Machine Learning Dynamical Models: Applications to Simulations and JAM

David Simon[1], Michele Cappellari[1], Shude Mao[2], Jiani Chu[2], Dandan Xu[2]

1. Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
2. Department of Astronomy, Tsinghua University, Beijing, 100084, China

Integral Field Spectroscopy

ATLAS3D

BLUE muse
multi unit spectroscopic explorer

muse

sami
galaxy survey

NIRSpec

MIRI

Key Questions

- Can Machine Learning be used to Improve Existing Methods?
- Can Machine Learning be used to Create New Methods?

Dynamical Modelling Methods

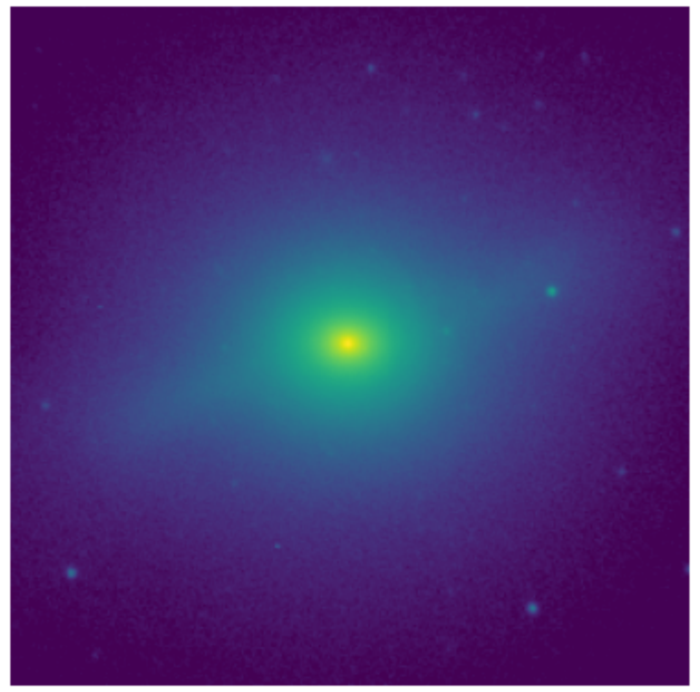
Schwarzschild Modelling

- Axisymmetric or Triaxial
- No anisotropy assumptions
- Uses full LOSVD
- Slow run time ~10 hours
- Non-smooth Chi2
- Lots of hyper-parameters

Jeans Modelling (JAM)

- Axisymmetric
- Cylindrically/spherically aligned velocity ellipsoid
- Uses 2nd moment
- Fast run time ~ 10 sec
- Smooth chi2
- Few hyper-parameters

JAM Modelling Pipeline



Photometry

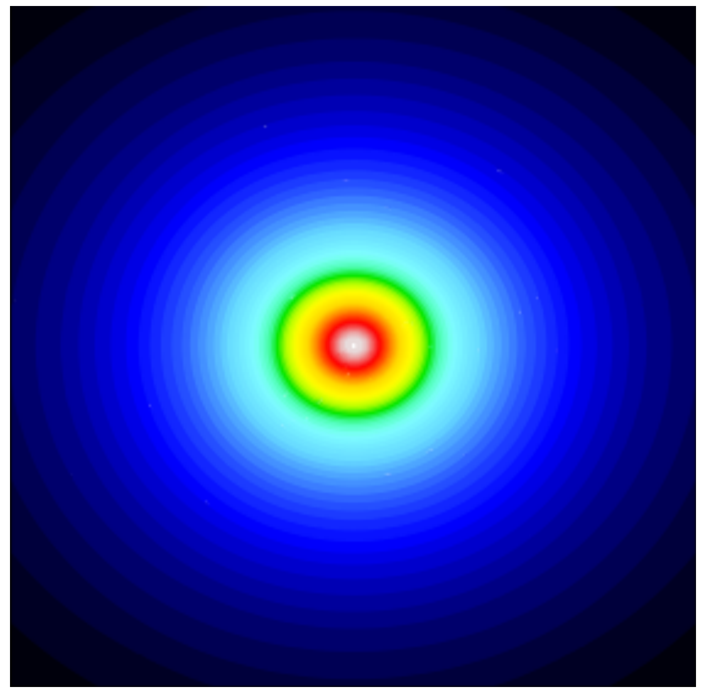
Gravitational Potential

Multi-Gaussian Expansion (MGE)

Light MGE

Mass MGE

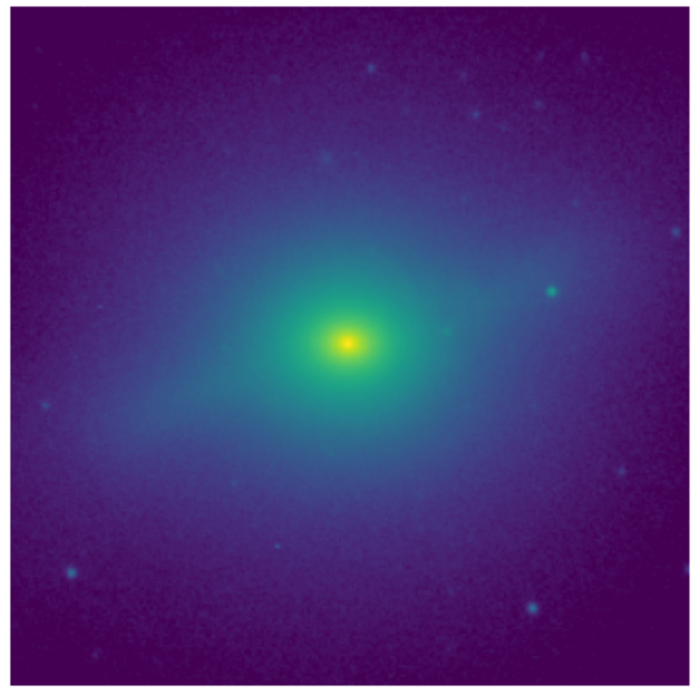
Jeans Anisotropic Modelling



Free Parameters

Inc, beta, M/L, Mbh, …

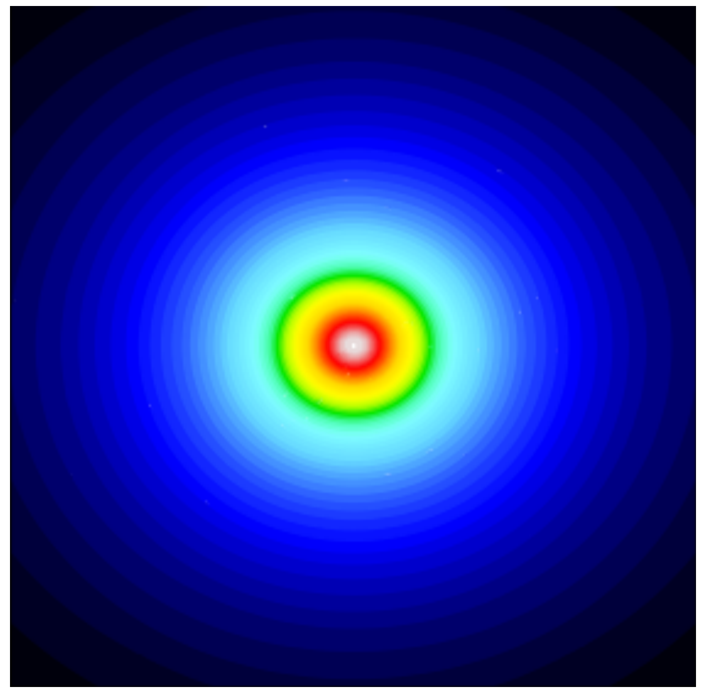
ML Modelling Pipeline



Photometry

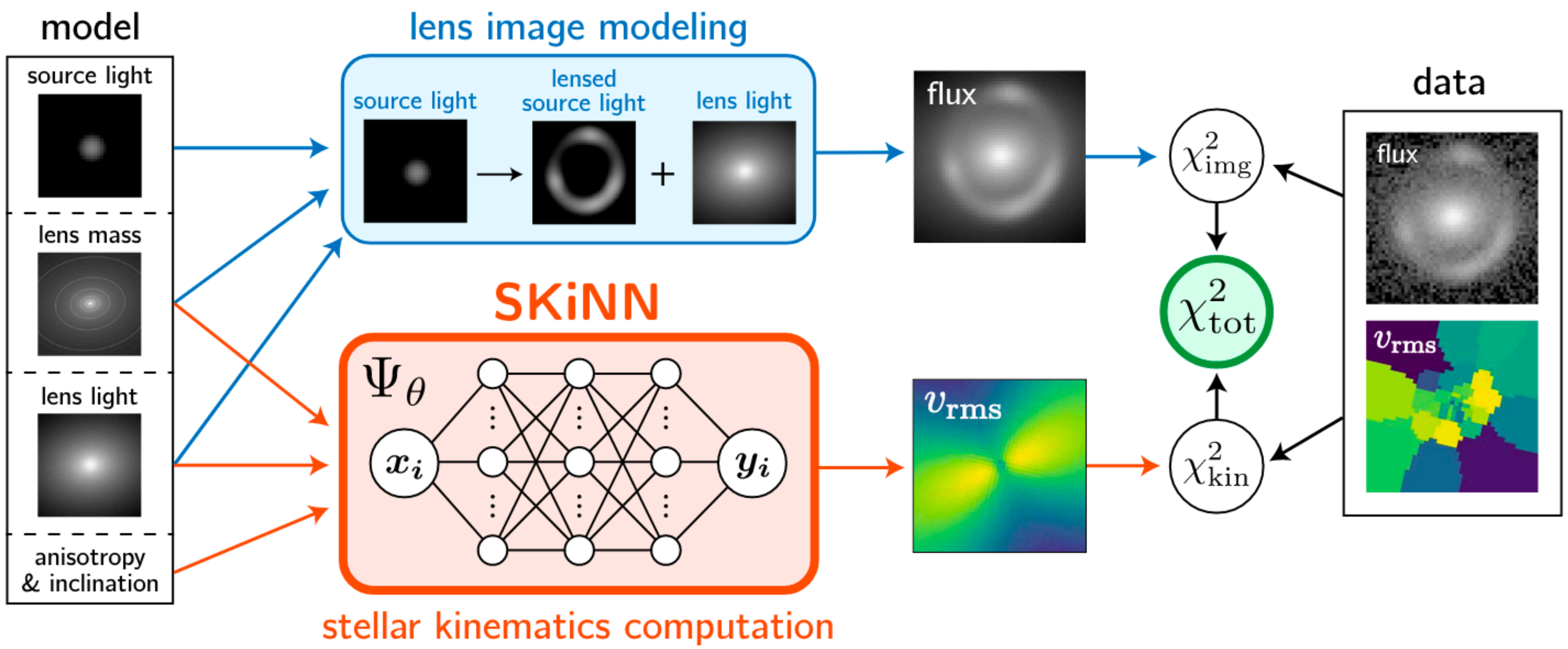
Gravitational Potential

Machine Learning Algorithm



Free Parameters

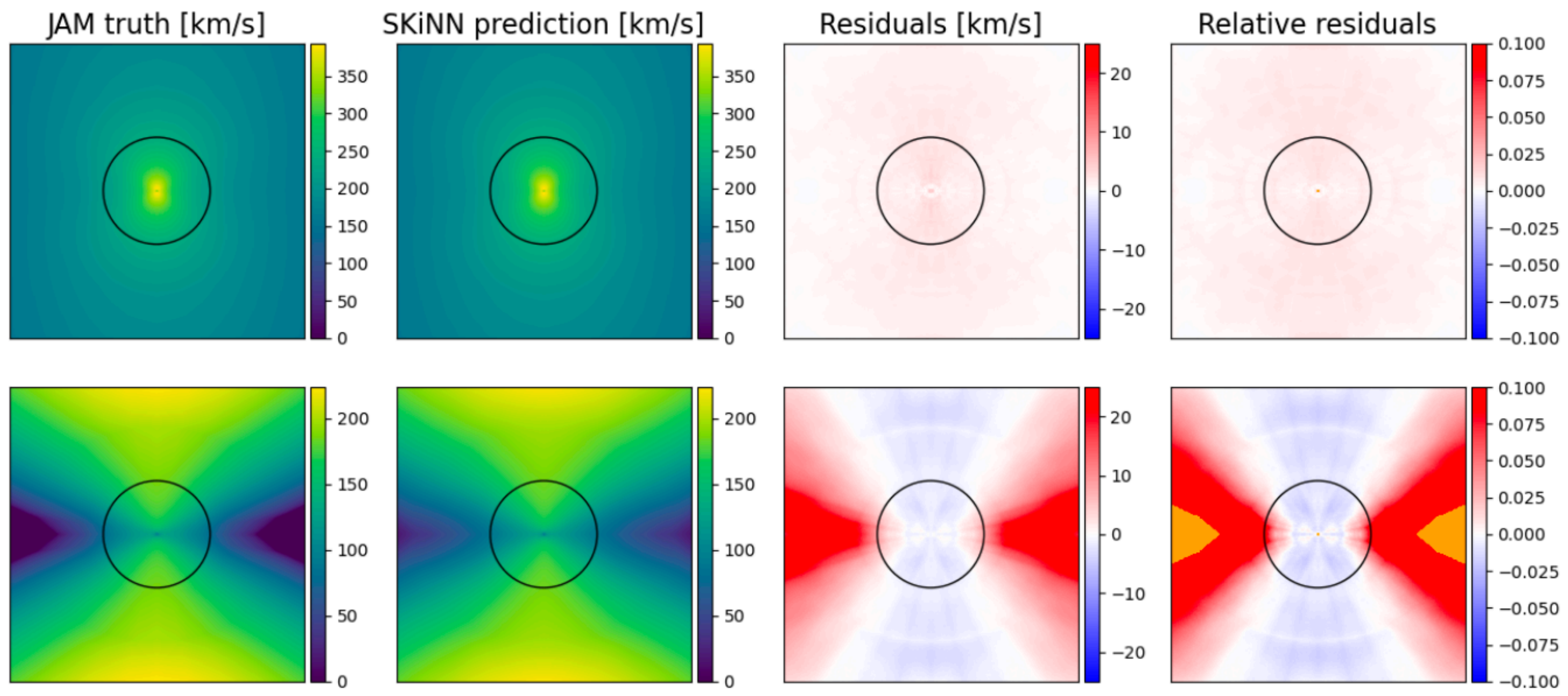
Gomer et al 2023 Architecture



- Joint lensing + kinematics inference
- Kinematics likelihood is a large bottle-neck (~500x slower than lensing likelihood)
- Conditionally Independent Pixel Synthesis (CIPS)
- Hardware requirements: GPU

(Gomer et al 2023, Fig. 1)

Gomer et al 2023 Training Set



Parameter	Description	Training set bounds
θ_E	Einstein radius	$[0.5, 2'']$
γ	2D PL slope (mass)	$[1.5, 2.5]$
q_M	Axis ratio (mass)	$[0.6, 1.0]$
q_L	Axis ratio (light)	$[0.6, 1.0]$
R_{Sersic}	Sérsic radius (light)	$[0.5\theta_E, \theta_E]$
n_{Sersic}	Sérsic index (light)	$[2, 4]$
β_z	Anisotropy	$[-0.4, 0.4]$
i	Inclination	$[\arccos(0.6), 90°]$
Map resolution	$0.02''$	
Map size	$11''$	

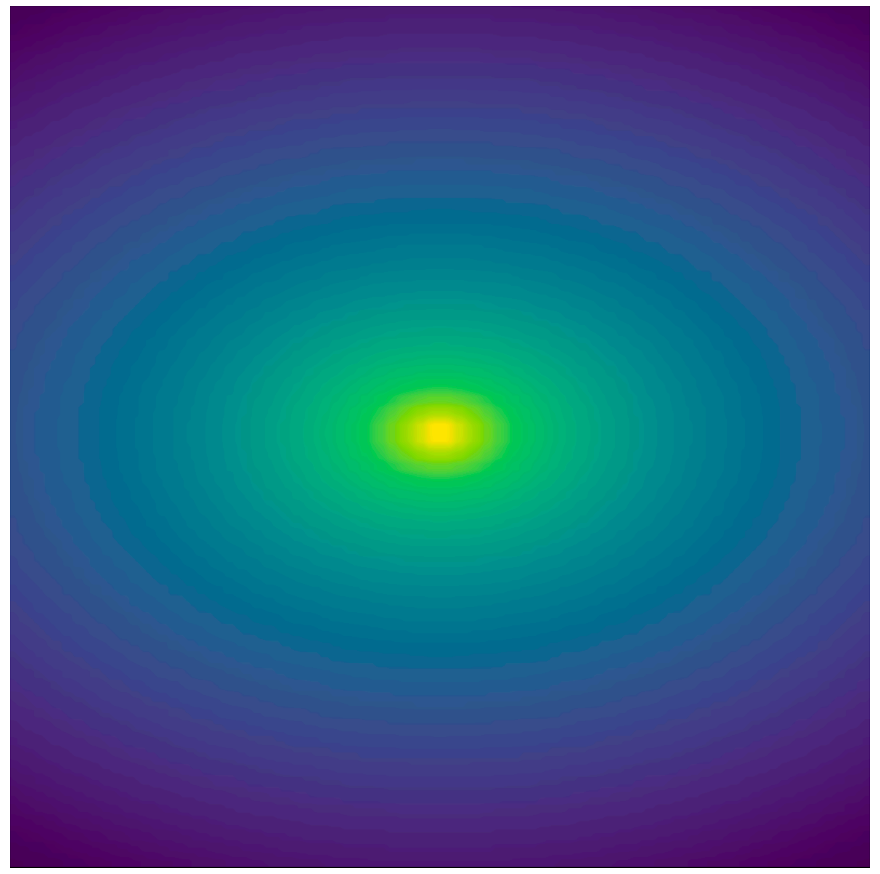
- 9000 Mock Galaxies
  - 4000 training
  - 1000 validation
  - 4000 test
- Uniformly sample Sersic galaxies
- Run time of ~100ms on GPU
- Better than 1% accuracy for most galaxies

(Gomer et al 2023, Fig.2 & Table 1)

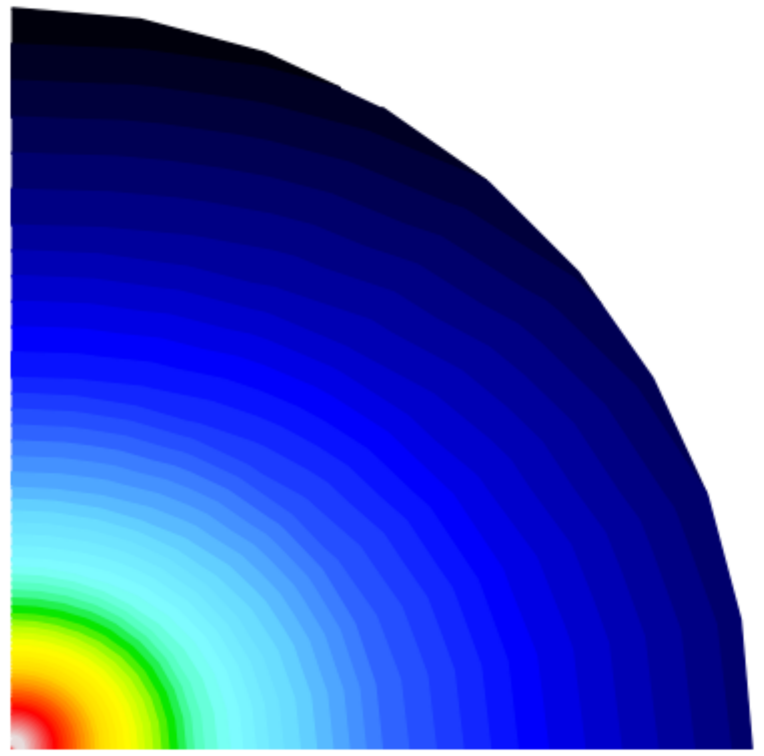
ML Architecture Requirements

- Directly use photometry
- Fast run time for MCMC
- Make use of symmetries (axisymmetry, $\overline{v_{\mathrm{los}}^2} \propto M$)
- Compatible with GPU and CPU

Training Set



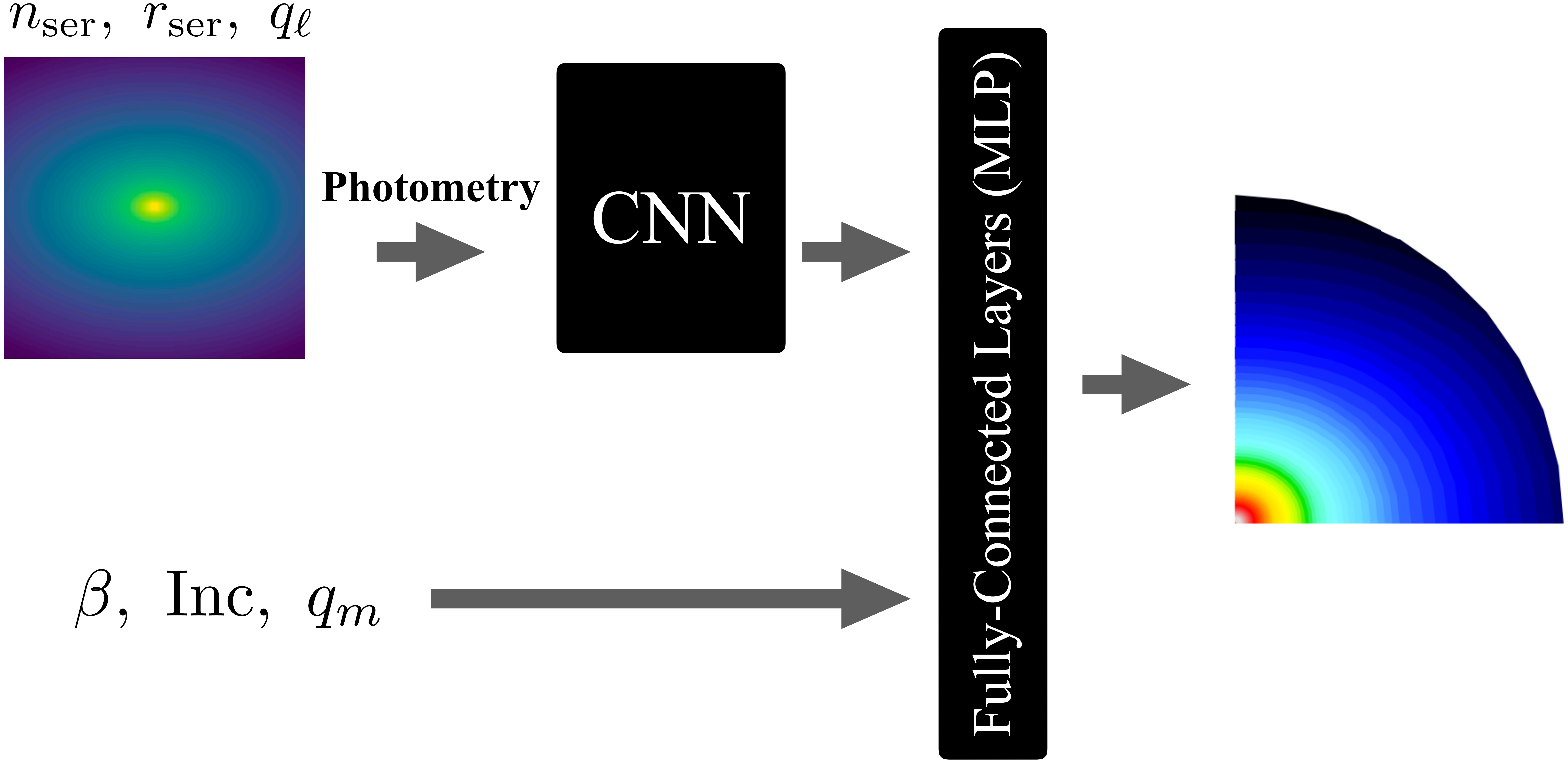
JAM Vrms



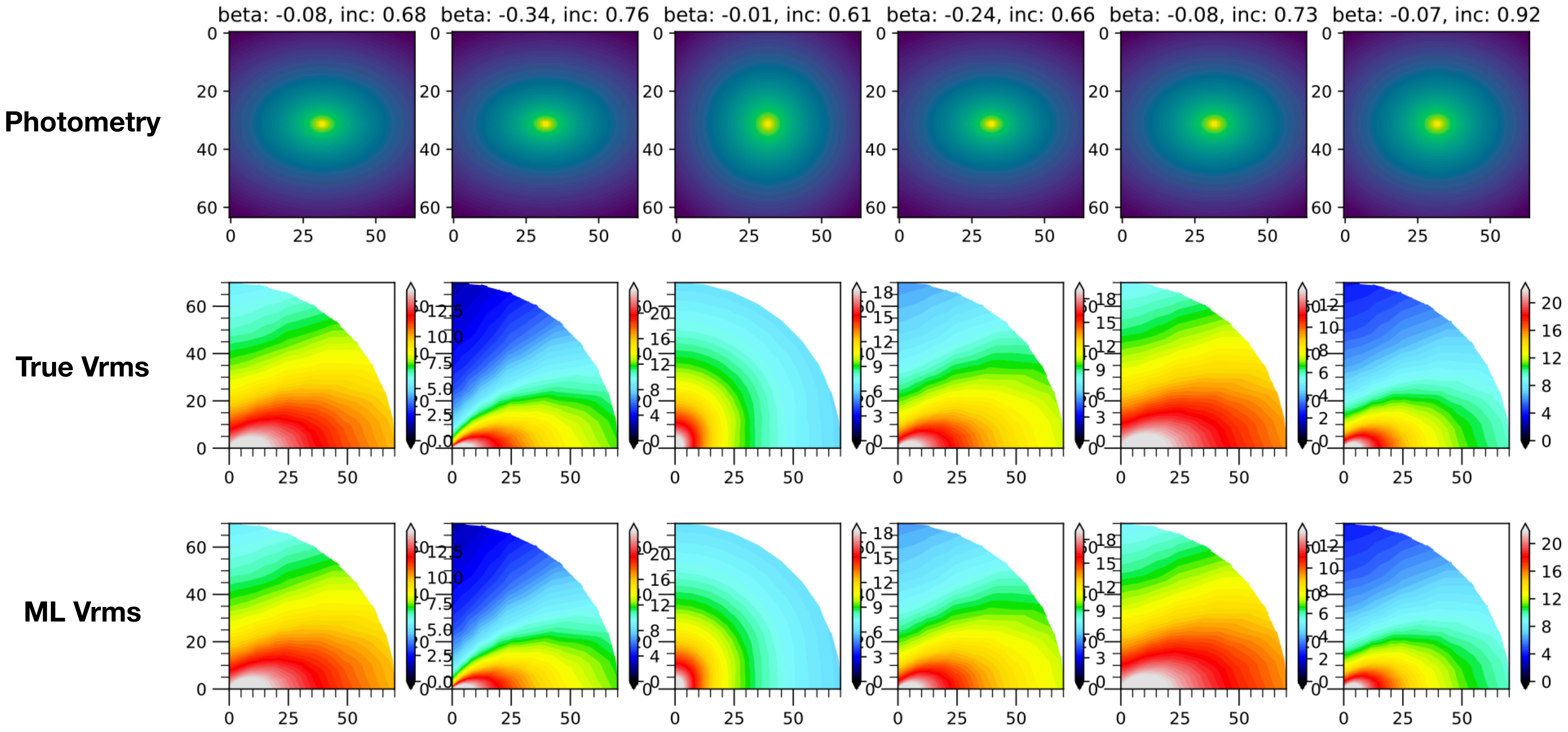
$$n_{\rm ser} \in [2, 4]$$
$$R_{\rm ser} \in [5", 20"]$$
$$q_\ell \in [0.6, 1]$$
$$\beta_{\rm z} \in [-0.4, 0.4]$$
$$q_{\rm m} \in [0.6, 1]$$
$$\mathrm{inc} \in [\arccos(0.6), 90°]$$
$$\theta \in [0°, 180°]$$

- 10,000 mock Galaxies
  - 8000 training
  - 2000 test
- Randomly sampled parameters while enforcing physical solutions
  - $\beta \leq 0.7\epsilon_{\rm intr}$ (See Wang et al 2021)

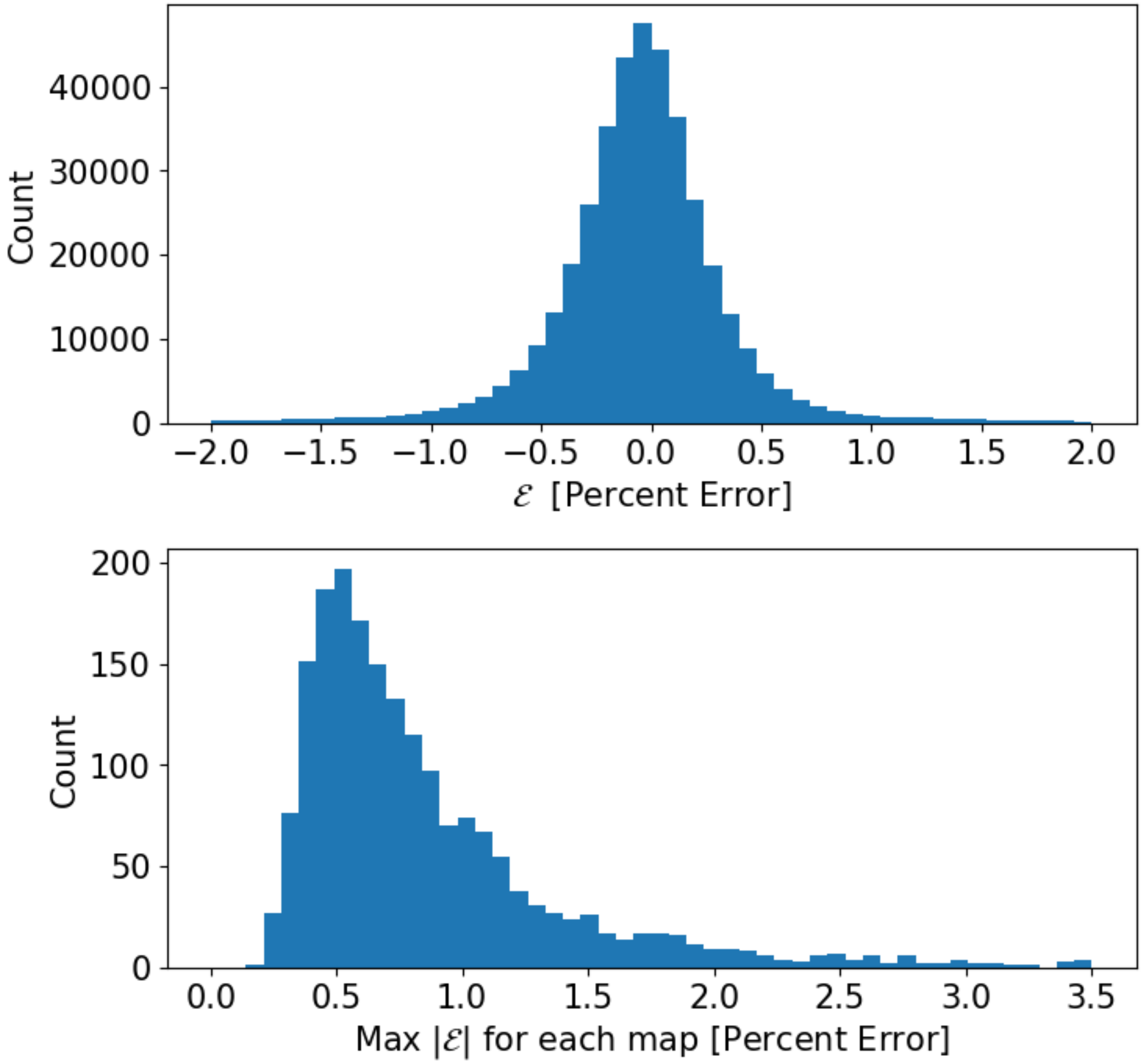
ML Modelling Pipeline



Outcomes: Maps



Outcomes: Accuracy

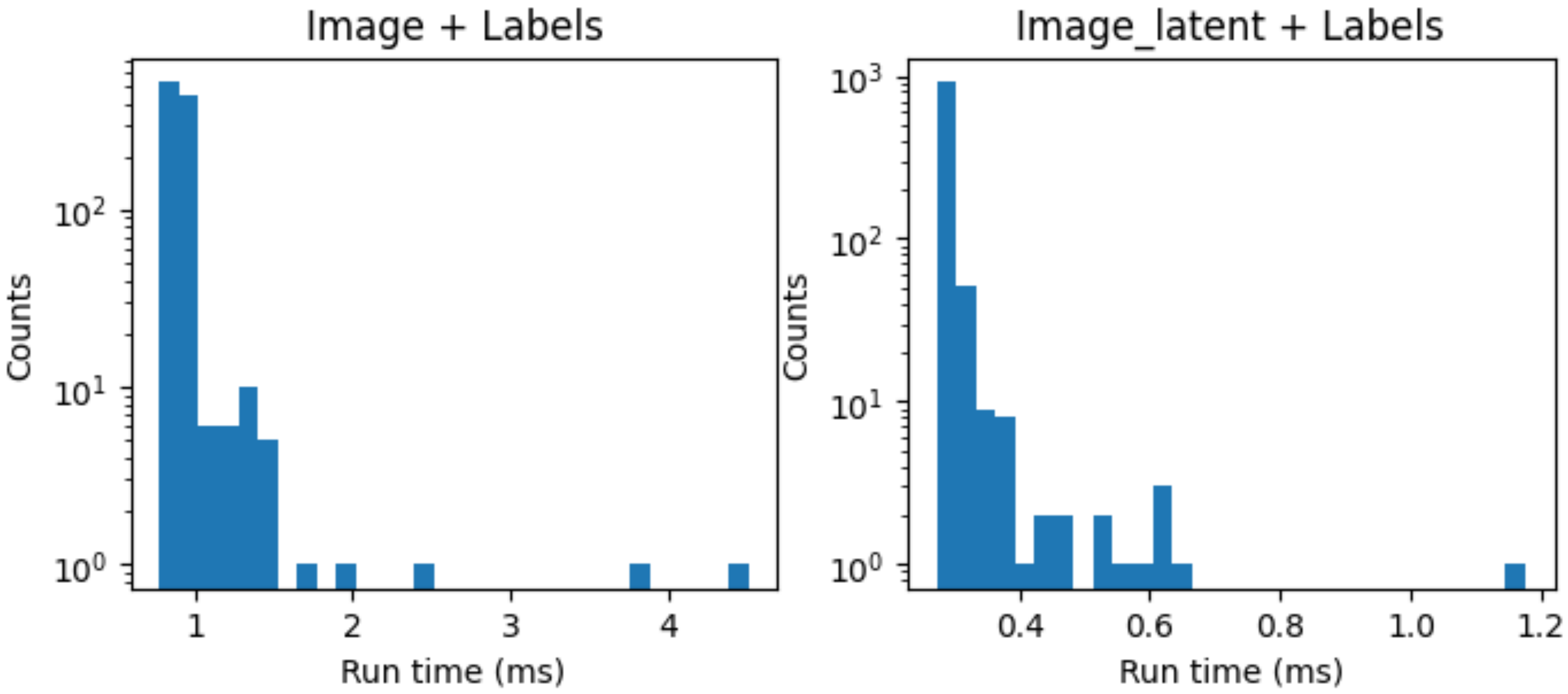


$$\mathcal{E} = \frac{\text{Model} - \text{Truth}}{\text{Truth}}$$

- The test set maps are accurate to within 1% and ninety percent of the maps have a max error less than 2%
- Real observations are accurate to within 6-7 km/s

Outcomes: Timing

- Training set can be generated in ~3 hours on a single CPU core
  - 2.5 days for Gomer et al 2023
- Model can be trained on a single NVIDIA V100 GPU in ~30 minutes
  - 3 days for Gomer et al 2023



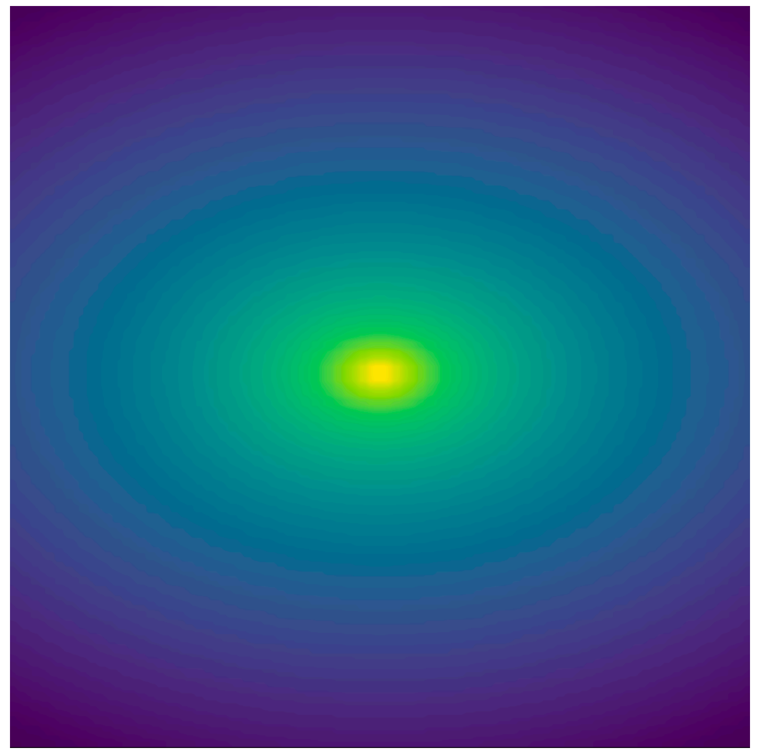
Almost 100x faster than (Gomer et al 2023) and on CPU

Key Questions

- ~~Can Machine Learning be used to Improve Existing Methods?~~
- Can Machine Learning be used to Create New Methods?

(Preliminary results!)

ML Modelling Pipeline

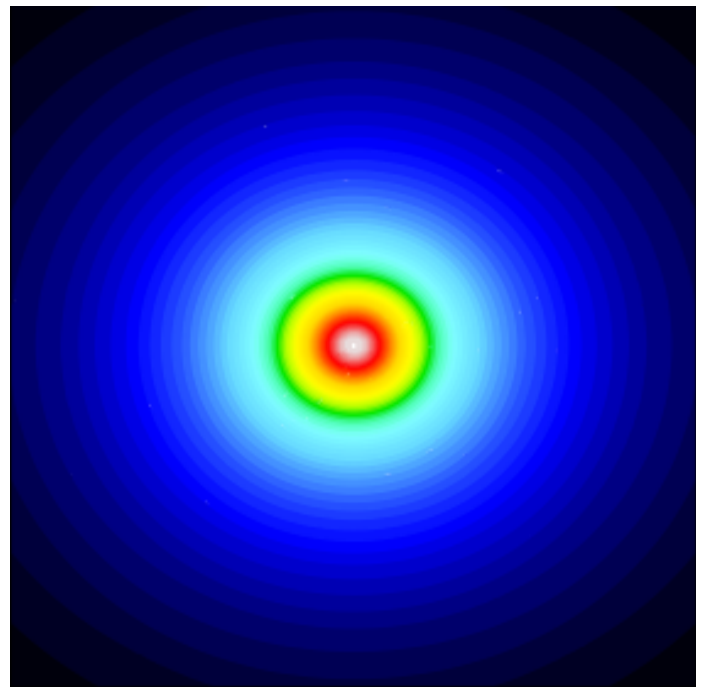


Photometry

Gravitational Potential

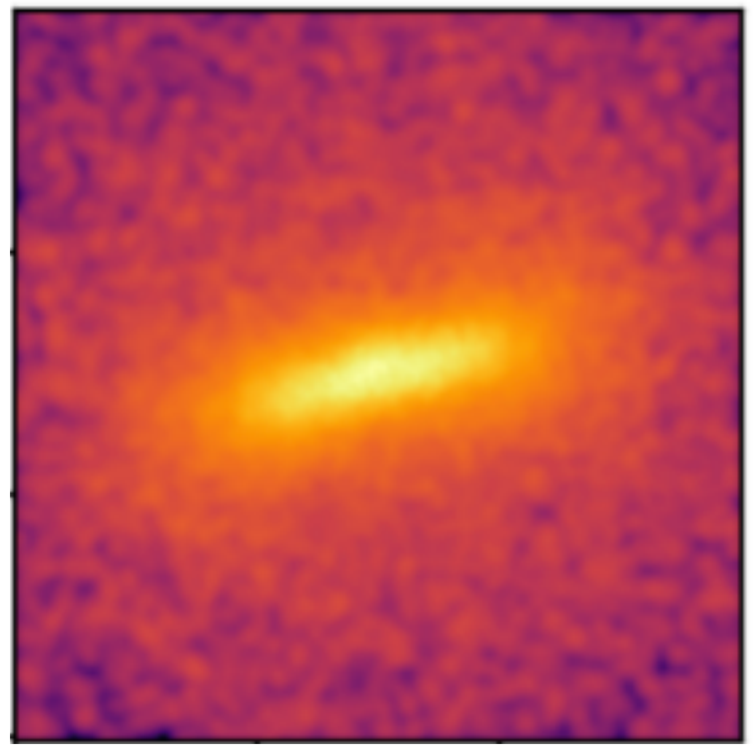
Free Parameters

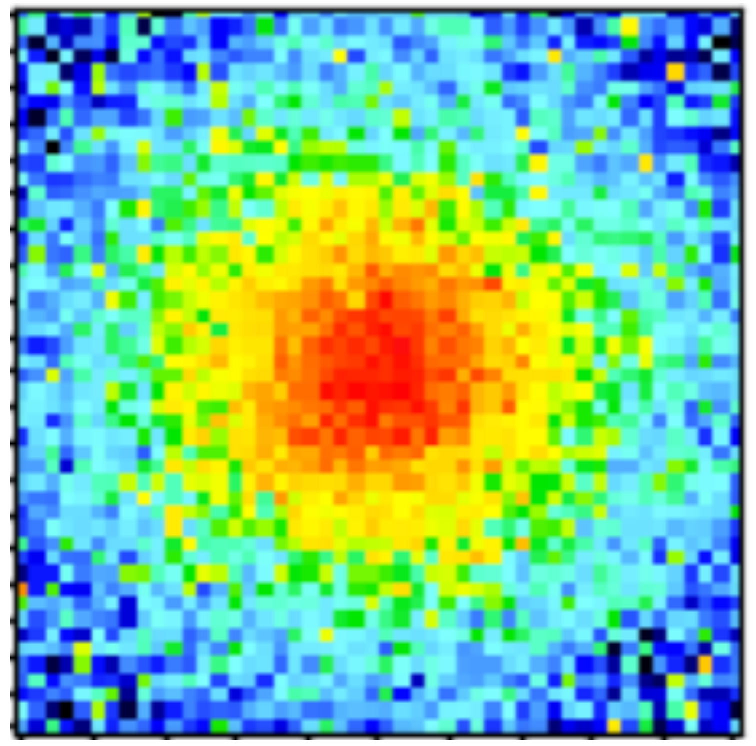
Machine Learning Algorithm



TNG100 Mock Observations

Training Set





Inc

$\beta(r < R_e)$

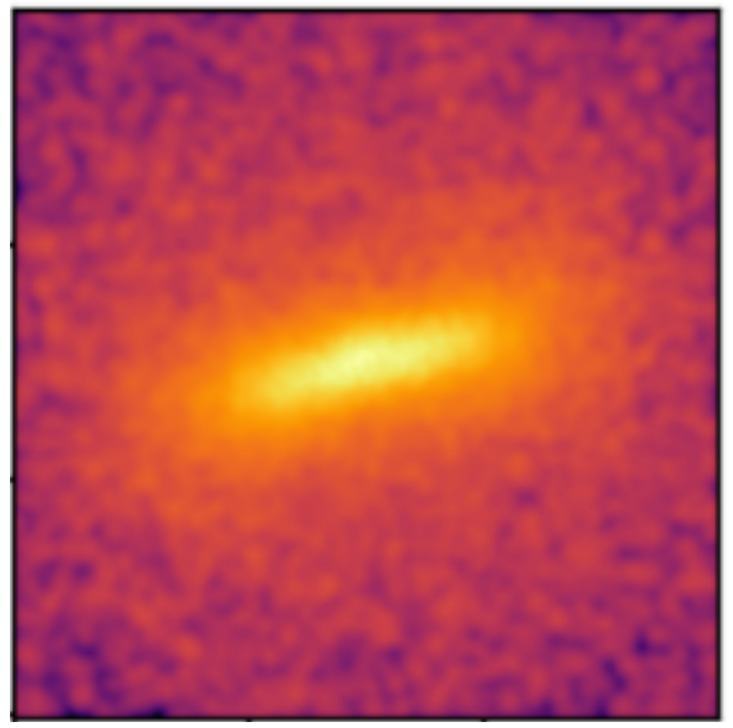
$M/L(r < R_e)$

$f_{\rm dm}(r < R_e),\ \gamma,\ r_s$

$\theta \in [0^\circ, 180^\circ]$

- TNG100 Mock Observations
  - 1.4e6 solar mass resolution for stars
  - Stellar mass within 30 kpc > 5e9 solar mass
  - Total mass < 1e14 solar mass
- ~5000 galaxies
- Mock observations along x,y, and z axis
- Field of view out to 3 Re

ML Modelling Pipeline



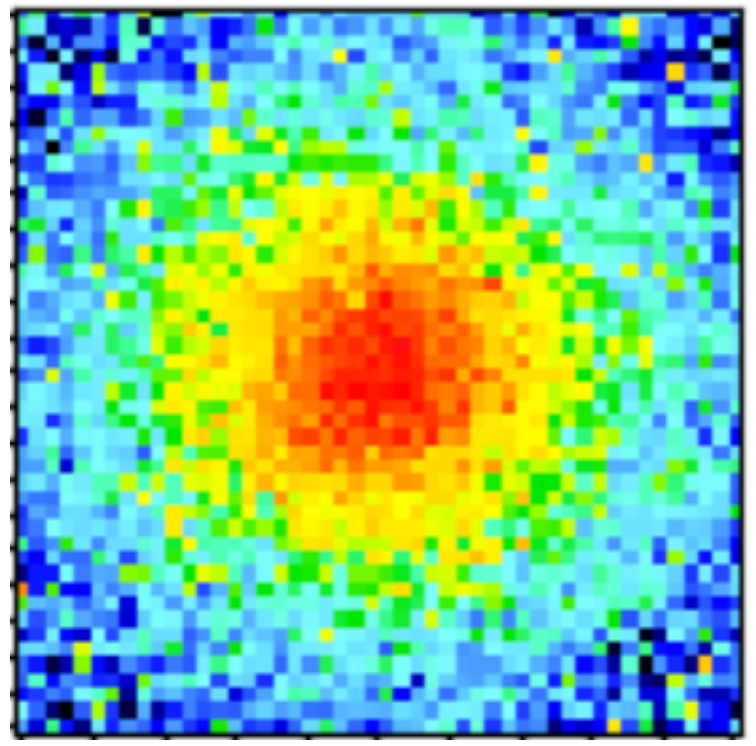
Photometry

CNN

Fully-Connected Layers

β, Inc, M/L,

$f_{\rm dm}$, γ, r_s, θ



Use non-standard loss to mask empty pixels

TNG100 Results

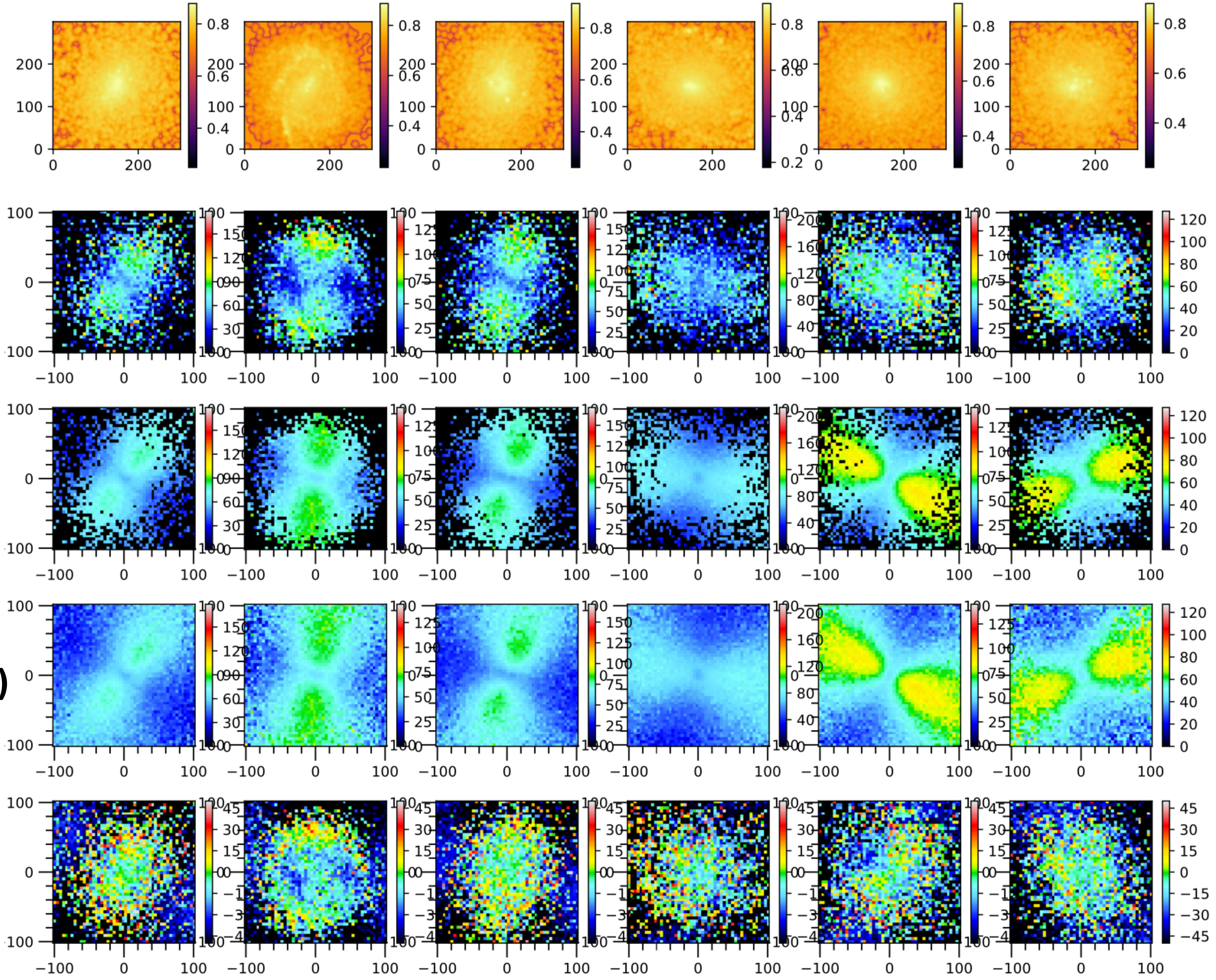
TNG100 Photometry

TNG100 Vrms

ML Vrms (with mask)

ML Vrms (without mask)

Residuals



Conclusions

- Simple machine learning algorithms can sometimes outperform complicated ones
- Close to 10,000 data points are required in order to machine learn dynamical models (less with more complicated models)
- Machine Learned dynamical models have the potential to outperform existing methods