Model-Based Search for Extended Emission-Line Regions in Astronomical Images

Charles Zhao[1] Advisor: Peter Melchior[2]

[1]Department of Computer Science, Princeton University, Class of 2021

[2]Department of Astrophysical Sciences, Princeton University

CENTER FOR STATISTICS AND MACHINE LEARNING

Motivation and Goal

Active galactic nuclei (AGN) feedback processes may explain several important phenomena in galaxy evolution, but are poorly understood. **Extended emission-line regions (EELRs)** can provide valuable information about AGN. Our goal is to develop an approach to imaging EELRs that:

- Is automatic and robust (to overlaps, noise, etc.)
- Does not require targeted observations
- Only relies on photometric observations

Background

The task of **deblending** is to model an image as the sum of contributions of multiple astronomical sources. **scarlet** [1] does this by generalizing non-negative matrix factorization to allow for general constraints. Given a B-band image with N pixels in each band, scarlet fits the following model:

$$\mathsf{M} = \sum_{k=1}^{K} \mathsf{A}_k \otimes \mathsf{S}_k = \mathsf{AS}$$

where $\mathsf{A}_k \in \mathbb{R}^B$ is the amplitude of component k across all bands and $\mathsf{S}_k \in \mathbb{R}^N$ is the morphology of that component.

Below is a schematic of an AGN, located at the center of a galaxy. An EELR is simply a particularly large narrow-line region.

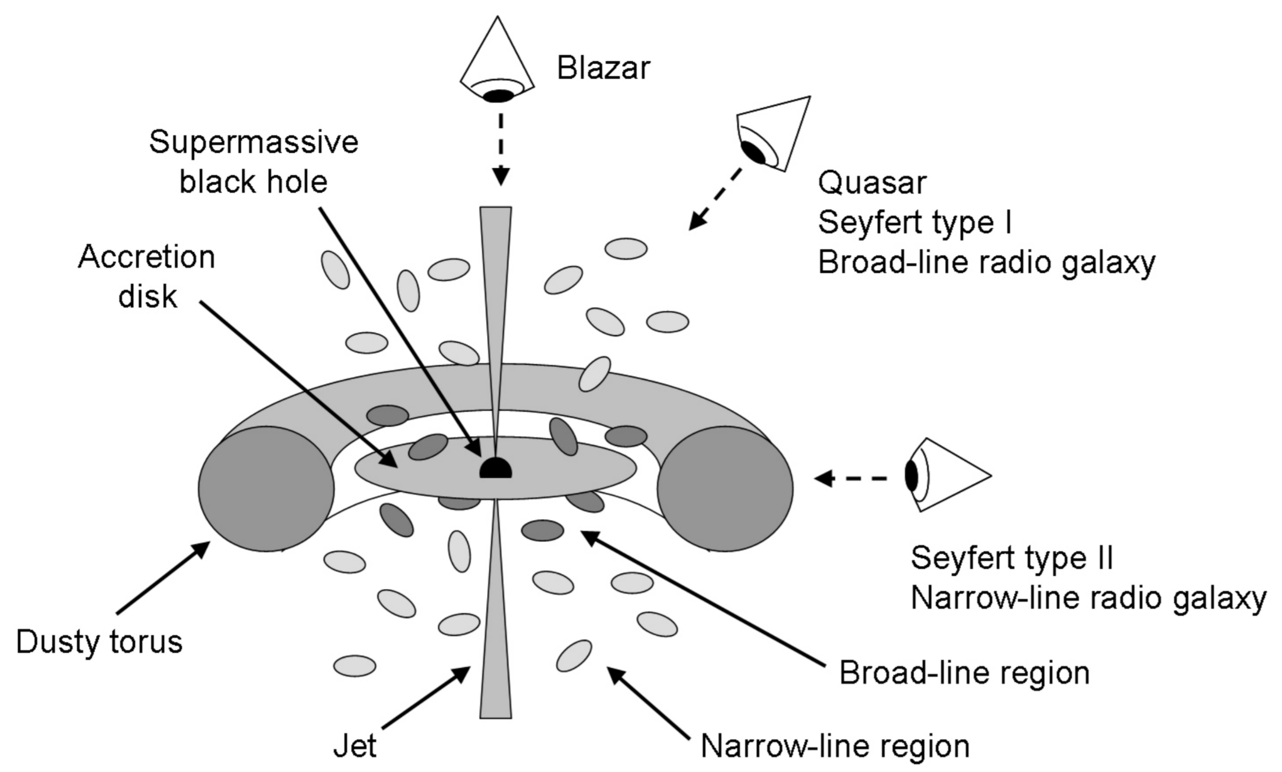

Figure 1. Unifed AGN model [3].

Related Work

- Sun et al. [2] developed a broad-band imaging technique, using the Subaru Hyper Suprime-Cam (HSC) Survey for broad-band images and the Sloan Digital Sky Survey (SDSS) for spectra
- They carefully subtract out the galactic stellar continuum from broad-band images, thus isolating the narrow-line emissions
- They image NLRs around 300 obscured AGN, finding 8 EELRs
- Their method does not require targeted observations beyond spectroscopic measurements

Data

- HSC for g, r, i, z, y broad-band images; SDSS for spectra and redshift
- Obscured type 2 AGN with redshift $z \sim 0.1 - 0.8$
- 444 observations

All code is available at `https://github.com/pmelchior/hsc_eelr`.

Approach

Since **spectroscopic measurements** are expensive to obtain and will have even less availability in future large-scale surveys, we use a sampling approach that essentially "guesses" the EELR spectrum from a **redshift measurement**.

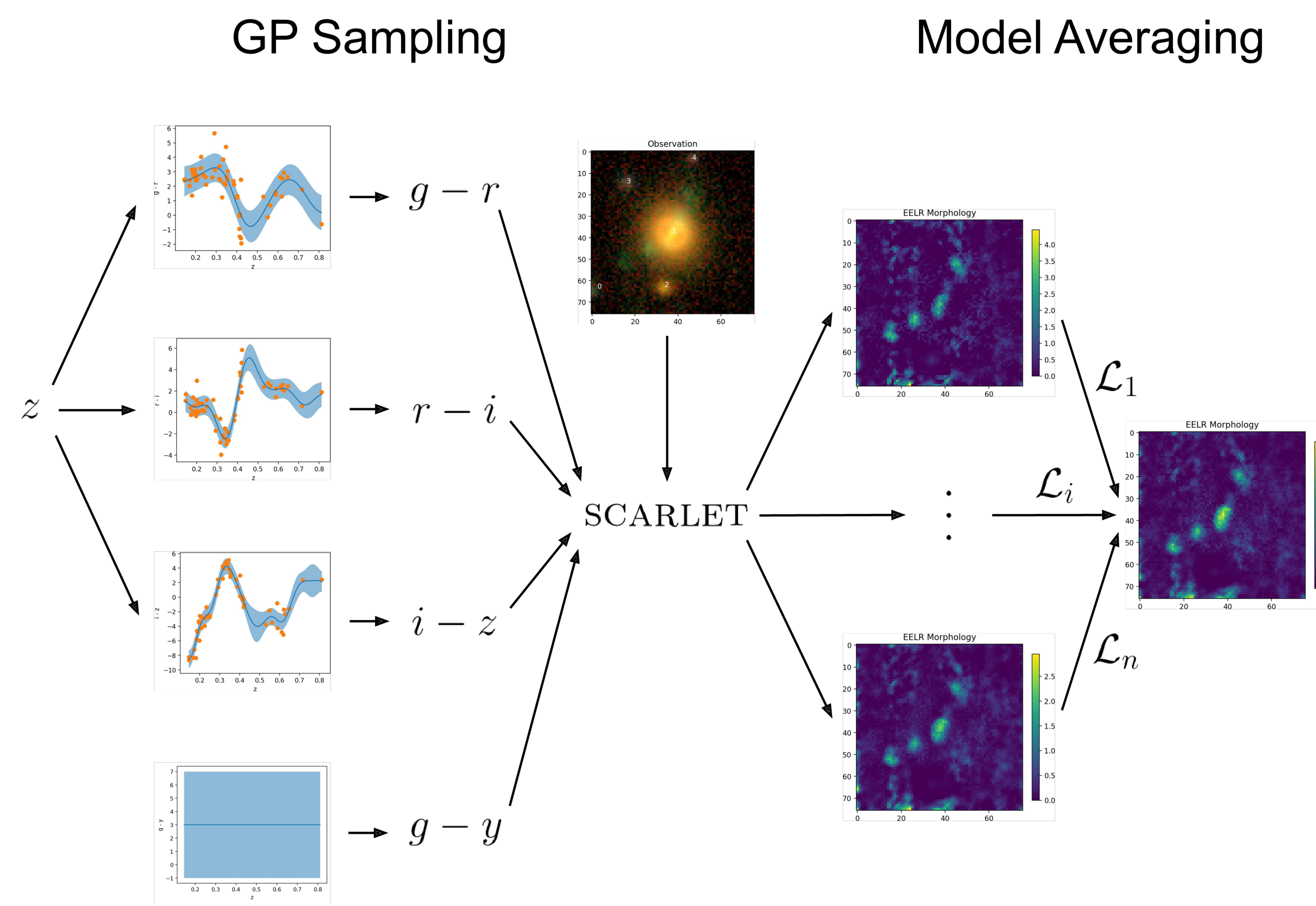

Figure 2. Summary of our inference procedure. Gaussian Process sampling is used to generate EELR spectra, which are used by scarlet as constraints when deblending, and the scarlet models are then combined using likelihood-weighted model averaging.

Implementation: Gaussian Process Regression

We use SDSS data to regress colors on redshift. Given a redshift, these GPs give us Gaussian distributions of colors from which to sample EELR spectra.

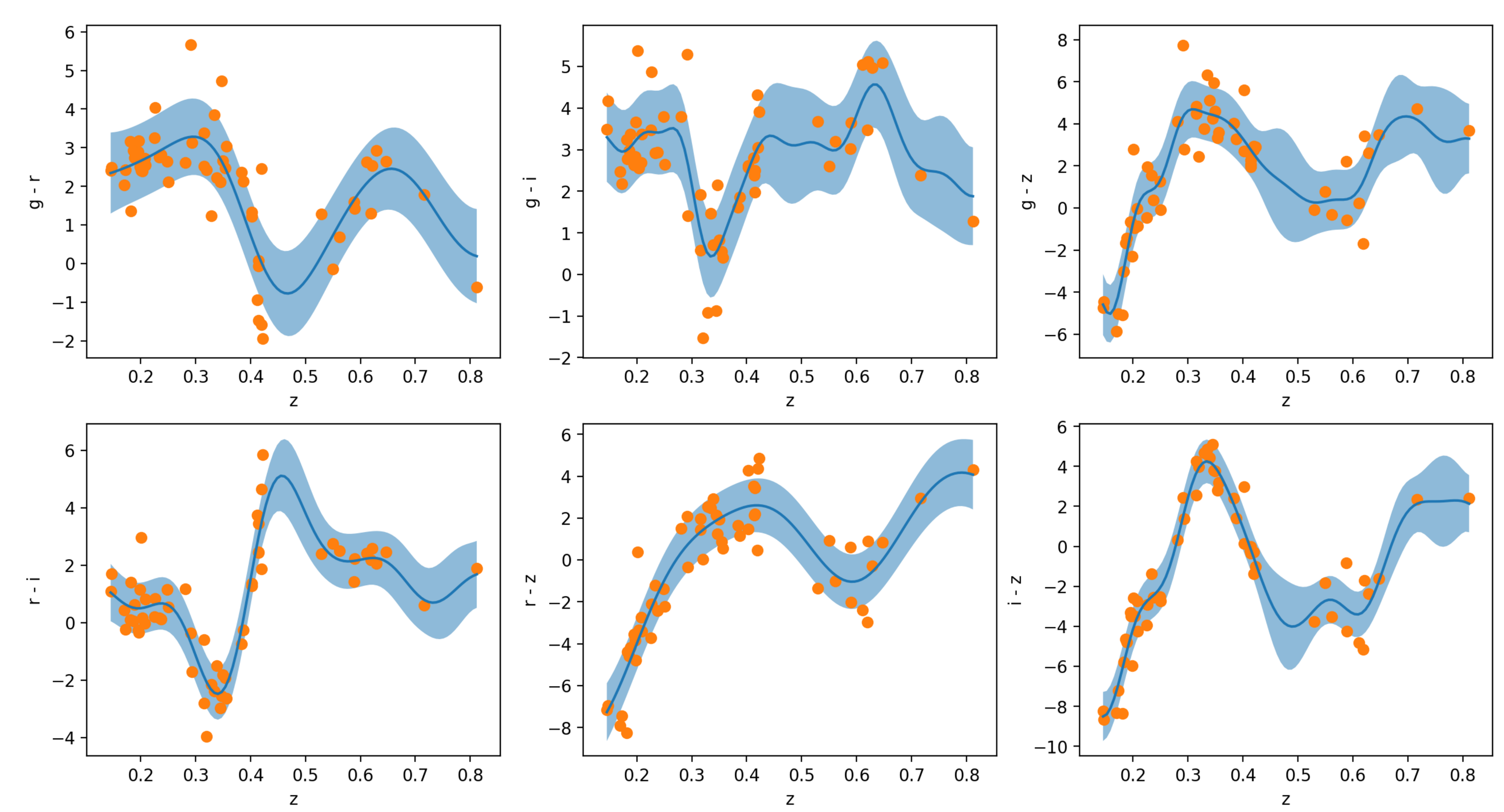

Figure 3. Color vs. redshift Gaussian Processes.

References

[1] P. Melchior *et al.*, "scarlet: Source separation in multi-band images by constrained matrix factorization," *Astronomy and Computing*, vol. 24, pp. 129 – 142, 2018.

[2] A.-L. Sun *et al.*, "Imaging extended emission-line regions of obscured AGN with the Subaru Hyper Suprime-Cam Survey," *Monthly Notices of the Royal Astronomical Society*, vol. 480, no. 2, pp. 2302–2323, 2018.

[3] E. Zackrisson, "Quasars and low surface brightness galaxies as probes of dark matter," Ph.D. dissertation, Department of Astronomy and Space Physics, Uppsala University, Sweden, 2005.

Implementation: Model Averaging

We weight each optimized model by its likelihood, thus giving more weight to models that produce a closer reconstruction of the original image. By combining samples in this way, we can reliably image EELRs despite uncertainty about the EELR spectra.

Our final estimator is:

$$\widehat{\mathbb{E}}(M \mid D) = \frac{1}{Z} \sum_i M_i \cdot \mathcal{L}(D \mid M_i)$$

where M is the model, D is the data (broad-band image and redshift), M_i are the model samples, and Z is a normalization factor.

The pixel-wise variance is:

$$\widehat{\sigma}^2(M \mid D) = \frac{\sum_i w_i (M_i - \widehat{\mathbb{E}}(M \mid D))^2}{V_1 - (V_2/V_1)}$$

where $w_i := \mathcal{L}(D \mid M_i)$, $V_1 := \sum_i w_i = Z$, and $V_2 := \sum_i w_i^2$.

Results

We use two performance metrics:

1. **Mean squared L2 error (MSE)** of the full model, which is proportional to the mean of the simplified negative log-likelihood assuming homoskedastic Gaussian error,

$$f(\mathsf{A}, \mathsf{S}) = \frac{1}{2} \|\mathsf{Y} - \mathsf{AS}\|_F^2$$

2. **Mean pixel intensity variance (MPIV)** of the EELR model, which is a measure of uncertainty about the model average

We compare three methods:

1. **Baseline**: uses SDSS spectra measurements
2. **GP**: uses spectroscopic redshift measurements
3. **Noisy GP**: same as the GP method, except uniform random noise is added to the redshift measurements to simulate photometric redshift

	Med. MSE	Med. MPIV
Baseline	**5.258**	**0.0098**
GP	5.276	0.0108
Noisy GP	5.281	0.0100

Table 1. Median performance metrics.

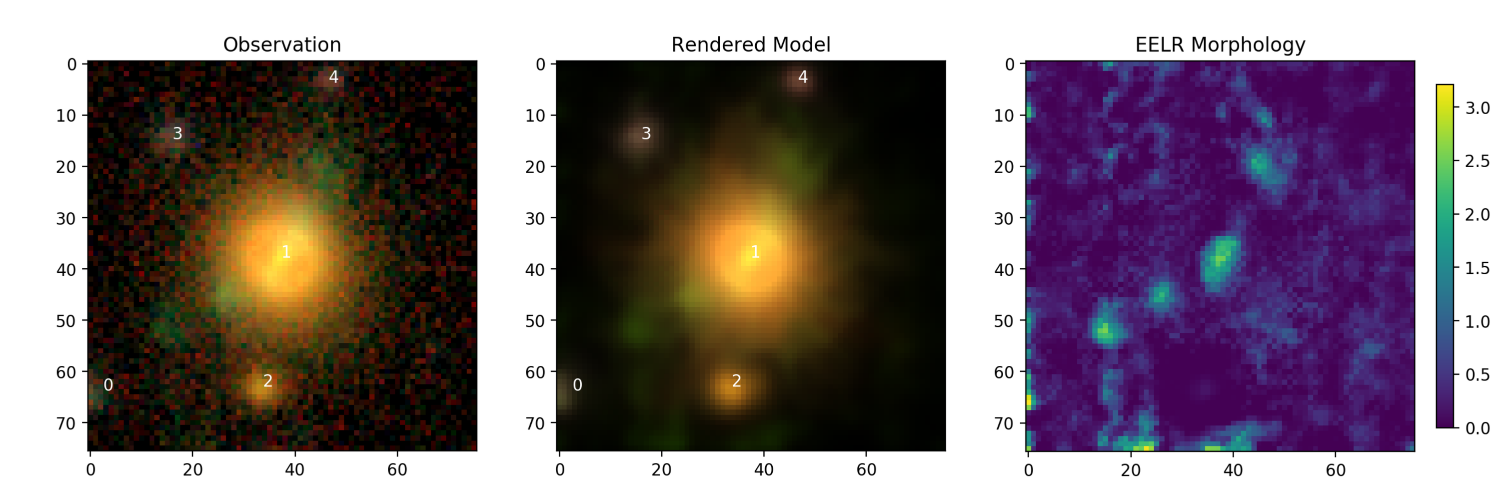

Figure 4. Example model averages using the GP method with noisy redshift.

Conclusion

- EELRs are promising sources of information about AGN
- We have developed a method of imaging EELRs using only photometric data
- Our method performs similarly to a method which is given spectra measurements rather than having to infer these from redshift
- Our method holds promise for detecting and imaging EELRs in future large-scale surveys