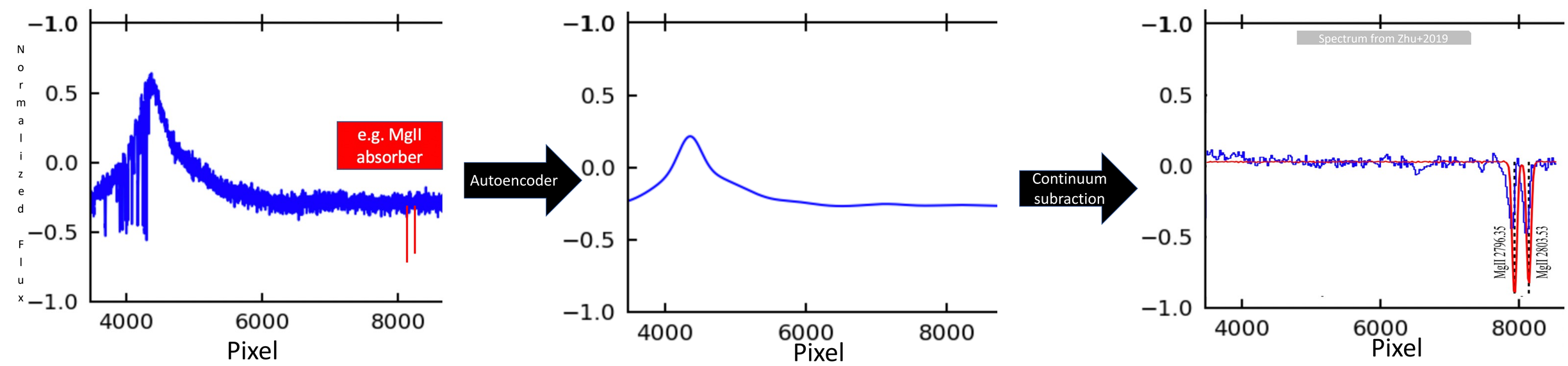
Detecting Metal Absorbers in Quasar Spectra: A Machine Learning Approach

Roland Szakacs, European Southern Observatory (Garching), roland.szakacs@eso.org

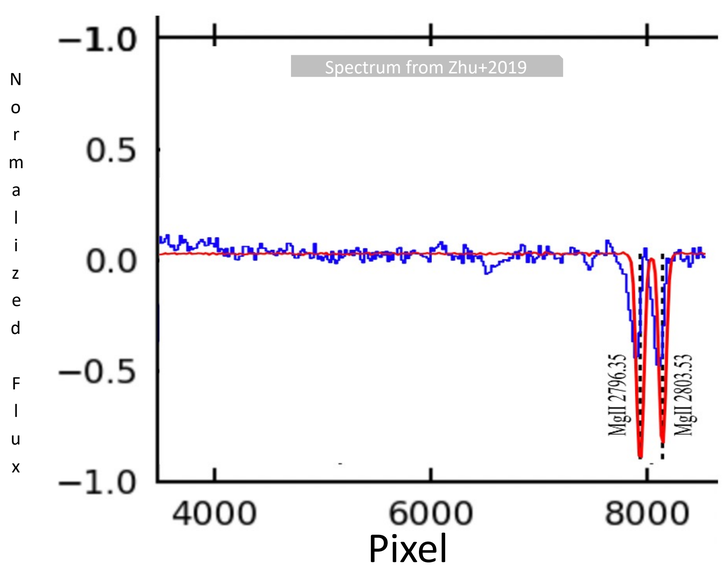
Introduction

Large scale surveys such as SDSS, DESI and upcoming 4MOST surveys provide an unprecedented amount of quasar spectra to explore. One of these surveys is our upcoming 4MOST community survey 4HI-Q (preliminary title). The survey will obtain ~ 1 million R=20 000 quasar spectra to search for HI and metal absorption line systems in quasar sightlines at low to high redshifts. Given the vast number of spectra, it is unfeasible to explore them individually to infer systems of interest. Therefore, in preparation for our survey, we are aiming to create a pipeline which efficiently fits the quasar continuum and subsequently identifies systems including metal absorption lines. Machine learning has become an important tool in astrophysics to approach such problems (e.g. Parks+2018, Wang+2022). Using TNG50 from the Illustris Project (e.g. Nelson+2019), we create realistic metal absorption lines and inject them into quasar mock spectra including Lyman-alpha forests. We use this dataset to train an autoencoder for quasar continuum fitting and a convolutional neural network (CNN) for the detection of metal absorption lines within quasar sightlines. This approach of deriving metal absorption lines from simulations, will also aid in exploring the physical parameters of the systems responsible for the absorption lines in the future.

Sketch of the Pipeline



Quasar continuum fit using an autoencoder architecture (see e.g. Bank+2021)



CNN

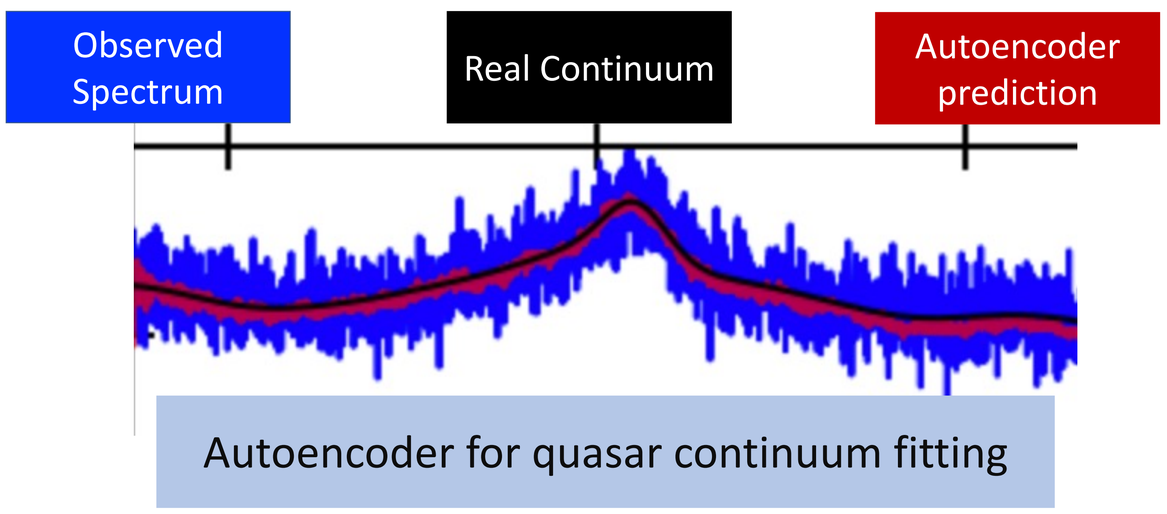
1

0

Or other possibilites:
- Classification of metal absorbers
- Redshift of metal absorbers
- Infering physical properties of absorbing gas or associated galaxies (e.g. based on TNG50)

Identification of systems with metal absorbers using a convolutional neural network architecture (see e.g. O'Shea+2015)

(Very Early) Preliminary Results



Autoencoder for quasar continuum fitting

First tests on non-normalized spectra and modified network of Zhao+2019:

Optimistic Sample (SNR = 100, $EW_{MgII,tot} > 0.02$):
98% accuracy

More Realistic Sample (SNR = 5, $EW_{MgII,tot} > 0.02$):
70-80% accuracy

CNN for metal absorption line detection

For more details (e.g. detailed architectures, our 4MOST survey) or just to chat / exchange ideas please don't hesitate to contact me at the workshop, through the workshop slack channel or through E-Mail (roland.szakacs@eso.org).

References:
Bank+2021, arXiv:2003.05991; Nelson+2019, MNRAS, 490, 3234; O'Shea+2015, arXiv:1511.08458; Parks+2018, MNRAS, 476, 1151; Wang+2022, ApJ, 259, 19; Zhao+2019, MNRAS, 487, 801;