SummerSOC 2025

Crete, Greece

Uncertainty-Aware Machine Learning for Astronomical Data Analysis

Gregory Tsagkatakis

Institute of Computer Science, FORTH
Computer Science Department, University of Crete

Funded by the European Union

SPL at a glance

2006

4 Researchers/Academics (permanent)
1 Postdoctoral Researchers
12 Postgraduate Students
2 Research Engineers

> 10M€ (PI/co-PI in EU & national projects)

FORTH

ICS + 8 Institutes

SPL + 7 Labs

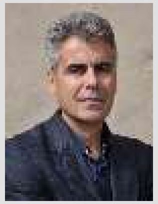
Research Institutes

Electronic Structure & Laser
Molecular Biology & Biotechnology
Computer Science
Applied and Computational Mathematics
Mediterranean Studies
Chemical Engineering Sciences
Astrophysics
Geoenegy
Biomedical Research

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

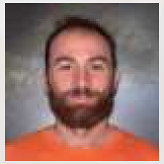
Collaborators & Funding

Panos Tsakalides
Signal Processing lab
FORTH



Jean-Luc Starck
CosmoStat lab
CEA, France

Mahta Moghaddam
MiXIL lab
USC

TITAN
ARTIFICIAL INTELLIGENCE
IN ASTROPHYSICS

RESEARCH EXECUTIVE AGENCY

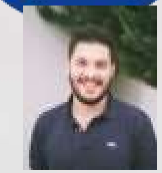
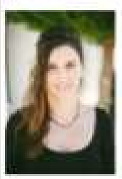
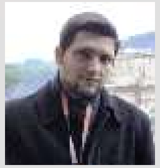
CALCHAS
Computational Intelligence for Multi-Source Remote Sensing Data Analytics

MARIE CURIE ACTIONS
European Commission | Horizon 2020 European Union funding for Research & Innovation

esa

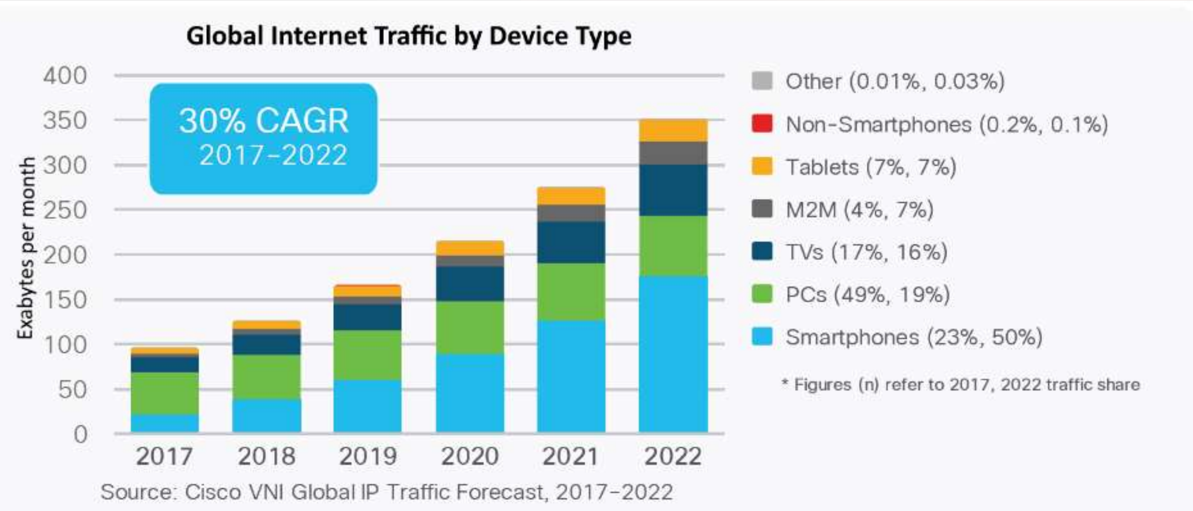
NASA

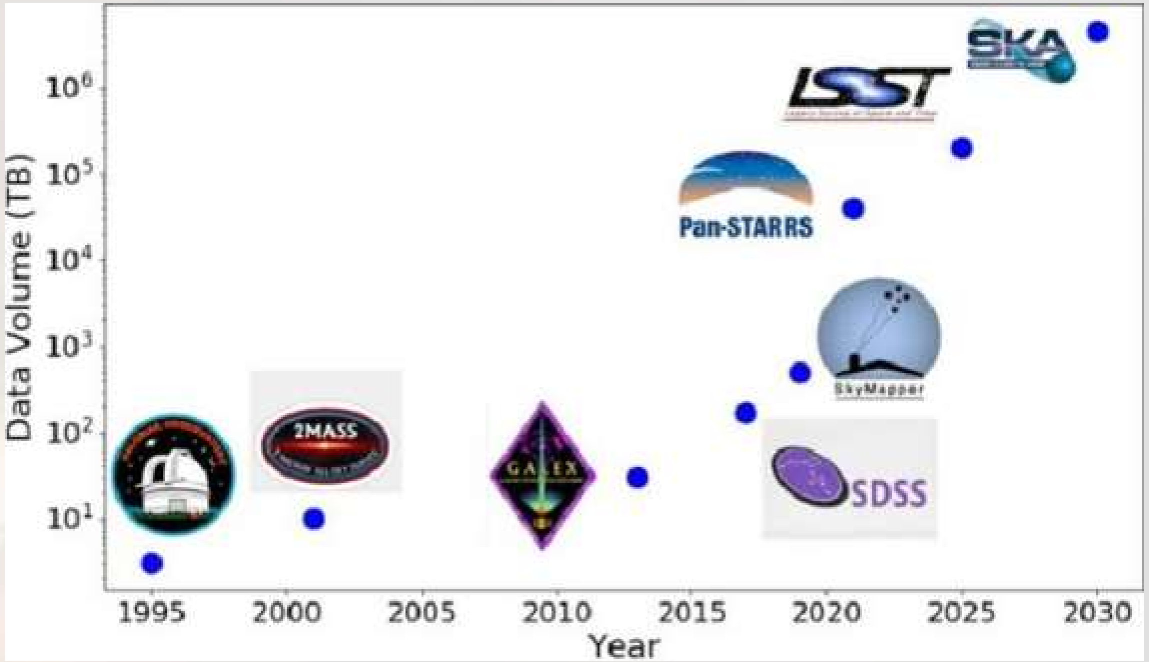
H.F.R.I.
Hellenic Foundation for Research & Innovation

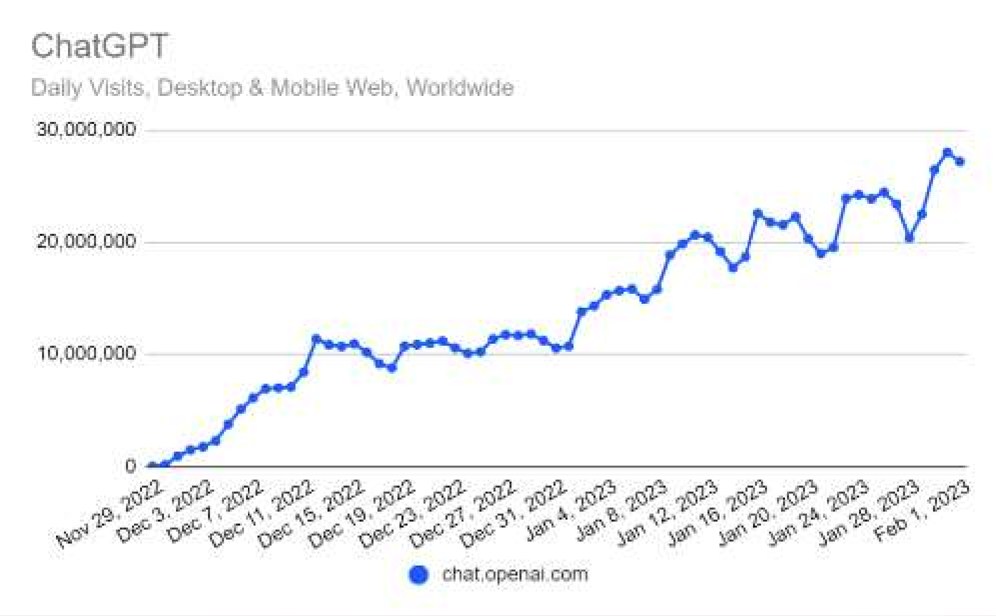
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS | FORTH INSTITUTE OF COMPUTER SCIENCE | Signal Processing Laboratory | COSMOSTAT | cea | Funded by the European Union

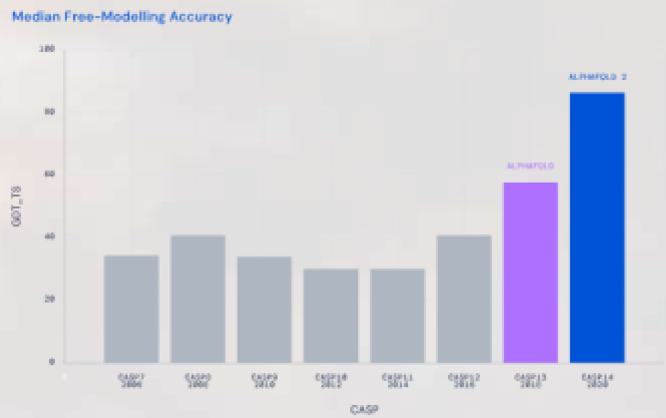
The Big Data Revolution





The AI Revolution

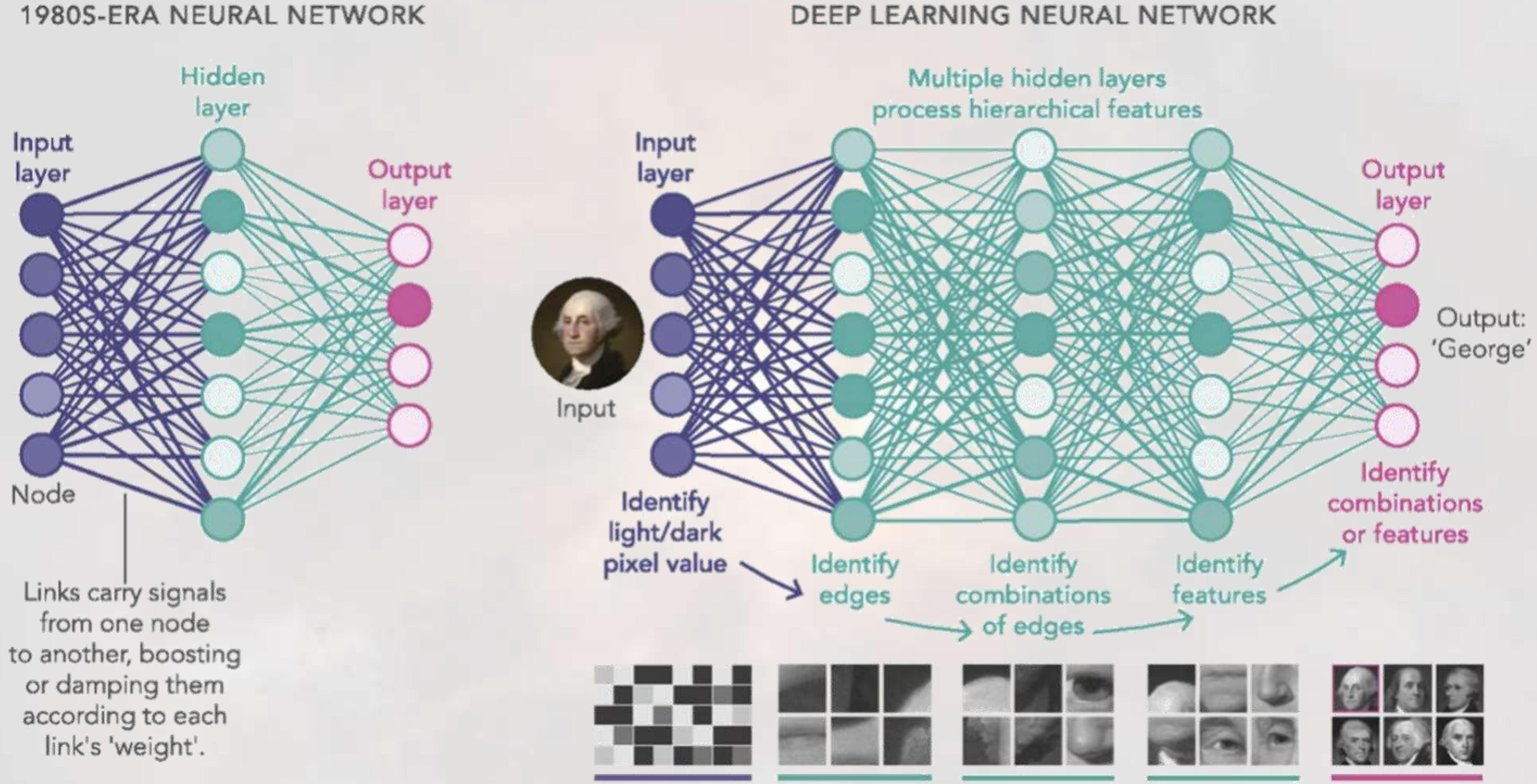




Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." *Nature.* 2021.

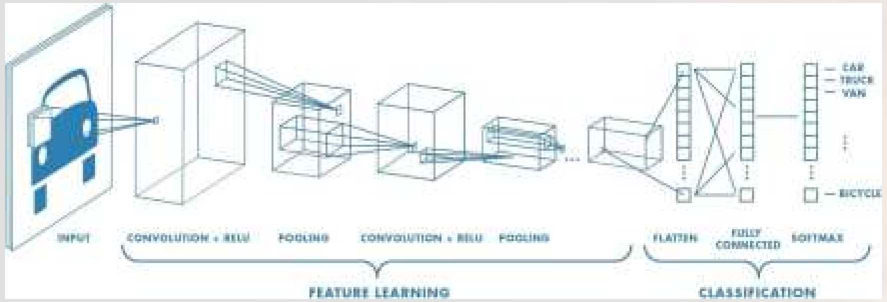
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

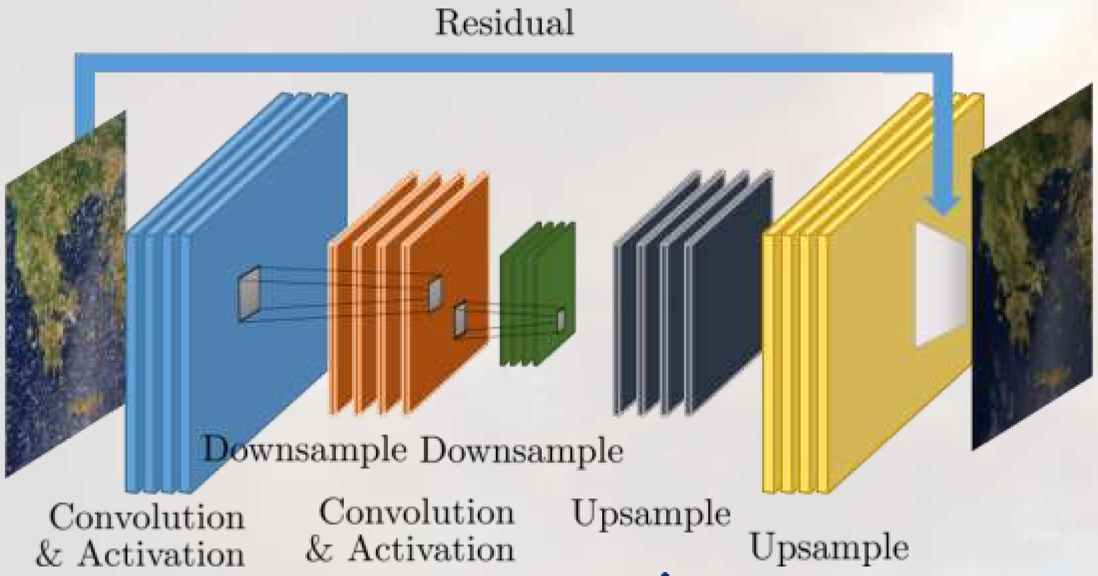
Deep Learning



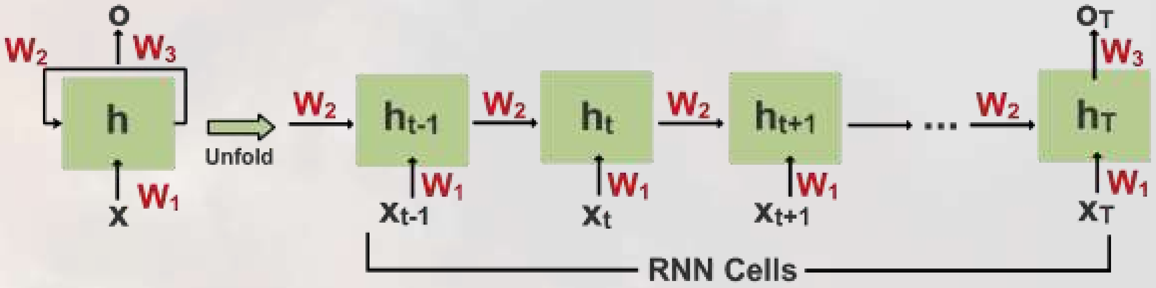
State-of-the-art in DL

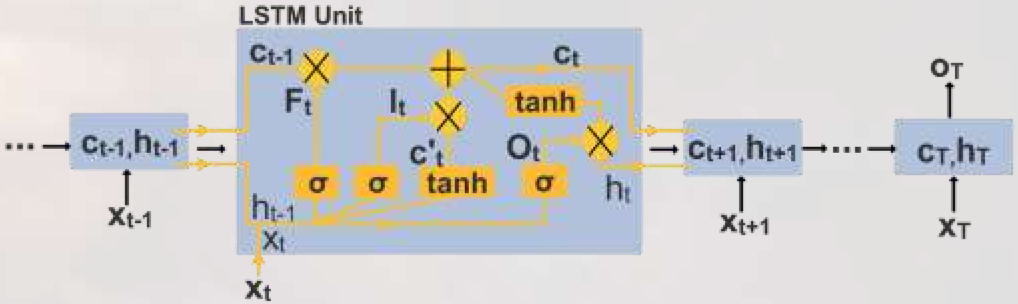
CNNs





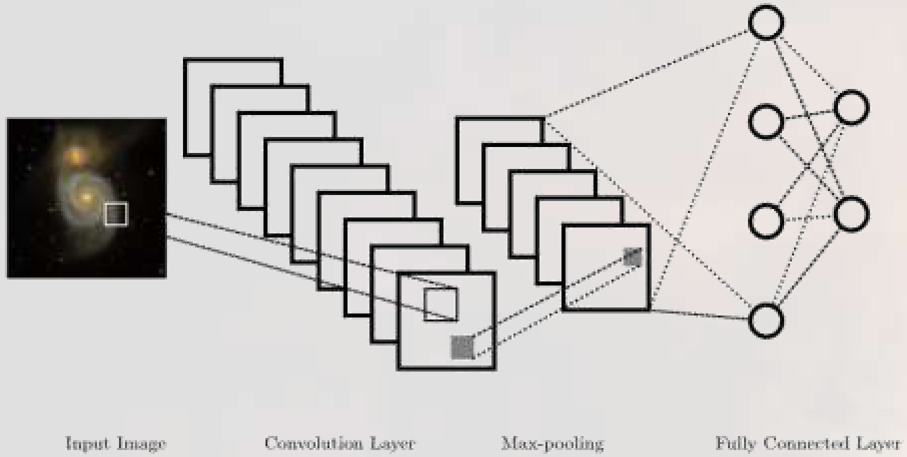
RNN/LSTMs

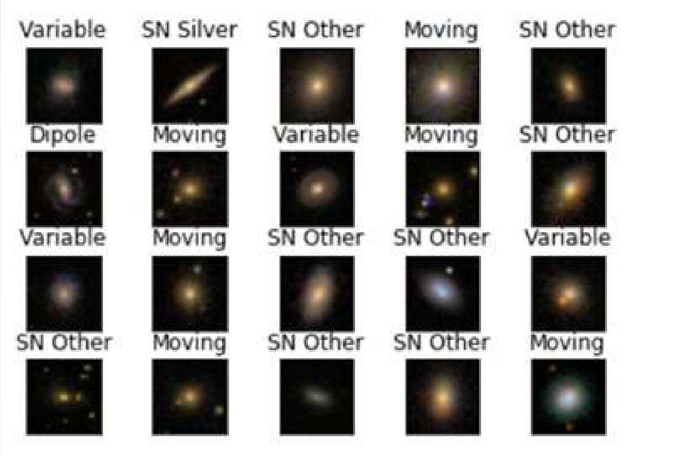


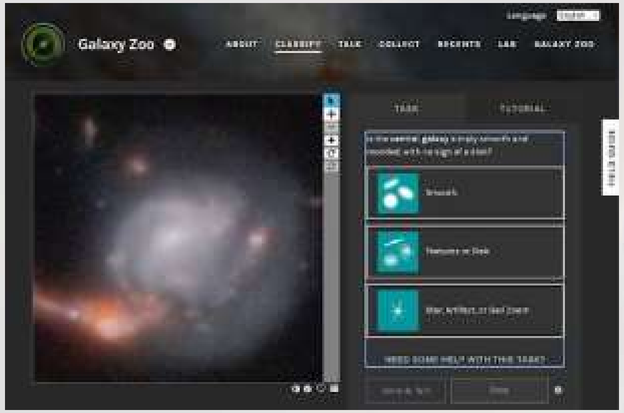


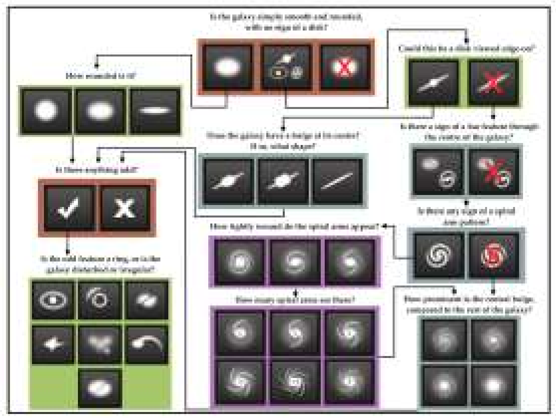
Transformers

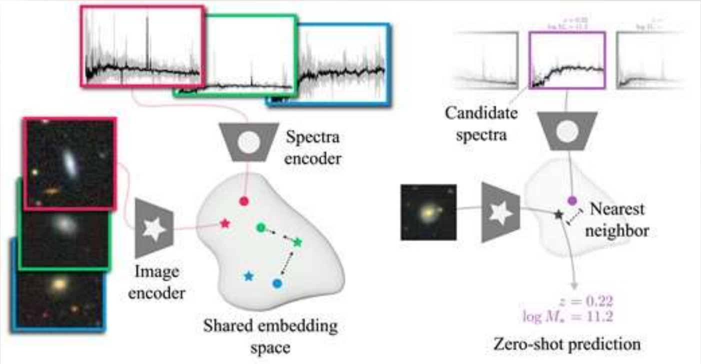
Deep Learning in Astronomy











TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Uncertainty (model)

alaskan malamute

siberian husky

siberian husky

Epistemic Uncertainty:

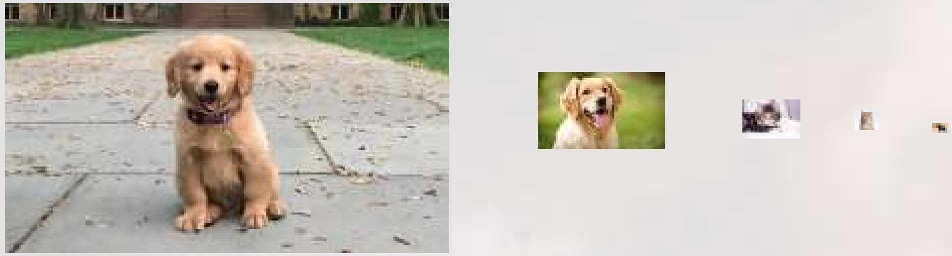
- ➢ Due to lack of knowledge about a system or process.
- ➢ Can be reduced as more knowledge is gained.

Uncertainty (data)




cat

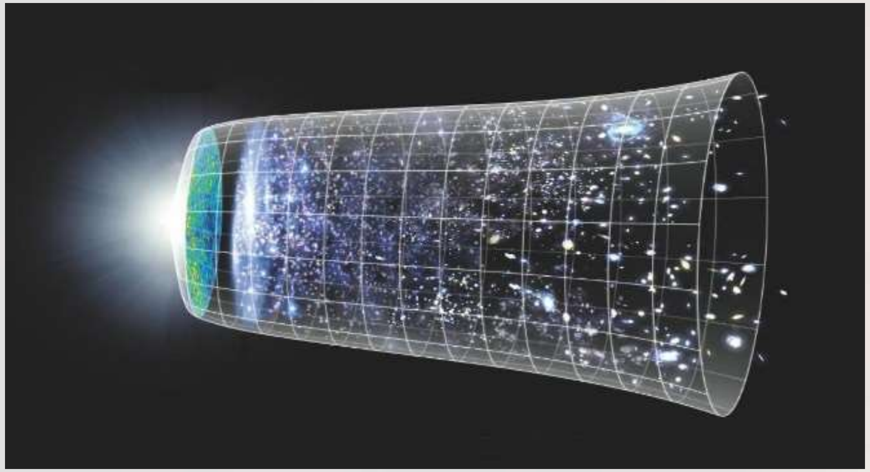
dog

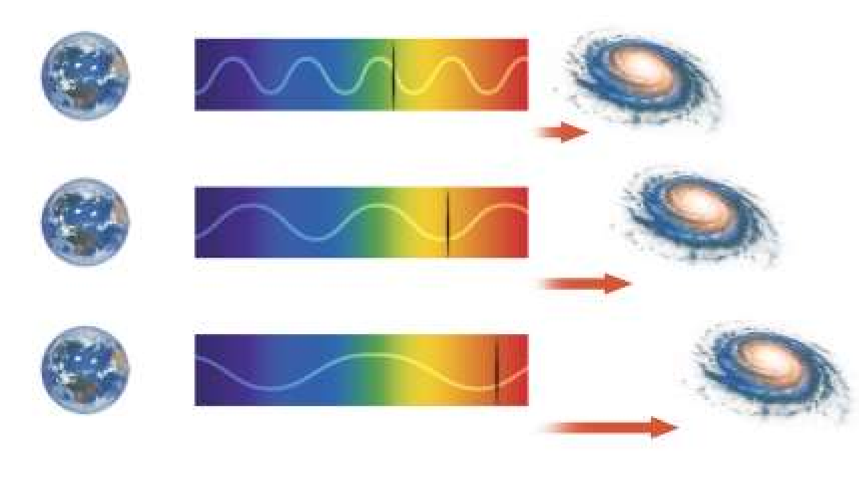


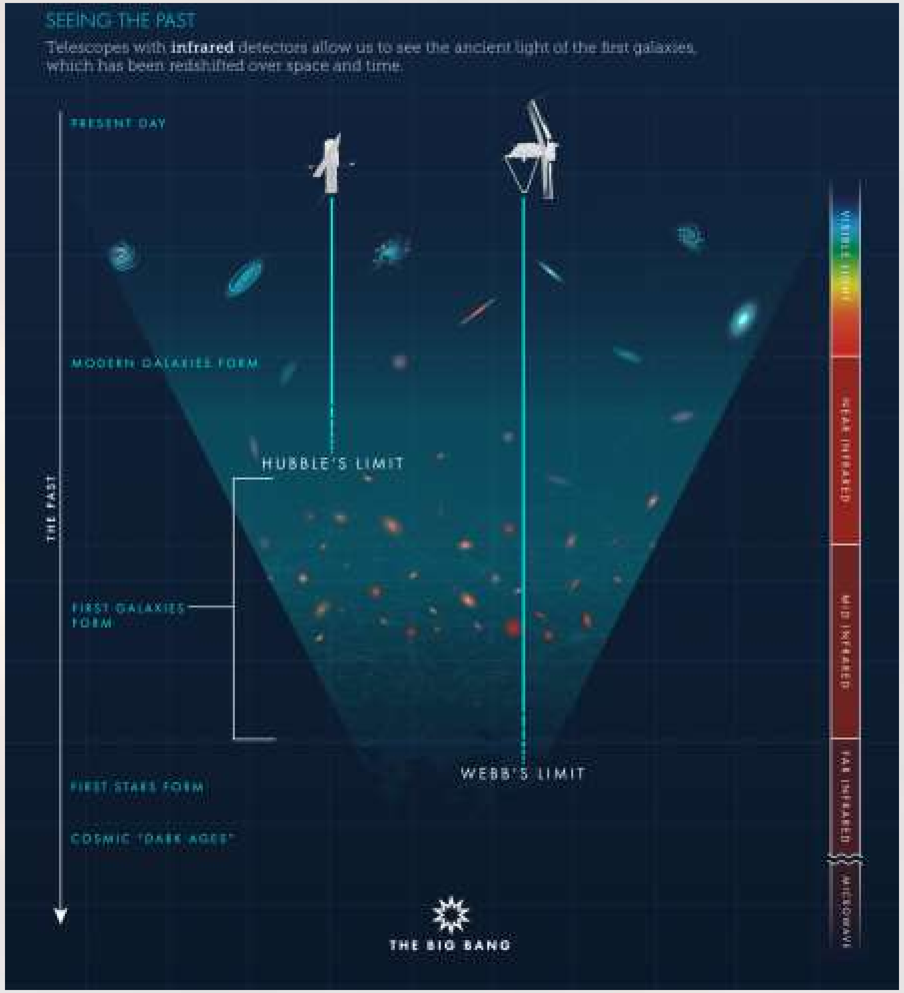
Aleatoric Uncertainty:

- inherent randomness in a system or process (flipping a coin
- cannot be reduced with more information or knowledge about the system.

Redshift estimation



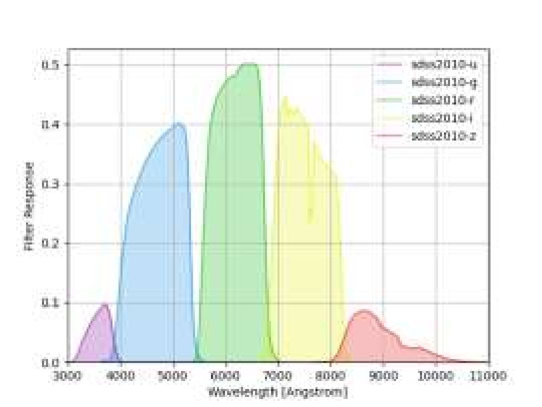


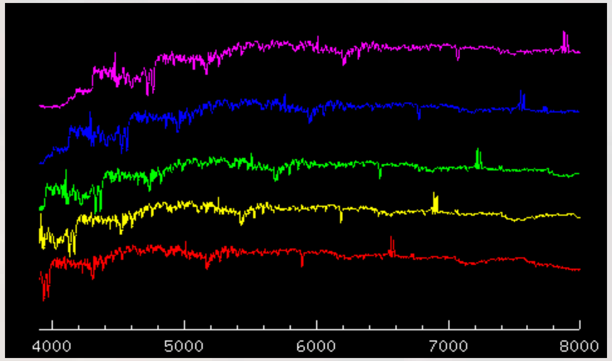


TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea

Funded by the European Union

Observations for Redshift estimation





Photometric

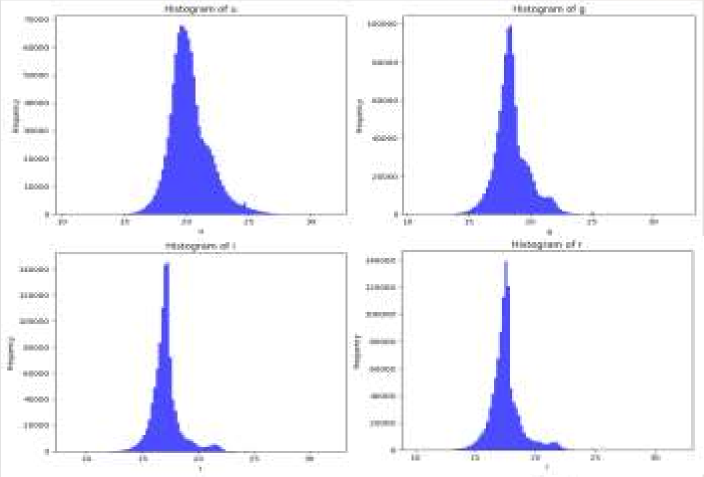
- 3-4 broad bands
- Cheap
- Inaccurate

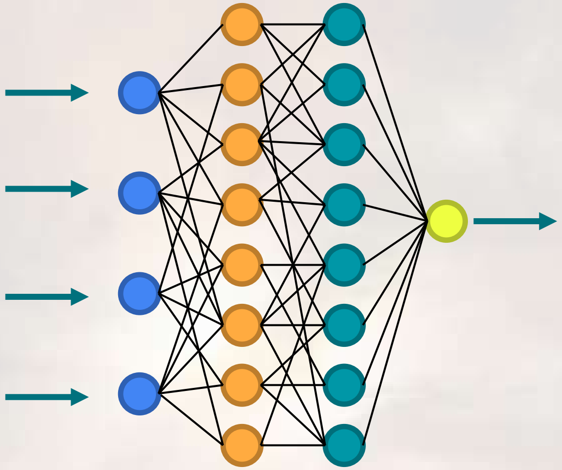
Spectroscopic

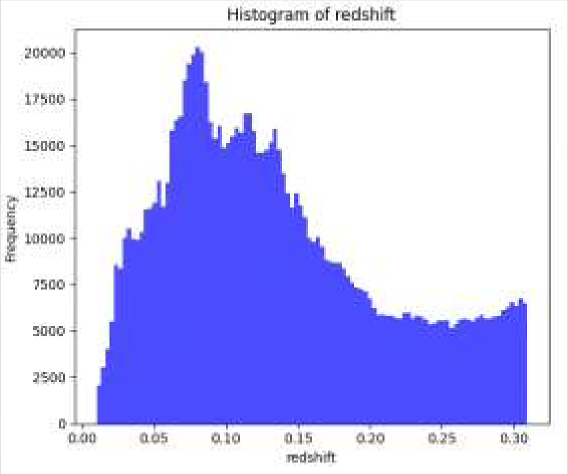
- Extended spectral range
- Expensive
- Accurate

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

ANNs for Photometric Redshift Estimation







TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH INSTITUTE OF COMPUTER SCIENCE

Signal Processing Laboratory

CosmoStat

cea

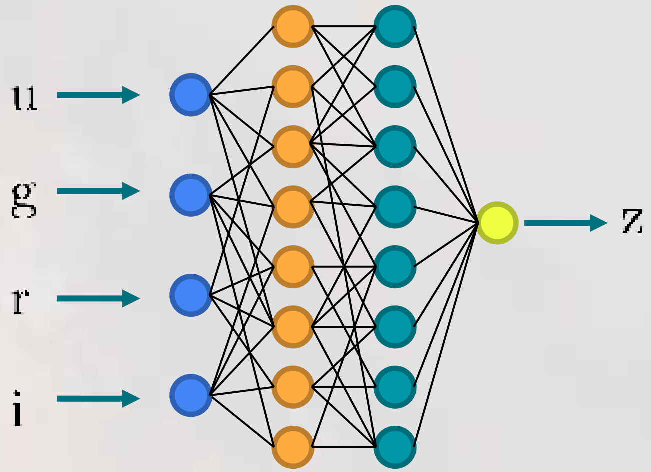
Funded by the European Union

ANN - Regression

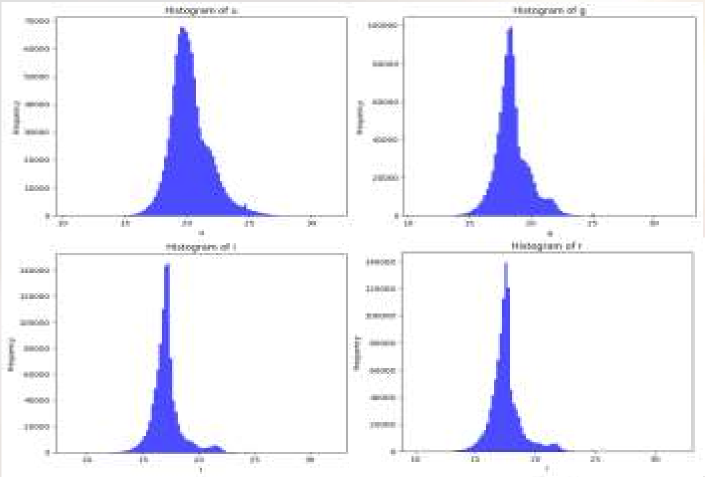
Baseline regression model $\hat{y}_i = f(x_i; \mathbf{w})$ where:

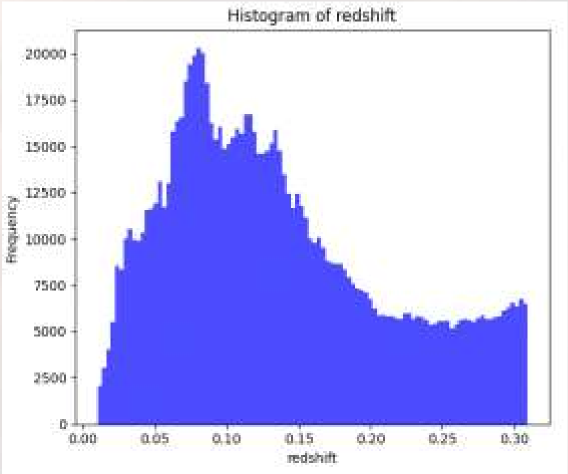
- $f(x_i; \mathbf{w})$ is the prediction for the i-th observation,
- y_i is the actual value for the i-th observation

Loss function: $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i; \mathbf{w}) - y_i)^2$



Dataset

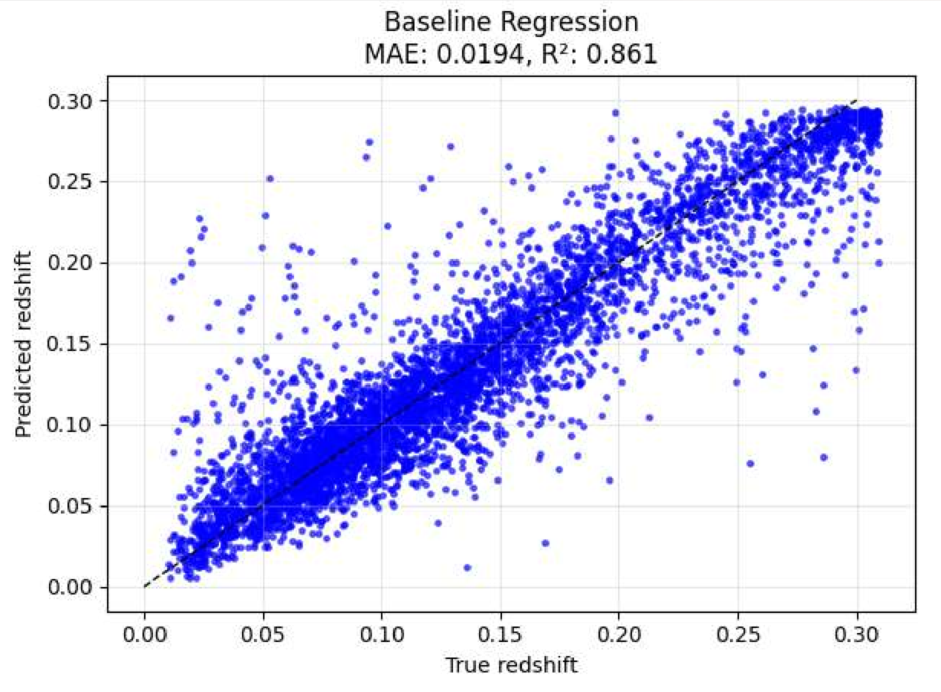


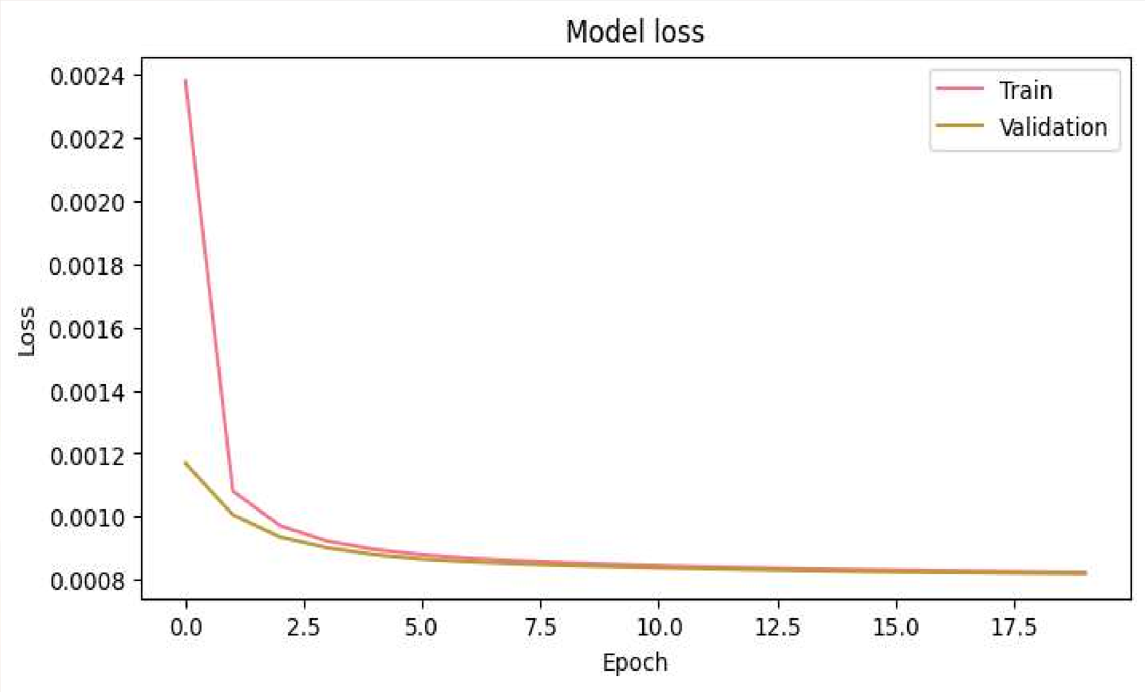


Training: 800K

Testing: 200K

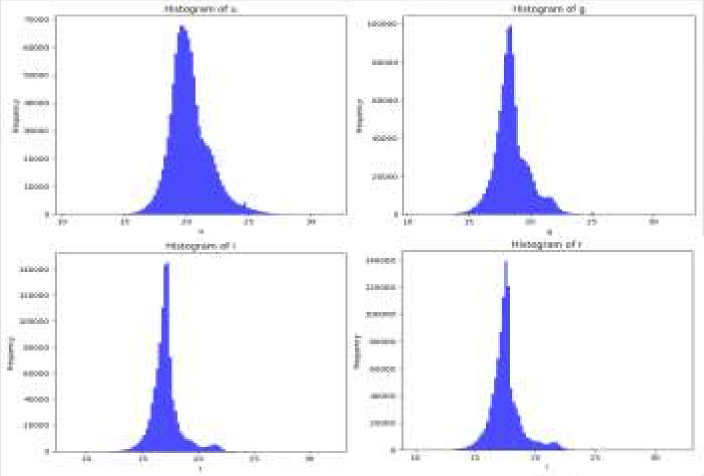
Baseline model & Data

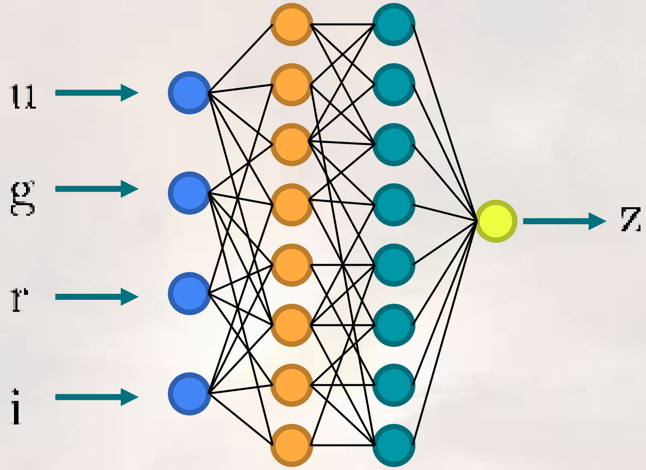


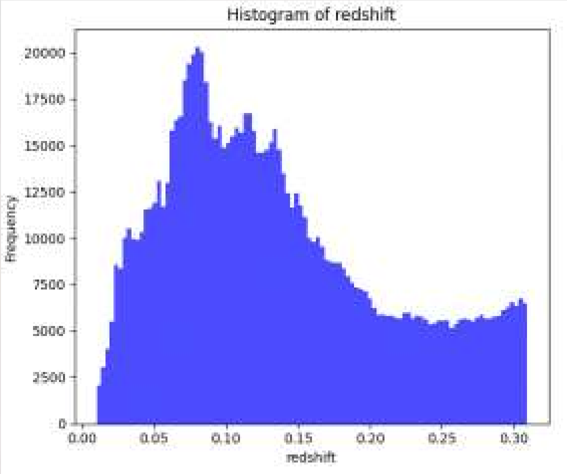


TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Uncertainties

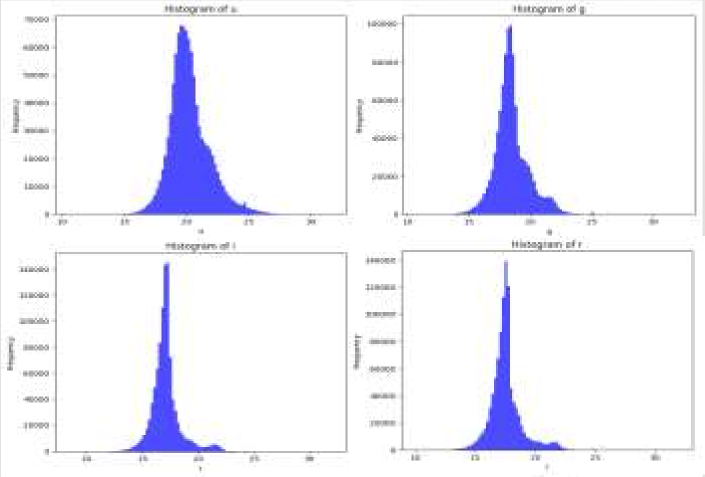


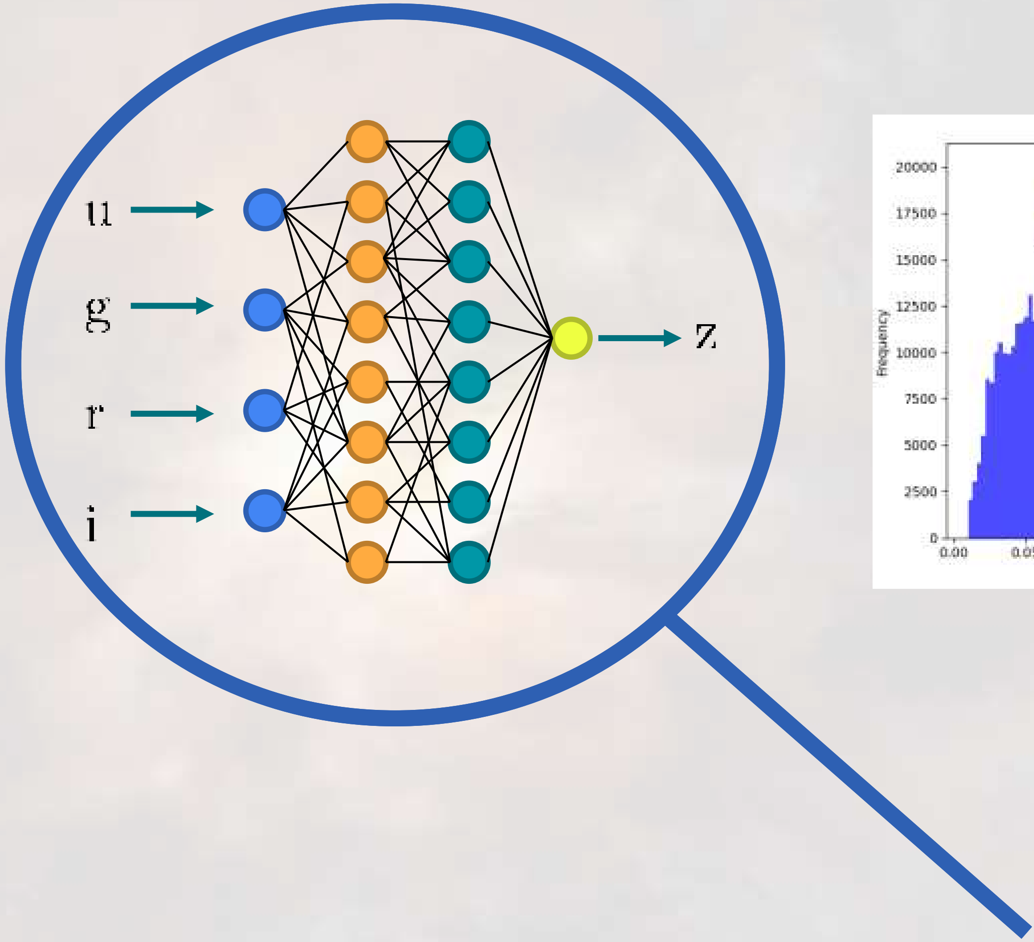


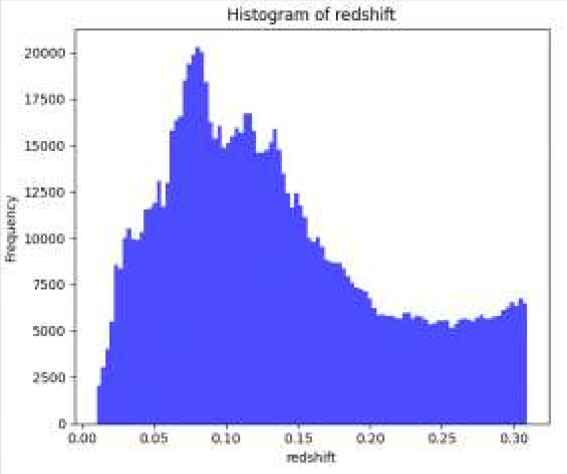


TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Uncertainties (model)





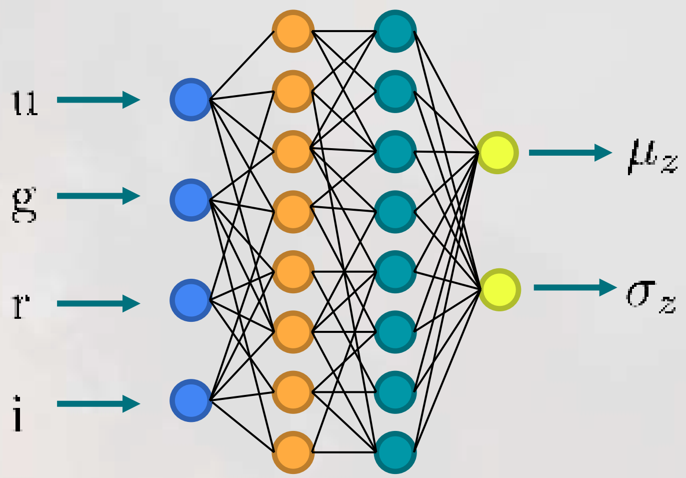


TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Gaussian Regression via ANN

- The target variable $y_i \sim \mathcal{N}(f(x_i; \mathbf{w}), \sigma^2)$
- The likelihood for each observation is:

$$p(y_i \mid x_i, \mathbf{w}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i; \mathbf{w}))^2}{2\sigma^2}\right)$$



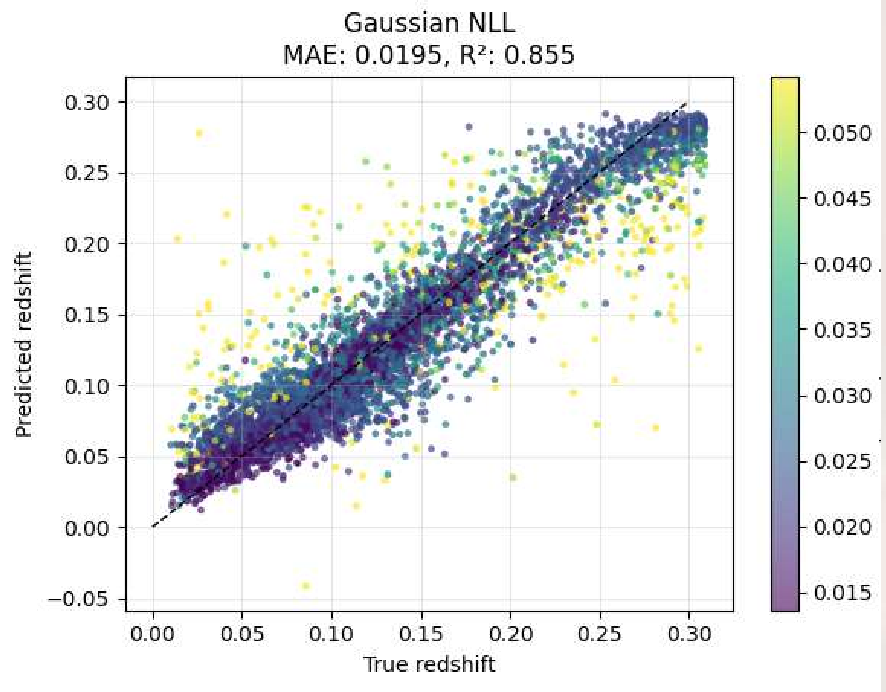
Loss: Negative Log-Likelihood (NLL) for N observations is:

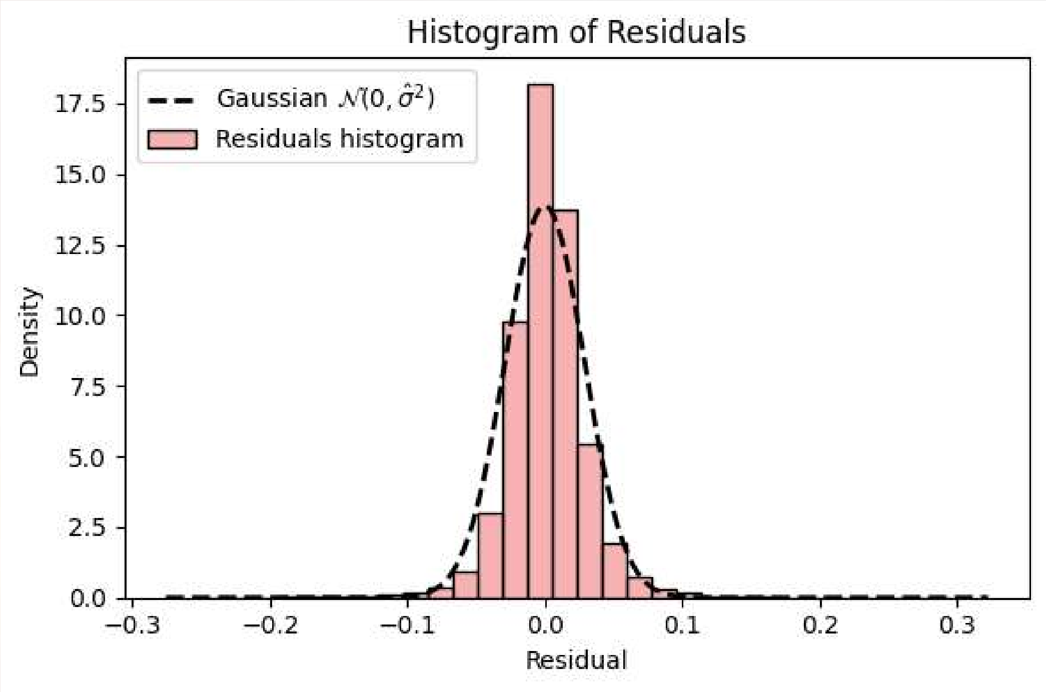
$$\mathcal{L}(\mathbf{w}) = -\sum_{i=1}^{N} \log p(y_i \mid x_i; \mathbf{w})$$

$$= \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - f(x_i; \mathbf{w}))^2$$

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

Gaussian Regression via ANN

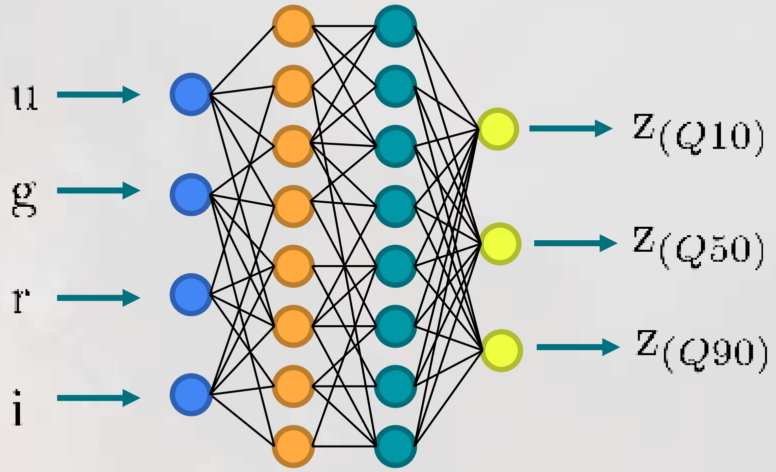




TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

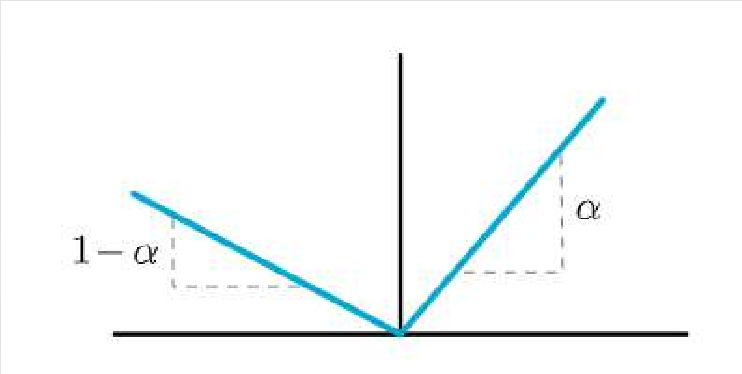
Quantile Regression

- Assume a set of target quantiles $\{\alpha_k\}_{k=1}^{K}$.
- $\hat{\mathbf{y}}_i = f(x_i; \mathbf{w}) = \left(\hat{y}_{i,\alpha_1}, \hat{y}_{i,\alpha_2}, \ldots, \hat{y}_{i,\alpha_K}\right)$

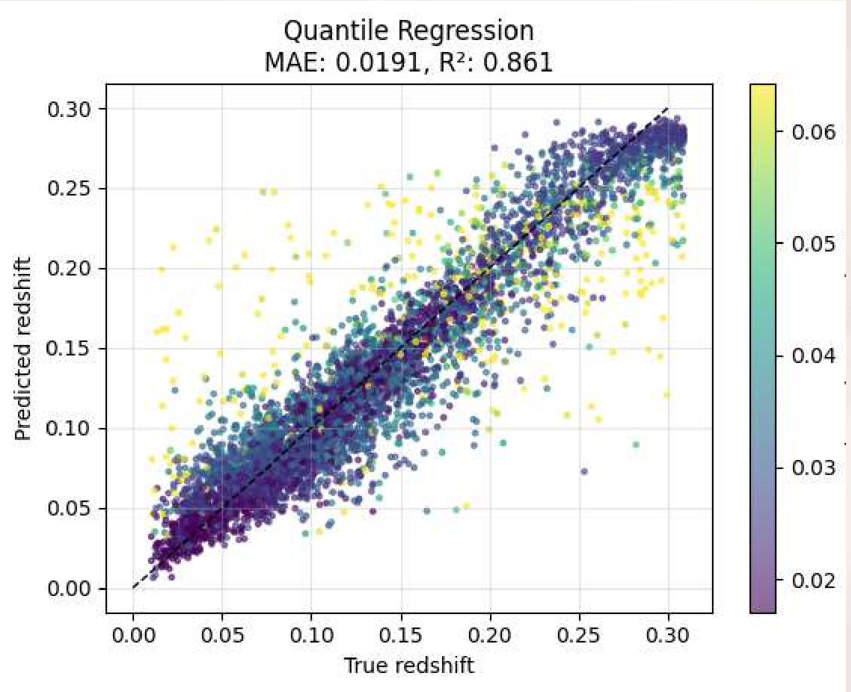


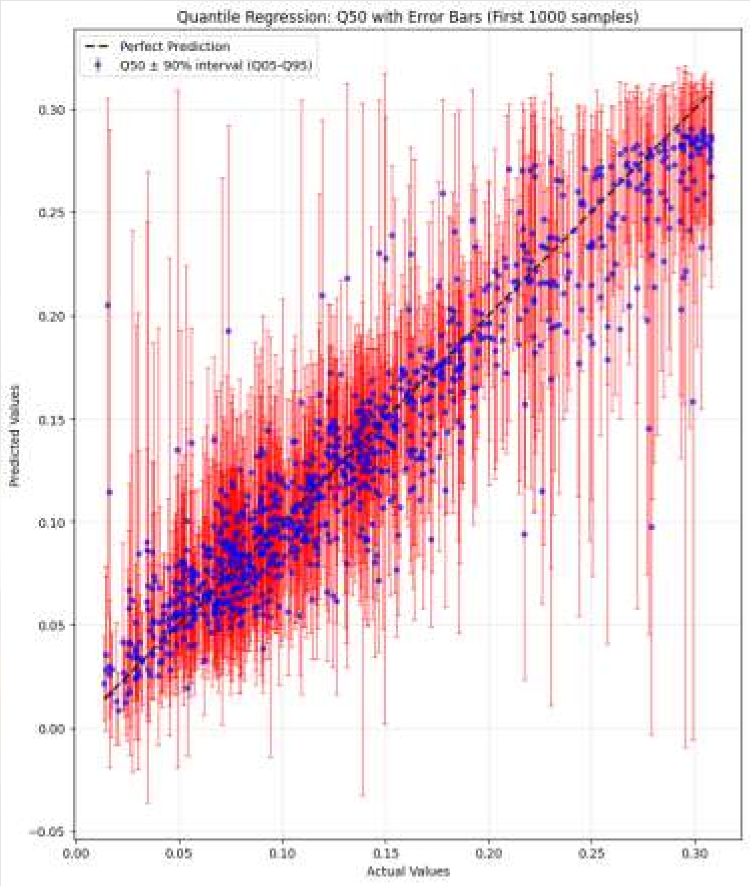
- *Mean pinball loss* (n samples, K quantiles):

$$L_\alpha\left(y, f(x)\right) = \begin{cases} \alpha\,|y - f(x)|, & y \geq f(x), \\ (1-\alpha)\,|y - f(x)|, & y < f(x). \end{cases}$$



Quantile Regression





90% Coverage: 0.922

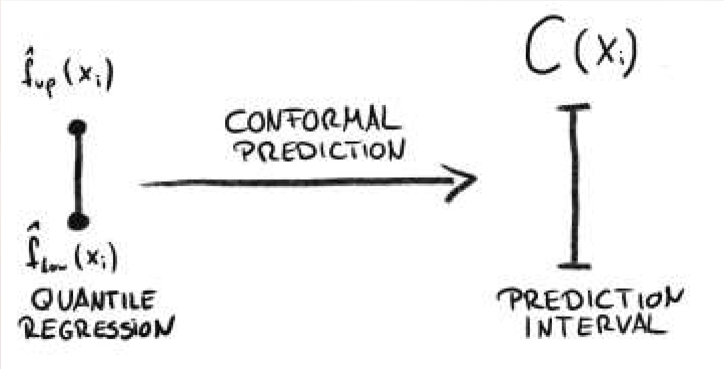
IQR $[Q_{075}(z)-Q_{0.25}(z)]$: 0.0319

Conformal Prediction

Regression task: age estimation

Model prediction: 24

MAPIE prediction interval: [20, 29]
(with 90% confidence)



TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

Conformal Prediction

1. Split training data to *proper training* set and *calibration*
2. Train lower and upper-quantile regressors $\widehat{Q}^{\alpha/2}$, $\widehat{Q}^{1-\alpha/2}$
3. On the calibration set, compute

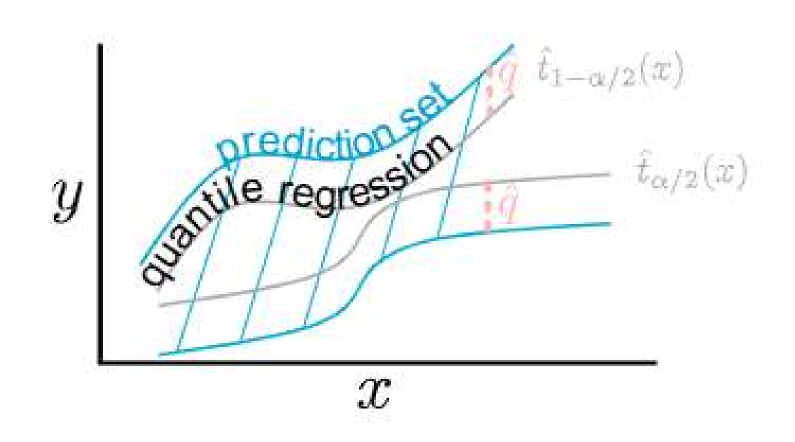
$$s_i^{\text{CQR}} = \max\left\{y_i^c - \widehat{Q}^{1-\alpha/2}(x_i^c),\ \widehat{Q}^{\alpha/2}(x_i^c) - y_i^c\right\}.$$

4. Compute the empirical $(1-\alpha)$-th quantile of the scores:

$$\widehat{q}^{\text{CQR}} = \begin{cases} s^{\text{CQR}}_{\lceil (1-\alpha)(n^c+1)\rceil}, & \text{if } \alpha \geq \dfrac{1}{n^c+1}, \\ \infty, & \text{otherwise.} \end{cases}$$

5. Output the coverage interval:

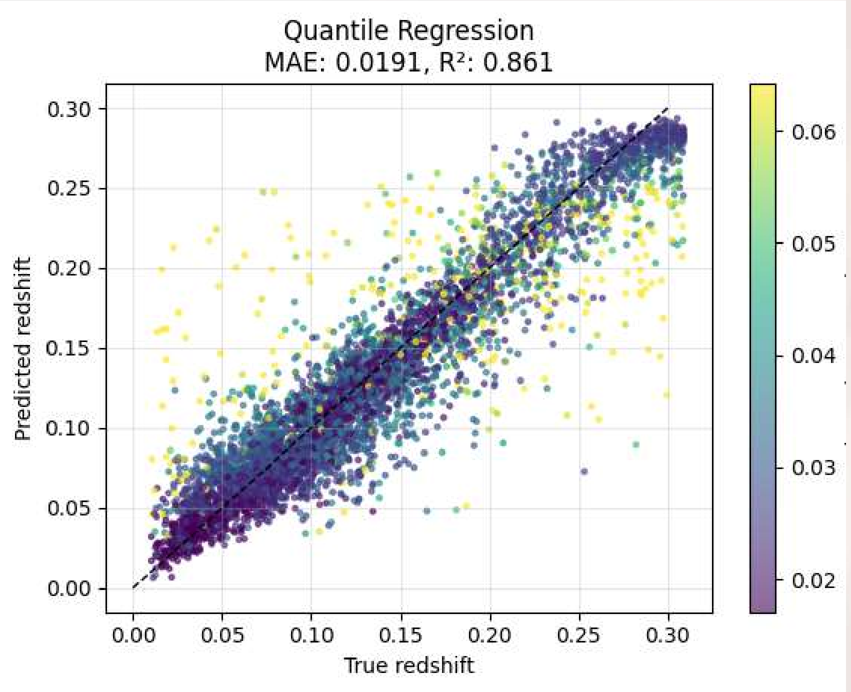
$$\mathcal{C}^{\text{CQR}}(x) = \left[\widehat{Q}^{\alpha/2}(x) - \widehat{q}^{\text{CQR}},\ \widehat{Q}^{1-\alpha/2}(x) + \widehat{q}^{\text{CQR}}\right]$$

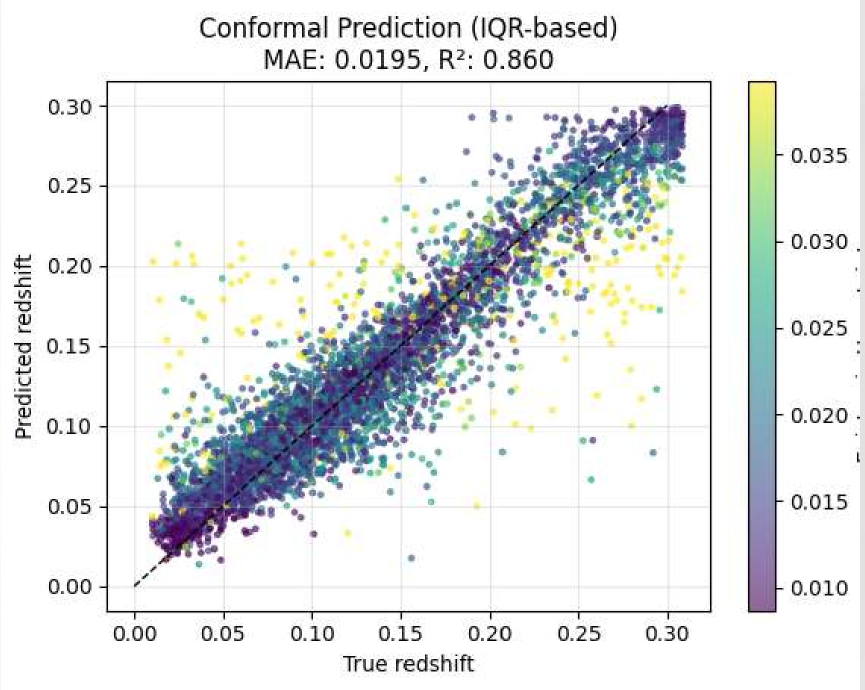


Romano, Y., Patterson, E., & Candes, E. Conformalized quantile regression. *NeurIPS 2029*.

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

Conformal Prediction





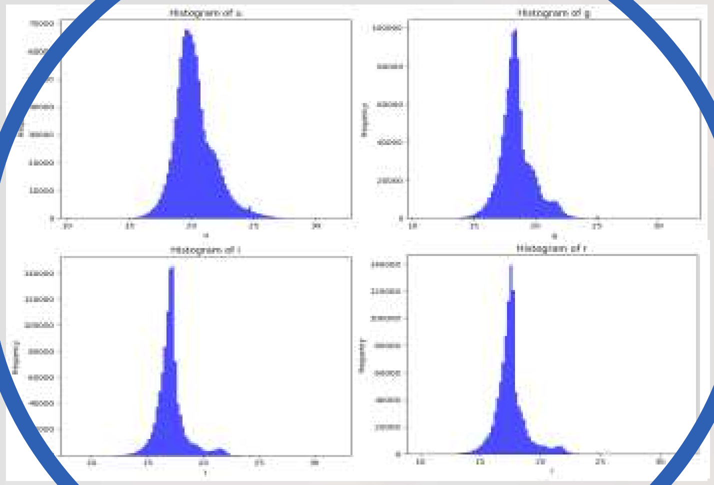
90% Coverage: 0.922
IQR: 0.0319

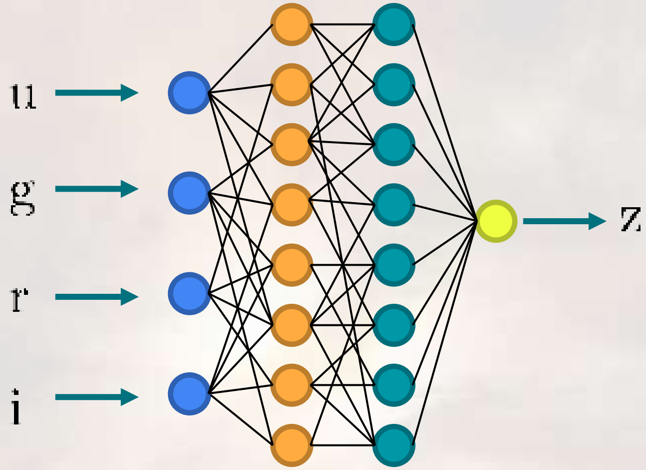
90% Coverage: 0.900
IQR: 0.0184

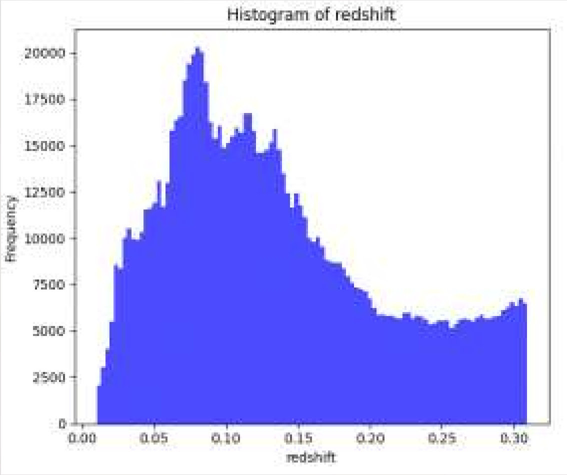
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea

Funded by the European Union

Uncertainties (data)







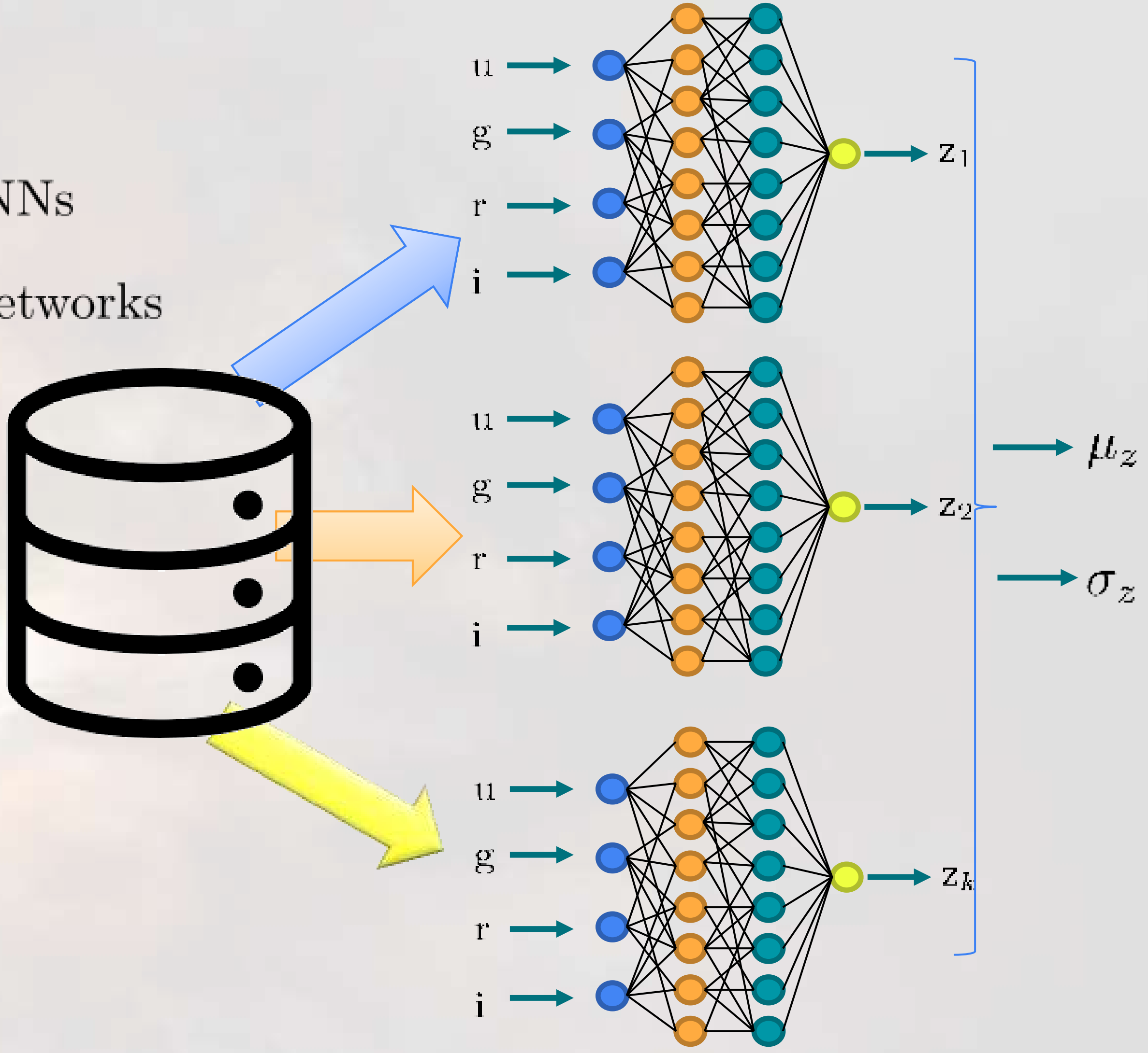
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Ensembles

- Let $f^{(m)}(\cdot\,;\mathbf{w}^{(m)})$ be M independently trained ANNs
- During inference, forward-propagate through all networks

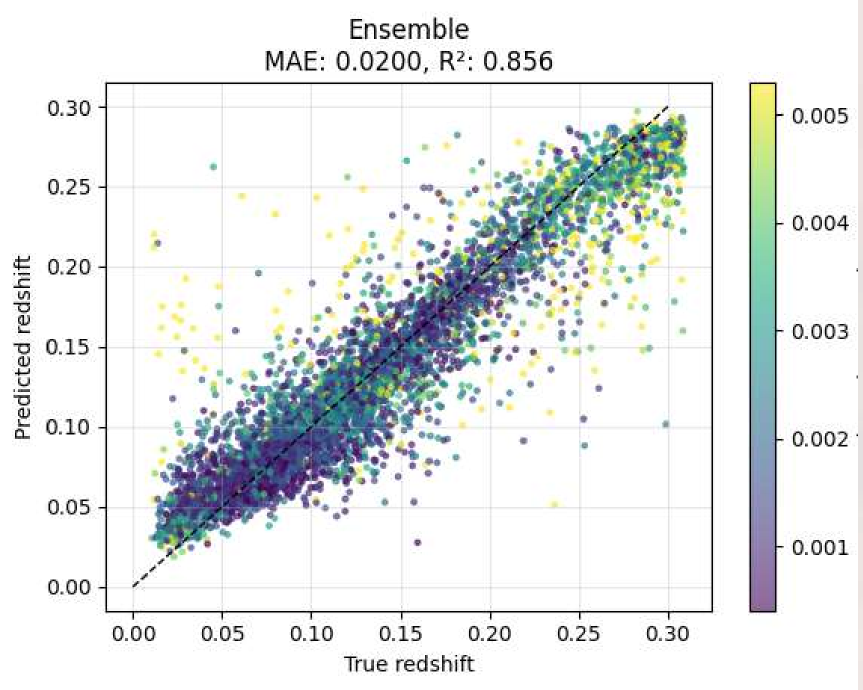
$$\mathbb{E}[y \mid x] \approx \bar{y}(x) = \frac{1}{M}\sum_{m=1}^{M} \hat{y}^{(m)}$$

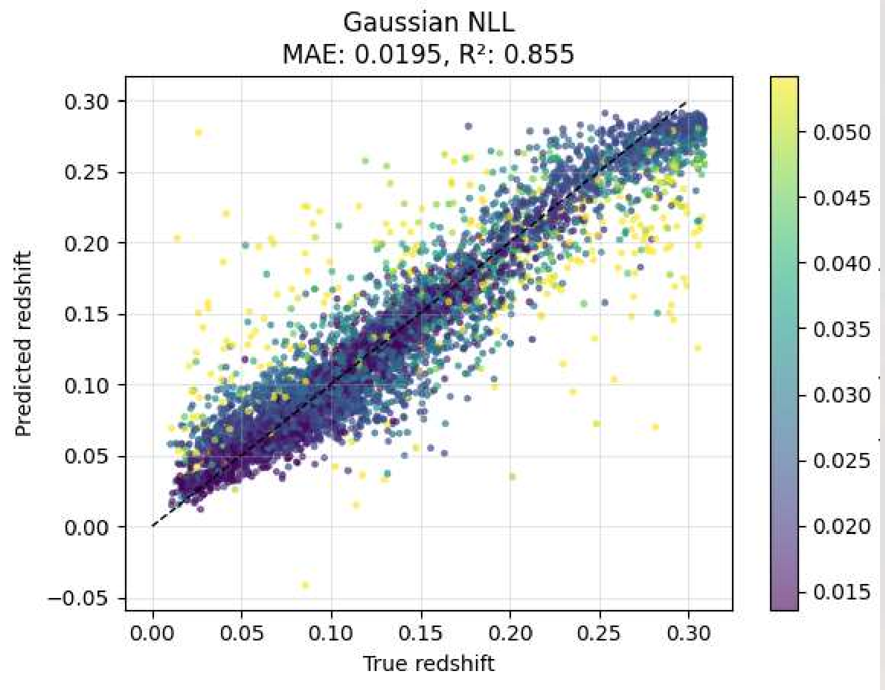
$$\mathrm{Var}[y \mid x] \approx \frac{1}{M}\sum_{m=1}^{M} \left(\hat{y}^{(m)}\right)^2 - \bar{y}(x)^2.$$



Ensemble approach

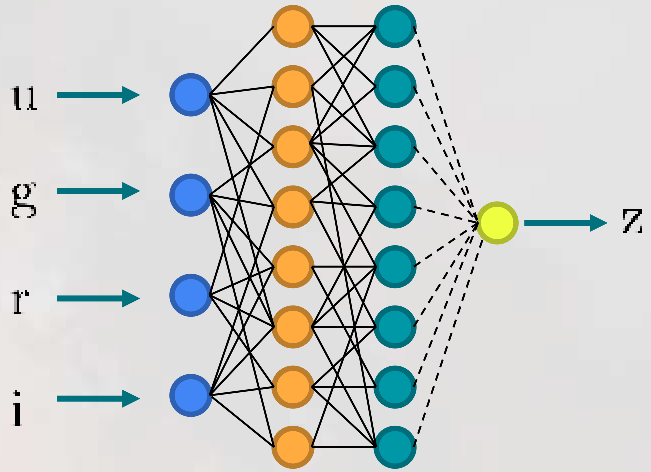
5 Models





TITAN Artificial Intelligence in Astrophysics · FORTH Institute of Computer Science · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

MCDropout



Apply *dropout* to forward pass

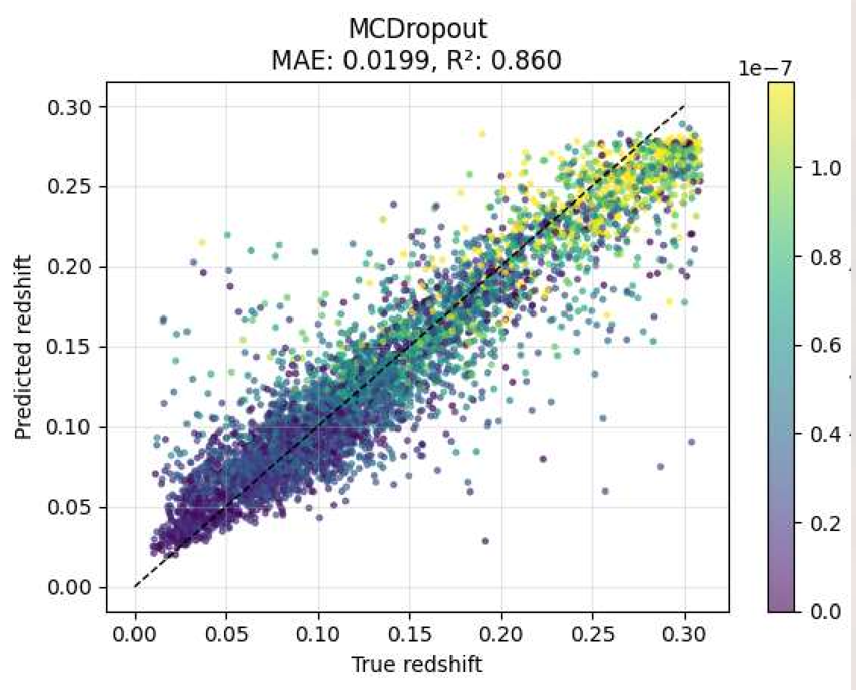
$$\hat{y}_i = f\big(x_i; \mathbf{w} \odot \mathbf{z}_i\big), \qquad \mathbf{z}_i \sim \mathrm{Bernoulli}(p),$$

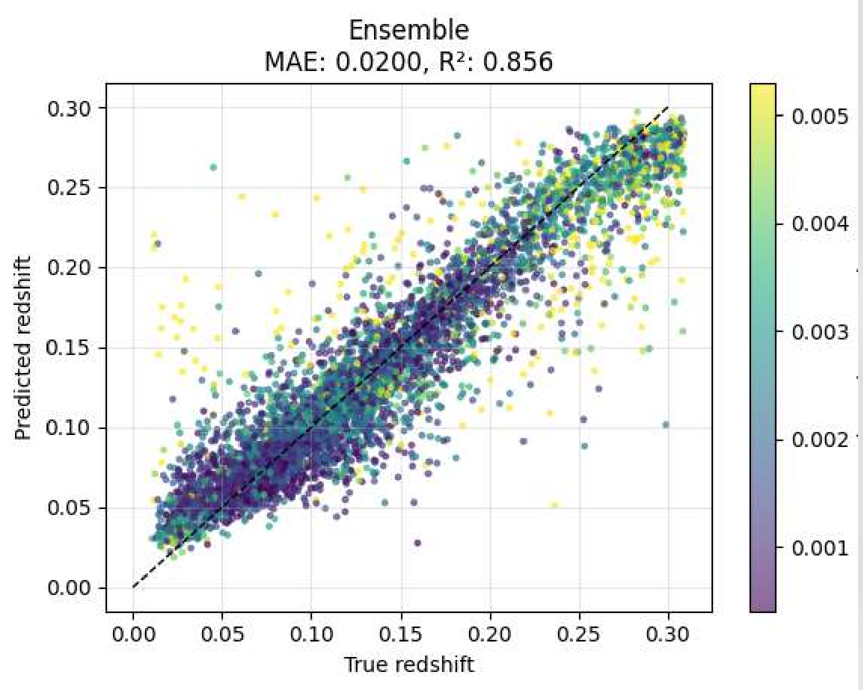
Predictive mean and variance

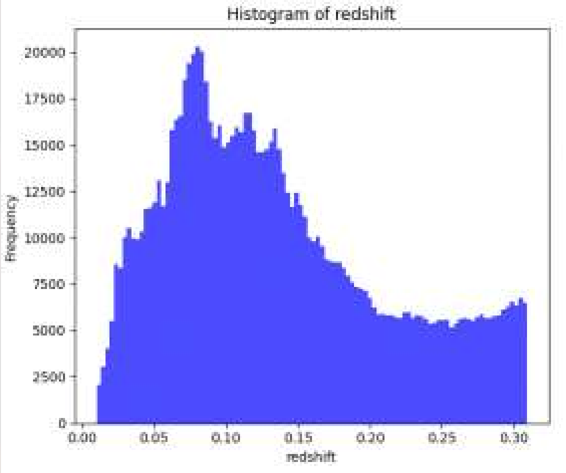
At test time keep dropout *on* and run S stochastic forward passes $\hat{y}^{(s)} = f\big(x; \mathbf{w} \odot \mathbf{z}^{(s)}\big)$, $s = 1, \dots, S$. The predictive posterior moments are approximated by

$$\mathbb{E}[y \mid x] \approx \bar{y} = \frac{1}{S}\sum_{s=1}^{S} \hat{y}^{(s)}, \quad \mathrm{Var}[y \mid x] \approx \frac{1}{S}\sum_{s=1}^{S}\big(\hat{y}^{(s)}\big)^2 - \bar{y}^{\,2} + \tau^{-1}.$$

MCDropout

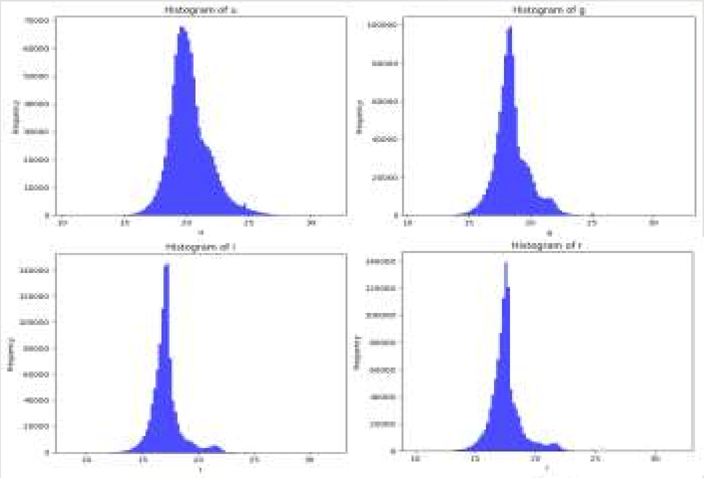


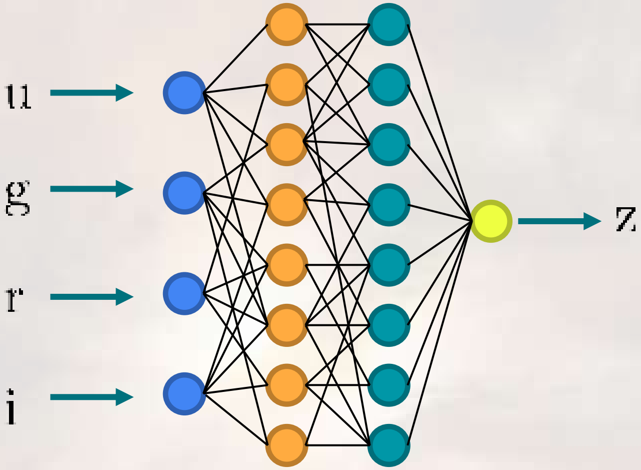


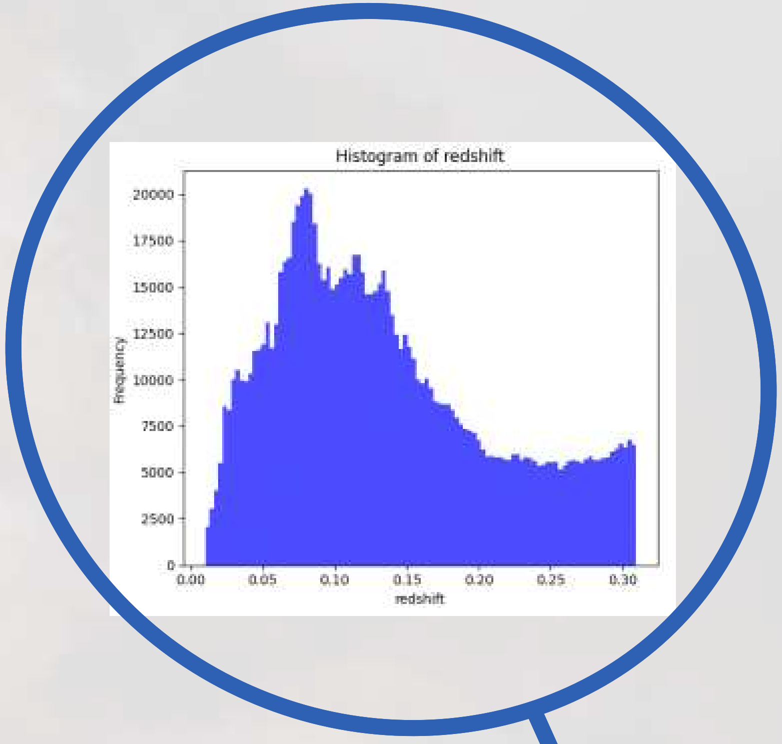


Funded by the European Union

Uncertainties (labels)

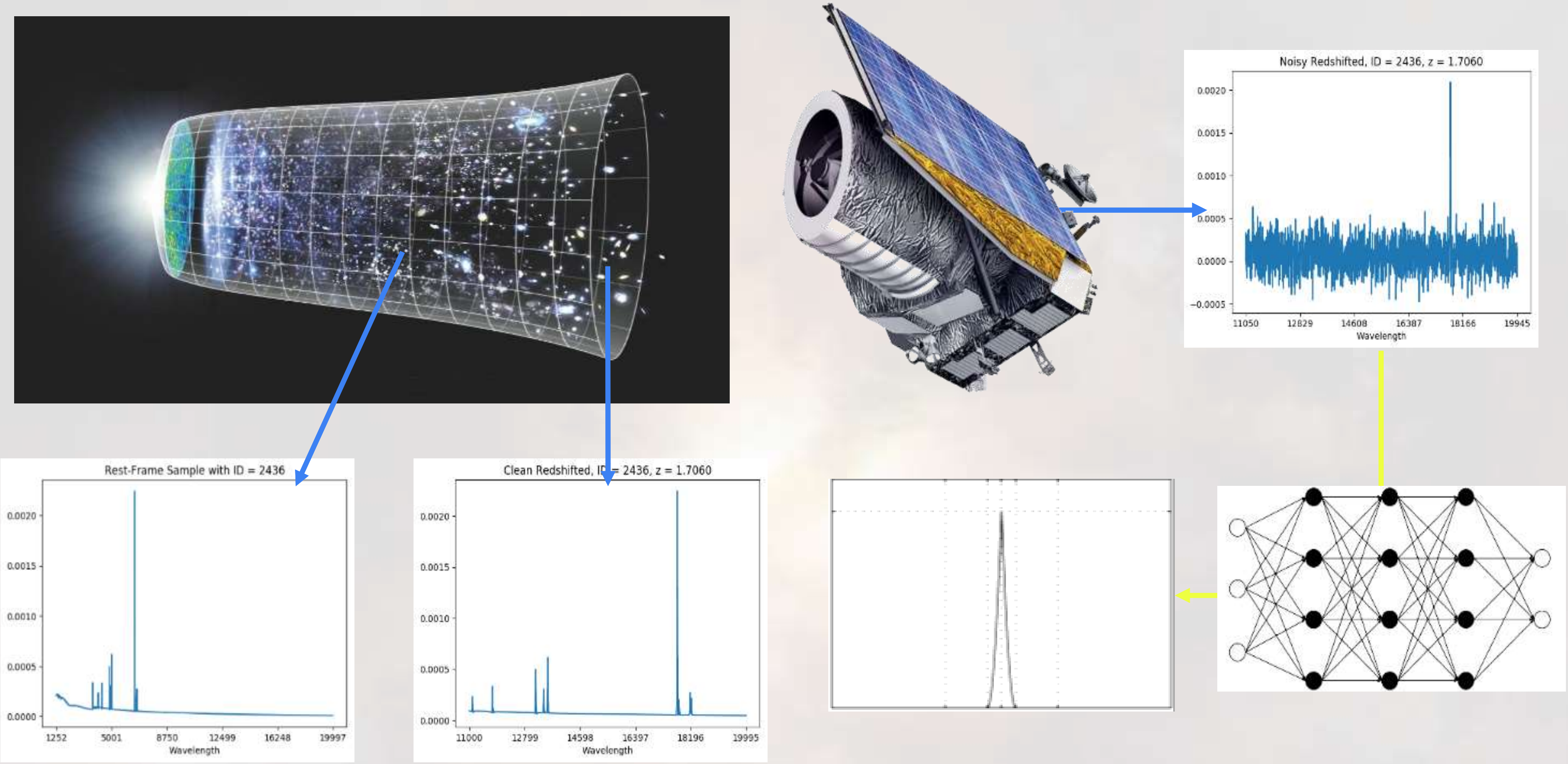






TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

Spectroscopic Red-Shift Estimation



Predictive model

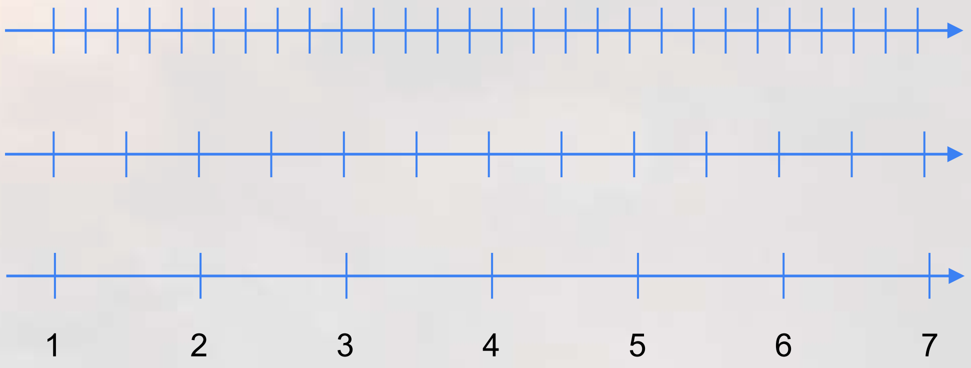
Redshift Estimation

Real-valued, non-negative number (z)

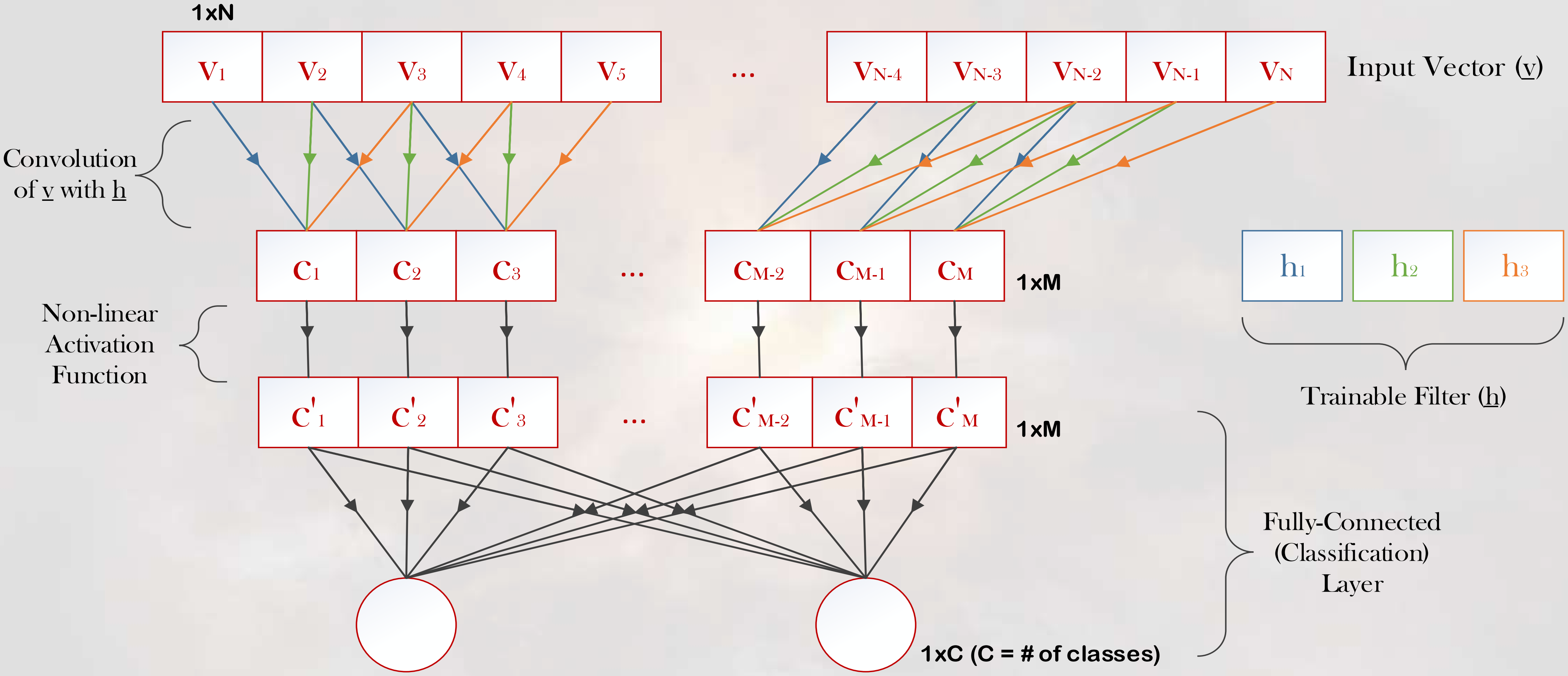
→ **Regression Analysis**

Split the examined redshift interval into ordinal classes, based on Euclid's characteristic resolution

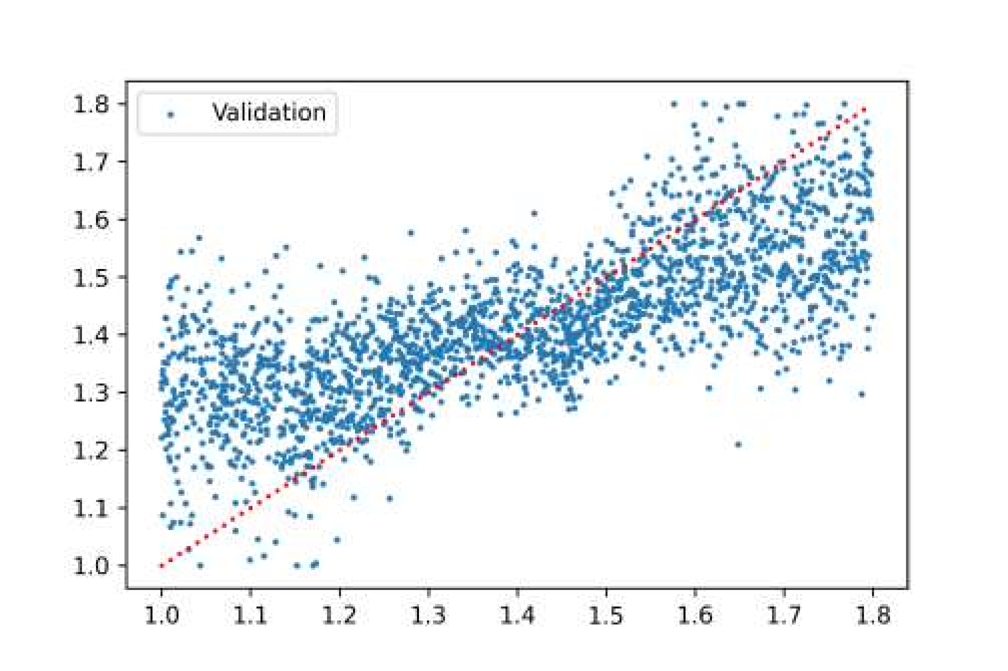
→ **Classification Problem**



1-Dimensional CNN - Classification

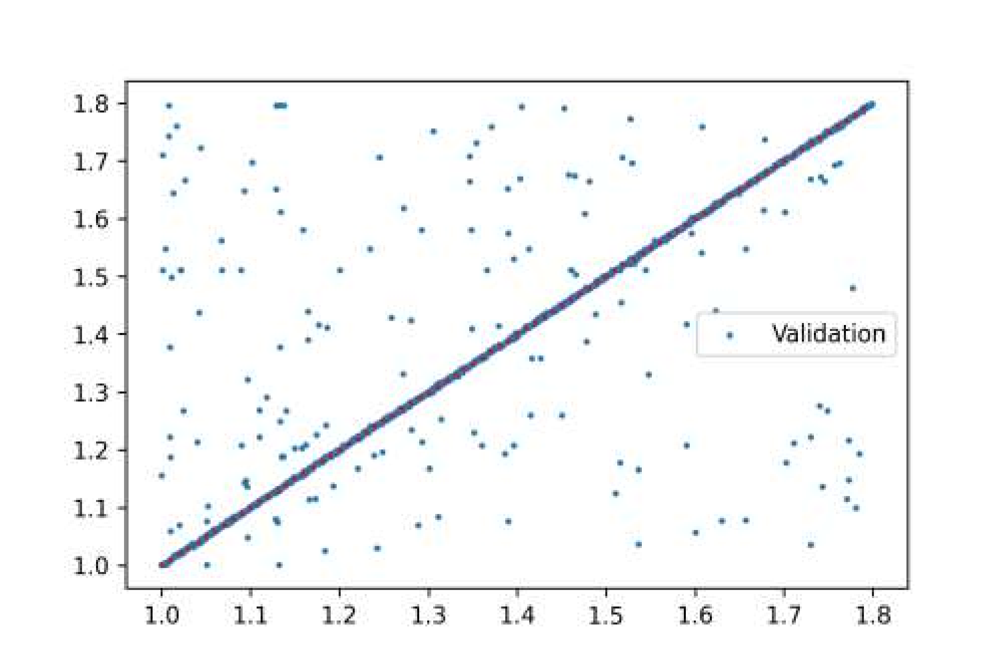


Regression



MSE 0.03811
MAE 0.16469
R^2 0.28188

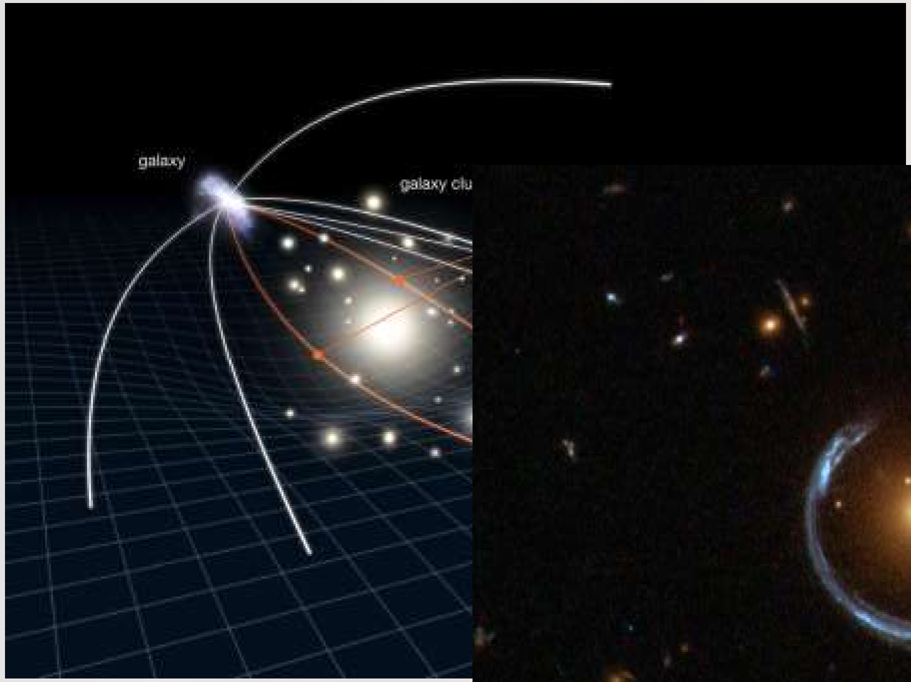
Classification (800 classes)

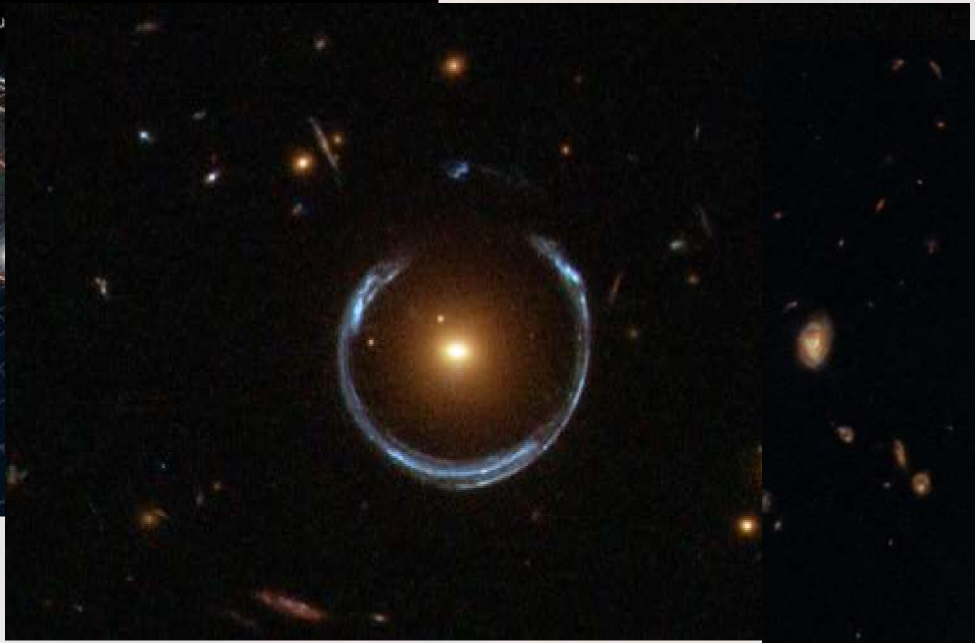


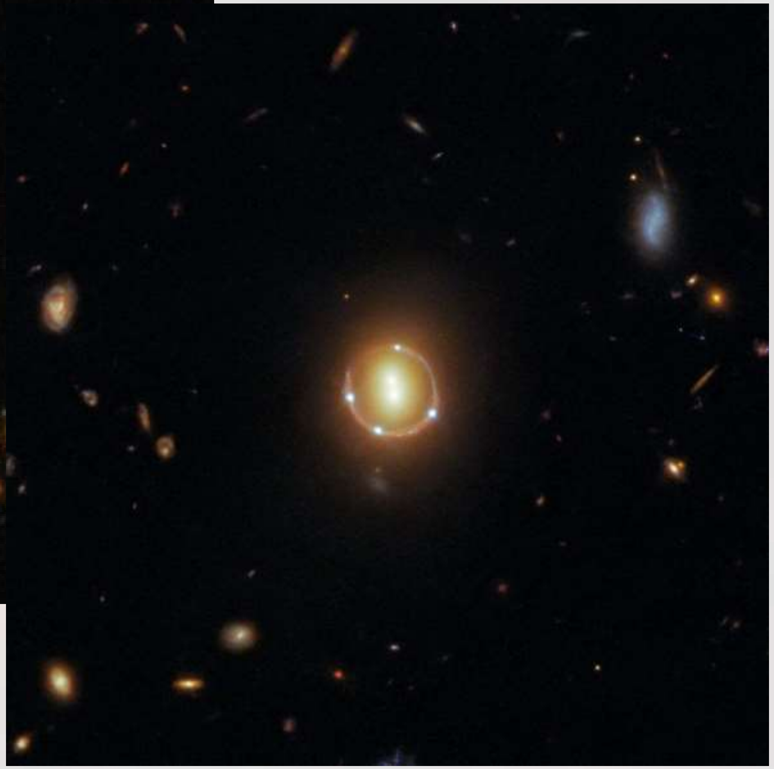
MSE 0.00791
MAE 0.02325
R^2 0.85082

TITAN Artificial Intelligence in Astrophysics · FORTH Institute of Computer Science · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

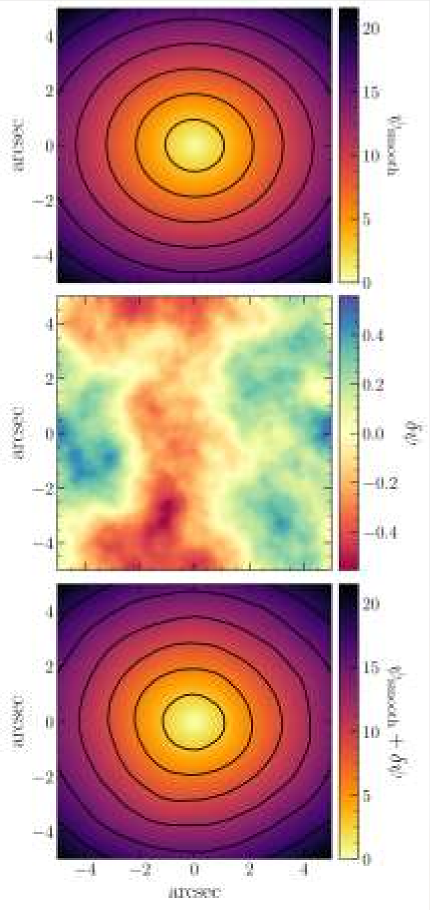
Gravitational Lensing







Modeling with uncertain labels



Singular Isothermal Ellipsoid parametric model

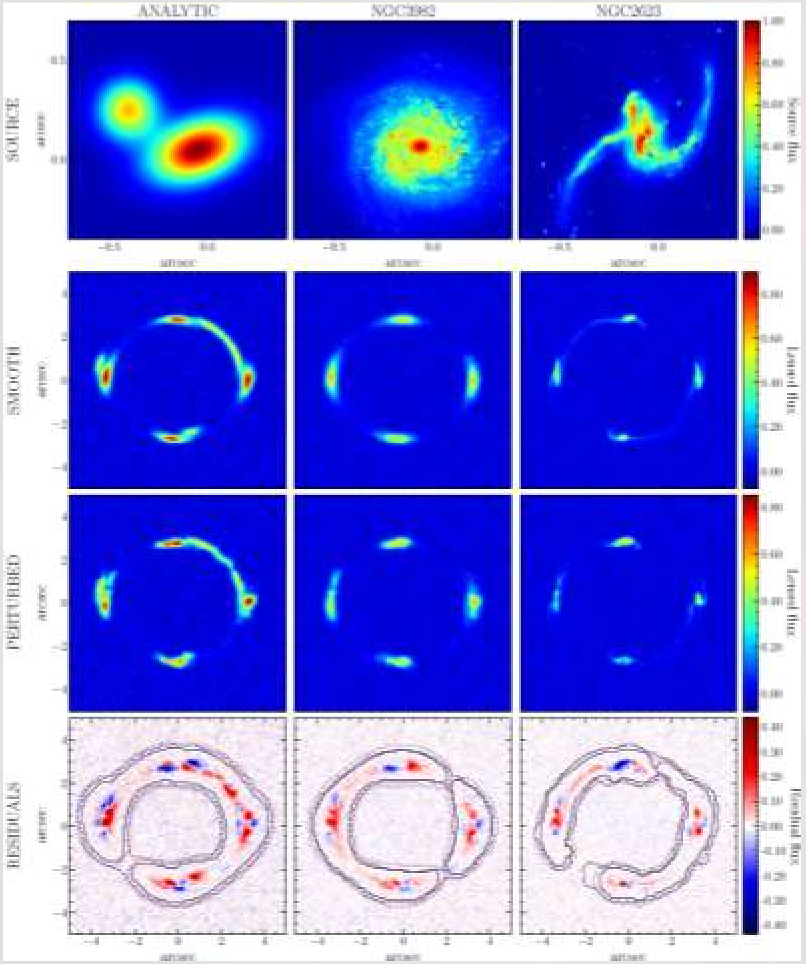
Realization of Gaussian Random Field perturbations

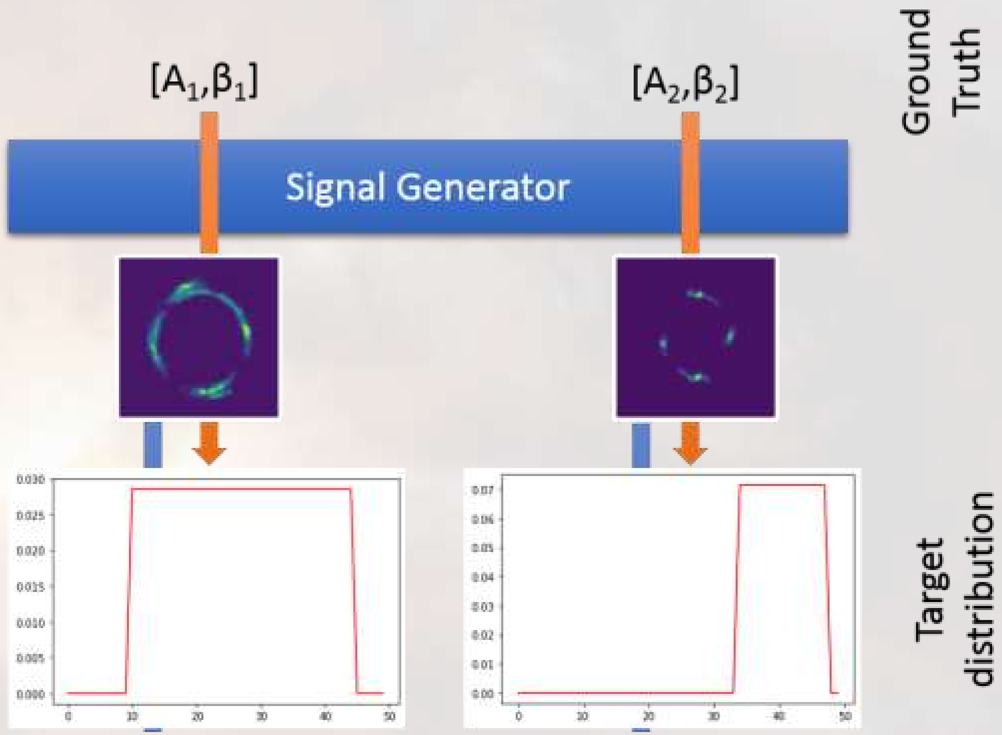
Perturbed lens potential

G. Vernardos, G. Tsagkatakis, and Y. Pantazis. "Quantifying the structure of strong gravitational lens potentials with uncertainty-aware deep neural networks." MNRAS. 2020.

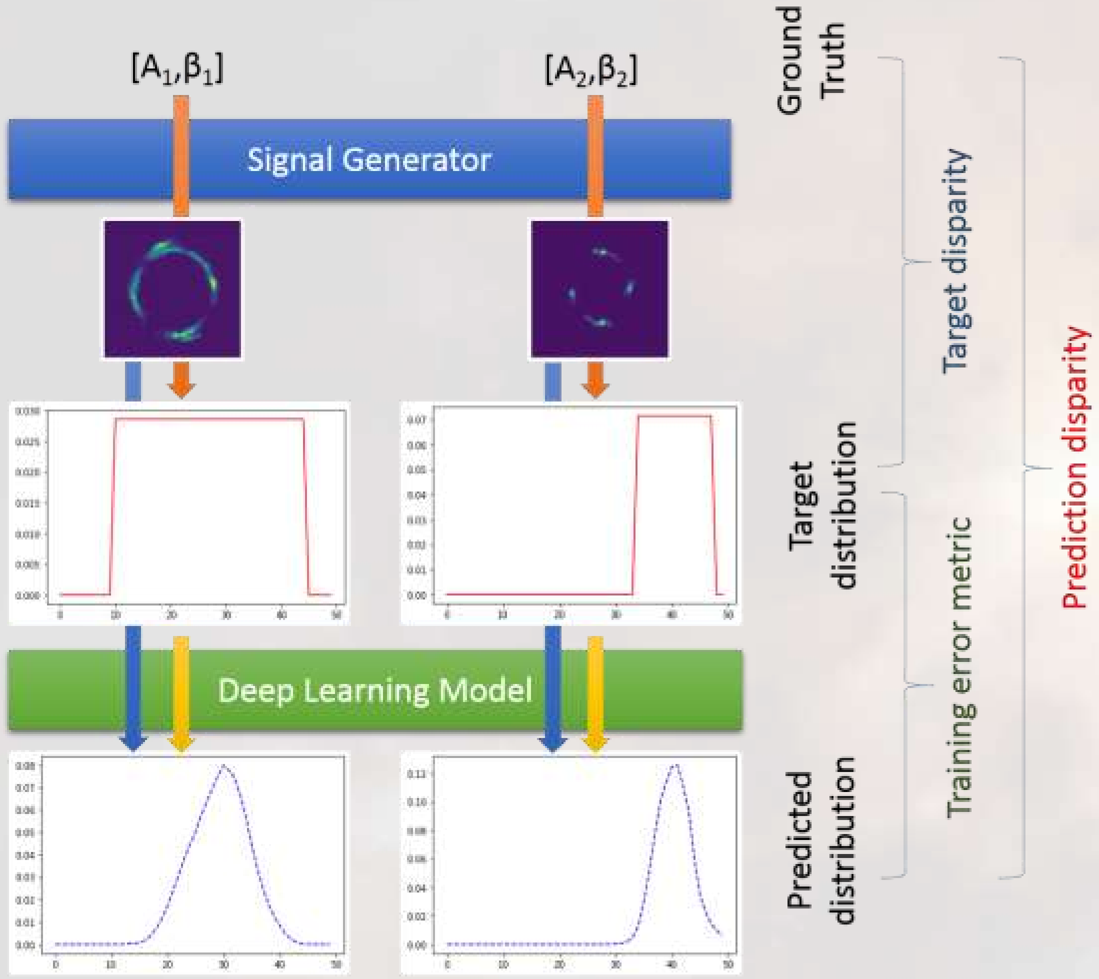
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Label uncertainty





Modeling with uncertain labels



TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

Modeling with uncertain labels

Given the predicted distribution P and the target distribution Q, the Jensen-Shannon divergence is defined by:

$$JS(Q,P) = \frac{1}{2}KL(P\|M) + \frac{1}{2}KL(Q\|M), M = \frac{1}{2}(P+Q)$$

Formally, the entropy of the predicted distribution $H(P)$ is given by:

$$H(P) = -\sum_{x\in\mathcal{X}} P(x)\log(P(x)).$$

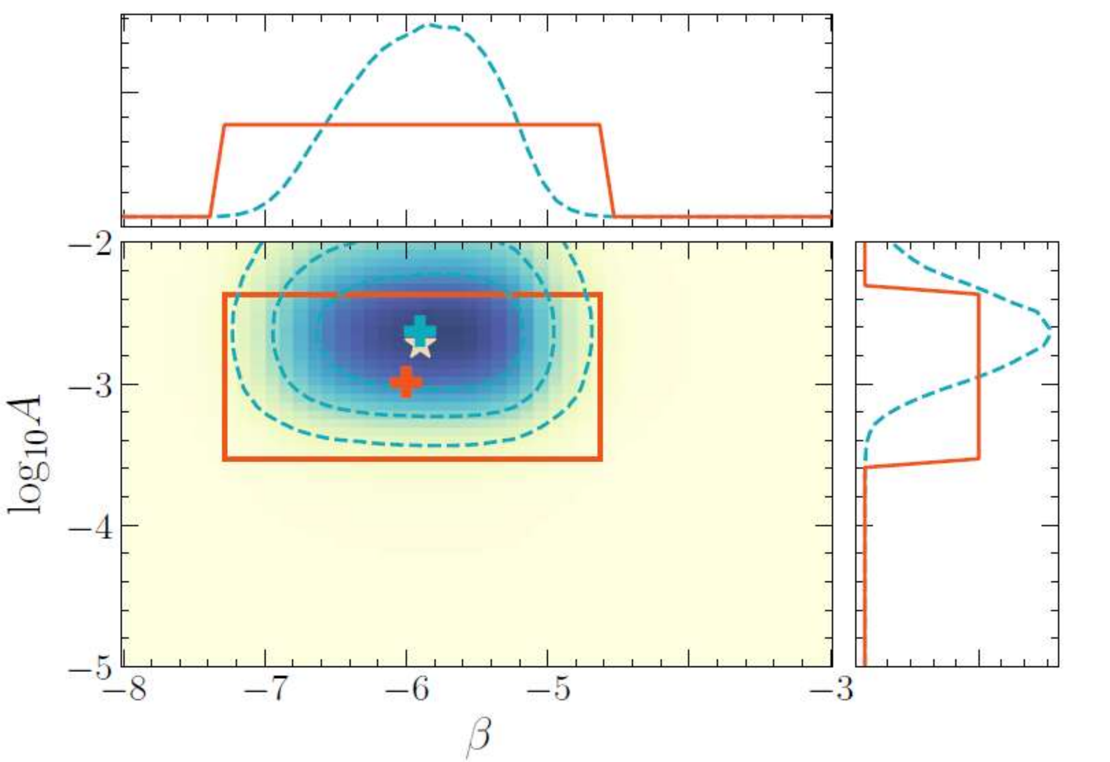
Entropy-regularized version of the JS divergence and is given by:

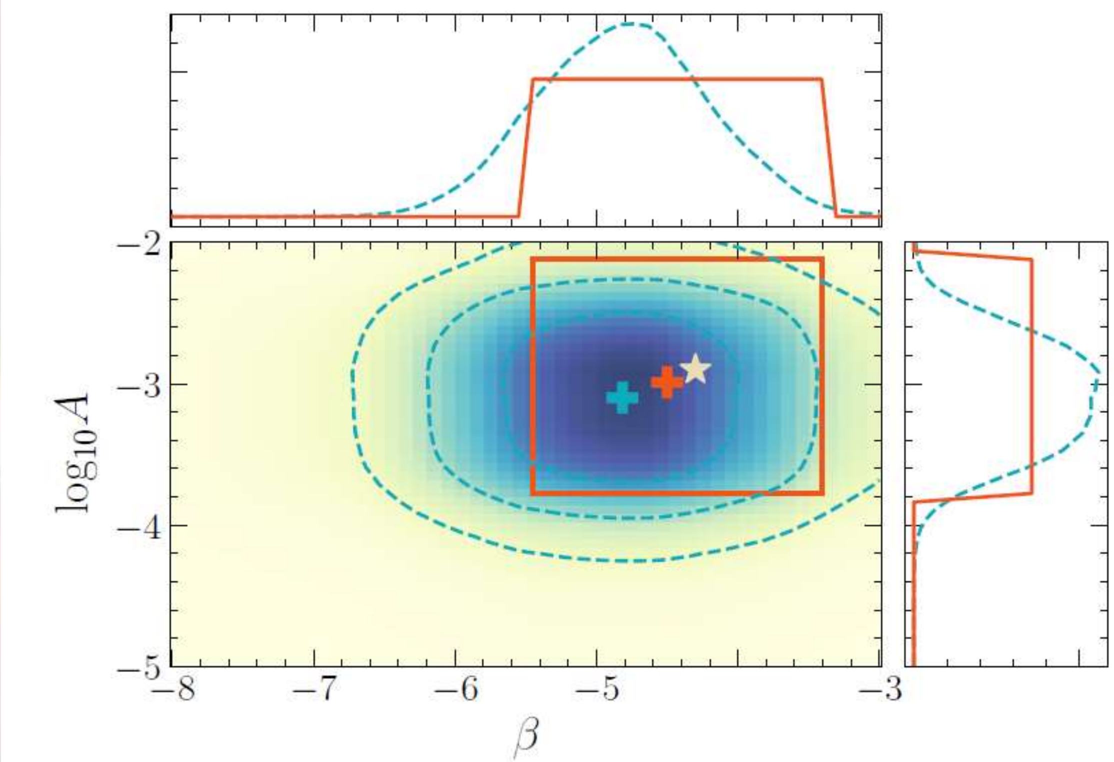
$$\mathcal{L}(P,Q) = \lambda_1 JS(P,Q) + \lambda_2 H(P),$$

where λ_1 and λ_2 control the impact of the two terms.

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH INSTITUTE OF COMPUTER SCIENCE · Signal Processing Laboratory · CosmoStat · cea · Funded by the European Union

“Label-super-resolution”





TITAN
ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH
INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Take-home messages

- Many flavors of uncertainty (decoupling, Bayesian, ordinal regression)
- “Limited” investigation in scientific data analysis
- The case of time-domain astronomy
- The case of spatially resolved observations
- The promise of multi-modality

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

MINOAS – Machine Intelligence for iNverse imaging, Observation Analysis and Sensing Workshop

Dates: 24-26 September 2025

Location: FORTH, Crete

Funded by the European Union