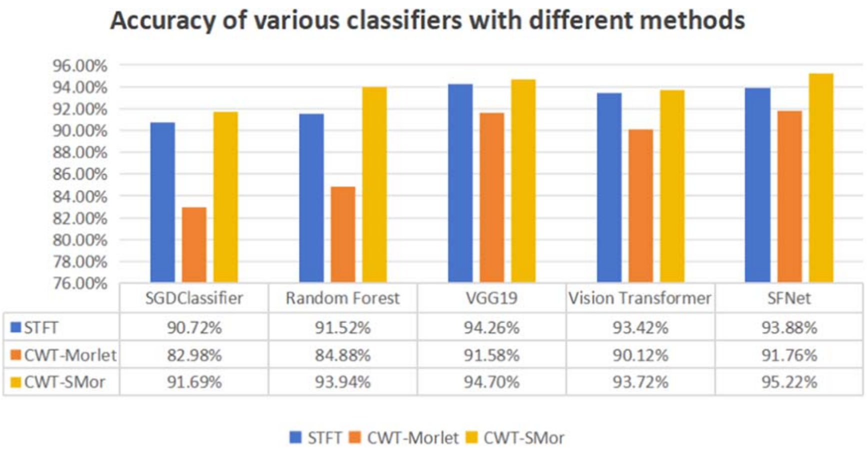
The ViT model (Dosovitskiy et al. 2020) employs a deep learning architecture based on the self-attention mechanism. During the data preprocessing stage, images are resized to 224×224 pixels and enhanced through random horizontal flipping and normalization to augment the model's generalization capability. The model training utilizes the AdamW optimizer, combined with a learning rate scheduler to optimize the training process. The initial learning rate is set at 2e-5 with a weight decay of 5e-2. Over the course of 100 epochs, the learning rate is gradually reduced and a warm-up strategy is implemented at the beginning of training, effectively balancing convergence speed with training stability.

For these classification methods, our comparative experiments are divided into three parts.

In the first part of our study, we opted for the Short-Time Fourier Transform (STFT) method (Li-Ting et al. 2021), commonly used in the spectral domain and previously applied to stellar classification tasks, with specific parameter settings consistent with those used by Liting Du. We retained the other preprocessing methods and data loading techniques to illustrate the impact of different transformation methods on the final classification accuracy. The results, as shown in Table 1, indicate that the accuracy of the five classification methods ranged from 90.72% to 94.26%. Notably, the SFNet designed in this study is more suited to classifying wavelet images generated by CWT and does not exhibit a feature extraction advantage with images produced by STFT.

In the second part of our study, we modified the preprocessing approach by employing the CWT, utilizing the commonly used Morlet wavelet function, while maintaining other preprocessing methods such as one-dimensional convolution processing and normalization unchanged. The specific results, as shown in Table 2, demonstrate that the accuracy of the five classification methods ranged from 82.98% to 91.76% across the same data set. The experimental outcomes indicate that the CWT method using only the Morlet wavelet function does not exhibit competitive performance. However, compared to the STFT method, the SFNet designed in this study demonstrates advantages in classifying stellar spectral images generated by CWT.

In the third part of our experiment, we utilized the SMor wavelet function proposed in this study as the wavelet function for CWT, while all other preprocessing methods remained unchanged. The specific results, as presented in Table 3, show that the accuracy of the five classification methods ranged from 91.69% to 95.22%. Each classification method experienced improvements in classification performance with images generated by CWT using the SMor wavelet function. Notably, the Random Forest method saw an increase in accuracy of 9.06%, while the other classification methods also exhibited at



	SGDClassifier	Random Forest	VGG19	Vision Transformer	SFNet
🟦 STFT	90.72%	91.52%	94.26%	93.42%	93.88%
🟧 CWT-Morlet	82.98%	84.88%	91.58%	90.12%	91.76%
🟨 CWT-SMor	91.69%	93.94%	94.70%	93.72%	95.22%

Figure 10. Accuracy of various classifiers with different methods.

least a 3% increase in accuracy. These experimental results, superior to those using the Morlet wavelet function for CWT, indicate that the SMor wavelet function proposed in this study is more effective in retaining information beneficial for classification in the transformation of stellar spectra. Additionally, the accuracies of all five classifiers were also higher than those achieved with the STFT method, demonstrating the feasibility of the CWT approach.

It is noteworthy that when comparing the CWT method using the SMor wavelet function with the STFT method, although the SMor approach resulted in increased accuracy across all classification methods, the improvements for the deep learning models VGG19 and ViT were relatively modest, at only 0.44% and 0.3%, respectively. In contrast, the SFNet demonstrated a more significant increase of 1.34%. This indicates that the SFNet designed in this study has a distinct advantage in handling classification tasks involving the conversion of stellar spectra to images using CWT. Figure 10 shows the accuracy of various classifiers with different methods.

5. Conclusion

In this paper, we aim to classify stars using deep learning methods. Initially, we analyzed the number of spectra for each stellar category, selecting five types of stars (B, A, F, G, and K) for our study and ensuring equal experimental numbers for each. This approach was taken to eliminate the potential impacts of data set imbalance and enhance the persuasiveness of the results. Furthermore, we propose the use of CWT to convert stellar spectra into wavelet images, introducing an SMor wavelet function specifically designed for stellar spectra, which preserves more spectral features and improves classification accuracy. Lastly, we designed the SFNet, a model that optimizes multi-scale feature extraction from local to global levels. This model demonstrates advantages in feature extraction for images generated using the SMor wavelet function, further enhancing classification accuracy. In the experimental phase, we compared the results of three methods (STFT, CWT-Morlet, and CWT-SMor) across five classifiers (SGDClassifier, Random Forest, VGG19, ViT, and SFNet), validating the superiority of the proposed SMor wavelet function and SFNet in the task of stellar classification.

In summary, this study has enhanced preprocessing steps through innovative wavelet functions and improved deep learning models to better suit feature extraction tasks for wavelet images, ultimately increasing the accuracy of stellar classification tasks. In future work, we aim to further enhance classification performance and subdivide the data set into more detailed stellar subcategories for experimental research.

ORCID iDs

Hao Fu https://orcid.org/0009-0000-2184-5035
Peng Liu https://orcid.org/0000-0003-4123-376X
Xuan Qi https://orcid.org/0000-0003-1118-9008
Xue Mei https://orcid.org/0000-0002-4751-0644

References

Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, ApJS, 267, 44
Alpar, O., Dolezal, R., Ryska, P., & Krejcar, O. 2022, Appl Intell, 52, 15237
Ambarwari, A., Adrian, Q. J., Herdiyeni, Y., et al. 2020, RESTI, 4, 117
Ba, J. L., Kiros, J. R., & Hinton, G. E. 2016, arXiv:1607.06450
Bottou, L. 2010, in Proc. COMPSTAT'2010: 19th Int. Conf. Comput. Stat., ed. Y. Lechevallier & G. Saporta (Berlin: Springer), 186
Breiman, L. 2001, Mach. Learn., 45, 5
Brown, W. R., Prieto, C. A., Beers, T. C., et al. 2003, AJ, 126, 1362
Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, RvMP, 29, 547
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 297
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Daniel, S. F., Connolly, A., Schneider, J., VanderPlas, J., & Xiong, L. 2011, AJ, 142, 203

Díaz-Hernández, R., Peregrina-Barreto, H., Altamirano-Robles, L., González-Bernal, J., & Ortiz-Esquivel, A. 2014, ExA, 38, 193

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2020, in 9th Int. Conf. on Learning Representations (arXiv:2010.11929), https://openreview.net/forum?id=YicbFdNTTy

Guo, W., Xu, G., Liu, B., & Wang, Y. 2022, LGRS, 19, 5

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in IEEE Conf. Computer Vision and Pattern Recognition (Las Vegas, CA: IEEE), 778

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, K. Q. 2016, in Computer Vision–ECCV 2016: 14th European Conf., ed. B. Leibe, J. Sebe, & M. Welling (Berlin: Springer), 661

Kaya, Y., Kuncan, F., & Ertunç, H. M. 2022, Turkish J. Electrical Eng. Computer Sci., 30, 1851

Kesseli, A. Y., West, A. A., Veyette, M., et al. 2017, ApJS, 230, 16

Kim, E. J., & Brunner, R. J. 2016, MNRAS, 464, 4475

LaSala, J. 1994, The MK Process at 50 Years: A Powerful Tool for Astrophysical Insight (San Francisco, CA: ASP), 312

LeCun, Y., & Bengio, Y. 1995, The Handbook of Brain Theory and Neural Networks (Cambridge, MA: MIT Press), 258

Li, X.-R., Lin, Y.-T., & Qiu, K.-B. 2019, RAA, 19, 111

Li-Ting, D., Li-Hua, H., Jin-Tao, Y., et al. 2021, ChA&A, 45, 352

Lin, M., Chen, Q., & Yan, S. 2013, in Int. Conf. Learning Representations

Lin, T.-Y., Dollár, P., Girshick, R., et al. 2017, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (Honolulu: IEEE), 944

Liu, Z., Lin, Y., Cao, Y., et al. 2021, in Proc. IEEE/CVF Int. Conf. Computer Vision (Montreal: IEEE), 10002

Liu, Z., Mao, H., Wu, C.-Y., et al. 2022, in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (New Oreleans, LA: IEEE), 11976

Liu, Z., Song, L., & Zhao, W. 2016, MNRAS, 455, 4289

Loshchilov, I., & Hutter, F. 2017, in 7th Int. Conf. on Learning Representations, https://openreview.net/forum?id=Bkg6RiCqY7

Mahdi, B. 2008, BASI, 36, 54

Mandala, S., Pratiwi Wibowo, A. R., Adiwijaya, et al. 2023, ApSci, 13, 3036

Morgan, W., Keenan, P. C., & Kellman, E. 1942, An Atlas of Stellar Spectra, with an Outline of Spectral Classification (Chicago, MA: Univ. Chicago Press)

Navarro, S., Corradi, R., & Mampaso, A. 2012, A&A, 538, A76

Qi, X., Wei, Y., Mei, X., Chellali, R., & Yang, S. 2023, in Int. Conf. Artificial Neural Networks (Berlin: Springer), 185

Shao, H., Xia, M., Wan, J., & de Silva, C. W. 2021, IEEE/ASME Trans. Mechatron., 27, 24

Sharma, K., Kembhavi, A., Kembhavi, A., et al. 2020, MNRAS, 491, 2280

Shi, J.-H., Qiu, B., Luo, A.-L., et al. 2023, MNRAS, 520, 2269

Simonyan, K., & Zisserman, A. 2015, in 3rd Int. Conf. on Learning Representations, ed. Y. Bengio & Y. LeCun (San Diego, CA: ICLR) http://arxiv.org/abs/1409.1556

Smiljanic, R., Korn, A. J., Bergemann, M., et al. 2014, A&A, 570, A122

Torrence, C., & Compo, G. P. 1998, BAMS, 79, 61

Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Advances in Neural Information Processing Systems, ed. I. Guyon, U. von Luxburg, S. Bengio et al. (Long Beach, CA: NIPS), 6008, https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Von Hippel, T., Storrie-Lombardi, L., Storrie-Lombardi, M., & Irwin, M. 1994, MNRAS, 269, 97

Wang, K., Guo, P., & Luo, A.-L. 2017, MNRAS, 465, 4311

Woo, S., Debnath, S., Hu, R., et al. 2023, in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (Vancouver: IEEE), 16142