A&A, 715, A101 (2026)
https://doi.org/10.1051/0004-6361/202660827
© The Authors 2026

Astronomy & Astrophysics

Detection of metal absorption lines in quasar spectra

A neural network approach using U-Net

Elena Sofia Mangola[1,2,⋆], Francesco Pistis[2,3,4], Michele Fumagalli[2,5], Matteo Fossati[2,6], Ting-Yun Cheng[7], Ryan J. Cooke[8], Rajeshwari Dutta[9], Ignasi Pérez-Ràfols[10], Matthew Pieri[1], and Emanuel Gafton[11]

[1] Aix Marseille Univ. CNRS, CNES, LAM, Marseille, France
[2] Dipartimento di Fisica 'G. Occhialini', Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
[3] National Centre for Nuclear Research, ul. Pasteura 7, 02-093 Warsaw, Poland
[4] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, I-40129 Bologna, Italy
[5] INAF – Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34143 Trieste, Italy
[6] INAF – Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Via E. Bianchi 46, 23807 Merate, Italy
[7] Kapteyn Astronomical Institute, University of Groningen, Landleven 12 (Kapteynborg, 5419), 9747 AD Groningen, The Netherlands
[8] Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK
[9] IUCAA, Postbag 4, Ganeshkind, Pune 411007, India
[10] Departament de Física, EEBE, Universitat Politècnica de Catalunya, c/Eduard Maristany 10, 08930 Barcelona, Spain
[11] Isaac Newton Group of Telescopes, Apartado 321, 38700 Santa Cruz de la Palma, Tenerife, Spain

Received 7 May 2026 / Accepted 18 September 2026

ABSTRACT

Context. Current and future large spectroscopic surveys are significantly enhancing the volume and resolution of quasar spectra that are observed, motivating the creation of efficient and precise automated techniques to detect absorption features.
Aims. This study focuses on the detection of metal absorption features using a novel U-Net model on WEAVE-like mock spectra in the quasar rest-frame wavelength interval 1230 Å $\leq \lambda_{\rm RF} \leq$ 3095 Å. We test the network performance for absorption detection both on ideal data and after simulating the continuum fitting step as applied on real data. The ability to recover the correct line centers is also studied.
Methods. The performance of these architectures is evaluated by the completeness, purity, and F1 score reached in bins of signal-to-noise ($S/N_{\rm line}$) for the absorption lines and with the absolute fractional flux error for the continuum.
Results. The U-Net model reaches scores of ≈90% for all metrics (completeness, purity, and F1 score) at $S/N_{\rm line} \approx 4$. All false positive detections with $S/N_{\rm line} \geq 5$ fall in the tails of the broad Lyα absorbers distribution of damped Lyα systems. The combination of continuum fitting and line detections has negligible effects on the detection performance at $S/N_{\rm line} \geq 4$.
Conclusions. Our proposed U-Net architecture offers a competitive tool for the analysis of absorption lines in current and upcoming large spectroscopic surveys and is well suited to the identification of any absorption line feature.

Key words. methods: data analysis – intergalactic medium – quasars: absorption lines – quasars: general – large-scale structure of Universe

1. Introduction

Quasar spectra enable the study of the intergalactic medium (IGM) through the footprints left in the form of absorption lines by hydrogen and metals. For example, the neutral hydrogen distribution within the IGM can be observed at high redshift ($z \gtrsim 1.5$) as a dense forest of absorber systems at wavelengths lower than 1216 Å in the quasar rest-frame, the so-called Lyman-α forest (Lyα, Lynds 1971; Bahcall & Goldsmith 1971; McDonald et al. 2006; Shull et al. 2012). Quasar spectroscopy further enables the study of the composition and kinematics of the diffuse, often metal-enriched, circumgalactic medium (CGM), through the measurement of (partially) optically thick hydrogen lines (e.g., Prochaska & Wolfe 2002; Prochaska et al. 2005; Rao et al. 2011; Lehner et al. 2013; Morrison et al. 2024) or strong ions such as C IV, Mg II, or Si IV (e.g., Hasan et al. 2020; Cooksey et al. 2013; Zhu & Ménard 2013; D'Odorico et al. 2016, 2022; Gontcho A Gontcho et al. 2018; Dutta et al. 2020; Galbiati et al. 2023, 2024; Yu et al. 2025).

The growing volume of data from large spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009), the VIMOS Public Extragalactic Redshift Survey (VIPERS; Scodeggio et al. 2018), the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2022), and the European Space Agency's Euclid mission (Laureijs et al. 2011), along with upcoming projects such as the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012, 2014, 2016; Jin et al. 2024) and the 4-meter Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019), has made the task of detecting the absorption lines challenging for traditional detection techniques, such as kernel filtering (e.g., Anand et al. 2021; Zou et al. 2021; Zhu & Ménard 2013), and spectral template-fitting methods (e.g., Brodzeller et al. 2025). Even though these approaches can provide accurate detections,

⋆ Corresponding author: elena.mangola@lam.fr

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

their application to large spectroscopic catalogs can be limited due to the computational cost of applying filters to large datasets. Therefore, it is essential to turn to fully automated algorithms that can deliver accurate and precise detection in extensive catalogs while using reduced computational resources and time.

Machine learning (ML) techniques have become ideal tools for fast and accurate detection of absorption lines in quasar spectra, as demonstrated by an increasing number of applications that can be found in the literature. Examples of the ML techniques used include Gaussian processes for the identification of damped Lyα systems (DLAs; Garnett et al. 2017) and C IV absorption lines detection (Monadi et al. 2023), random forest classifiers for searches of Lyman limit systems (LLSs; Fumagalli et al. 2020), and artificial neural networks (ANN or NN) for the identification of DLAs (Parks et al. 2018; Wang et al. 2022), Lyα forest lines (Cheng et al. 2022), and Mg II absorption lines (Zhao et al. 2019; Szakacs et al. 2023).

In this work, we build a new NN model aimed at detecting metal absorption lines in quasar spectra. For this task, we selected a U-Net architecture (Ronneberger et al. 2015) for its ability to track both local and global structures in the data. U-Net models have the ability to produce accurate semantic segmentations of images. They have been applied to several astrophysics problems such as denoising (Vojtekova et al. 2021), reconstructing spectra and density fields (Zhong et al. 2025; Aragon-Calvo 2019; Shi et al. 2025), and object detection (Nagam et al. 2024; Silver et al. 2025).

The model was trained and tested using mock spectra for the WEAVE-QSO survey. The same mock data were used in the companion paper by Pistis et al. (2025) to derive a model for the quasar continuum instead. WEAVE (Jin et al. 2024; Rogers et al. 2014) is a multiobject wide-field spectroscopic survey conducted at the 4.2 m William Herschel Telescope at the Roque de los Muchachos Observatory in La Palma, Spain (Dalton et al. 2016, 2020). The WEAVE spectrograph is fed by ∼1000 fibers positioned within a 2-degree field of view, enabled by a prime focus corrector, and it covers the 366−969 nm range at resolutions $R \sim 5000$ and 20 000. The WEAVE survey comprises eight individual surveys that address a variety of scientific goals, ranging from stellar to extragalactic astronomy. In particular, the WEAVE-QSO survey (WQ; Pieri et al. 2016) is of interest for this work. The aim of the WEAVE-QSO survey is to observe $\gtrsim 10^5$ high-redshift ($z \geq 2.2$) quasars over an area of $\approx$6000 deg^2 (termed the WQ-Wide area; Jin et al. 2024). In this work, we refer primarily to this survey, but our results can easily be generalized to other surveys.

The paper is organized as follows. In Sect. 2, we briefly describe the WEAVE mock data used in this work, in Sect. 3 we outline the data preprocessing and the construction of the NN model, while in Sect. 4, we detail the application of the model to the mock spectra. In Sect. 5, we discuss and compare our work with prior literature. Finally, we discuss and summarize our results in Sect. 6.

2. The WEAVE quasar mock data

Neural networks fall under the category of supervised learning, requiring labeled datasets of quasar spectra with known absorber information to train and accurately predict new and unseen data. In preparation for analyzing WEAVE data, this work relies on simulated quasar spectra. We anticipate, however, that the model can be deployed with no to minimal modifications on other datasets that share similar characteristics. Specifically, for building and profiling the performance of the U-Net, we used 50 000 mock WEAVE-like spectra as done in the companion paper by Pistis et al. (2025). The creation of a mock spectrum follows three steps: the generation of the sightline and the absorbers' profiles, the injection of a Lyα forest spectrum, and the addition of noise.

Table 1. Metal ion names and rest wavelengths for transitions inserted in the mock spectra.

Ion name	Restframe λ_{RF} (Å)
O I$^{@\dagger}$	1302
C II$^{@\dagger}$	1334
Al II$^{@\dagger}$	1670
Zn II$^{@}$	2026
Si IV$^{@\dagger\circ}$	1393, 1402
C IV$^{@\dagger\circ}$	1548, 1550
Ni II$^{@}$	1709, 1741
Mn II$^{@}$	2576, 2594
Mg II$^{@\dagger\circ}$	2796, 2803
S II$^{@}$	1250, 1253, 1259
Fe II$^{@\dagger\circ}$	1608, 2382, 2600
Si II$^{@\dagger}$	1260, 1304, 1526, 1808

Notes. Neutral hydrogen absorption systems class: $^{@}$DLA (Damped Ly-α Absorbers $\log(N_{\rm H\,I}/{\rm cm}^{-2}) \geq 20.3$), †sub-DLA ($19 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) < 20.3$), $^{\circ}$LLS (Lyman Limit Systems $17 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) < 19$).

In the first step, a quasar redshift between $2.1 < z_{\rm QSO} < 4.8$ and an observed r-magnitude between $16.3 < r < 21.8$ is assigned to each spectrum. Then, sightlines are populated with strong neutral hydrogen absorbers ($0.2 < z_{\rm abs} < 4.8$ and $17 < \log(N_{\rm H\,I}/{\rm cm}^{-2}) < 22.5$) and the associated metal absorption lines, depending on the class of the system, as detailed in Table 1. Physical parameters, distributions, and frequencies of the absorption systems are retrieved from observed data (Pâris et al. 2018; Shu et al. 2019; Lyke et al. 2020).

Metal absorption line profiles are created using a profile maker based on nonnegative matrix factorization (NMF; Lee & Seung 2000). This NMF profile maker (NMFPM; Longobardi et al. 2023) successfully reproduces realistic line profiles based on physical information, such as the ion producing the line, its rest-frame center, and the value of the column density. The NMFPM can generate a set of noise-free metal profiles at the desired resolution. To avoid sampling effects, the velocity resolution in the NMFPM profiles (1 km s^{-1}) is significantly finer than the effective WEAVE velocity sampling at the redshift of the targets, which corresponds to an observed-frame pixel scale value of 18 km s^{-1}. For each NMFPM set of lines (i.e., a multiplet), equivalent widths (EWs) and respective errors are determined, and the cumulative distribution functions of the optical depth are computed. Each multiplet is then inserted into the respective spectrum. The position of each ion is encoded in a binary label array, with positive values assigned to pixels that encompass the interval 5%−95% of the optical depth profile of each absorption line.

In the second step, we add low column density hydrogen lines. Low column density systems ($\log(N_{\rm H\,I}/{\rm cm}^{-2}) < 17$) are statistically represented by the Lyα forest. Lyα forest mocks are created using Lyα Cosmological Lofty Realization (LyaCoLoRe; Farr et al. 2023), a software that uses CoLoRe simulations (Ramírez-Pérez et al. 2022) to create realistic skewers, 1D line-of-sight samples through a 3D cosmological volume

that reproduce the physics of transmitted flux caused by the forest that shapes quasar absorption spectra. The forest is added while ensuring independence of the absorbers along different sightlines. As most of these absorbers have associated absorption lines that are weak and buried in the noise of WEAVE-like spectra, for simplicity, we did not track their associated metal content.

In the third step, flux-dependent noise is included to simulate spectra similar to those obtained in WEAVE observations at the observational condition with an air mass of $a = 1.107$ and an apparent magnitude of the sky $b_{\rm sky} = 20.92\,{\rm mag/arcsec^2}$. The simulated spectra have been updated to include atmospheric emission skylines, in contrast to their previous use in Pistis et al. (2025). The spectra are resampled and convolved with an appropriate line-spread function model to match the low-resolution mode ($R \sim 5000$) of the WEAVE survey. Finally, a continuum shape for the quasar is added. The continuum profiles are reconstructed using principal component analysis from Pâris et al. (2011). The noise is computed by predicting photon counts from the target and sky, following standard S/N calculations based on the instrument specifications. Using the noise-free model continuum, each spectrum is then normalized to unity over the whole wavelength range.

3. Absorption line identification with the U-Net architecture

In this work, we aim to develop a new U-Net architecture for detecting metal absorbers in quasar spectra. In this section, we describe the data preprocessing (Sect. 3.1) and the development of the architecture used for the detection (see Sect. 3.2). We also briefly describe the autoencoder developed for the continuum fitting by Pistis et al. (2025). This code was be used to assess the stability of the U-Net classifier's performance once uncertainties stemming from the continuum normalization step of quasar spectra are taken into account (see Sect. 4.3).

3.1. Data preprocessing

The preprocessing of the data is an important aspect when dealing with NNs (Goodfellow et al. 2016, Section 12.2.1). This step can have a significant impact on the performance of the algorithms used. For the detection of absorbers, we have considered all generated spectra, regardless of their signal-to-noise ratio $\mathrm{S/N_{spectrum}}$, computed as the median of the per-pixel signal-to-noise ratios over all the original wavelength ranges. The decision to train on all spectra, as opposed to focusing only on those with higher $\mathrm{S/N_{spectrum}}$, is discussed in Appendix A.

The spectra and flux uncertainties are shifted and linearly interpolated onto a common rest frame wavelength array encompassing the wavelength range $[1230, 3095]$ Å, with spacing of ~0.114 Å to maintain the smallest resolution element out of all the observed frames. The upper and lower limits are dictated by the need to include the lower redshift spectra flux and to exclude the Lyα forest in addition to continuum-modeling effects, respectively. We exclude the Lyα forest region due to the blending with hydrogen lines that hinder the detection of metal lines and also due to the presence of a current limitation in the mocks concerning the modeling of the Lyα emission line. In particular, we want to avoid the Lyα broad line emission because we observed that the normalized continuum does not remain at the expected level of unity in this region. Using a common quasar rest frame ensures fixed wavelength ranges where metal absorption lines appear.

The binary labels are also interpolated over the common wavelength array to align with the positions of the absorption lines. During preprocessing of data, wavelengths outside the spectral range (NaN values) are replaced by a flux value of 1 and their corresponding labels are set to 0. These regions are not explicitly masked in the loss function; so they are treated as continuum during training. For a given line, the $\mathrm{S/N_{line}}$ is defined as its EW over the error on the EW itself:

$$\mathrm{S/N_{line}} \equiv \mathrm{EW}/\Delta\mathrm{EW}. \tag{1}$$

The EW and its associated error are computed following the derivation in Vollmann & Eversberg (2006), assuming that the continuum does not have an error and discretizing over the pixel values to obtain

$$\mathrm{EW} = \Delta\lambda \sum_{i=\mathrm{px}_i}^{\mathrm{px}_f} (1 - f_i) \quad \Delta\mathrm{EW} = \Delta\lambda \sqrt{\sum_{i=\mathrm{px}_i}^{\mathrm{px}_f} \sigma_{f,i}^2}, \tag{2}$$

where f is the normalized flux, σ_f the error associated with the flux, $\Delta\lambda$ is the binning spacing of the wavelength array, and $\mathrm{px}_{i,f}$ are the initial and final spectral bins of the lines. Out of all the injected absorption lines in the spectra, we consider relevant only those with $\mathrm{S/N_{line}} \geq 1.5$; this threshold ensures only potentially detectable absorption lines are taken into consideration, given the typical quality of a mock spectrum. The absorption lines with $\mathrm{S/N_{line}} < 1.5$ are considered undetected from the spectral noise and, as such, their corresponding classification label is set at 0. Here and throughout, we adopt the signal-to-noise integrated over the line ($\mathrm{S/N_{line}}$) when considering detections, as it provides a better metric than the spectrum signal-to-noise ($\mathrm{S/N_{spectrum}}$) to determine which absorption lines are detectable because it encompasses both the notion of the error of the flux and the strength of the line itself. An additional step is required to handle the continuum when using spectra that are not already normalized to unity. To test the effect of residuals relating from imperfect continuum subtraction, we performed the steps of continuum fitting and then the step of line detection. An additional step before the continuum fitting was to scale all the spectra to the same range to avoid the specific features of some spectra dominating over others. Each quasar spectrum was normalized by the median flux at $\lambda = 1450$ Å within a window of $\Delta\lambda = 50$ Å.

3.2. The U-Net architecture: A model for absorber detection

The U-Net model, first described by Ronneberger et al. (2015), has a convolutional NN architecture characterized by an encoder, a symmetrical decoder path, and a bottleneck linking the two. The encoder progressively decreases spatial resolution, forcing the U-Net to learn a compressed representation of the image. In contrast, the decoder path increases the spatial resolution, reconstructing the image from its representation. The bottleneck, as the name suggests, is the bottom narrowest part of the network, connecting the encoder and the decoder. An architecture that consists of an encoder and a decoder are known as autoencoders (Hinton & Salakhutdinov 2006); however, they often suffer from poor location accuracy. The U-Net, instead, employs skip connections, concatenating the higher resolution features from the encoder to the upsampled features of the decoder, propagating information to higher resolution layers. Skip connections help to preserve fine details of the spectrum while incorporating high-level features captured in deeper layers, preserving the spatial

Table 2. Hyperparameters of the optimized U-Net for line identification.

Hyperparameter	Value
Base size	4
Increment 1	1
Increment 2	0
Increment 3	1
Dropout rate	0
Weight	2
Kernel size	5

localization of features of the image. In a classification task, such as that of detecting metal absorption lines (Zhao et al. 2019; Xia et al. 2022; Szakacs et al. 2023), the choice of using a U-Net architecture is supported by the encoder-decoder structure capable of high location accuracy without abandoning the use of the surrounding context.

Despite the rapid emergence of transformer and attention-based architectures capable of spectral feature extraction (e.g., Koblischke & Bovy 2024; Fortino et al. 2026), the use of a U-Net model is preferred for precise segmentation. The U-Net architecture provides a strong spatially located inductive bias that enables the precise localization of narrow absorption features such as those imprinted by metal ions.

As the model handles spectra, the input is a 1D image with a length equal to the number of pixels in the common rest-frame defined above. For each pixel, the model predicts whether it belongs to a metal absorption line, operating in a binary classification. The input spectrum is processed in the first block of the encoder path, which consists of a double convolution immediately followed by a max pooling operation with a fixed pool size and stride of 2. In the convolutions, the feature maps are padded to maintain their spatial dimensions across layers. The operations of the first block are repeated for three layers. Between each block, a dropout layer is inserted to prevent overfitting. At the bottom of the model, there is the bottleneck path that links the encoder and decoder. The bottleneck layer consists of a double convolution on the most condensed (low-resolution) spatial representation of the feature maps.

The decoder is symmetrical with respect to the encoder, with transposed convolution layers in place of max pooling layers. At each block, it doubles the spatial size and maintains the same number of filters as the corresponding encoder level, mirroring the trend of the encoder path. After every step of the encoder path, skip connections concatenate feature maps at the same resolution level at corresponding layers between the encoder and decoder. To restore the original input dimension, the final operation is a convolution with a linear activation function, where the number of filters is reduced to match the number of classes the model is trying to predict. In the case of a binary classification, this means reducing the output to 2 filters, each with a kernel size of 1.

The specific choices in building the architecture come from a hyperparameter optimization done using the Bayesian optimization implemented within Keras Tuner 1.4.7 (O'Malley et al. 2019) on 50 trials, aiming for the lowest validation loss value. The search space encompasses the number of filters present in each layer, expressed as a power of 2, starting from the power of the base size and incrementing cumulatively at each encoder level. Since the decoder path is symmetrical with the encoder path, it has a mirrored number of filters. Additional hyperparameters include the kernel size and dropout layer probability, which are fixed for all blocks. A final hyperparameter is the class weight assigned to positive labels in the loss function. Since the metal line features are underrepresented compared to continuum values, we used weights to penalize the loss in the case of a misdetection of a metal line. The loss function used was a weighted sparse categorical cross-entropy function:

$$L = -w_c \log(\hat{y}_c), \tag{3}$$

where $\hat{y}_c$ is the probability of the predicted label and w_c the class weight for the true class. The values of the hyperparameters of the best model are summarized in Table 2. A few interesting observations can be made on the results of the hyperparameter optimization. The dropout rate probability of 0 indicates that all nodes are considered essential, and therefore, regularization via node dropout does not play a significant role. The second increment is 0, suggesting that it is not necessary to increase the number of filters to better learn representation at a higher level of abstraction in the deepest layers. Ultimately, the weight assigned in the loss function is 2, which is the lowest admissible value, since a weight of 1 would correspond to equal importance being assigned to both classes. This means that the model can learn the absorption features without requiring heavy penalization in the loss function. The final U-Net architecture constructed with these hyperparameters is described in Fig. 1.

The 50 000 mock spectra, created as explained in Sect. 2, are divided in a 70%–15%–15% proportion for train, validation, and test sets. To improve train efficiency the ∼35 000 spectra of the training dataset were processed with a batch size of 32 and pre-fetched, meaning that the model processed the data taking 32 spectra at a time and prepared next batches of data in advance. For each set of hyperparameters, the model was trained for 1000 epochs. This large number allowed the model to train indefinitely until an early stopping based on the validation loss, using a patience of 5 epochs and automatic restoration of the best weights. The ∼7500 spectra of the validation dataset are used to evaluate the performance during training. Lastly, the ∼7500 test spectra are used to assess the model performance based on the metrics described in Sect. 3.4.

3.3. Autoencoder for continuum fitting

Mock data have the artificial, yet desirable, property that the continuum level of the quasars is known with no uncertainty. This valuable feature of the mock library enables development the U-Net and assess its intrinsic performance without confusion arising from artifacts of inexact continuum estimation. However, we are also interested in assessing the purity and completeness of absorption line recovery in real-life applications, where the true continuum is unknown, and we thus repeat the study of the performance of the network starting on mocks with a realistic quasar continuum, to which we apply a normalization step.

The continuum normalization step was performed on the WEAVE mocks using an autoencoder. The choice of using an autoencoder is supported by the results of Pistis et al. (2025), in which an autoencoder outperformed a CNN and a U-Net in predicting quasar continua. We optimized that autoencoder architecture for our problem. In fact, unlike previous works (Liu & Bordoloi 2021; Turner et al. 2024; Pistis et al. 2025), we do not limit the spectra in the restframe wavelength range $1020 < \lambda(\text{Å}) < 2000$, but we modify the wavelength range considered to that of the mock spectra ([1230, 3095] Å). This, combined with the chosen smaller spectral resolution with respect to

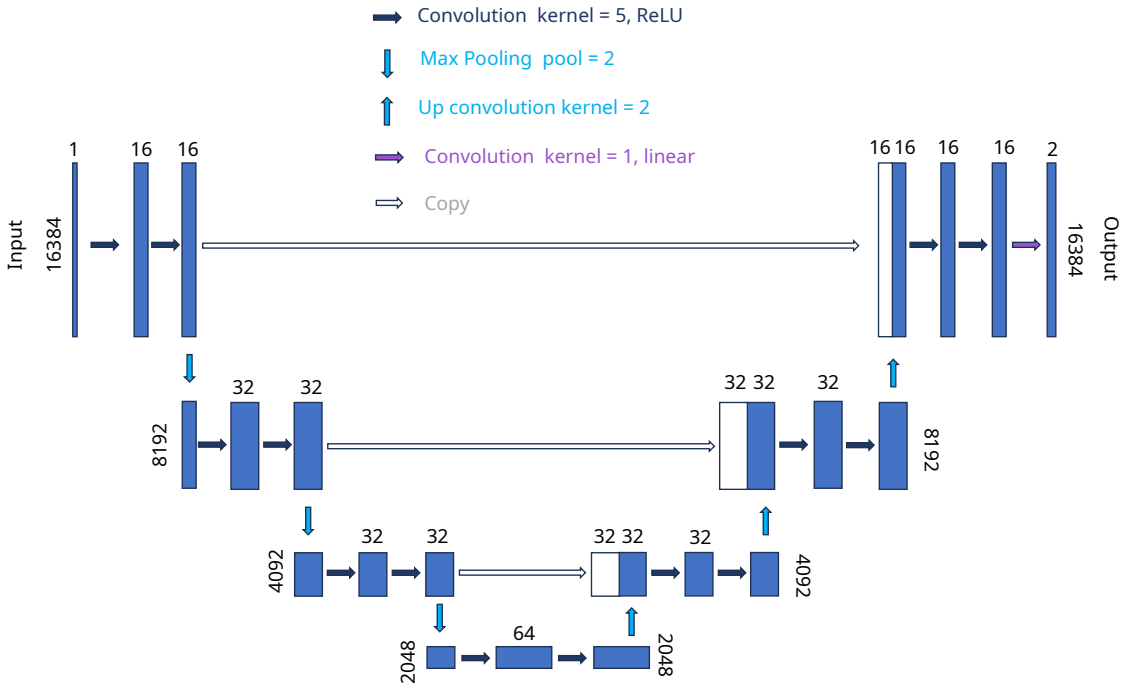


Fig. 1. Scheme of the architecture of the optimized U-Net model applied to WEAVE mock data for metal lines detection. The blocks are the different layers in the U-Net, with the white blocks being the ones copied by the skip connections. The different colored arrows represent operations between layers. The activation function used is, in all cases, the ReLU function. The number above the blocks denotes the number of feature maps in the layer, while the number on the side is the size of each layer.

Table 3. Hyperparameters of the optimized autoencoder for continuum normalization.

Hyperparameter	Value
Number of layers	2
Size encoder 1	384
Activation encoder 1	ELU
Size encoder 2	384
Activation encoder 1	ELU
Size bottleneck	320
Activation bottleneck	ReLU
Size decoder 1	256
Activation decoder 1	ReLU
Size decoder 2	640
Activation decoder 2	ReLU
Learning rate	10^{-4}

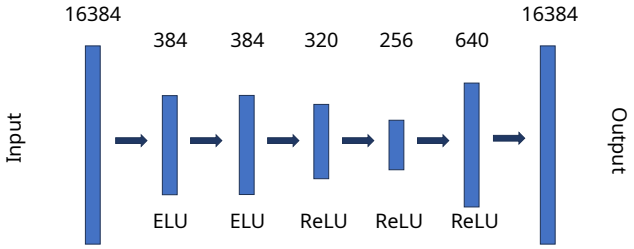


Fig. 2. Scheme of the architecture of the optimized autoencoder model applied to WEAVE mock data for the continuum estimation. The blocks are the different layers in the autoencoder. The number above the blocks denotes the size of the layer. The activation function used between each layer is written below the blocks. All the layers represented are dense layers.

the previous work (Pistis et al. 2025), meant we needed to train a new autoencoder.

The autoencoder has a classical architecture, which includes a part that compresses the input spectra into a smaller representation, the encoder, and a part that reconstructs the original input from that, the decoder. A masking layer allows the autoencoder to ignore NaN values of the spectra. During the optimization step, the size of each layer was not constrained to the size of the previous layer. This choice led to an unconventional autoencoder where the encoder and the decoder have asymmetrical sizes.

Each layer could vary its dimension between 128 and 1024, with a step of 128. Additionally, the activation function is left free to change between layers, allowing it to be either a rectified linear unit function (ReLU; Agarap 2018) or an exponential linear unit function (ELU; Clevert et al. 2015). Between the encoder and the decoder, there is a bottleneck layer. A last hyperparameter is the learning rate of the model, chosen between 10^{-2}, 10^{-3}, 10^{-4}. We perform a hyperparameter optimization using a random search tuner on 50 trials, aiming for the lowest validation loss value, with a loss function based on the mean squared error. In Table 3, the optimized values obtained from the hyperparameter search are summarized. The specifics of the model's architecture are illustrated in Fig. 2.

3.4. Metrics for assessing the network performance

We used different metrics to quantify the performance of our NNs according to their respective goals. For the detection of absorbers, we used the completeness, defined as the fraction of absorption lines retrieved among all the ones with $S/N_{line} \geq 1.5$, and purity, defined as the fraction of the real lines among all the retrieved ones. Completeness (C) and purity (P) can be expressed as

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

and

$$\text{Purity} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

where TP represents the true positive detections, FN the false negative ones, and FP the false positive ones.

The U-Net model described in Sect. 3.2 outputs an array of labels with the exact spatial dimensions as the input spectrum, meaning that the predictions are made pixel by pixel. Given that the ultimate goal is to detect absorption lines, the evaluation of the model through metrics computation should focus on the detection of entire lines, rather than individual pixels. For this reason, a true positive detection is recorded when at least one predicted positive pixel coincides with the corresponding positions of the ground truth labels of an absorption line. Otherwise, the detection is classified as a false negative. The U-Net model might predict a common label for close absorptions or blended lines. Since the metric calculations are based on the true and predicted labels and we calculate a single centroid (see Sect. 4.2 for the error on the position of the line centroids) for each predicted absorption, in both cases we count TPs for all the multiple lines and blended lines. Similarly, a false positive is defined as a predicted detection that does not correspond to a system in the ground truth labels. Additionally, to provide a more balanced evaluation of the U-Net's performance, accounting for both completeness and purity, we used the F1 score metric:

$$\text{F1 score} = 2 \cdot \frac{\text{Completeness} \cdot \text{Purity}}{\text{Completeness} + \text{Purity}}. \tag{6}$$

True metal absorption lines not completely within the boundaries of the spectra are not considered as lines in the metric computation.

To assess the quality of the continuum estimation, we used the absolute fractional flux error (AFFE; e.g., Liu & Bordoloi 2021; Turner et al. 2024; Pistis et al. 2025) defined as:

$$|\delta F| = \int_{\lambda_1}^{\lambda_2} \left| \frac{F_{\text{pred}}(\lambda) - F_{\text{true}}(\lambda)}{F_{\text{true}}(\lambda)} \right| d\lambda \Big/ \int_{\lambda_1}^{\lambda_2} d\lambda \,, \tag{7}$$

where F_{pred} is the predicted output and F_{true} is the true continuum of the simulated quasar spectra.

4. Application to the WEAVE mock data

From the 50 000 mock spectra, the training and validation spectra are used to optimize the U-Net model as described in Sect. 3.2 and to optimize the autoencoder for continuum fitting as described in Sect. 3.3.

The optimized models are then applied to the test subset to evaluate their performance. Firstly, we evaluated the U-Net performance on the spectra with the true underlying continuum (see Sects. 4.1 and 4.2). Then, to simulate the detection process starting from real observed data, the U-Net model is tested on the same test spectra, but with the continuum produced by the autoencoder (see Sect. 4.3). The training and testing of the U-Net and autoencoder models were performed on three NVIDIA A2 GPUs, each with 15 GB of memory, and were implemented using Python 3.12.0 and the open-source machine learning libraries Keras 3.9.2 (Chollet 2018) and TensorFlow 2.19.0 (Abadi et al. 2015). Application of the autoencoder and of the U-Net on the ∼7500 test spectra took ∼2.7 s and ∼21 s on 1 NVIDIA A2 GPU, respectively.

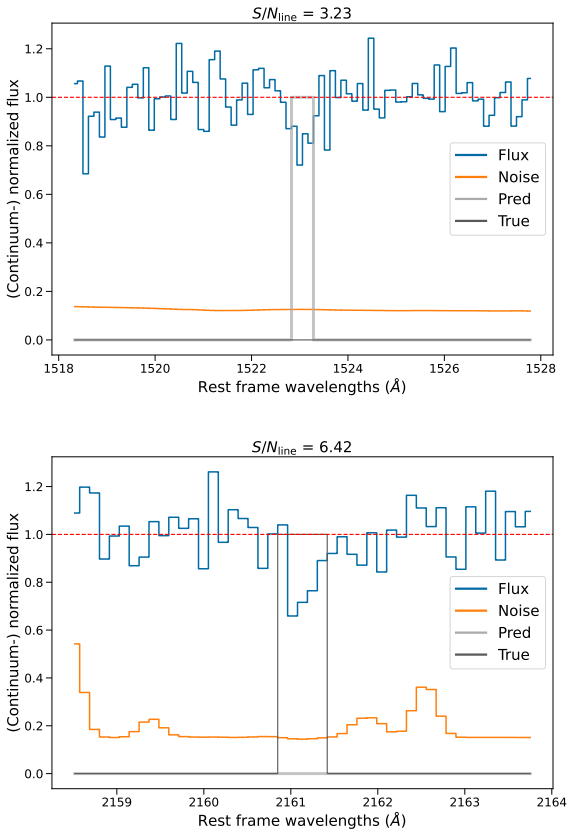


Fig. 3. Examples of false positive detection with $S/N_{line} = 3.23$ (top) and a false negative detection of a line with $S/N_{line} = 6.42$ (bottom). The spectrum (blue) has been normalized to the continuum, with the dashed red line marking unity. The orange line represents noise, light gray is for the predicted labels, and dark gray shows the true labels.

4.1. Detection of metal lines

We applied the optimized U-Net model to the test spectra, predicting pixel-wise labels for binary detection of the absorbers. The metrics described in Sect. 3.4 are employed to determine the effectiveness of the model. Focusing on only the predictions within the wavelength range of the spectra, we then computed the number of TP, FP, and FN. For each detection, real or false, we study the distribution of the metrics as a function of the signal-to-noise ratio (S/N_{line}) of the line. For a TP or FN detection, the S/N_{line} is the one given in the spectra creation. For a FP detection, instead, the S/N_{line} is computed on the predicted positive pixels group, always using Equation (2). Therefore, the metrics used in this work are simulation-based diagnostics and the actual values for an observational catalog might vary when using a uniform measurement of the S/N_{line}. In Fig. 3 there are displayed examples of a false positive detection

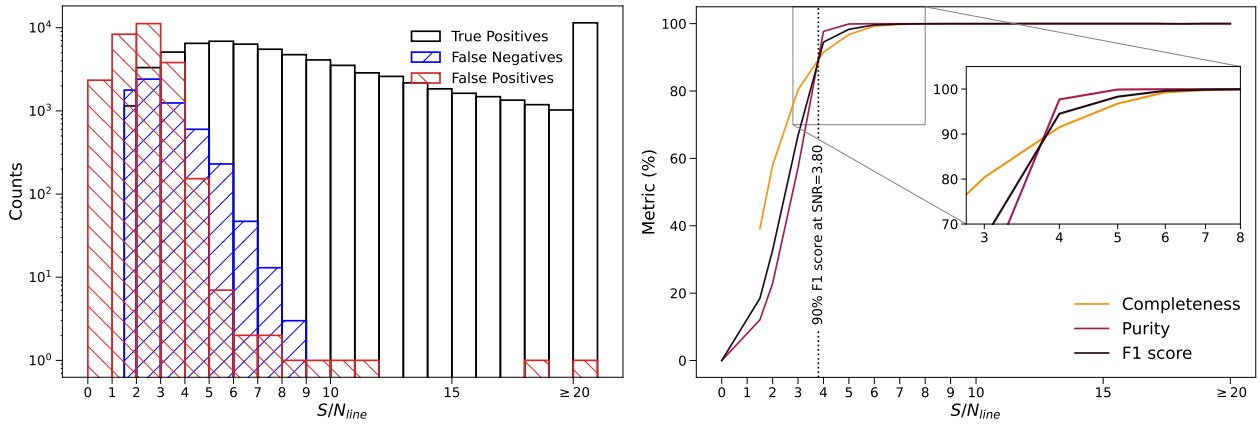


Fig. 4. Performance of the U-Net for the line detection as a function of the S/N_{line}. Left panel: Histograms for true positives (black), false positives (red), and false negatives (blue) in S/N_{line} bins. The last bin is cumulative for $S/N_{line} \geq 20$. Right panel: Curves of different metrics (completeness in yellow, purity in dark red, and F1 score in black) as a function of the S/N_{line}. The dotted line highlights the S/N_{line} level at which the F1 score reaches 90%.

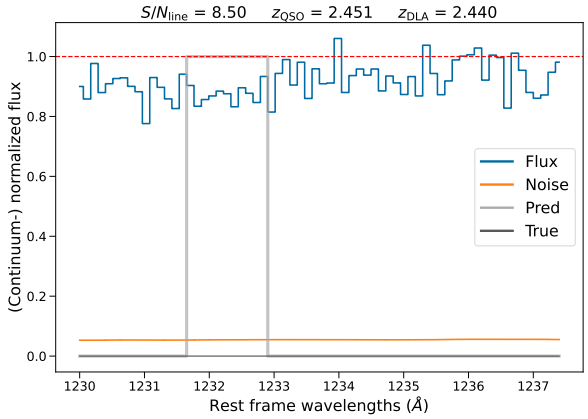


Fig. 5. Example of a false positive detection caused by the presence of a DLA absorption feature at $z_{dla} = 2.440$ for a quasar at $z_{qso} = 2.451$. The presence of the DLA is apparent since the average continuum level is below unity (red dashed line). The conventions of the lines are the same as Fig. 3.

and a false negative missed by the model. Figure 4 shows the performance of the U-Net. In the left panel, a histogram shows that FNs and FPs are found typically at $S/N_{line} \leq 5$, except for a tail that extends at higher signal-to-noise. Based on a visual inspection of these 17 more extreme cases of FPs detections with $S/N_{line} > 5$, all of these are located near DLA absorption features in the spectrum. When DLA absorption wings are present, they are often the cause of false positive detections, also for lower S/N_{line} lines, due to a drop in the continuum that leads to misdetections. For all FPs with $S/N_{line} > 5$, the redshifts of the DLAs present in the spectra are close to the redshift of the quasar, with an average of $\Delta z = z_{qso} - z_{dla} \simeq 0.015$. An example of this effect is shown in Fig. 5, where the false positive detection is due to the presence of a DLA wing, recognizable thanks to the continuum level that is consistently below unity. Appropriate identification and either masking or subtraction of DLAs (using, e.g., Prochaska & Herbert-Fort 2004; Noterdaeme et al. 2009; Garnett et al. 2017) is thus recommended as a further step. The observed decrease in TP at higher S/N_{line} is, instead, due to the fact that there are fewer high S/N_{line} lines available to detect. The right panel of Fig. 4 shows both completeness and purity reaching a 90% level around $S/N_{line} \approx 4$, with the F1 score reaching 90% at $S/N_{line} = 3.8$. The model is able to perform with high completeness and purity not only in the high S/N_{line} regime ($S/N_{line} \geq 8$, where all the metrics reach 100%) but also at lower values. At $S/N_{line} \sim 4$, the model is still performing at satisfactory levels (metrics $\sim 90\%$). These high values of completeness, purity, and F1 score indicate that the model successfully identifies the most relevant samples and that most of the identified samples are accurate. For lines with $S/N_{line} \lesssim 4$ the performance drops. The model is nevertheless able to detect a large number of lines; however, the purity declines and it is hard to make use of this low S/N_{line} sample. Figure 6 shows examples of true positive detections of lines with different S/N_{line}.

4.2. Reliability of the detection: Position of the absorbers

As an additional test of the reliability of the detection, we measured the center of the detected lines. The line center is defined as the weighted mean wavelength over the detection range:

$$\lambda_{\rm det} = \frac{\sum_{i={\rm px}_i}^{{\rm px}_f} \lambda_i \, (1 - f_i)}{\sum_{i={\rm px}_i}^{{\rm px}_f} (1 - f_i)}, \tag{8}$$

where f is the normalized flux, λ is the associated wavelength, and $px_{i,f}$ are the initial and final bins of the detected lines. Figure 7 shows the Bland–Altman plot showing the agreement between the true and predicted centers. The average of the difference between true and predicted centers is $\langle \Delta\lambda_C \rangle \approx -0.028$ Å, which corresponds to 25% of a pixel size, with a standard deviation $\sigma_{\lambda_C} \sim 0.205$ Å. The mean difference and its variations are minimal, indicating a good match between the true center of the line and the computed weighted center on the predicted line position. 73.5% of the predicted centers are recovered within 1 pixel, and 49.7% within 1/2 of a pixel. The kernel density estimator (KDE) contour lines highlight the most dense regions near a zero mean difference. Being mostly symmetrical, the lines further demonstrate the reliability of the model's detections. It is worth noting that some of the metal absorption lines are blended. This can cause some of the dispersion seen, but overall it is a minor

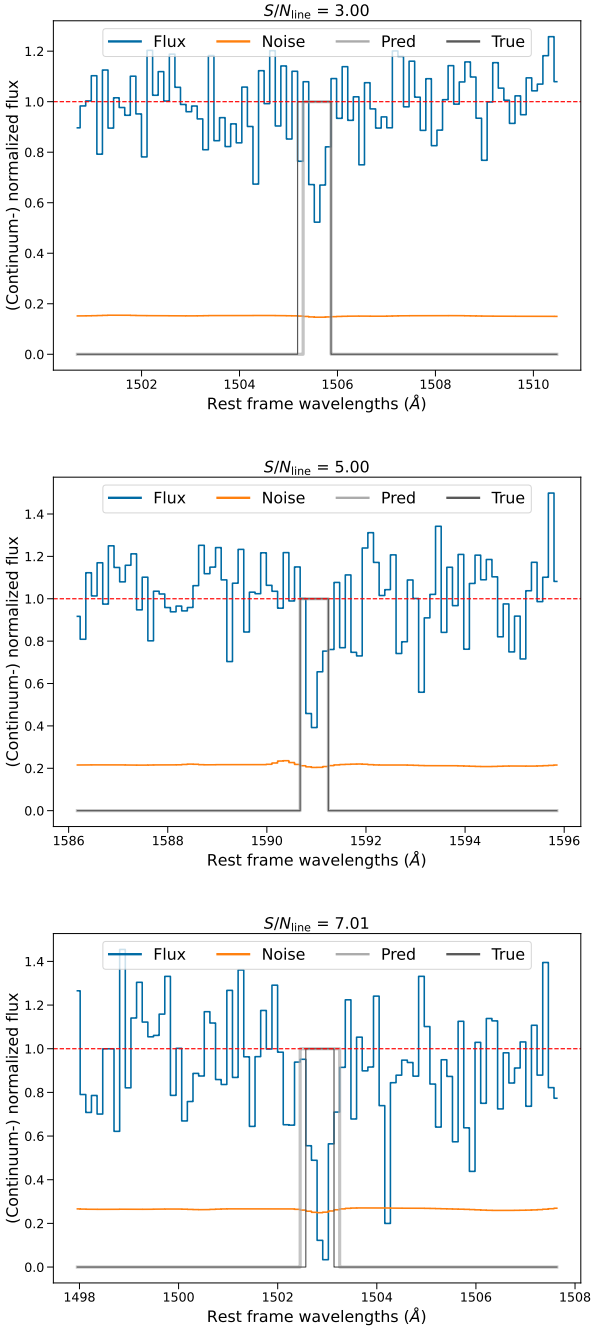


Fig. 6. Examples of true positive detections of lines with $S/N_{line} = 3$ (top), $S/N_{line} = 5$ (center), and $S/N_{line} = 7.01$ (bottom). The conventions of the lines are the same as Fig. 3.

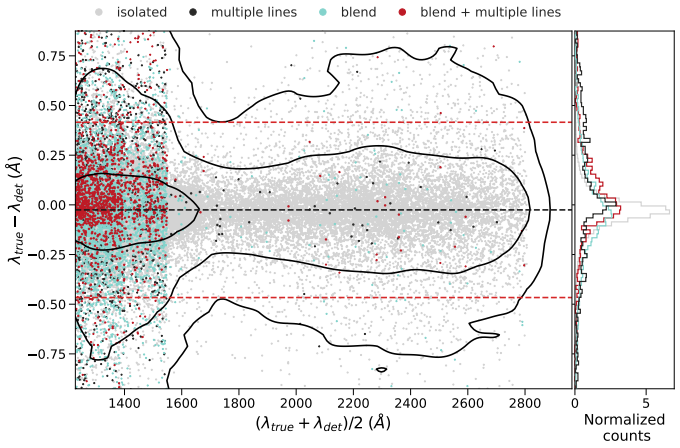


Fig. 7. Bland–Altman plot for the true and predicted centers of the absorption lines. The mean difference is marked by a dashed black line, and the 95% confidence levels with red dashed lines. KDE contour levels at 1σ, 2σ, and 3σ levels (increasing inward) are represented with solid black lines. Blended lines are shown as light blue circles, multiple true line predictions are the black circles, and red circles mark cases where both occur. Other isolated detections of lines are denoted in small light grey circles. On the right, a normalized histogram of the count of occurrences of the different cases.

effect, since we compute the rate of blends over the entirety of lines to be 1.8%. Even though blended lines are considered as separate systems in the metric computation, the predicted center is often associated with two different lines. Our analysis method for the position of the absorbers has the current limit of not being able to distinguish between lines in case of blended features. In Fig. 7, we can also notice the presence of multiple lines detections. These are the cases where, in the range of wavelengths over which the model predicted a line, there is more than one true absorption line. In this case where more true absorption lines are present in one single detection range, the lines are not necessarily blended. Likewise, the prediction range of a line that is blended often does not include the second blended line; therefore not falling into the category of multiple lines. However, the two cases can occur simultaneously, when in one single detection the line detected has a blend, and the blend is present under the predicted range. Both blends and multiple lines findings are more common in the bluest side of the spectra, contributing to the dispersion mostly in that region.

4.3. Detectability after continuum fitting

Up to now, we have focused on the performance of the network on idealized data for which the continuum level was perfectly known. As real applications also require a continuum-fitting step, we are interested next in understanding the U-Net performance on data that are continuum-normalized, starting from a complete set of mocks where the quasar continuum has been injected and then determined using the autoencoder presented by Pistis et al. (2025).

The autoencoder we use in this work is a slightly modified and retrained version of the one in Pistis et al. (2025). In Fig. 8 we reassess the performance of the autoencoder for the continuum fitting. The model reproduces the true continuum with a median error of $\langle \mathrm{AFFE} \rangle = 0.011$, computed over all the rest-frame wavelength range, that is in agreement with the value found in Pistis et al. (2025). Having established that the retrained autoencoder yields comparable performances to the original one, we finally tested the full pipeline (continuum fitting plus line detection) needed for observational data. To enable comparison with the results presented in earlier sections, we defined the difference in a given metric (purity, completeness, and F1 score) as

$$\Delta M = \left| \frac{M_{\mathrm{d}} - M_{\mathrm{c+d}}}{M_{\mathrm{d}}} \right| \tag{9}$$

where M corresponds to the chosen metric while the suffixes d and $c + d$ correspond to detection and continuum plus detection pipelines, respectively. The horizontal axis, for the central and bottom panels, represents the line S/N

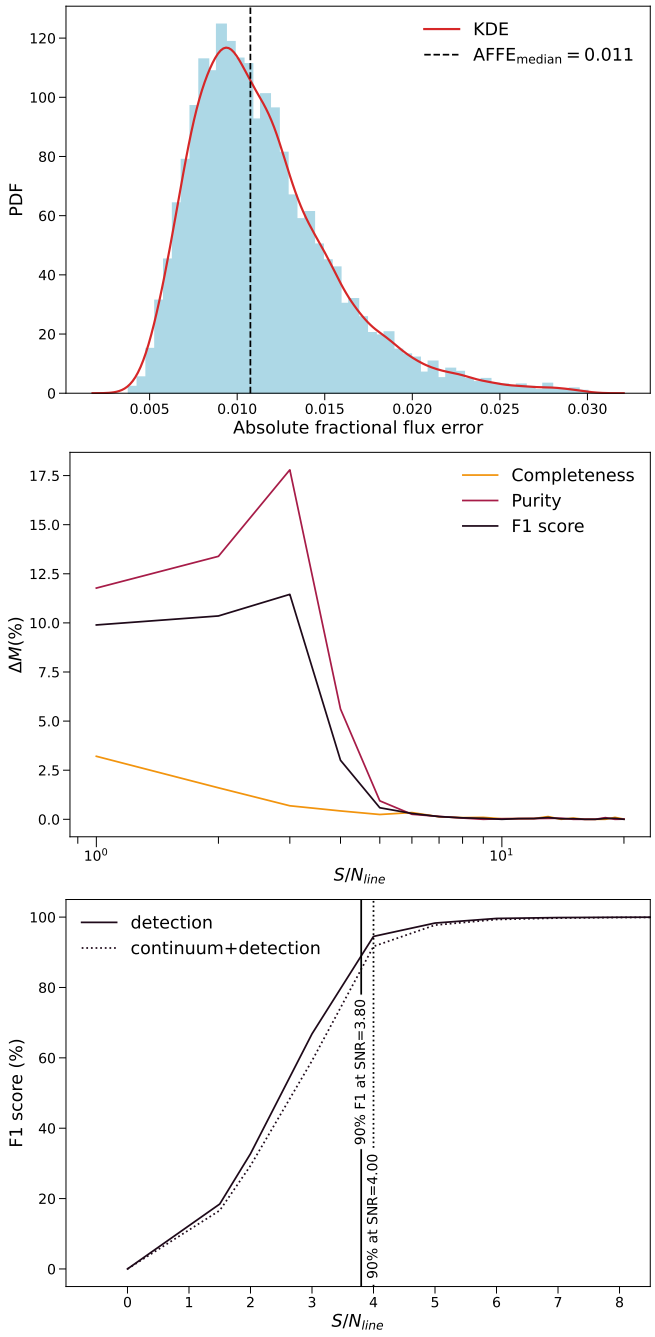


Fig. 8. Performance of the sequential use of the models for the continuum fitting and line detection. Top panel: Histogram of the absolute fractional flux error of the autoencoder used for the continuum fitting. Middle panel: Fractional difference of the various metrics (completeness, purity, F1 score) as a function of the signal-to-noise of the line (S/N_{line}). Bottom panel: F1 score curves for the detection-only and continuum plus detection pipelines as a function of the line signal-to-noise ratio (S/N_{line}). Values for $S/N_{line} > 8$ are omitted, as they are all approximately 100%.

under the ideal-continuum assumption, rather than a measurement significance incorporating continuum-fitting uncertainty. The small median AFFE is encouraging, but does not by itself establish that local continuum errors are negligible for all weak lines, exactly where the performances are deteriorated.

The middle panel of Fig. 8 shows the change of performance ΔM when the step of the continuum fitting is added in the analysis. We note that the U-Net has not been retrained at this step, so the training is still performed on idealized data with no errors on the continuum level. The bottom panel of Fig. 8 shows, for lower S/N_{line}, the different F1 metrics for the detection only step, thus using the idealized continuum from the mocks, and for the continuum plus detection pipeline, using the estimated continuum from the autoencoder. This comparison shows that using the estimation of the continuum from the autoencoder yields, as expected, slightly worse results than using mock spectra with a perfect normalization. However, the difference in performance is limited to low S/N_{line} values, with $\Delta M < 10\%$ for the F1 score when $S/N_{line} > 3$. From $S/N_{line} \approx 6$, the delta metric tends to 0; at $S/N_{line} \approx 4$, which we have established above as a reasonable threshold where the U-Net performs reliably, $\Delta M \lesssim 5\%$ for the F1 score. The relative difference for lines of $S/N \geq 4$ is reasonably small for purity ($\lesssim$10%) and negligible for completeness. This test shows that using the continuum estimation from the autoencoder does not introduce significant deterioration in the model's ability to recover absorption lines. Given this, we can expect to apply the existing trained models of continuum fitting and detection of metal absorption lines to future WEAVE-QSO data consecutively.

5. Discussion and comparison with other networks

We presented a U-Net model able to successfully identify metal absorption lines in mock quasar spectra. Our work adds a new architecture, the U-Net, to the growing body of literature that is employing machine learning for metal lines detection. For instance, CNNs have been used for similar tasks (e.g., Zhao et al. 2019; Xia et al. 2022; Szakacs et al. 2023; Liu et al. 2025). Some of these works have explored the use of deep neural networks to find narrow absorption lines in quasar spectra, but have focused on a single metal species. Zhao et al. (2019) and Szakacs et al. (2023) focused on Mg II metal absorbers, while Xia et al. (2022) and Liu et al. (2025) targeted Ca II absorbers. A common trait in these works is the use of CNN as the architecture of choice, due to its widely established ability to detect features in images. A general conclusion from these applications is that using a first filter size comparable to the typical width of the absorption feature can improve the network performance. This choice of the filter width implies that the networks should be tailored to the spectral resolution of the dataset when the line spread function dominates the profile (in moderate and lower resolution data). A question, however, arises on the performance of these models on much higher resolution data, when absorption lines will be dominated by a variety of line widths that are intrinsic and no longer driven by the resolution of the spectrograph.

In our case, since we explore multiple metal ion species with varying EW, the filter size was allowed to vary (see Sect. 3.2). This does not pose a problem for the success of our results; in fact, the U-Net uses skip-connections to help preserve information about the context around the lines to be able to compensate for the loss of spatial information typical of CNNs.

The previously cited works obtained high performances, with an accuracy of $a \sim 94\%$ for Zhao et al. (2019), an accuracy and F1 score of ~95.9% for Xia et al. (2022), an accuracy of $a \sim 98.6\%$ for Szakacs et al. (2023), and $a \sim 99.77\%$ for Liu et al. (2025). The accuracy is the ratio between the correct classifications over the total classifications. These values are compatible with our results taking into consideration only lines for which the $S/N_{line} \geq 4$, obtaining total metrics of F1 = $C = P \sim 99\%$.

Table 4. Comparison of the pixel size, S/N selection, metal ion targets, training set size, and computed metrics (F1 score: F1, Completeness: C, Purity: P, Accuracy: a) in this work against those of the works discussed in Sect. 5.

Work	Pixel size	S/N selection	Metal ion targets	Training set size	Metric
This work	~0.114 Å	$S/N_{line} \geq 1.5$	(refer to Table 1)	35 000	F1 ~ 82% $C = 92\%$ $P = 74\%$
		$S/N_{line} \geq 4$			F1 = $C = P \sim 99\%$
Zhao et al. (2019)	~0.4 Å	$EW_{2796/2803} \geq 0.3$ Å	Mg II	34 379[†]	$a \sim 94\%$
Szakacs et al. (2023)	$R = 20\,000$[*]	$S/N_{spectrum} \geq 3$	Mg II	108 000	$a \sim 98.6\%$
Xia et al. (2022)	1 Å	$S/N_{\lambda 3934} \geq 2.5$ $S/N_{\lambda 3969} \geq 2$	Ca II	680 000	$a = F1 \sim 95.9\%$
Liu et al. (2025)	1 Å	$S/N_{\lambda 3934} \geq 2.5$ $S/N_{\lambda 3969} \geq 2$	Ca II	216 000	$a = F1 = C = P$ ~99.77%

Notes. [(*)]Szakacs et al. (2023) relies on high-resolution mock spectra with the indicated resolution R. [(†)]Zhao et al. (2019) use five training sets of the same size with the same spectra containing Mg II, but different control spectra. The indicated number is the total sum of the five training sets.

Overall, we obtained a F1 score of F1 = 82%, completeness of $C = 92\%$, and purity of $P = 74\%$, slightly lower results compared to the previous works. Our total metrics are derived from predictions over quasars spectra regardless of their $S/N_{spectrum}$, performing a selection only over the metal absorption lines signal-to-noise ratio ($S/N_{line} \geq 1.5$, Sect. 3.1). They are, therefore, dominated by the majority of low S/N_{line} lines.

Nevertheless, previous works were also subject to selection. Szakacs et al. (2023) worked with spectra having a $S/N_{spectrum}$ taken from a discrete uniform distribution with a minimum $S/N_{spectrum}$ equal to 3. Furthermore, Zhao et al. (2019) filtered out Mg II absorption lines with $EW_{2796/2803} < 0.3$ Å, on the other hand Xia et al. (2022) and Liu et al. (2025) instead inserted only Ca II lines with $S/N_{\lambda 3934} \geq 2.5$ and $S/N_{\lambda 3969} \geq 2$ in their artificial spectra dataset.

Given the substantial differences in spectral resolution, targeted ion species, training-set selection, and task generality, the comparison in Table 4 should be read as contextual rather than as a controlled ranking of methods. It should be noted that the metrics used in Szakacs et al. (2023), Liu et al. (2025) measure spectrum-level binary classification, i.e., whether an absorber is present in a given spectrum; whereas our completeness, purity, and F1 are computed at the level of individual detected lines across the full spectral range and multiple ion species. The fraction of correctly labeled pixels or events over all examples is conceptually distinct from the line-level completeness, purity, and F1 score we report. In particular, accuracy is sensitive to class imbalance (most pixels are non-absorbers) and can appear artificially high even when rare absorption features are poorly recovered. Completeness and purity are instead computed at the level of individual detected lines and directly quantify the network's ability to find real absorbers and suppress false positives, respectively.

Concerning the wavelength accuracy, we obtain a mean absolute error MAE $\simeq$ 0.11 Å for a pixel size of ~0.114 Å, measuring the line center as defined in Equation (8), whereas Szakacs et al. (2023) obtained an overall accuracy of MAE = 6.9 Å with a resolution of the spectra of $R = \frac{\lambda}{\Delta\lambda} = 20000$.

A comparison of performances also needs to take into account the degree of specialization of the various networks. Szakacs et al. (2023) considered only synthetic spectra containing only one Mg II doublet within each spectrum, excluding the presence of both other Mg II absorbers in the spectra and other metals' absorption lines. Applicability to more general cases is likely to require new training or changes to the architecture. The models by Xia et al. (2022) and Liu et al. (2025) require instead knowledge of the positions of the Mg II absorption lines to be able to detect Ca II lines. Lastly, Zhao et al. (2019) searched for Mg II features only in the spectral region between the C IV and Mg II emission lines.

As illustrated in Sect. 3.2, we preferred the U-Net model in comparison to attention based architectures. It is important to consider the fact that Liu et al. (2025)'s final choice of architecture involves the introduction of an attention mechanism at the end of each residual block's final convolution unit. They argue that this choice improves their detection abilities compared to previous work also on new unseen data, obtaining a precision of 92.6% compared to that of 20.3% of Xia et al. (2022). Given this result, the introduction of an attention mechanism in a future updated version of this work could be beneficial.

Compared to these previous works, our U-Net aims to generalize the problem to the detection of any metal absorption lines (see Table 1 for the list of ions included in our mocks) and considers the full spectral range redward of the Lyα emission line. This is a more general and hence challenging task; thus, the U-Net performance appears satisfactory even if purity and completeness start to drop below 90% for $S/N_{line} \lesssim 4$. Unlike other works, however, our architecture is specifically designed to identify absorption features without recognizing and classifying the specific ions. The U-Net model we have developed is intended to form part of a pipeline for the identification of metal ions absorption lines in quasar spectra. The pipeline comprises three distinct steps, which are delineated in Pistis et al. (2025), this paper, and Pistis et al. (2026), respectively. Pistis et al. (2025), as mentioned throughout the manuscript, describes the continuum fitting step procedure for the fitting of the quasar continuum with an autoencoder architecture. This procedure constitutes a significant preprocessing step, which has also been applied in this study (see Sect. 4.3). The work presented here serves as the next preprocessing step: using the continuum given as the output of the autoencoder, the spectra are normalized and used as the input of the U-Net model, which in return flags the pixels corresponding to metal absorption lines. Lastly, the original flux, the autoencoder continuum, and the pixel-level detection map are entered into a classification algorithm, as delineated in Pistis et al. (2026), which performs an automatic classification of metal absorption lines.

Pistis et al. (2026) also provide a validation of the pipeline using real observational data, specifically from the first public data release of DESI (DESI DR1 DESI Collaboration 2026). The characteristics of DESI survey differ from the WEAVE-like mocks considered here in terms of resolution ($R_{\rm DESI}$ = 2000−5500), noise properties, and sample size. While our model is evaluated under idealized conditions, the application to DESI data incorporates observational effects such as bad pixels and sky residuals. Another aspect to consider is the presence of broad absorption line (BAL) quasars, both in future WEAVE-QSO and DR1 spectra. According to Ennesser et al. (2022), ∼12−16% of quasar spectra in large survey present BAL features, which can be a potential source of contamination. Pistis et al. (2026) tackles this problem by using a BAL finding tool to select BAL-free spectra before applying the full pipeline. The validation presented in the work of Pistis et al. (2026) demonstrates that the pipeline maintains good performance when applied to real spectra, having an average 1.2σ tension with the literature catalogs for C IV detections (Cooksey et al. 2013; Hasan et al. 2020), while also highlighting the additional uncertainties associated with real data. This population-level consistency is encouraging. However, the WEAVE-like mocks lack metal lines associated with systems with $14 \leq \log(N_{\rm H\,I}/{\rm cm}^{-2}) \leq 16$, which can produce a non-detection bias for such systems.

6. Conclusions

This study focuses on building an NN model for the detection of absorption lines in quasar spectra arising from metals in the IGM or CGM along the line of sight. We select the U-Net as our architecture for its ability to classify small-scale features in the context of larger-scale properties of the dataset. This capability appears well-suited to the task of identifying narrow absorption lines within the quasar spectrum at greater wavelengths than the Lyα emission line, where metal lines are present.

We trained the model on realistic mock spectra that mimic the data products of the WEAVE surveys and include a wide variety of ions, which have real-looking absorption profiles and are distributed in wavelengths according to observed redshift distributions. It is possible in future to introduce further realism to the mocks, expanding the metal absorption lines systems injected into the spectra. After training on idealized spectra that do not include the quasar continuum (i.e., corresponding to a perfect continuum level), the model reaches high levels of completeness and purity, above 90%, at $S/N_{\rm line} \approx 4$. At high signal-to-noise, the U-Net is fully complete and suffers only from the detection of occasional false positives that lie in the wings of strong absorbers, such as DLAs (see Fig. 5).

The model is also able to accurately center the absorption lines. The simple calculation of the centers from the weighted (using the absorption level as weight) mean of the wavelengths within a line find an average difference between true and predicted centers $\langle \Delta\lambda_{\rm C} \rangle = -0.028$ with a standard deviation $\sigma_{\lambda_{\rm C}} = 0.205$ (see Fig. 7).

We also explore, without retraining, the U-Net performance to applications in which the quasar continuum is measured on the data, and hence not fully idealized. This is achieved by applying the quasar continuum to our mocks, and renormalizing the spectra with the autoencoder by Pistis et al. (2025), which we specialize to work across the entire wavelength range redward of the quasar Lyα line.

Even with this dataset, the performance of the network remains satisfactory. Compared to the more idealized case of a perfect continuum, the U-Net shows no appreciable difference in completeness, a purity that is within ≈10% of the idealized case, and a F1 score within ≈5% for $S/N_{\rm line} \approx 4$. By $S/N_{\rm line} \approx 6$, the performance with and without continuum normalization becomes identical (see Fig. 8). Our models can thus detect lines with good precision of their centers while maintaining high performance despite uncertainty in the continuum fit.

This work expands on the set of tools available for the identification of absorption lines in large spectroscopic surveys, extending previous models in the capability to identify any absorption feature. Our U-Net is, however, not trained to distinguish and classify different ions, a task we will explore in a companion paper (Pistis et al. 2026).

Data availability

Supplementary materials with the trained models for all architectures are available on Zenodo for both quasar continuum fitting and metal absorption line detection at https://doi.org/10.5281/zenodo.22711575.

Acknowledgements. We thank the referee for insightful comments that have improved the content and presentation of this work. This work has been supported by the European Union – Next Generation EU, Mission 4, Component 1 CUP H53D23011030001. IPR was supported by funding from the grant PID2023-151122NA-I00 by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Agarap, A. F. 2018, ArXiv e-prints [arXiv:1803.08375]
Anand, A., Nelson, D., & Kauffmann, G. 2021, MNRAS, 504, 65
Aragon-Calvo, M. A. 2019, MNRAS, 484, 5771
Bahcall, J. N., & Goldsmith, S. 1971, ApJ, 170, 17
Brodzeller, A., Wolfson, M., Santos, D. M., et al. 2025, Phys. Rev. D, 112, 083510
Cheng, T.-Y., Cooke, R. J., & Rudie, G. 2022, MNRAS, 517, 755
Chollet, F. 2018, Astrophysics Source Code Library [record ascl:1806.022]
Clevert, D. A., Unterthiner, T., & Hochreiter, S. 2015, ArXiv e-prints [arXiv:1511.07289]
Cooksey, K. L., Kao, M. M., Simcoe, R. A., O'Meara, J. M., & Prochaska, J. X. 2013, ApJ, 763, 37
Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, SPIE Conf. Ser., 8446, 84460P
Dalton, G., Trager, S., Abrams, D. C., et al. 2014, SPIE Conf. Ser., 9147, 91470L
Dalton, G., Trager, S., Abrams, D. C., et al. 2016, SPIE Conf. Ser., 9908, 99081G
Dalton, G., Trager, S., Abrams, D. C., et al. 2020, SPIE Conf. Ser., 11447, 1144714
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3
DESI Collaboration (Abareshi, B., et al.) 2022, AJ, 164, 207
DESI Collaboration (Abdul Karim, M., et al.) 2026, AJ, 171, 285
D'Odorico, V., Cristiani, S., Pomante, E., et al. 2016, MNRAS, 463, 2690
D'Odorico, V., Finlator, K., Cristiani, S., et al. 2022, MNRAS, 512, 2389
Dutta, R., Fumagalli, M., Fossati, M., et al. 2020, MNRAS, 499, 5022
Ennesser, L., Martini, P., Font-Ribera, A., & Pérez-Ràfols, I. 2022, MNRAS, 511, 3514
Farr, J., Font-Ribera, A., du Mas des Bourboux, H., et al. 2023, Astrophysics Source Code Library [record ascl:2312.005]
Fortino, W. F., Bianco, F. B., Protopapas, P., Muthukrishna, D., & Brockmeier, A. J. 2026, ApJ, 1000, 14
Fumagalli, M., Fotopoulou, S., & Thomson, L. 2020, MNRAS, 498, 1951
Galbiati, M., Fumagalli, M., Fossati, M., et al. 2023, MNRAS, 524, 3474
Galbiati, M., Dutta, R., Fumagalli, M., Fossati, M., & Cantalupo, S. 2024, A&A, 690, A7
Garnett, R., Ho, S., Bird, S., & Schneider, J. 2017, MNRAS, 472, 1850
Gontcho A Gontcho, S., Miralda-Escudé, J., Font-Ribera, A., et al. 2018, MNRAS, 480, 610
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press), http://www.deeplearningbook.org
Hasan, F., Churchill, C. W., Stemock, B., et al. 2020, ApJ, 904, 44

Hinton, G. E., & Salakhutdinov, R. R. 2006, Science, 313, 504
Jin, S., Trager, S. C., Dalton, G. B., et al. 2024, MNRAS, 530, 2688
Koblischke, N., & Bovy, J. 2024, ArXiv e-prints [arXiv:2411.04750]
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Lee, D., & Seung, H. S. 2000, Advances in Neural Information Processing Systems, 13
Lehner, N., Howk, J. C., Tripp, T. M., et al. 2013, ApJ, 770, 138
Liu, B., & Bordoloi, R. 2021, MNRAS, 502, 3510
Liu, Y., Li, J., Gao, L., et al. 2025, ApJS, 276, 37
Longobardi, A., Fossati, M., Fumagalli, M., et al. 2023, RAS Tech. Instrum., 2, 470
Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8
Lynds, R. 1971, ApJ, 164, L73
McDonald, P., Seljak, U., Burles, S., et al. 2006, ApJS, 163, 80
Monadi, R., Ho, M.-F., Cooksey, K. L., & Bird, S. 2023, MNRAS, 526, 4557
Morrison, S., Som, D., Pieri, M. M., Pérez-Ràfols, I., & Blomqvist, M. 2024, MNRAS, 532, 32
Nagam, B. C., Koopmans, L. V. E., Valentijn, E. A., et al. 2024, MNRAS, 533, 1426
Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087
O'Malley, T., Bursztein, E., Long, J., et al. 2019, KerasTuner, https://github.com/keras-team/keras-tuner
Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, A&A, 530, A50
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Parks, D., Prochaska, J. X., Dong, S., & Cai, Z. 2018, MNRAS, 476, 1151
Pieri, M. M., Bonoli, S., Chaves-Montero, J., et al. 2016, in SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, eds. C. Reylé, J. Richard, L. Cambrésy, et al., 259
Pistis, F., Fumagalli, M., Fossati, M., et al. 2025, A&A, 698, A292
Pistis, F., Fumagalli, M., Fossati, M., et al. 2026, A&A, submitted
Prochaska, J. X., & Herbert-Fort, S. 2004, PASP, 116, 622
Prochaska, J. X., & Wolfe, A. M. 2002, ApJ, 566, 68
Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123
Ramírez-Pérez, C., Sanchez, J., Alonso, D., & Font-Ribera, A. 2022, JCAP, 2022, 002
Rao, S. M., Belfort-Mihalyi, M., Turnshek, D. A., et al. 2011, MNRAS, 416, 1215
Rogers, K., Stuik, R., Steele, I. A., et al. 2014, SPIE Conf. Ser., 9147, 91476H
Ronneberger, O., Fischer, P., & Brox, T. 2015, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (Cham: Springer International Publishing), 234
Scodeggio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84
Shi, F., Wang, Z., Yang, X., et al. 2025, ApJS, 280, 53
Shu, Y., Koposov, S. E., Evans, N. W., et al. 2019, MNRAS, 489, 4741
Shull, J. M., Smith, B. D., & Danforth, C. W. 2012, ApJ, 759, 23
Silver, E., Wang, R., Huang, X., et al. 2025, ApJ, 994, 117
Szakacs, R., Péroux, C., Nelson, D., et al. 2023, MNRAS, 526, 3744
Turner, W., Martini, P., Karaçaylı, N. G., et al. 2024, ApJ, 976, 143
Vojtekova, A., Lieu, M., Valtchanov, I., et al. 2021, MNRAS, 503, 3204
Vollmann, K., & Eversberg, T. 2006, Astron. Nachr., 327, 862
Wang, B., Zou, J., Cai, Z., et al. 2022, ApJS, 259, 28
Xia, I., Ge, J., Willis, K., & Zhao, Y. 2022, MNRAS, 517, 4902
Yu, X., Qu, Z., Cai, Z., et al. 2025, ApJ, 985, 64
Zhao, Y., Ge, J., Yuan, X., et al. 2019, MNRAS, 487, 801
Zhong, F., Napolitano, N. R., Heneka, C., et al. 2025, MNRAS, 543, 691
Zhu, G., & Ménard, B. 2013, ApJ, 770, 130
Zou, S., Jiang, L., Shen, Y., et al. 2021, ApJ, 906, 32

Appendix A: Training data and model architecture selection

To train the U-Net model discussed in Sect. 3.2, we utilized all generated mock spectra instead of only those with $S/N_{spectrum} > 3$, which would feature clearer absorption lines. We also tested a modified U-Net architecture to include noise and flux as inputs to enhance detection. Figure A.1 shows F1 score metrics for the different training data with both U-Net models. The model with noise performs better with high $S/N_{spectrum}$ spectra, whereas the noise-free model shows improved results with all available spectra. When using the noise model, the noise input channel introduces confusion when the $S/N_{spectrum}$ is low, resulting in a poorer performance compared to its use on higher $S/N_{spectrum}$ spectra. In the comparison, the model without noise is preferred overall, particularly when using all spectra. For $S/N_{spectrum} > 3$, the noise model is favorable only up to $S/N_{spectrum} = 3$; beyond that, the noise-free model outperforms it. Since we consider absorption lines with $S/N_{line} \geq 1.5$, the model without noise is considered the preferred model. We also previously consider a series of U-Net models, with architectures similar to the one we take into consideration, trained on subsets of data. Spectra were divided into discrete ranges of magnitude and, following the same idea of considering part of the continuum absorption lines lower than the S/N_{line} threshold, into bins of thresholds of S/N_{line} of the absorption lines. Training in different magnitude conditions and considering only some absorption lines makes it difficult to decide which model can generalize better to all different S/N_{line} lines. We also tested if the choice of shifting all the spectra into a common reference frame of the quasar rest wavelength range has an advantage over using the observed wavelength range. In the rest frame there are fixed ranges of wavelengths in which the absorption lines from a given metal ion appear, always at a shorter wavelength than the corresponding rest frame emission peak. In the observed frame, instead, these regions are at different wavelengths. We, indeed, observed that the common rest frame yields better results than using the different observed frames.

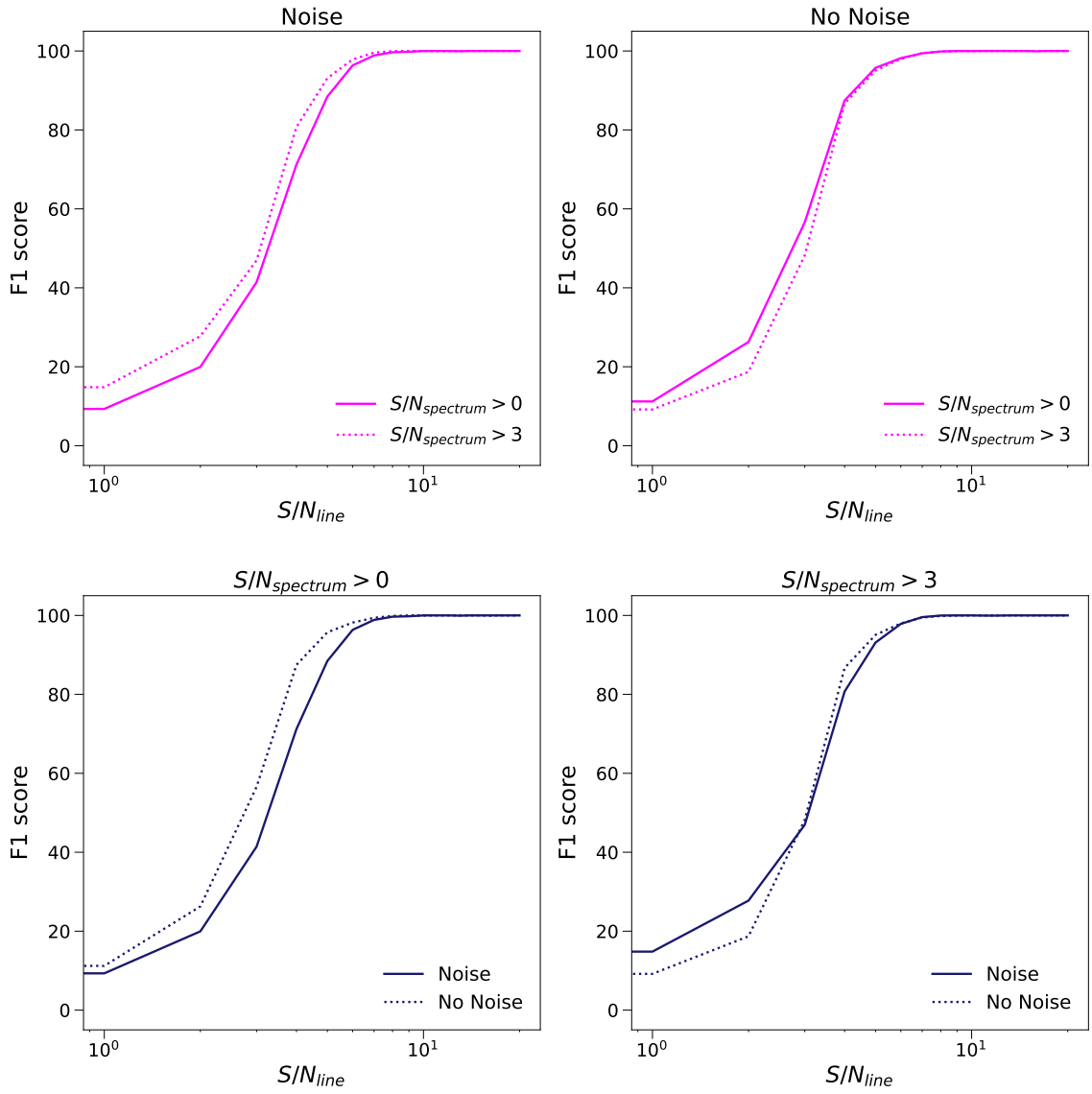


Fig. A.1. F1 score metric trends in function of the S/N_{line} of the metal absorption lines. Top plots: F1 score of the sample with $S/N_{spectrum} > 3$ spectra (dotted line) and with all spectra (solid line) using the model with (left panel) and without noise (right panel). Bottom plots: F1 score using the model without (dotted line) and with noise (solid line) on the data with all the spectra (left panel) and with only the high $S/N_{spectrum} > 3$ spectra (right panel).