DENOISING MEDIUM RESOLUTION STELLAR SPECTRA WITH U-NET CONVOLUTIONAL NEURAL NETWORKS*

Balázs Pál ⓘ

Department of Physics of Complex Systems, ELTE Eötvös Loránd University
Budapest, Hungary
and
Institute for Particle and Nuclear Physics
HUN-REN Wigner Research Centre for Physics
Budapest, Hungary

László Dobos ⓘ

Department of Physics & Astronomy, The Johns Hopkins University
Maryland, USA
and
Department of Information Systems, Eötvös Loránd University
Budapest, Hungary

We investigated the use of a U-Net convolutional neural network for denoising simulated medium-resolution stellar spectroscopic observations, generated under realistic observational conditions resembling the Subaru Prime Focus Spectrograph (PFS). We found that our U-Net model effectively captured spectral features, achieving an average relative error of around 1% across diverse stellar parameters, despite a limited training set of only 1000 observations and a relatively short training period. Although U-Net did not reach the performance previously demonstrated by denoising autoencoders (DAEs) trained extensively on larger datasets, it outperformed DAEs trained under similarly constrained conditions. These results indicate that the U-Net architecture offers rapid, robust feature learning and may be particularly advantageous in scenarios involving initial denoising, subsequently refined by more accurate, but otherwise slower deep-learning models.

---

1. Introduction

Modern stellar spectroscopic surveys often face substantial variations in data quality due to large differences in target brightness, resulting in spectra spanning 6–7 magnitudes in apparent brightness and widely varying signal-to-noise ratios (S/N). Faint, low S/N spectra challenge conventional analysis techniques, such as continuum normalization and template fitting, often leading to non-convergent fits or erroneous parameter estimates. Overcoming these limitations is crucial for Galactic Archaeology (GA), which relies on spectroscopy to reconstruct the chemical and dynamical history of the Milky Way and nearby galaxies. New survey instruments (*e.g.* DESI [1, 2] or Subaru PFS [3, 4]) can now observe very distant and faint stars — including red giants in the Andromeda Galaxy [2] — offering unprecedented opportunities to study small-scale dark matter dynamics and the early formation of galaxies [4].

To address these observational challenges, machine-learning techniques such as denoising autoencoders (DAEs) have been successfully employed, demonstrating the capability to recover clean stellar spectra from noisy observations at S/N as low as 10 with approximately 1% average accuracy [5]. Trained on simulated stellar spectroscopic data, DAEs quickly reconstruct noise-free spectra from noisy inputs, enabling near-real-time denoising. This rapid inference capability allows DAEs to be integrated into observation pipelines as immediate on-the-fly tools or as initial approximation steps to enhance the convergence and accuracy of traditional spectroscopic analyses.

In this study, we explore an alternative deep-learning framework for denoising stellar spectroscopic observations: a U-Net convolutional neural network. The U-Net architecture — originally developed for image segmentation [6] — employs an encoder–decoder structure with skip connections, potentially capturing broad continuum features and sharp spectral lines at the same time more effectively than the simpler fully-connected DAE used in our previous study. We demonstrate that a U-Net model quickly learns robust spectral features early in the training process and achieves reasonable denoising performance even with a very limited training set (*e.g.*, 1000 observations in fewer than 20 epochs). However, in the long term, it does not match the performance of dense autoencoders. These results suggest that U-Net may be advantageous in knowledge distillation scenarios, where stellar spectroscopic observations are initially denoised using a U-Net model and subsequently refined using a dense autoencoder.

2. Simulating stellar spectroscopic observations

To train our neural network model, we generated simulated stellar spectra using the BOSZ synthetic spectrum grid [7]. Observational conditions

(seeing, target zenith angle, moon illumination) and instrument parameters were selected to closely resemble typical observing conditions at Mauna Kea Observatory using the Subaru Prime Focus Spectrograph (PFS) in medium-resolution mode ($R \approx 5000$, spectral range 710–885 nm).

Fundamental stellar parameters (T_{eff}, $\log g$, metallicity, α-element abundances, Doppler shift, magnitude, extinction) were uniformly sampled across broad intervals, covering stellar types M to G. We also included cooler stars to increase spectral variability, to provide a more robust test of the denoising capabilities. Some parameter combinations were non-physical but were intentionally kept to further enhance model robustness.

Each simulated spectrum was produced in two main steps: first, interpolation of the BOSZ grid to randomly selected stellar parameters, applying sub-pixel Doppler shifts; second, application of observational effects including interstellar extinction, instrumental broadening, detector pixelization, and sky noise. The Subaru PFS Exposure Time Calculator [8] was adapted and precomputed to efficiently simulate large spectral datasets, ensuring realistic noise modeling.

During training, further data augmentation included generating new photon-noise realizations in each epoch and introducing wavelength-dependent flux calibration distortions (up to 2%) to simulate systematic calibration errors. Finally, spectra were normalized to a median flux of 0.5 within 6500–9500 Å. We present a typical example from our simulated dataset in Fig. 1.
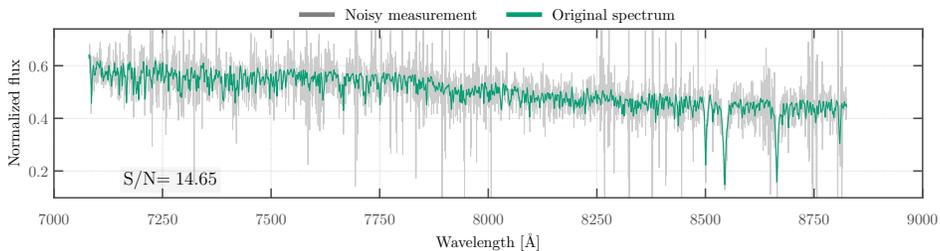

Fig. 1. An example for a typical simulated spectroscopic observation of a star with stellar parameters [Fe/H] = 0.75, $T_{\text{eff}} = 5000$ K, $\log g \sim 2.2$, [α/Fe] = 0.0, with and without a random realization of the noise added to the flux (grey and black curves, respectively). The signal-to-noise value indicated in the bottom left corner is the median pixel-wise signal-to-noise ratio over all pixels.

3. Training

We implemented a U-Net convolutional neural network to denoise stellar spectroscopic observations, closely following the original U-Net architecture.

Our network consists of four encoder–decoder levels, each comprising convolutional blocks with filter counts of 32, 64, 128, and 256 at successive depths, complemented by max-pooling operations and a bottleneck layer between the encoder and decoder. Our simulated spectra contain 4096 wavelength bins, which are gradually downsampled to 256 dimensions per filter at the bottleneck layer and subsequently upsampled back to 4096 bins at the output.

The training procedure closely mirrored that of our previous study. However, in this study, both the training and validation datasets contained 1,000 synthetic spectra each, which were augmented at each epoch as detailed in Section 2. We optimized the U-Net using a stochastic gradient descent optimizer, minimizing the mean squared error (MSE) loss function over 50 epochs. Batch normalization and ReLU activations were applied in all layers, except the output layer, where we employed a Sigmoid activation. Additionally, we applied a dropout rate of 0.3 exclusively in the bottleneck layer to mitigate overfitting. The model was trained on a single NVIDIA V100 GPU, with a batch size of 100 and an initial learning rate of 0.1, which was reduced by a factor of 10 after 20 epochs.

We did not continue training the model beyond 50 epochs, as we observed that the training and validation losses consistently plateaued after this point.

4. Results

We characterize the denoising performance by evaluating the relative errors across the entire validation dataset, comparing the denoised spectra to the original noiseless spectra. Specifically, we calculate the pixel-wise relative error as

$$|R_\lambda| = \frac{|D_\lambda[\hat{F}_\lambda] - F_\lambda|}{F_\lambda} \, , \tag{1}$$

where $D_\lambda[\hat{F}_\lambda]$ denotes the denoised spectrum, $\hat{F}_\lambda$ the noisy input, and F_λ the noiseless simulated spectrum.

In Fig. 2, we present the pixel-wise average and maximum of these relative errors as a function of four key stellar parameters: signal-to-noise ratio (S/N), effective temperature (T_{eff}), stellar surface gravity ($\log g$), and metallicity ([Fe/H]). As expected, the denoising accuracy improves with increasing S/N, showing a pronounced degradation in performance at lower S/N values. The performance is relatively stable across the investigated temperature, gravity, and metallicity ranges, with errors slightly ramping up at higher temperatures. Similarly, errors increase significantly at higher metallicities, reflecting the richer spectral line structure in metal-rich stars.
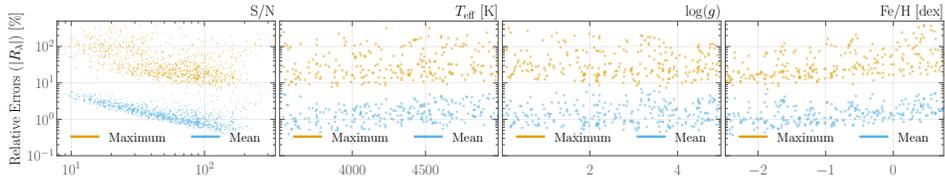
Fig. 2. The pixel-wise average and maximum of the relative errors from denoising, evaluated across the entire validation dataset, are shown as a function of relevant parameters such as signal-to-noise ratio (S/N), effective temperature (T_{eff}), stellar surface gravity ($\log g$), and metallicity ([Fe/H]).

5. Conclusions

We demonstrated that the U-Net architecture is effective at denoising stellar spectroscopic observations, achieving average relative errors around 1% across a wide range of fundamental stellar parameters and signal-to-noise ratios, before it reaches a knowledge plateau. While this level of performance is notable given the limited training dataset and relatively short training duration, it remains approximately an order of magnitude higher than what was achieved previously using denoising autoencoders (DAEs) for a dataset 100 times larger and trained for 7500 epochs, but an order of magnitude lower than the performance of DAEs trained on the small dataset for just 50 epochs.

Our results suggest that although the U-Net does not reach the same precision as dense DAEs in the long term, it offers rapid feature learning and robustness even with limited data. Consequently, U-Net architectures may be particularly beneficial for knowledge distillation approaches, providing efficient initial denoising that can be subsequently refined by more specialized autoencoder models.

Analysis was carried out on the Volta server, as well as on the Ampere A100 GPU server of the Wigner Scientific Computing Laboratory (WSCLAB) at the HUN-REN Wigner Research Centre for Physics, Hungary.

REFERENCES

[1] DESI Collaboration (B. Abareshi *et al.*), *Astron. J.* **164**, 207 (2022), arXiv:2205.10939 [astro-ph.IM].

[2] A. Dey *et al.*, *Astrophys. J.* **944**, 1 (2023), arXiv:2208.11683 [astro-ph.GA].

[3] H. Sugai *et al.*, «Progress with the Prime Focus Spectrograph for the Subaru Telescope: a massively multiplexed optical and near-infrared fiber spectrograph», in: S.K. Ramsay, I.S. McLean, H. Takami (Eds.) «Proc. SPIE, Ground-based and Airborne Instrumentation for Astronomy V» Vol. 9147, July 2014, arXiv:1408.2825 [astro-ph.IM].

[4] M. Takada *et al.*, *Publ. Astron. Soc. Jpn.* **66**, R1 (2014), arXiv:1206.0737 [astro-ph.CO].

[5] B. Pál, L. Dobos, *Astron. Nachr.* **345**, e20240049 (2024), arXiv:2409.11625 [astro-ph.IM].

[6] O. Ronneberger, P. Fischer, T. Brox, arXiv:1505.04597 [cs.CV].

[7] R.C. Bohlin *et al.*, *Astron. J.* **153**, 234 (2017), arXiv:1704.00653 [astro-ph.SR].

[8] C.M. Hirata *et al.*, arXiv:1204.5151 [astro-ph.IM].