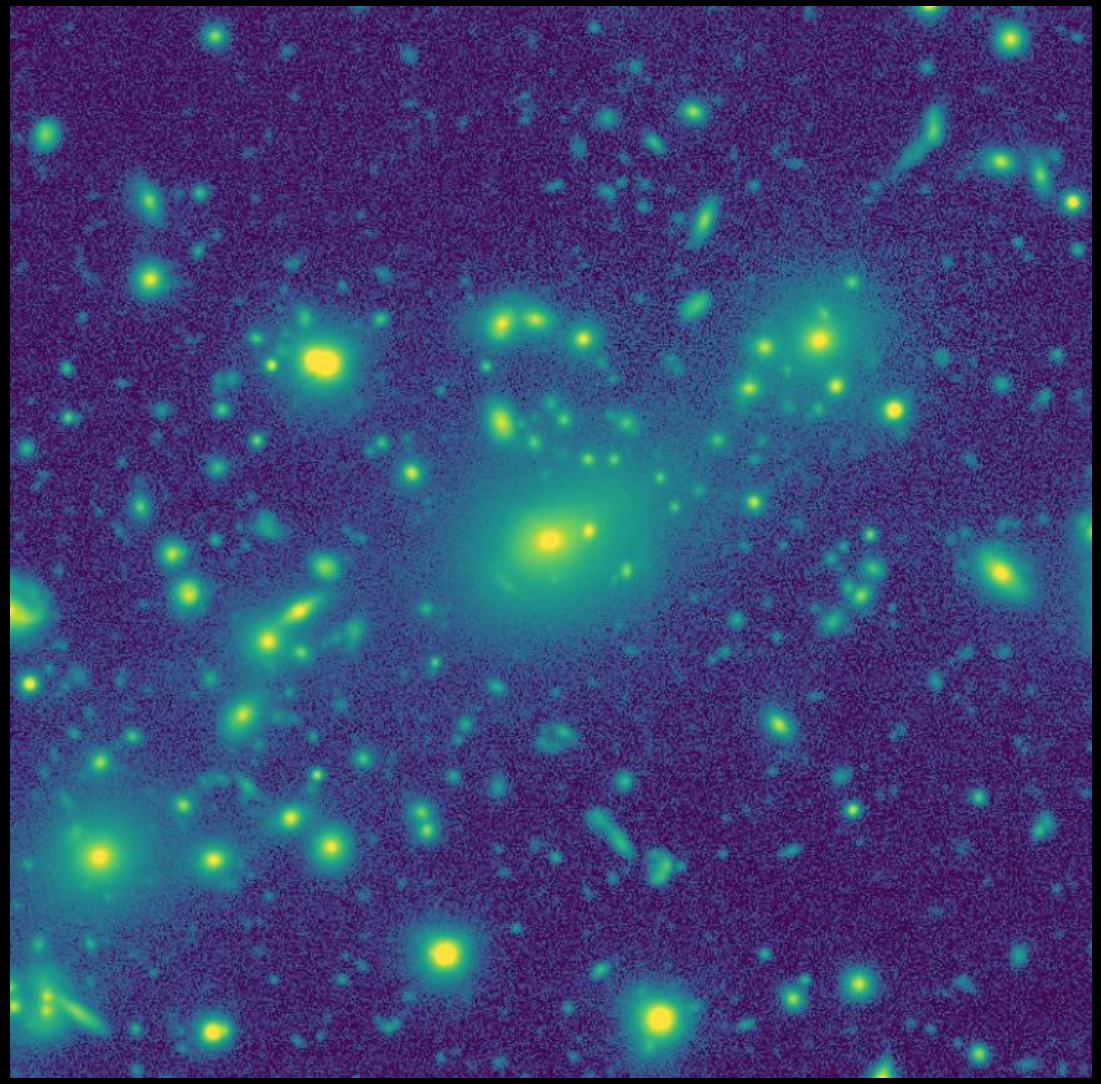
Measuring intracluster light with machine learning

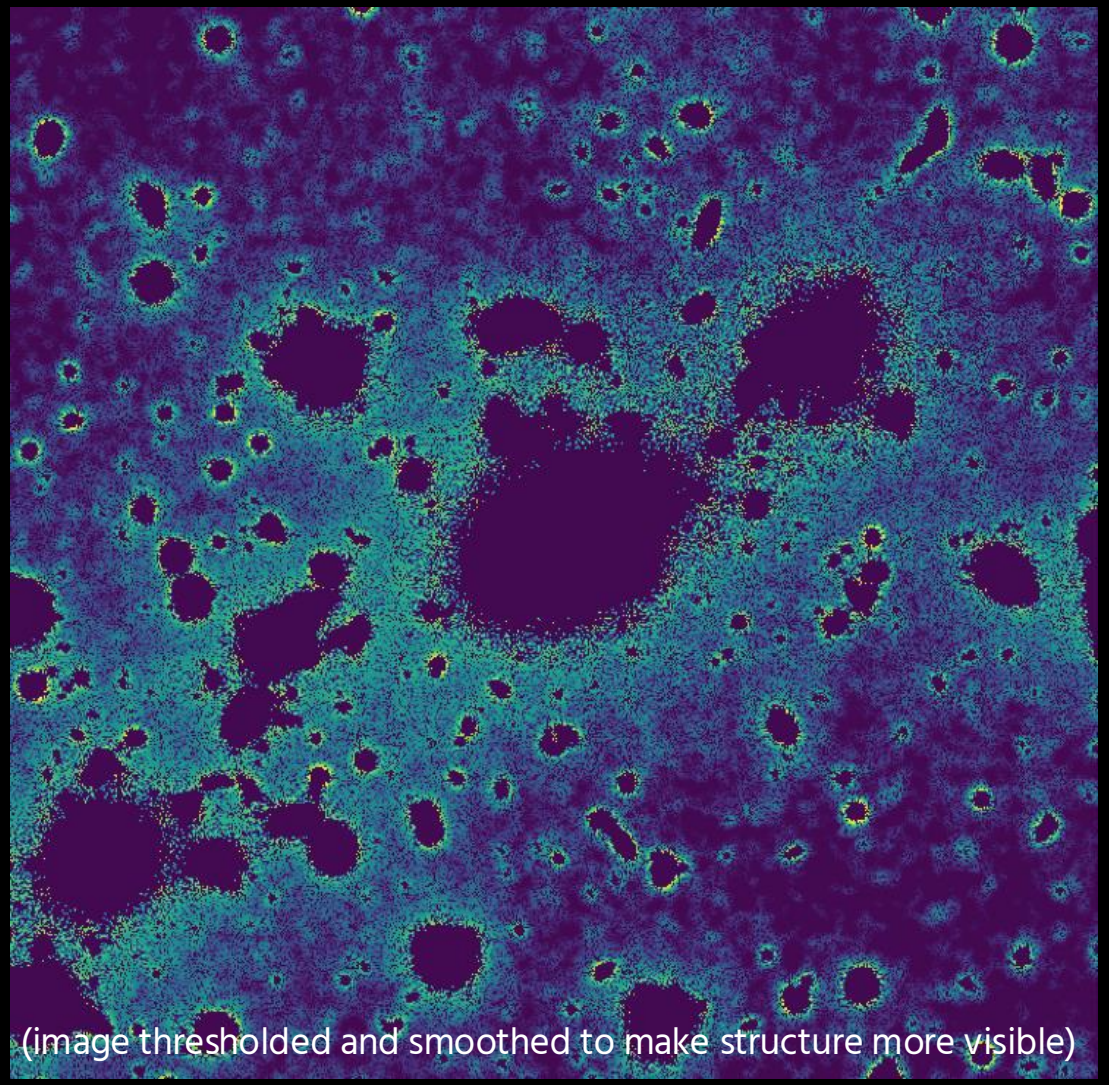
Louisa Canepa,

Sarah Brough, Francois Lanusse, Mireia Montes, Nina Hatch

What is ICL?



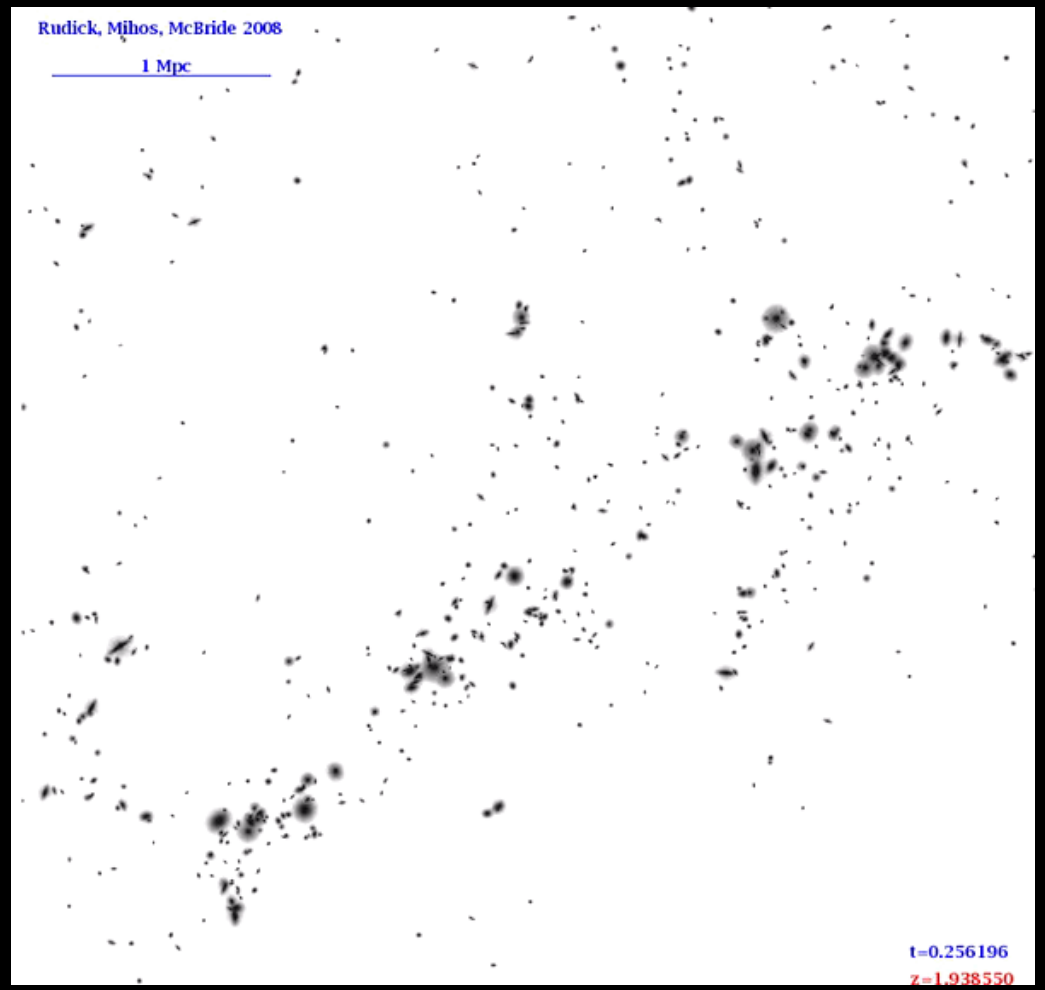
What is ICL?



(image thresholded and smoothed to make structure more visible)

- Diffuse, low surface brightness light in the centres of galaxy clusters
- Made up of stars that are not bound to any particular galaxy

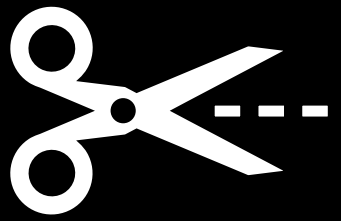
But what is it?



- Produced by tidal stripping and merging of galaxies
- Traces the history of interactions within the cluster (e.g. Rudick+11)
- Despite this, there's a lot we don't know...
  - Progenitors and formation mechanisms?
  - Relationship with cluster properties?

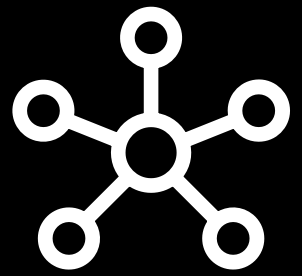
How do we measure it?

- Measure the ICL fraction
- How to separate the ICL from the BCG?
- Not standardised and generally hard to scale :(

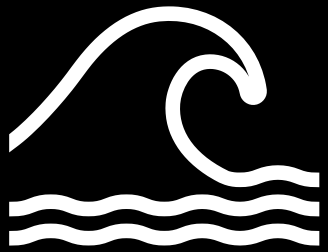


Surface brightness cut
(e.g. Montes+21)

Composite model
(e.g. Martinez-Lombilla+23)



Multi-galaxy fitting
(e.g. Jimenez-Teja+16)



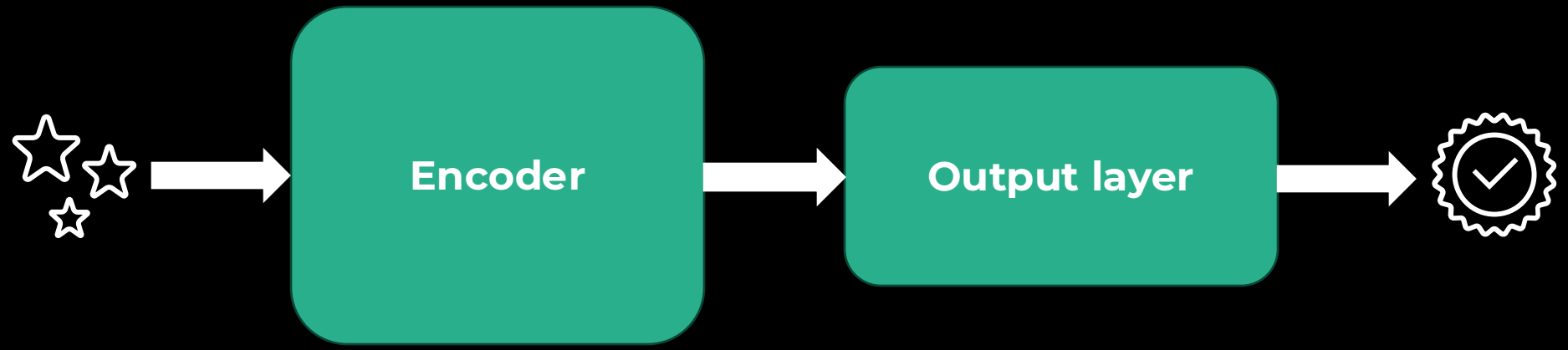
Wavelet decomposition
(e.g. Ellien+21)

… + machine learning?

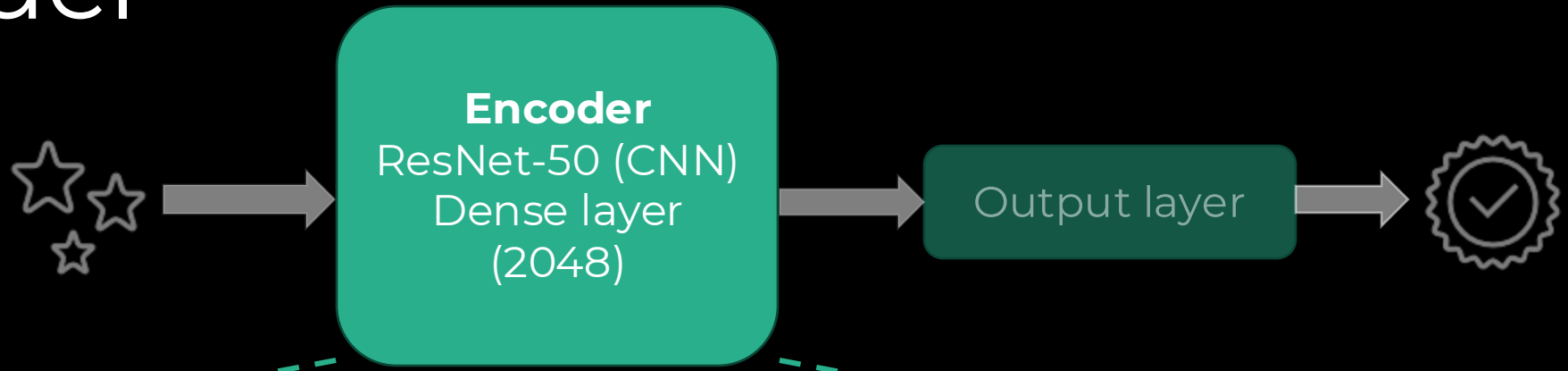
Why machine learning?

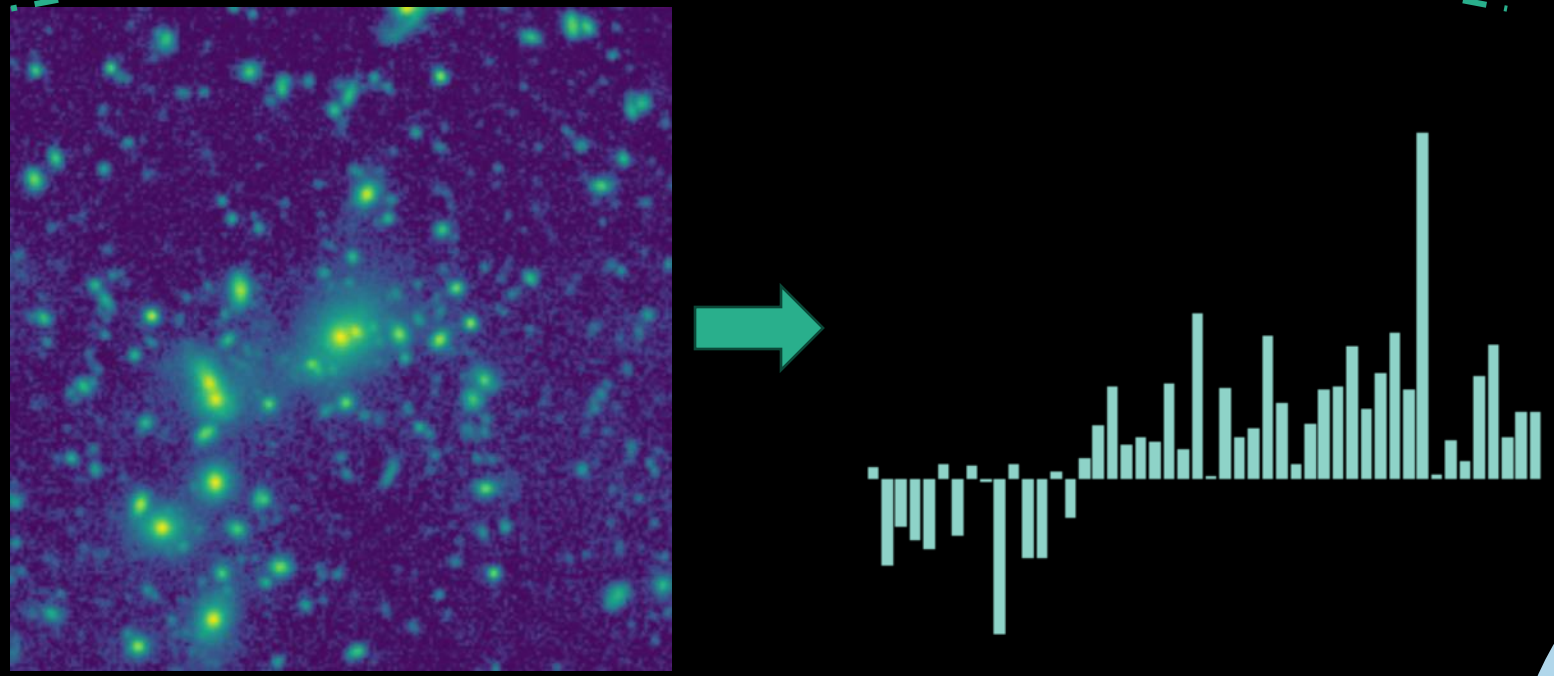
- Big data is coming, and we have better things to do than ultra-repetitive measurement tasks
- Machine learning is an option that could help
  - High throughput
  - Flexible
  - Can be applied to many problems with some understanding of its capabilities
- But there are some hurdles to applying ML to astronomy

The Model

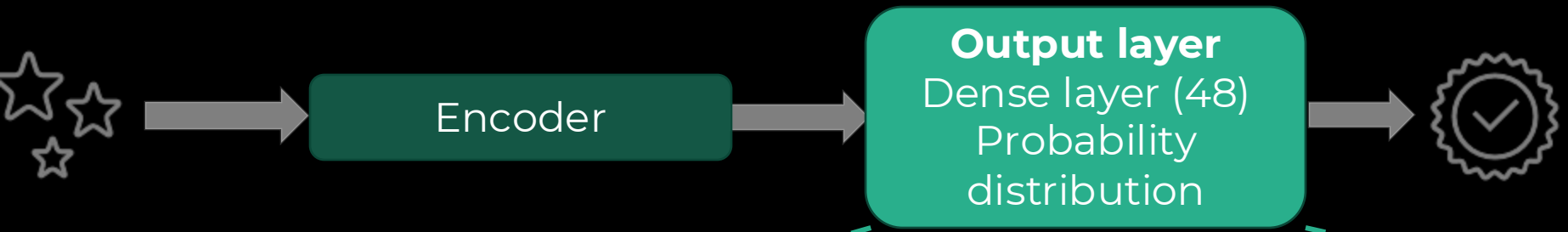


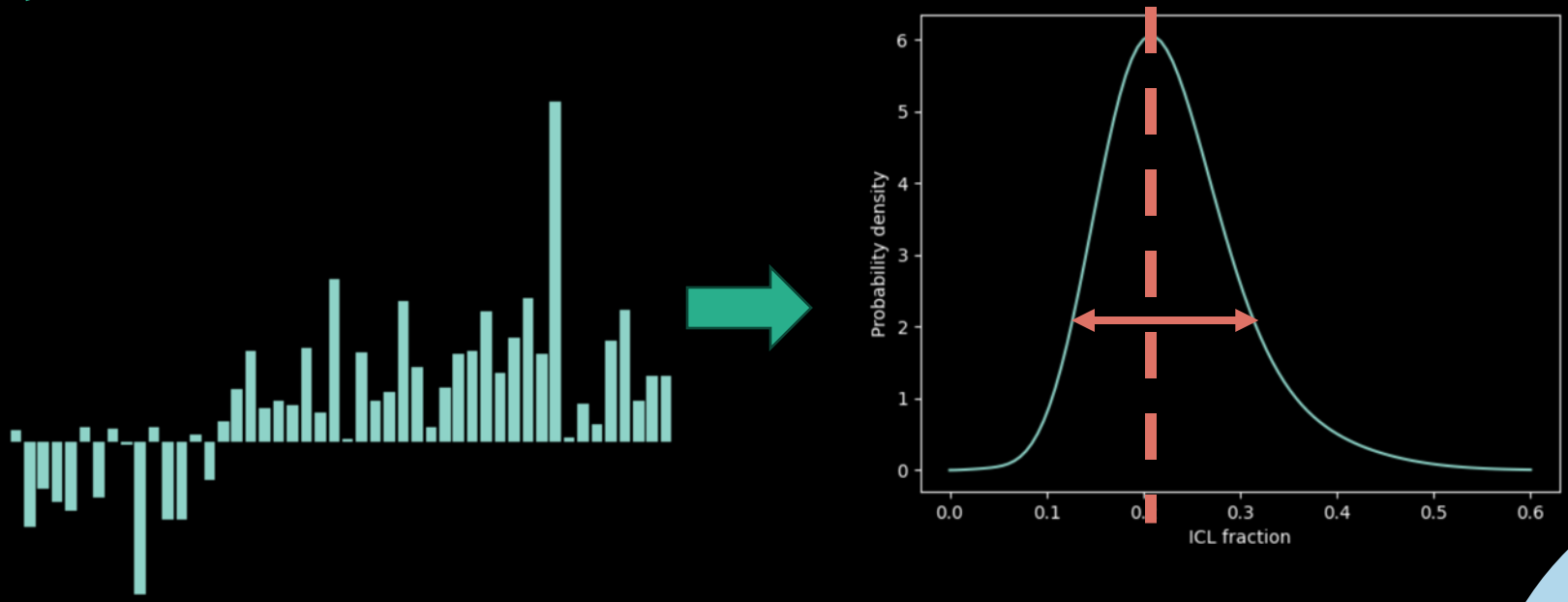
The Model





The Model

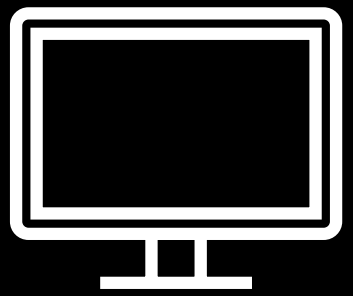

Encoder
Output layer
Dense layer (48)
Probability distribution


Probability density
ICL fraction
0.0
0.1
0.3
0.4
0.5
0.6
0
1
2
3
4
5
6

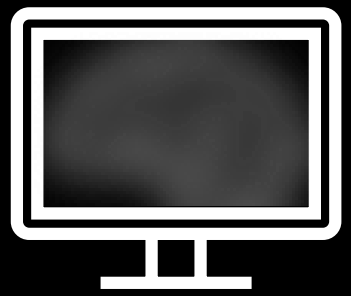
Training data

- We need a LOT of training data
- Not feasible to manually measure thousands of clusters
- Use transfer learning!

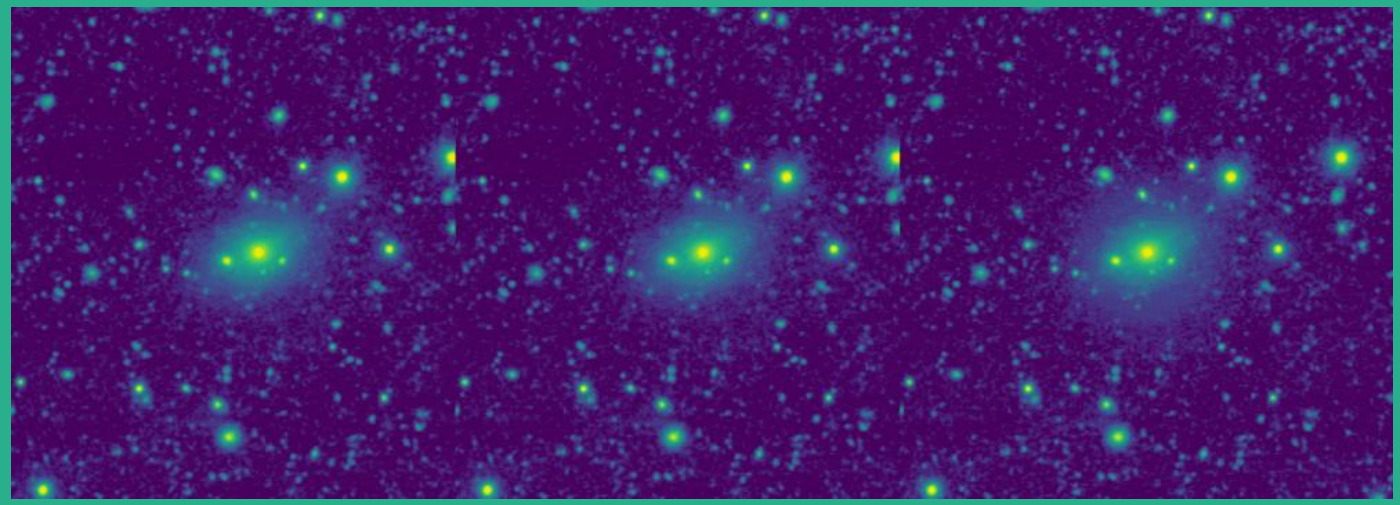
Training data



Training data



Large dataset
(50 000 examples)

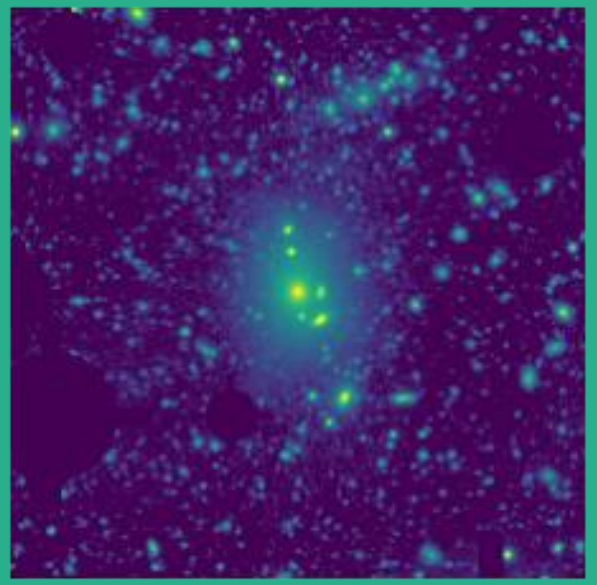


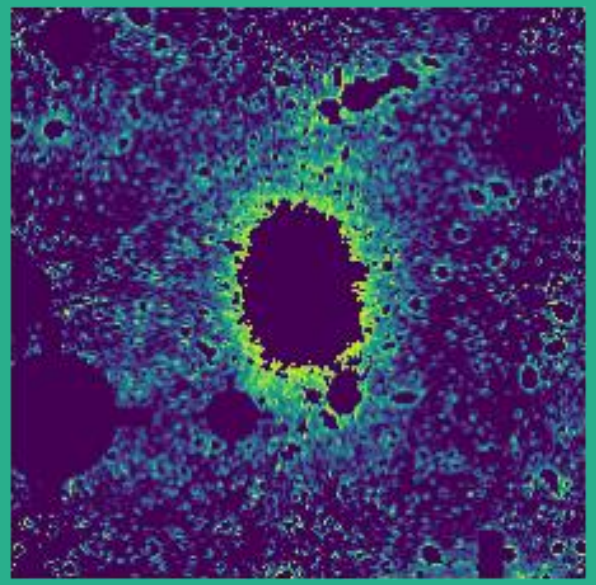
This data will look a bit different to real data, and measurements will be sub-optimal

Training data

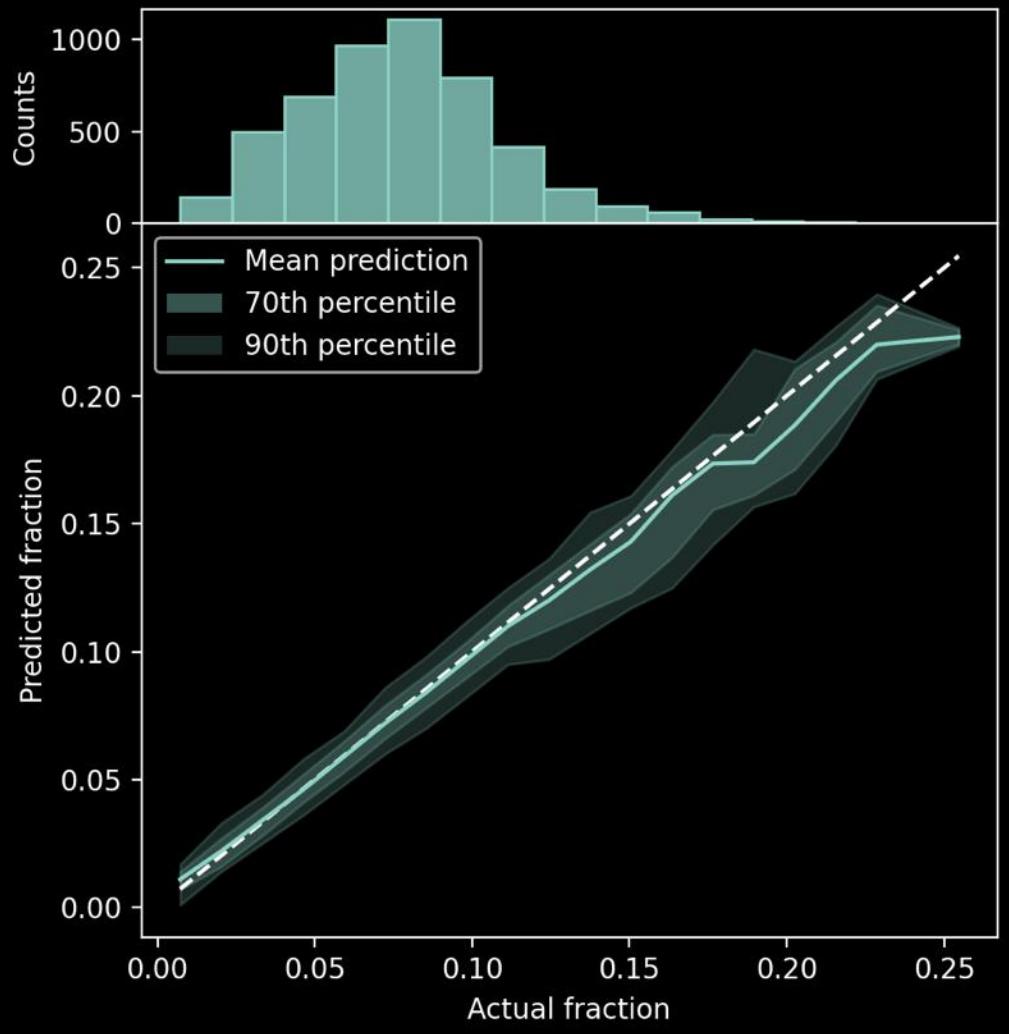
Finetuning dataset
(~100 examples)

Manually measure the real, deep cluster images



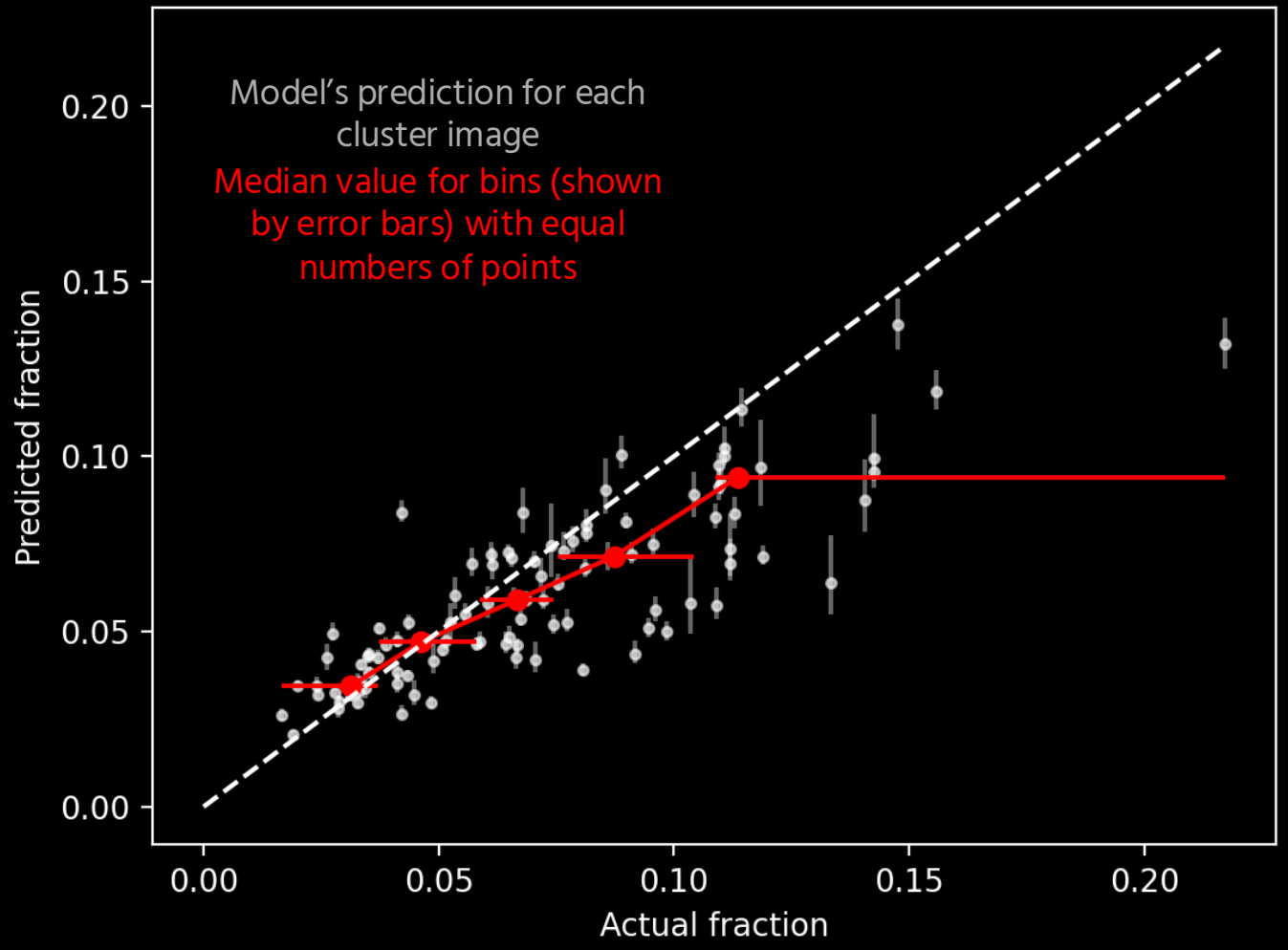


Results – training dataset



Mean Absolute Error (MAE) = 0.00511

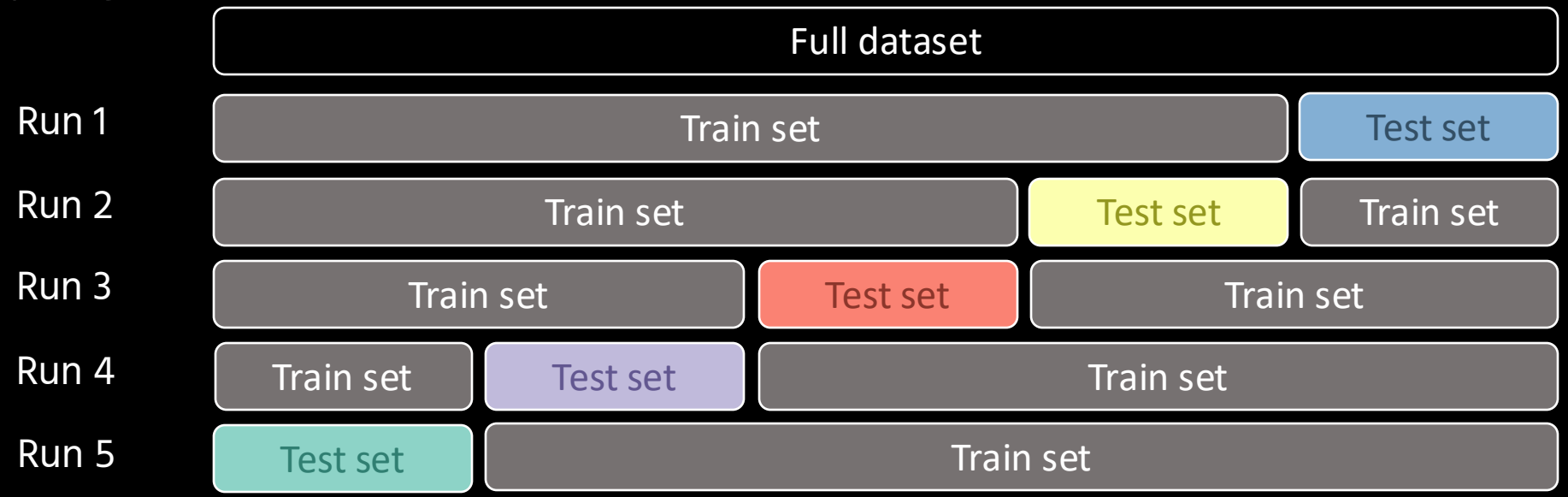
Results – before finetuning



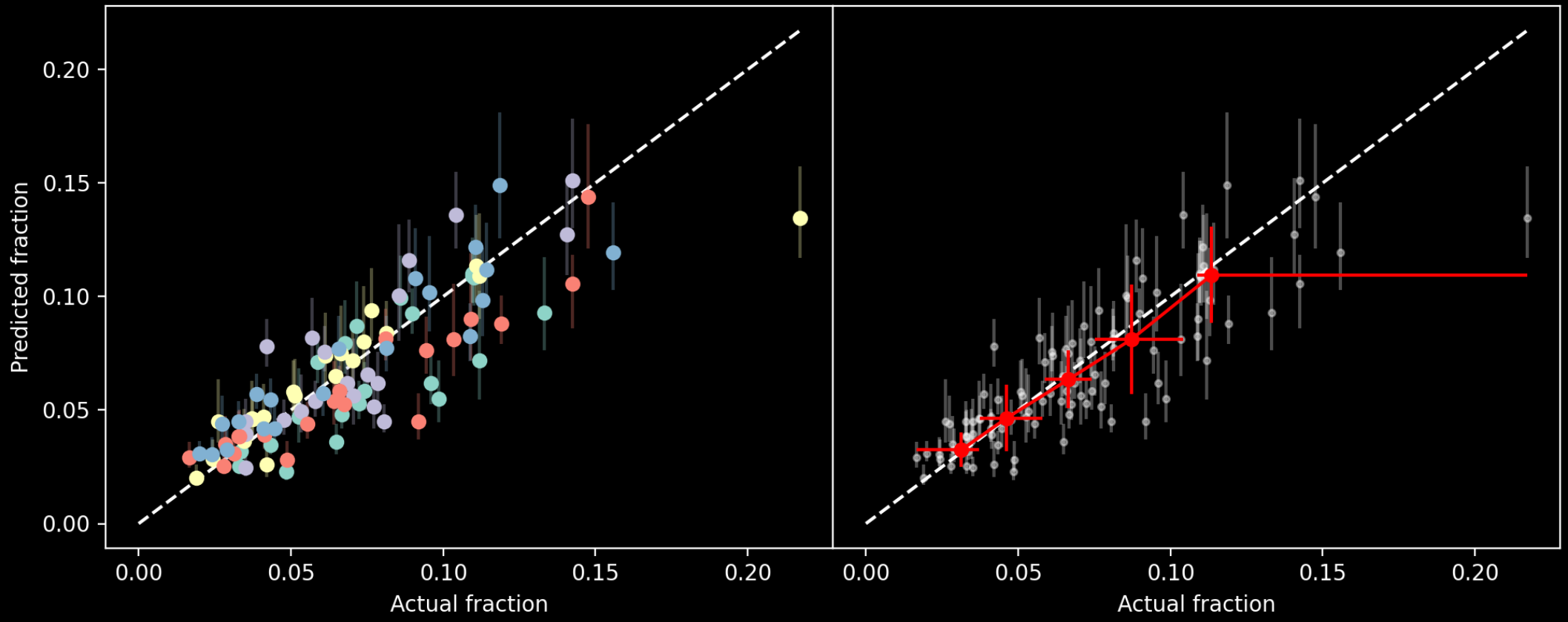
MAE = 0.0163

K-fold cross-validation

- We only have 101 finetuning samples, or ~20 samples for testing
- Train the model k times, reserving different data for testing each time



Results – after finetuning

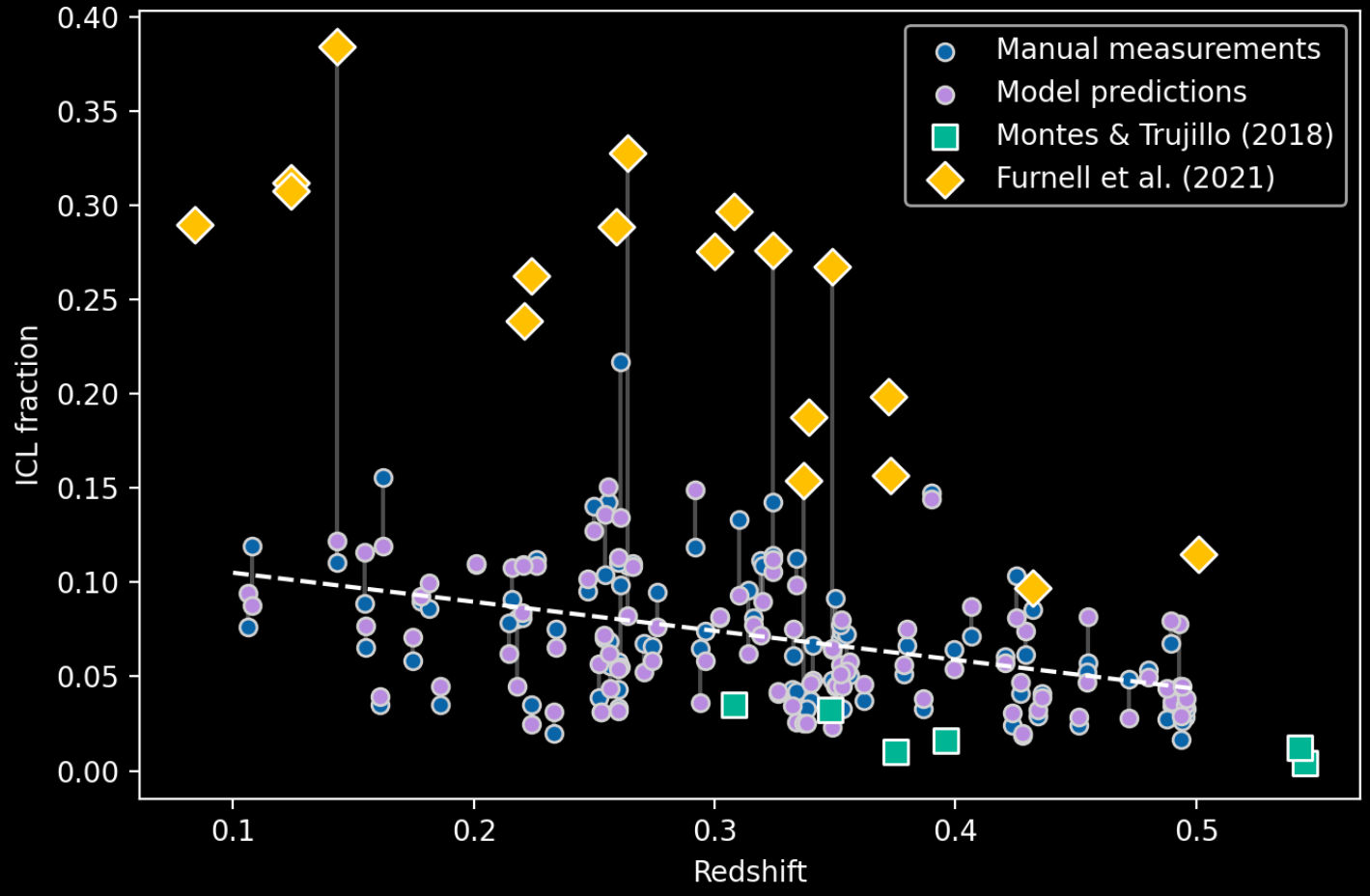


MAE = 0.0139

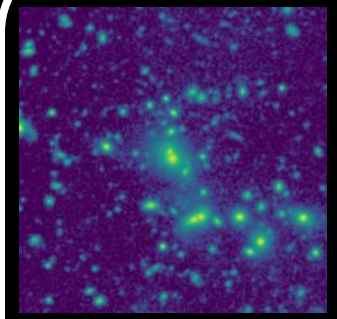
Larger samples!

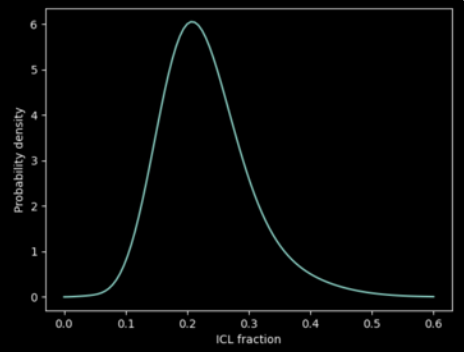
- Now we’re closer to taking advantage of LSST amounts of data
- 500 samples in seconds on a GPU, or a couple of minutes on CPU
- This allows us to easily collate much larger samples (and expand easily!)

Larger samples!

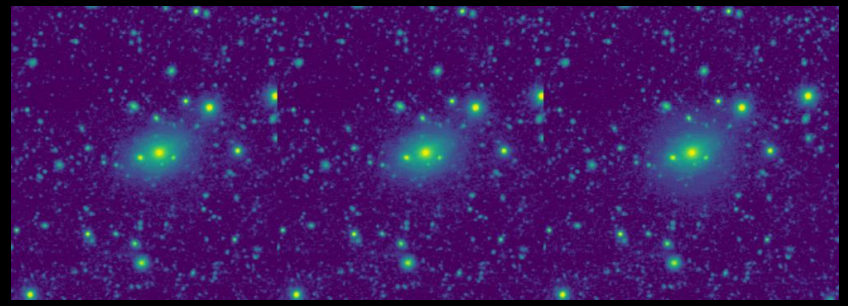


Summary



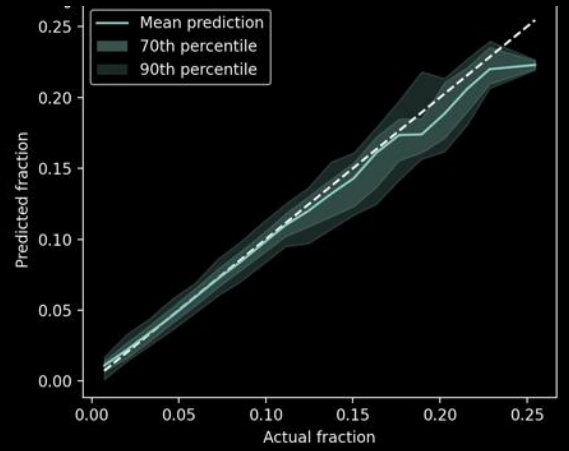


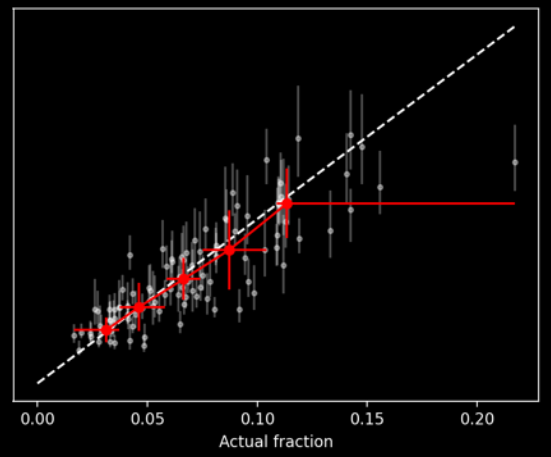
An image regression ML model



Trained on mostly artificial data, and only a small amount of real data

Accurately learns to predicts ICL fraction in artificial and real data





Bring on the data!

Stay tuned for the paper!

l.canepa@unsw.edu.au

lpcan.github.io

lpcan