AI FOR COSMOLOGY

Nesar Ramachandra
nramachandra@anl.gov
Computational Science Division
Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

LSST DESC Dark Energy Science Collaboration

SPHEREx

Argonne NATIONAL LABORATORY

INTRODUCTION

This talk highlights a subset of the AI-for-Cosmology efforts at Argonne.

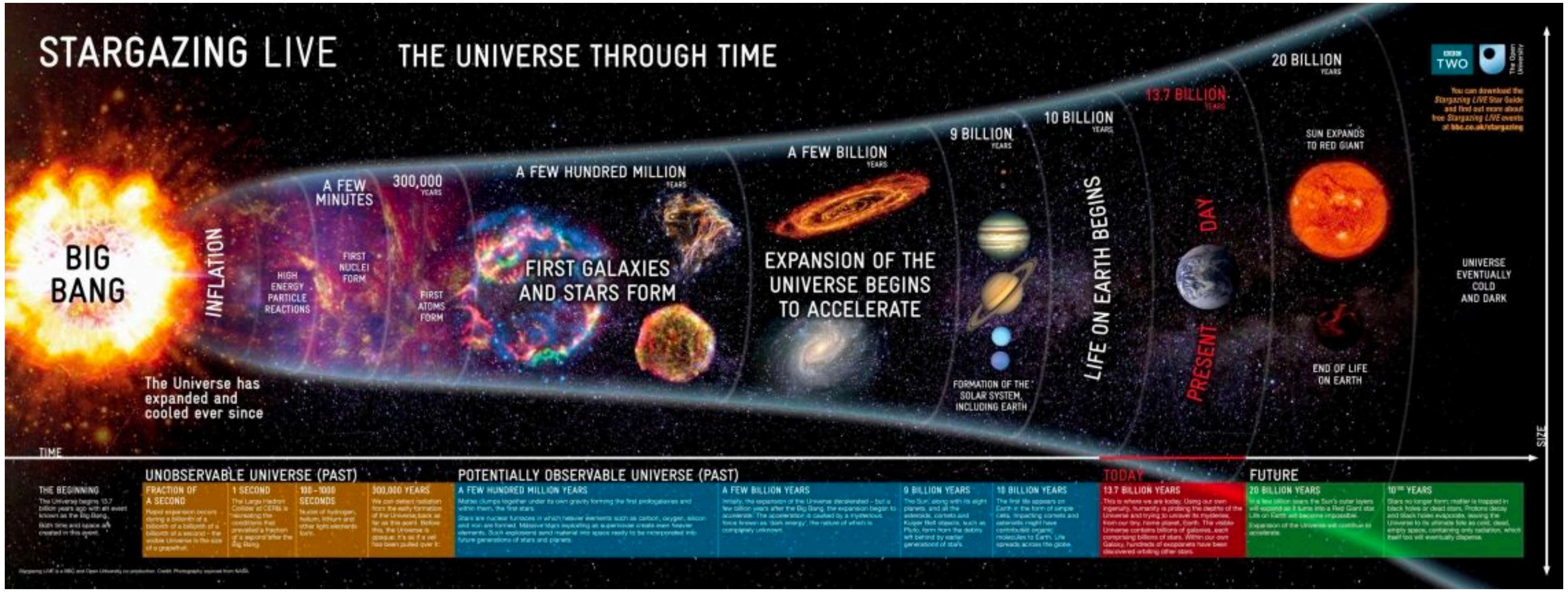
Common themes here are:

- Synthetic/simulation data to enhance/replace real astronomical observations.
- Bayesian/probabilistic schemes rather than point-predictions.
- Explainability of the AI algorithms.

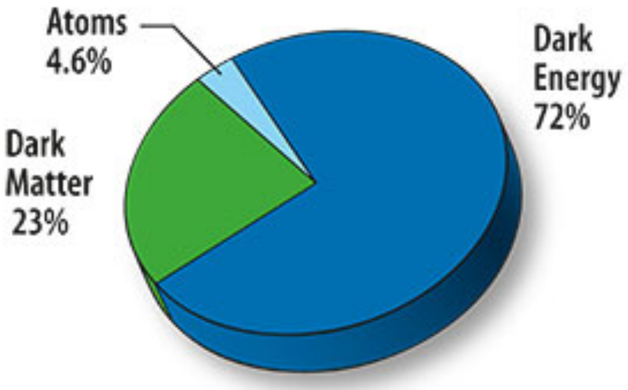
Case study:

- Image processing pipelines for de-noising, de-blending etc.
- Probabilistic classification and regression
- Latent space exploration

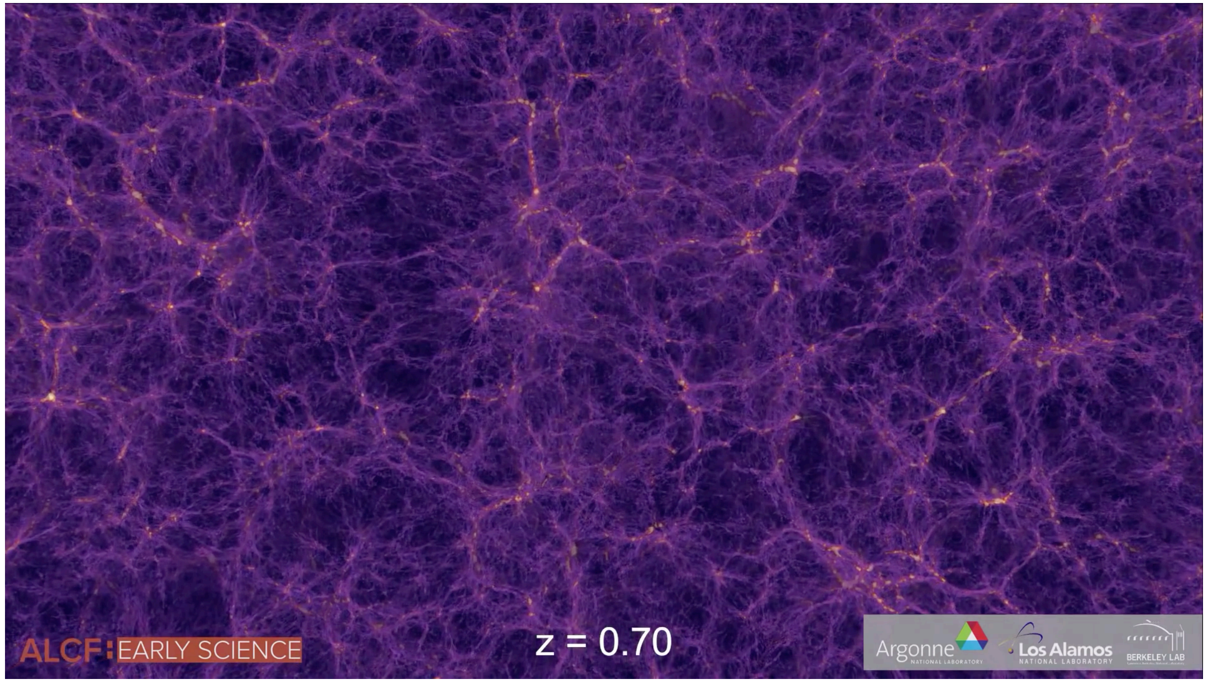
SHORT INTRO TO COSMOLOGY



Source: open.edu

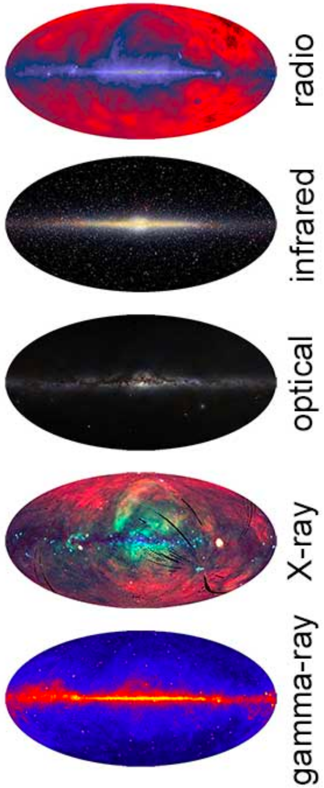


Credit: nasa.gov

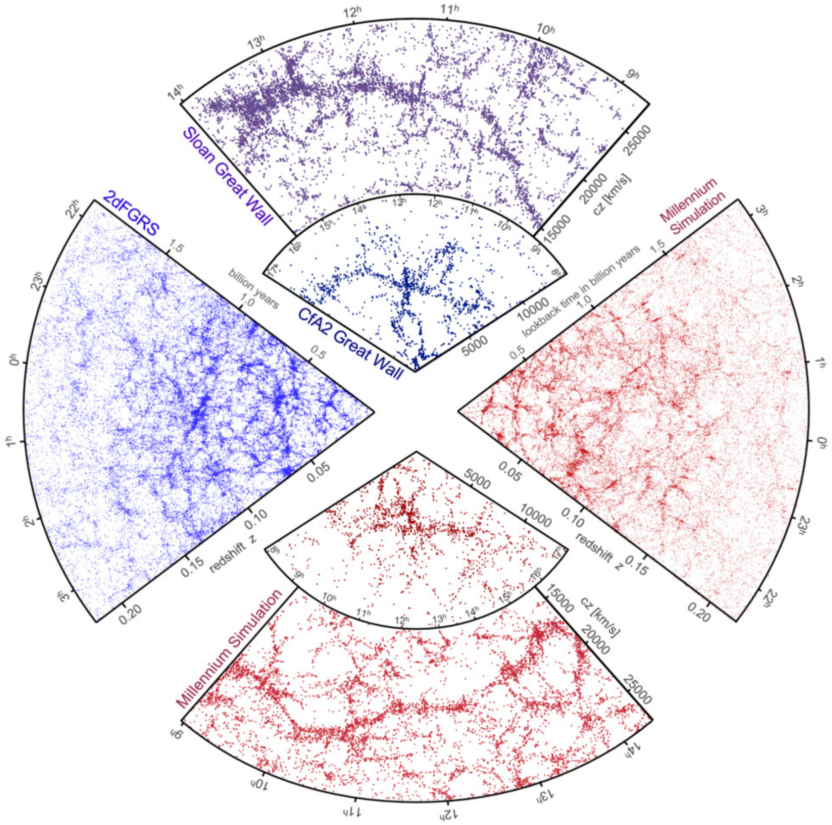


Name (Symbol)	Description	Value
Hubble Constant (H_0)	Current rate of expansion of the universe.	67.4 ± 0.5 km/s/Mpc
Cosmological Constant (Λ)	Energy density of space, or vacuum energy.	1.1056×10^{-52} m^{-2}
Dark Energy Density ($\Omega\Lambda$)	Fraction of the universe's energy density consisting of dark energy.	0.6847 ± 0.0073
Dark Matter Density (Ωc)	Fraction of the universe's energy density consisting of dark matter.	0.264 ± 0.013
Baryon Density Parameter ($\Omega b\, h^2$)	Density of ordinary matter (baryons) relative to the critical density.	0.0224 ± 0.0001

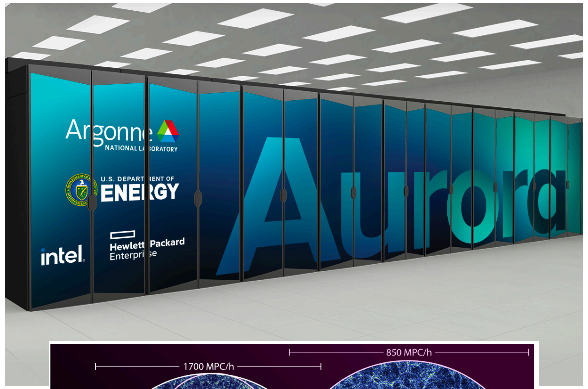
STUDYING THE UNIVERSE: JOINT EFFORTS

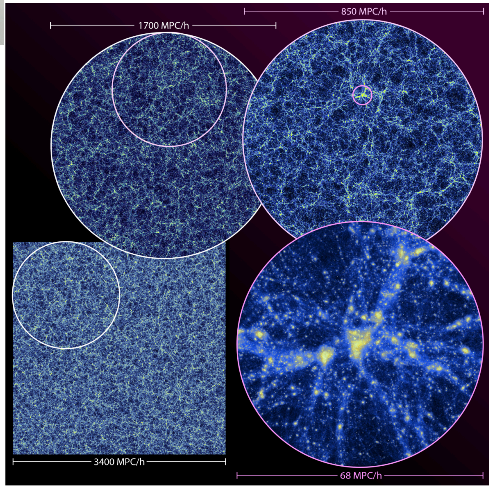


Observational Astronomy



Zavala, J.; Frenk, C.S. Dark Matter Haloes and Subhaloes. *Galaxies* **2019**, *7*, 81.



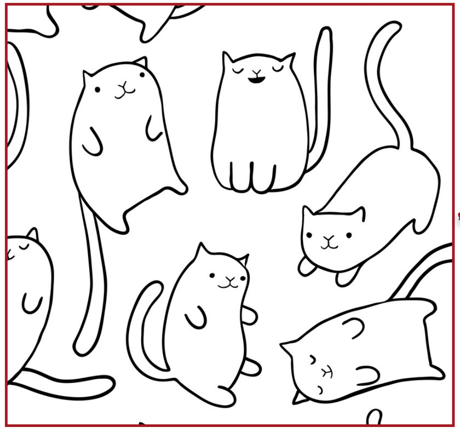


Theoretical and Computational Cosmology

CAVEATS IN COSMOLOGICAL STUDIES

- Multi-modal data:
  - Graphs, images, time-series, summary vectors, scalars, text.
- Expensive:
  - Both simulations and observations are from expensive science campaigns
- Multi-fidelity:
  - Data from different sources have different resolutions, approximations and systematic effects.
  - Transfer of knowledge is not straightforward.
- Data coverage:
  - Gaps, biased datasets are common. No assumption of a ‘fair’ sampling.

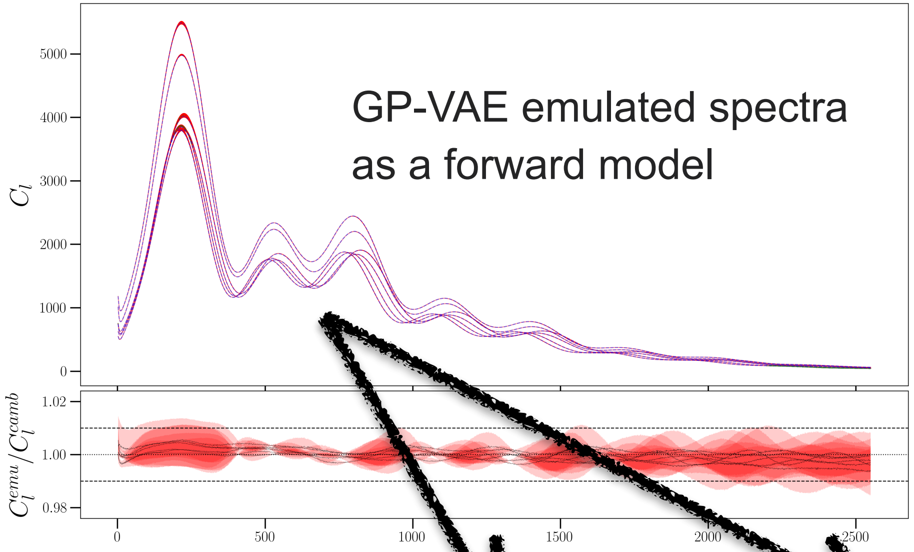
- Analysis requirements
  - Precision cosmology has high error requirements
  - Traditional statistical analyses have been highly successful.
- Prior domain knowledge:
  - Studies assume known physics, conservation laws.



TRAIN

TEST

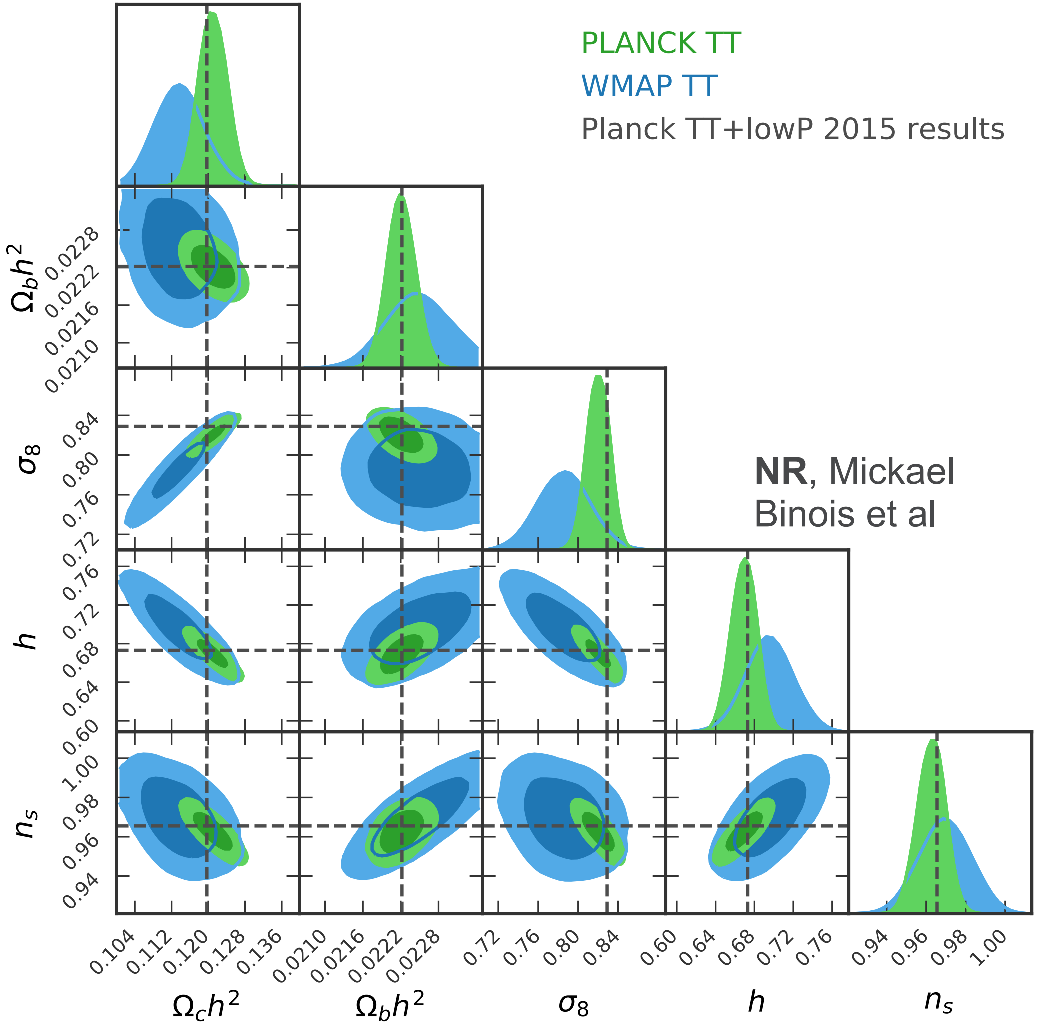
A FEW EXAMPLES: BAYESIAN INFERENCE WITH EMULATORS



$$\mathscr{L}(D\,|\,\theta) \propto \exp\left[-\frac{1}{2}\sum_{i,j}\left(D - f(\theta)\right)_i C_{ij}^{-1}\left(D - f(\theta)\right)_j\right]$$

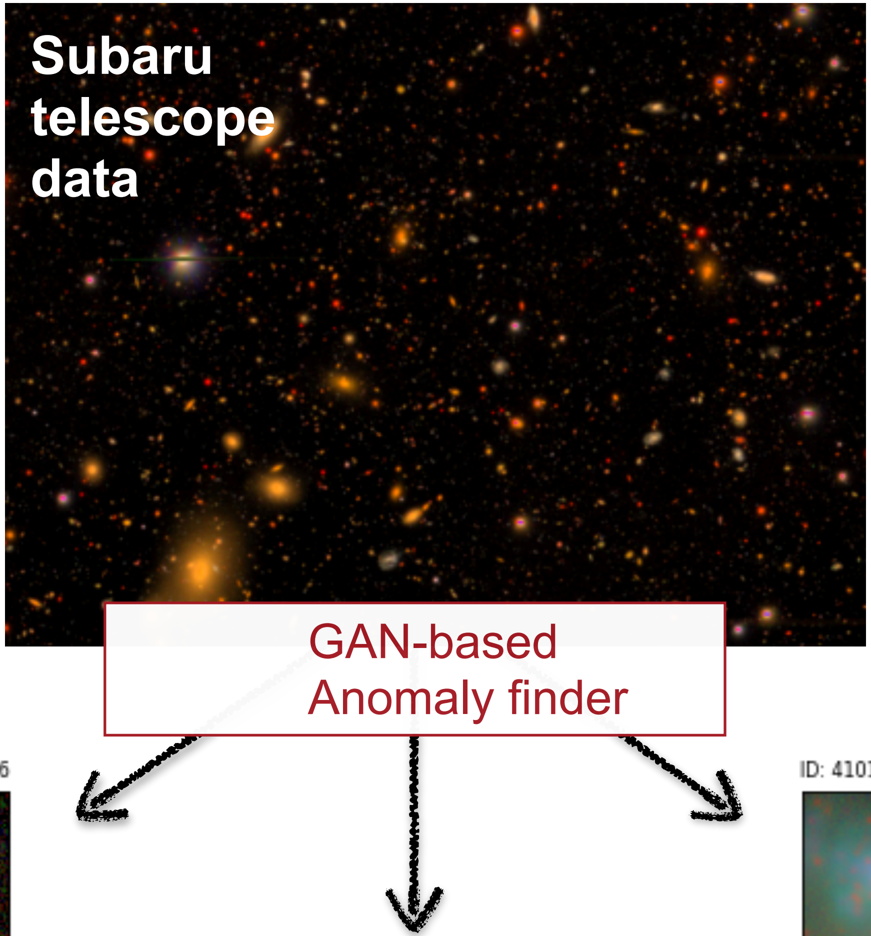
$$P(\theta\,|\,D) \propto \mathscr{L}(D\,|\,\theta)P(\theta)$$

MCMC sampling for PLACK/WMAP data

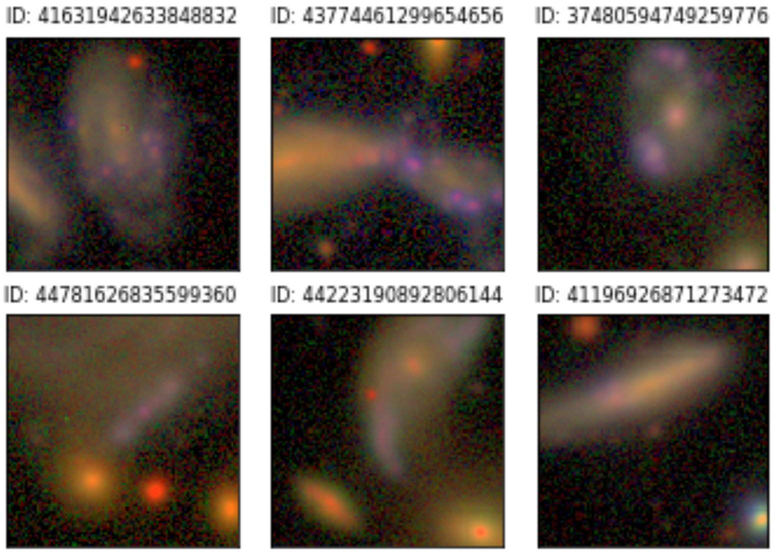


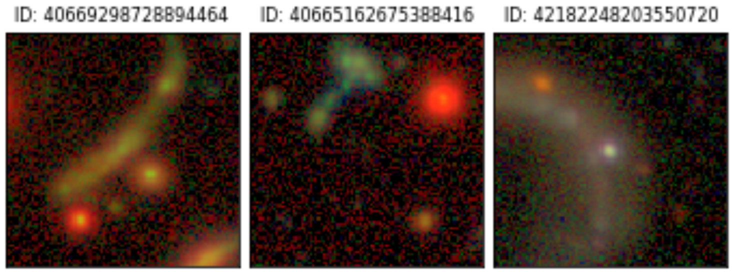
NR, Mickael Binois et al

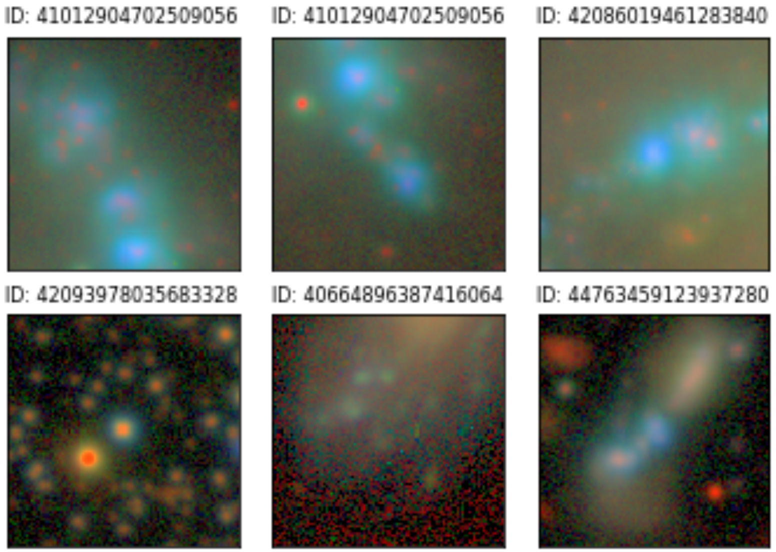
A FEW EXAMPLES: FINDING UNKNOWN UNKNOWNS OBJECTS IN THE SKY



https://arxiv.org/abs/2012.08082

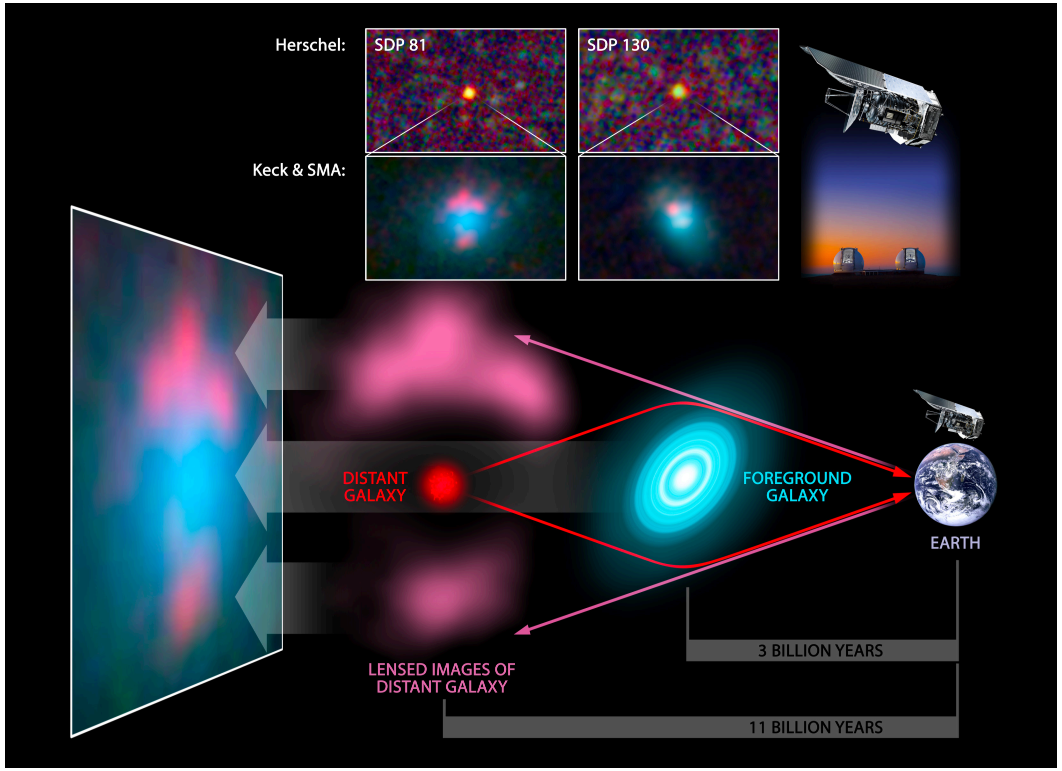


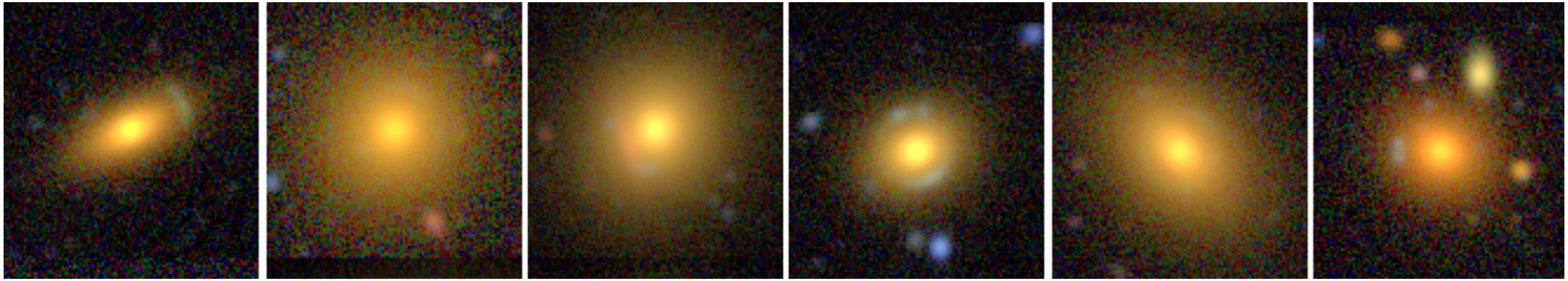




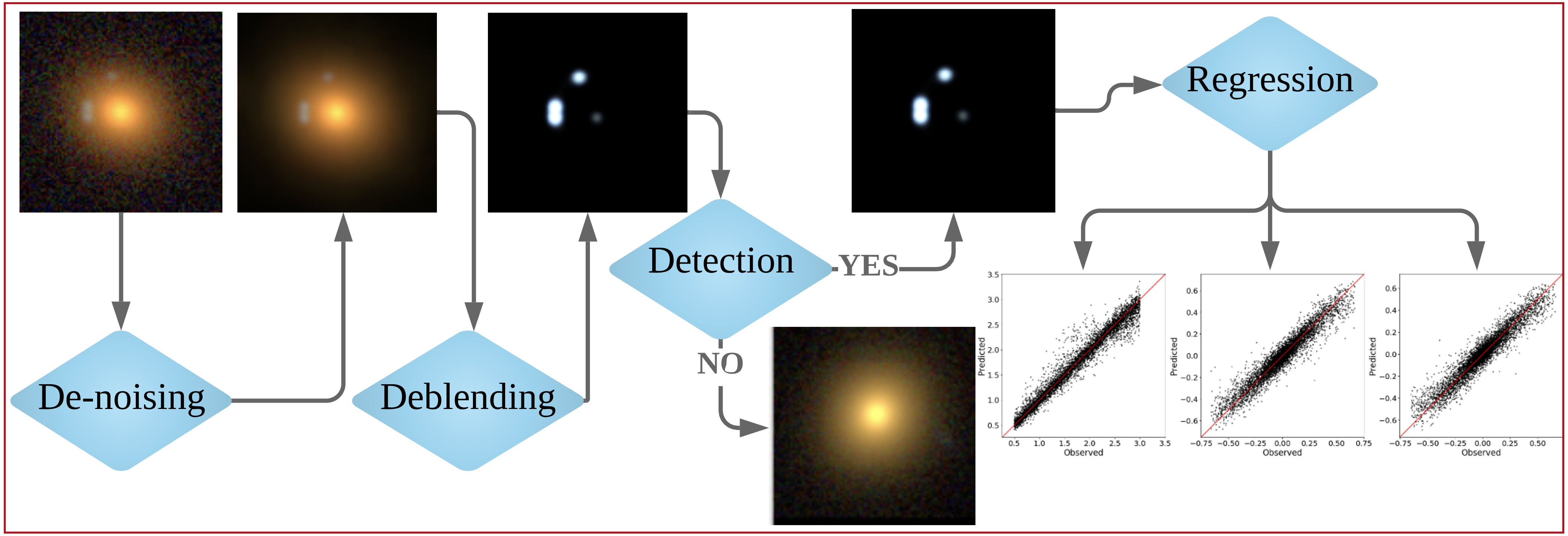
CASE STUDY: GALAXY-SCALE STRONG LENSING

- Strong-lenses are rare objects.
  - But understanding them is key to several questions: Distribution of dark matter, expansion of the Universe.
- Discrepancy with current amount of observed data vs future data
  - Observed data is/will be a highly imbalanced dataset
  - Tractable physical models



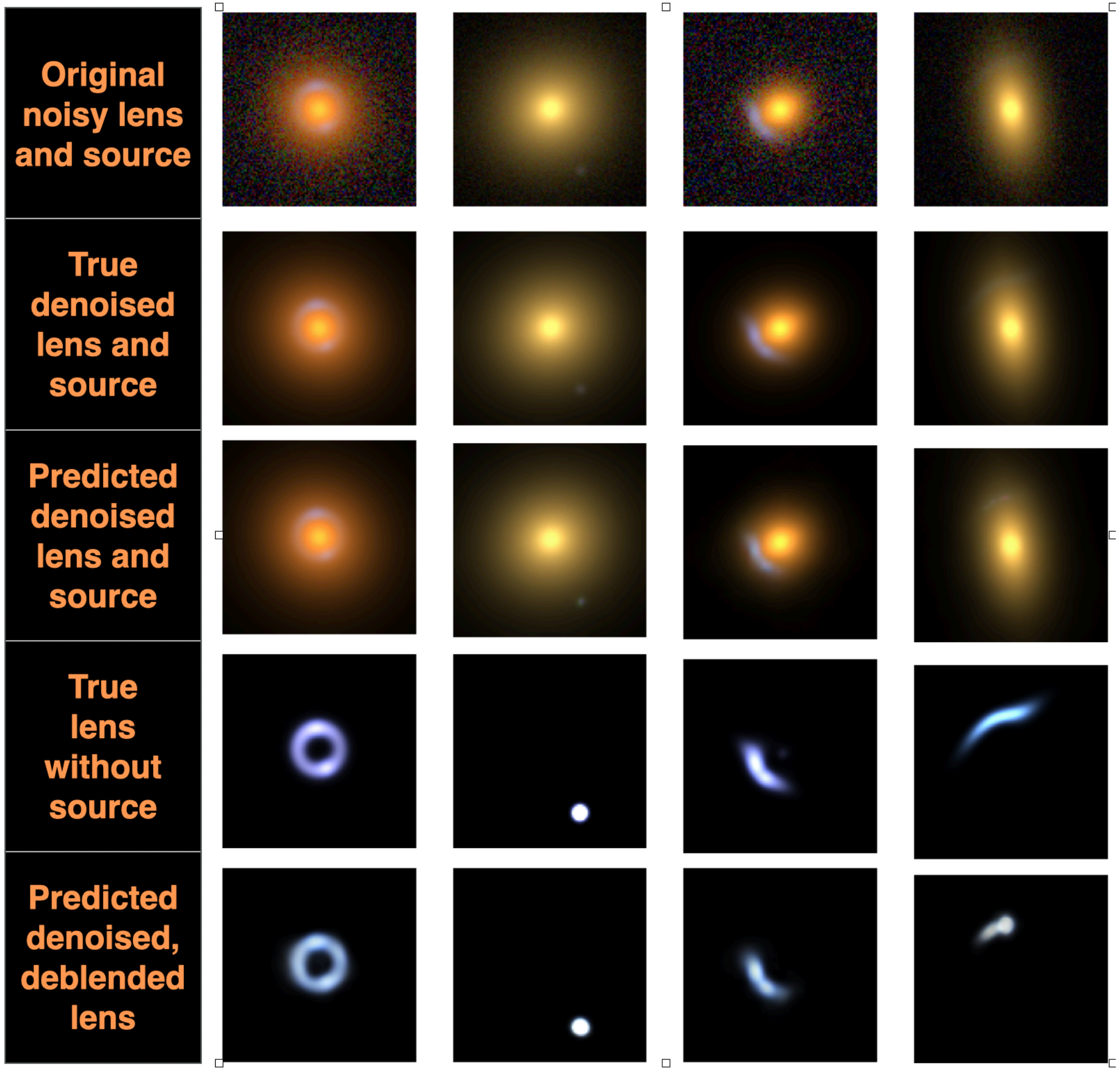


INTERPRETABLE LEARNING PIPELINES



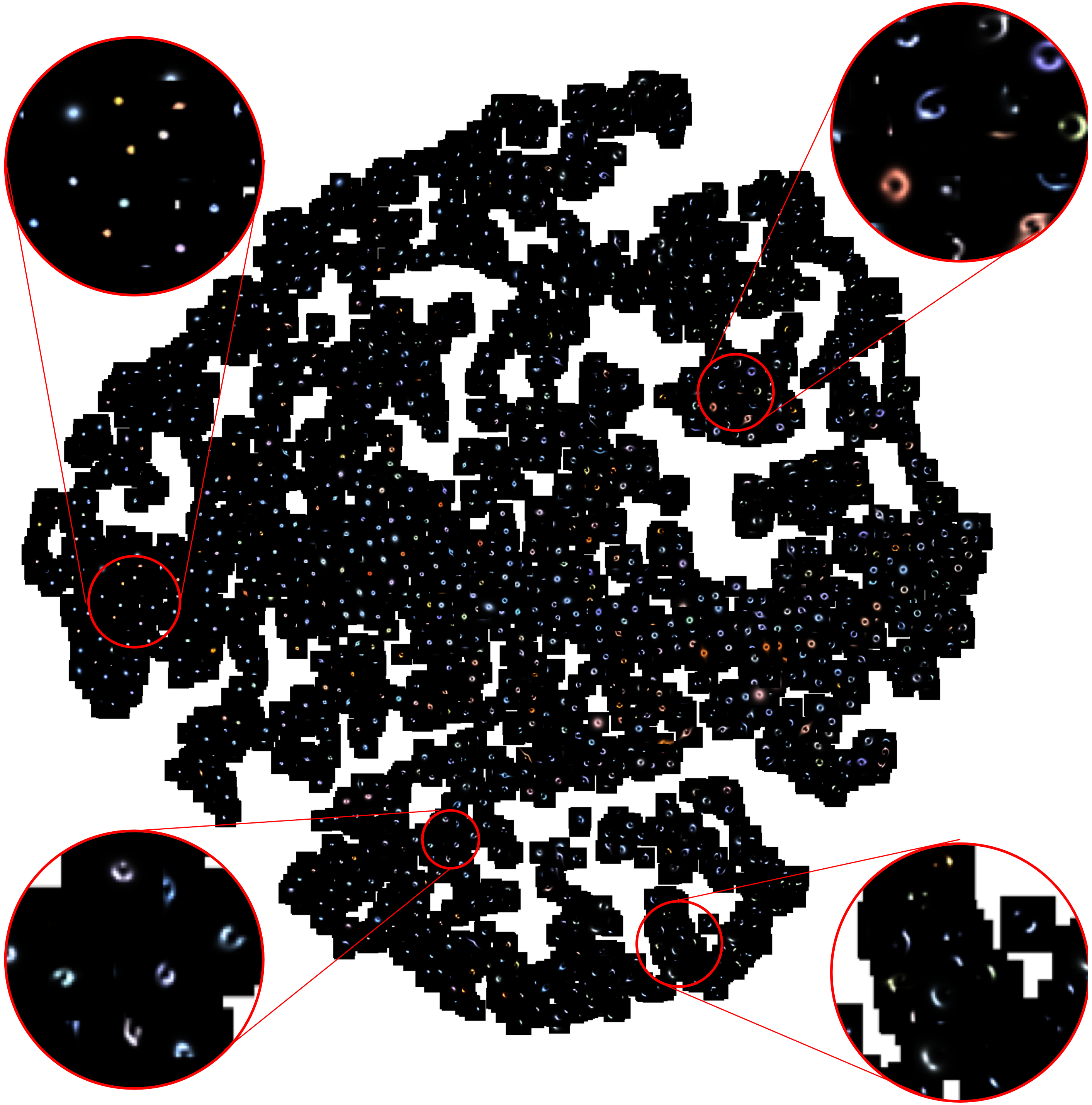
- Each AI-module is independently trained and validated. (Super-resolution modules for Denoising and deblending, Information bottleneck design for detection and regression)
- Synthetic data allows one to train modular pipelines that enable better control over systematics than end-to-end training methods

INTERPRETABLE STRONG LENS END-TO-END ANALYSIS PIPELINE

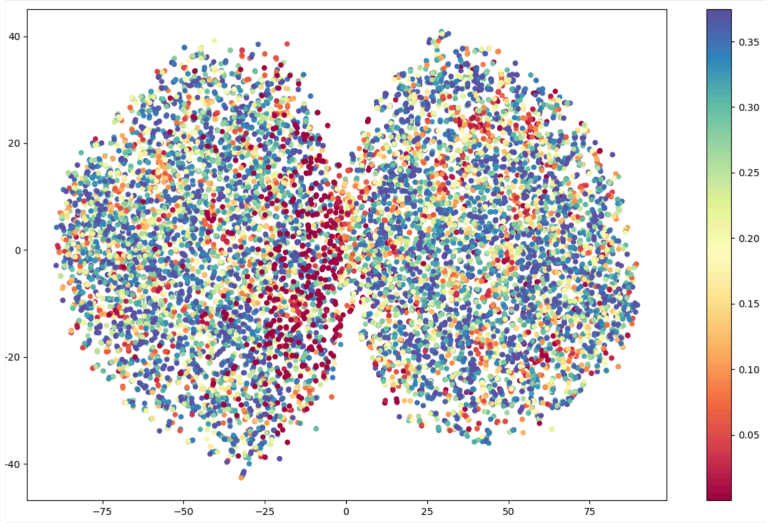


Sandeep Madireddy, **NR** et al**:**
arxiv.org:1911.03867

Argonne NATIONAL LABORATORY



Variational Information Bottleneck and representation learning



Uncertainty quantification for classification

U.S. DEPARTMENT OF ENERGY | Office of Science

CONCLUSIONS

Argonne NATIONAL LABORATORY

- Cosmological studies involve variety of data modalities, with vast amount of data. This makes data-driven AI-models extremely valuable.
- Synthetic datasets are often a necessity in Cosmological analysis.
- Careful experimental design, robust data creation, extensive validations are all required while dealing with synthetic data.
- Interpretable, uncertainty quantified models are still very important, probably even more so while using synthetic data in training.