COSMO21- 22/5/2024

Chania, Crete

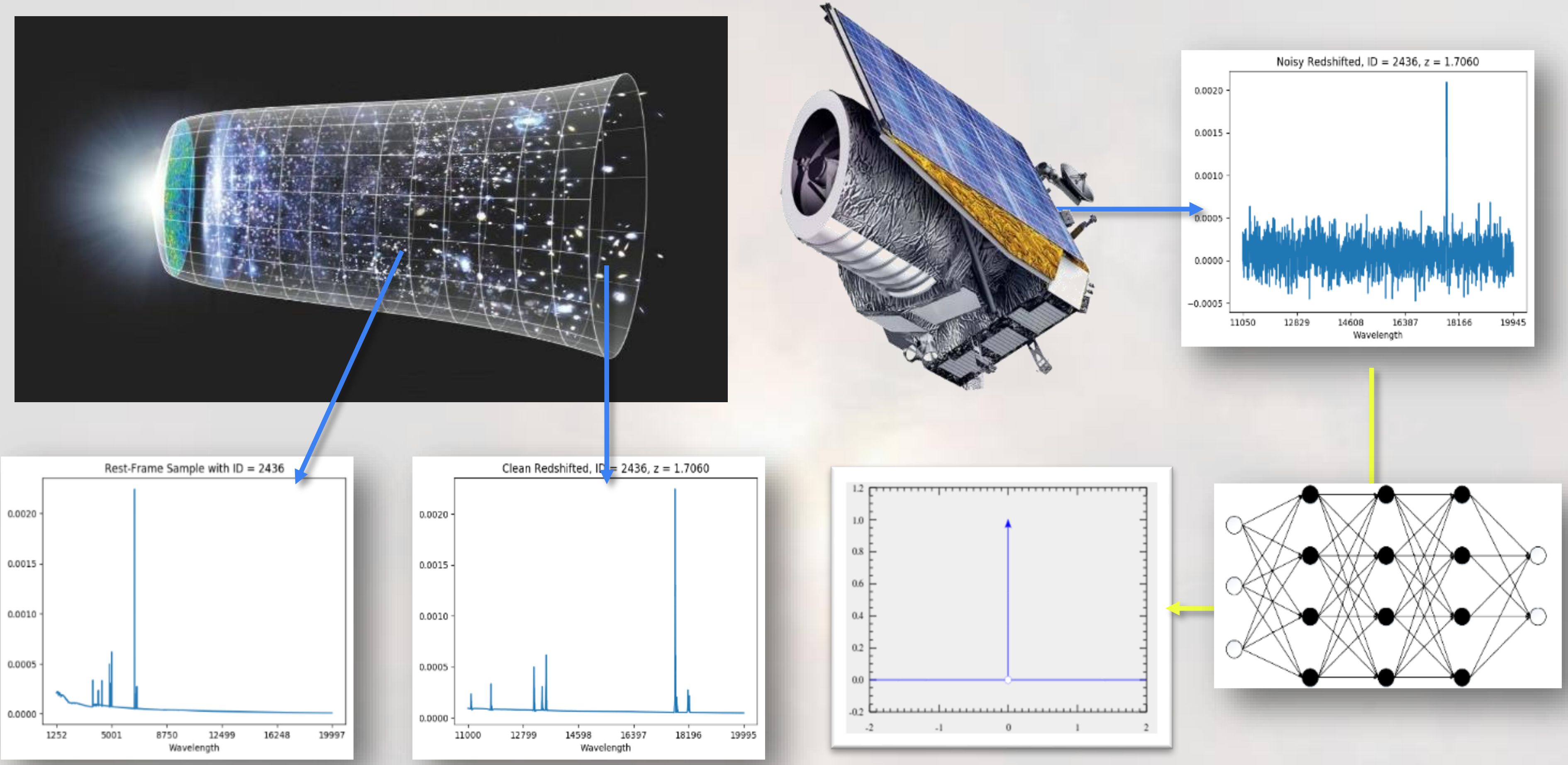
A Comparative Analysis of Deep Learning Methods for Spectroscopic Redshift Estimation

Greg Tsagkatakis

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

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

Spectroscopic Red-Shift Estimation

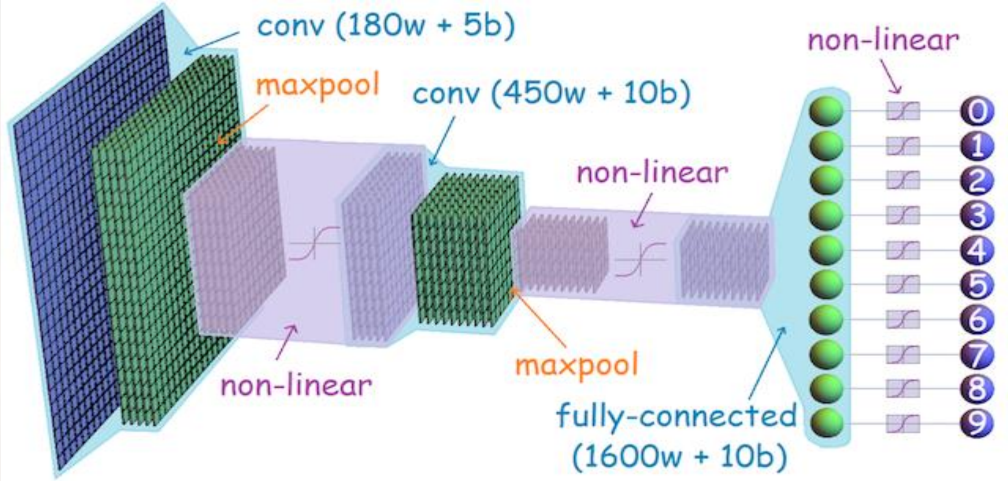


R. Stivaktakis, G. Tsagkatakis, B. Moraes, F. Abdalla, J-L Starck, P. Tsakalides, "Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data", IEEE Transactions on Big Data (Special Issue on Big Data From Space), 2020

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

Convolutional Neural Networks

- Inspired by the concept of "visual receptive fields"



- Automated feature extractors

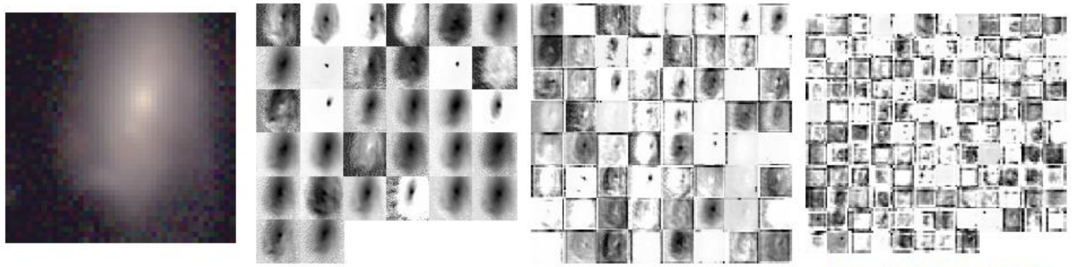
- Exhibit spatial correlations of the given input

- less prone to overfitting
  - local-connectivity property & weight-sharing lead to a dramatic decrease in total parameters
  - effective utilization of Big Data results in a high-generalization capacity
  - Dropout & Batch Normalization

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

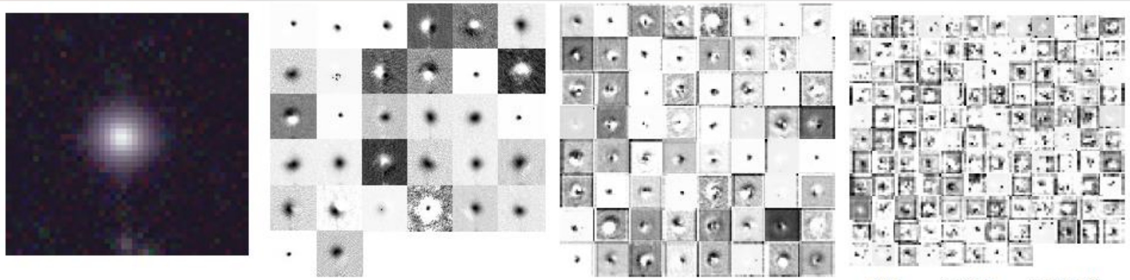
Applications of Deep Learning methods

Galaxy



(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

Star



(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

type	filters	filter size	padding	non-linearity	initial weights	initial biases
convolutional	32	5×5	-	leaky ReLU	orthogonal	0.1
convolutional	32	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	64	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
convolutional	128	3×3	1	leaky ReLU	orthogonal	0.1
pooling	-	2×2	-	-	-	-
fully-connected	2048	-	-	leaky ReLU	orthogonal	0.01
fully-connected	2048	-	-	leaky ReLU	orthogonal	0.01
fully-connected	2	-	-	softmax	orthogonal	0.01

Kim, Edward J., and Robert J. Brunner. "Star-galaxy classification using deep convolutional neural networks." *Monthly Notices of the Royal Astronomical Society* (2016): stw2672.

Predictive model

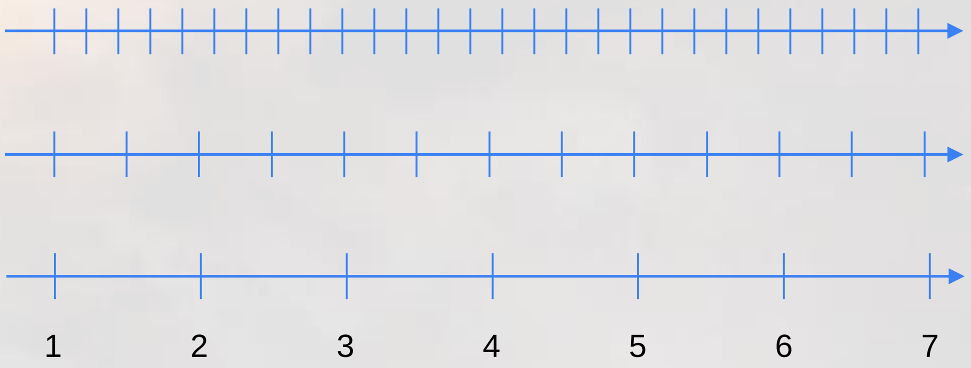
Redshift Estimation

Real-valued, non-negative number (z)

Regression Analysis

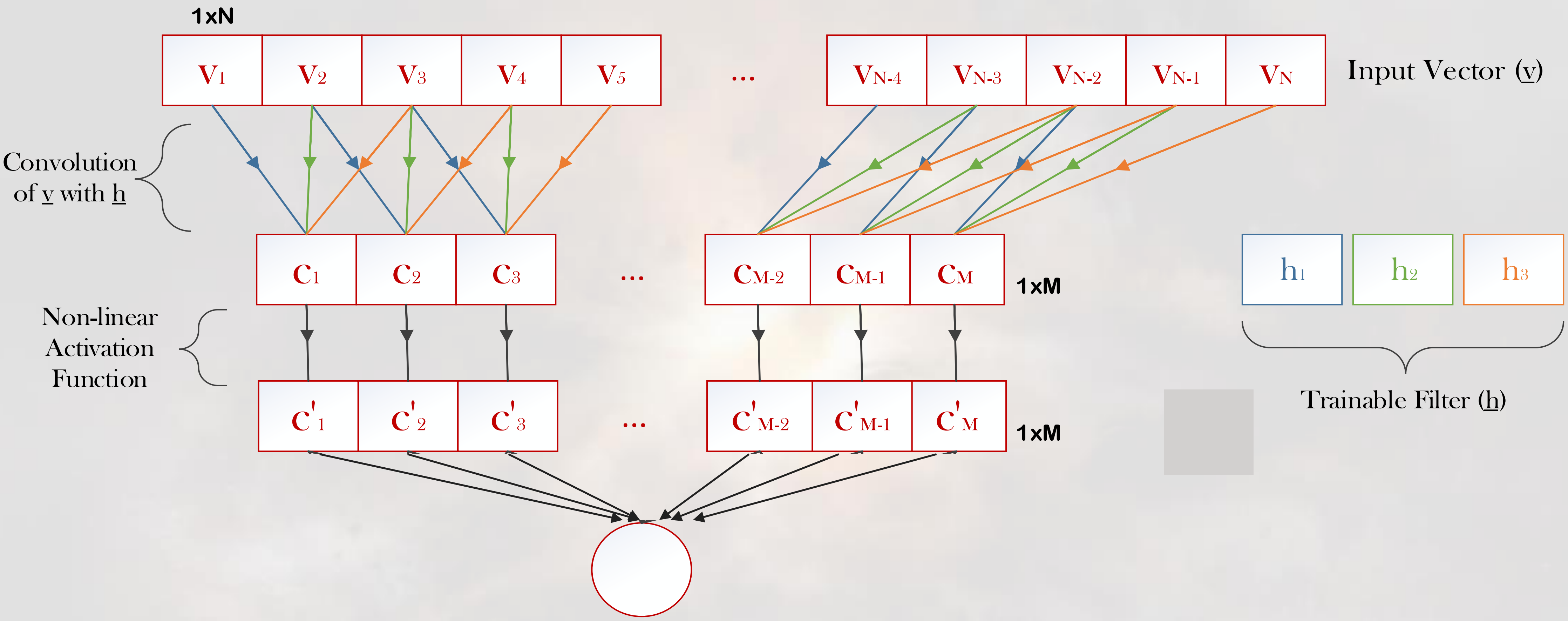
Split the examined redshift interval into ordinal classes, based on Euclid's characteristic resolution

Classification Problem



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

1-Dimensional CNN - Regression



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

1-Dimensional CNN - Regression

Consider a regression model where the predicted value $\hat{y}_i$ is the output of a function f, parameterized by weights $\mathbf{w}$, for the input x_i:

$$\hat{y}_i = f(x_i; \mathbf{w})$$

The mean squared error (MSE) between the predicted values $\hat{y}_i$ and the actual values y_i over n observations is defined as:

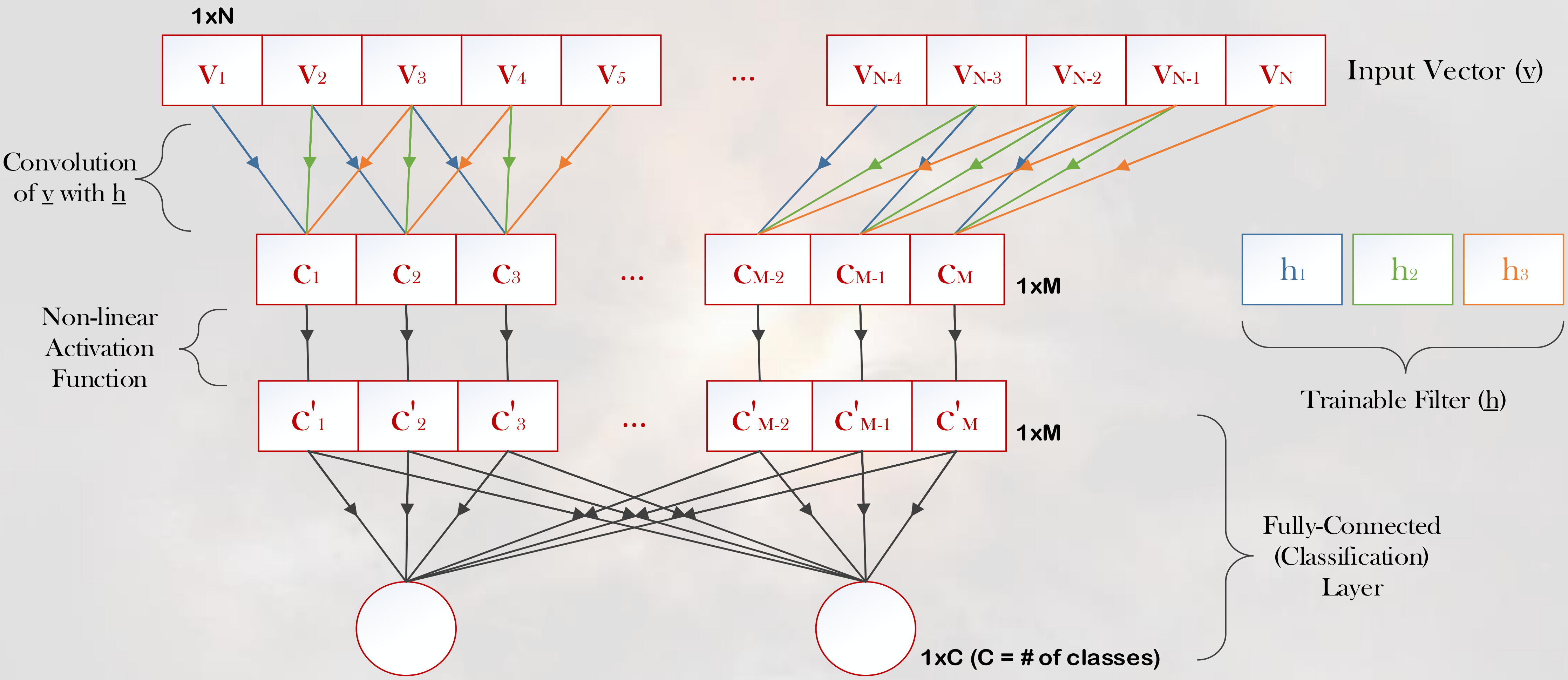
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i; \mathbf{w}) - y_i)^2$$

where:

- $f(x_i; \mathbf{w})$ is the prediction from the regression function for the i-th observation,
- y_i is the actual value for the i-th observation,
- n is the number of observations.

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

1-Dimensional CNN - Classification



Funded by the European Union

Classification specifics

1. **Discretization:** Divide the range of input values $\{x_1, x_2, \ldots, x_n\}$ into k equal-sized bins. Each value x_i is mapped to a discrete bin b_j such that:

$$b_j = \left\lfloor \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \cdot k \right\rfloor$$

   where $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, and $\lfloor \cdot \rfloor$ denotes the floor function.

2. **Feature Encoding:** Convert the discrete bins b_j into a one-hot encoded vector $\mathbf{v}_i$, which serves as the input to the neural network.

3. **Neural Network Classification:** Apply a neural network f parameterized by weights $\mathbf{w}$ to the encoded vectors $\mathbf{v}_i$:

$$\hat{y}_i = f(\mathbf{v}_i; \mathbf{w})$$

   where $\hat{y}_i$ is the predicted class label for the input x_i.

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

Classification specifics

The categorical cross-entropy loss between a true distribution p and a predicted distribution q over C classes is defined as:

$$H(p, q) = -\sum_{c=1}^{C} p(c) \log q(c)$$

where:

- $p(c)$ is the true probability of class c,
- $q(c)$ is the predicted probability of class c.

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

DATASET & METHODOLOGY

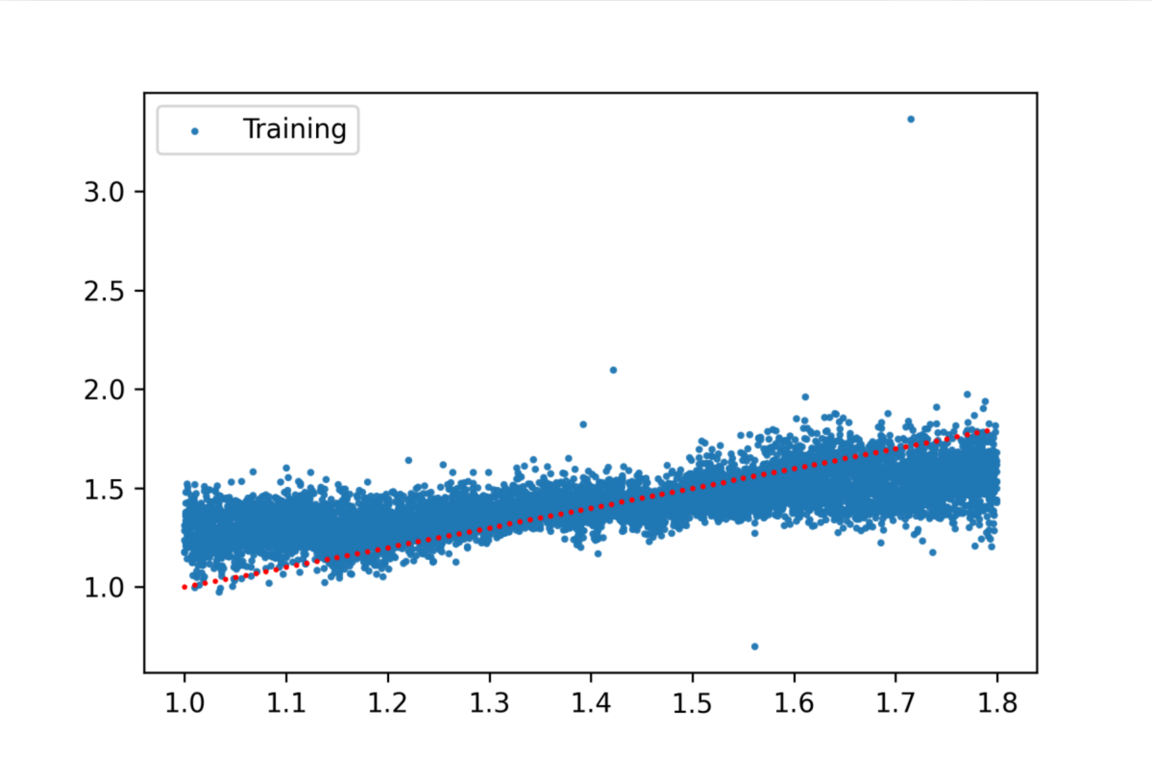
- 10K initial clean rest-frame spectral profiles (SEDs)

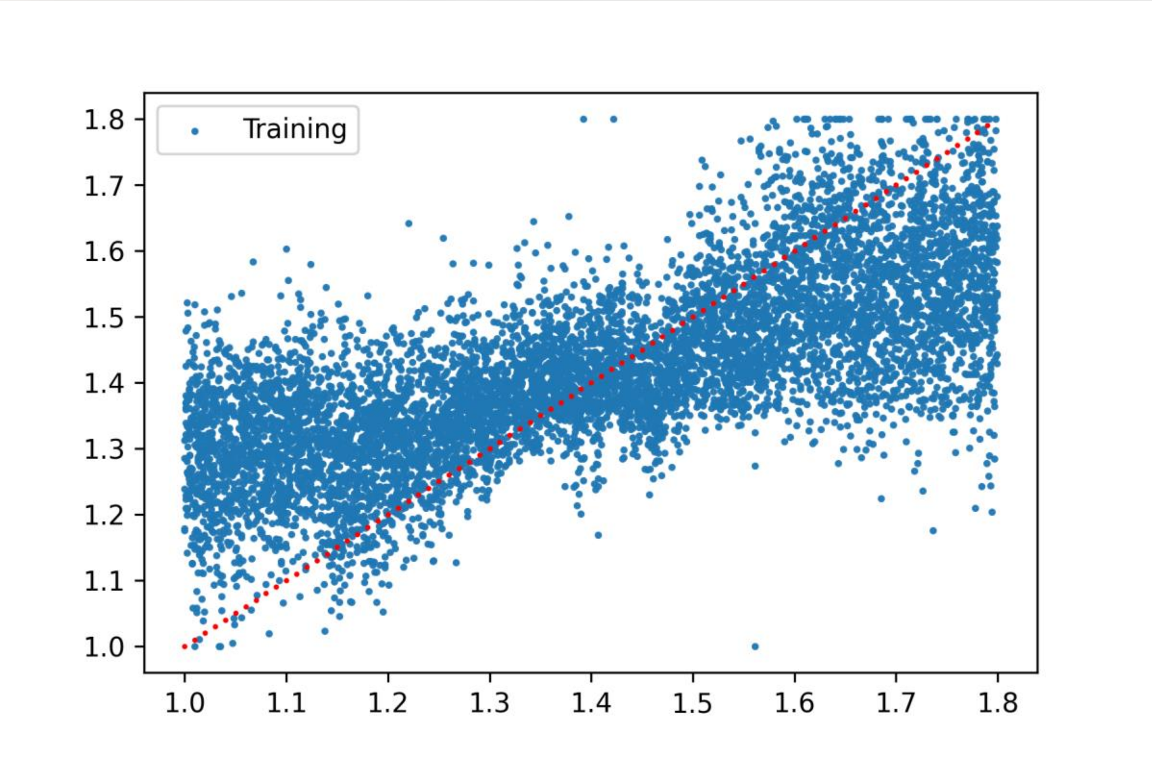
- Generation of randomly redshifted examples
    - $z = [1, 1.8)$, similar to Euclid's specification
    - $\log(1+z) = \log(\lambda_{observed}) - \log(\lambda_{emit}) \leftrightarrow 1 + z = \frac{\lambda_{observed}}{\lambda_{emit}}$
    - optional addition of white Gaussian noise (idealistic vs. realistic)

- Quantization of the utilized redshift range
    - split into 800 discrete classes, which implies a resolution of 0.001

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

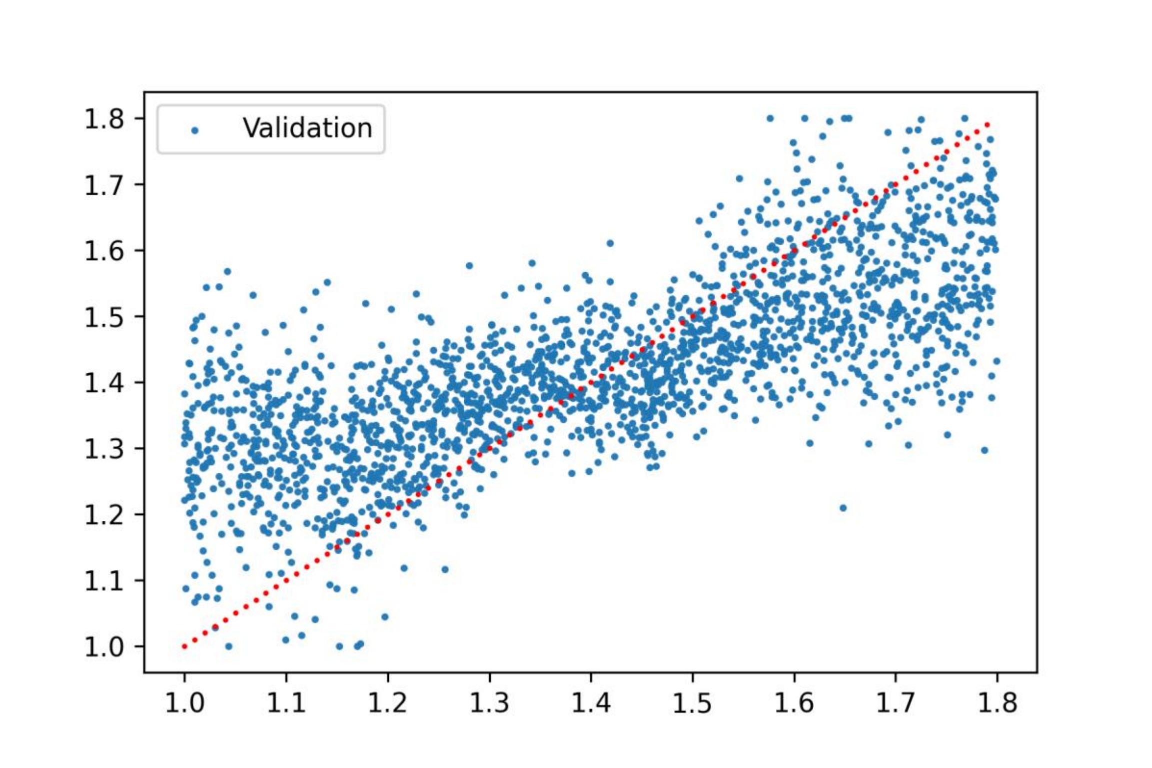
Regression





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

Regression

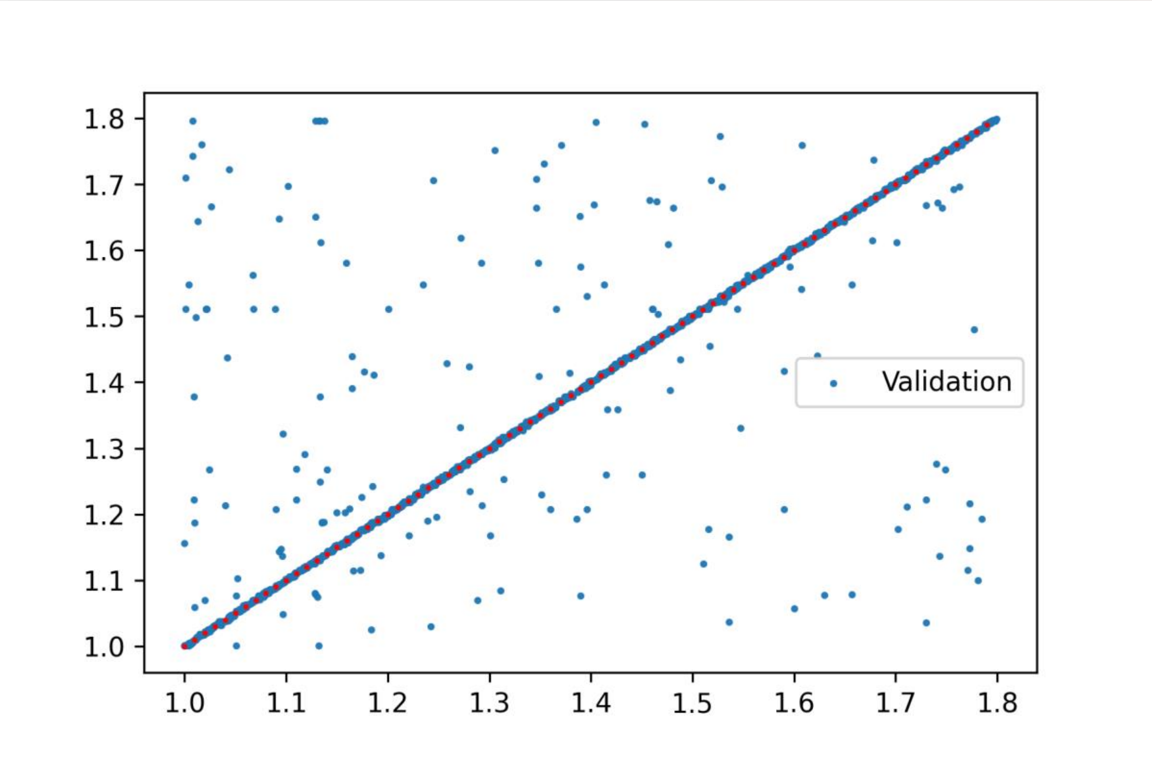


MSE	0.03811
MAE	0.16469
R^2	0.28188

Signal Processing Laboratory

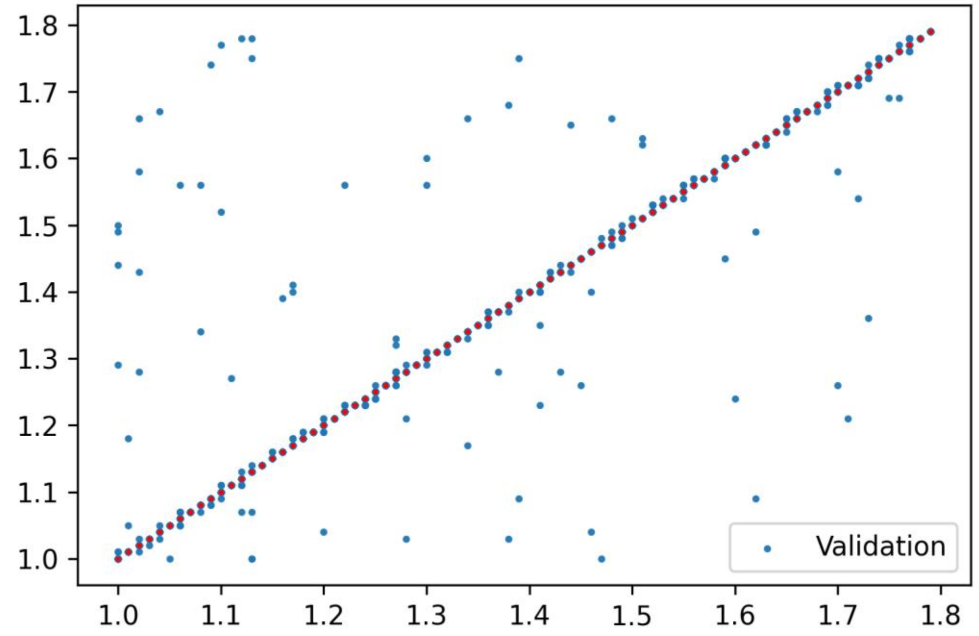
Funded by the European Union

Classification: 800 classes



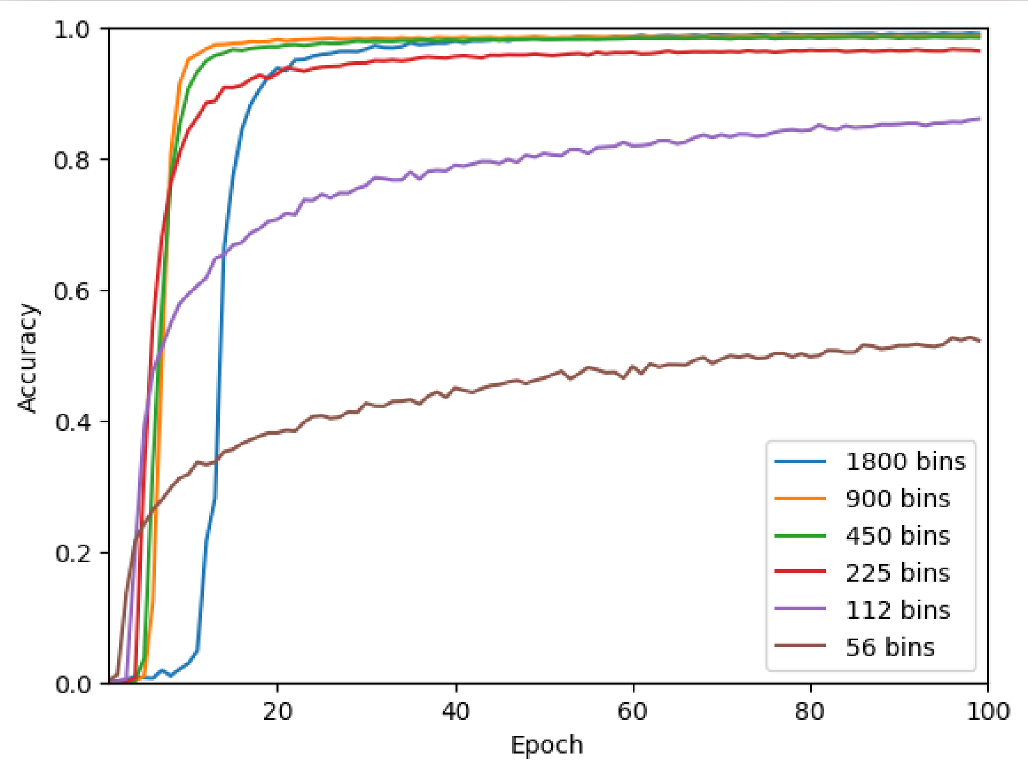
MSE 0.00791
MAE 0.02325
R^2 0.85082

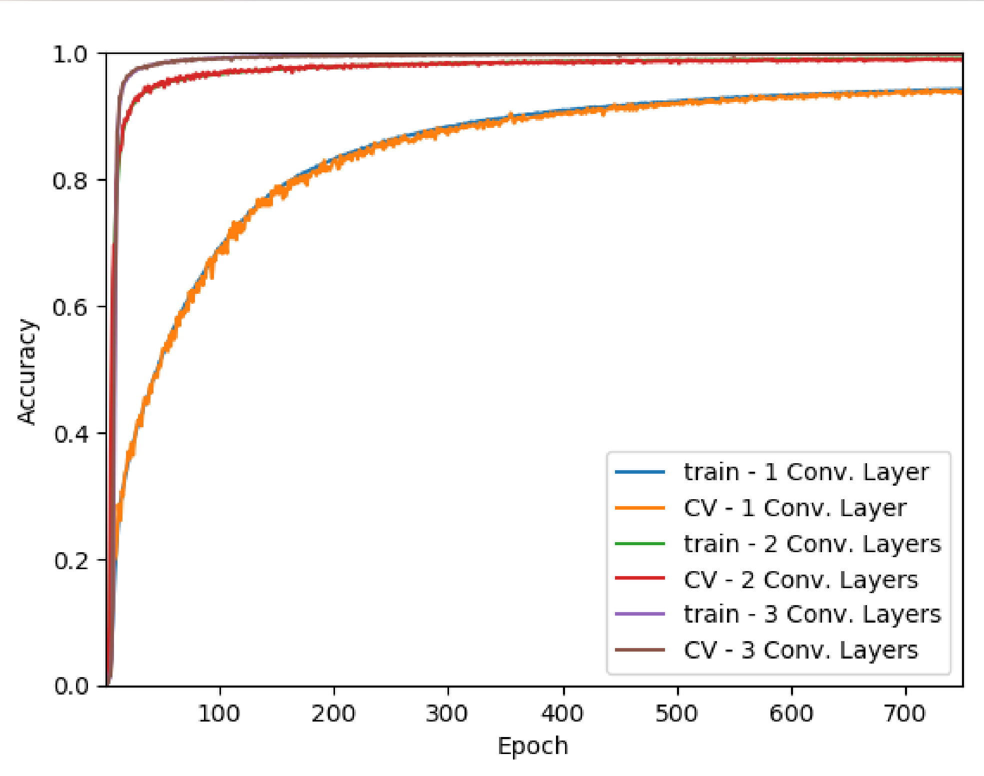
Classification: 80 classes



MSE 0.00364
MAE 0.00968
R^2 0.93128

Impact of network architecture





TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH INSTITUTE OF COMPUTER SCIENCE

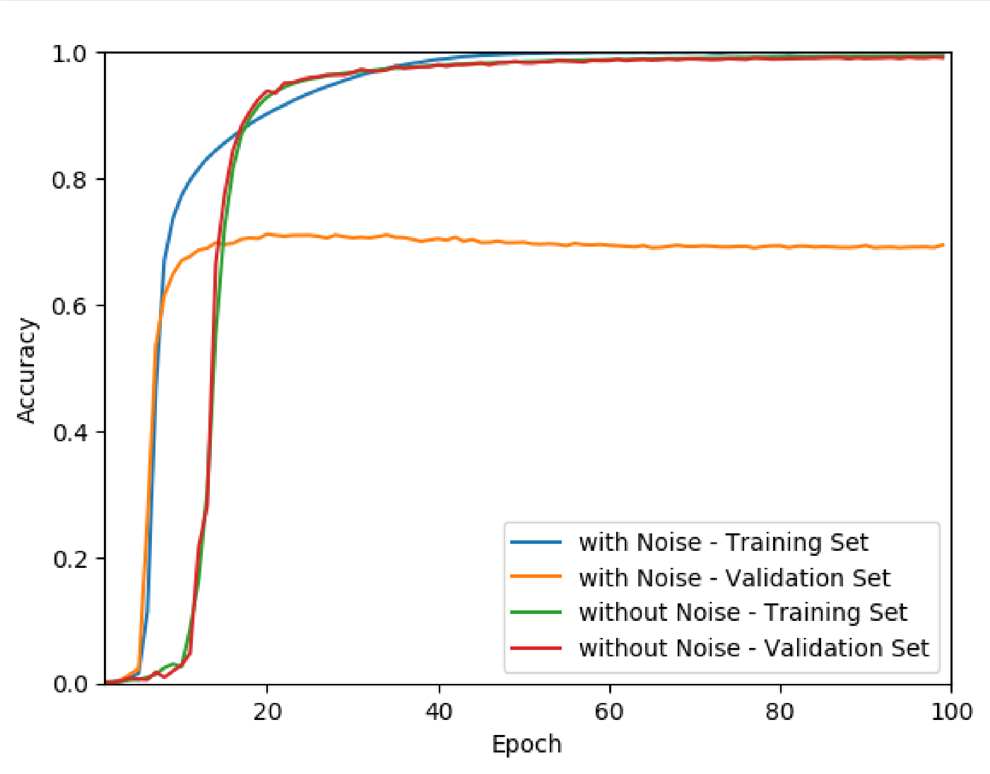
Signal Processing Laboratory

CosmoStat

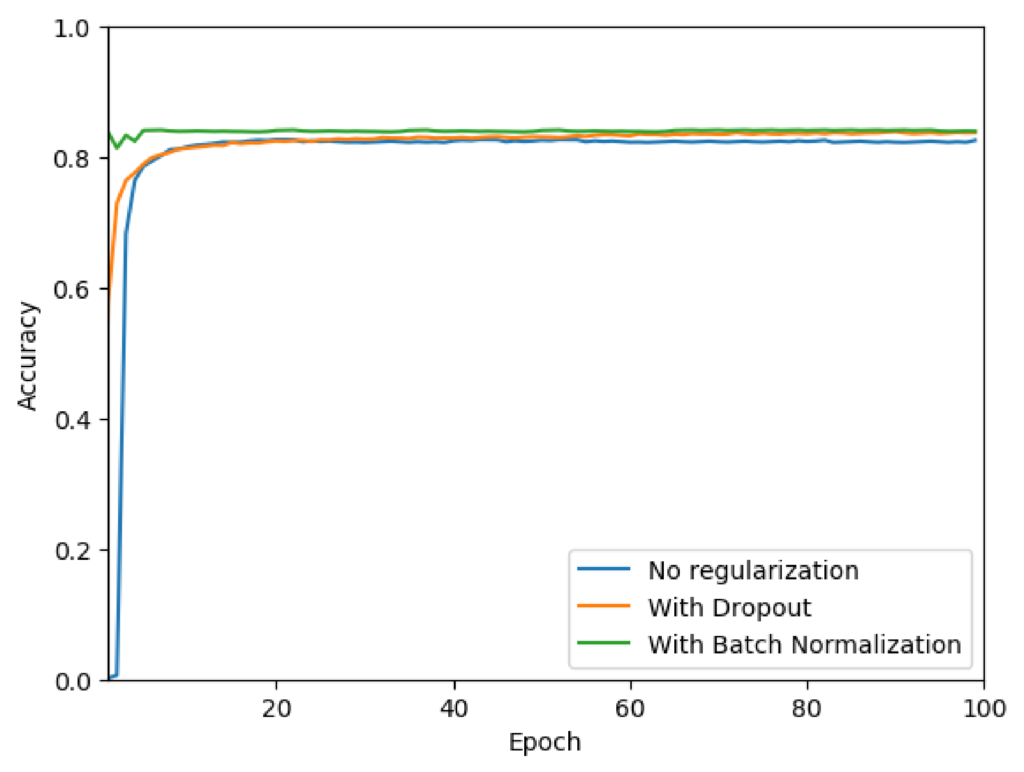
cea

Funded by the European Union

Impact of training set size



400,000 Tr. Samples



4,000,000 Tr. Samples

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH INSTITUTE OF COMPUTER SCIENCE

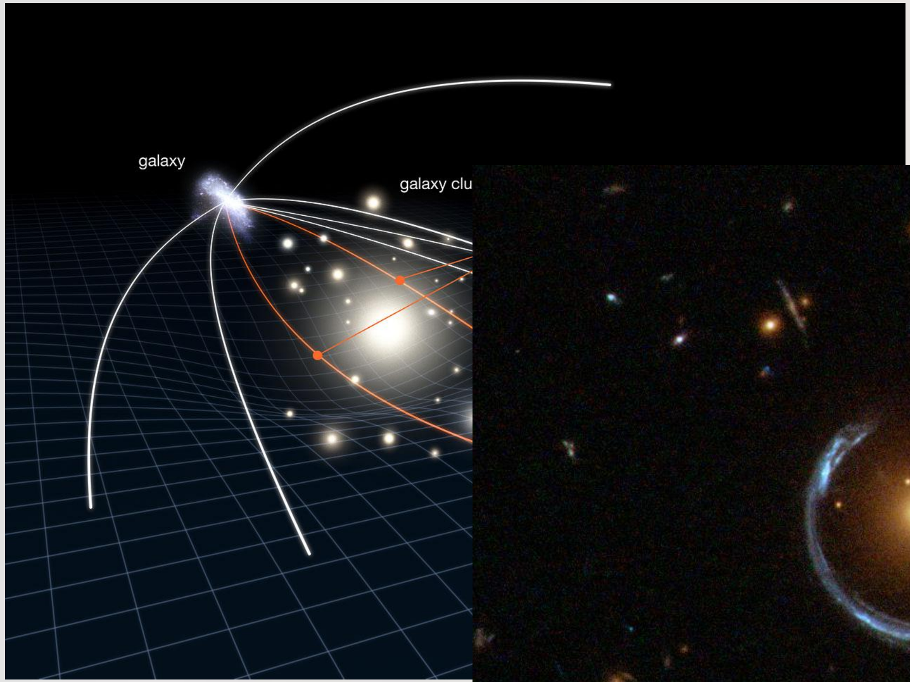
Signal Processing Laboratory

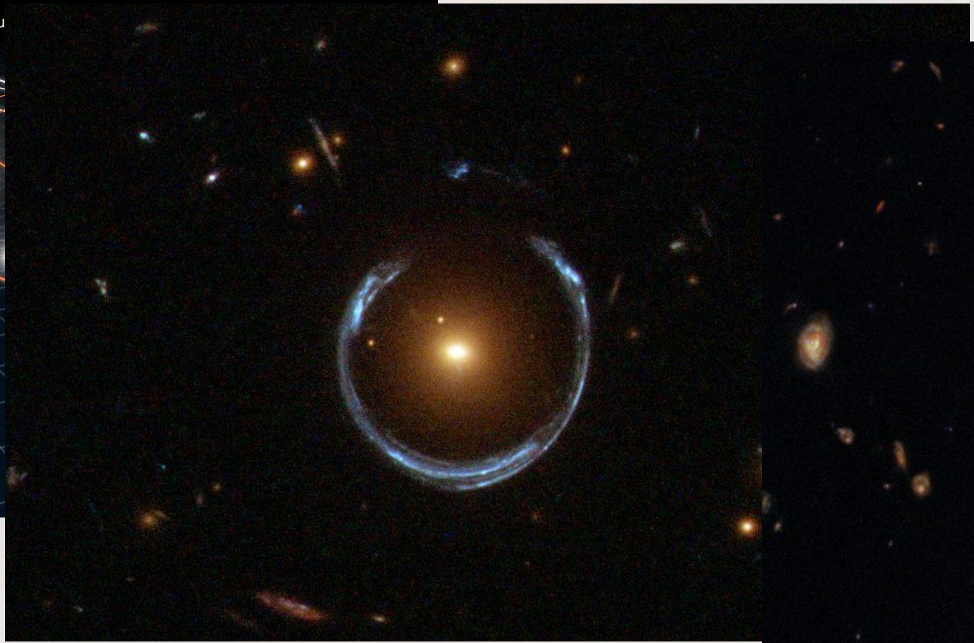
CosmoStat

cea

Funded by the European Union

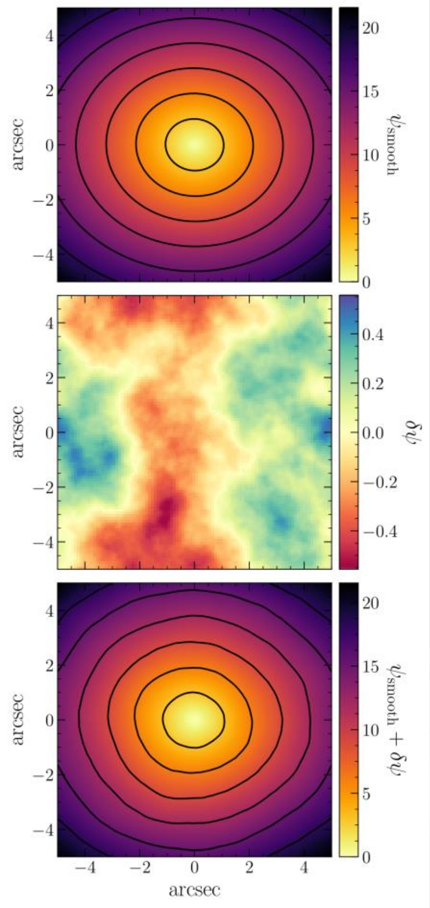
Gravitational Lensing





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

Modeling with uncertain labels



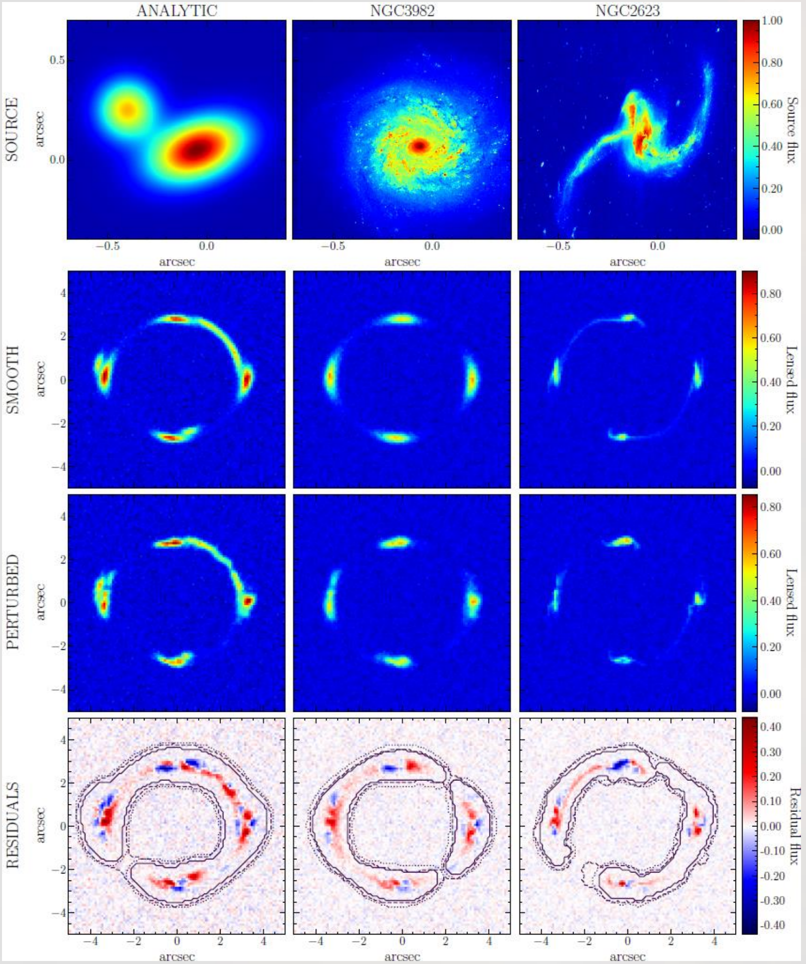
Singular Isothermal Ellipsoid parametric model

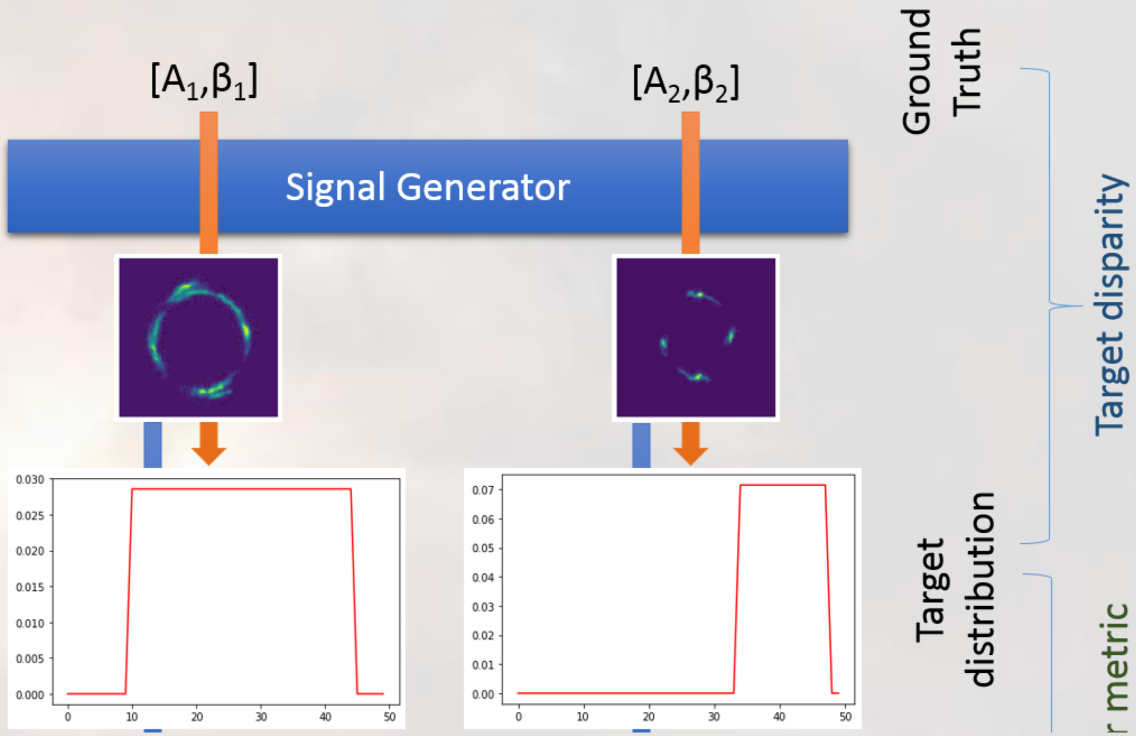
Realization of Gaussian Random Field perturbations

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.

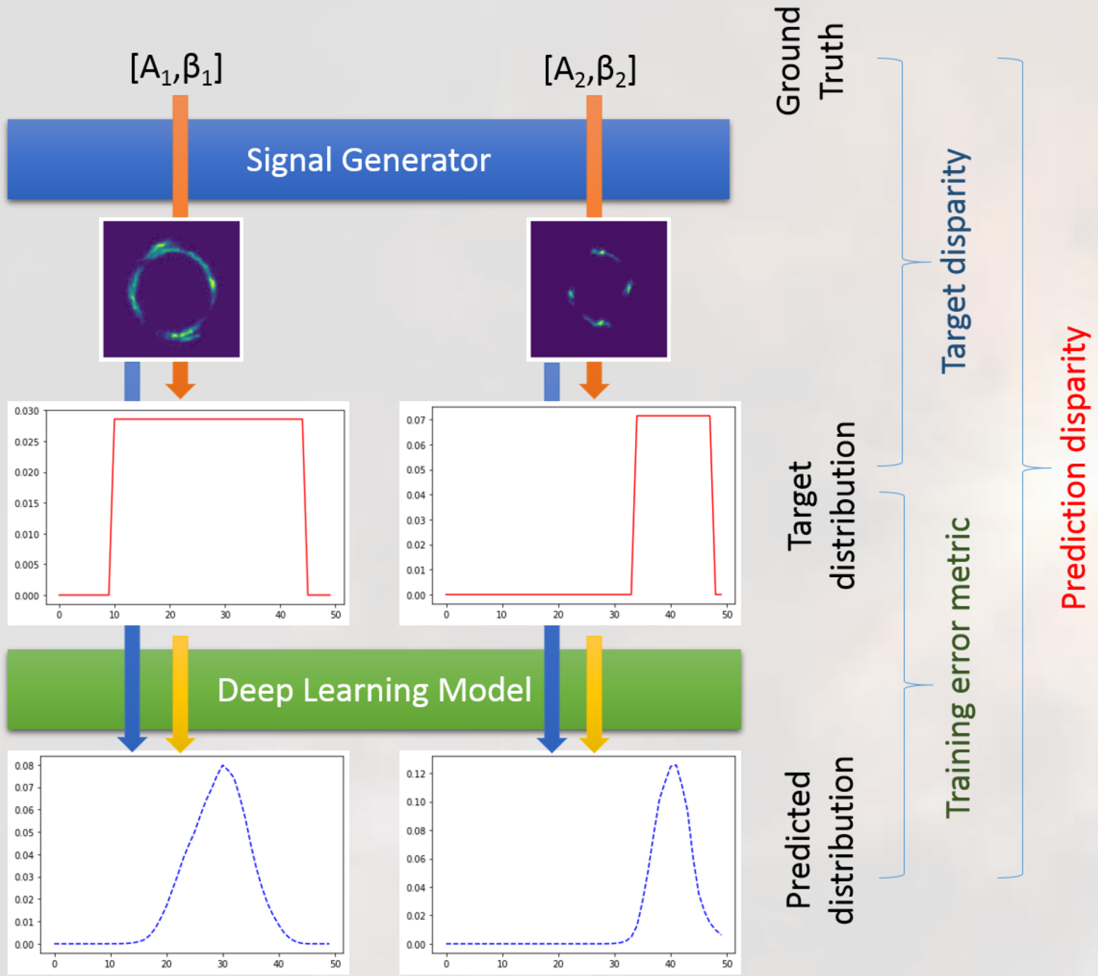
Label uncertainty





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

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:

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

where $M = \frac{1}{2}(P + Q)$.

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

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

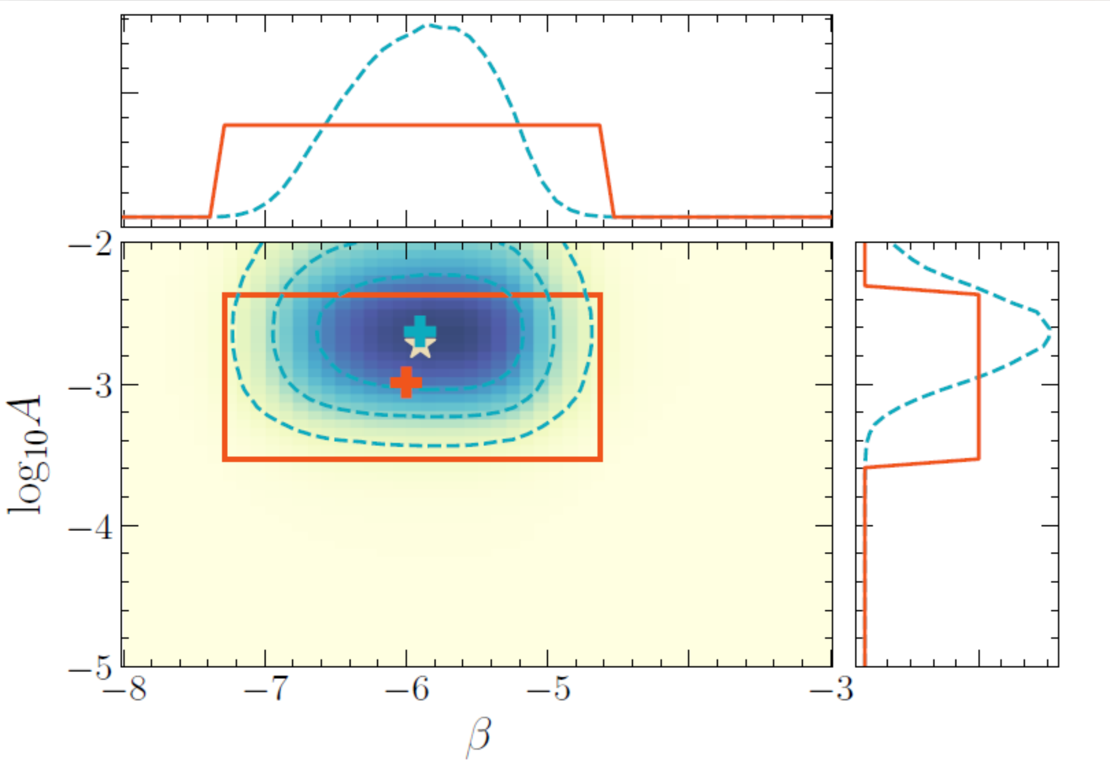
Entropy-regularized version of the JS divergence and is given by:

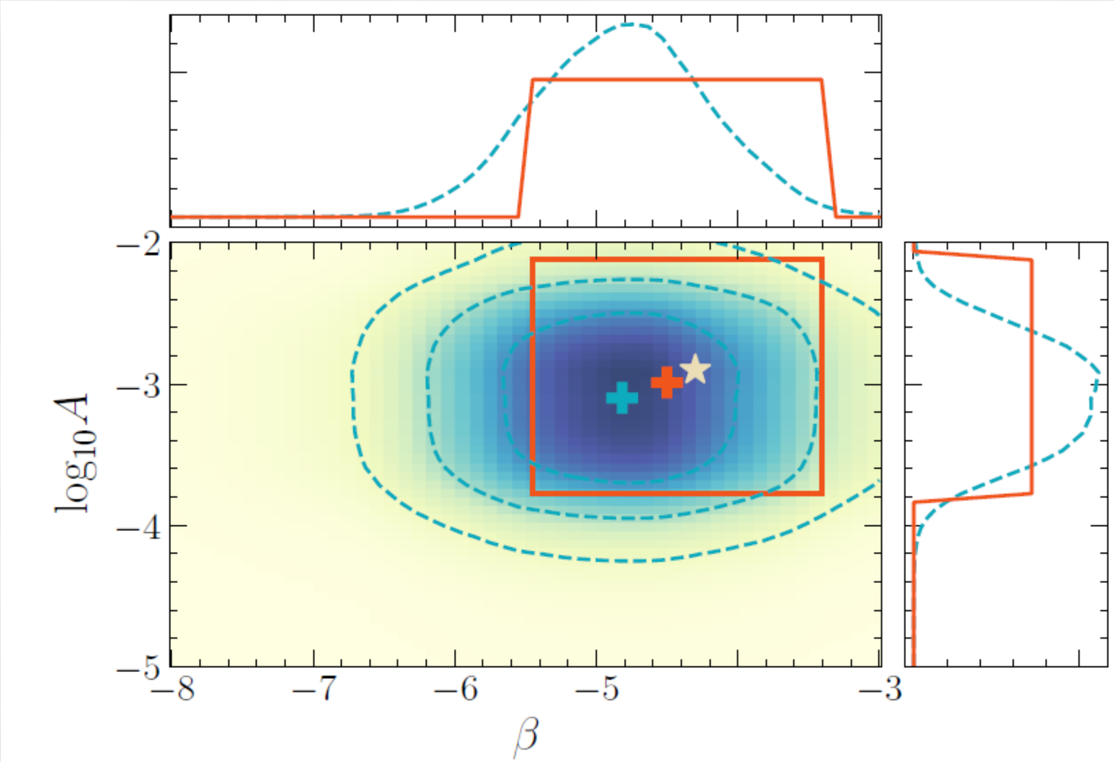
$$\mathcal{L}(P, Q) = \lambda_1 JS(P, Q) + \lambda_2 H(P),$$

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

ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
INSTITUTE OF COMPUTER SCIENCE
Laboratory
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

Discussion - Uncertainty

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.

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

Discussion - Uncertainty




cat

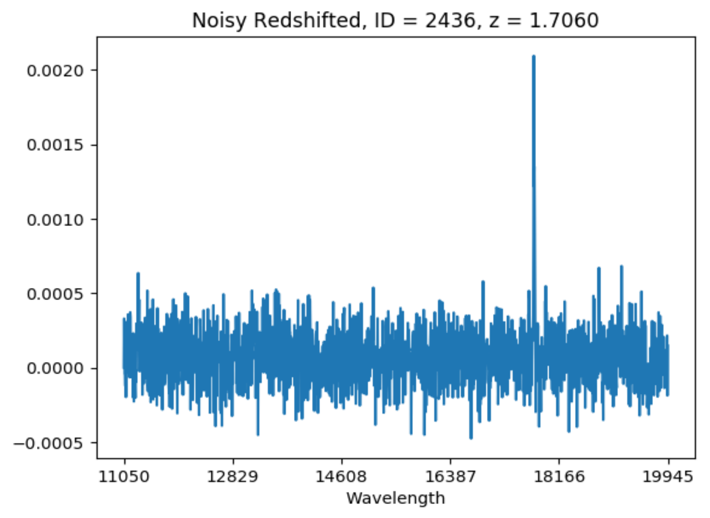
dog

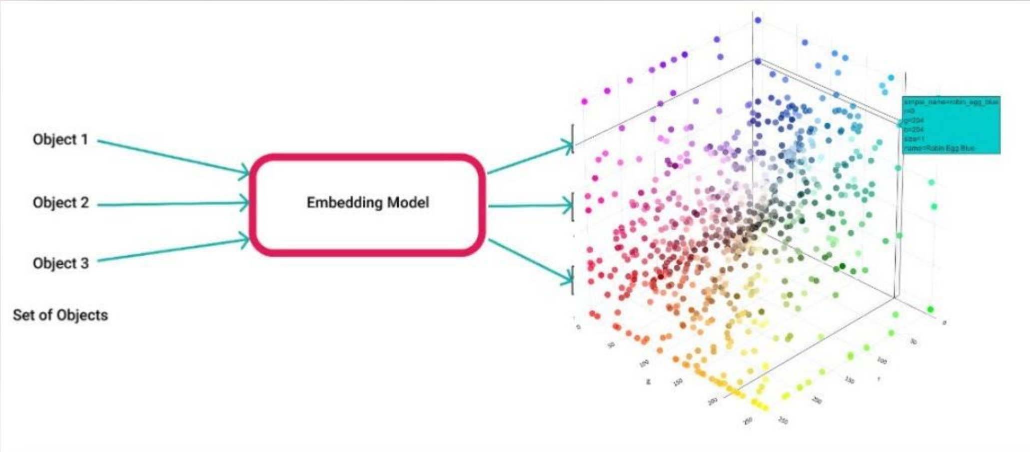


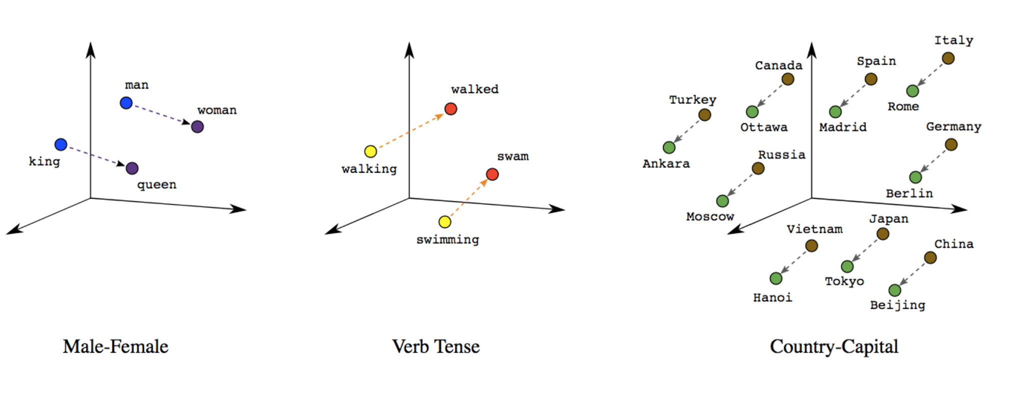
Aleatoric Uncertainty:

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

Discussion – Vector databases







Thank you

TITAN
ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH
INSTITUTE OF COMPUTER SCIENCE

Signal Processing Laboratory

CosmoStat

cea

Funded by the European Union