

SPT-GloCal: Enhancing [O/Fe] and [Mg/Fe] determinations in metal-poor stars with UV-extended low-resolution CSST spectra

Mengmeng Zhang¹, Yude Bu^{1,*}, Jianrong Shi^{2,3,4}, Hongliang Yan^{2,3}, Jingzhen Sun¹, Jiangchuan Zhang⁵, Ke Wang¹, Yuhang Zhang¹, Shanshan Li¹, Siqi Wang¹, Zhenping Yi⁶, Meng Liu⁶, and Xiaoming Kong⁶

¹ School of Mathematics and Statistics, Shandong University, Weihai 264209, Shandong, PR China

² CAS Key Lab of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences A20 Datun Road, Chaoyang, Beijing 100101, PR China

³ School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, PR China

⁴ School of Physics and Technology, Nantong University, Nantong 226019, PR China

⁵ School of Space Science and Technology, Shandong University, Weihai 264209, PR China

⁶ School of Aerospace Science and Engineering, Shandong University, Weihai 264209, PR China

Received 18 July 2025 / Accepted 28 November 2025

ABSTRACT

Metal-poor stars (MP, [Fe/H] < -1.0) retain the chemical signatures of the early Universe, making their α -element invaluable for tracing the Galactic chemical evolution. Traditional optical spectroscopic methods and machine learning approaches often struggle at low metallicities and neglect the diagnostic power of ultraviolet (UV) spectral features (e.g., OH bands, Mg II/Mg I lines) that are accessible to the forthcoming Chinese Space Station Telescope (CSST). Using over 1.8×10^5 simulated CSST spectra, we developed the spectral transformer (SPT)-GloCal model to quantify the impact of UV (2550–4000 Å) low-resolution ($R \approx 200$) spectra on the precision of [O/Fe] and [Mg/Fe] estimates in MP stars. The model improves local attention through score-aware competitive filtering and integrates it with global attention via a learnable gating mechanism. We compared models trained on full spectra versus optical-only spectra. Incorporating UV spectra halved the mean absolute error (MAE) of [O/Fe] predictions from 0.0785 to 0.0367 dex and reduced the scatter (σ) from 0.135 to 0.063 dex. For [Mg/Fe], MAE decreased from 0.0010 to 0.0006 dex and σ from 0.054 to 0.0068 dex. Error analyses demonstrated that UV data lead to more stable and accurate estimates, especially at a high log g and low T_{eff} . SPT-GloCal outperforms both its predecessor (SPT) and tree-based regressors, confirming the effectiveness of its global-local attention design. Furthermore, tests on spectra with simulated noise show that the model maintains consistent performance across a range of signal-to-noise ratios ($S/N = 30$ – 50), with marginal gains over the SPT model at the lower S/N end. UV spectral features, even at low resolution ($R \approx 200$), enhance α -element abundance determinations. The SPT-GloCal framework provides a scalable solution for upcoming space-based surveys. Future work will apply this model to real CSST data, thus extending it to other elements.

Key words. methods: data analysis – methods: statistical – techniques: spectroscopic – astrometry – ultraviolet: stars

1. Introduction

Metal-poor (MP) stars of [Fe/H] < -1.0 dex in the Milky Way are old, and they provide a unique observational window into the conditions of the early Universe, including the physical and chemical properties of the gas from which the first generations of stars and galaxies formed. Over the past 12–13 billion years, the atmospheres of these old stars have preserved the local chemical signatures of the gas clouds from which they originated (e.g., Frebel & Norris 2015; Frebel & Ji 2023). As such, MP stars play an important role in the field of stellar archaeology. By investigating their chemical composition, we can reconstruct the formation and assembly history of the Milky Way and reveal the nature of the first stars, their supernova explosions, the onset of cosmic chemical evolution, the physics of nucleosynthesis, and early metal- and gas-mixing processes (e.g., Beers & Christlieb 2005; Frebel et al. 2005; Iwamoto et al. 2005; Frebel 2010; Joggerst et al. 2010; Joggerst & Whalen 2011; Susa et al. 2014; Zhao et al. 2016; Chen et al. 2017; Ishigaki et al. 2018; Hartwig et al. 2019; Chiaki & Tominaga 2020; Kolborg et al. 2022; Shi et al. 2025).

Oxygen (O) and magnesium (Mg) are typical α -elements, synthesized primarily through the explosions of massive stars as supernovae during their evolution (Timmes et al. 1995; Kobayashi et al. 2006; Ting et al. 2018; Palla et al. 2022; Santos-Peral et al. 2023). The [O/Fe] and [Mg/Fe] abundance ratios, representing the ratio of α -elements to iron, are widely used as important chemical fossils for tracing the chemical evolution of the Galaxy in Galactic archaeology (Arnone et al. 2005; Carretta et al. 2009; Marino et al. 2011). In MP stars, the [O/Fe] and [Mg/Fe] ratios are typically enhanced. These two elements have shorter nucleosynthesis timescales compared to iron from supernova Ia, resulting in distinct abundance distributions. By studying the abundances of these two elements in MP stars, we can reveal the physical and chemical mechanisms of early star formation, infer the initial mass function (IMF) of massive stars, estimate the contributions of different types of supernova explosions to stellar metallicity, and provide crucial information on the chemical enrichment of the early MW (Gratton et al. 2000; Krauss & Chaboyer 2003; Beers & Christlieb 2005; Mashonkina 2013; Bond et al. 2013; Li et al. 2022a; Frebel & Norris 2015; Sneden et al. 2023; Xing et al. 2023; Welsh et al. 2024; Guigliano 2025; Bonifacio et al. 2025).

Currently, the methods for estimating the [O/Fe] and [Mg/Fe] in MP stars can be primarily divided into traditional and machine

* Corresponding author: buyude@sdu.edu.cn

learning. Traditional methods include spectral line index analysis, full spectral fitting, and non-local thermodynamic equilibrium (non-LTE) analysis. These methods typically rely on high resolution spectral observations, which are generally very expensive in terms of telescope time, thus limiting their application to large stellar samples. By analyzing spectral features, these methods derive the abundance ratios of [O/Fe] and [Mg/Fe] directly or indirectly. For instance, Carretta et al. (2000) used high resolution spectra and non-LTE analysis to measure the abundances of O and Mg, discussing their ratios in comparison to iron. Milone et al. (2011) used spectral index analysis to measure the [Mg/Fe] ratio in 76.3% of stars in the MILES spectral library. Ishigaki et al. (2012) conducted high resolution spectral analysis of 97 MP stars, applying one-dimensional local thermodynamic equilibrium (LTE) abundance analysis code to investigate the relationships between [O/Fe], [Mg/Fe], [Fe/H], and stellar dynamics. Magrini et al. (2017) used UVES spectral data from the *Gaia*-ESO survey to study the elemental abundance distribution of open clusters and young field stars in the Galactic disk, including [O/Fe] and [Mg/Fe]. Palla et al. (2022) employed the AMBRE: HARPS dataset to provide more precise estimates of [Mg/Fe] and reliable stellar ages, investigating the bimodality of [α /Fe] and the chemical evolution of the Galactic disk. These classical spectroscopic techniques are not confined to the Milky Way and have been successfully applied to extragalactic systems in order to measure the abundances in stars within nearby dwarf galaxies and globular clusters (e.g., Larsen et al. 2018, 2022; Hernandez et al. 2018). Moreover, in an effort to improve observational efficiency, some recent studies have started to push these techniques to “lower” spectral resolutions (e.g., Hernandez et al. 2017), with $R \approx 5000$.

With the advancement of astronomical observation technologies, particularly the accumulation of large-scale spectral data, machine learning methods (Wu et al. 2023; Zhang et al. 2024c) have experienced widespread application due to their ability to handle high-dimensional data and efficiently process large samples. Crucially, this technique pushes this type of analysis to low-resolution data (e.g., $R \approx 1800$), which is a huge step forward, as it enables the chemical characterization of millions of stars. Within Galactic archaeology, Xiang et al. (2019) employed the DD-Payne data driven framework to jointly predict 16 elemental abundances (including [O/Fe] and [Mg/Fe]) from LAMOST DR5 low-resolution ($R \approx 1800$) spectra. Their results achieve an internal precision of 0.03–0.1 dex over the general metallicity range. Li et al. (2022b) used low-resolution LAMOST spectral data and convolutional neural networks (CNNs) to estimate the abundances of various elements, including [O/Fe] and [Mg/Fe]. The average absolute error (MAE) on the test set was approximately 0.03 dex for both [O/Fe] and [Mg/Fe]. Ting & Weinberg (2022) used normalizing flow methods to reconstruct the probability distributions of 15 elemental abundances in the Galactic disk, conditioned on [Fe/H] and [Mg/Fe]. The study found that the abundance residuals for most elements, conditioned on [Fe/H] and [Mg/Fe], were dispersed by less than 0.02 dex. Ambrosch et al. (2023) utilized CNNs to predict elemental abundances such as [Mg/Fe], [Al/Fe], and [Fe/H] from the GIRAFFE spectra of the *Gaia*-ESO project. The standard deviation (σ) for [Mg/Fe] was 0.03 dex. Ferreira Lopes et al. (2025) used multi-band photometric data from Southern Photometric Local Universe Survey (S-PLUS) in combination with deep neural networks (DNNs) and random forest (RF) methods to estimate the [Mg/Fe] and [O/Fe] ratios for approximately five million stars. Furthermore, these powerful techniques are now being extended to extragalactic studies. Asa’ d et al. (2024)

applied machine learning methods to *Multi-Unit Spectroscopic Explorer* (MUSE) spectral data for the first time, successfully deriving the abundances of [O/Fe], [Mg/Fe], and other elements in the NGC1856 star cluster.

However, the aforementioned methods have certain limitations, particularly because most of them rely solely on spectral lines or photometric information in the optical range (≈ 4000 – 8000 Å), with very limited utilization of the ultraviolet (UV) region (≤ 4000 Å). Traditional line-based abundance determinations typically depend on weak or medium strength absorption features in the optical to measure O and Mg abundances. For example, for O, the [O I] 6300 Å forbidden line is extremely weak at very low metallicities and is susceptible to contamination from nearby Ni I lines. The O I 7771–7775 Å infrared triplet, while more sensitive in MP stars, depends on non-LTE corrections and is sensitive to effective temperature (T_{eff}) and surface gravity ($\log g$) (Kiselman 2001; Asplund 2005). For Mg, the Mg b triplet (≈ 5170 Å) and the Mg I 8806 Å weak line are commonly used, but in stars with [Fe/H] ≤ -2.5 dex, the strength of these lines approaches the noise level, leading to significantly increased measurement errors, and they are also susceptible to inaccuracies in the background continuum (Cayrel et al. 2004; Bergemann et al. 2017). Moreover, existing machine learning methods are largely based on low-resolution optical or near-infrared spectra (≈ 3500 – 9000 Å) from surveys such as the Sloan Digital Sky Survey (SDSS) (Kollmeier et al. 2017; Kollmeier & The SDSS-V Collaboration 2025), and LAMOST (Zhao et al. 2012; Deng et al. 2012; Liu et al. 2014; Yan et al. 2022) do not fully exploit the critical information available in the UV region. This results in a lack of feature extraction capabilities for key UV absorption features such as the OH molecular bands, Mg II 2796/2803 Å doublet, and Mg I 2852 Å line, which are relatively stronger and have better line separation in MP environments, providing more sensitive diagnostic information for O and Mg abundances (Roederer et al. 2014; Barklem & Collet 2016; Amarsi et al. 2019).

The Chinese Space Station Telescope (CSST), scheduled for launch around 2027 (CSST Collaboration 2025), is poised to fill a critical gap in current stellar archives. Our search for existing observations revealed a significant scarcity of large-scale spectroscopic data for MP stars covering the NUV range, as ground-based surveys are restricted by the atmospheric cutoff and previous space missions lacked the survey efficiency for such vast samples. CSST addresses this limitation with several unique observational advantages. It is equipped with slitless grism spectroscopy, which covers three bands: GU (255–420 nm), GV (400–650 nm), and GI (620–1000 nm), spanning from 2550 to 10 000 Å with a spectral resolution power of $R \approx 200$. This enables low-resolution spectra ranging from the near-ultraviolet (NUV) to the optical wavelengths. By operating in space, CSST can make it easier than ground-based surveys to detect weak absorption features and key UV spectral lines between 2550–4000 Å without being limited by the atmospheric cutoff at short wavelengths (Zhang et al. 2024a). Although its resolution is only $R \approx 200$, existing simulation analyses indicate that this resolution is sufficient to coarsely distinguish important features such as the OH bands, the Mg II doublet, and Mg I absorption, providing a foundation for deep learning models to extract absorption information. Furthermore, the large field of view of the CSST (~ 1.1 deg²) and high limiting magnitude enable it to cover a larger area of the sky at high Galactic latitudes, making it possible to obtain a large sample of UV-inclusive MP stars.

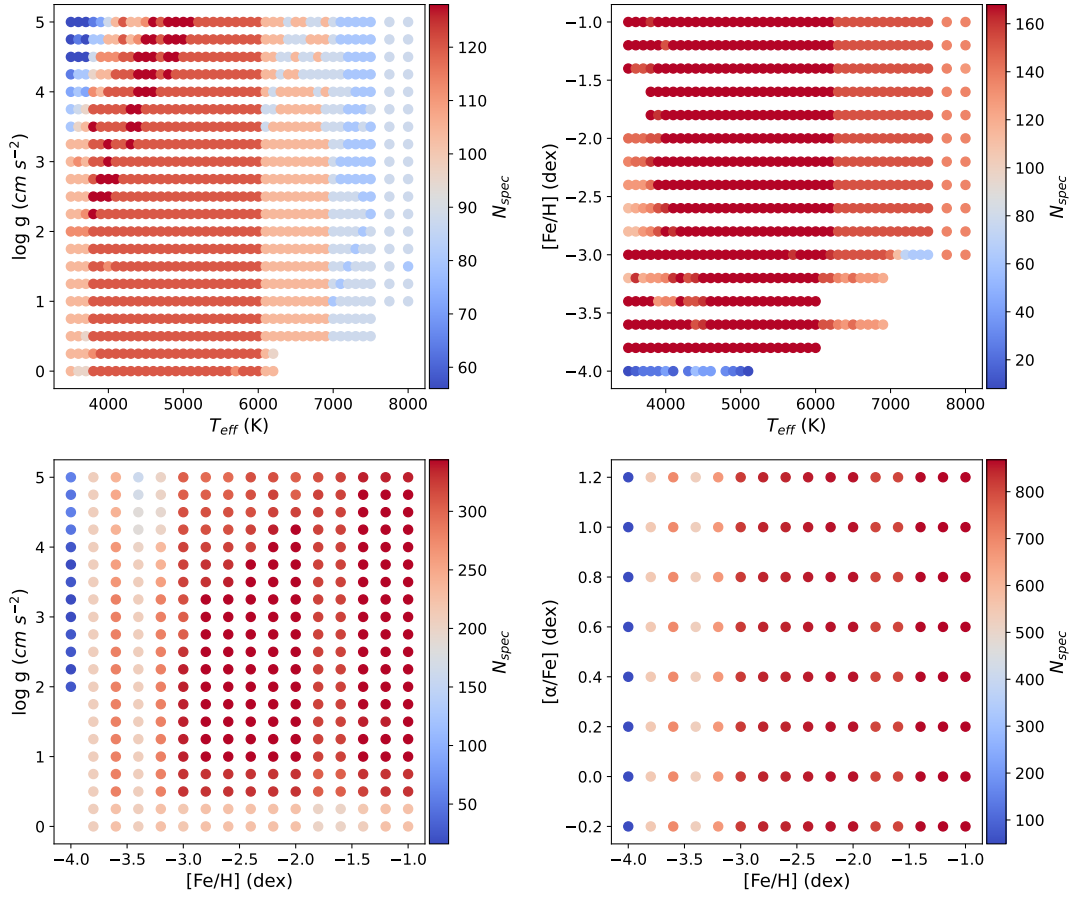


Fig. 1. Stellar atmospheric parameters of the synthetic spectra grid. Here, N_{spec} represents the number of spectra at each parameter combination.

Zhang et al. (2024b) used synthetic PHOENIX spectra data to simulate CSST observations in the 2550–10 000 Å range with $R \approx 200$. By comparing inputs with UV data (2550–4000 Å) to inputs restricted to the optical range (4000–10 000 Å), they found that adding the UV band data improves the recall in [Fe/H], $[\alpha/\text{Fe}]$, and [C/Fe] predictions even the S/N as low as 10–20. Building on this, we further quantified the importance of the UV band data in estimating the abundances of [O/Fe] and [Mg/Fe] in MP stars by developing an improved deep learning algorithm. We also systematically investigated how UV band information affects regression accuracy under varying T_{eff} , [Fe/H], and $\log g$.

The paper is organized as follows: in Section 2, we introduce the data used in this study, including the synthetic spectra and their processing methods. Section 3 provides an overview of the research method, with a focus on the self-attention mechanism and the introduction of the SPT-GloCal model. Section 4 presents the experimental results, demonstrating that the inclusion of UV data in the spectra improves prediction accuracy, as well as an error analysis across different parameter spaces. Finally, Section 5 provides a summary of the research findings and a discussion on the implications of these results for future studies.

2. Data

2.1. Synthetic spectra

This work employed the SPECTRUM v2.77 synthesis code (Gray & Corbally 1994) to perform radiative transfer calculations on

the ATLAS9 grid of stellar atmospheric models of Castelli & Kurucz (2003). We first constructed a three-dimensional grid of model atmospheres covering $T_{\text{eff}} = 3500\text{--}8000$ K ($\Delta T_{\text{eff}} = 100$ K for $T_{\text{eff}} = 3500\text{--}6000$ K, then 250 K steps to 8000 K), $\log g = 0.00\text{--}5.00$ dex ($\Delta \log g = 0.25$ dex), and [Fe/H] = -4.0 to -1.0 dex ($\Delta[\text{Fe}/\text{H}] = 0.2$), yielding 11 763 stellar atmospheric models (see Figure 1).

In the radiation transfer phase, we independently adjusted [O/Fe] and [Mg/Fe] with values of $-0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$, and 1.2 dex. The process is as follows.

First, the solar logarithmic abundances of O or Mg are read from `stdatom.dat` as $A_X^\odot = 12 + \log(N_X/N_H)$, and the metallicity [M/H] from the stellar atmospheric model header is added to all metal elements (in this study, we set $[M/H] = [\text{Fe}/\text{H}]$), resulting in

$$A_X^{\text{model}} = A_X^\odot + [\text{Fe}/\text{H}]. \quad (1)$$

Therefore, to achieve a non-zero [X/Fe], an additional offset must be applied to the abundance of the element before reading.

Then, for each grid point ($T_{\text{eff}}, \log g, [\text{Fe}/\text{H}]$) and the target differential $[X/\text{Fe}]_{\text{target}}$, the corresponding abundance in `stdatom.dat` is modified:

$$A_X^{\text{file}} = A_X^\odot + [X/\text{Fe}]_{\text{target}}. \quad (2)$$

After reading, SPECTRUM adds the [Fe/H] value, resulting in the final abundance:

$$A_X^{\text{model}} = A_X^\odot + [\text{Fe}/\text{H}] + [X/\text{Fe}]_{\text{target}}. \quad (3)$$

All spectra were computed with the `nita` switch and without the `f` option. Therefore, before output, the SPECTRUM program solves the radiative transfer equation with only continuous absorption, using the same stellar atmospheric model. The resulting composite spectrum is normalized by dividing the intensity spectrum I_λ by the pure continuum spectrum $I_{c,\lambda}$, yielding a normalized intensity spectrum with $F/F_c = 1$. If the `f` switch is used, absolute flux is output instead. The distinction between the default “normalized-intensity spectrum” mechanism and the `f` switch is detailed in the document (Gray & Corbally 1994). For further explanation of the physical background and observational normalization techniques, see Gray (2021).

Additionally, we assigned microturbulent velocities v_{turb} based on $(T_{\text{eff}}, \log g, [\text{Fe}/\text{H}])$ in the following segments (García Pérez et al. 2016; Holtzman et al. 2018): For $\log g \geq 3.5$ dex and $T_{\text{eff}} \geq 5000$ K,

$$v_t = 1.1 + 10^{-4}(T_{\text{eff}} - 5500) + 4 \times 10^{-7}(T_{\text{eff}} - 5500)^2 \text{ km s}^{-1}; \quad (4)$$

for $\log g \geq 3.5$ dex and $T_{\text{eff}} < 5000$ K,

$$v_t = 1.0 \text{ km s}^{-1}; \quad (5)$$

for $\log g < 3.5$ dex and $[\text{Fe}/\text{H}] \geq -1$,

$$v_t = 10^{0.226 - 0.0228 \log g + 0.0297 \log g^2 - 0.0113 \log g^3} \text{ km s}^{-1}; \quad (6)$$

for $\log g < 3.5$ dex and $[\text{Fe}/\text{H}] < -1$,

$$v_t = 2.478 - 0.325 \log g \text{ km s}^{-1}. \quad (7)$$

The initial resolution of the synthetic spectra is 0.02 \AA , covering a wavelength range from 2550 to 10 000 \AA . Each stellar atmospheric model corresponds to 16 different values of $[\text{O}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, totaling approximately $11\,763 \times 16 \approx 1.88 \times 10^5$ spectra for subsequent regression analysis.

2.2. Low-resolution processing

As described in Section 1, CSST operates with a low spectral resolution of approximately $R \approx 200$. Its wide field of view ($\sim 1.1 \text{ deg}^2$) and high limiting magnitude allow it to survey high Galactic latitude regions, thus identifying a large sample of MP stars together with crucial UV spectra. To evaluate the impact of including the UV band data on regression models for MP stars $[\text{O}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, this study quantifies the performance improvement enabled by UV spectra coverage and investigates the underlying spectroscopic reasons. To this end, we simulate low-resolution spectra as observed by CSST to provide data more closely resembling real observations, thereby supporting the development of deep learning algorithms. These algorithms are specifically designed to predict $[\text{O}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$ from spectra that include UV spectra coverage under low resolution ($R \approx 200$), which is introduced in detail in Section 3.

The core idea of the resolution degradation process is to use a convolution based method to blur the fine structures in the original high resolution spectra to match the target resolution. In this process, we construct the output wavelength grid $\{\Lambda_j\}$ using a constant resolution $R \equiv \Lambda/\Delta\Lambda$. Starting from $\Lambda_0 = \lambda_{\text{min}}$, the grid is constructed recursively as

$$\Lambda_{j+1} = \Lambda_j + \frac{\Lambda_j}{R}, \quad j = 0, 1, \dots, N-1, \quad (8)$$

until $\Lambda_{N-1} \leq \lambda_{\text{max}}$. Each wavelength point has an effective width of $\Delta\Lambda_j \approx \frac{\Lambda_j}{R}$ to maintain constant resolution over the entire wavelength range.

For each target wavelength Λ_j , we performed a local Gaussian convolution on the trimmed high resolution spectrum $\{\lambda_i, F_i\}$ to mimic the point spread function (PSF) of the spectrograph. The full width at half maximum (FWHM) of the Gaussian kernel is given by

$$\text{FWHM}(\Lambda_j) = \frac{\Lambda_j}{R}, \quad (9)$$

and the kernel is normalized to unit area as

$$G_j(\lambda) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left[-\frac{(\lambda - \Lambda_j)^2}{2\sigma_j^2}\right], \quad (10)$$

where $\sigma_j = \frac{\text{FWHM}(\Lambda_j)}{2\sqrt{2 \ln 2}}$.

The downsampled spectral flux at each Λ_j is computed by

$$\tilde{F}(\Lambda_j) = \frac{\int_{\Lambda_j-5\text{FWHM}}^{\Lambda_j+5\text{FWHM}} F(\lambda) G_j(\lambda) d\lambda}{\int_{\Lambda_j-5\text{FWHM}}^{\Lambda_j+5\text{FWHM}} G_j(\lambda) d\lambda}. \quad (11)$$

To accelerate the processing of nearly 1.9×10^5 spectra, Python `multiprocessing.Pool` was employed for parallel computation. For each target wavelength $\{\Lambda_j\}$, we first prepared extended subarrays $\{\lambda_i, F_i\}$, then distributed the convolution tasks across multiple CPU cores, and finally collected the downsampled spectra $\{\Lambda_j, \tilde{F}\}$. To reduce the computational load, the convolution range was truncated to $\pm 5\text{FWHM}$, which significantly cuts down computation time. In practice, continuous integration is replaced by discrete summation over the clipped original wavelength grid. The convolution for each Λ_j point was computed using the kernel function (e.g., Gaussian function), and normalized by dividing by the total kernel weight.

Figure 2 presents a comparison between the original high resolution spectra and the degraded low-resolution version. The distinction between the two spectra is evident, highlighting the effect of resolution degradation on spectral features.

2.3. Spectral analysis in different parameter spaces

Before quantitatively estimating the experiments, we first performed a qualitative analysis of the simulated CSST low-resolution spectra in the UV band, investigating the spectral features under different $[\text{O}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$ abundance conditions to assess the potential contribution of the UV band to abundance estimation. Overall, the UV band data exhibited significant spectral line features in all of parameter spaces.

Figure 3 shows the variation of the OH molecular band and Mg lines in the UV band at different T_{eff} . The depth of the OH molecular band initially increases and then decreases with rising T_{eff} , with the peak occurring around K-type stars (approximately 4000–5000 K). At low T_{eff} (around 3500 K, M-type stars), OH molecules are abundant, but the UV radiation is weak, causing the spectral lines to be inconspicuous. At intermediate T_{eff} (around 4000–5000 K, from K-type to early G-type stars), UV radiation increases, and OH molecules still remain abundant, leading to a peak in line strength. At high T_{eff} (> 5500 K, from G-type to F- and A-type stars), the temperature rise causes OH molecules to dissociate extensively, weakening the spectral lines.

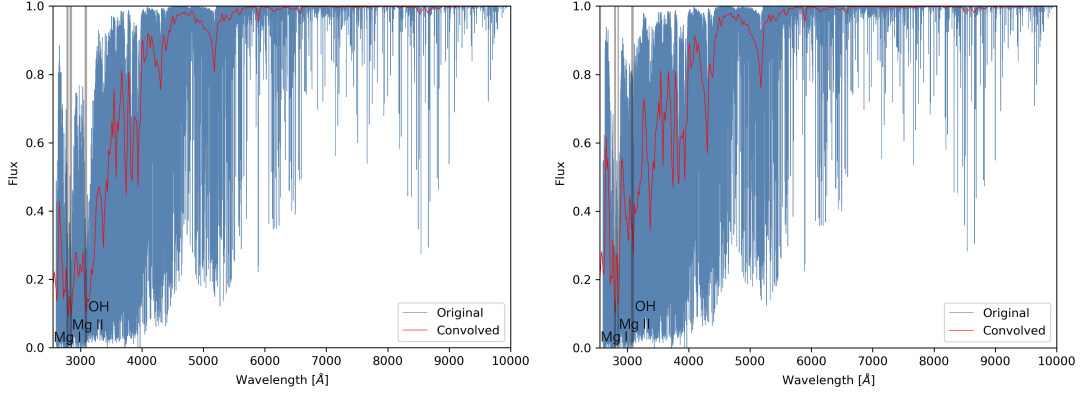


Fig. 2. Comparison between the original high resolution spectra (blue) and the low-resolution spectra (red) degraded to $R \approx 200$ through Gaussian convolution. Both panels cover the 2550–10 000 Å range, with the left panel showing an MP star with $[O/Fe] = 0.4$, $[Mg/Fe] = 0.0$, $T_{\text{eff}} = 4700$ K, $\log g = 5.0$, and $[Fe/H] = -2.0$, and the right panel showing an MP star with $[O/Fe] = 0.0$, $[Mg/Fe] = 0.4$, and the same other parameters.

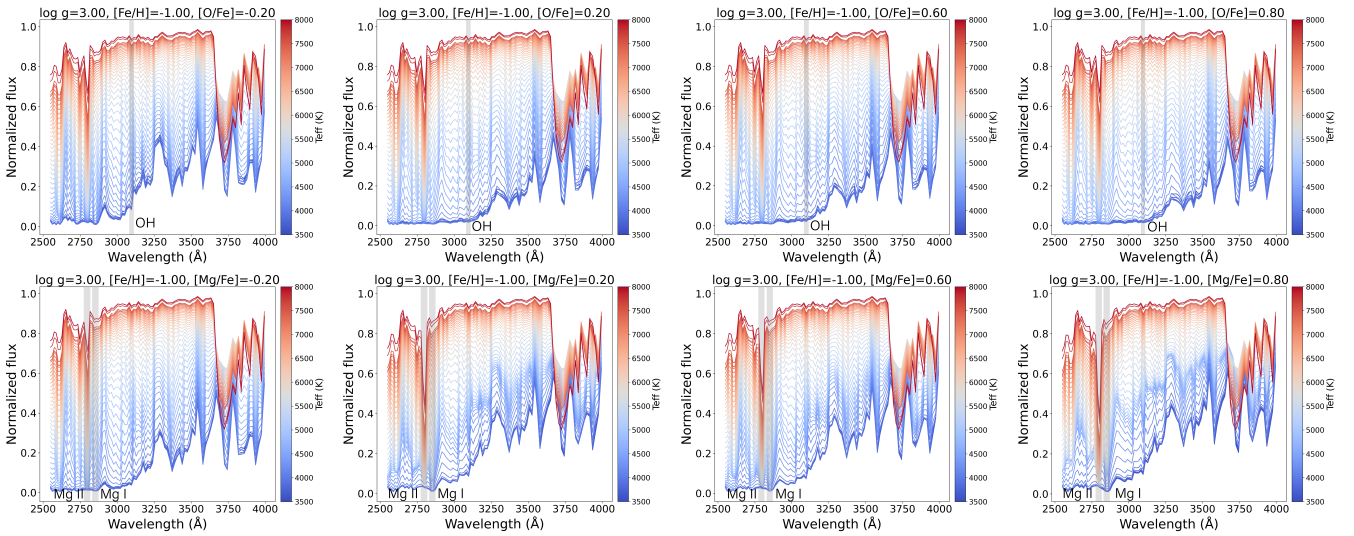


Fig. 3. Variation of UV spectral flux as a function of T_{eff} with fixed parameters ($\log g = 3.0$, $[Fe/H] = -1.0$). Upper row: variation for the $[O/Fe]$ ratio of -0.2 , $+0.2$, $+0.6$, and $+0.8$. Lower row: same as the upper row but for $[Mg/Fe]$ ratios. The depth of the OH molecular band first increases and then decreases with rising T_{eff} , with the peak occurring around K-type stars (approximately 4000–5000 K). Similarly, the depth of the Mg I line increases and then decreases with rising T_{eff} , with the peak occurring around K-type to G-type stars (approximately 4000–5800 K), while the Mg II line continuously strengthens with increasing T_{eff} .

By late F-type and A-type stars (approximately 7000–8000 K), the OH molecular band virtually disappears.

Similarly, the depth of the Mg I line first increases and then decreases with increasing T_{eff} , with the peak occurring around K-type to G-type stars (approximately 4000–5800 K). At low T_{eff} (around 3500 K, M-type stars), magnesium predominantly exists in its neutral Mg I form. At intermediate T_{eff} (approximately 4000–5800 K, from K-type to G-type stars), as the T_{eff} increases, the excitation effect dominates, enhancing the line strength; however, the ionization effect begins to emerge. At high T_{eff} (>6000 K, from F-type to A-type stars), the ionization effect becomes dominant, reducing the amount of neutral Mg I, causing the line strength to decrease. By 8000 K in A-type stars, the Mg I line becomes extremely weak. Similarly, the Mg II line strength also follows a trend of first increasing and then decreasing with rising T_{eff} .

Figure 4 shows the effect of different $\log g$ on the OH molecular band and Mg absorption lines in the UV band. As the $\log g$ increases, the depth of the OH molecular band becomes deeper and stronger. An increase in $\log g$ results in higher

atmospheric pressure, which promotes the formation of OH molecules, thereby enhancing the strength of the OH absorption band in the UV band.

At the same time, the intensity of the Mg I line increases with increasing $\log g$, while the Mg II line weakens. This is because, as $\log g$ increases, the higher atmospheric pressure leads to a more significant recombination of Mg II to Mg I, resulting in a decrease in Mg II and an increase in Mg I. Therefore, for stars with similar T_{eff} , the Mg I lines of dwarf stars (high $\log g$) are typically stronger than those of giant stars (low $\log g$), while the Mg II lines of giant stars are typically stronger than those of dwarf stars.

Figure 5 illustrates the impact of $[Fe/H]$ on key UV spectral features. It is important to note that the variations observed in the OH and Mg lines as a function of $[Fe/H]$ arise from the design of the synthetic spectra. Specifically, the absolute elemental abundances are coupled through the standard relation, $[Mg/H] = [Mg/Fe] + [Fe/H]$, where $[Mg/Fe]$ better reflects the established empirical trends of Galactic chemical evolution (GCE).

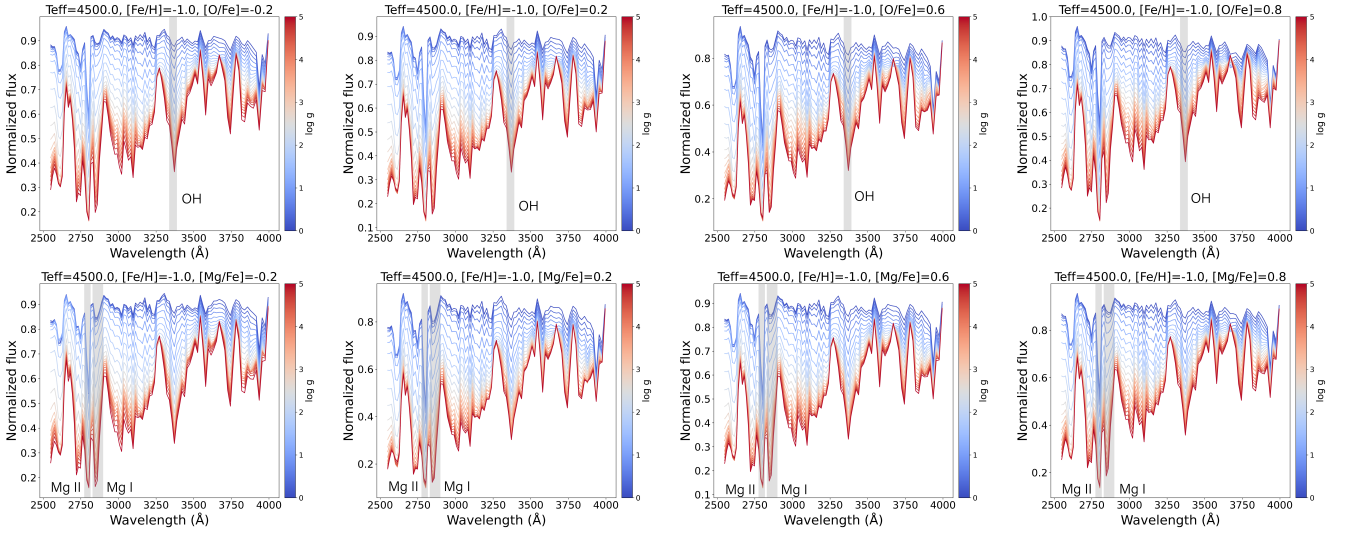


Fig. 4. Similar to Figure 3, but showing the variation of UV spectral flux as a function of $\log g$. The synthetic spectra were calculated for fixed parameters ($T_{\text{eff}} = 4500$ K, $[\text{Fe}/\text{H}] = -1.0$). Upper row: variation for the $[\text{O}/\text{Fe}]$ ratios of -0.2 , $+0.2$, $+0.6$, and $+0.8$. Lower row: same as the upper row but for $[\text{Mg}/\text{Fe}]$ ratios. As the $\log g$ increases, the depth of the OH molecular band becomes deeper. The intensity of the Mg I line increases with increasing $\log g$, while the Mg II line weakens.

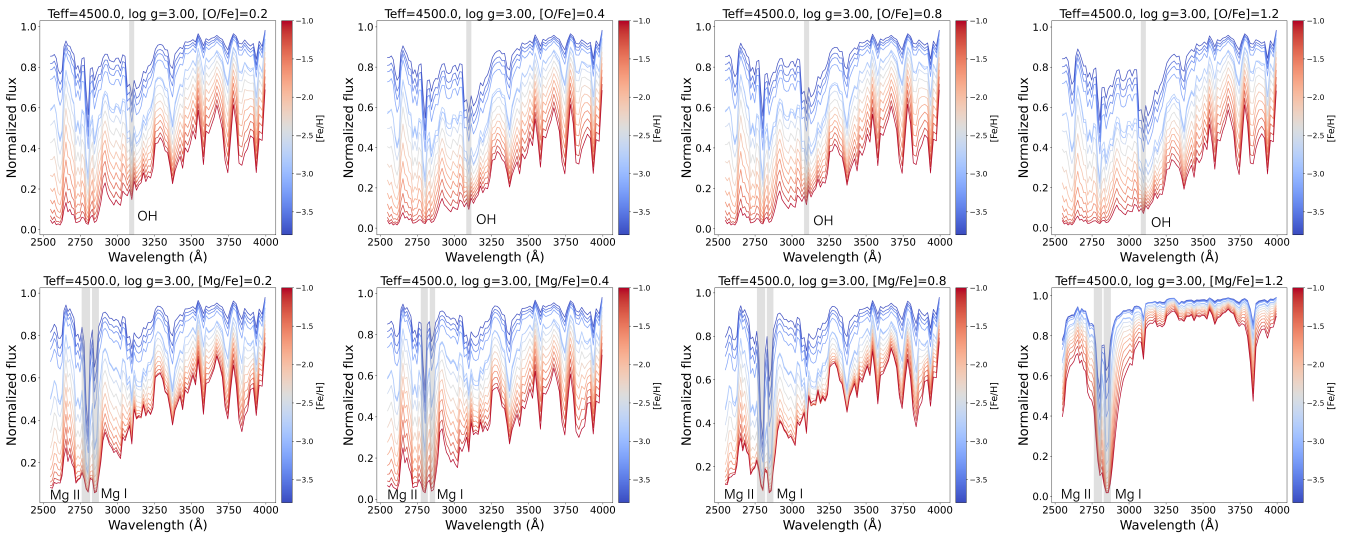


Fig. 5. Similar to Figures 3 and 4, but showing the variation of UV spectral flux as a function of $[\text{Fe}/\text{H}]$. The synthetic spectra were calculated for fixed parameters ($T_{\text{eff}} = 4500$ K, $\log g = 3.00$). Upper row: variation for the $[\text{O}/\text{Fe}]$ ratios of $+0.2$, $+0.4$, $+0.8$, and $+1.2$. Lower row: same as the upper row but for $[\text{Mg}/\text{Fe}]$ ratios. As $[\text{Fe}/\text{H}]$ decreases, the strength of the OH molecular band, Mg II doublet, and Mg I line gradually increases, with the Mg II doublet showing a more significant increase. When $[\text{Fe}/\text{H}] < -2.8$, the strength of the Mg II doublet surpasses that of the Mg I line, and as $[\text{Fe}/\text{H}]$ decreases further, the strength difference between the two increases.

As depicted in the figure, with decreasing $[\text{Fe}/\text{H}]$, the strengths of OH molecular band, the Mg II doublet, and the Mg I line progressively increase, with the enhancement of the Mg II doublet being the most significant. For metallicities where $[\text{Fe}/\text{H}] < -2.8$, the strength of the Mg II doublet surpasses that of the Mg I line. As $[\text{Fe}/\text{H}]$ decreases further, the difference in strength between these two features is amplified.

3. Methods

3.1. Preliminary

Self-attention, a cornerstone of the Transformer architecture (Vaswani et al. 2017), is a mechanism that allows a model to weigh the importance of different elements in an input sequence

when processing a specific element. It enables the model to capture long range dependencies and contextual relationships within the sequence itself.

Let the input sequence be represented by a matrix $\mathbf{X} \in \mathbb{R}^{n \times d_{\text{model}}}$, where n is the sequence length and d_{model} is the dimension of the embedding for each token. From this input, we derive three distinct matrices: the query (\mathbf{Q}), key (\mathbf{K}), and value (\mathbf{V}) matrices. These are produced by multiplying the input matrix \mathbf{X} with three separate, learnable weight matrices $\mathbf{W}^{\mathbf{Q}} \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $\mathbf{W}^{\mathbf{K}} \in \mathbb{R}^{d_{\text{model}} \times d_k}$, and $\mathbf{W}^{\mathbf{V}} \in \mathbb{R}^{d_{\text{model}} \times d_v}$: $\mathbf{Q} = \mathbf{X}\mathbf{W}^{\mathbf{Q}}$; $\mathbf{K} = \mathbf{X}\mathbf{W}^{\mathbf{K}}$; $\mathbf{V} = \mathbf{X}\mathbf{W}^{\mathbf{V}}$. Here, d_k is the dimension of the keys and queries, and d_v is the dimension of the values.

The core of the self-attention mechanism is the Scaled Dot-Product Attention function. The output is computed as a

weighted sum of the values, where the weight assigned to each value is determined by the dot-product similarity of the query with its corresponding key. The formula is as follows:

$$\mathbf{O}_{original} = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (12)$$

In this equation, the term \mathbf{QK}^T calculates the dot product between every query vector in \mathbf{Q} and every key vector in \mathbf{K} , resulting in a matrix of attention scores that reflects the similarity between each pair of positions in the sequence. To prevent the dot products from growing excessively large and pushing the softmax function into regions with vanishing gradients, the scores are scaled by a factor of $\frac{1}{\sqrt{d_k}}$. The softmax function is then applied to normalize the scaled scores into a probability distribution, yielding the attention weights. Finally, this matrix of attention weights is multiplied by the Value matrix \mathbf{V} to produce the output of the self-attention layer. This output provides a new representation for each token that is a weighted average of all token values in the sequence, thereby encoding rich contextual information.

3.2. GloCal-Attention

To address the issue of scattered attention in traditional self-attention mechanisms, Zhang et al. (2024c) proposed an attention-enhanced operator and constructs the spectral transformer (SPT) model based on it. The SPT model was designed to estimate the age and mass of red giant stars from their spectra. Its core innovation is the multi-head Hadamard Self-Attention (multi-head HSA) mechanism. Unlike conventional self-attention that relies on dot-product similarity, the HSA operator computes attention scores using the Hadamard product (i.e., element-wise multiplication) between normalized Query and Key matrices. This is combined with an enhanced_softmax function to create a sharper and more focused attention distribution, better suited for continuous spectral data. However, as the sequence length increases, self-attention can still suffer from the problem of scattered attention. A typical example in transformers is the degradation of their performance and an increased tendency for hallucination as the input context length grows in large language models.

It can be fundamentally attributed to the mathematical properties of the self-attention mechanism, the core component of the Transformer architecture. The primary bottlenecks arise from attention score dilution. A critical vulnerability emerges from the interplay between the \mathbf{QK}^T operation and the subsequent softmax normalization, especially as the sequence length, n , grows. The softmax function transforms the raw attention scores into a probability distribution, forcing the sum of weights across all n tokens to equal one. As n becomes exceedingly large (e.g., hundreds of thousands of tokens), the denominator of the softmax function, $\sum_{j=1}^n e^{z_j}$, inflates due to the sheer volume of tokens. This leads to a phenomenon known as attention dilution. The weights assigned to a few genuinely critical tokens are suppressed, as the finite probability mass is distributed across a vast number of irrelevant or noisy tokens. The cumulative effect of these numerous, albeit individually small, weights on their corresponding Value vectors (\mathbf{V}) can introduce significant noise that overwhelms the signal from a few pertinent tokens. This loss of focus prevents the model from concentrating its computational resources on the most relevant parts of the context, thereby increasing the likelihood of generating outputs that are unfaithful to the source material.

Therefore, to alleviate the issue of scattered attention in self-attention mechanisms under long-sequence spectra, we propose a self-attention method called SPT-GloCal, which combines a score-aware backbone attention with global attention. This approach aims to help self-attention better focus on the phenomenon of ‘‘perceptual isolation’’ caused by long sequences, where the attention scores across the sequence tend to approach zero more easily.

Score-aware information flow. In the context of long sequence data processing, the conventional softmax function, by its nature of normalizing scores into a probability distribution that sums to unity, can inadvertently dilute the influence of highly relevant features. The fixed total probability mass must be distributed across all elements in the sequence, which can lead to significant information being obscured by a long tail of low weight, noisy elements. To address this limitation, we introduce a score-aware information flow mechanism. This methodology is engineered to iteratively prune insignificant candidate items and re-emphasize the most salient information, thereby refining the information flow within the attention computation. The general algorithmic procedure is depicted in Figure 6.

First, we defined the model’s hyperparameters: the number of window layers n , the input sequence length l , and the number of competitive filtering layers p . The function Ψ refers to the enhanced softmax as detailed in the work by Zhang et al. (2024c). Within the i -th sliding window and at the r -th competitive layer, we have:

$$\text{Score}(Q_i, K_i)_r = \Psi(\text{ReLU}(Q_i) \circ \text{ReLU}(K_i)d^*). \quad (13)$$

In this equation, $Q_i \in \mathbb{R}^{n_w \times d^*}$ and $K_i \in \mathbb{R}^{n_w \times d^*}$ are the query and key matrices within the i -th window of length n_w . The operator \circ denotes the Hadamard product (i.e., element-wise multiplication), and d^* is the feature dimension of the query and key vectors at this competitive layer.

For each window, we select the top- n elements that satisfy a cumulative score contribution threshold α to proceed to the next competitive layer. The set of indices \mathbf{I}_{ir} is identified, and the scores for the subsequent layer are composed exclusively of the elements corresponding to these indices. This is formally expressed as:

$$\mathbf{I}_{ir} = \left\{ j \mid \sum \text{rank}(\text{Score}(Q_i, K_i)_r)[j] < \alpha \right\} \quad (14)$$

$$\text{Score}(Q_i, K_i)_{r+1} = \{(\text{Score}(Q_i, K_i)_r)[j] \mid j \in \mathbf{I}_{ir}\}. \quad (15)$$

Finally, the competitive process terminates when the number of candidate elements falls below a threshold γ . The output of the Score-aware information flow can then be represented as:

$$\mathbf{O}_{flow} = \text{Score}(Q_i, K_i)_p \circ V_p d^* \quad (16)$$

where V_p denotes the value vectors corresponding to the attention scores retained in the final layer. In essence, the proposed procedure decouples the competitive selection mechanism of the softmax function from its inherent normalization constraint. This is achieved by first acquiring principal information from local regions through a layer wise competitive filtering process, followed by a final normalization step.

Overall information flow. To integrate the global context with the salient local information derived from our proposed mechanism, the final output is formulated as a weighted combination:

$$\mathbf{O}_{final} = \sum_{c \in \mathcal{C}} g^c \mathbf{O}_c, \quad (17)$$

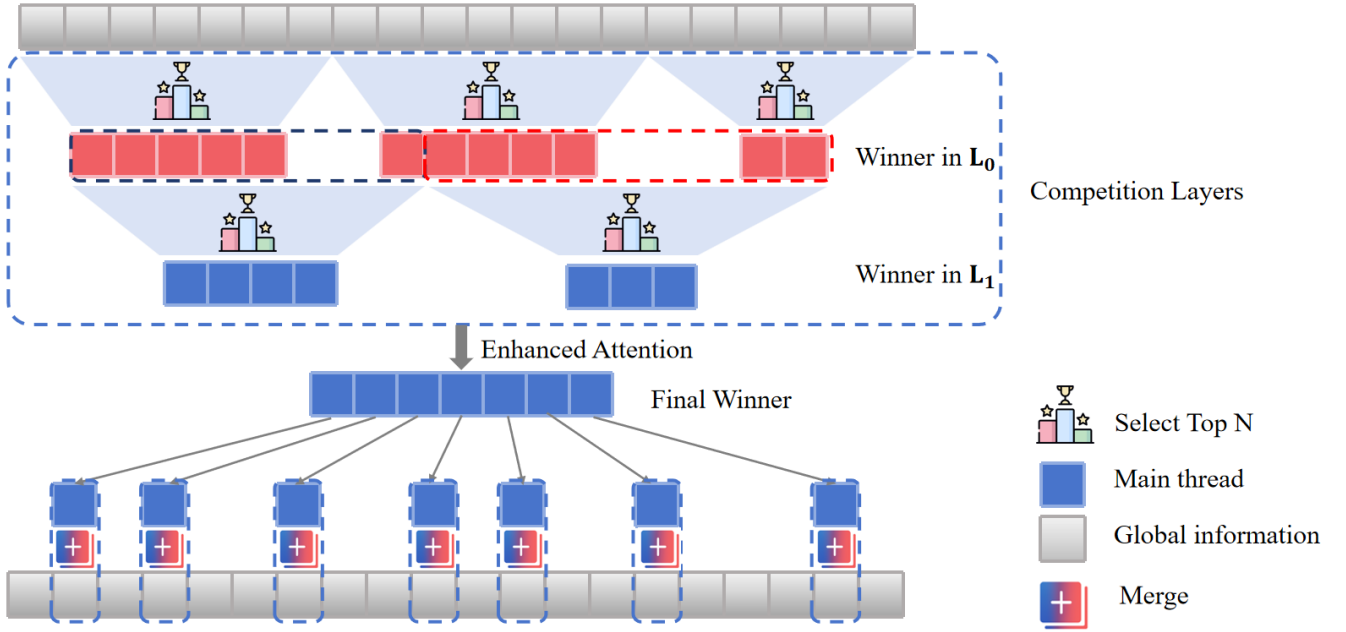


Fig. 6. Model framework. It employs a multi-round competitive selection mechanism. The process begins with input tokens (gray squares). In long sequences, the standard softmax function introduces a competition that can diminish the informational contribution of many tokens to near zero. To counteract this and decouple the selection process (“competition”) from the intrinsic value of the information, our model iteratively identifies and discards less salient features through multiple rounds. The “winner” tokens from each round are highlighted within the boxes outlined with blue dashes. The final set of “Winner” tokens, representing the most critical local information, is then integrated with the initial sequence to enrich the model’s overall perceptual field.

where the component set is $C = \{\text{original, flow}\}$. The term $g^c \in [0, 1]$ is a learnable gating parameter that adaptively balances the contribution from the original attention output ($\mathbf{O}_{\text{original}}$) and the score-aware flow output (\mathbf{O}_{flow}).

3.3. Training details

To train and evaluate our SPT-GloCal model, we first randomly partitioned the full dataset of approximately 188 000 synthetic spectra into a training set (80%), a validation set (10%), and an independent test set (10%). The model is designed to take a low-resolution spectrum as input and regress the [O/Fe] and [Mg/Fe] abundance ratios.

The model architecture is based on the Vision Transformer (ViT), with the standard self-attention mechanism replaced by our proposed GloCal-Attention. The overall configuration adopts an embedding dimension of 768, 12 transformer blocks, and 12 attention heads per block, resulting in approximately 85.2 million trainable parameters. For model training, we adopted the following strategies:

Optimization. We employed the Adam optimizer for the training process. Its parameter settings were fixed at $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ to ensure stable and efficient convergence.

Learning rate policy. An initial learning rate of 10^{-4} was used. To refine convergence in later stages, we applied a decay strategy, reducing the learning rate by a factor of 0.8 every 100 epochs.

Regularization. To mitigate overfitting, L2 regularization with a weight decay coefficient of 5×10^{-4} was applied. This encourages the model to learn smoother and more generalizable weight distributions.

Weight initialization. All model weights were initialized using a truncated normal distribution. This strategy was chosen to facilitate a stable starting point for the optimization.

Early stopping. An early stopping mechanism was implemented to prevent overfitting and save computational resources. If the model’s performance on the validation set did not show improvement over ten consecutive epochs, the training was terminated. The final model parameters were taken as the average of the weights from these last ten epochs to enhance robustness.

To investigate the impact of the UV data, two separate models were trained with identical configurations: one using the full spectral range (2550–10 000 Å) and another using only the optical range (4000–10 000 Å).

3.4. Evaluation metrics

To quantitatively assess the predictive accuracy of our models, we used two standard statistical metrics: the mean absolute error (MAE) and the standard deviation of the residuals (σ). All performance evaluations were conducted on the independent test set. The metrics are defined as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (18)$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N ((y_i - \hat{y}_i) - \bar{r})^2}, \quad (19)$$

where N is the total number of samples in the test set, y_i is the true abundance ratio for the i -th sample, \hat{y}_i is the corresponding model prediction, and \bar{r} is the mean of all residuals ($y_i - \hat{y}_i$).

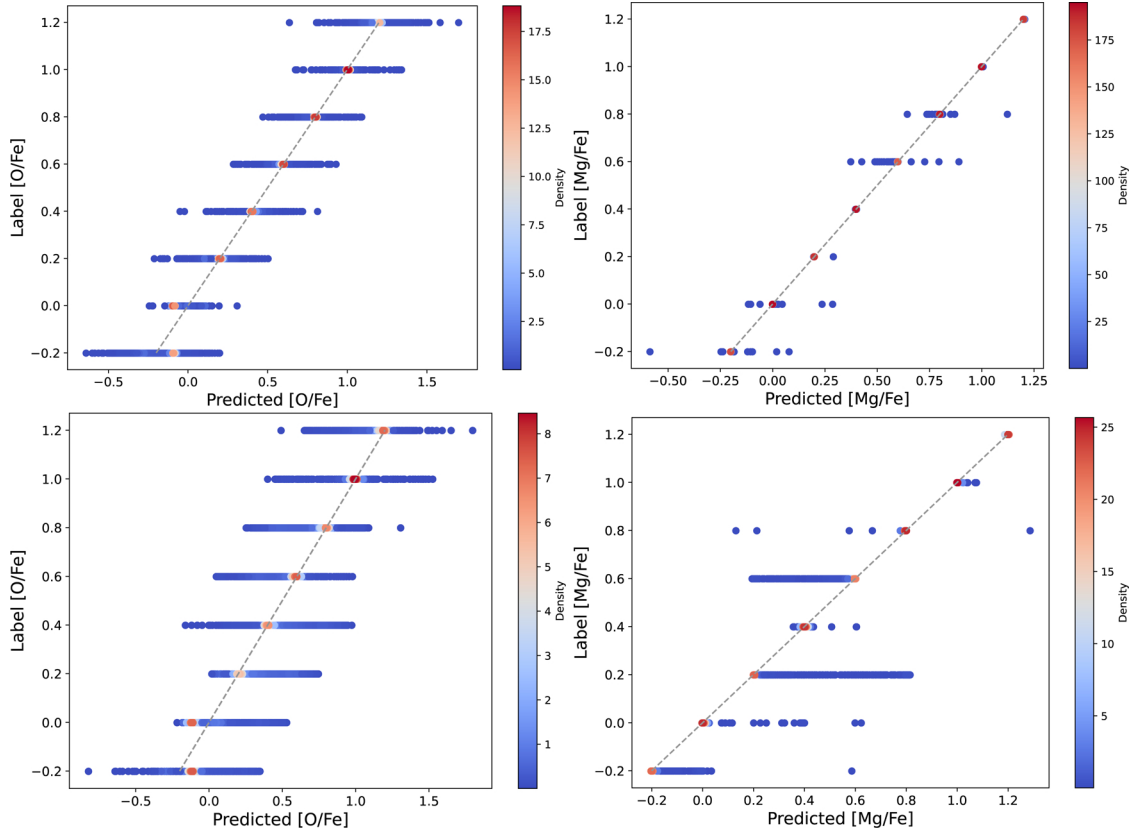


Fig. 7. Predicted versus true [O/Fe] (left) and [Mg/Fe] (right) on the test set. Upper row: model trained on 2550–10 000 Å spectra. Lower row: model trained on 4000–10 000 Å spectra. Colors encode the local point density; the dashed line marks the 1:1 relation.

Table 1. Statistical comparison of prediction errors for [O/Fe] and [Mg/Fe] over the 2550–10 000 Å and 4000–10 000 Å spectral ranges.

Band (Å)	[O/Fe]		[Mg/Fe]	
	σ	MAE	σ	MAE
2550–10 000	0.0629	0.0367	0.0068	0.0007
4000–10 000	0.1346	0.0785	0.0541	0.0010

Notes. σ is the standard deviation of the residuals and MAE is the mean absolute error, both computed on the test set.

4. Results

4.1. Global regression performance of SPT-GloCal

We evaluated the performance of the SPT-GloCal model on the independent test set using the metrics defined in Section 3.4. Table 1 summarizes the prediction accuracy achieved by SPT-GloCal when the UV band (2550–4000 Å) is included or removed from the input spectra. For [O/Fe] the standard deviation of the residuals decreases from $\sigma = 0.135$ to $\sigma = 0.063$, and the MAE is halved, from 0.078 to 0.037. The effect is even more striking for [Mg/Fe]: the error scatter shrinks by almost an order of magnitude (from $\sigma = 0.054$ to $\sigma = 0.0068$) and the MAE drops from 0.010 to only 6.5×10^{-4} .

Figure 7 shows the regression results for [O/Fe] and [Mg/Fe]. The upper row present the full band model (2550–10 000 Å), where the majority of the predicted values for both

[O/Fe] and [Mg/Fe] closely follow the 1:1 relation. The lower row, based on a model trained only with the optical band (4000–10 000 Å), show systematic deviations in the predictions for [O/Fe] across all abundances, while [Mg/Fe] exhibits a broader tail at low abundances. These results highlight that, although the classical indicators for oxygen and magnesium are primarily found in the optical band, the diagnostic information provided by the UV range is crucial for improving prediction accuracy.

4.2. Error analysis in different parameter spaces

In different physical parameter spaces, we analyzed the predictive performance of [O/Fe] and [Mg/Fe]. As shown in Figures 8 and 9, the data that included the UV band demonstrated a better predictive performance for [O/Fe] and [Mg/Fe] in terms of T_{eff} , $\log g$, [Fe/H], and elemental distribution compared to the data without the UV band, with smaller corresponding mean squared error (MSE) and σ . Moreover, the prediction uncertainties in the data with UV band exhibited relatively smaller fluctuations, resulting in more stable predictions that were less affected by atmospheric parameters.

In detail, Figure 8 shows that the improvement in the UV band is most pronounced in the 4500–6000 K range, with the difference being significantly higher than in other temperature ranges. This indicates that the UV band spectra effectively reduces the estimation errors of [O/Fe] at these temperatures. For $\log g$, the MSE and σ values for the non-UV band data increase significantly with higher $\log g$, while the MSE and σ of the [O/Fe] estimates with the UV band data show a slight decreasing trend and smaller fluctuations. From the difference in the error curves, it is evident that the impact of the UV

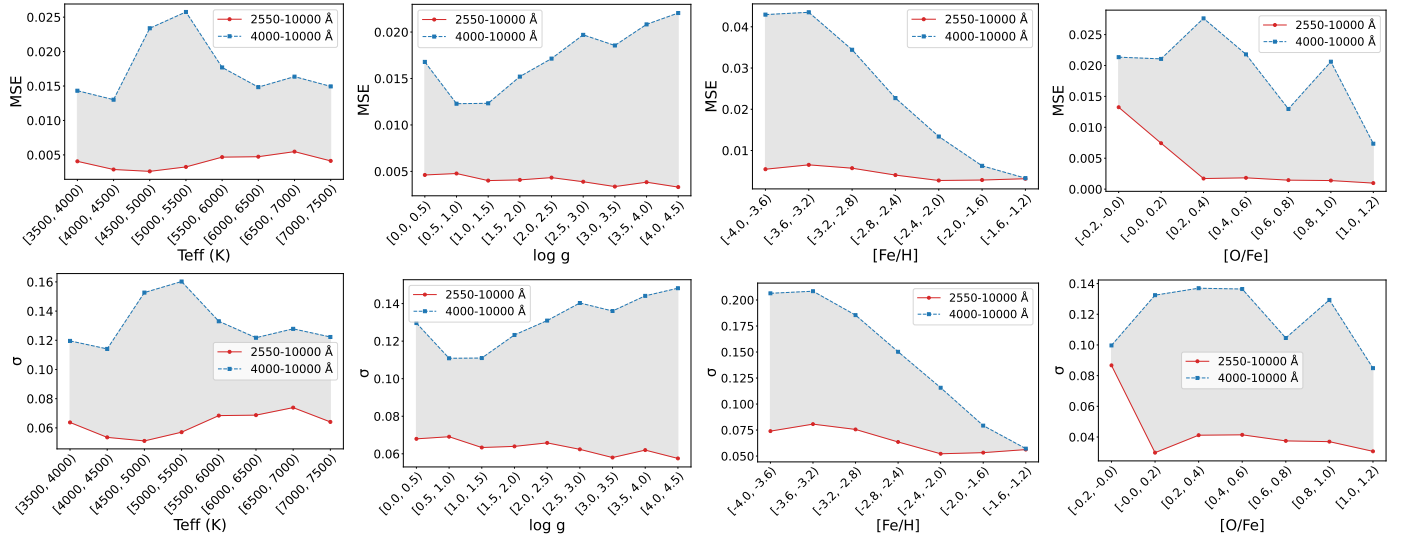


Fig. 8. Mean squared error (left) and σ (right) of the $[\text{O}/\text{Fe}]$ prediction error across the four-dimensional parameter space (T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{O}/\text{Fe}]$). Red solid line: including the UV band data (2550–10 000 Å); blue dashed line: optical only (4000–10 000 Å).

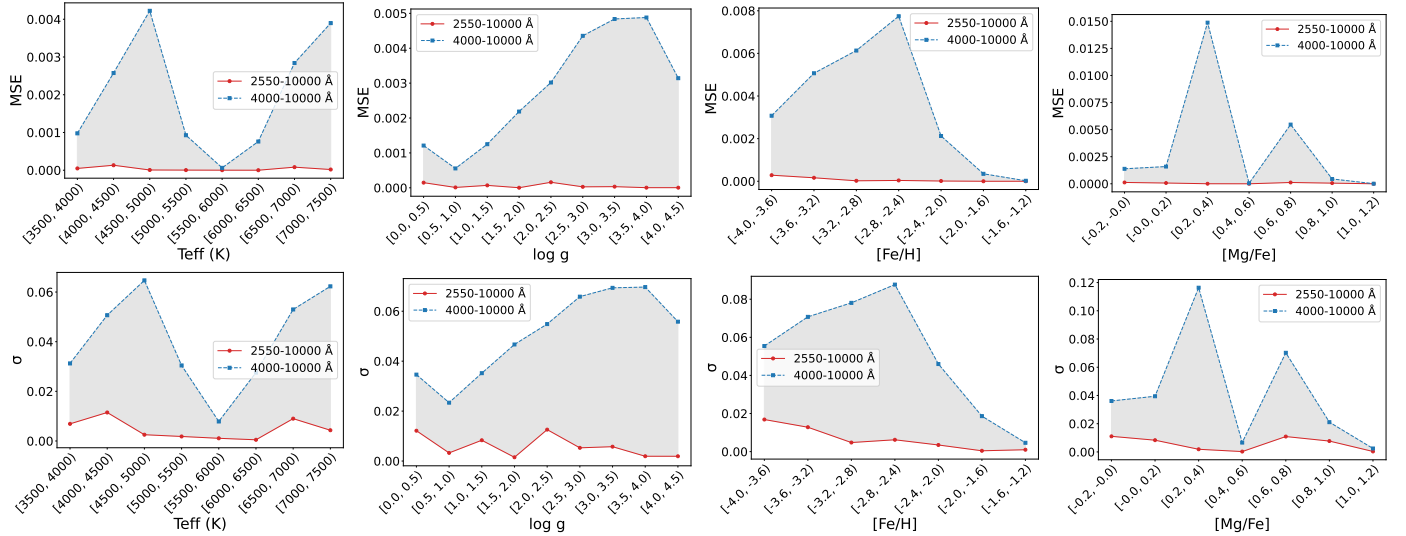


Fig. 9. Same as Figure 8, but showing the $[\text{Mg}/\text{Fe}]$ prediction error.

band spectra on improving the estimation accuracy of $[\text{O}/\text{Fe}]$ becomes more significant as $\log g$ increases (with the difference between the curves increasing). Similarly, for $[\text{Fe}/\text{H}]$, as its value decreases, the estimation errors of $[\text{O}/\text{Fe}]$ without the UV band spectra increase monotonically in terms of MSE and σ . This result supports the general finding in the literature that $[\text{O}/\text{Fe}]$ estimations for MP stars are typically challenging and subject to significant uncertainties, particularly when relying solely on standard optical diagnostics (e.g., Zhao et al. 2016; Amarsi et al. 2019). However, the introduction of the UV band spectra leads to a reduction in both MSE and σ , which provides new confidence for future estimates of $[\text{O}/\text{Fe}]$ for MP stars and highlights the important role in the UV band plays in this process. Additionally, we observe that as the value of $[\text{O}/\text{Fe}]$ increases, the estimation accuracy improves, which is consistent with the trend that the estimation accuracy of $[\text{O}/\text{Fe}]$ increases as $[\text{Fe}/\text{H}]$ decreases, since lower $[\text{Fe}/\text{H}]$ generally correlates with higher abundance of α -elements.

In Figure 9, for T_{eff} , the improvement with the UV band data is more noticeable in the 3500–5300 K and 6300–8000 K ranges,

Table 2. Performance of SPT-GloCal versus tree boosting baselines in estimating $[\text{O}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$ over the 2550–10 000 Å spectral ranges.

Method	$[\text{O}/\text{Fe}]$		$[\text{Mg}/\text{Fe}]$	
	σ	MAE	σ	MAE
XGBoost	0.1262	0.0615	0.0401	0.0071
LightGBM	0.1358	0.0640	0.0521	0.0096
CatBoost	0.1269	0.0617	0.0401	0.0073
SPT	0.0736	0.0453	0.0079	0.0007
SPT-GloCal	0.0629	0.0367	0.0068	0.0006

Notes. The optimal results are **bolded**.

indicating that the UV band spectra is more effective at reducing the estimation errors of $[\text{Mg}/\text{Fe}]$ in these temperature ranges. In comparison, the UV band data improves the $[\text{Mg}/\text{Fe}]$ estimation accuracy less than $[\text{O}/\text{Fe}]$, which is expected in our study.

Table 3. Comparison of abundance determination performance for [Mg/Fe] and [O/Fe] across different methods and signal-to-noise ratios (S/N).

[X/Fe]	S/N	XGBoost		LightGBM		CatBoost		SPT		SPT-GloCal	
		σ	MAE	σ	MAE	σ	MAE	σ	MAE	σ	MAE
[O/Fe]	30	0.6826	0.8314	0.6951	0.8412	0.6862	0.8438	0.3676	0.1375	0.3526	0.1314
	40	0.3674	0.1793	0.3779	0.1808	0.3817	0.1728	0.3221	0.1327	0.3131	0.1221
	50	0.3258	0.1229	0.3081	0.1326	0.3271	0.1340	0.2271	0.0907	0.2215	0.0776
[Mg/Fe]	30	0.2923	0.0942	0.2806	0.0934	0.2905	0.0891	0.1415	0.0600	0.1357	0.0543
	40	0.1860	0.0590	0.2031	0.0633	0.1957	0.0717	0.0923	0.0515	0.0921	0.0497
	50	0.1654	0.0476	0.1533	0.0406	0.1323	0.0391	0.0568	0.0313	0.0487	0.0223

Notes. The metrics shown are the σ and MAE. The best results from our proposed method are **bolded**.

This is because, for [Mg/Fe], the estimation is already better with or without the UV band spectra than for [O/Fe], and thus the potential for further improvement is smaller. For $\log g$, similar to [O/Fe], the MSE and σ for the data without the UV band increase with increasing $\log g$, while the [Mg/Fe] estimates with the UV band data show smaller fluctuations. The difference in error curves indicates that the effect of the UV band spectra on improving [Mg/Fe] estimation accuracy becomes more significant as $\log g$ increases. For [Fe/H], as its value decreases, the MSE and σ curves for the data with the UV band exhibit smaller fluctuations, with only a small difference compared to the non-UV data. For the errors in the distribution of [Mg/Fe] values, the non-UV data show larger fluctuations and a bimodal pattern, while the errors with the UV band data are much smaller, indicating a more stable prediction.

4.3. Model comparison

We compared the performance of the SPT-GloCal model with the SPT model proposed by Zhang et al. (2024c), which is designed to process low-resolution spectral data with UV band to predict [Fe/H] and $[\alpha/\text{Fe}]$, as well as several benchmark models based on tree boosting (XGBoost, LightGBM, and CatBoost) for predicting the [O/Fe] and [Mg/Fe] ratios of MP stars.

Table 2 presents the performance of different models in estimating the [O/Fe] and [Mg/Fe] abundance ratios over the spectral range of 2550–10 000 Å. The best results are highlighted in the table. As shown, SPT-GloCal outperformed the other models in terms of prediction accuracy for both [O/Fe] and [Mg/Fe]. Specifically, SPT-GloCal achieves a σ of 0.0629 and a MAE of 0.0367 for [O/Fe], outperforming XGBoost ($\sigma = 0.1262$, MAE = 0.0615), LightGBM ($\sigma = 0.1358$, MAE = 0.0640), and CatBoost ($\sigma = 0.1269$, MAE = 0.0617). Similarly, for [Mg/Fe], SPT-GloCal also performs exceptionally well, with a σ of 0.0068 and an MAE of 0.0006, surpassing the other benchmark models.

Compared to the SPT model, SPT-GloCal shows an improvement in performance. For [O/Fe], the σ of SPT is 0.0736 and the MAE is 0.0453, both higher than the 0.0629 and 0.0367 of SPT-GloCal. For [Mg/Fe], the σ of SPT is 0.0079 and the MAE is 0.0007, still higher than those of SPT-GloCal. The introduction of a global attention mechanism in SPT-GloCal enhances the ability of the SPT model to process long sequence information, particularly in low-resolution spectral data. This improvement allows the model to capture key information more effectively, leading to enhanced prediction accuracy.

4.4. Performance on noisy spectra

To assess the robustness of our model under more realistic observational conditions, we conducted an additional analysis by

introducing noise to the model spectra. We generated three versions of the test set by adding Gaussian noise to simulate average signal-to-noise ratios of 30, 40, and 50 per pixel. We then evaluated the performance of SPT-GloCal and all baseline models on these noisy spectra.

The results are summarized in Table 3. As expected, the precision of all models deteriorates as the S/N decreases. For instance, for [O/Fe], the SPT-GloCal model yields predictions where the σ increases from 0.2215 at S/N=50 to 0.3526 at S/N=30. A similar trend is observed for all other methods and for [Mg/Fe] predictions.

Despite the general performance degradation, SPT-GloCal consistently outperforms all other models across all tested S/N. For [O/Fe] at a challenging S/N of 30, SPT-GloCal achieves a σ of 0.3526, which is marginally better than its predecessor SPT ($\sigma = 0.3676$) and substantially superior to the tree-based models (e.g., XGBoost with $\sigma = 0.6826$). For [Mg/Fe], SPT-GloCal also maintains a clear advantage, demonstrating its superior ability to extract meaningful features even from noise-contaminated, low-resolution spectra. This analysis confirms that the architectural improvements in SPT-GloCal provide a tangible advantage in noise-limited scenarios, reinforcing its potential for reliable application to real CSST survey data.

5. Conclusion

This study presents a comprehensive framework for improving the abundance estimation of [O/Fe] and [Mg/Fe] in MP stars through the integration of UV spectral information and a novel deep learning model, SPT-GloCal. Motivated by the known limitations of traditional and machine learning methods in the optical only regime, especially for MP stars, we explored the diagnostic power of UV band spectra (2550–4000 Å), which is accessible to upcoming CSST survey.

To simulate realistic low-resolution observations, we generated over 180 000 synthetic spectra using SPECTRUM v2.77, spanning a wide stellar parameter grid. These spectra were processed to match the resolution of CSST ($R \approx 200$) and were used to train and evaluate our models. Our analyses revealed that the key UV absorption features including OH molecular bands and Mg II/Mg I lines provide robust diagnostic signals for [O/Fe] and [Mg/Fe], especially in parameter regimes where optical lines are weak or absent.

We introduced the SPT-GloCal model, which extends the self-attention backbone with a score-aware information flow mechanism and a global-local integration strategy. This design addresses the attention dilution problem in long sequence spectral inputs, enabling the model to better preserve and extract relevant signals from complex spectra.

Quantitative results demonstrate the superiority of our approach. The inclusion of UV data reduced the MAE of [O/Fe] predictions from 0.0785 to 0.0367, and for [Mg/Fe] from 0.0010 to 0.0006. Error analyses across T_{eff} , $\log g$, and [Fe/H] further confirmed that UV band spectra enhance predictive stability, especially in challenging regions such as high surface gravity or low metallicity. Compared to the baseline SPT model and tree based regressors (XGBoost, LightGBM, CatBoost), SPT-GloCal consistently achieved lower residual errors, validating the effectiveness of both its architectural design and the UV inclusive spectral input.

This work highlights the underutilized potential of the UV spectra domain in stellar spectroscopy and demonstrates that even at low resolution, UV spectral features can play a critical role in stellar chemical analysis. The SPT-GloCal framework offers a scalable and physically informed solution for abundance regression, well suited to the next generation of space based spectroscopic surveys. In future work, we plan to apply this model to real CSST and UV survey data, extend its application to other α - and iron-peak elements, and incorporate uncertainty quantification into the model outputs.

Acknowledgements. This work is supported from the Strategic Priority Research Program of Chinese Academy of Sciences, grant No. XDB1160101. This study is also supported by the Natural Science Foundation of Shandong Province under grant No. ZR2024MA063, No. ZR2022MA076, and No. ZR2022MA089, the science research grants from the China Manned Space Project with No. CMS-CSST-2021-B05 and CMS-CSST-2021-A08, and the National Natural Science Foundation of China (NSFC) under grants No. 12090044, 11873037, 11803016, and 12573109.

References

- Amarsi, A. M., Nissen, P. E., Asplund, M., Lind, K., & Barklem, P. S. 2019, *A&A*, **622**, L4
- Ambrosch, M., Guiglion, G., Mikolaitis, Š., et al. 2023, *A&A*, **672**, A46
- Arnose, E., Ryan, S. G., Argast, D., Norris, J. E., & Beers, T. C. 2005, *A&A*, **430**, 507
- Asa'd, R., Hernandez, S., John, J. M., et al. 2024, *AJ*, **167**, 265
- Asplund, M. 2005, *ARA&A*, **43**, 481
- Barklem, P. S., & Collet, R. 2016, *A&A*, **588**, A96
- Beers, T. C., & Christlieb, N. 2005, *ARA&A*, **43**, 531
- Bergemann, M., Collet, R., Amarsi, A. M., et al. 2017, *ApJ*, **847**, 15
- Bond, H. E., Nelan, E. P., VandenBerg, D. A., Schaefer, G. H., & Harmer, D. 2013, *ApJ*, **765**, L12
- Bonifacio, P., Caffau, E., François, P., & Spite, M. 2025, arXiv e-prints [arXiv:2504.06335]
- Carretta, E., Gratton, R. G., & Sneden, C. 2000, *A&A*, **356**, 238
- Carretta, E., Bragaglia, A., Gratton, R., & Lucatello, S. 2009, *A&A*, **505**, 139
- Castelli, F., & Kurucz, R. L. 2003, in *IAU Symposium*, 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, A20
- Cayrel, R., Depagne, E., Spite, M., et al. 2004, *A&A*, **416**, 1117
- Chen, K.-J., Whalen, D. J., Wollenberg, K. M. J., Glover, S. C. O., & Klessen, R. S. 2017, *ApJ*, **844**, 111
- Chiaki, G., & Tominaga, N. 2020, *MNRAS*, **498**, 2676
- CSST Collaboration (Gong, Y., et al.) 2025, arXiv e-prints [arXiv:2507.04618]
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *Res. Astron. Astrophys.*, **12**, 735
- Ferreira Lopes, C. E., Gutiérrez-Soto, L. A., Ferreira Alberice, V. S., et al. 2025, *A&A*, **693**, A306
- Frebel, A. 2010, *Astron. Nachr.*, **331**, 474
- Frebel, A., & Ji, A. P. 2023, arXiv e-prints [arXiv:2302.09188]
- Frebel, A., & Norris, J. E. 2015, *ARA&A*, **53**, 631
- Frebel, A., Aoki, W., Christlieb, N., et al. 2005, *Nature*, **434**, 871
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, **151**, 144
- Gratton, R. G., Carretta, E., Matteucci, F., & Sneden, C. 2000, *A&A*, **358**, 671
- Gray, D. F. 2021, *The Observation and Analysis of Stellar Photospheres*, 4th edn. (Cambridge University Press)
- Gray, R. O., & Corbally, C. J. 1994, *AJ*, **107**, 742
- Guiglion, G. 2025, *A&A*, **699**, A191
- Hartwig, T., Ishigaki, M. N., Klessen, R. S., & Yoshida, N. 2019, *MNRAS*, **482**, 1204
- Hernandez, S., Larsen, S., Trager, S., Groot, P., & Kaper, L. 2017, *A&A*, **603**, A119
- Hernandez, S., Larsen, S., Trager, S., Kaper, L., & Groot, P. 2018, *MNRAS*, **476**, 5189
- Holtzman, J. A., Hasselquist, S., Shetrone, M., et al. 2018, *AJ*, **156**, 125
- Ishigaki, M. N., Chiba, M., & Aoki, W. 2012, *ApJ*, **753**, 64
- Ishigaki, M. N., Tominaga, N., Kobayashi, C., & Nomoto, K. 2018, *ApJ*, **857**, 46
- Iwamoto, N., Umeda, H., Tominaga, N., Nomoto, K., & Maeda, K. 2005, *Science*, **309**, 451
- Joggerst, C. C., & Whalen, D. J. 2011, *ApJ*, **728**, 129
- Joggerst, C. C., Almgren, A., Bell, J., et al. 2010, *ApJ*, **709**, 11
- Kiselman, D. 2001, *New A Rev.*, **45**, 559
- Kobayashi, C., Umeda, H., Nomoto, K., Tominaga, N., & Ohkubo, T. 2006, *ApJ*, **653**, 1145
- Kolborg, A. N., Martizzi, D., Ramirez-Ruiz, E., et al. 2022, *ApJ*, **936**, L26
- Kollmeier, J., & The SDSS-V Collaboration 2025, in *American Astronomical Society Meeting Abstracts*, 245, American Astronomical Society Meeting Abstracts, 159.21
- Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints [arXiv:1711.03234]
- Krauss, L. M., & Chaboyer, B. 2003, *Science*, **299**, 65
- Larsen, S. S., Brodie, J. P., Wasserman, A., & Strader, J. 2018, *A&A*, **613**, A56
- Larsen, S. S., Eitner, P., Magg, E., et al. 2022, *A&A*, **660**, A88
- Li, H., Aoki, W., Matsuno, T., et al. 2022a, *ApJ*, **931**, 147
- Li, Z., Zhao, G., Chen, Y., Liang, X., & Zhao, J. 2022b, *MNRAS*, **517**, 4875
- Liu, X. W., Yuan, H. B., Huo, Z. Y., et al. 2014, in *IAU Symposium*, 298, Setting the scene for Gaia and LAMOST, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 310
- Magrini, L., Randich, S., Kordopatis, G., et al. 2017, *A&A*, **603**, A2
- Marino, A. F., Villanova, S., Milone, A. P., et al. 2011, *ApJ*, **730**, L16
- Mashonkina, L. 2013, *A&A*, **550**, A28
- Milone, A. D. C., Sansom, A. E., & Sánchez-Blázquez, P. 2011, *MNRAS*, **414**, 1227
- Palla, M., Santos-Peral, P., Recio-Blanco, A., & Matteucci, F. 2022, *A&A*, **663**, A125
- Roederer, I. U., Preston, G. W., Thompson, I. B., et al. 2014, *AJ*, **147**, 136
- Santos-Peral, P., Sánchez-Blázquez, P., Vazdekis, A., & Palicio, P. A. 2023, *A&A*, **672**, A166
- Shi, J., Zhao, G., Liu, S., et al. 2025, *ApJ*, **983**, 127
- Sneden, C., Boesgaard, A. M., Cowan, J. J., et al. 2023, *ApJ*, **953**, 31
- Susa, H., Hasegawa, K., & Tominaga, N. 2014, *ApJ*, **792**, 32
- Timmes, F. X., Woosley, S. E., & Weaver, T. A. 1995, *ApJS*, **98**, 617
- Ting, Y.-S., & Weinberg, D. H. 2022, *ApJ*, **927**, 209
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Asplund, M. 2018, *ApJ*, **860**, 159
- Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in *Advances in Neural Information Processing Systems*, 30, eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Curran Associates, Inc.)
- Welsh, L., Cooke, R., Fumagalli, M., Pettini, M., & Rudie, G. C. 2024, *A&A*, **691**, A285
- Wu, F., Bu, Y., Zhang, M., et al. 2023, *AJ*, **166**, 88
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, *ApJS*, **245**, 34
- Xing, Q.-F., Zhao, G., Liu, Z.-W., et al. 2023, *Nature*, **618**, 712
- Yan, H., Li, H., Wang, S., et al. 2022, *The Innovation*, **3**, 100224
- Zhang, J., Tang, B., Chang, J., et al. 2024a, *Res. Astron. Astrophys.*, **24**, 015011
- Zhang, M., Bu, Y., Wu, F., et al. 2024b, *A&A*, **691**, A21
- Zhang, M., Wu, F., Bu, Y., et al. 2024c, *A&A*, **683**, A163
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *Res. Astron. Astrophys.*, **12**, 723
- Zhao, G., Mashonkina, L., Yan, H. L., et al. 2016, *ApJ*, **833**, 225