Machine learning / AI in astronomy

ICTP Physics Without Frontiers

Stephen Serjeant , Jane Bromley, Hugh Dickinson, Lynge Lauritsen, Ruby Pearce-Casey, James Pearson, Chris Sorrell, Josh Wilde, Laura Hunt

School of Physical Sciences & School of Computing and Communication

Wednesday 1st October 2025

Machine learning / AI in astronomy

ICTP Physics Without Frontiers

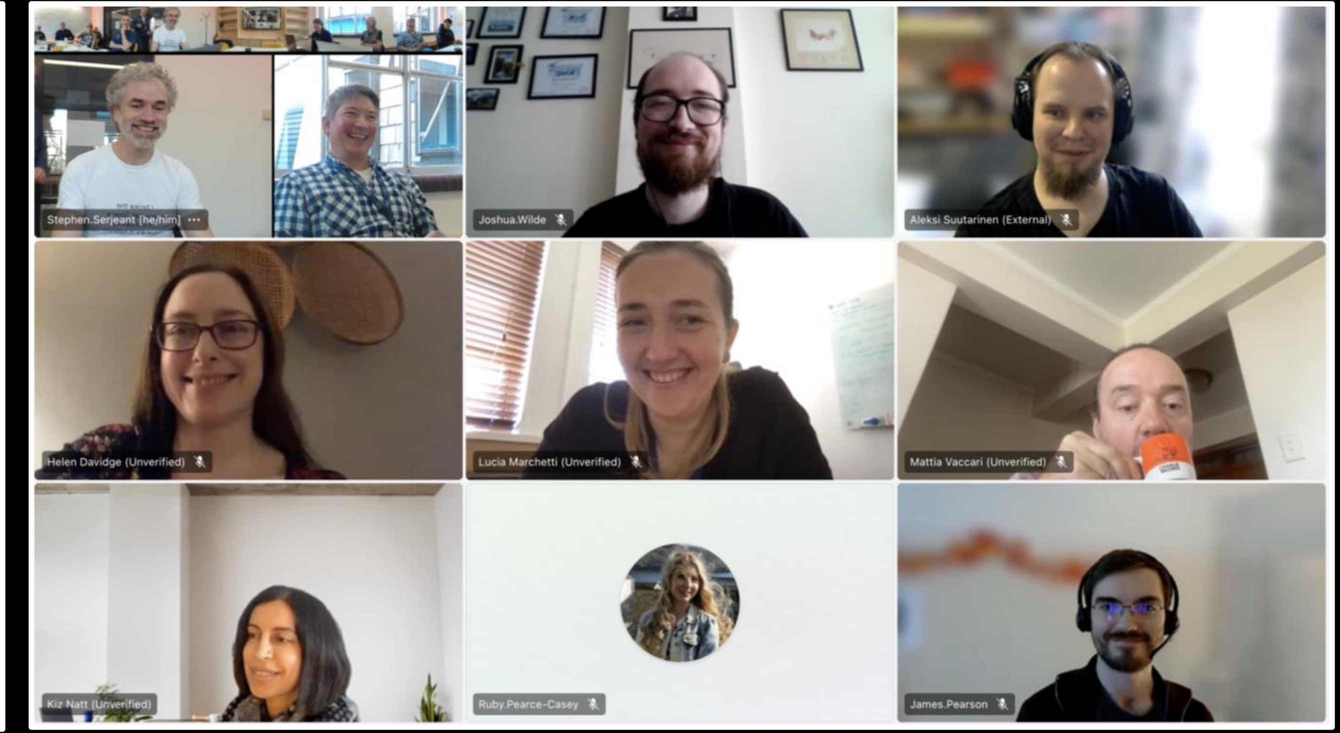
Stephen Serjeant, Jane Bromley, Hugh Dickinson, Lynge Lauritsen, Ruby Pearce-Casey, James Pearson, Chris Sorrell, Josh Wilde, Laura Hunt

School of Physical Sciences & School of Computing and Communication

Wednesday 1st October 2025

1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	Citizen Science, Galaxy Zoo and Zoobot
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	Infrared gravitational lenses and deconvolution
6	Other astro ML, and Conclusions

The Open University


Stephen.Serjeant (he/him)
Joshua.Wilde
Aleksi Suutarinen (External)
Helen Davidge (Unverified)
Lucia Marchetti (Unverified)
Mattia Vaccari (Unverified)
Kiz Natt (Unverified)
Ruby.Pearce-Casey
James.Pearson

1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	Citizen Science, Galaxy Zoo and Zoobot
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	Infrared gravitational lenses and deconvolution
6	Other astro ML, and Conclusions

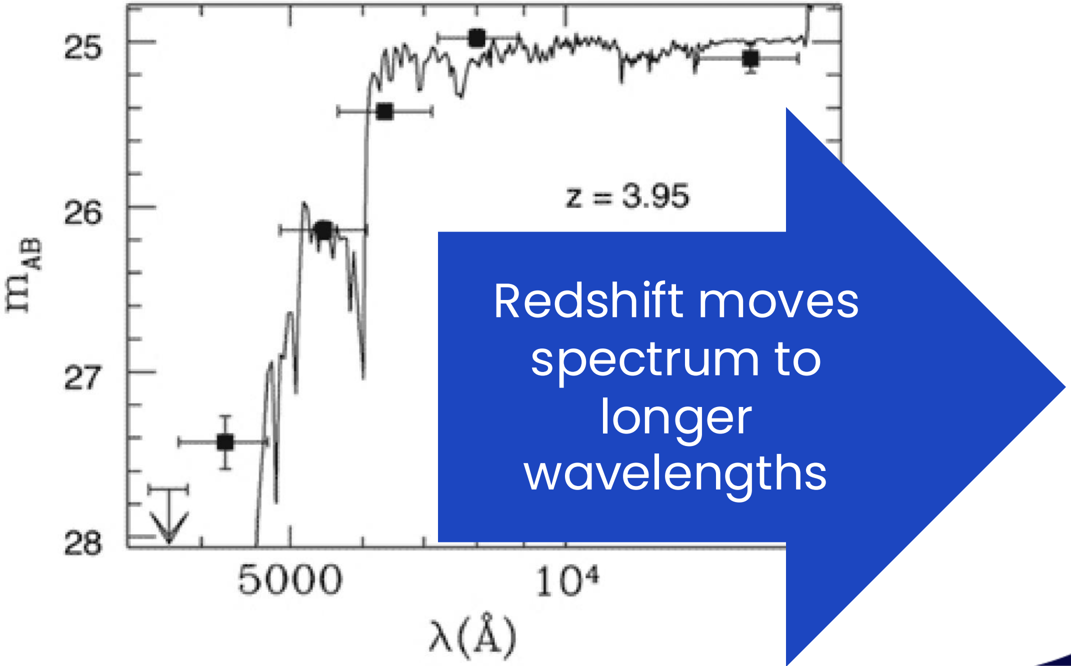
The Open University

My own first forays into machine learning (2005-7)

Estimating Photometric Redshifts Using Genetic Algorithms

Stephen Serjeant
Physical Sciences
Open University,
Milton Keynes, Bucks,
MK7 6AA, UK

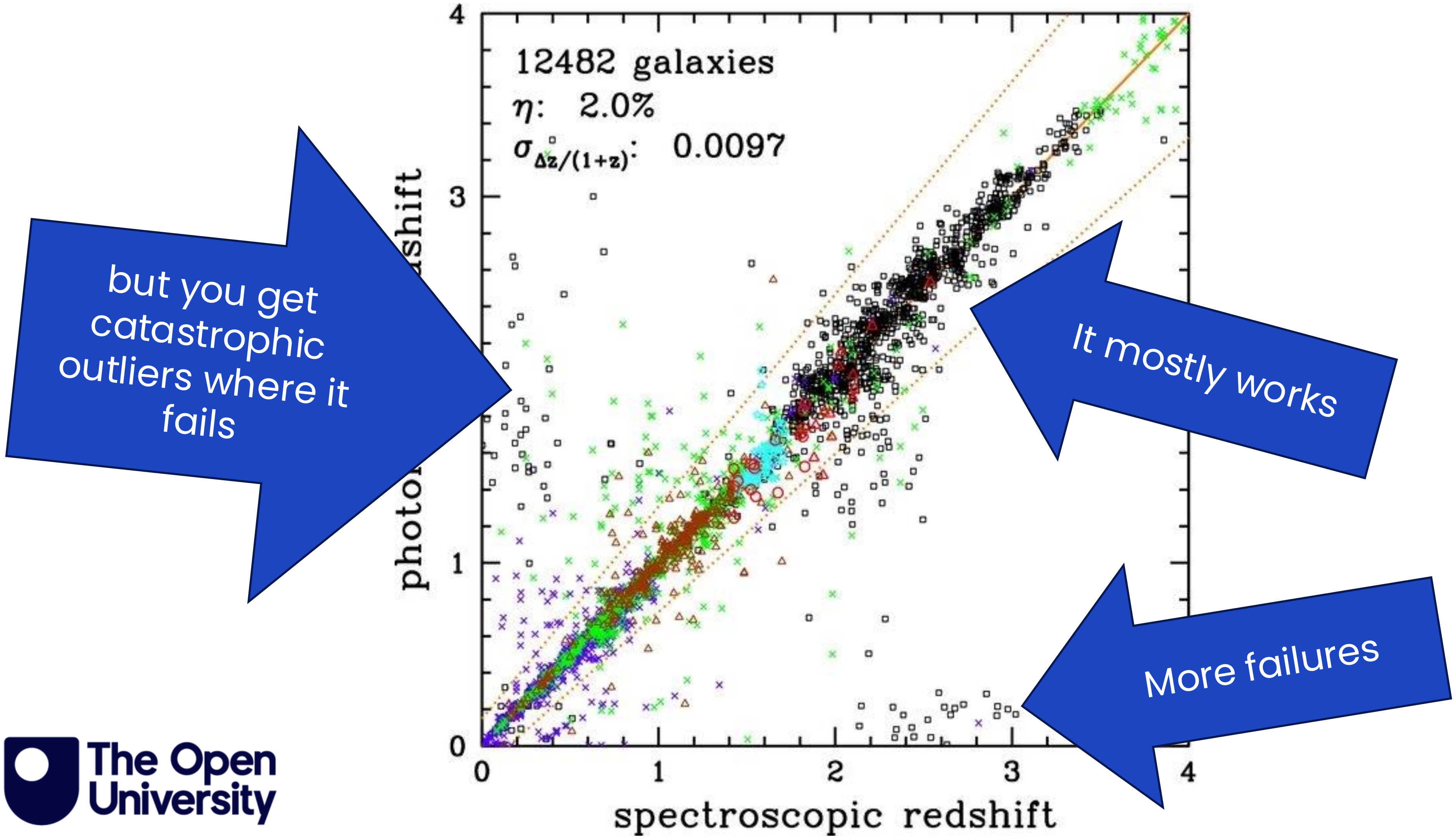
The problem: measure the redshift (from cosmic expansion) without getting a spectrum, using just the broad-band colours



The Open University

(1) Applications and Innovations in Intelligent Systems XIV. ISBN 978-1-84628-665-0. Springer-Verlag London, 2007, p. 75
(2) GECCO'06, July 8–12, 2006, Seattle, Washington, USA
ACM 1-59593-186-4/06/0007

My own first forays into machine learning (2005-7)



Scoville+13 Astrophysical Journal Supplement Series 206(1):3

The Open University

My own first forays into machine learning (2005-7)

Estimating Photometric Redshifts Using Genetic Algorithms

Nicholas Miles
Computer Science Dept
University of Kent,
Canterbury, Kent,
CT2 7NF, UK
nick@terado.co.uk

Alex Freitas
Computer Science Dept
University of Kent,
Canterbury, Kent,
CT2 7NF, UK
A.A.Freitas@kent.ac.uk

Stephen Serjeant
Physical Sciences
Open University,
Milton Keynes, Bucks,
MK7 6AA, UK

The idea: see if we can create a decision tree to find these failures

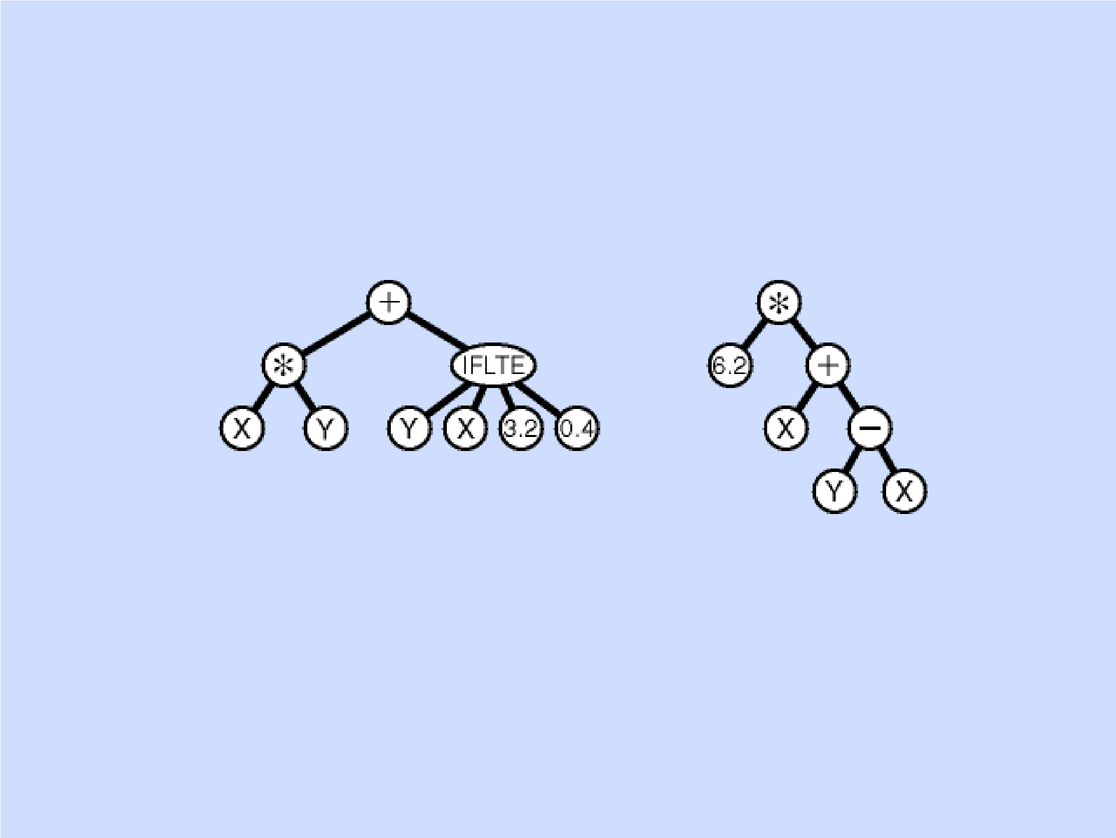
We could use more information than just the fluxes and colours, e.g. sizes, environments

(1) Applications and Innovations in Intelligent Systems XIV. ISBN 978-1-84628-665-0. Springer-Verlag London, 2007, p. 75
(2) GECCO'06, July 8–12, 2006, Seattle, Washington, USA ACM 1-59593-186-4/06/0007

The Open University

My own first forays into machine learning (2005-7)

The technology: genetic programming. Evolve a population of decision trees, using mutation and crossover. Find the best!



Source: Wikimedia commons

The Open University

My own first forays into machine learning (2005-7)

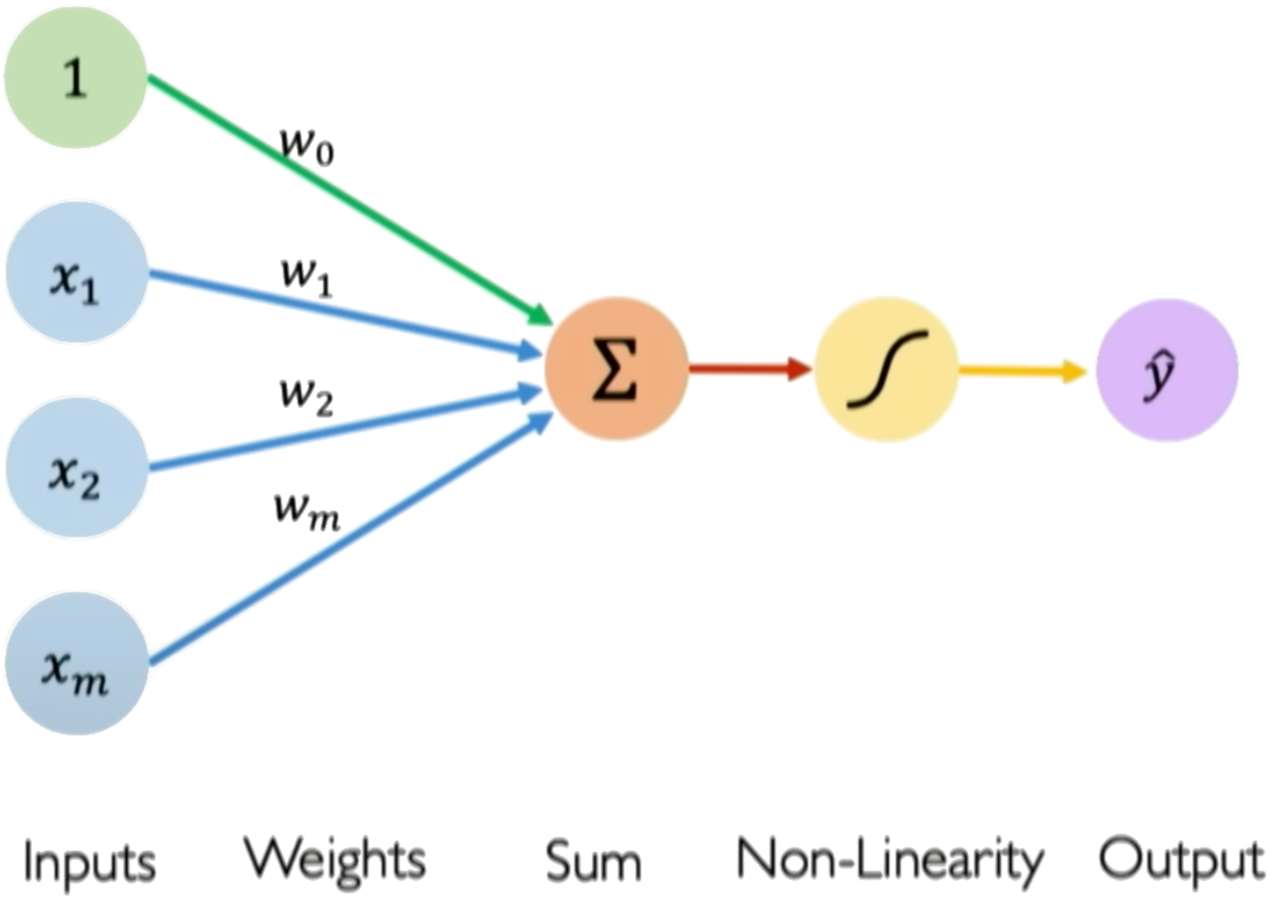
The problem: it's quite a 'spiky' surface in which to find the minimum. For example, swapping > for < completely changes the decision tree

Mitigated by having many in each generation

It worked, sort of (better than a benchmark computer science decision tree from 2005) but it was tricky to work with

The Open University

A better way to optimise: neural nets



Output

Linear combination of inputs

$$\hat{y} = g\left(w_0 + \sum_{i=1}^{m} x_i \, w_i \right)$$

Non-linear activation function

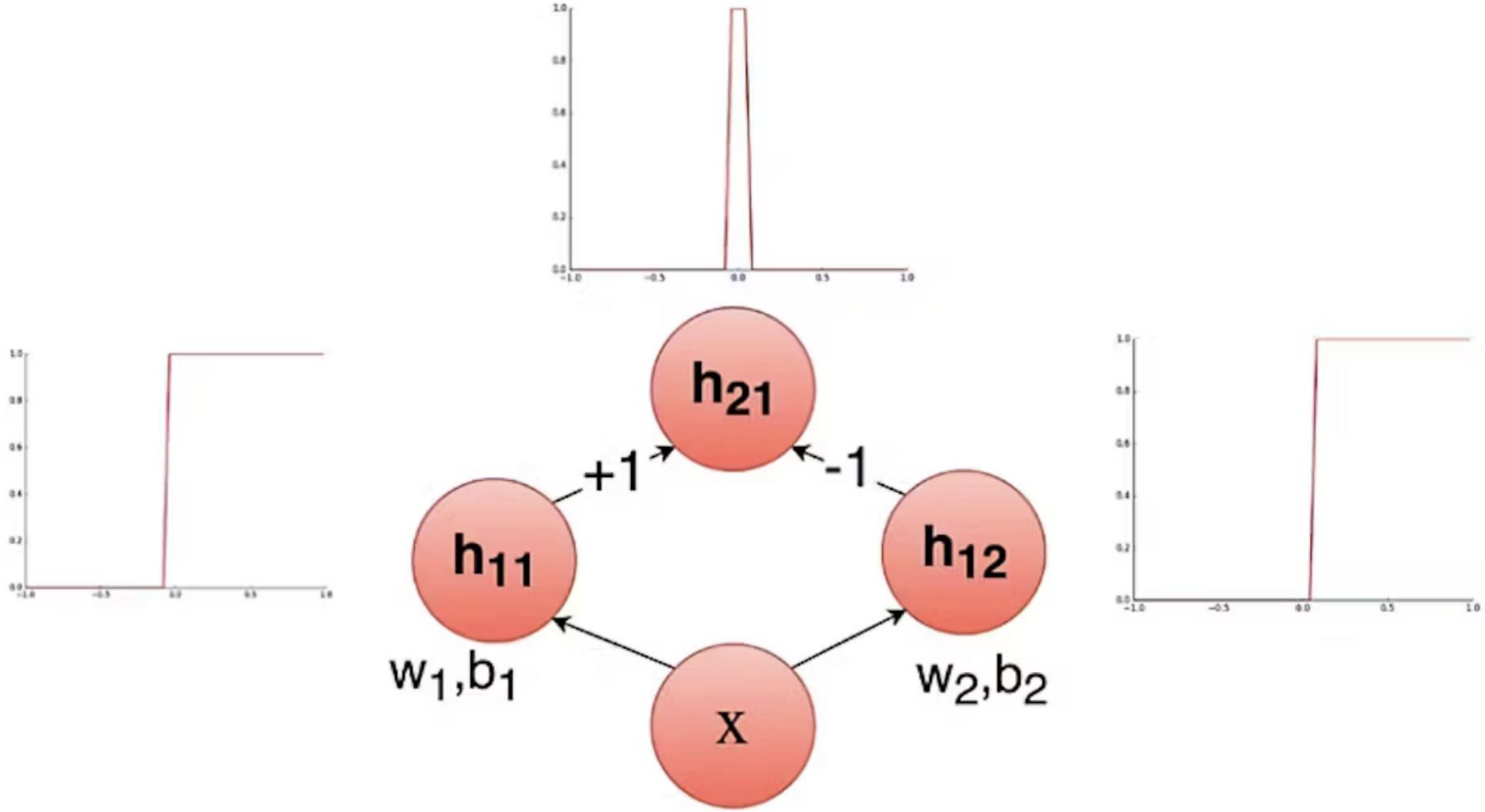
Bias

The Open University

Credits: Alexander Amini, MIT

A better way to optimise: neural nets

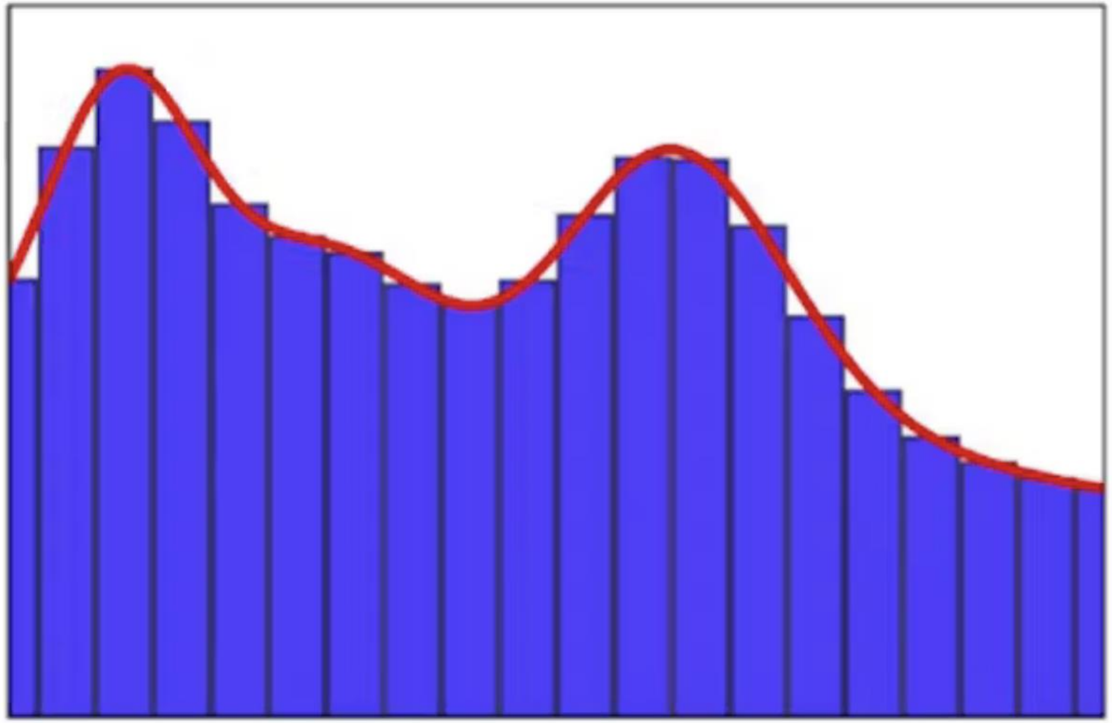
Imagine we only have one variable input, x, and let's use two neurons to output a box function

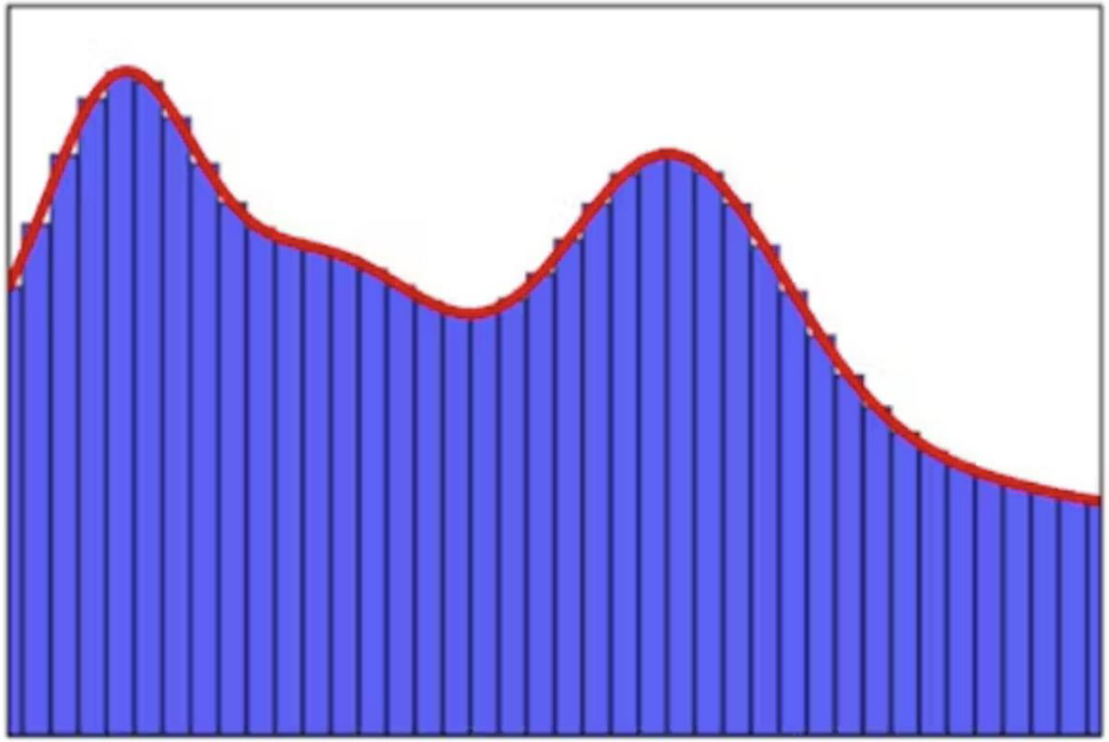


Credits: hackernoon.com

A better way to optimise: neural nets

With enough boxes we can approximate any one-dimensional function

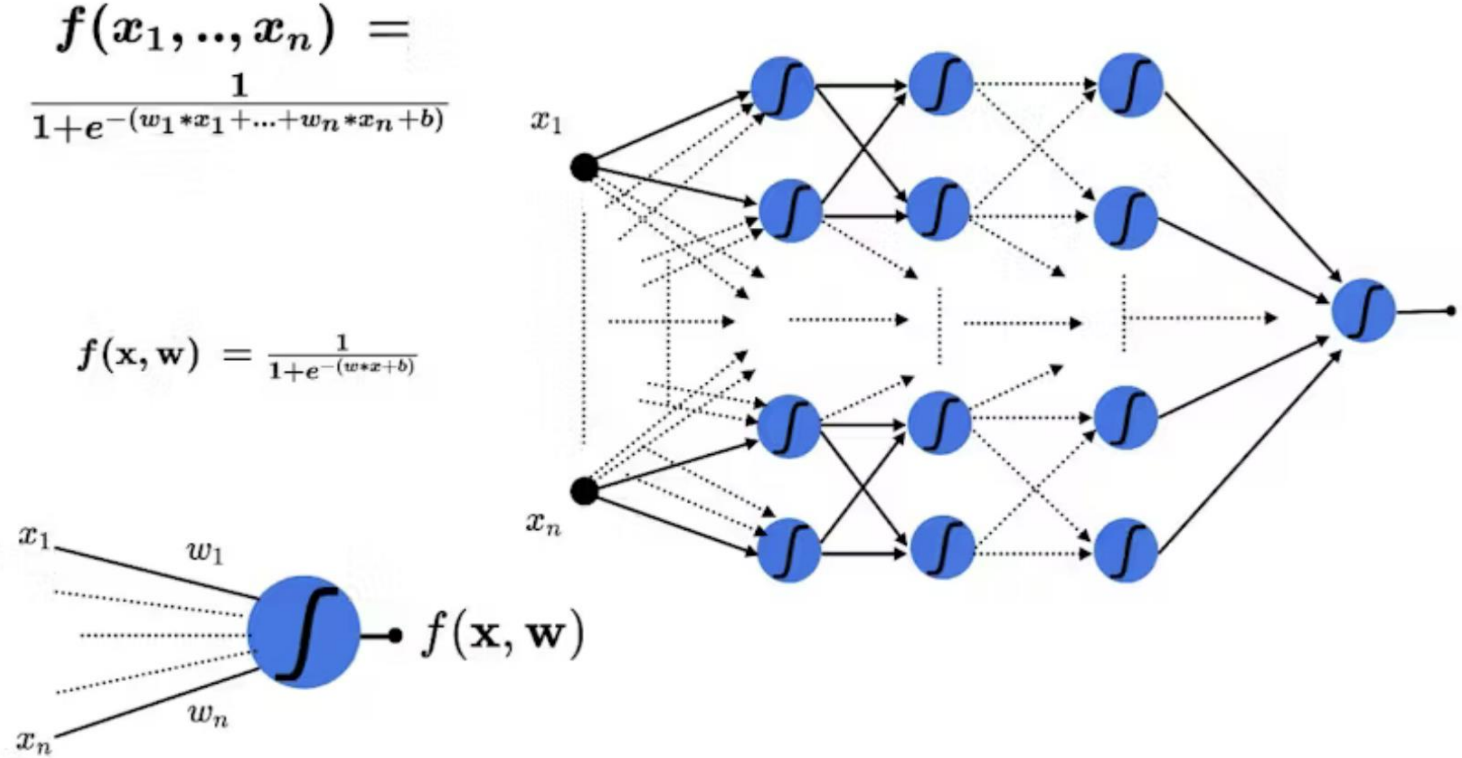


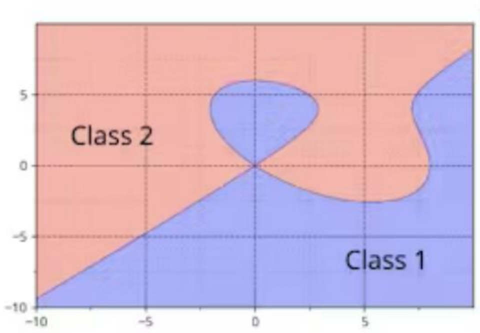


Credits: hackernoon.com

A better way to optimise: neural nets

This generalises to higher dimensional functions: **the universal approximation theorem.**
A much easier way to explore 'function space'!



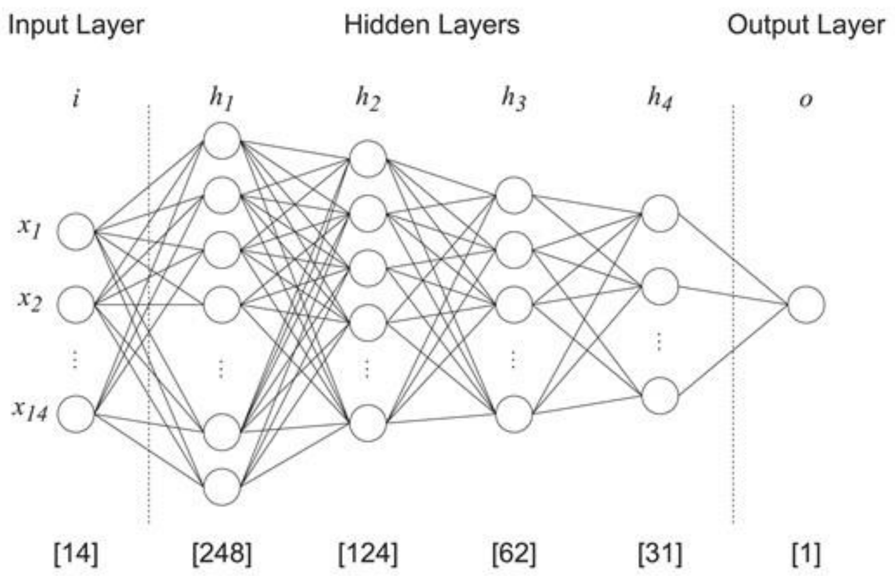


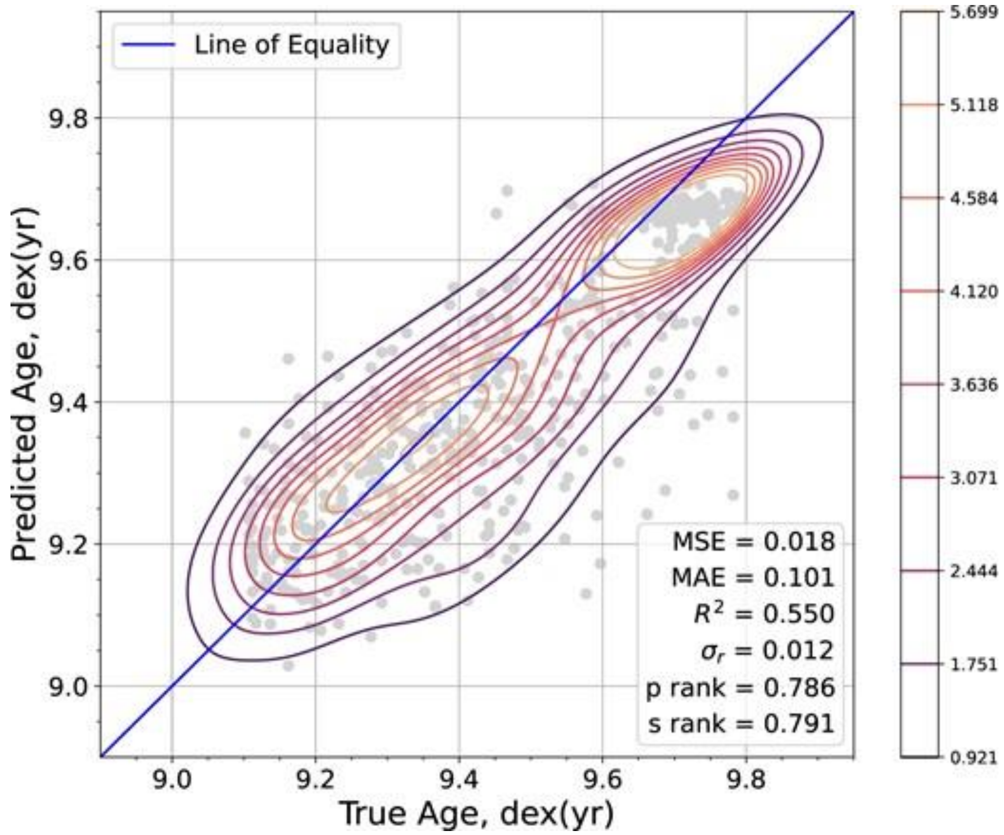
Optimising the weights and biases is a big computational linear algebra problem, solved by calculating numerical derivatives though "back propagation" (backprop) often on GPUs – ask a computer scientist for details!

Credits: hackernoon.com, Facebook: Meme for mathematicians

A better way to optimise: neural nets

Very powerful & versatile. Example: predicting galaxy stellar population ages from colours





Dr Laura Hunt

The Open University

Hunt+24 MNRAS 529, 479

1 My own first forays into machine learning

2 Astronomy's big data problem, gravitational lensing and convnets

3 Citizen Science, Galaxy Zoo and Zoobot

4 Clumpy galaxies: Zoobot, RCNNs and cGANs

5 Infrared gravitational lenses and deconvolution

6 Other astro ML, and Conclusions

The Open University

1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	Citizen Science, Galaxy Zoo and Zoobot
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	Infrared gravitational lenses and deconvolution
6	Other astro ML, and Conclusions

The Open University

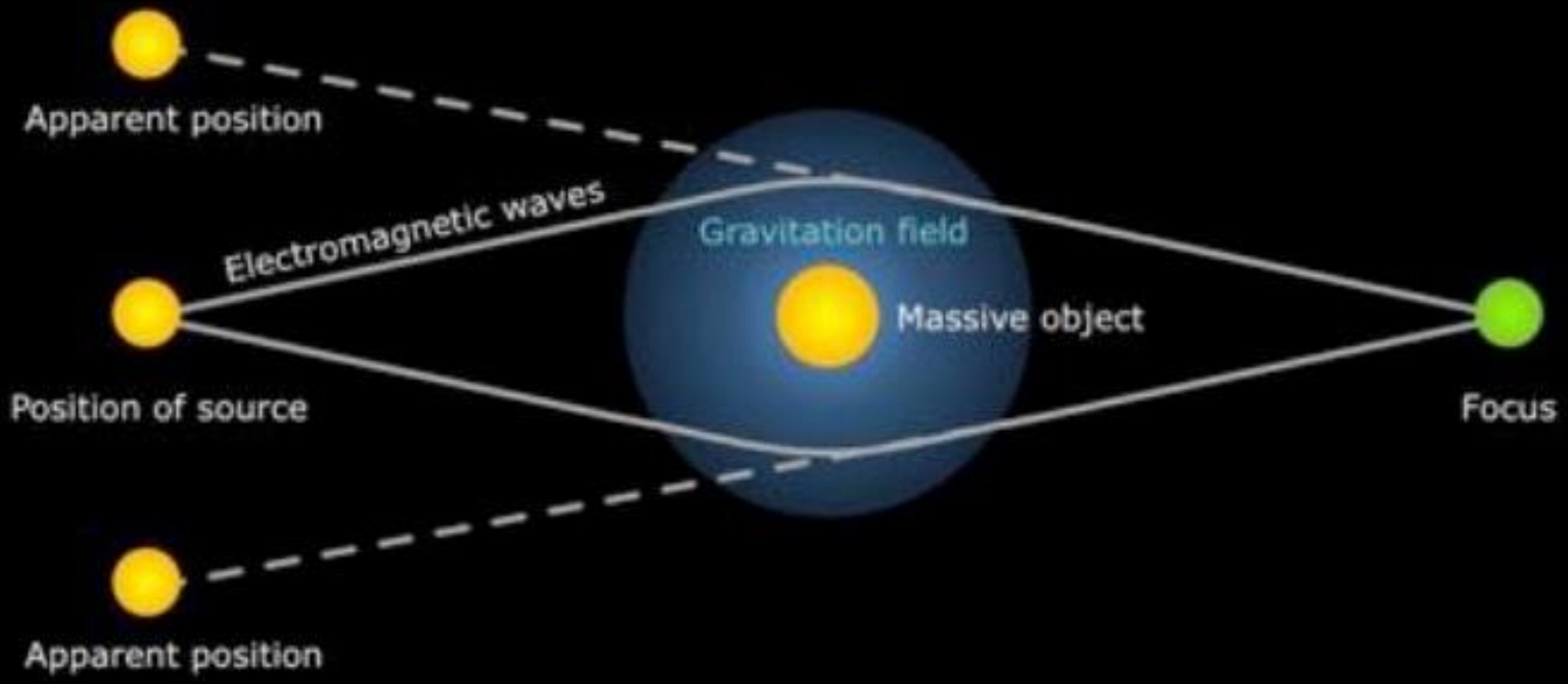
SKA telescope to generate more data than entire Internet in 2020

An exaflop-capable supercomputer, storage of at least 1.5 petabytes and data centres around the world will be required say scientists

Hamish Barwick (Computerworld)
07 July, 2011 12:05


esa
euclid


esa
euclid


Apparent position
Electromagnetic waves
Gravitation field
Massive object
Position of source
Focus
Apparent position
nasa.gov

Tells you where the dark matter is

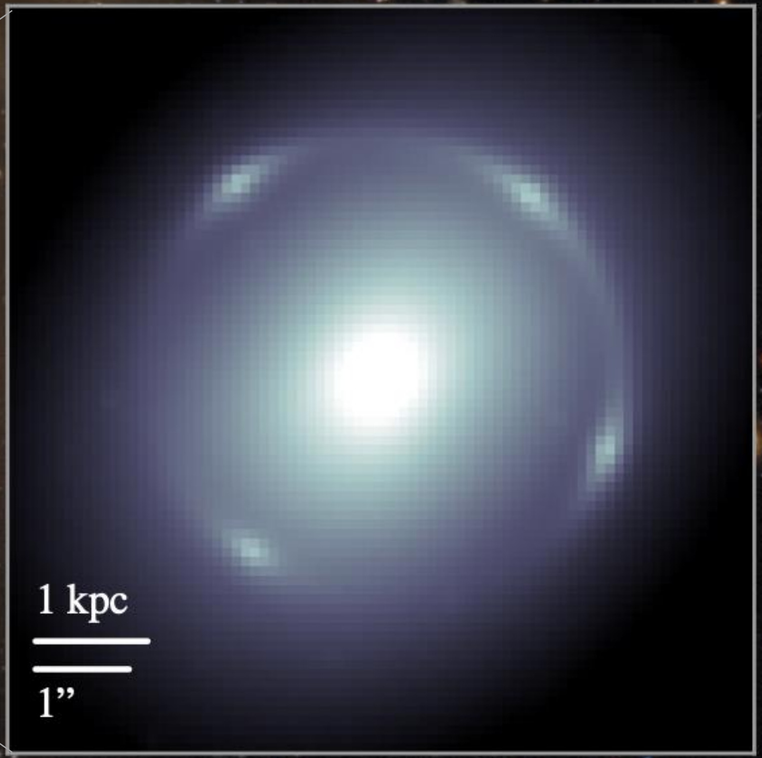
Constrains dark matter theories

Time delays between images depend on Hubble constant (expansion rate of Universe)

Magnifies background galaxies

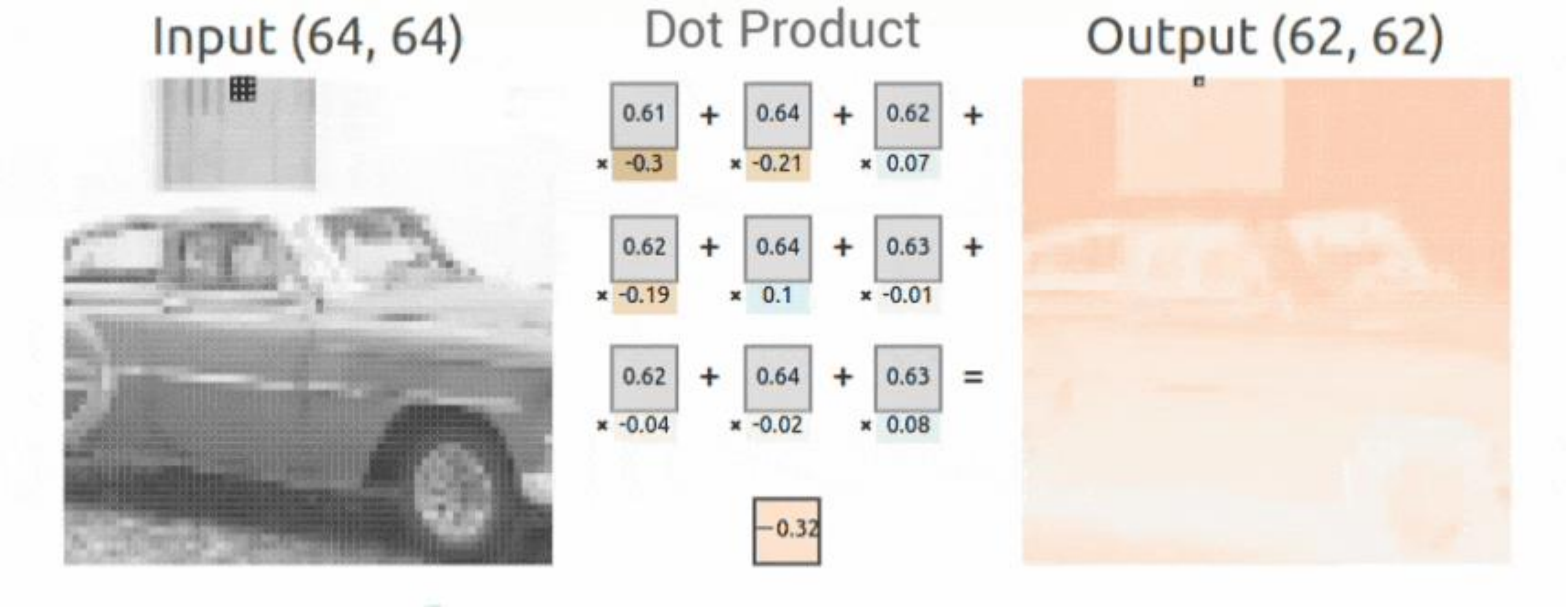
But they are rare! Need chance alignments

Euclid: 100000 strong lenses! But among a billion other galaxies



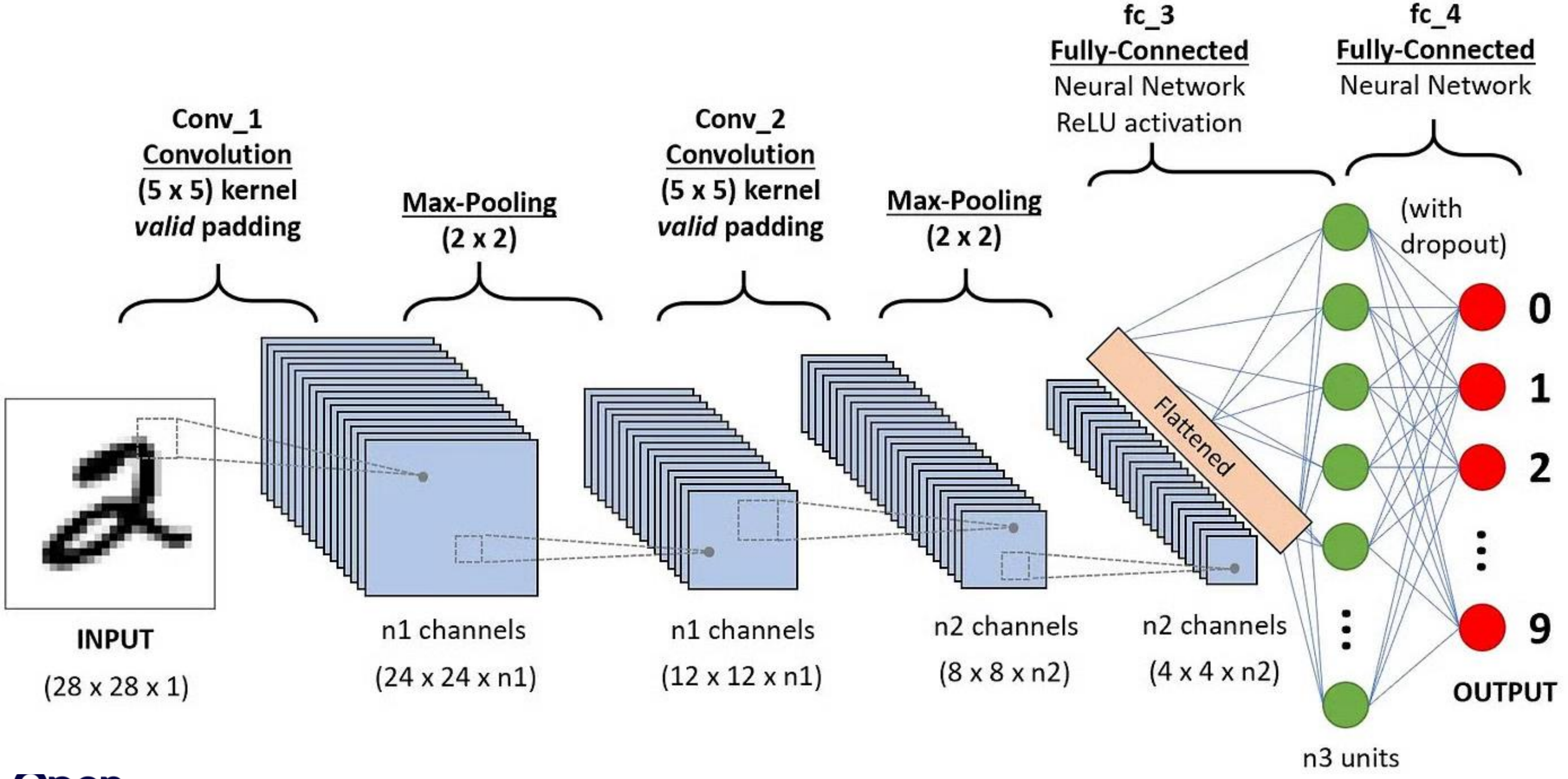
O'Riordan+25 Astronomy & Astrophysics 694, A145

How to explore these giant imaging data sets: convolutional neural networks



learnopencv.com

How to explore these giant imaging data sets: convolutional neural networks

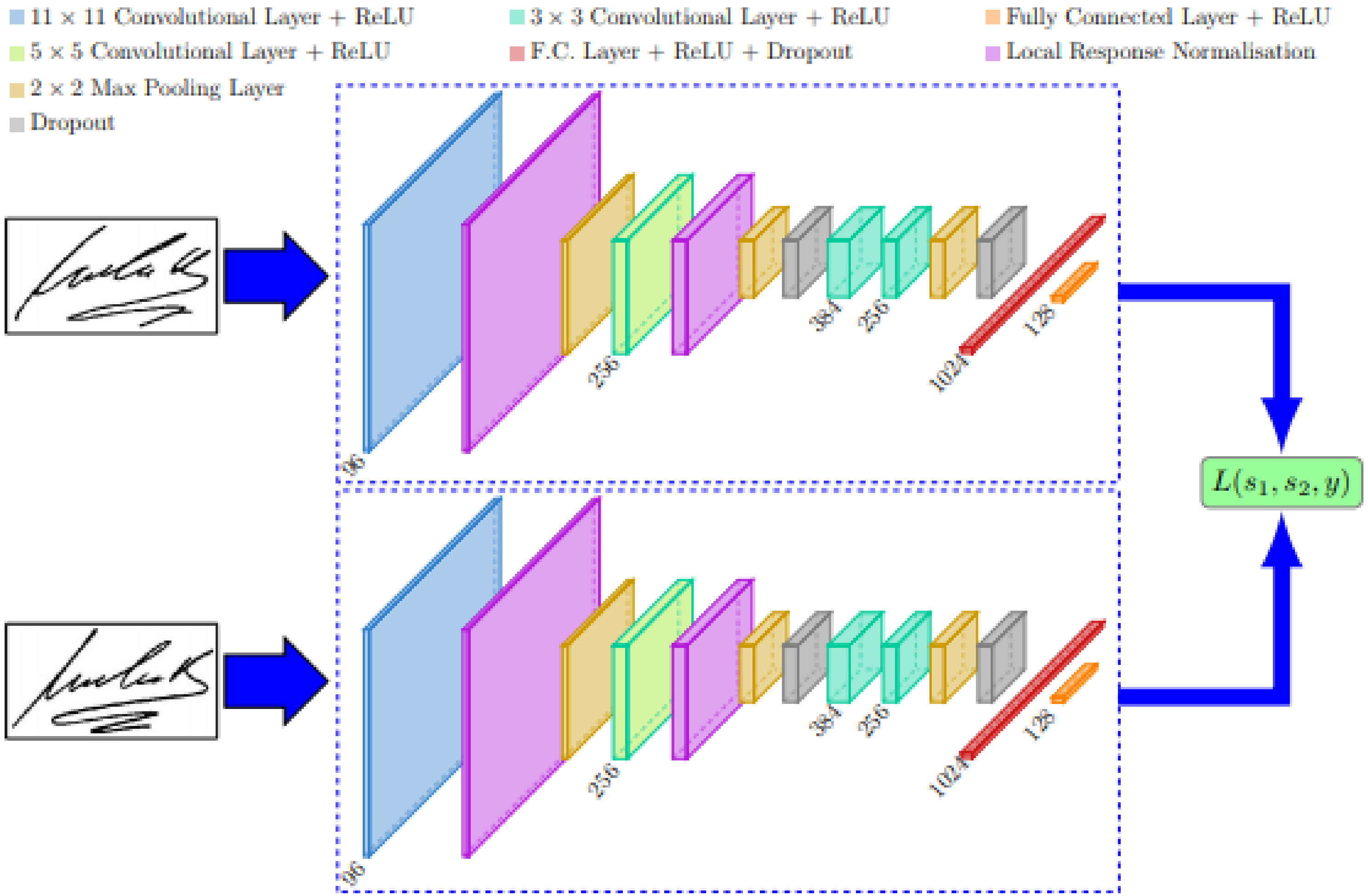


The Open University

datacamp.com

How to explore these giant imaging data sets: convolutional neural networks

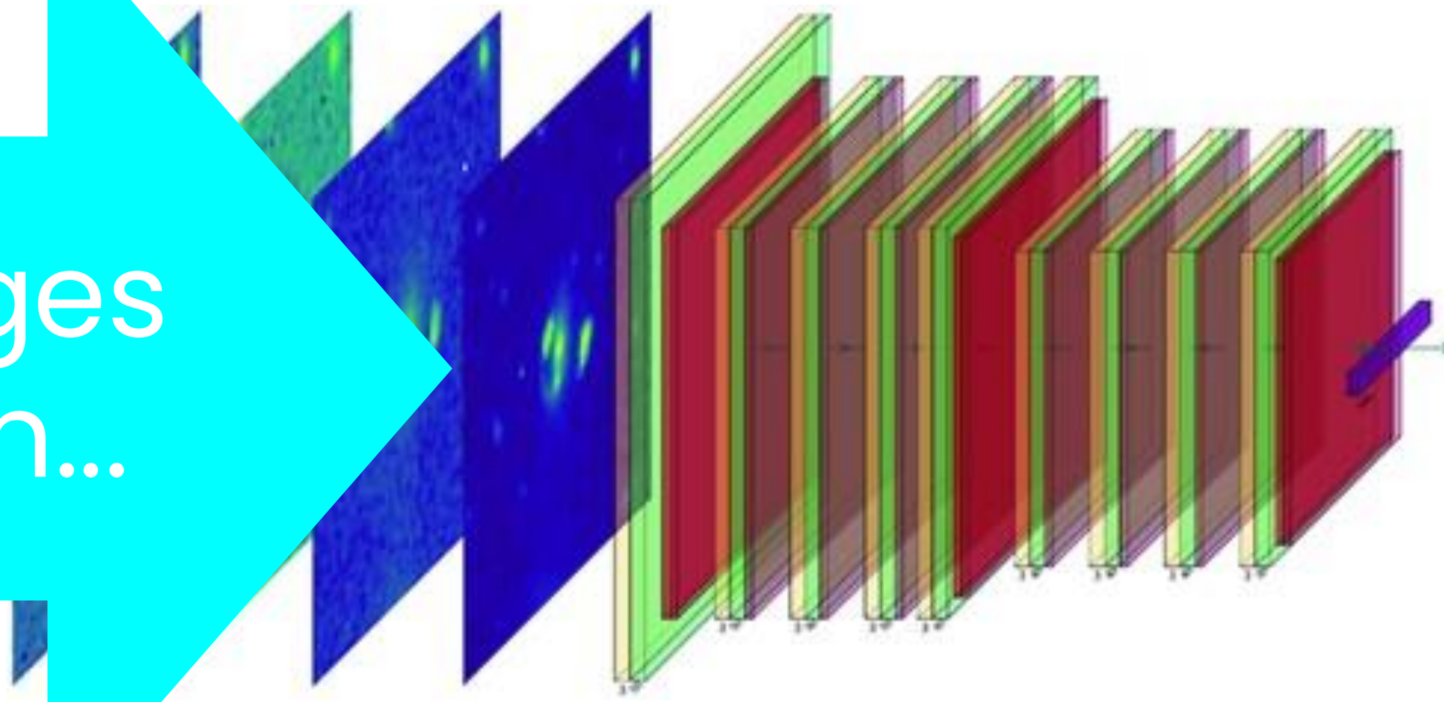
Dr Jane Bromley



Finding strong gravitational lenses

CNNs; Davies+19, Wilde+22

Images go in...

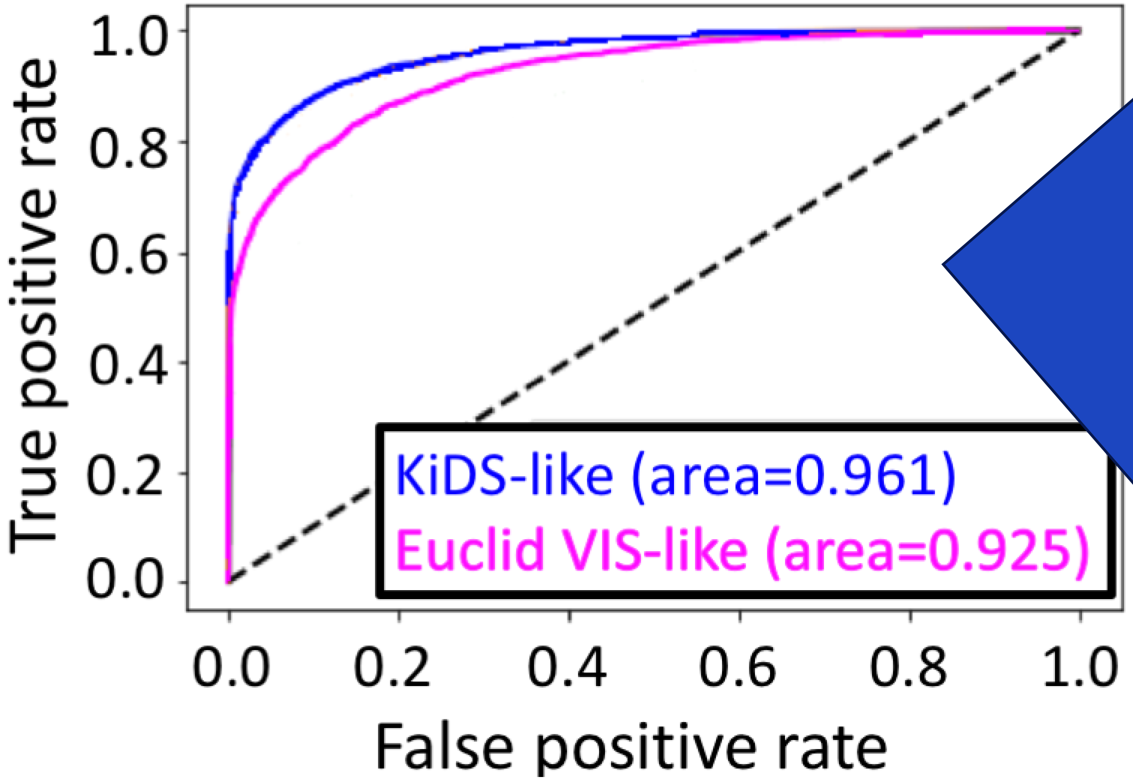


...and a classification comes out

Conv 2D | Dropout | ReLu | MaxPool 2D

Classification: finding strong gravitational lenses

CNNs; Davies+19, Wilde+22

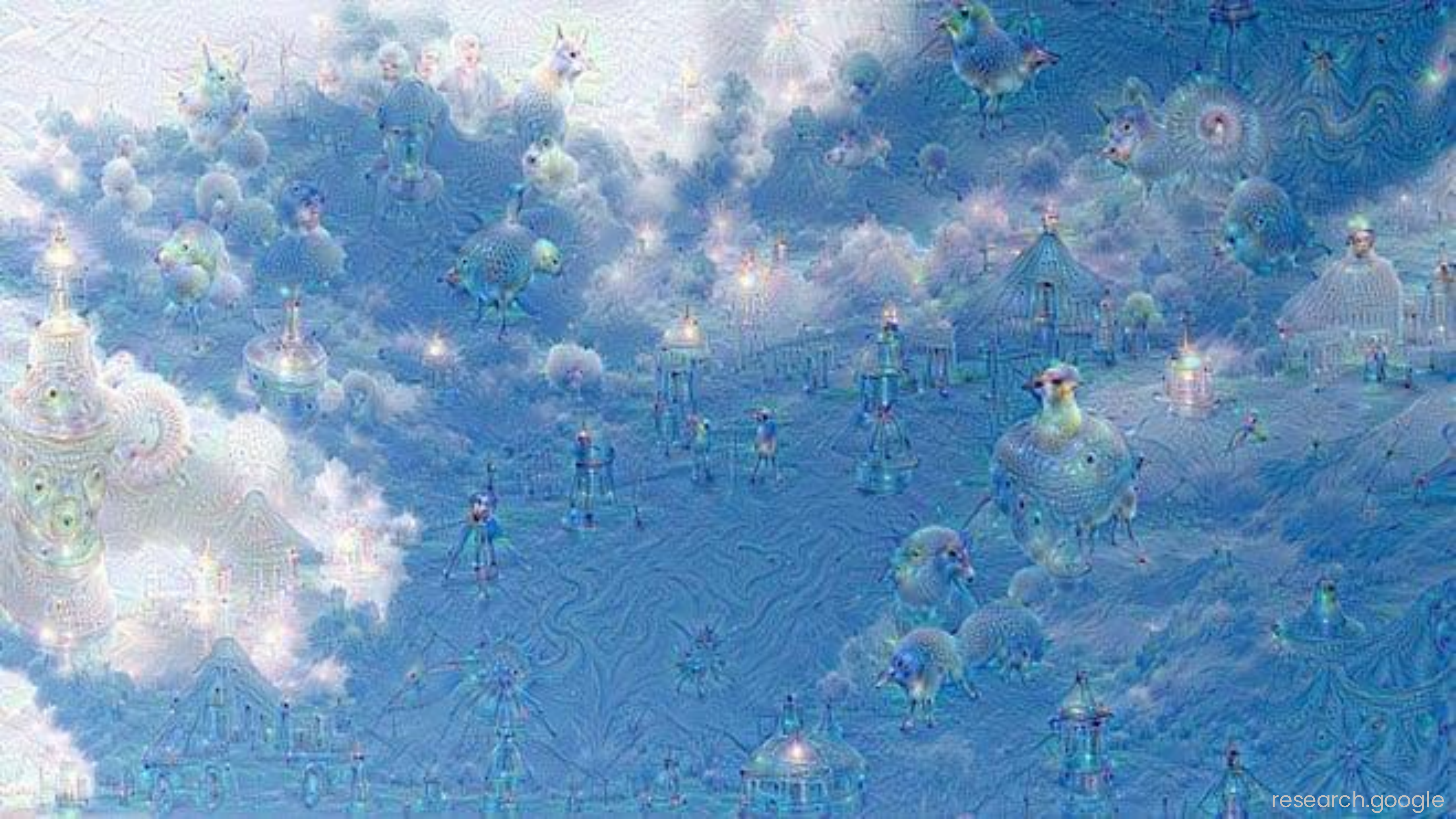


Good performance, but not quite good enough for 100000 lenses among one billion non-lenses

Dr Josh Wilde

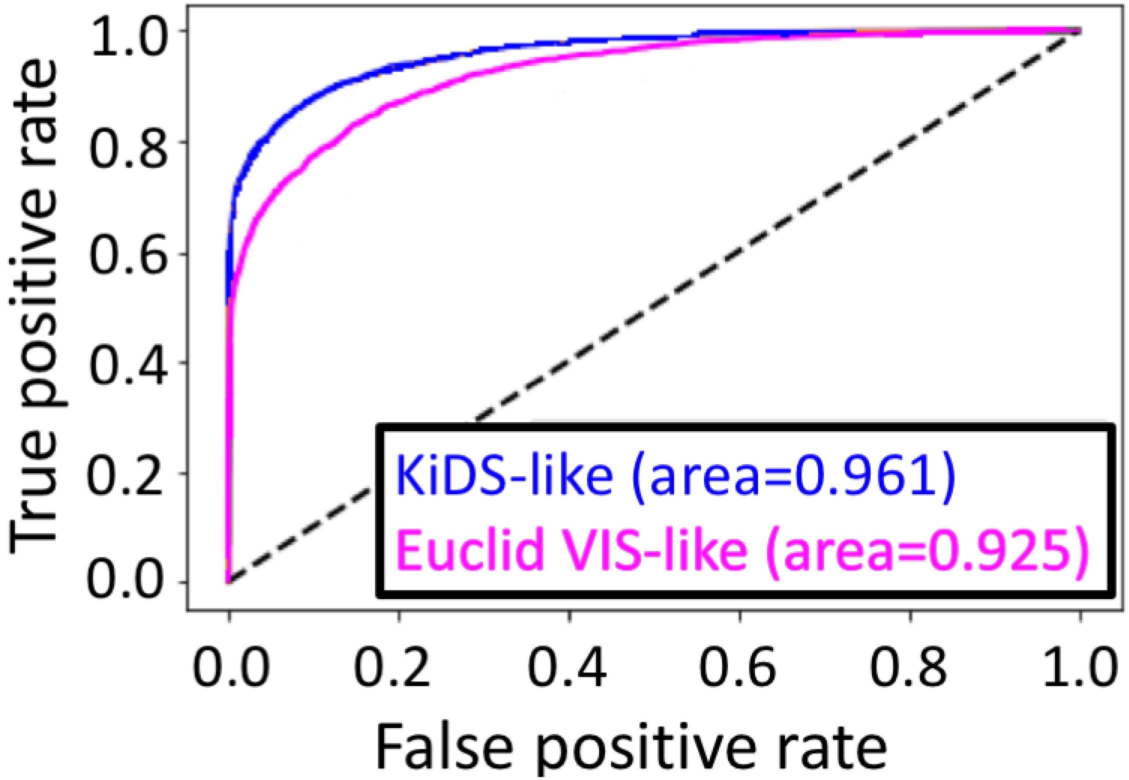
The Open University

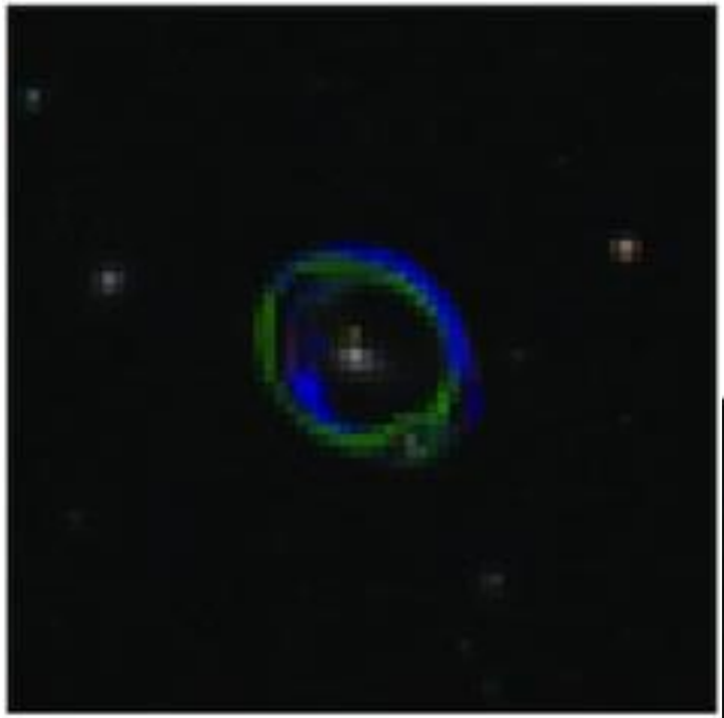

research.google

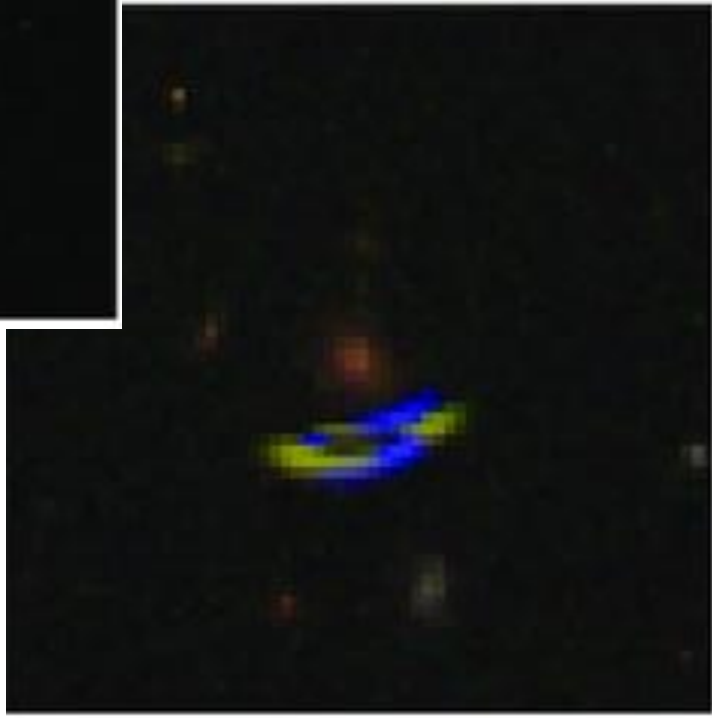

research.google

Classification: finding strong gravitational lenses

CNNs; Davies+19, Wilde+22







The Open University

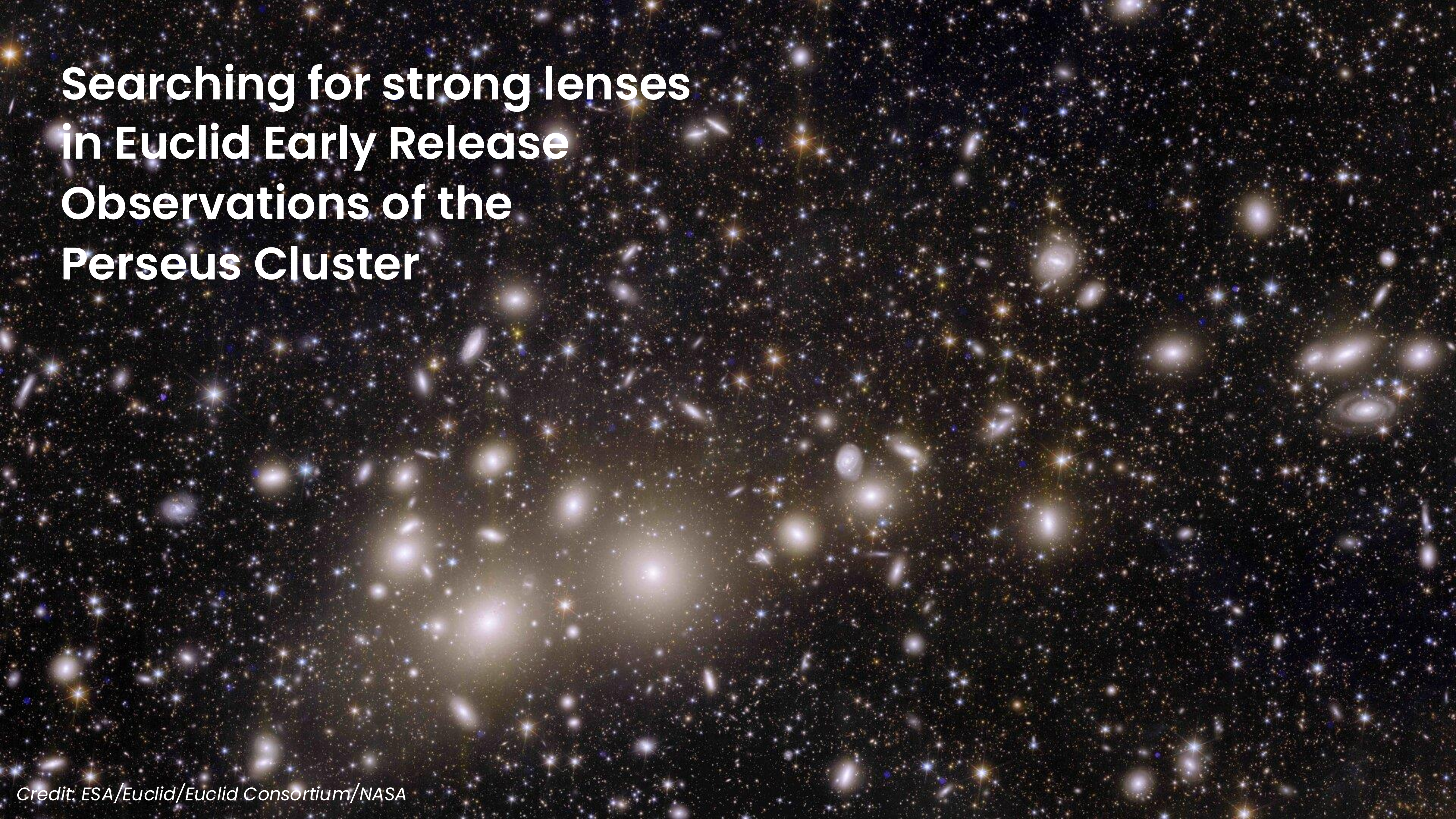
The Open University

The Open University

Citizen Science

Dickinson+22, Serjeant+23, Serjeant+24, Pearson+25.

Searching for strong lenses in Euclid Early Release Observations of the Perseus Cluster



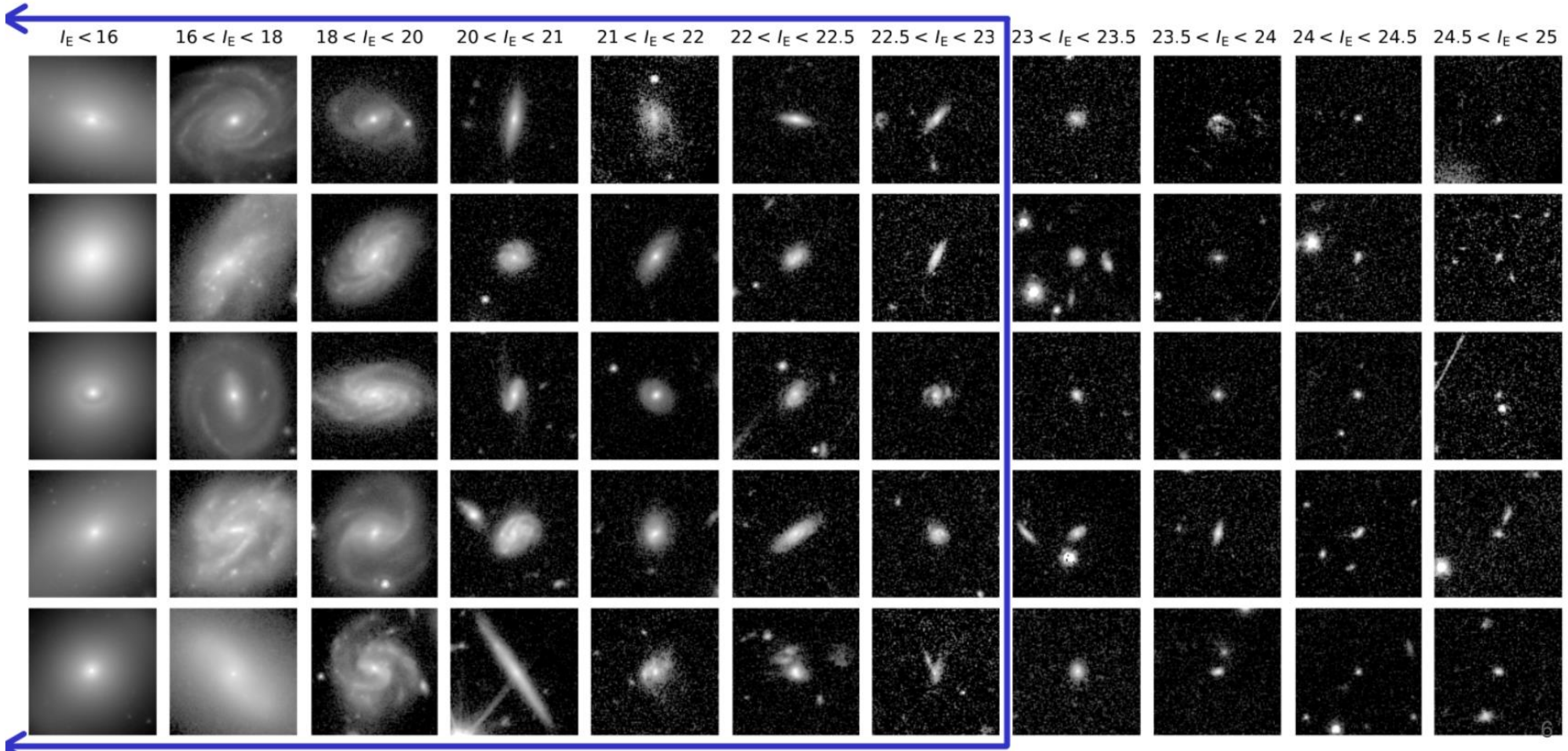
Credit: ESA/Euclid/Euclid Consortium/NASA

Searching for strong lenses in Euclid Early Release Observations of the Perseus Cluster

Credit: ESA/Euclid/Euclid Consortium/NASA

The ERO Lens Finding Experiment (ELSE)

12 086 sources selected for visual inspection with VIS I_E<23

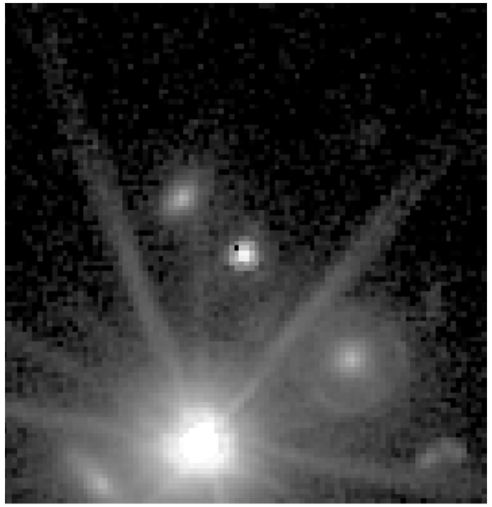


Slide by Ruby Pearce-Casey

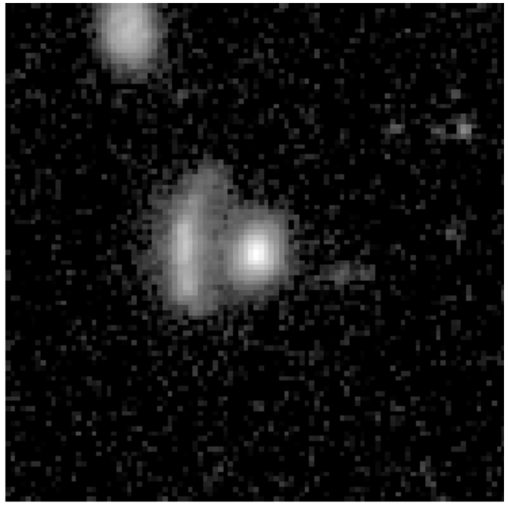
The Grading System

The 12,086 stamps were classified into one of the following non-overlapping categories:

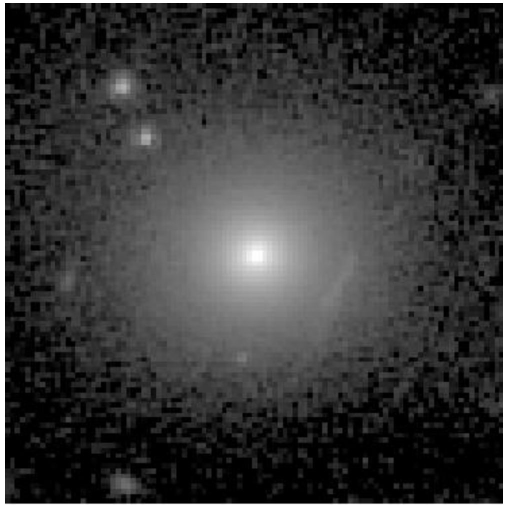
- **A:** *definite lens*. Clear lensing features present.
- **B:** *probable lens.* Lensing features present but additional information is needed to confirm this.
- **C:** *possible lens.* Some lensing features but this could be explained by other phenomena.
- **I:** *interesting*. Definitely not a lens but shows interesting features/morphology.
- **X:** definitely not a lens.



Grade A



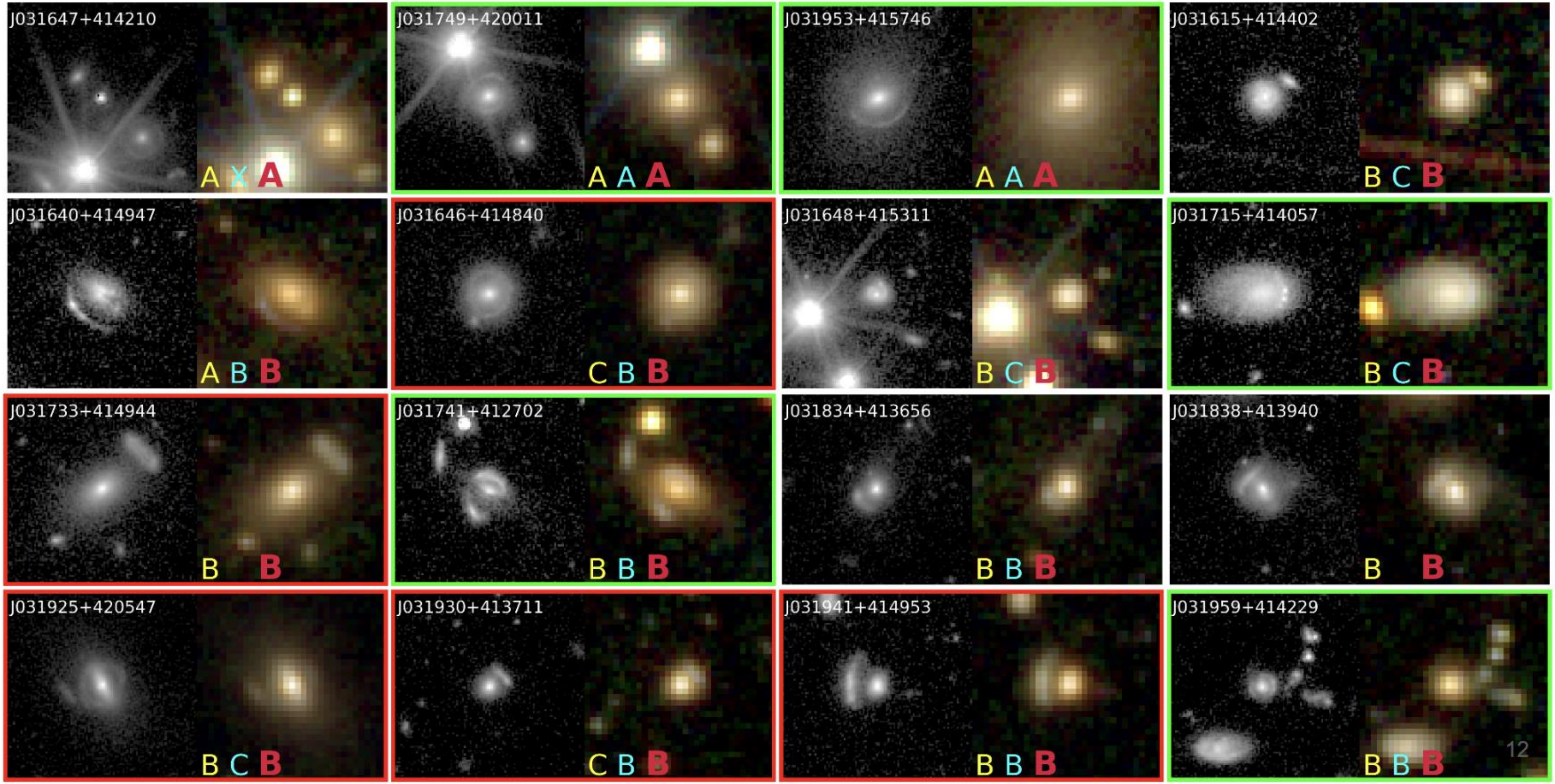
Grade B



Grade C

Slide by Ruby Pearce-Casey

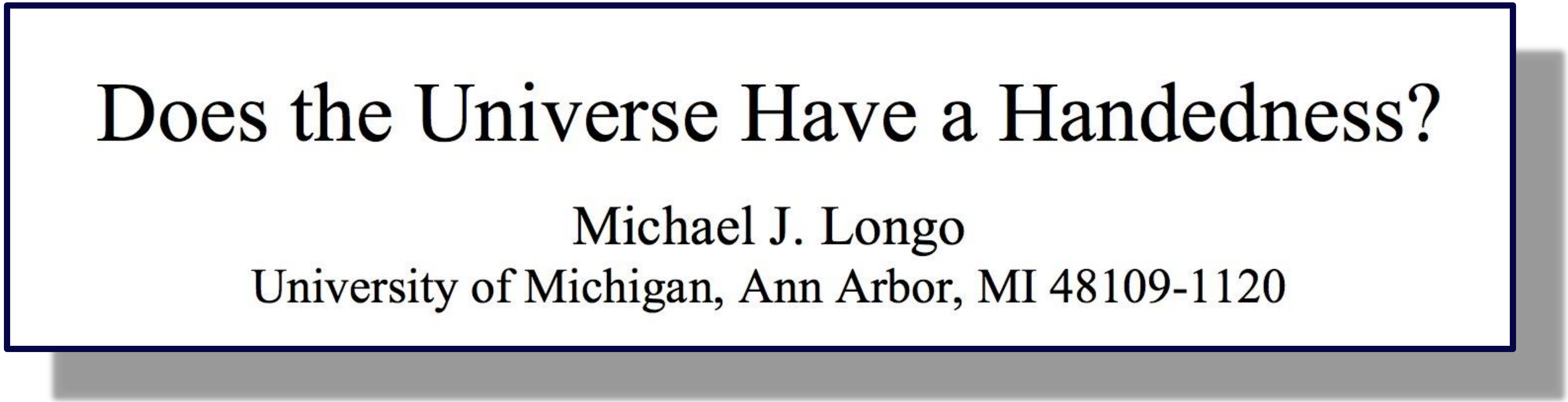
The lens candidates



12

Slide by Ruby Pearce-Casey

Using humans to classify galaxies is not new...


Does the Universe Have a Handedness?

Michael J. Longo

University of Michigan, Ann Arbor, MI 48109-1120

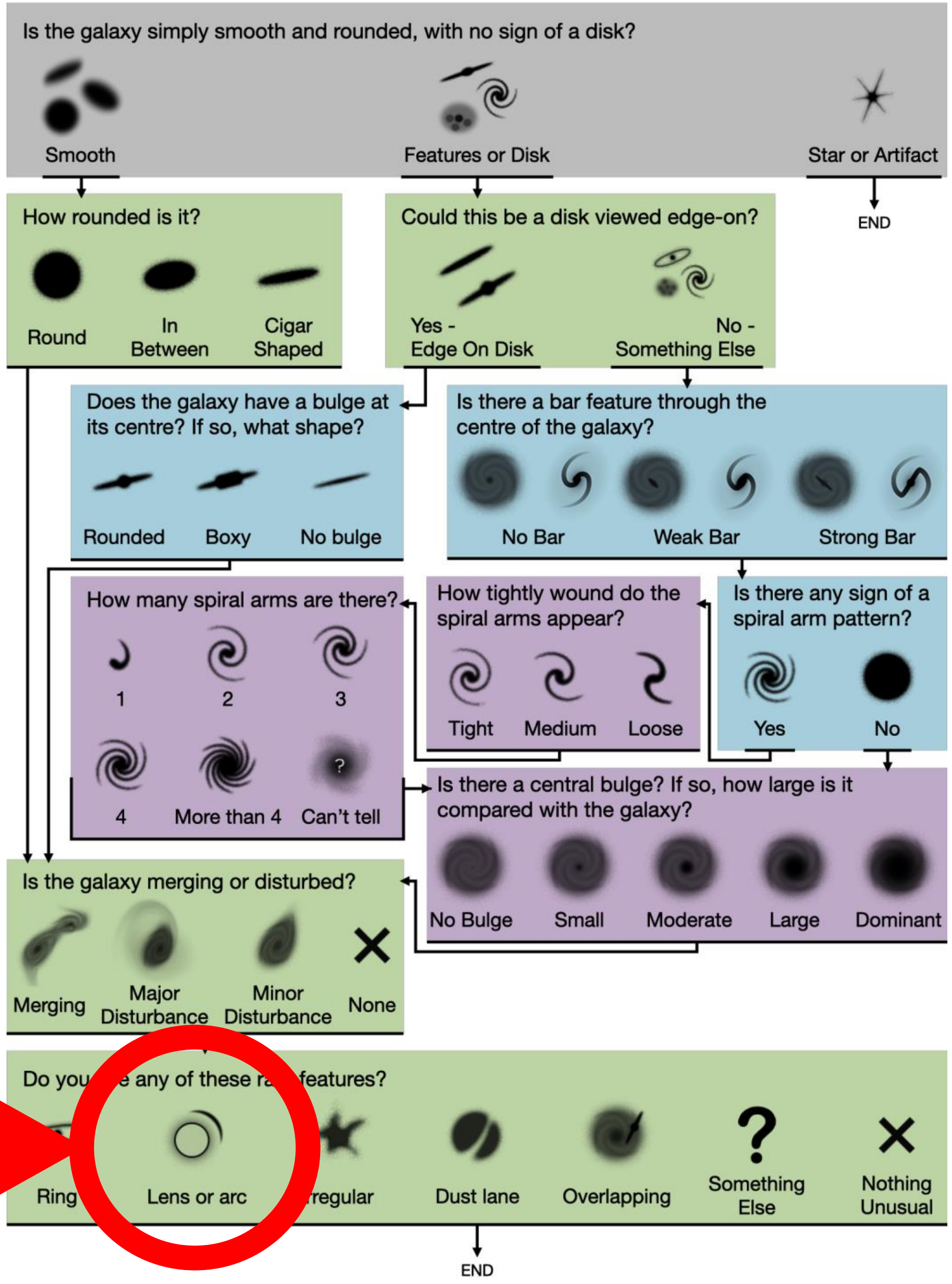
astro-ph/0703325

Jarrett+17 AJ

- Task: classify 900,000 galaxies
- Would take an expert 3-5 years even working 24 hours a day, 7 days a week

- Expectation: 20–30,000 volunteers
- What happened: 100,000 volunteers making 40 million classifications

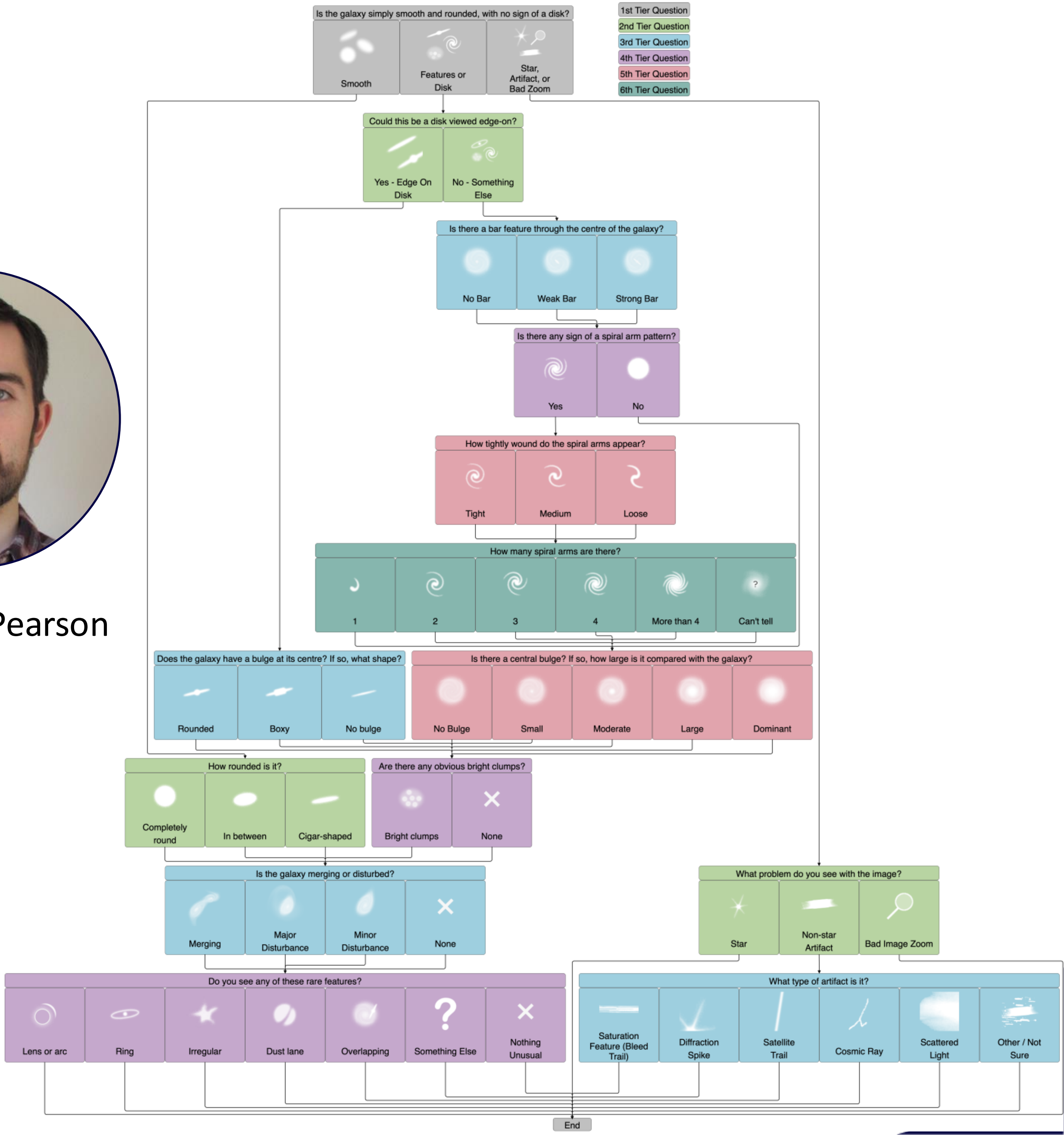
Galaxy Zoo workflow breaks down a complicated classification into lots of small, manageable volunteer decisions



Walmsley+21 MNRAS 509, 3966

Galaxy Zoo workflow breaks down a complicated classification into lots of small, manageable volunteer decisions

Dr James Pearson



The Open University

Pearson+25 MNRAS submitted

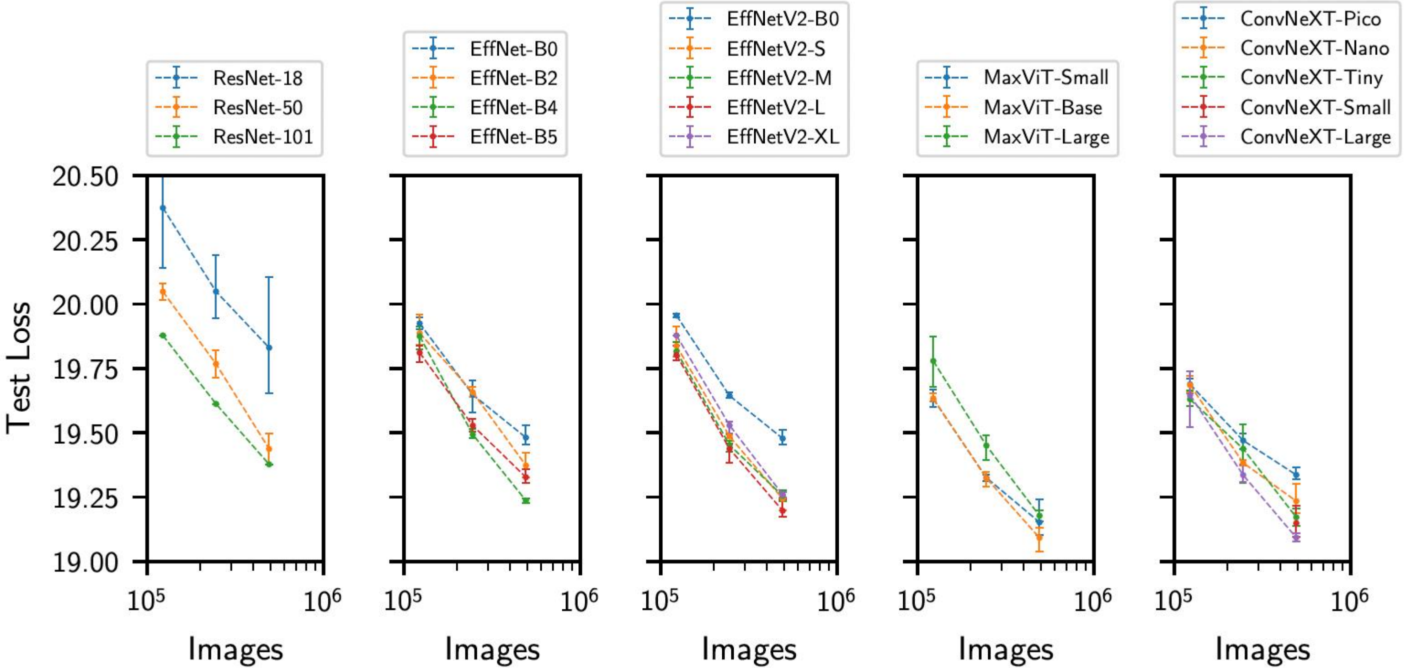
Zoobot: foundation models for galaxies

14 years of Galaxy Zoo annotations of 847k images and >100M classifications
This is more than half of all human galaxy annotations ever collected!

Tried every popular architecture (ConvNexT, EffNetV2, MaxViT, etc.) and every model size, from 1M to 200M parameters

Adding more labels helps: all models have same power-law improvements.
Size helps only so far: improvements plateau after ~100M parameters

Walmsley+24 arXiv:2404.02973



The Open University

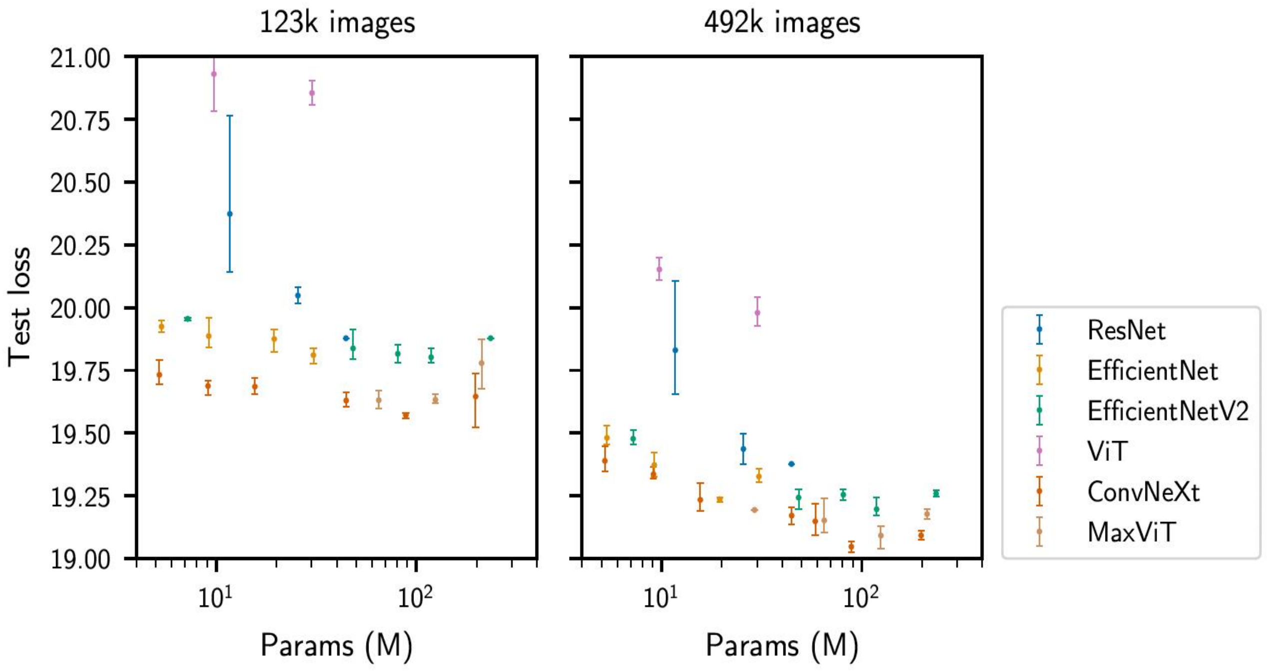
Zoobot: foundation models for galaxies

14 years of Galaxy Zoo annotations of 847k images and >100M classifications
This is more than half of all human galaxy annotations ever collected!

Tried every popular architecture (ConvNexT, EffNetV2, MaxViT, etc.) and every model size, from 1M to 200M parameters

Adding more labels helps: all models have same power-law improvements.
Size helps only so far: improvements plateau after ~100M parameters

Walmsley+24 arXiv:2404.02973

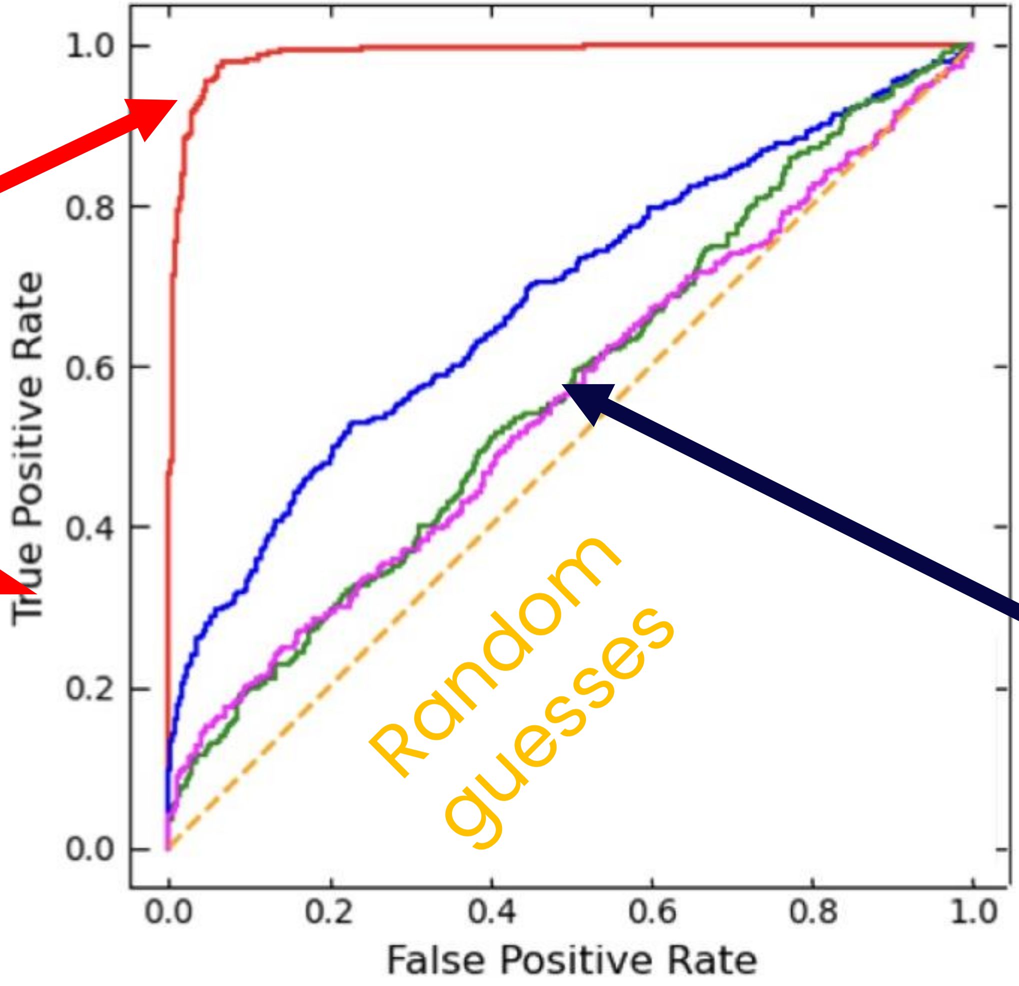


Can this be used to find gravitational lenses?

Using Zoobot to find gravitational lenses in Euclid

Fine-tune Zoobot EfficientNetB0 model: remove 'head', freeze weights of the remaining layers, train new head on simulated lenses and non-lenses

Our EfficientNet-B0 model outperforms all others



Probably because it's trained on 100 million human classifications!

Ruby Pearce-Casey

Example competitor networks

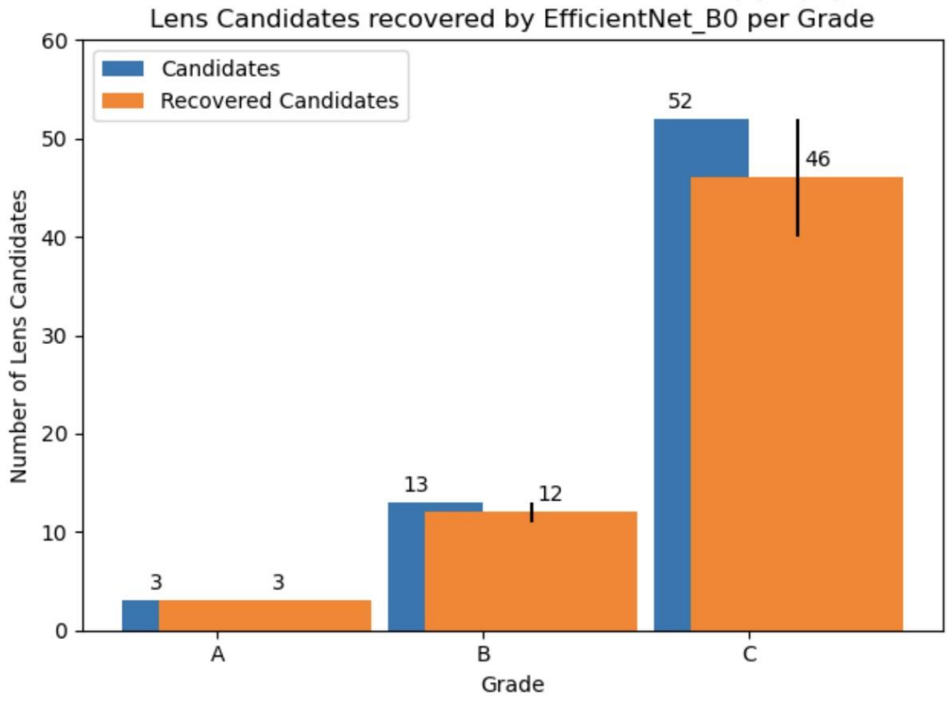
Pearce-Casey+25 Astronomy & Astrophysics, 696, A214

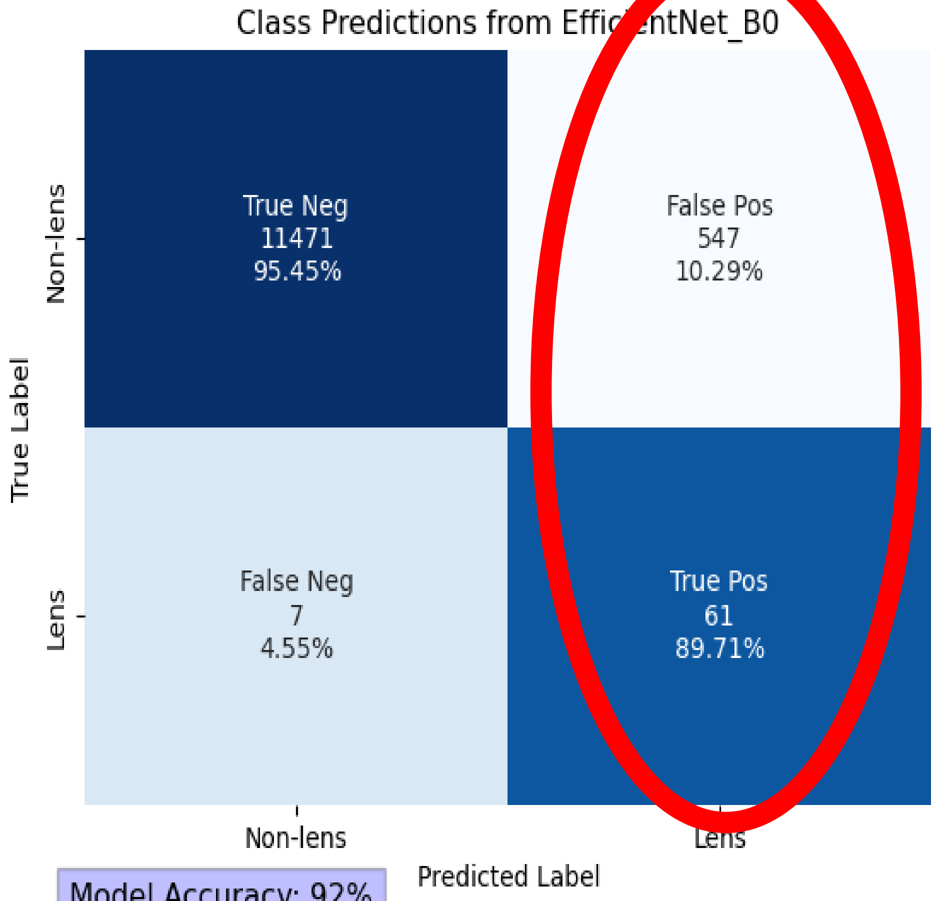
Classification: finding strong gravitational lenses

(Slide by Ruby Pearce-Casey)

Number of True Positives Recovered with

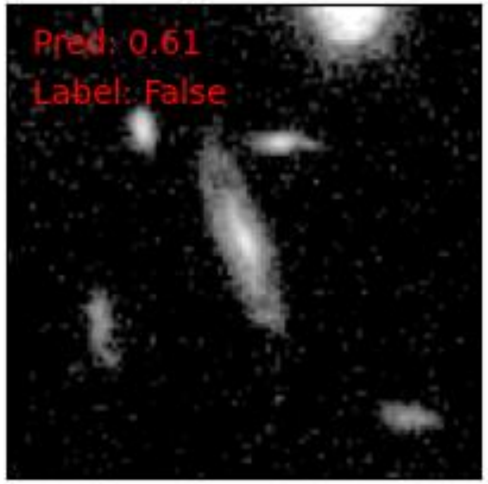
EfficientNet_B0 with $p_{threshold} > 0.5$



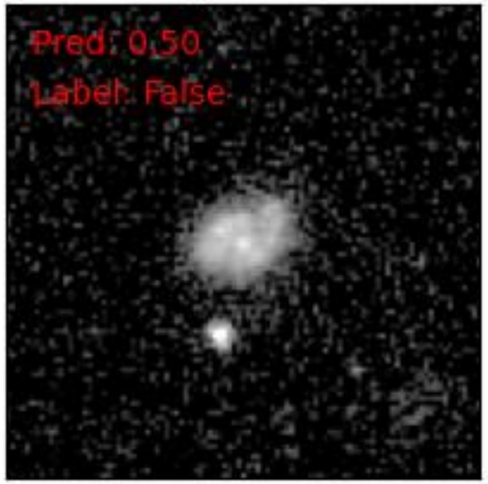


Model Accuracy: 92%

ID_120457



ID_50077

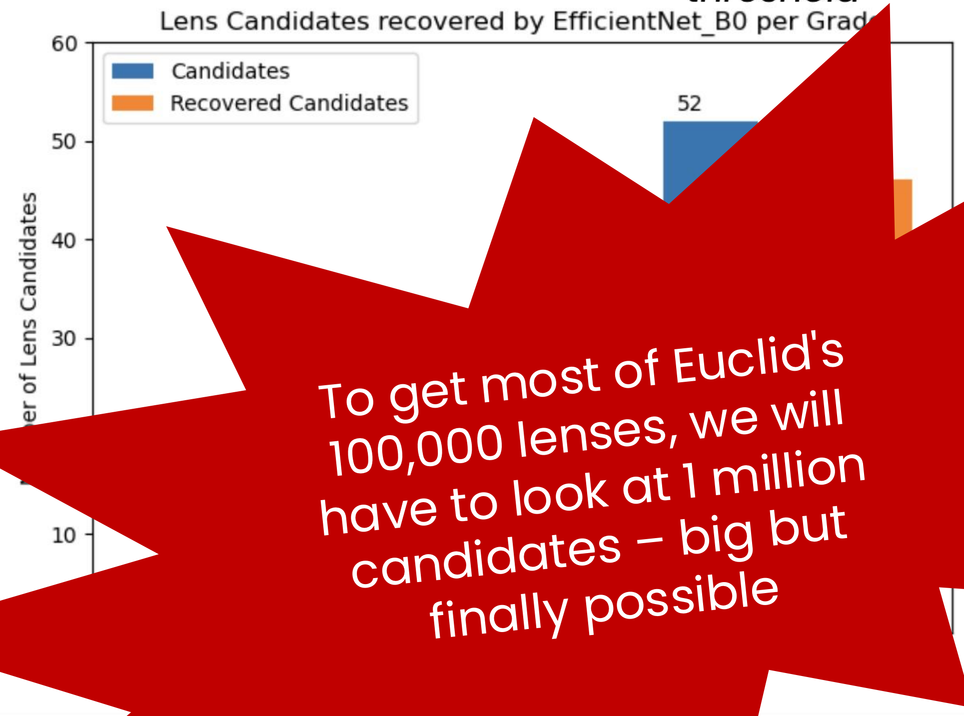


The Open University

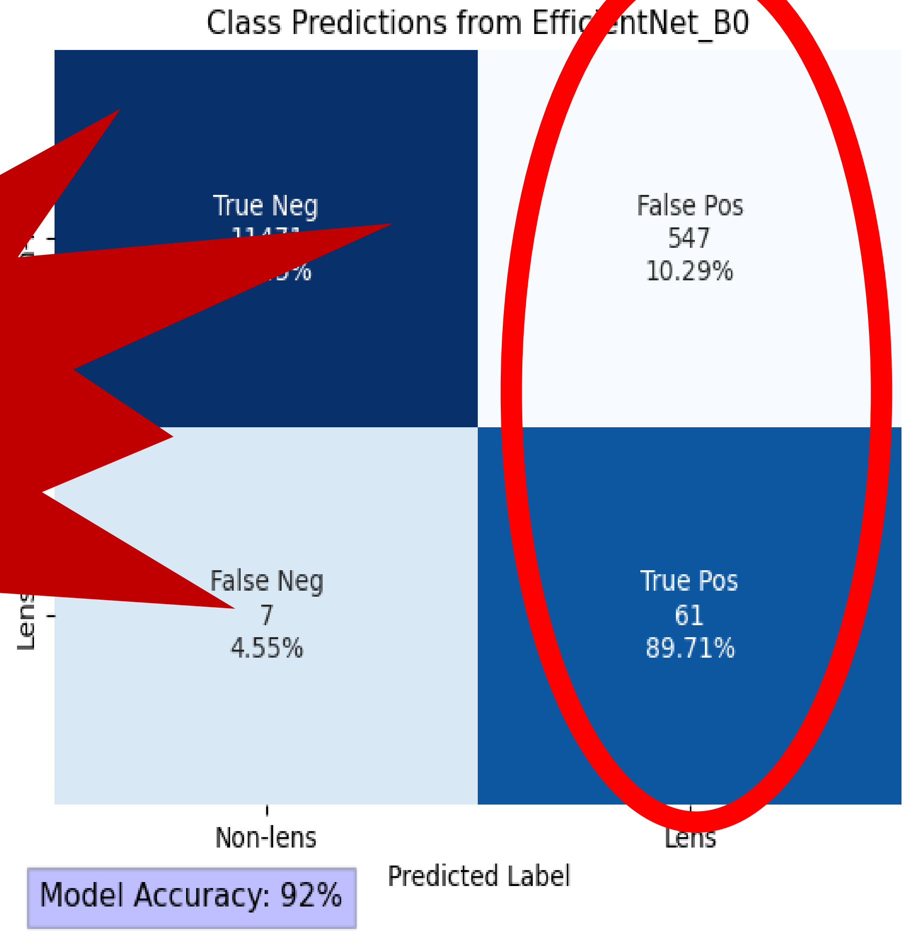
Classification: finding strong gravitational lenses

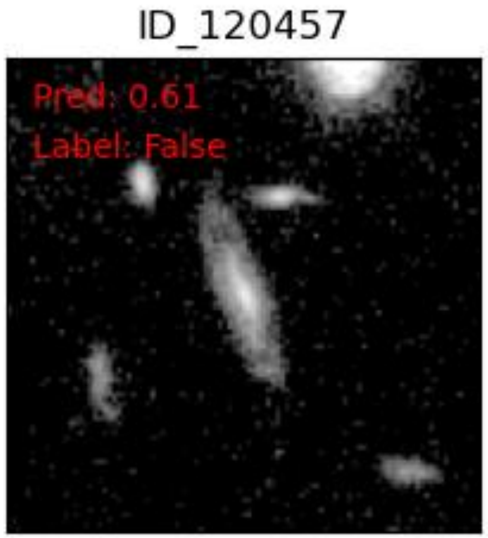
(Slide by Ruby Pearce-Casey)

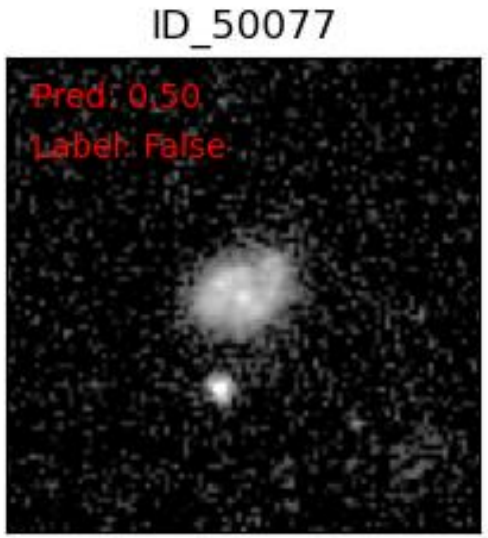
Number of True Positives Recovered with *EfficientNet_B0 with* $p_{threshold} > 0.5$



To get most of Euclid's 100,000 lenses, we will have to look at 1 million candidates – big but finally possible



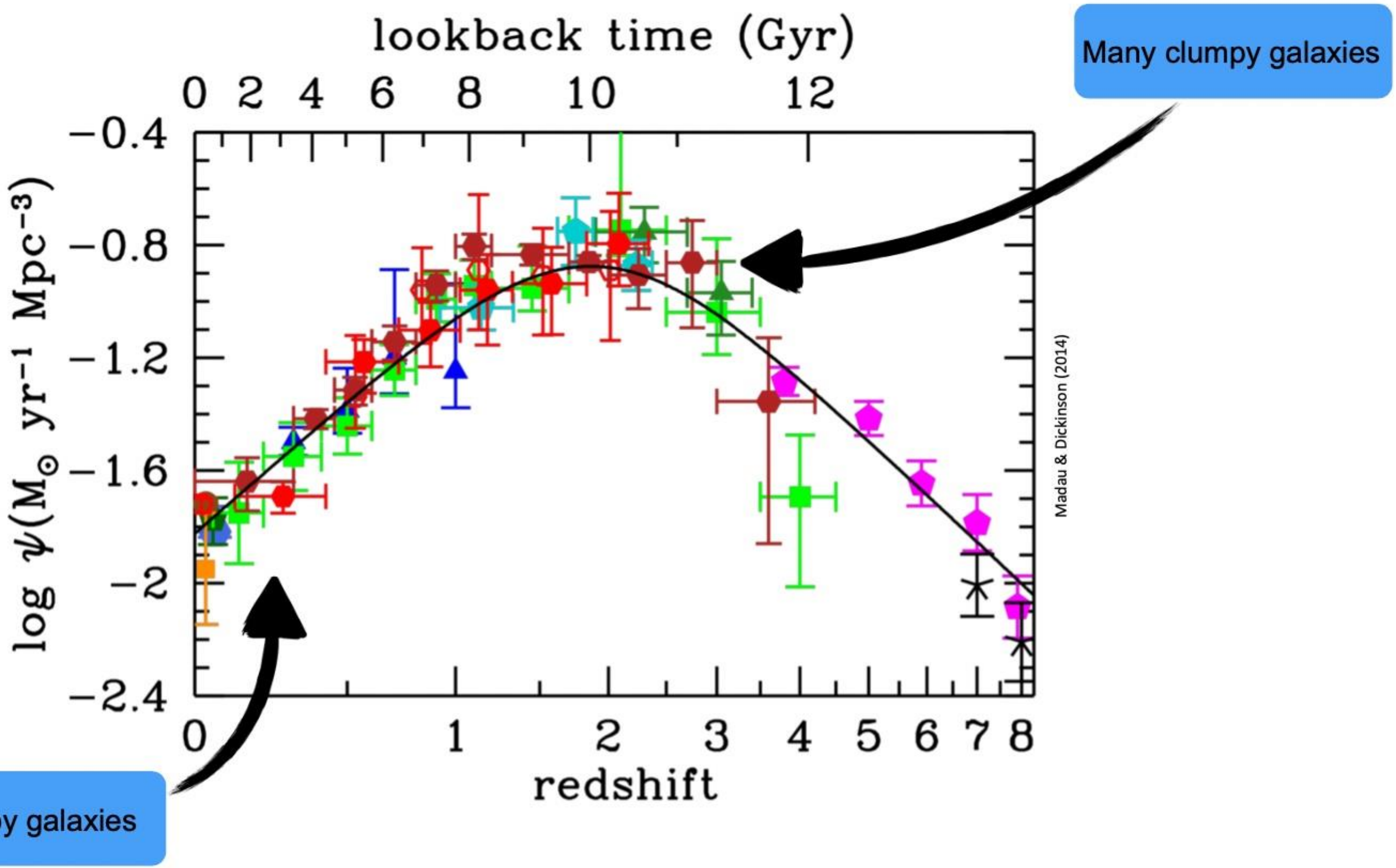




1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	**Citizen Science, Galaxy Zoo and Zoobot**
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	Infrared gravitational lenses and deconvolution
6	Other astro ML, and Conclusions

The Open University

The Open University



Slide from Hugh Dickinson

Language English

Galaxy Zoo: Clump Scout

ABOUT CLASSIFY TALK COLLECT RECENTS GALAXY ZOO

Thank you! With your help, we've compiled the first comprehensive catalog of clumpy galaxies in the local Universe. Stay tuned over the next few months for our first scientific results!

Great work! Looks like this project is out of data at the moment!
See the results or dismiss this message

ORGANIZATION: GALAXY ZOO

Help to find regions where stars are being born!

Learn more | Get started

Dr Hugh Dickinson

Image credit: Dr Hugh Dickinson

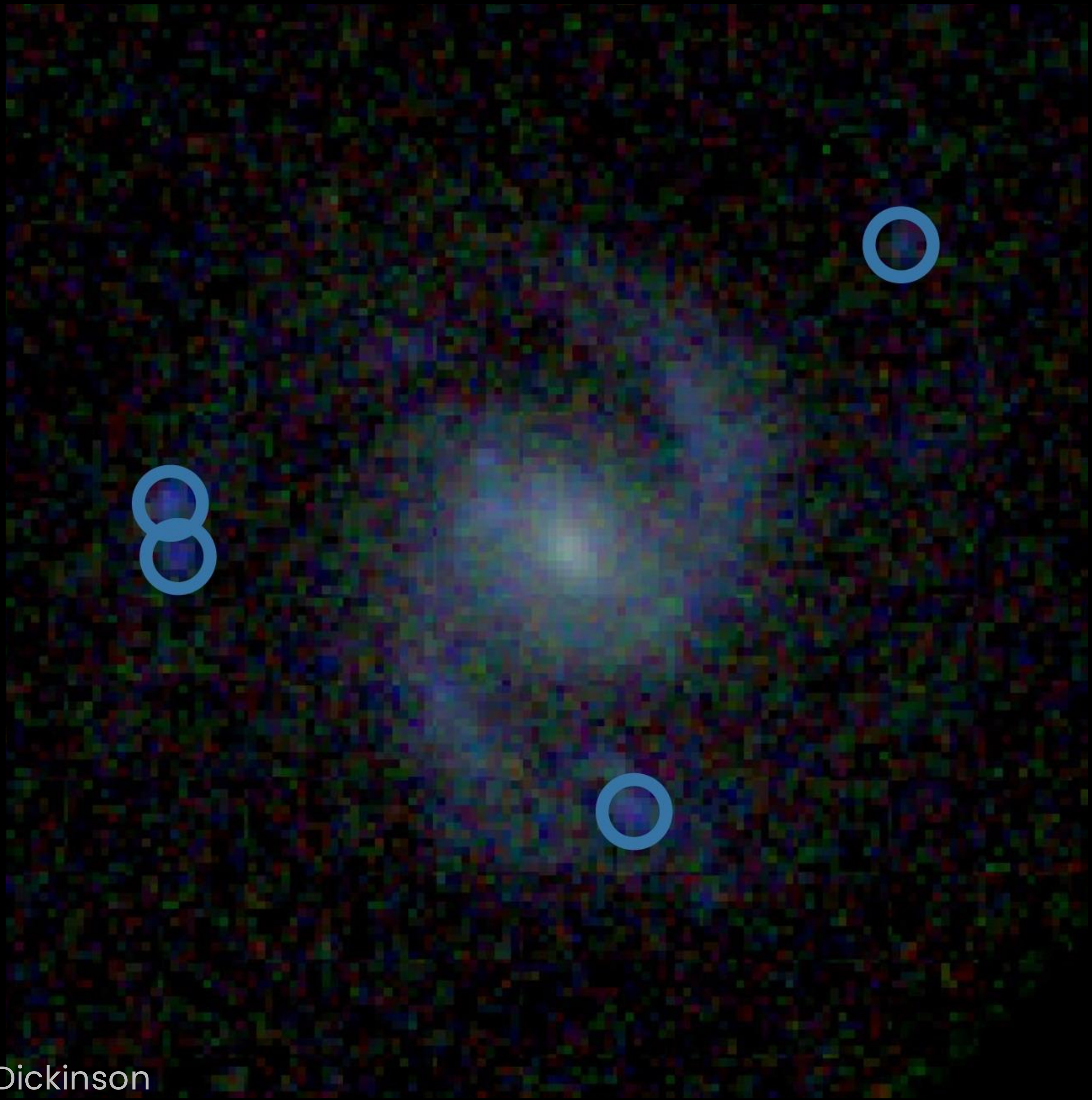


Image credit: Dr Hugh Dickinson

Image credit: Dr Hugh Dickinson

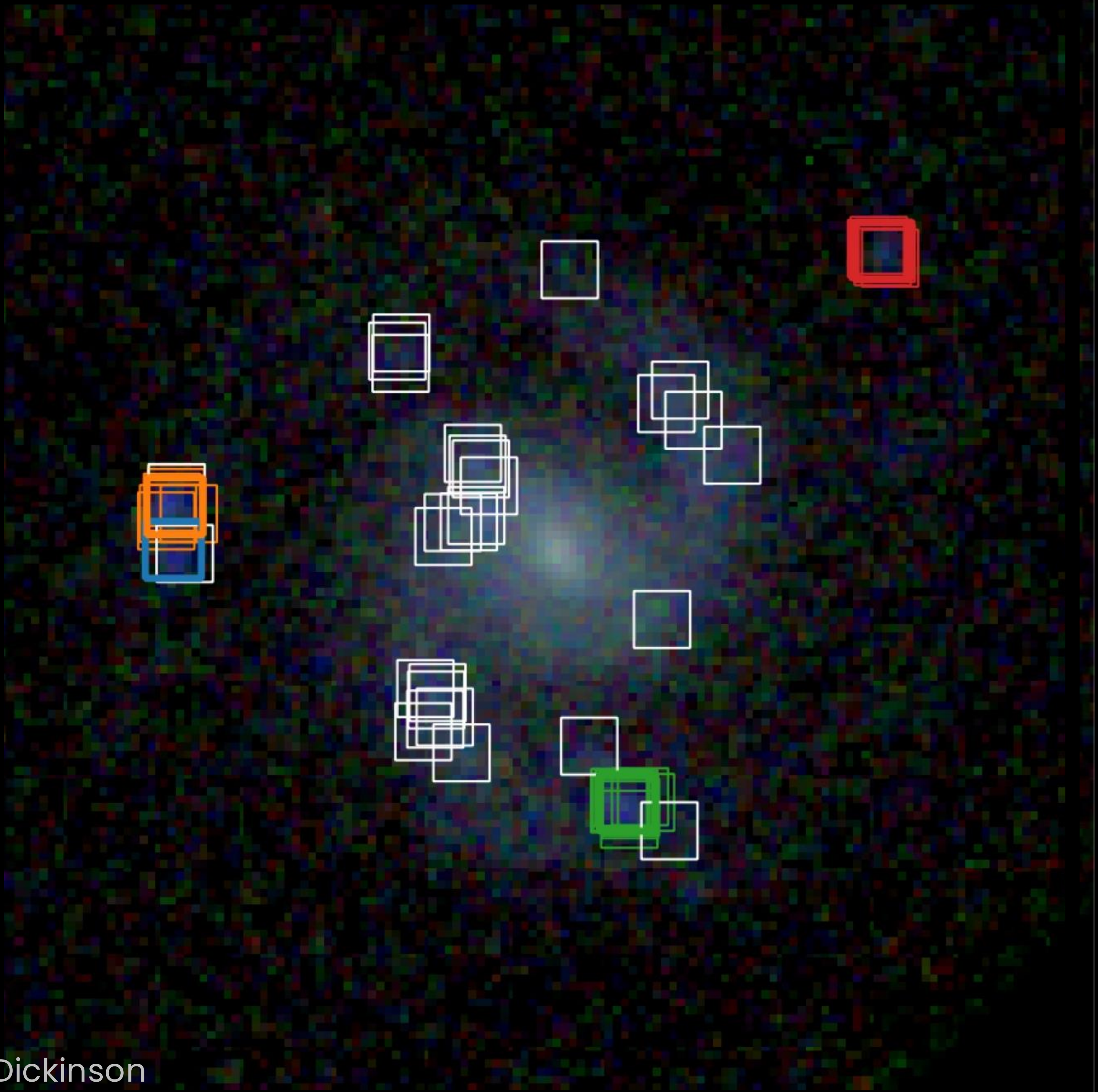
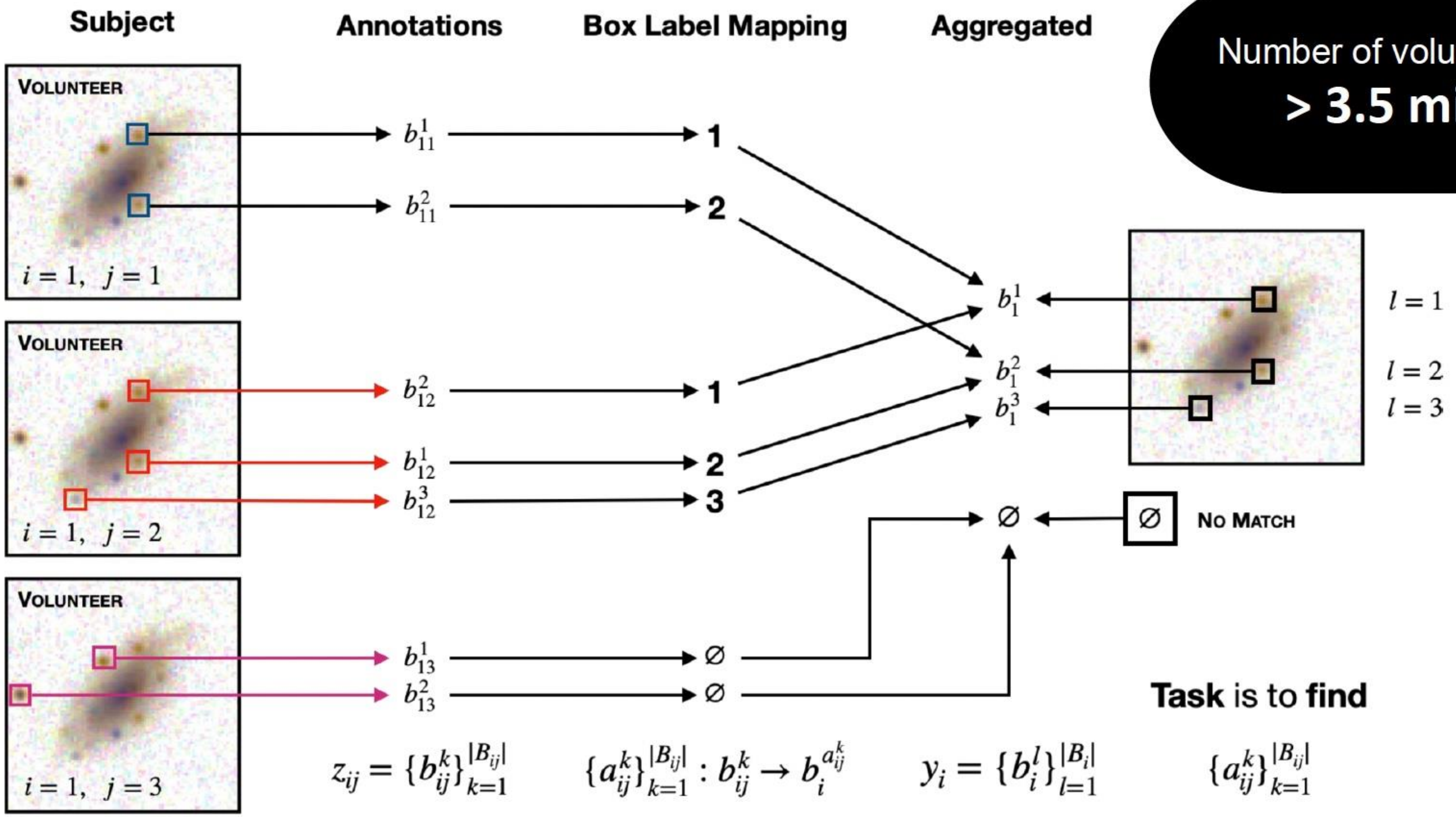


Image credit: Dr Hugh Dickinson



Number of volunteer clicks

> 3.5 million

github.com/ou-astrophysics/BoxAggregator

arXiv Dickinson et al (2022) - arxiv.org/abs/2210.03684

Image credit: Hugh Dickinson

Image credit: Dr Hugh Dickinson

Image credit: Dr Hugh Dickinson

Number of clumpy galaxies:

~35,000

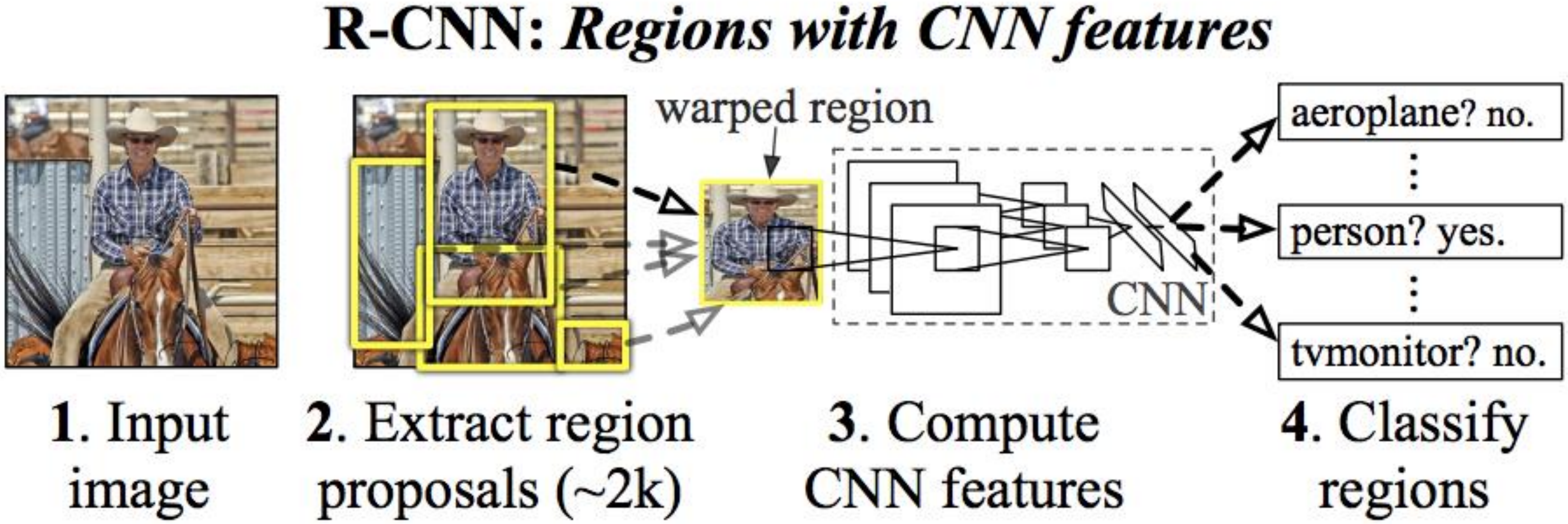
Number of potential clumps:

~100,000

First catalogue released!

arXiv Adams et al (2022) - arxiv.org/abs/2201.06581

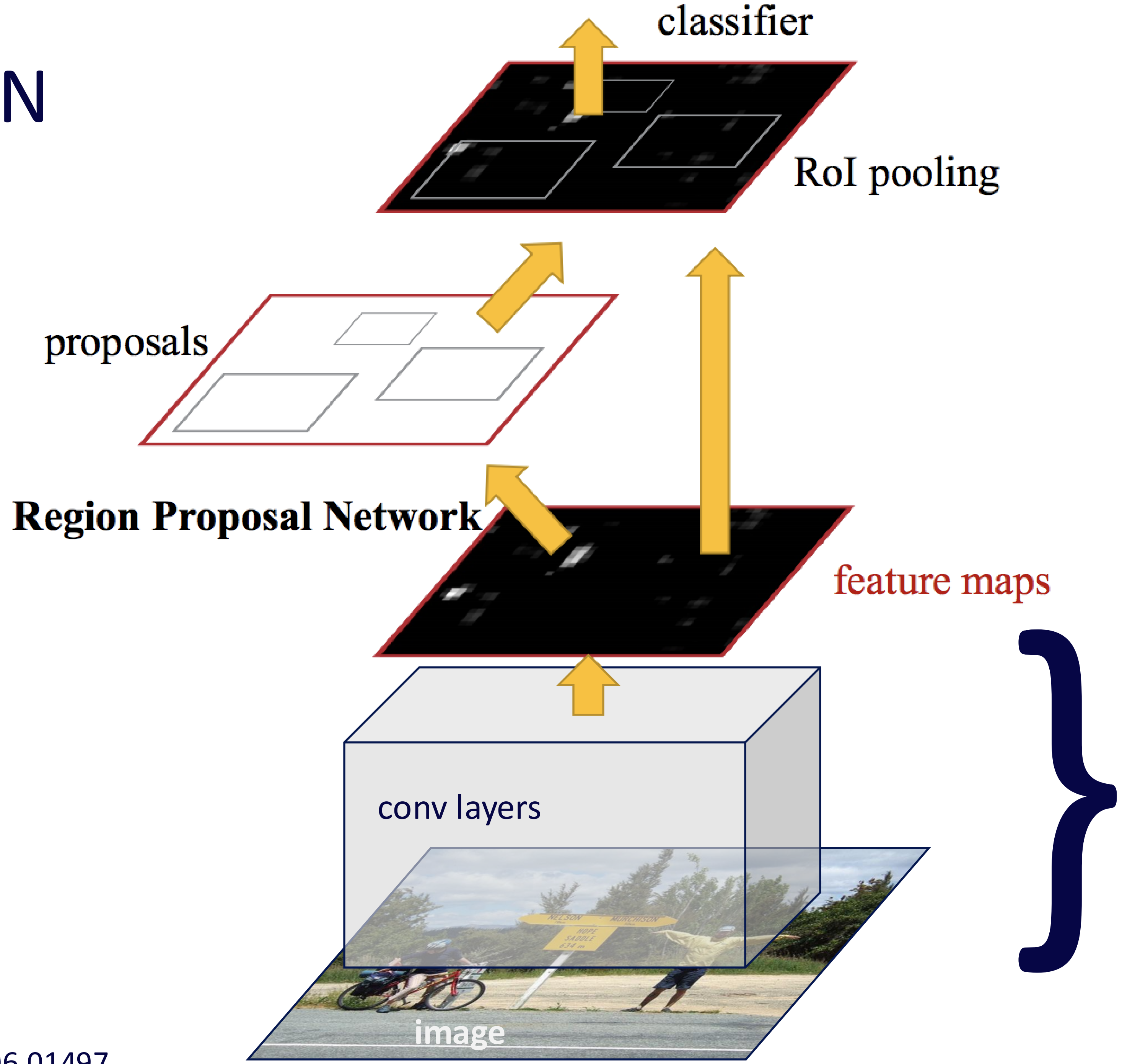
Image credit: Dr Hugh Dickinson



Girshick+13 arXiv:1311.2524

Faster R-CNN

Jürgen Popp



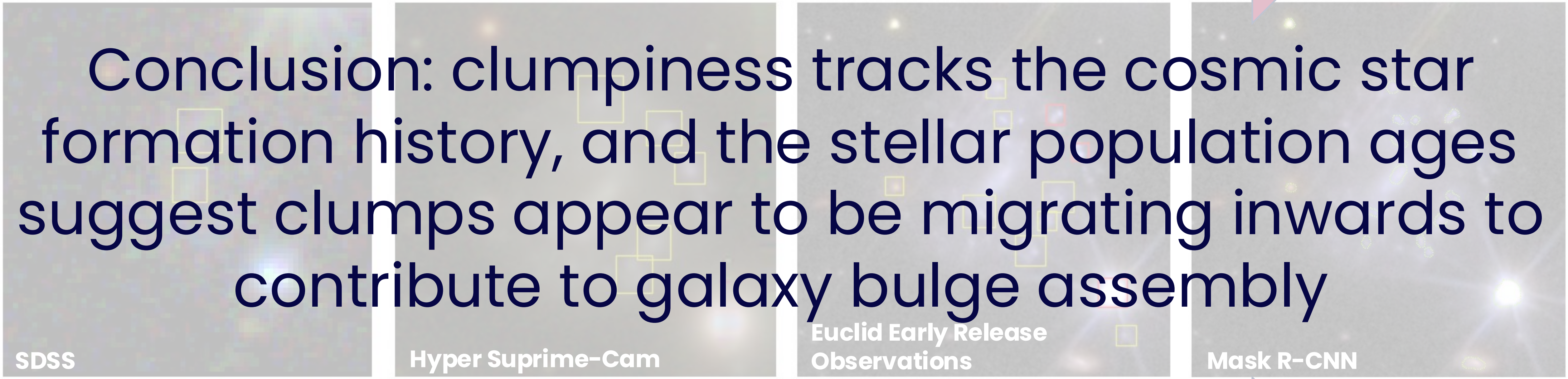
Zoobot

Adapted from Ren+15 arXiv:1506.01497

Region proposal: finding stellar nurseries

Model applications; Popp+24 RASTI 3, 174 and Popp+25 in prep.

Citizen science labels, to Faster RCNN, to Mask RCNN



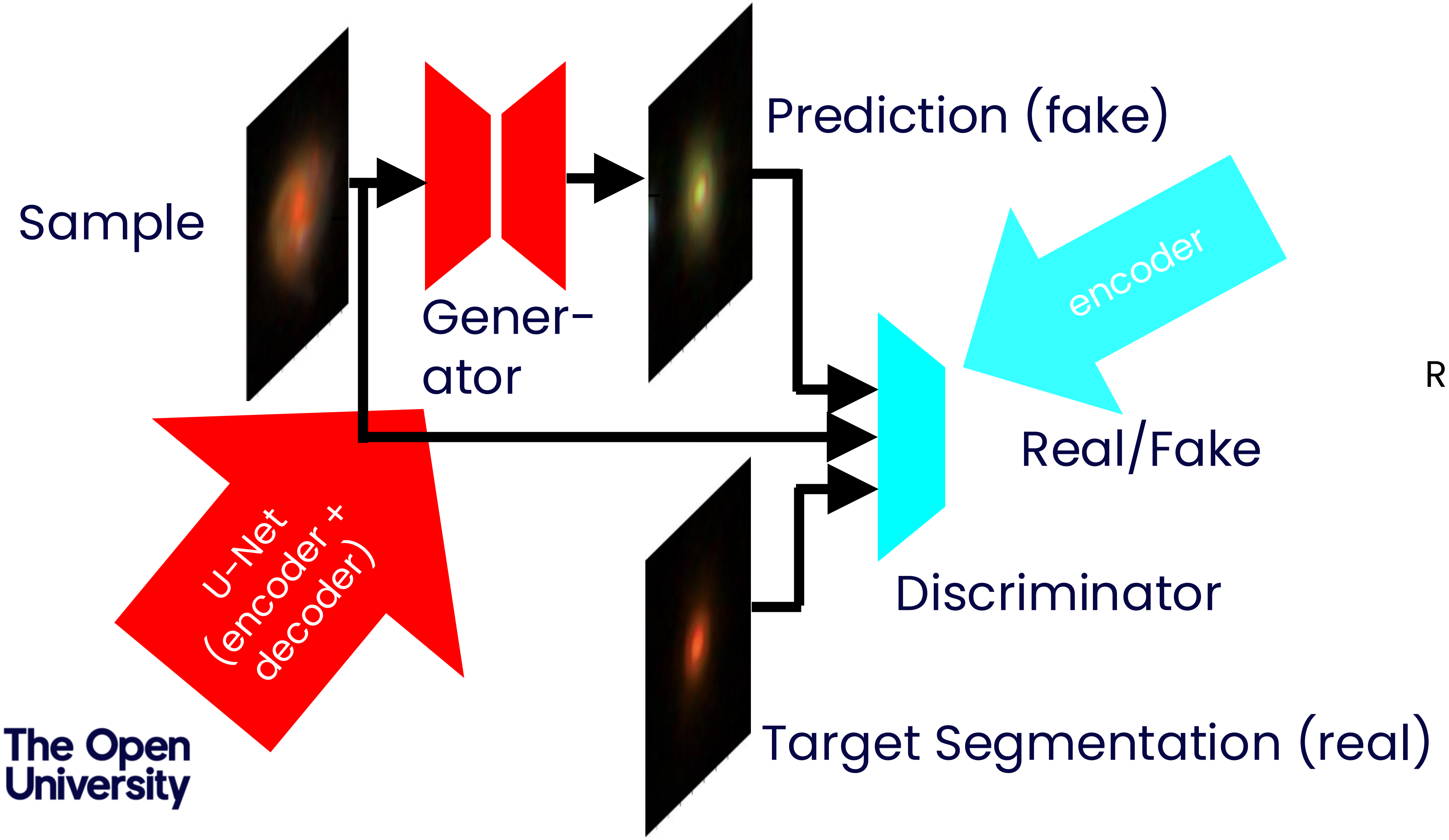
Conclusion: clumpiness tracks the cosmic star formation history, and the stellar population ages suggest clumps appear to be migrating inwards to contribute to galaxy bulge assembly

Ground-based data, better ground based, to space data with Euclid

The Open University

Image prediction for clumpy JWST galaxies

Conditional Generative Adversarial Networks (pix2pix): Pearce-Casey+23 RNAAS 7,217



Ruby Pearce-Casey

The Open University

Image prediction for clumpy JWST galaxies

Conditional Generative Adversarial Networks (pix2pix): Pearce-Casey+23 RNAAS 7,217

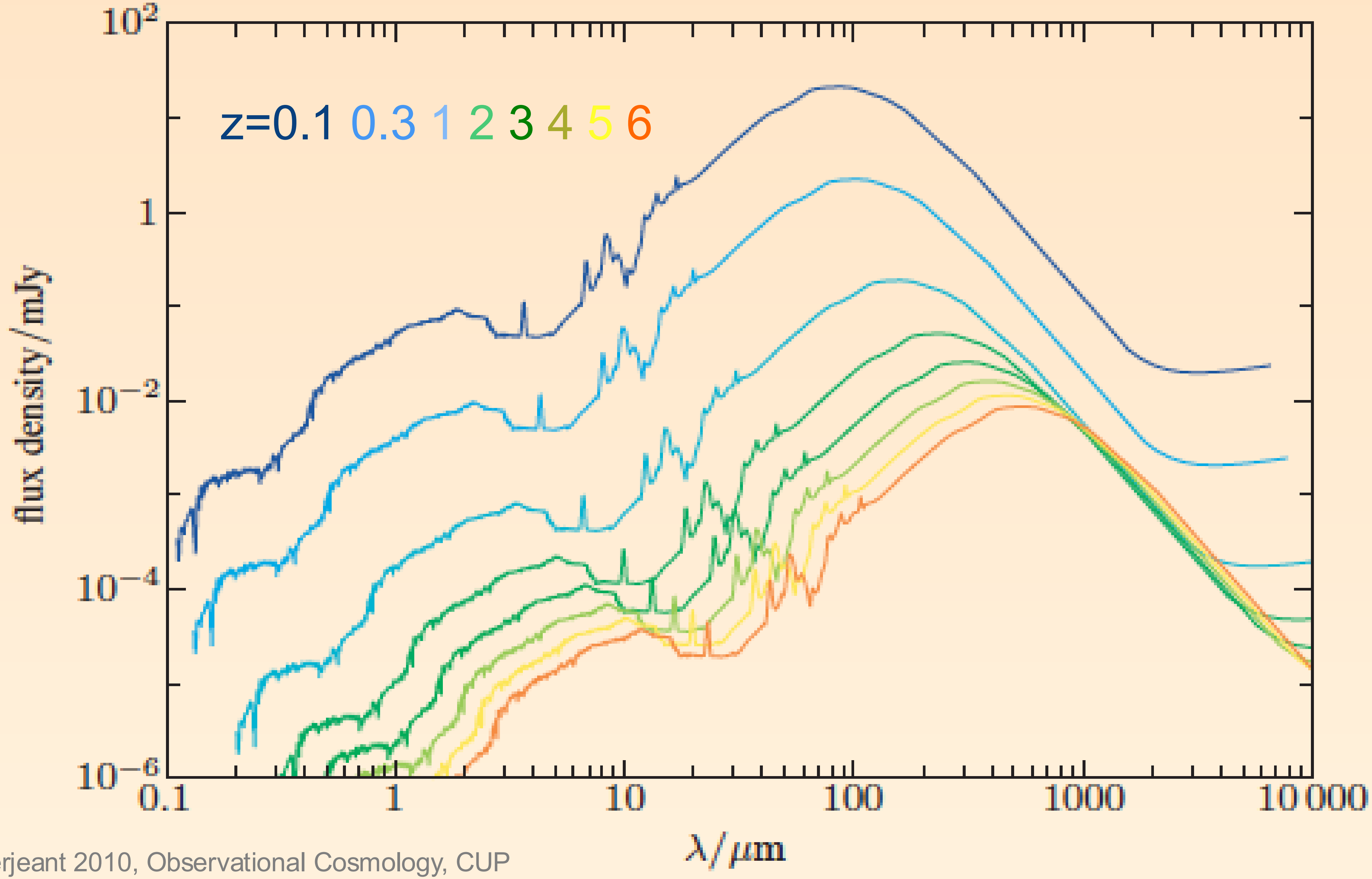
SW Channel [0.6-2.3μm]

Actual LW [2.4-5.0μm]

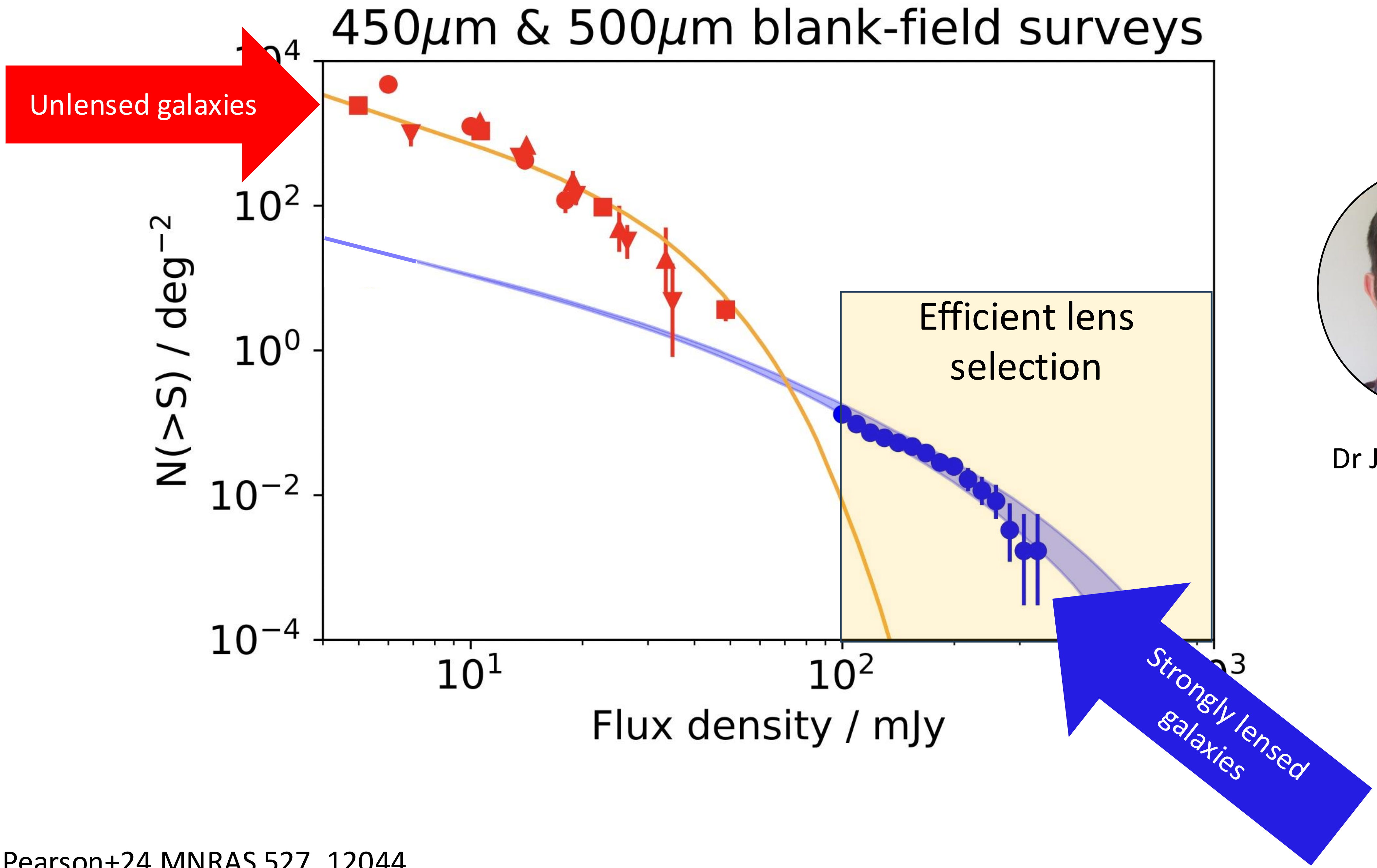
Generated LW

The Open University

The Open University



Serjeant 2010, Observational Cosmology, CUP



Dr James Pearson

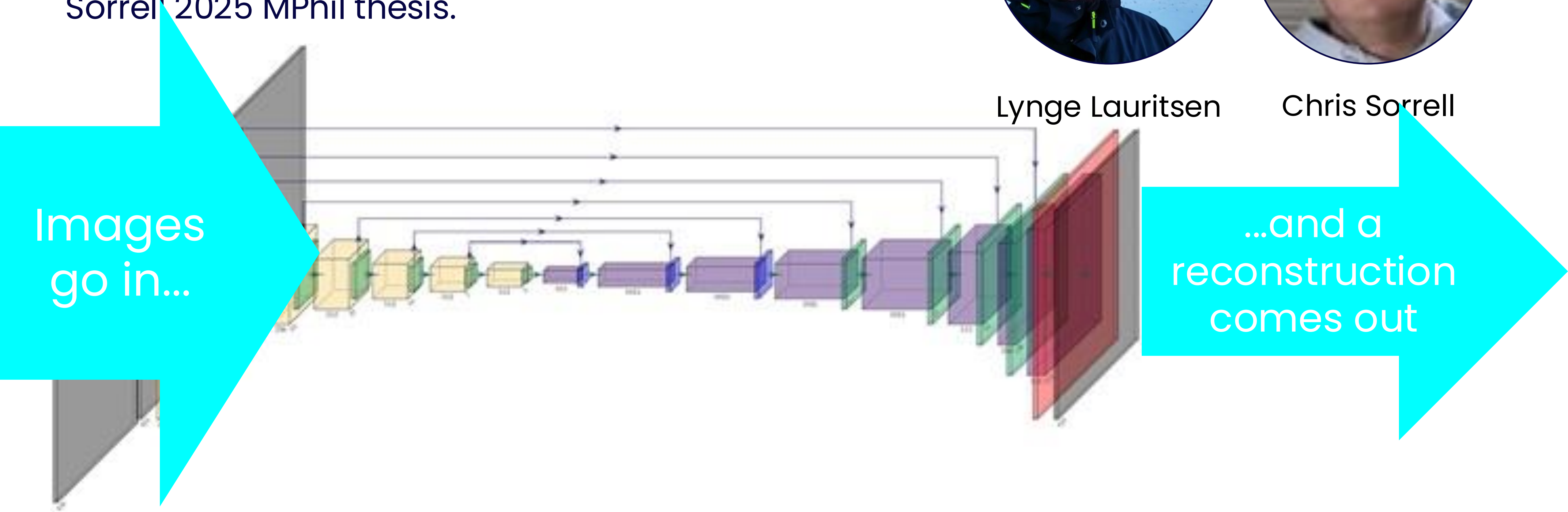
Pearson+24 MNRAS 527, 12044

Deconvolution

Denoising auto-encoder – Lauritsen+21, Donnellan+24, Sorrell 2025 MPhil thesis.

Lynge Lauritsen

Chris Sorrell



Convolutional layer

Sigmoid activation function

Deconvolutional layer

Combined BatchNorm and Leaky ReLu

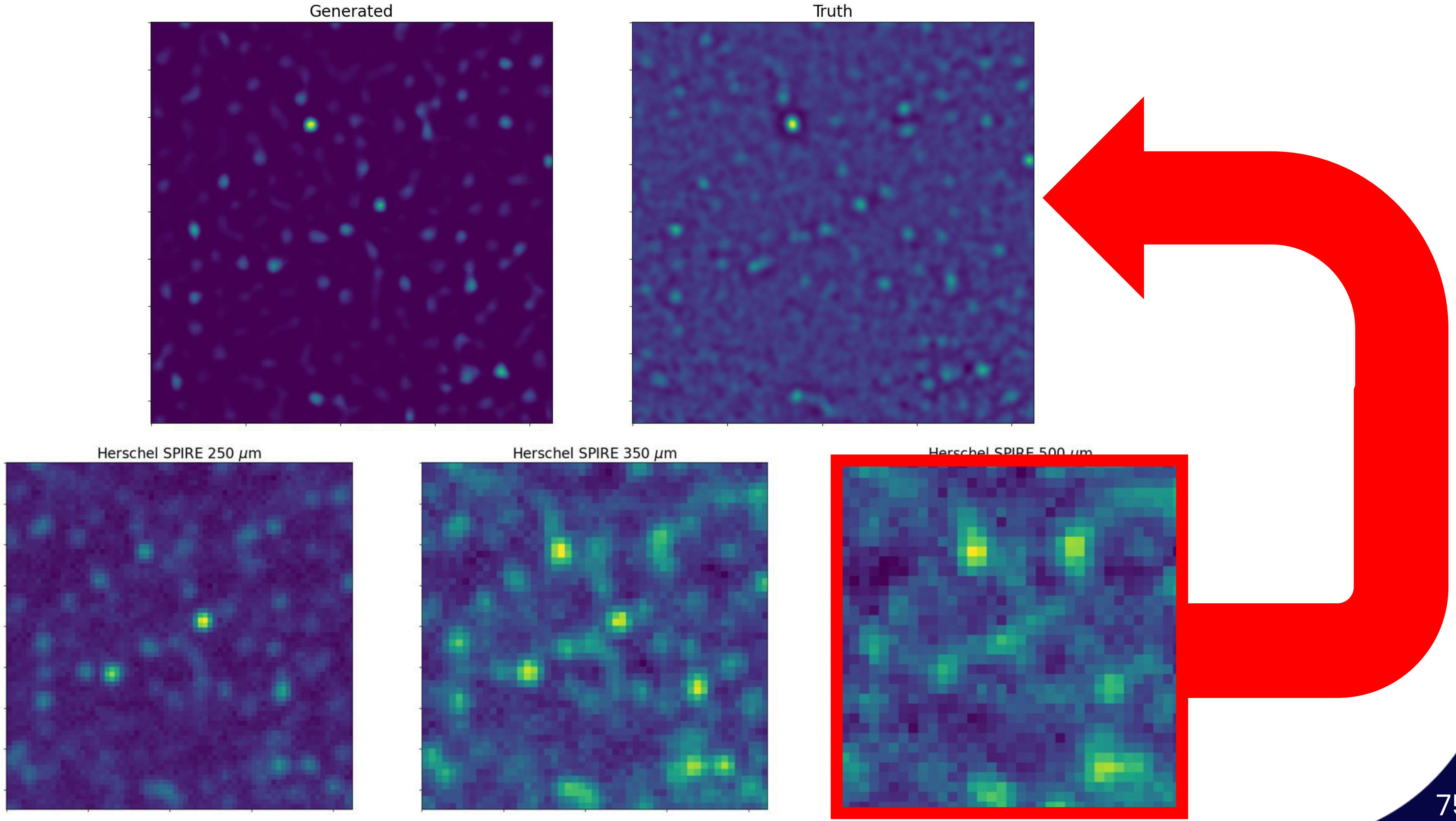
BatchNorm, dropout, and Leaky ReLu

Input and output images

The Open University

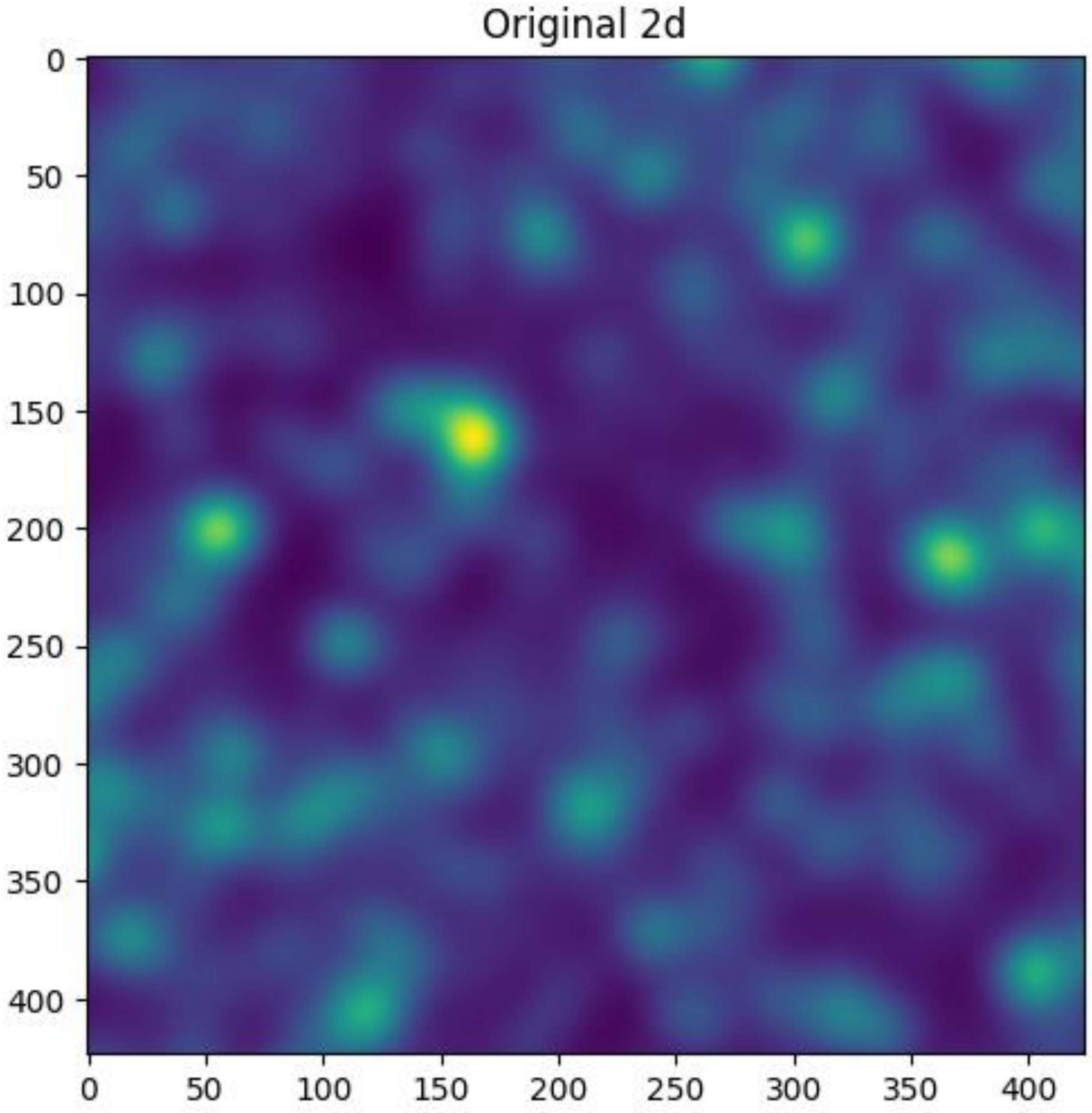
Deconvolution

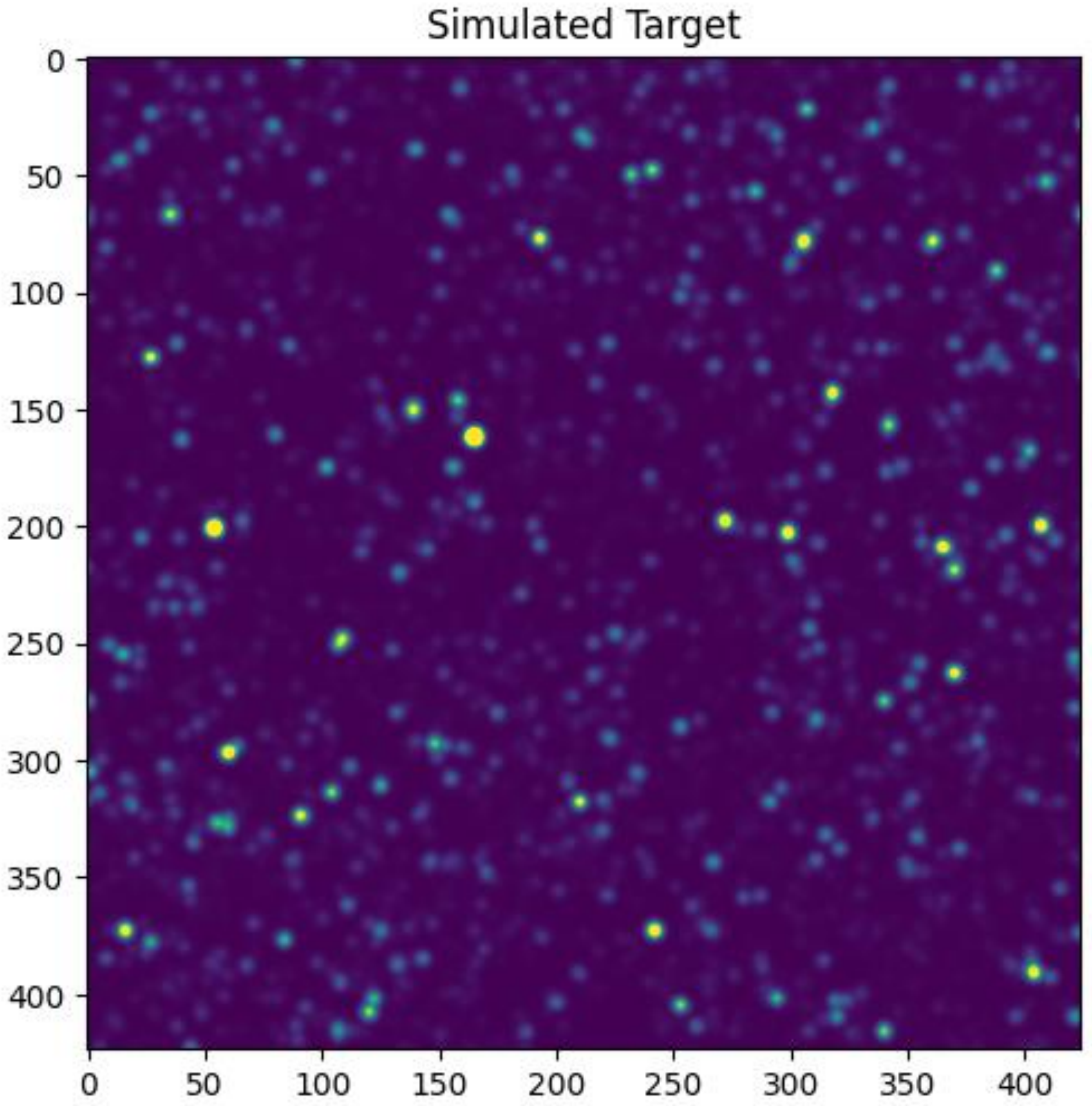
Denoising auto-encoder – Lauritsen+21, Donnellan+24, Sorrell+24 in prep.



Deconvolution

Simulated imaging of the proposed NASA PRIMA far-infrared observatory





1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	Citizen Science, Galaxy Zoo and Zoobot
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	**Infrared gravitational lenses and deconvolution**
6	Other astro ML, and Conclusions

1	My own first forays into machine learning
2	Astronomy's big data problem, gravitational lensing and convnets
3	Citizen Science, Galaxy Zoo and Zoobot
4	Clumpy galaxies: Zoobot, RCNNs and cGANs
5	Infrared gravitational lenses and deconvolution
6	**Other astro ML, and Conclusions**

Convolutional Neural Nets

Deep Dream

U-nets

Region proposal networks

De-noising autoencoders

Generative adversarial networks

Citizen science

Guided Grad CAM

t-SNE

UMAP

Image deconvolution: finding distant galaxies

Image prediction: finding rare galaxy populations

Image classification: finding warps in space and time

Image annotation: finding star-forming clumps in galaxies

The Open University

Selected recent publications

Davies, A., et al., 2019, Using convolutional neural networks to identify gravitational lenses in astronomical images, Monthly Notices of the Royal Astronomical Society, Volume 487, Issue 4, p.5263-5271

Dickinson, H., et al., 2022, Galaxy Zoo: Clump Scout - Design and first application of a two-dimensional aggregation tool for citizen science, Monthly Notices of the Royal Astronomical Society, Volume 517, Issue 4, pp.5882-5911

Donnellan, J., et al., 2024, Overcoming Confusion Noise with Hyperspectral Imaging from PRIMAger, Monthly Notices of the Royal Astronomical Society, in press (arXiv:2404.06935)

Lauritsen, L., et al., 2021, Super-resolving Herschel imaging: a proof of concept using Deep Neural Networks, Monthly Notices of the Royal Astronomical Society, Volume 507, Issue 1, pp.1546-1556

Pearce-Casey, R., et al., 2023, Using cGANs for Anomaly Detection: Identifying Astronomical Anomalies in JWST Imaging, Research Notes of the AAS, Volume 7, Issue 10, id.217

Pearce-Casey, R., et al., 2025, Euclid: Searches for strong gravitational lenses using convolutional neural nets in Early Release Observations of the Perseus field, Astronomy & Astrophysics 696, A214

Popp, J., et al., 2023, Transfer learning for galaxy feature detection: Finding giant star-forming clumps in low-redshift galaxies using Faster Region-based Convolutional Neural Network, RAS Techniques and Instruments, Volume 3, Issue 1, pp.174-197

Serjeant, S., et al., 2024, Citizen science in European research infrastructures, The European Physical Journal Plus, Volume 139, Issue 5, article id.418

Serjeant, S., 2023, Citizen Science in the European Open Science Cloud, Europhysics News, Volume 54, Issue 2, 2023, pp.20-23

Wilde, J., et al., 2022, Detecting gravitational lenses using machine learning: exploring interpretability and sensitivity to rare lensing configurations, Monthly Notices of the Royal Astronomical Society, Volume 512, Issue 3, May 2022, Pages 3464–3479





Dr Jane Bromley

Jürgen Popp

Chris Sorrell

Ruby Pearce-Casey

Dr Laura Hunt

Prof Stephen Serjeant

Thank you




Dr Hugh Dickinson

Dr James Pearson

Dr Josh Wilde

Dr Lynge Lauritsen

Funded by the European Union

The Open University

"**ELSA: Euclid Legacy Science Advanced analysis tools**" (Grant Agreement no. 101135203) and **ACME: Astrophysics Center for Multimessenger studies in Europe** (Grant Agreement no. 101131928) are **funded by the European Union.** UK participation is funded through the UK Horizon guarantee scheme under Innovate UK grant 10093177. **ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures** has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 824064. **ASTERICS** is a project supported by the European Commission Framework Programme Horizon 2020 Research and Innovation action under grant agreement n. 653477. **REINFORCE** has received funding from the European Union's Horizon 2020 project call H2020-SwafS-2018-2020 under Grant Agreement no. 872859. The **EOSC Future** project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020 - Grant Agreement Number 101017536. The **OSCARS project** has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751. **Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union** or Innovate UK. **Neither the European Union nor the granting authority can be held responsible for them.** We thank the Science and Technology Facilities Council for support under grant ST/T506321/1.