



Contribution ID: 144

Type: **Poster Presentation**

Estimating Bayesian Posteriors for Galaxy Morphological Parameters using Machine Learning

Tuesday, 31 May 2022 14:50 (3 minutes)

Galaxy morphology is connected to various fundamental properties of a galaxy and its environment, such as galaxy mass, star formation rate, stellar kinematics, merger history, etc. Thus, studying the morphology of large samples of galaxies can be a crucial clue to understanding galaxy formation and evolution.

In the past few years, although machine learning has been increasingly used to determine the morphology of galaxies, most previous works have provided only broad morphological classifications (without parameter estimation or uncertainty quantification). They have also required large sets of pre-classified training data.

We have developed Galaxy Morphology Posterior Estimation Network (GaMPEN), a machine learning framework that can estimate the Bayesian posteriors for a galaxy's bulge-to-total light ratio, effective radius, and flux. The computation of full Bayesian posteriors is crucial for drawing scientific inferences that properly account for uncertainty (e.g., deriving robust scaling relations, performing tests of theoretical models using morphology).

To predict posteriors, GaMPEN uses the Monte Carlo Dropout technique and incorporates the full covariance matrix in its loss function. The latter novel step empowers GaMPEN to simultaneously predict accurate posteriors for all three variables by incorporating the structured relationships between the output variables into its predictions.

GaMPEN also contains a novel Spatial Transformer Network (STN) that automatically crops input galaxy frames to an optimal size before determining their morphology. The STN trains along with the rest of the framework, with no additional supervision, and will be crucial in applying GaMPEN to new survey data with no radius measurements.

We have used GaMPEN to determine robust morphological parameter posteriors for ~ 1 million $z < 0.6$ galaxies in the Hyper Suprime-Cam Wide survey. In order to not require a large pre-classified training set of real galaxies, we first trained GaMPEN on realistic simulations of galaxies and then performed transfer-learning/domain adaptation using a small amount of real data.

We have demonstrated that GaMPEN's predicted posteriors are well-calibrated and accurate ($\leq \pm 4\%$ of corresponding confidence levels). Testing has shown that GaMPEN predictions become less precise for especially small or faint galaxies, where the algorithm correctly predicts correspondingly larger uncertainties. We have also demonstrated that by predicting qualitative labels (instead of quantitative values) in certain small regions of the parameter space where GaMPEN's residuals are higher, we can achieve accuracies of $\geq 95\%$.

GaMPEN is the first machine learning framework for determining joint posterior distributions of multiple morphological parameters, does not require large amounts of real training data, and is also the first application of an STN in astronomy.

Main Topic

Deep learning

Secondary Topic

Classification and regression

Participation mode

In person

Primary authors: GHOSH, Aritra (Yale University); URRY, C. Megan (Yale University); RAU, Amrit (Yale University); TIAN, Chuan (Yale University); PERREAULT-LEVASSEUR, Laurence (Univesité de Montréal | Flatiron Institute)

Presenter: GHOSH, Aritra (Yale University)

Session Classification: Poster Session Day 2