Estimating Bayesian Posteriors for Galaxy Morphological Parameters using Machine Learning

***Aritra Ghosh**, C. M. Urry, A. Rau, L. P. Levasseur, M. Cranmer, K. Schawinski, D. Stark, C. Tian, R. Ofman, T. Ananna, C. Auge, N. Cappelluti, D. B. Sanders, E. Treister*

Introduction

Studying the morphology of large samples of galaxies at different redshifts is crucial to understanding the physics of galaxy formation and evolution.

Although astronomers have increasingly used machine learning for morphology determination, most of these algorithms:-

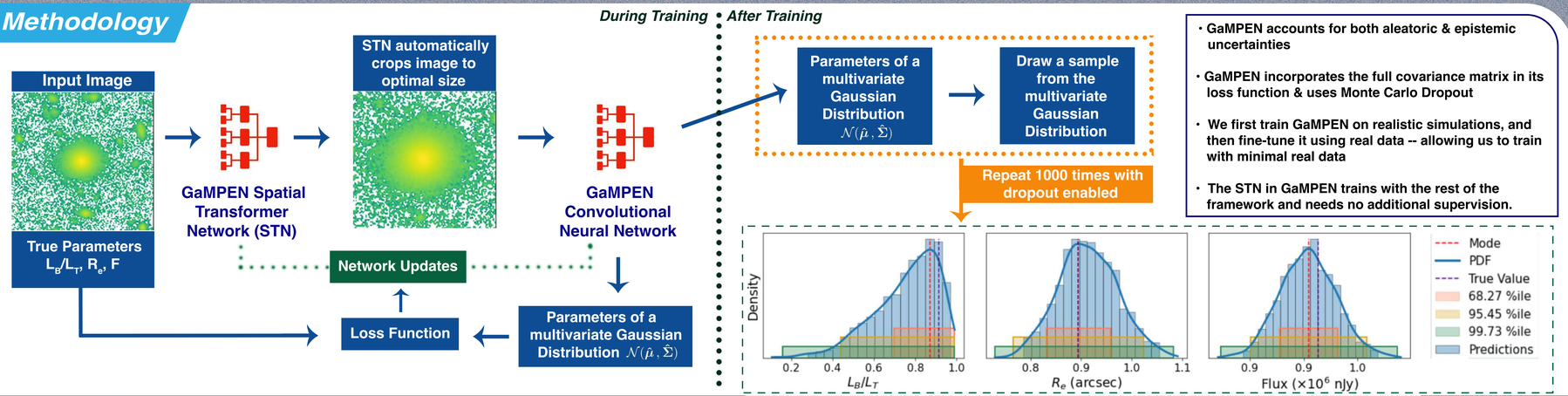
- **provide only broad classifications, without any parameter-estimation**
- **do not provide estimates of uncertainties**
- **need large amounts of pre-classified training data**
- **perform poorly in crowded fields due to the presence of secondary galaxies in the frame**

To address these challenges, we have developed the Galaxy Morpohlogy Posterior Estimation Network (GaMPEN). GaMPEN estimates values and uncertainties for a galaxy's bulge-to-total light ratio (L_B/L_T), effective radius (R_e), and flux (F).

GaMPEN is a deep learning framework that

- **Can accurately determine Bayesian posteriors for morphological parameters of galaxies**
- **Does not need large amounts of pre-classified data for training**
- **Automatically crops input images to focus on the galaxy of interest**
- **Has been tested extensively on HSC data and can be easily applied to other datasets**

Methodology



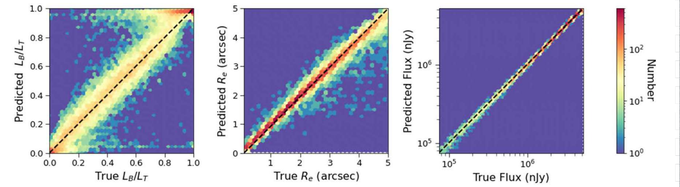
- **GaMPEN accounts for both aleatoric & epistemic uncertainties**
- **GaMPEN incorporates the full covariance matrix in its loss function & uses Monte Carlo Dropout**
- **We first train GaMPEN on realistic simulations, and then fine-tune it using real data -- allowing us to train with minimal real data**
- **The STN in GaMPEN trains with the rest of the framework and needs no additional supervision.**

Results

Coverage probabilities (i.e., the percentage of times the true value lies within a specific central confidence interval) are shown for various confidence intervals.

GaMPEN's estimates of posterior distributions are accurate & well-calibrated

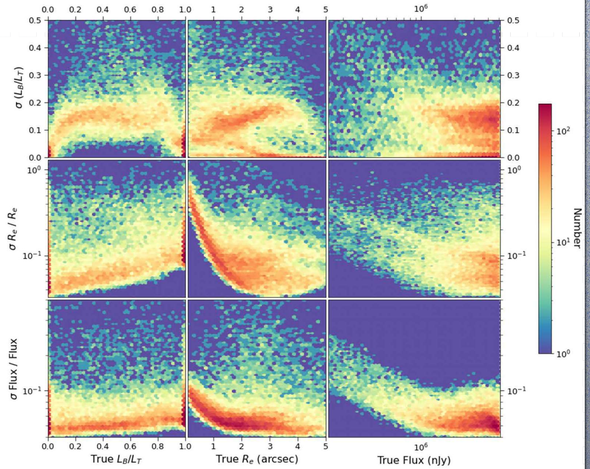
Parameter Name	68.27% Conf. Level	95.45% Conf. Level	99.73% Conf. Level
L_B/L_T	71.8%	96.9%	98.9%
R_e	68.1%	95.9%	98.3%
F	78.7%	98.2%	99.9%
Mean	72.9%	97.0%	99.0%

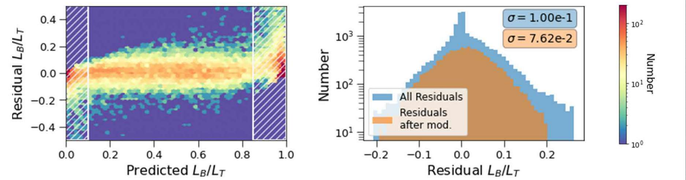


The most probable values of the predicted distributions closely track the true parameter values. The typical errors produced for L_B/L_T, R_e, and Flux are 0.1, 0.17", and 6.3 X 10^4 nJy, respectively.

The width of the 68.27% confidence interval for different parameters.

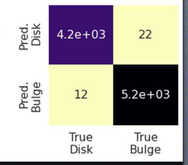
GaMPEN residuals are higher for smaller and fainter galaxies. GaMPEN correctly accounts for this by predicting appropriately higher uncertainties in these regions.





GaMPEN residuals are high for L_B/L_T values very close to 0 or 1. By making quantitative predictions only in the range $0.1 < L_B/L_T < 0.85$, and transforming the predictions in the rest of the space (white shaded region) to qualitative labels (e.g., "highly disk-dominated"), the typical error decreases.

The produced labels are > 99% accurate, as shown by the confusion matrix on the right.



Yale

Supported By:-

NSF NASA Y|CRC aws

GaMPEN will be publicly available in Fall 2022 along with trained models. For early access to the beta version, please contact me!