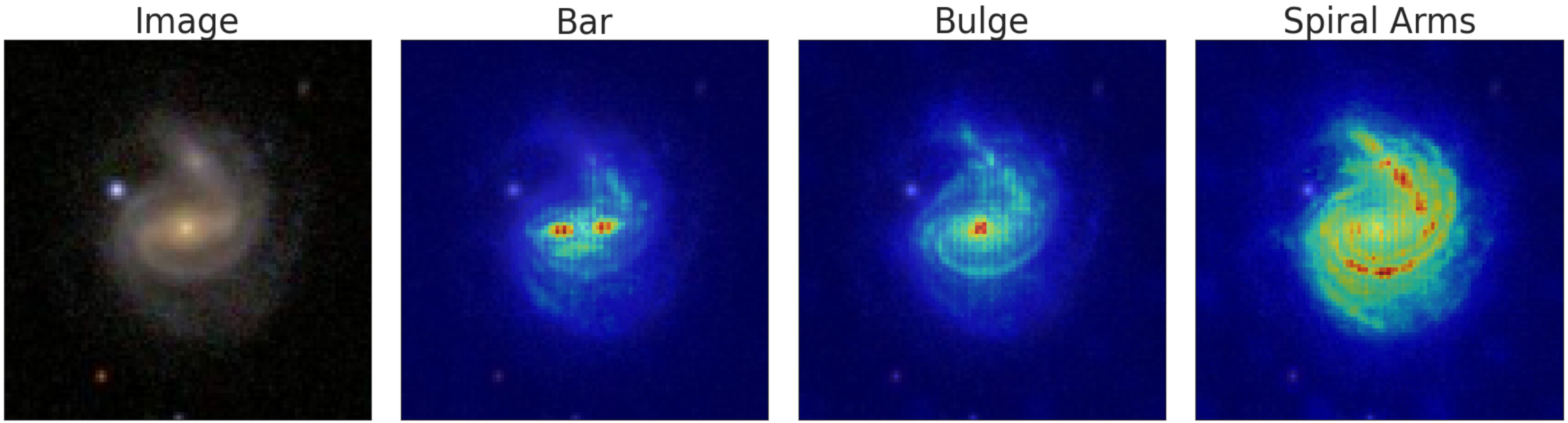
`Deep' vs. `Shallow' Learning in Galaxy Surveys

Ofer Lahav

University College London



Outline

- What is the gain in using Deep Learning vs. conventional Shallow Learning?
- How to understand/explain/interpret Deep Learning?

Test-cases, motivated by

SDSS, DES, KiDS, HSC, Euclid, Rubin-LSST,…

- XAI of galaxy morphology (Bhambra et al. 2110.08288 – today!)
- Photo-z from full images (Henghes et al. 2109.02503)
- Benchmarking and scalability (Henghes et al. 2104.01875)

UCL CDT DIS

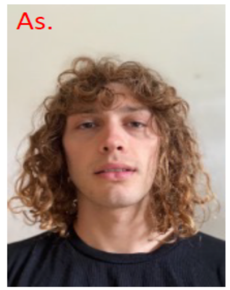
UCL CDT in Data Intensive Science

http://www.hep.ucl.ac.uk/cdt-dis/

UCL

DiRAC
PRIVITAR
The Economist Group
BBC
Hartree Centre
Science & Technology Facilities Council
QUANTUMBLACK
A MCKINSEY COMPANY
OCF
ASI DATA SCIENCE
Met Office
Particle Physics
Rutherford Appleton Laboratory
Lenovo
The Alan Turing Institute
BLUE SKIES SPACE
Transport for London
Science & Technology Facilities Council
Scientific Computing Department
asos
Quantemol
PGS
CERN openlab
UK Atomic Energy Authority
Mellanox TECHNOLOGIES
nccgroup
European Bank
for Reconstruction and Development



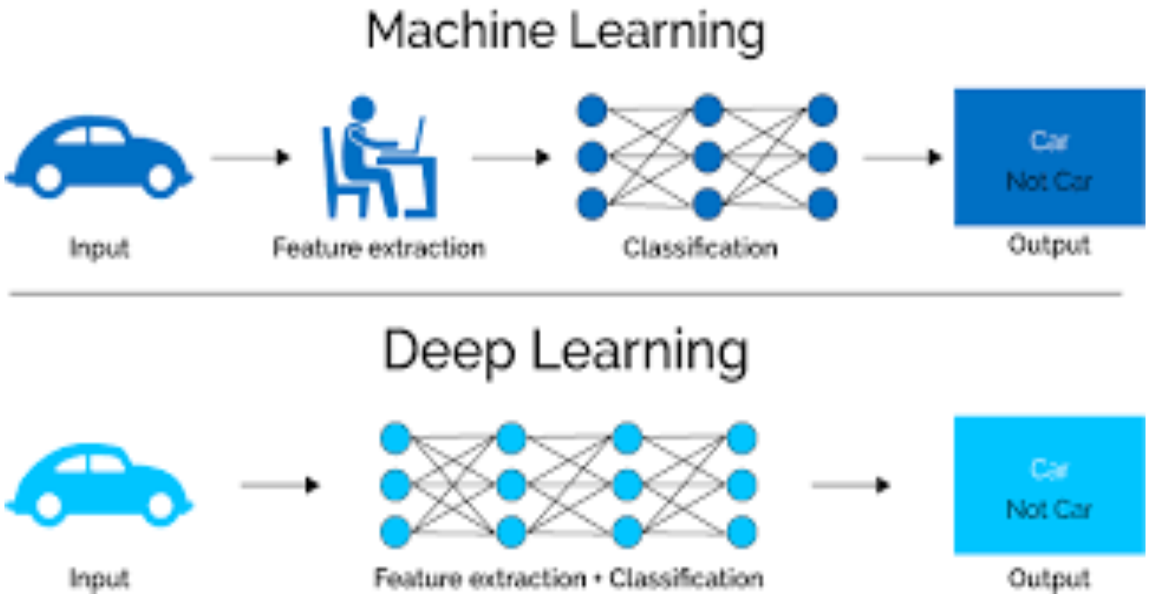



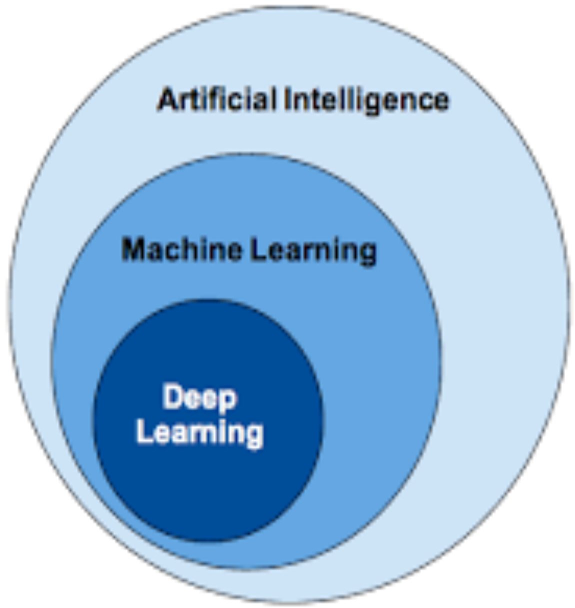
52 PhD students in five (2017-2021) cohorts
in a 4-yr programme (including 6m in AI industry)
(+ a spin-off UCL-Jordan programme)

Astro papers on the arXiv with `Deep Learning' in the title

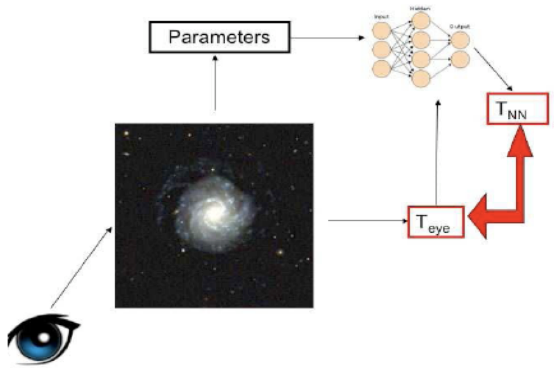
#Papers per year

2017 (23), 2018 (35), 2019 (83), 2020 (90), 2021 (81)



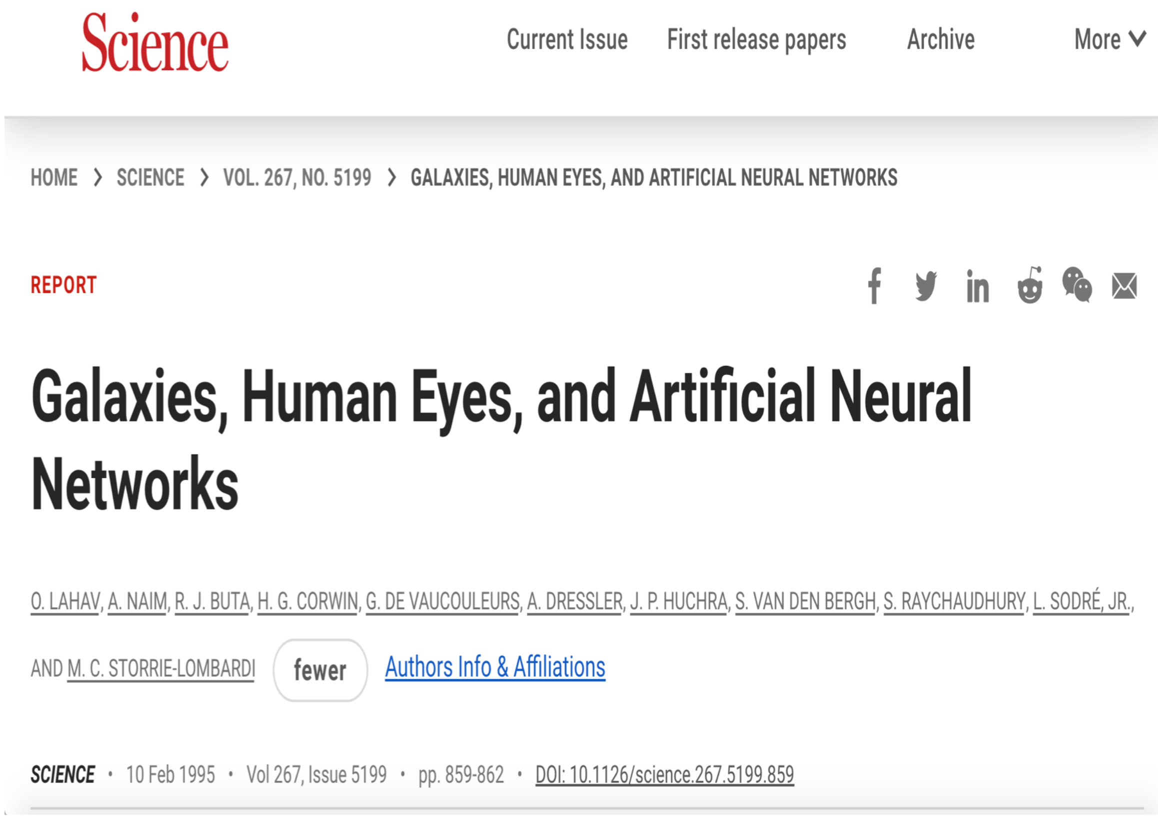


Galaxy classification with ANN (1990's)



- 840 APM galaxy images
- 6 'gurus'
- ANN reproduced the human classification to rms of 2 on the deV T-system [-5 to 10]

- Further applications to Galaxy Zoo etc.


Science Current Issue First release papers Archive More

HOME › SCIENCE › VOL. 267, NO. 5199 › GALAXIES, HUMAN EYES, AND ARTIFICIAL NEURAL NETWORKS

REPORT

Galaxies, Human Eyes, and Artificial Neural Networks

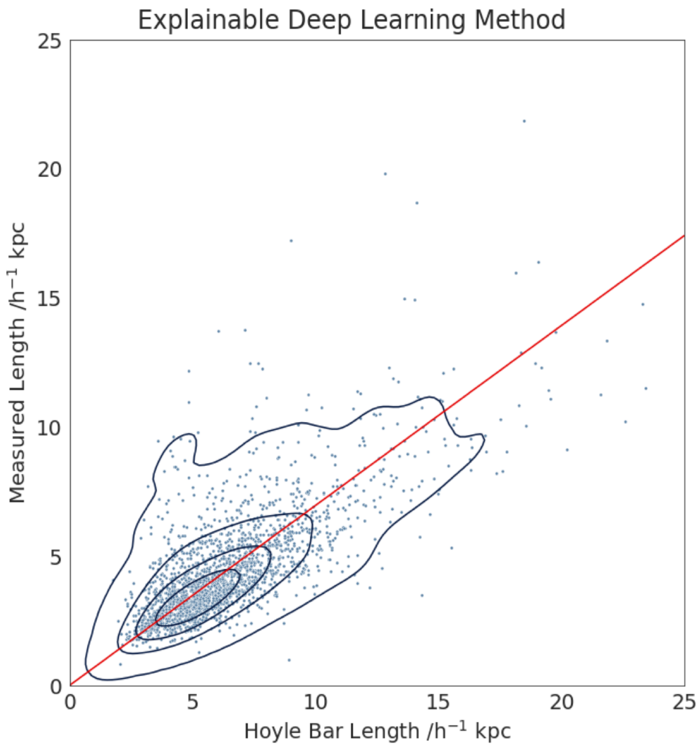
O. LAHAV, A. NAIM, R. J. BUTA, H. G. CORWIN, G. DE VAUCOULEURS, A. DRESSLER, J. P. HUCHRA, S. VAN DEN BERGH, S. RAYCHAUDHURY, L. SODRÉ, JR., AND M. C. STORRIE-LOMBARDI fewer Authors Info & Affiliations

SCIENCE • 10 Feb 1995 • Vol 267, Issue 5199 • pp. 859-862 • DOI: 10.1126/science.267.5199.859

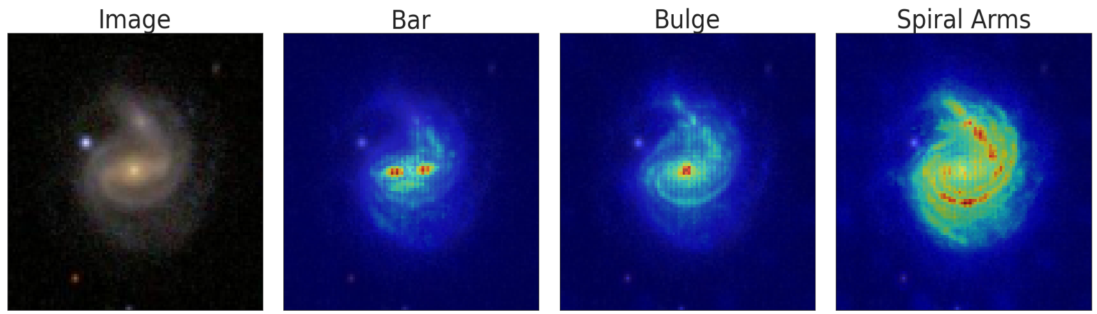
OL, Naim et al. (Science, 1995)
Banerji, OL et al. (2010)

Explaining Galaxy Morphology with Saliency Mapping

XAI vs. Galaxy Zoo cataloged bar length



Bhambra, Joachimi, OL
arXiv:2110.08288

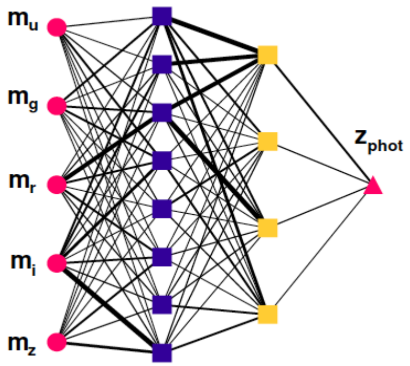


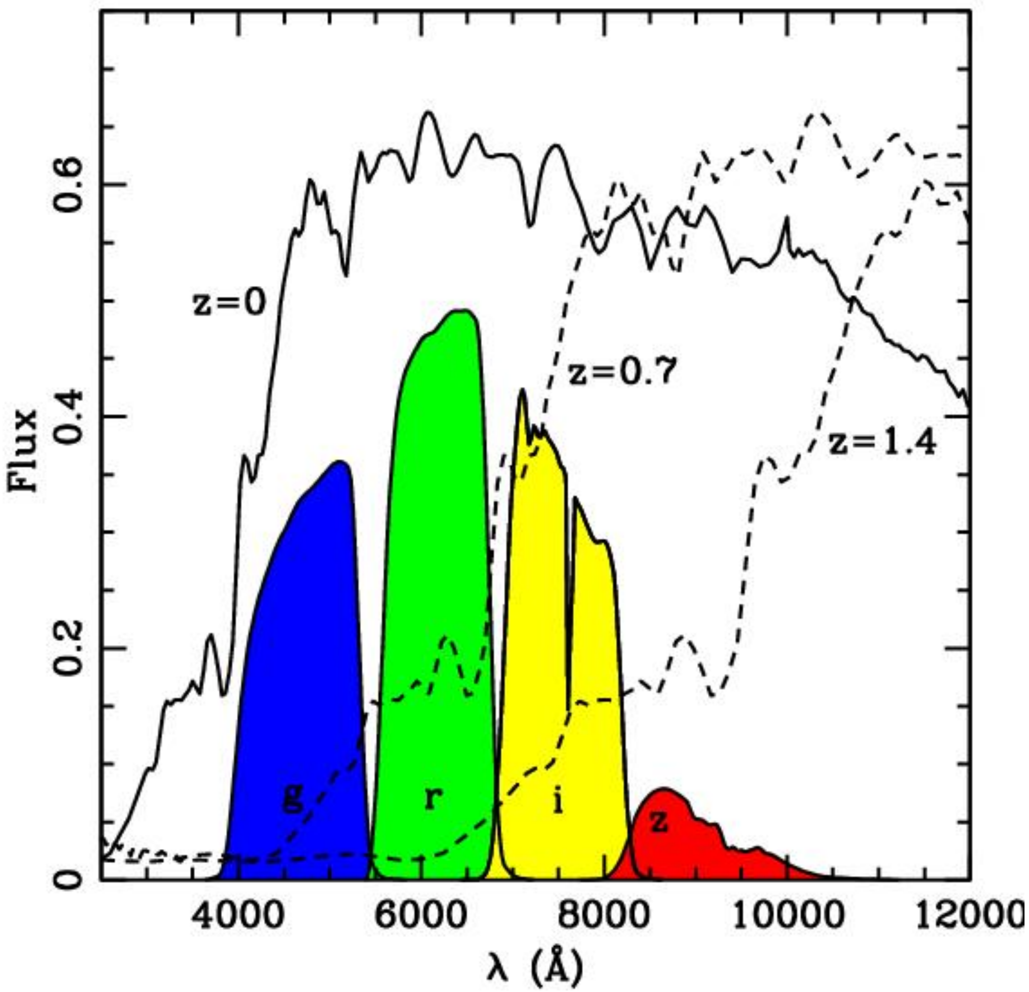
SmoothGrad: Calculate in each pixel the (smoothed) gradient of the score per class y^c wrt the pixel intensity x.
(note the internal architecture is bypassed.)

$$L^c(x) = \frac{\partial y^c}{\partial x}$$

Photometric redshift

Difference in flux through filters as the galaxy is redshifted



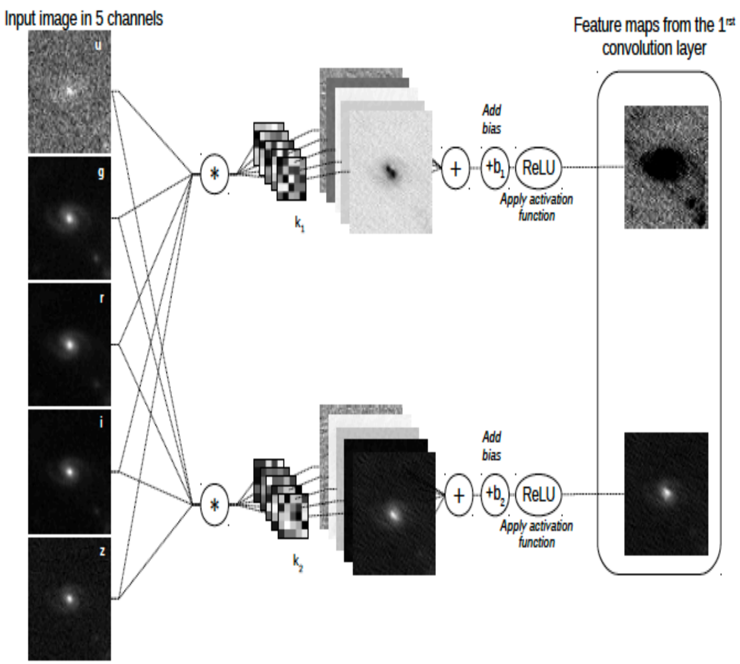


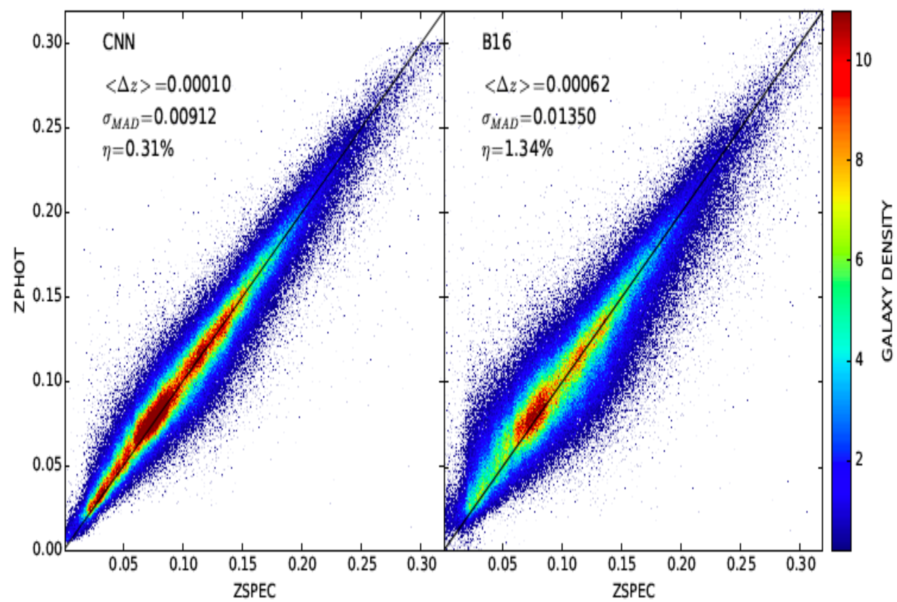
$z = f\,(m_1, m_2, \ldots)$

ANNz (Collister & OL 2004)
ANNz2 (Sadeh et al. 2016)

A dozen or so templae and ML methods are now available

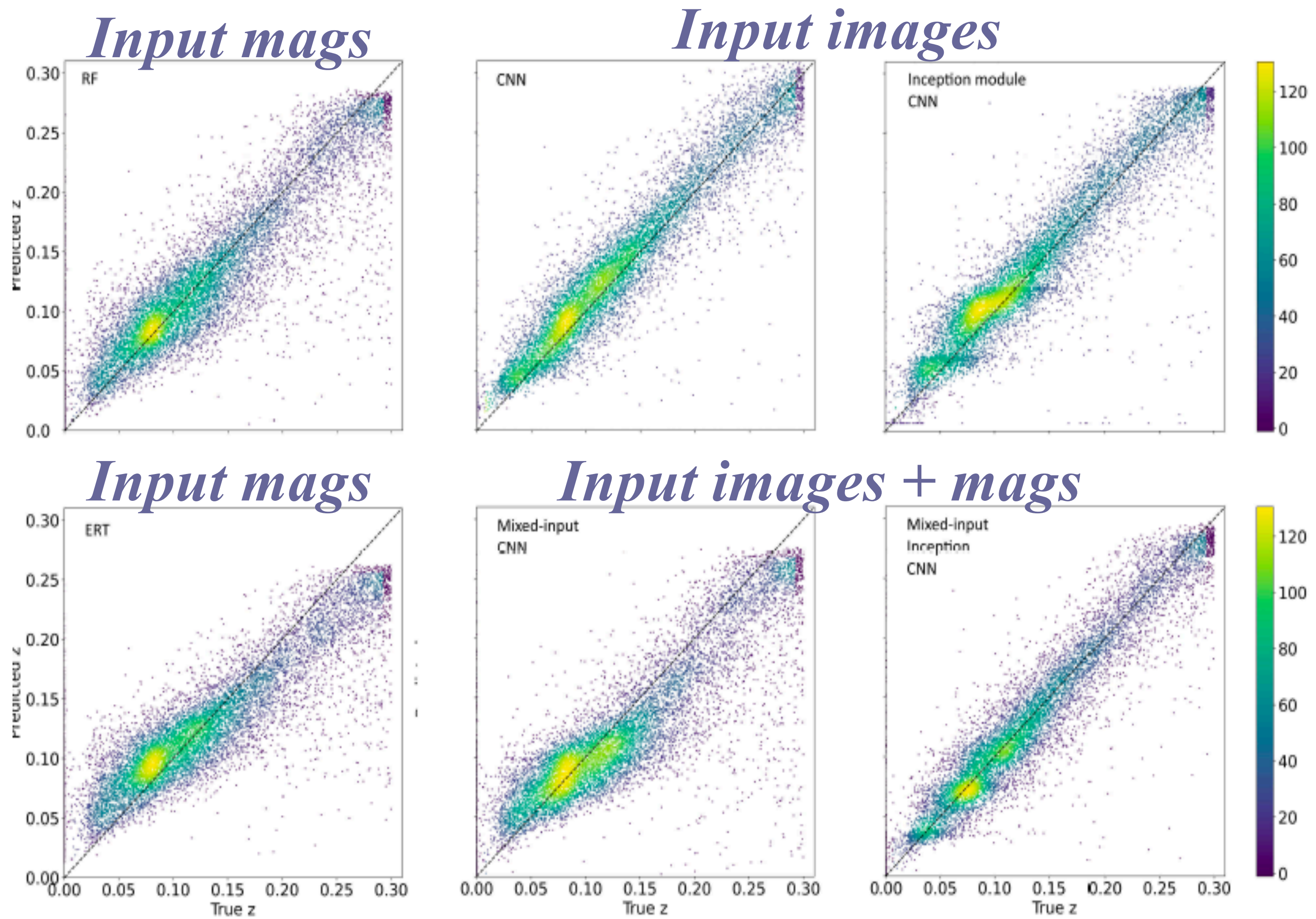
Photo-z from SDSS full images using CNN





Pasquet, Bertin, Treyer et al. (2019)

Photo-z from 1M SDSS full images: factor 2 improvement in MSE

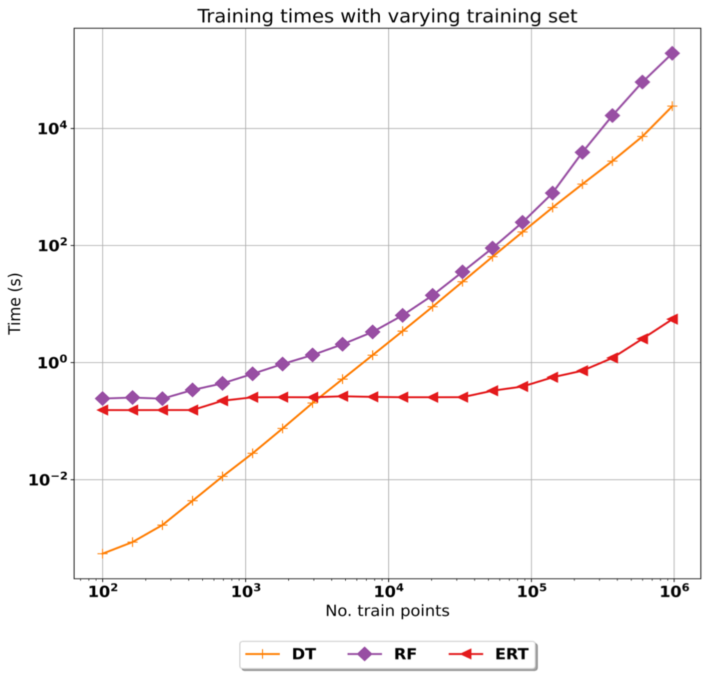


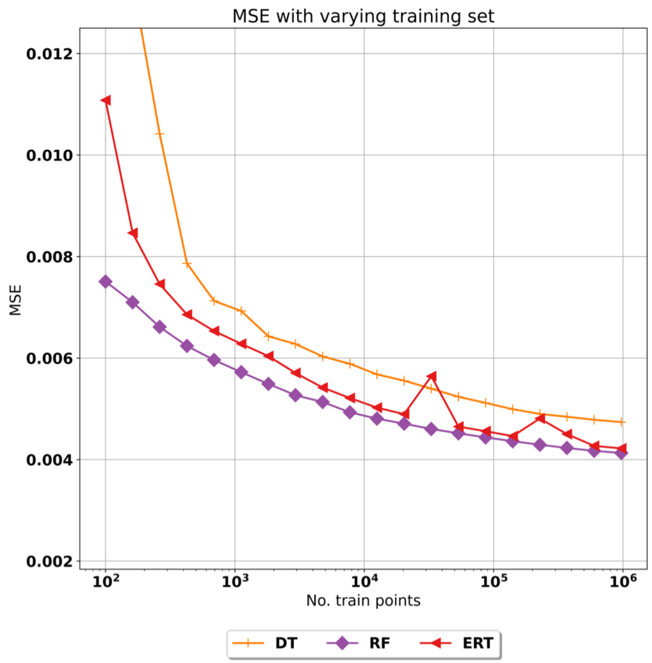
Henghes, Pettitt, Thiyagalingam, Hey, OL (2109.02503)

Benchmarking and Scalability of ML for Photo-z using 1M SDSS galaxies (ugriz filters)

a small concession of error can allow for a great improvement in efficiency.

Training/Testing times vs. accuracy





cf. BASE-ExCALIBUR Exa-scale

- Random Forest is best for MSE, but the slowest to train.

- Extremely Randomised Trees could be trained ~100x faster, with similar MSE.

Henghes et al. arXiv:2104.01875

AI/ML for galaxy surveys

- Benchmarking: assessing up-scaling of of ML algorithms to exa-scale
- Deep Learning from entire images improves photo-z
- Understanding/explaining/interpreting galaxy morphology

Challenges:

- Enhancing ‘deep’ vs. ‘shallow’ performance.
- Incorporating known Physics in the input and getting out new Physics
- Training the next generation of PhDs and Post-docs, beyond academia

“Only if you know how to make money without ML
you can make money with ML.”
(The Economist, 1990s)

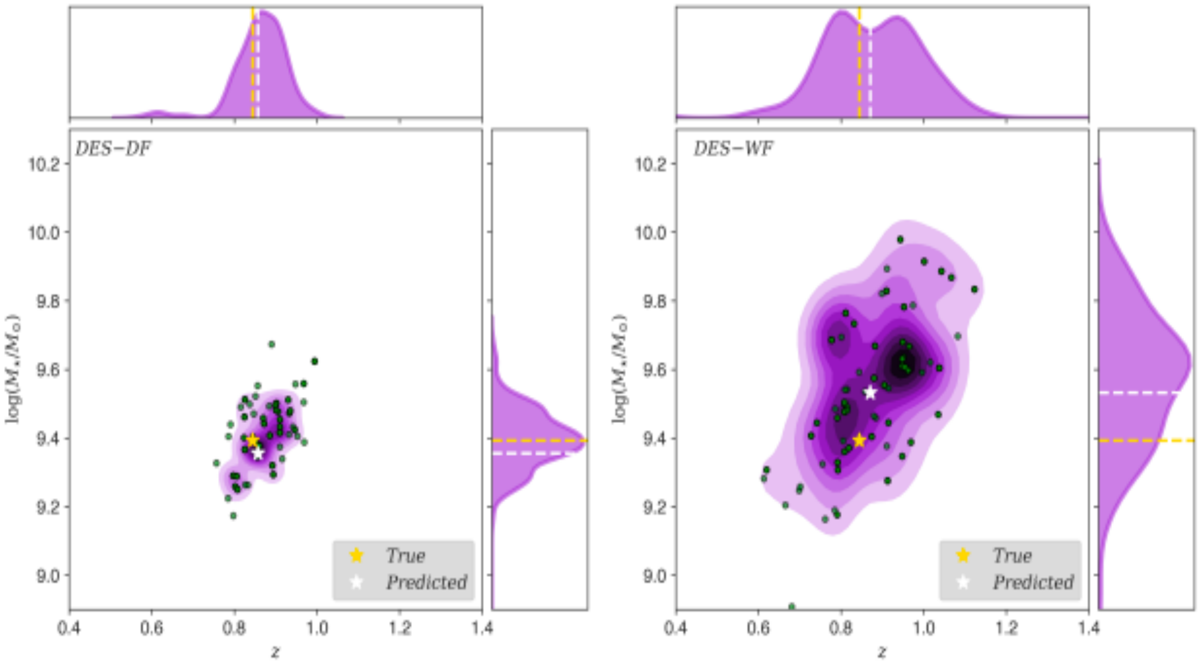
Extra Slides

Big Data in Astronomy

Survey	Data per night/day	Galaxies	Cost	Scientists
DES	1 TeraB	~300 Million	~$40M	~400
DESI	40 GigaB	~35 Million	~$70M	~600
Rubin-LSST	15 TeraB	~Billions	~$1.0B	~1000
Euclid	850 GigaB	~Billions	~$1.5B	~1500
SKA	1 PetaB	~Billions	~$1.3B	~1000

Joint pdf (photo-z, stellar mass) with Machine Learning (Random Forest) using DES (Cosmos) Deep Field and Wide Field

Stellar mass



Redshift

Mucesh, Hartley, Palmese, OL et al. (*2012.05928*)

Gold-true; White-predicted