# Deep Learning Applications in Astrophysics

GREGORY TSAGKATAKIS

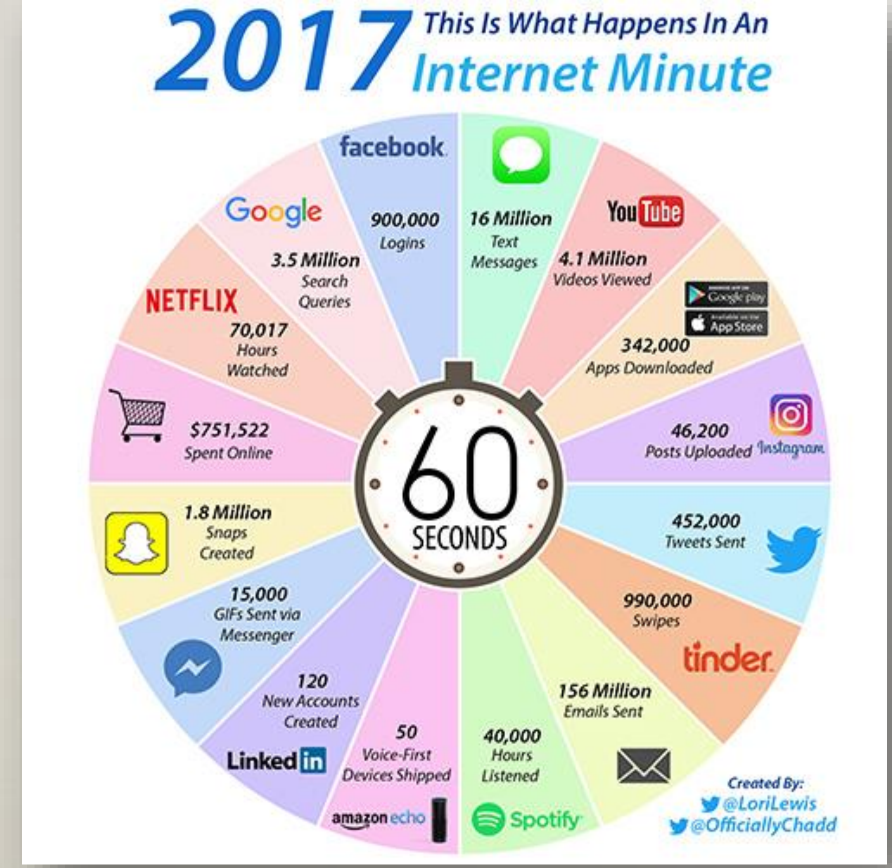
SIGNAL PROCESSING LAB

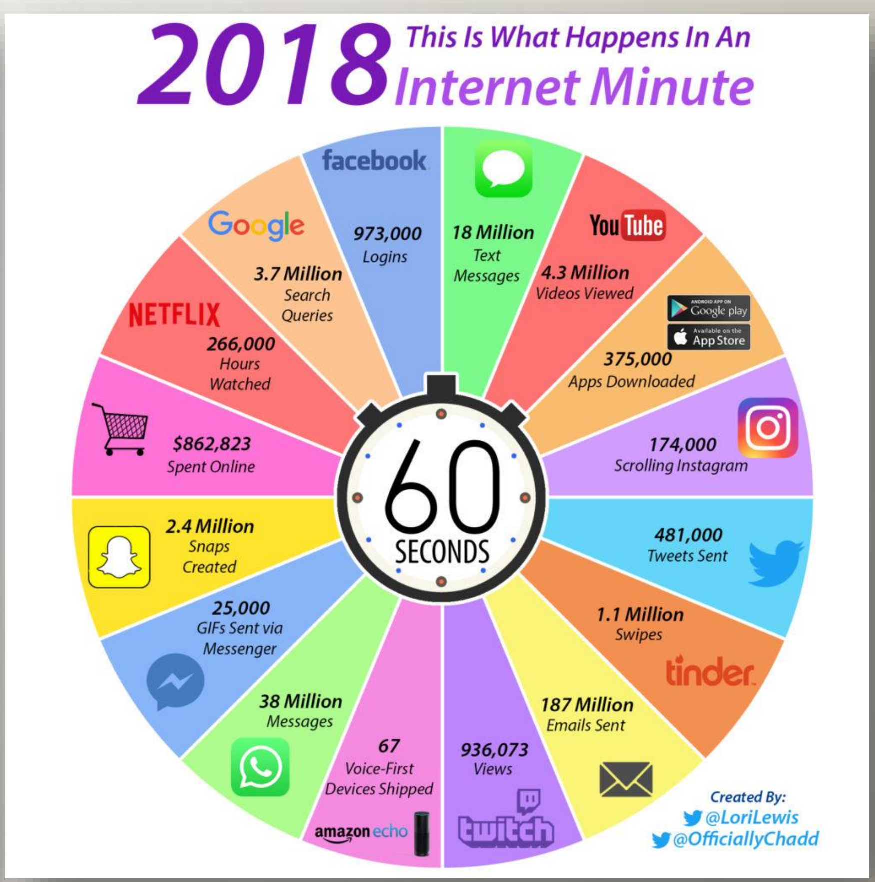
ICS - FORTH

# Outline

- The case of Big Data
- The advent of Deep Learning
- Applications in Astrophysics
- Spectroscopic red-shift estimation
- The future

# The Big Data era



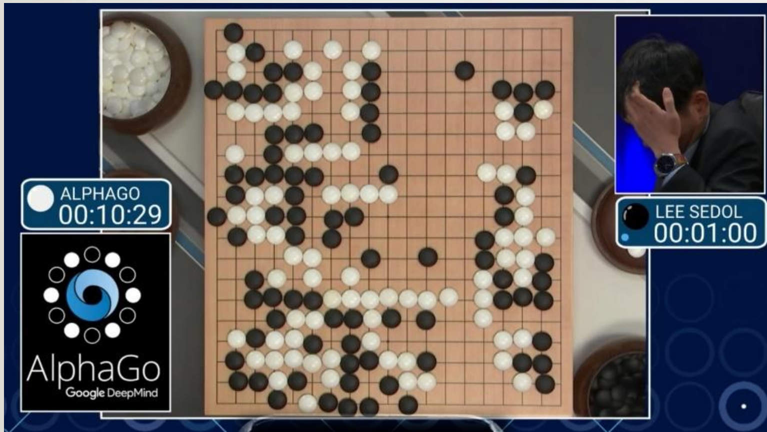



# The Big Data era in Astrophysics

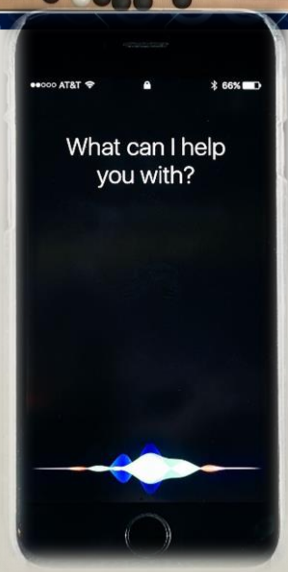
| Sky Survey Project | Volume | Velocity | Variety |
|---|---|---|---|
| Sloan Digital Sky Survey (SDSS) | 50 TB | 200 GB per day | Images, redshifts |
| Large Synoptic Survey Telescope (LSST ) | ~ 200 PB | 10 TB per day | Images, catalogs |
| Square Kilometer Array (SKA ) | ~ 4.6 EB | 150 TB per day | Images, redshifts |

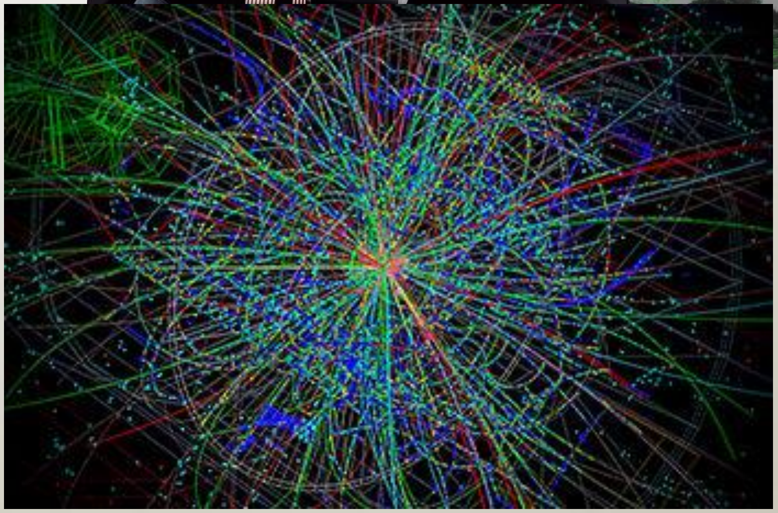
Garofalo, Mauro, Alessio Botta, and Giorgio Ventre. "Astrophysics and Big Data: Challenges, Methods, and Tools." Proceedings of the International Astronomical Union 12.S325 (2016)

# Advances in machine learning





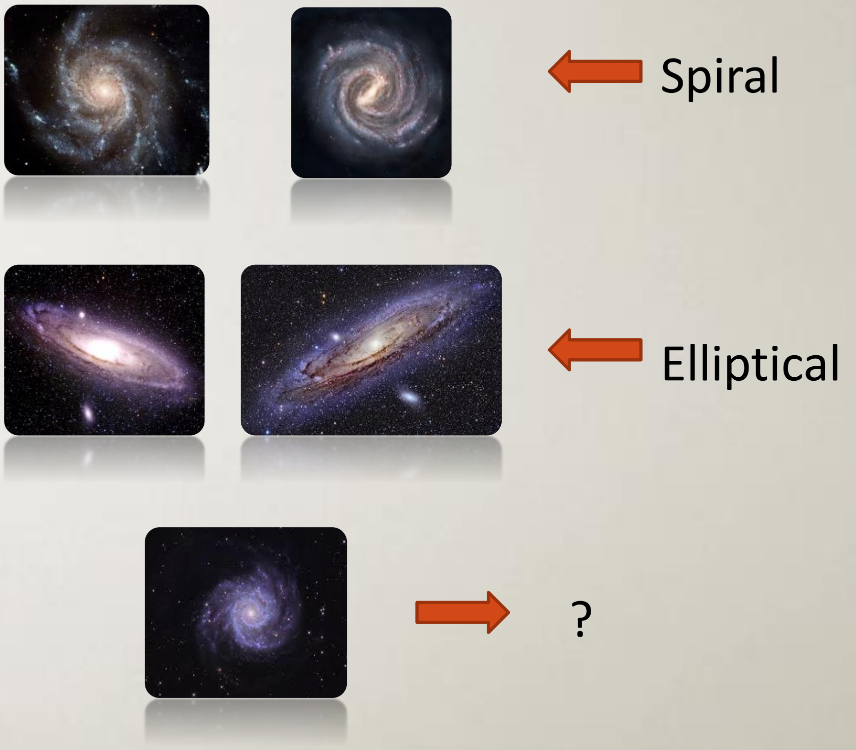


# Supervised Learning

Data
Labels

Model
Prediction

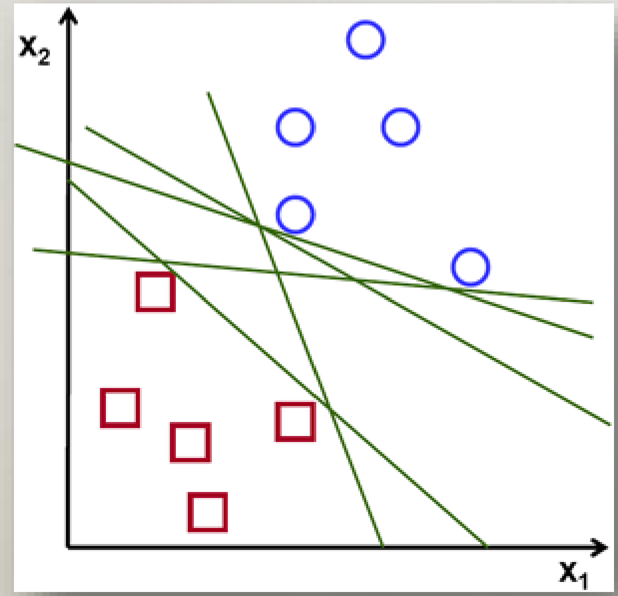


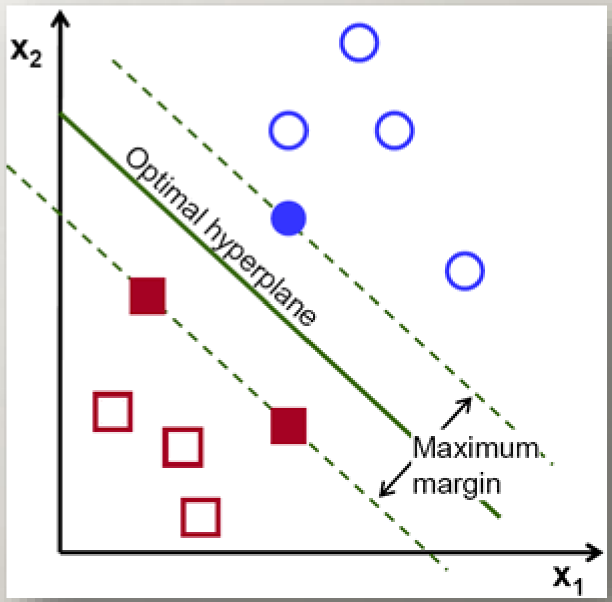

**Exploiting prior knowledge**

- Expert users
- Crowdsourcing
- Other instruments

# State-of-the-art (almost)

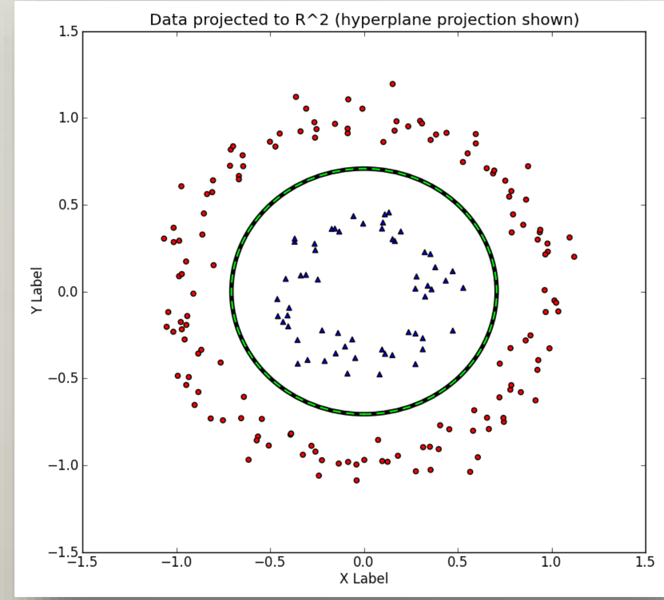
- Support Vector Machines
  - Binary classification

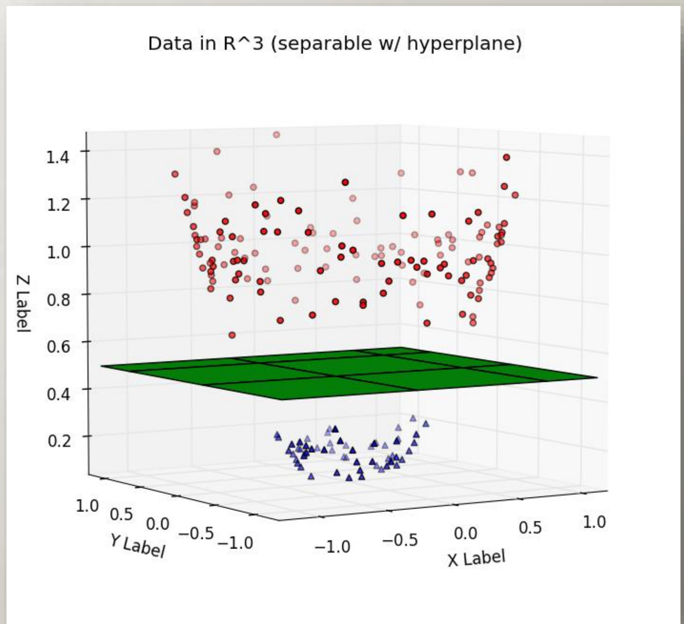


# State-of-the-art (almost)

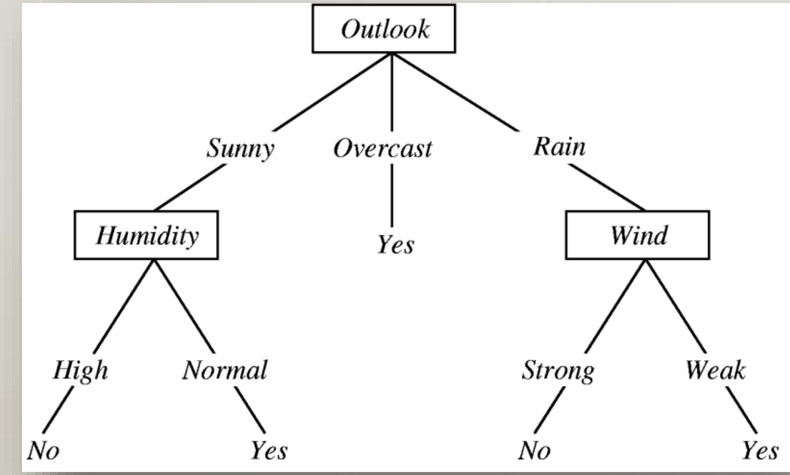
- Support Vector Machines
  - Binary classification
  - Kernels ↔ non-linearity

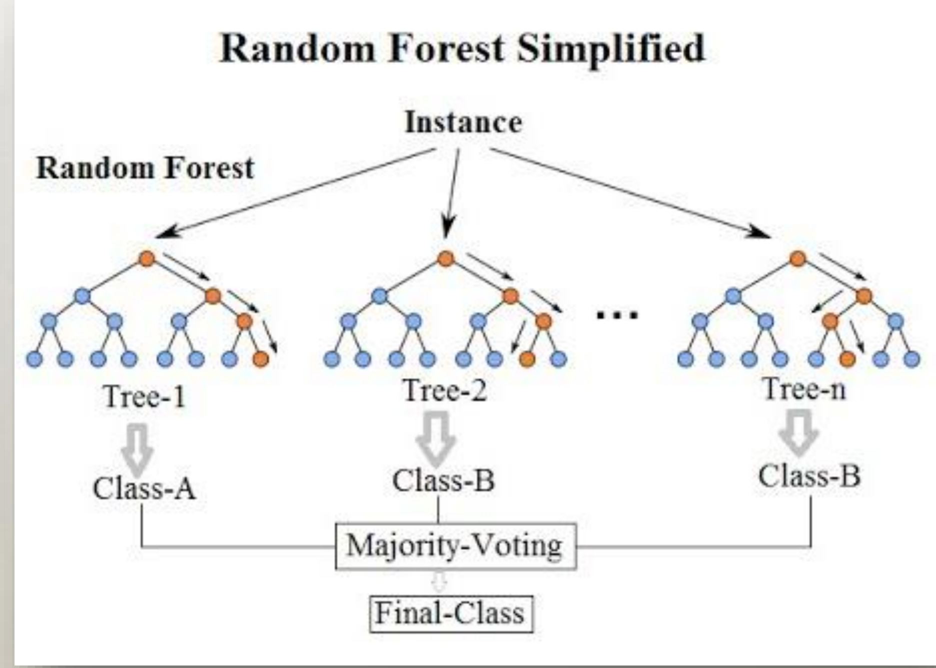


# State-of-the-art (almost)

- Support Vector Machines
  - Binary classification
  - Kernels ↔ non-linearity
- Random Forests
  - Multi-class classification

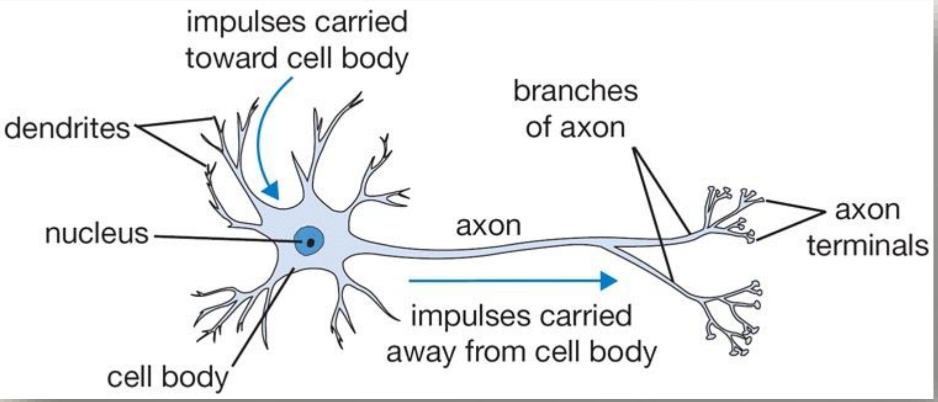


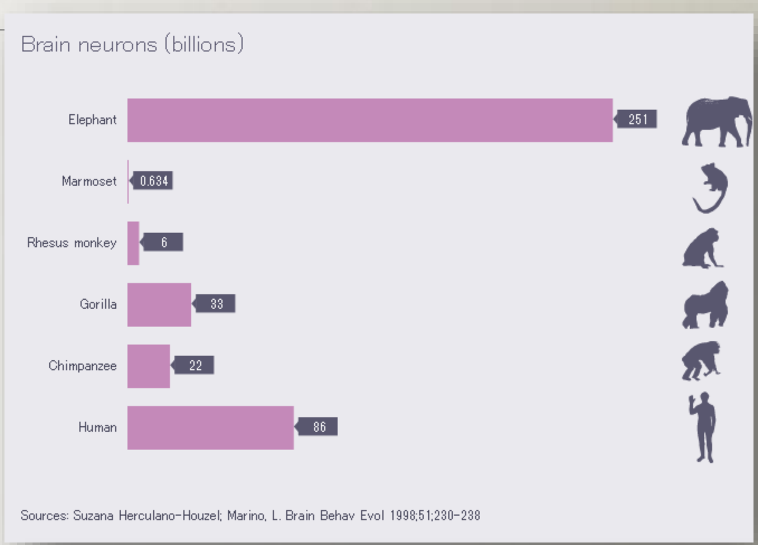



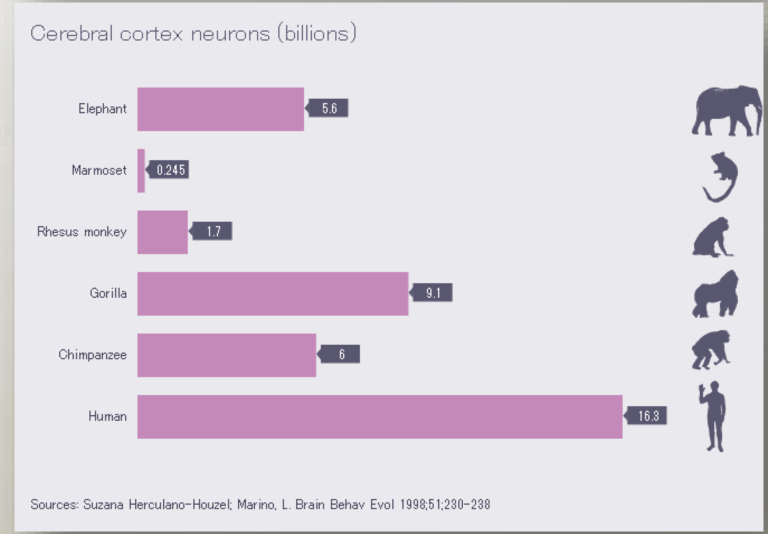
# Deep Learning

## Inspiration from brain

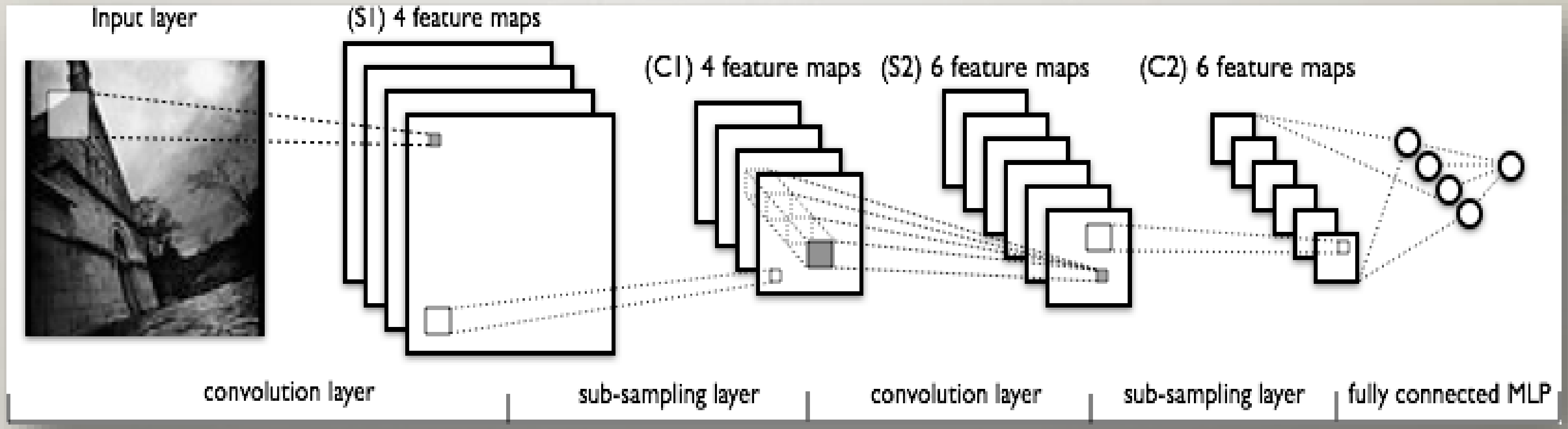
- 86Billion neurons
- $10^{14}$-$10^{15}$ synapses



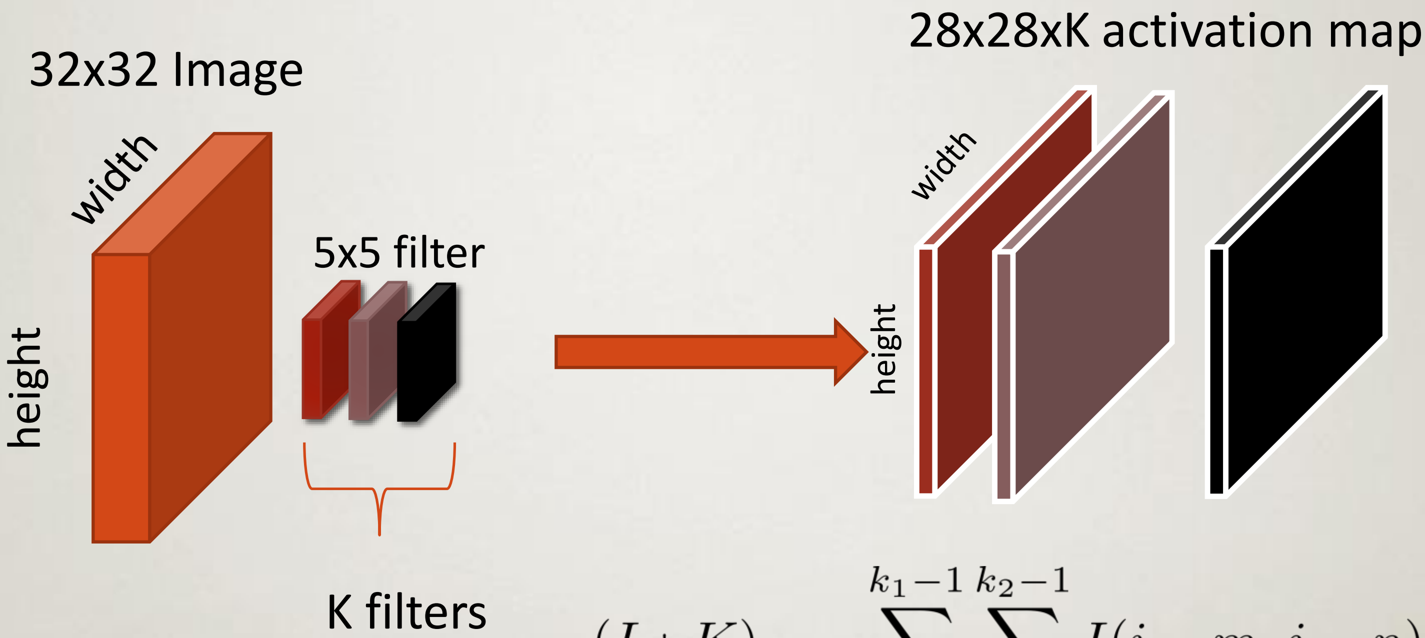
# Convolutional Neural Networks




(Convolution + Subsampling) + () … + Fully Connected

LeCun, Yann, and Yoshua Bengio. "Convolutional networks for images, speech, and time series." The handbook of brain theory and neural networks 3361.10 (1995): 1995.

# Convolutional Layers

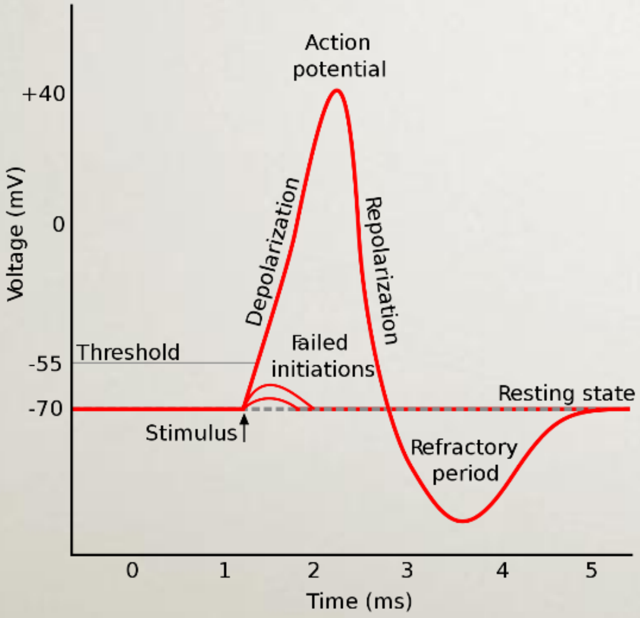


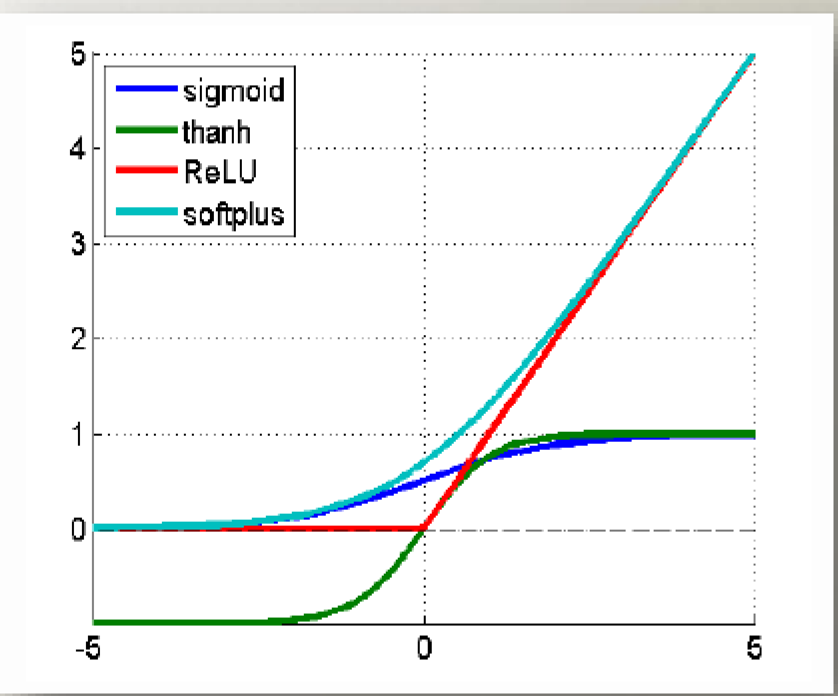

$$(I * K)_{ij} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i-m, j-n)K(m,n)$$
$$= \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} I(i+m, j+n)K(-m,-n)$$

# Activations

## Introduction of non-linearity

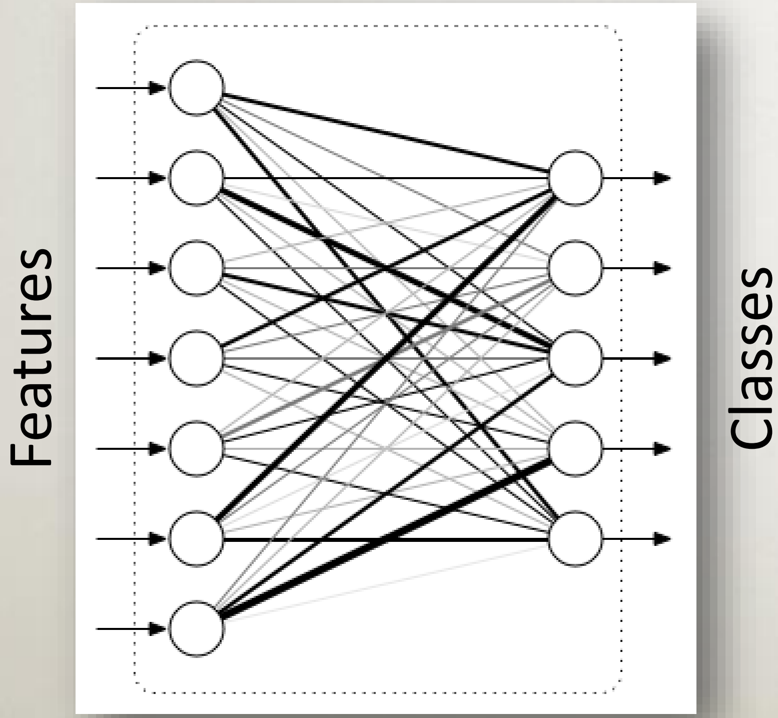
- Brain: thresholding -> spike trains



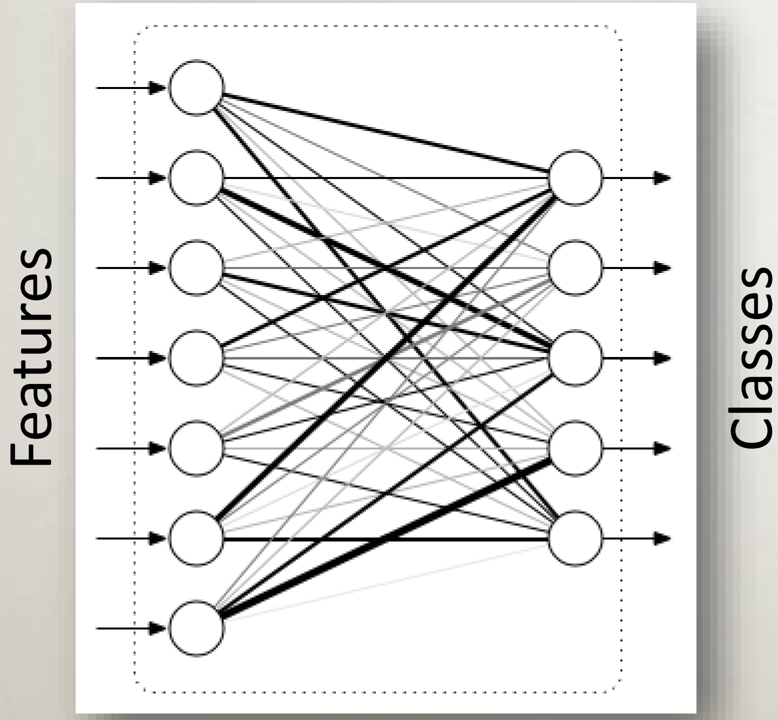
# Other types of layers

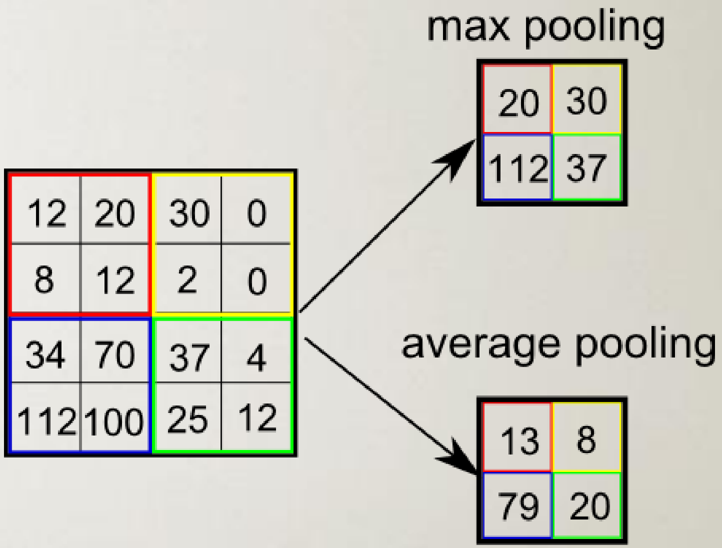
- Fully Connected Layers



# Other types of layers

- Fully Connected Layers

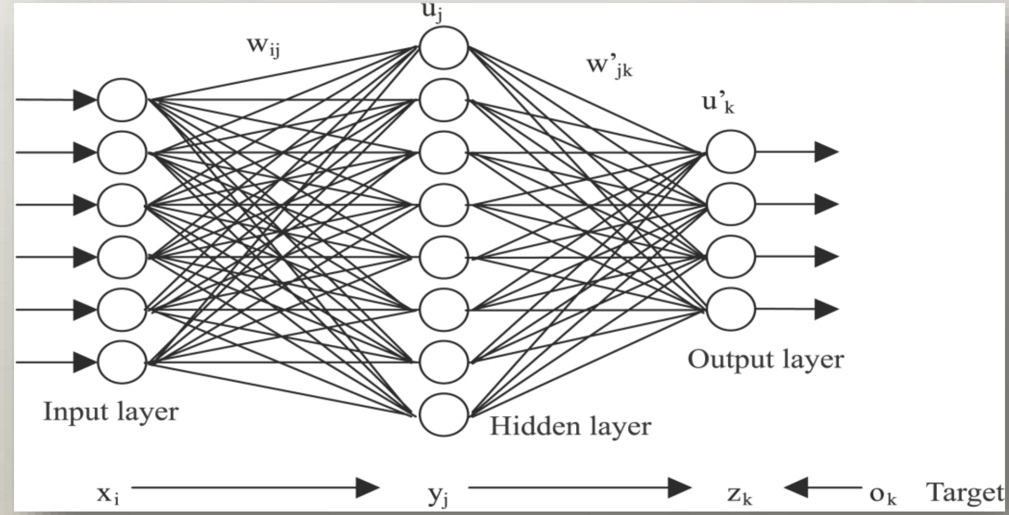



- Pooling layer

# Training DNN

1. Get batch of data
2. Forward through the network -> estimate loss
3. Backpropagate error
4. Update weights based on gradient




**Errors**

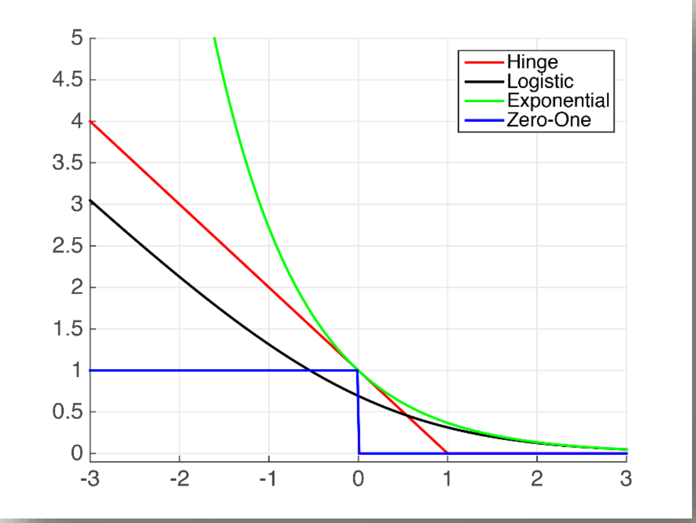
# Backpropagation

Chain Rule in Gradient Descent: Invented in 1969 by Bryson and Ho

**Defining a loss/cost function** $J(x, y; \theta) = \frac{1}{2} \sum (y - f(x;\theta))^2$

Assume a function

$$f(x;\theta) = \sigma(w^T x + b) \quad , \quad \theta = \{w, b\}$$




Types of Loss function

- Hinge $J(x, y) = max\{0, 1 - xy\}$
- Logistic $J(x, y) = log_2(1 + exp(-xy))$
- Cross-entropy $J(x, y) = -\frac{1}{N} \sum_{n=1}^{N} [xlog(y) + (1-y)log(1-x)]$

# Gradient Descent

➢ Minimize function J w.r.t. parameters θ

New weights

$$\theta^* := \theta - n \nabla J(y, x; \theta)$$

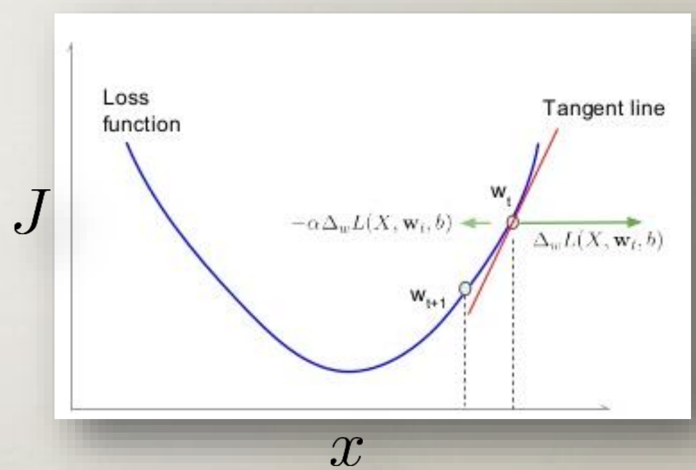
Gradient

Old weights

Learning rate

- Gradient

$$\nabla J(x) = \left(\frac{\partial J(x)}{\partial x_1}, \frac{\partial J(x)}{\partial x_2}, \ldots, \frac{\partial J(x)}{\partial x_n}\right)$$

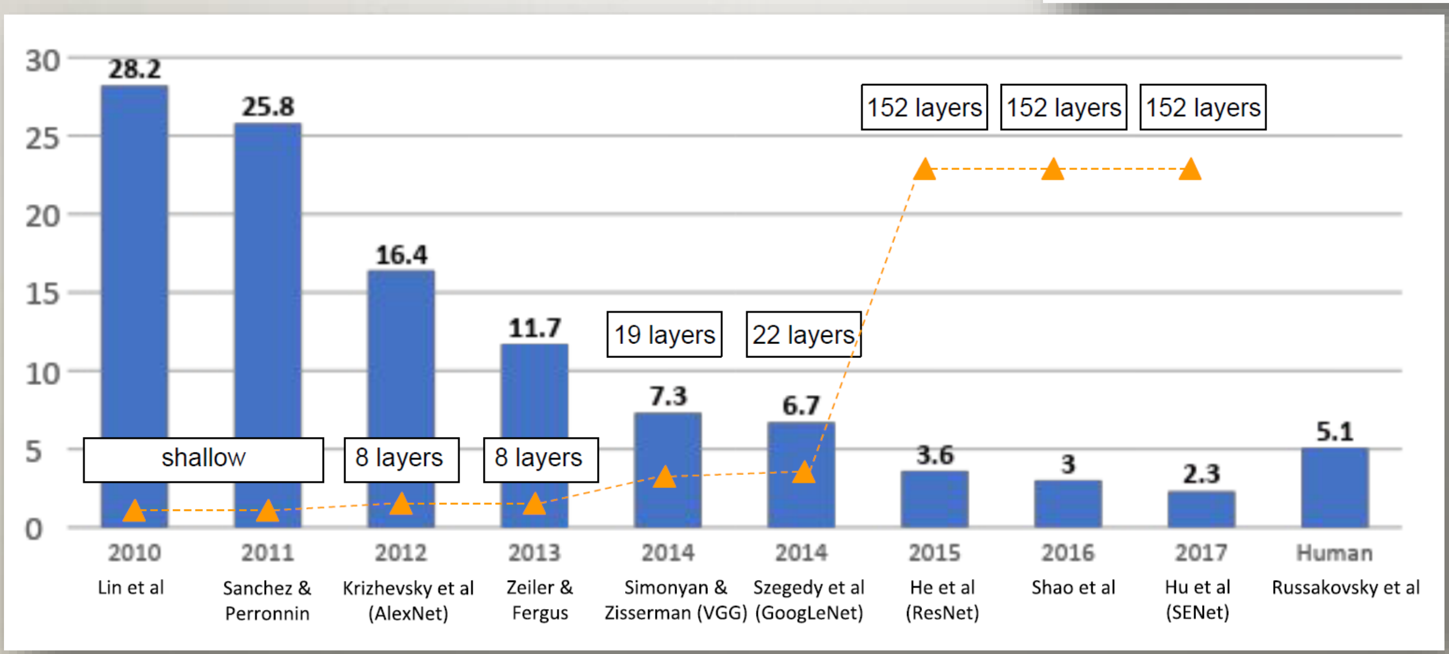



- Stochastic Gradient Descend

$$\theta^* := \theta - n \sum_{i=1}^{N} \nabla J(y_i, x_i; \theta)$$

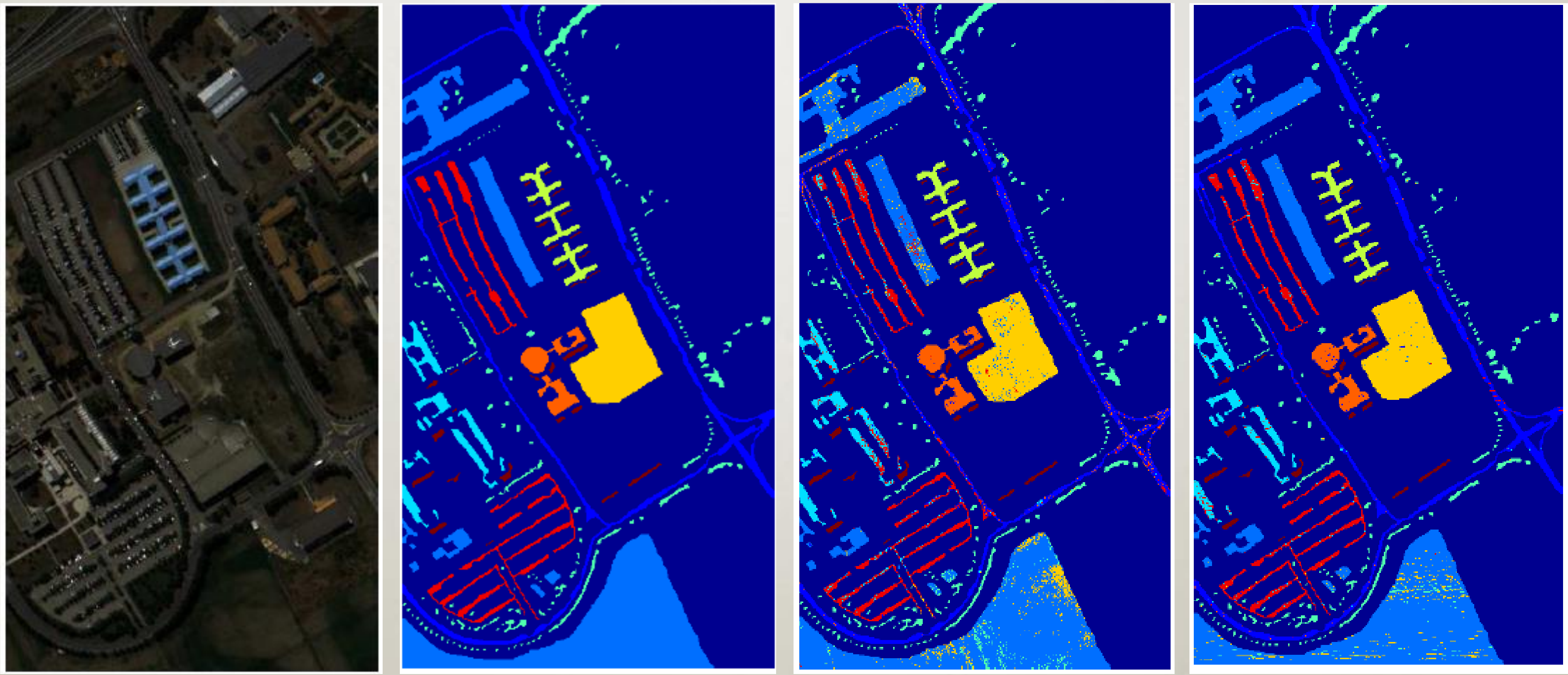
# Applications in imaging

- Generic image understanding

# Applications in imaging

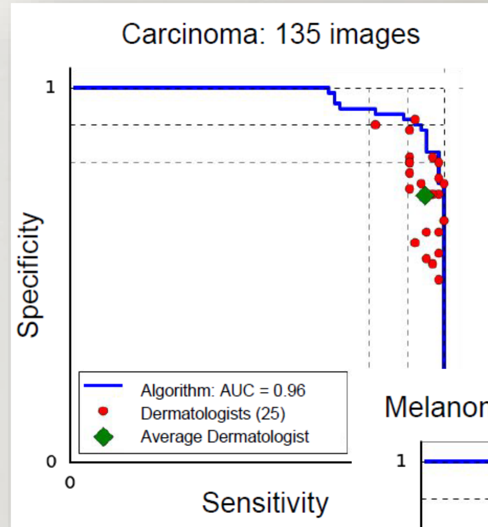
- Generic image understanding
- Remote Sensing

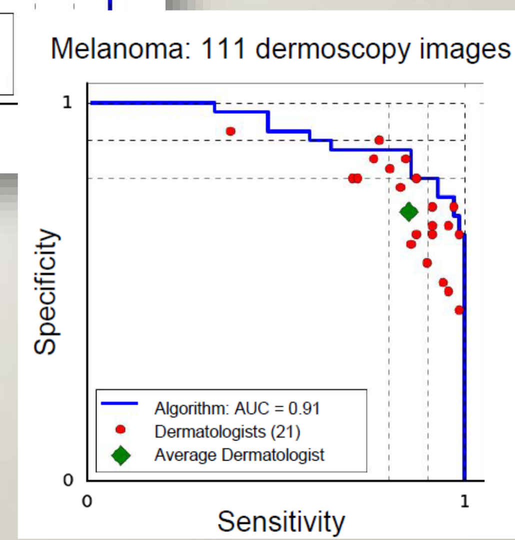


Optical | Ground Truth | SVM | CNN

# Applications in imaging

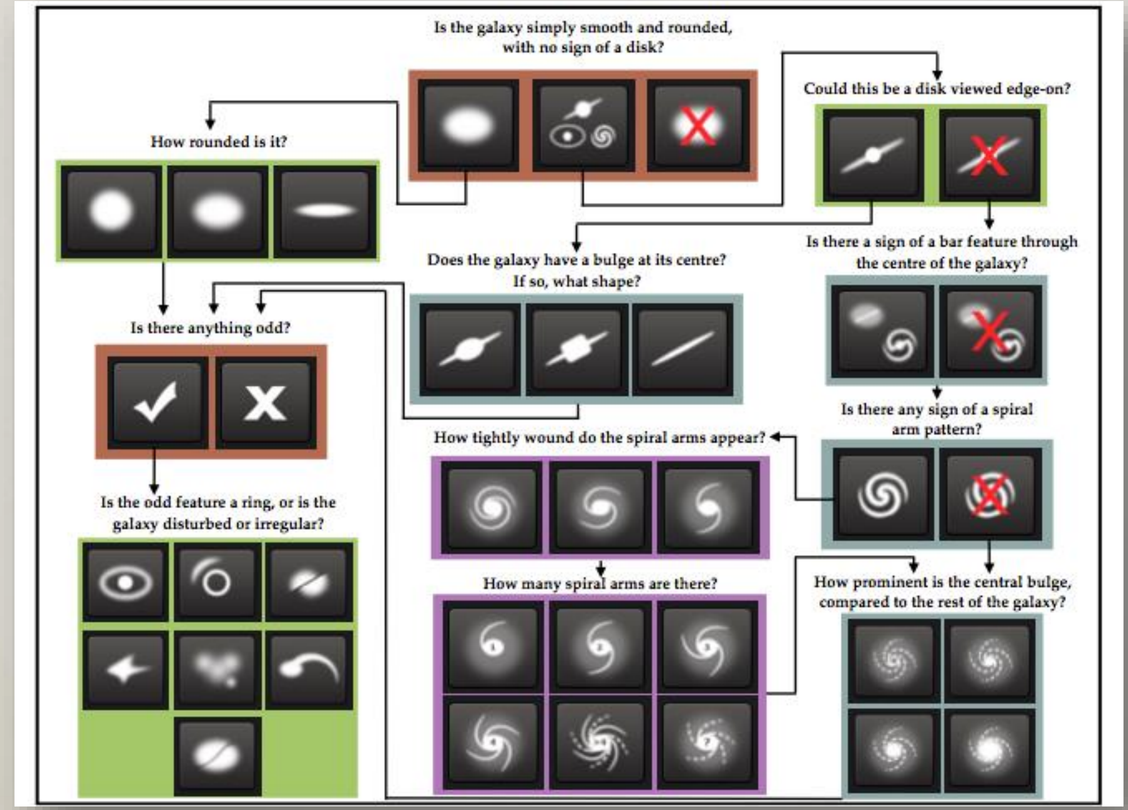
- Generic image understanding
- Remote Sensing
- Medical Imaging


Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115.

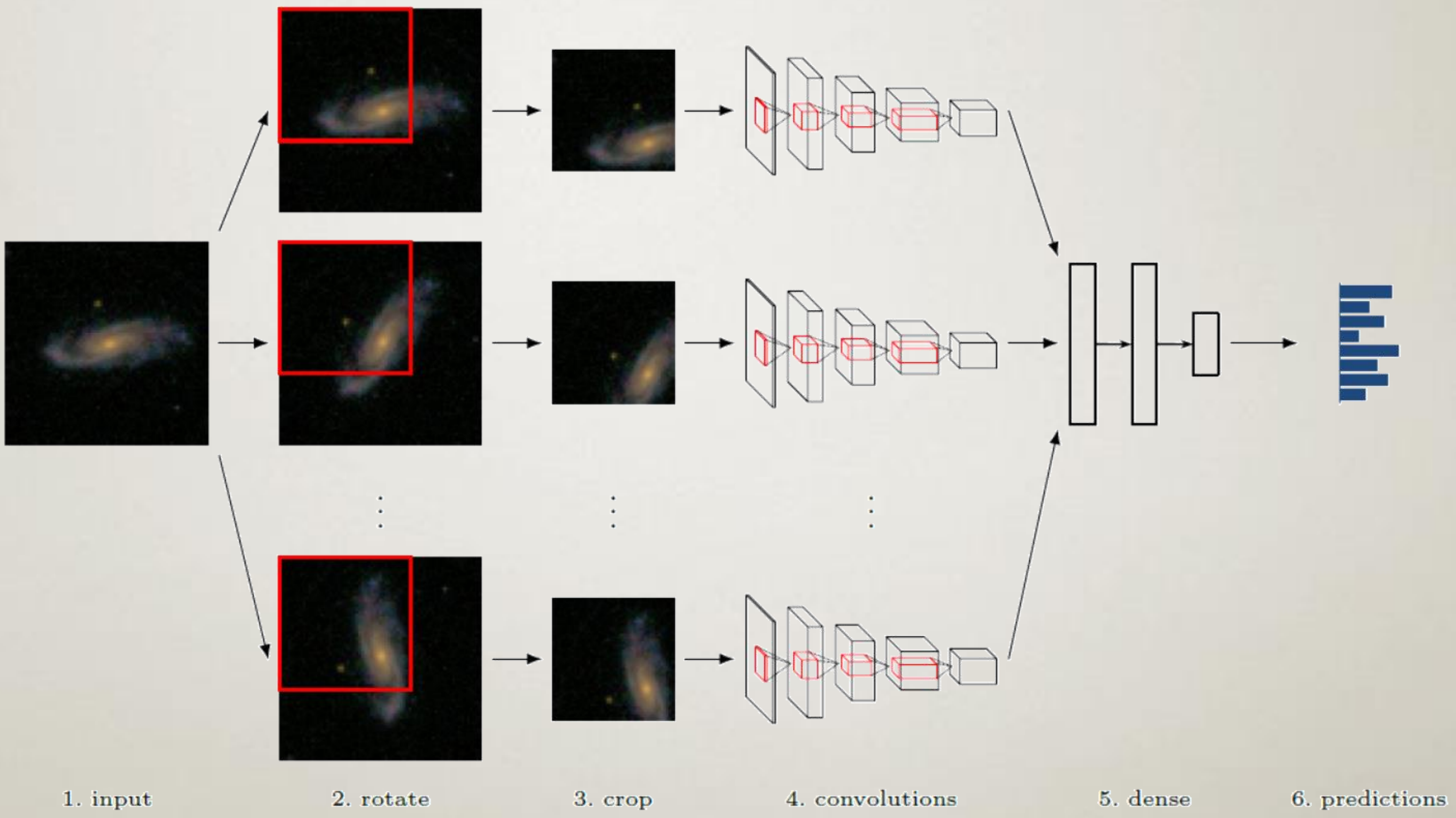
# The Galaxy zoo challenge




- Online crowdsourcing project
- users describe the morphology of galaxies based on color images
- 1 million galaxies imaged by the Sloan Digital Sky Survey (2007)

Darg, Daniel W., et al. "Galaxy Zoo: the fraction of merging galaxies in the SDSS and their morphologies." *Monthly Notices of the Royal Astronomical Society* 401.2 (2010)

# CNN in galaxy morphology

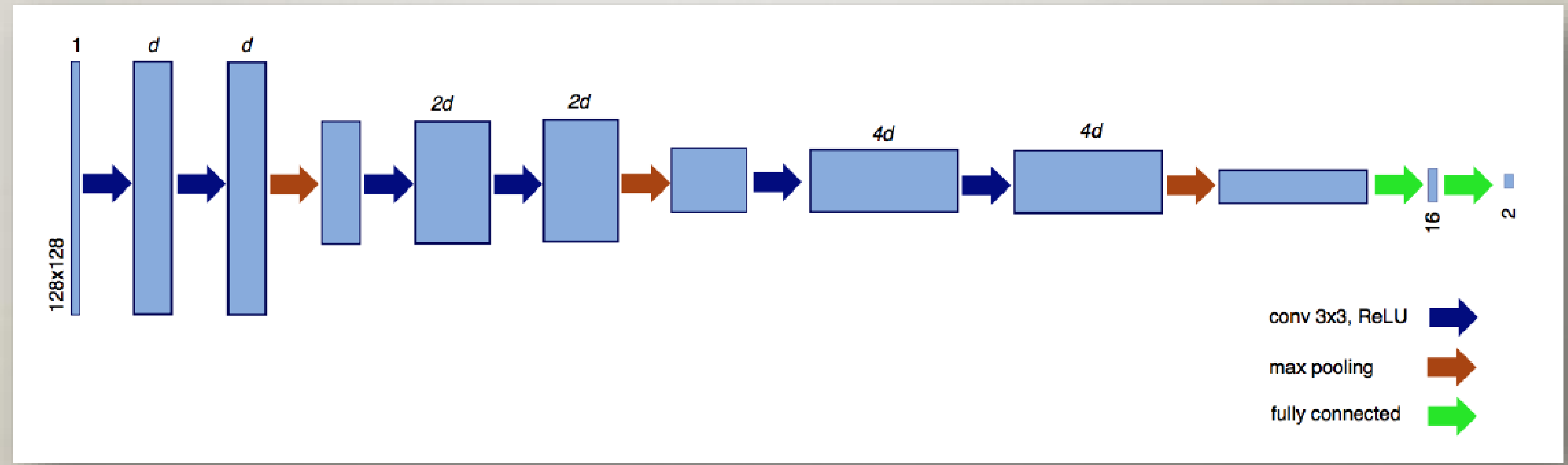



Dieleman, S., Kyle W. W., and Joni D.. "Rotation-invariant convolutional neural networks for galaxy morphology prediction." Monthly notices of the royal astronomical society, 2015

# DL for galaxy morphology

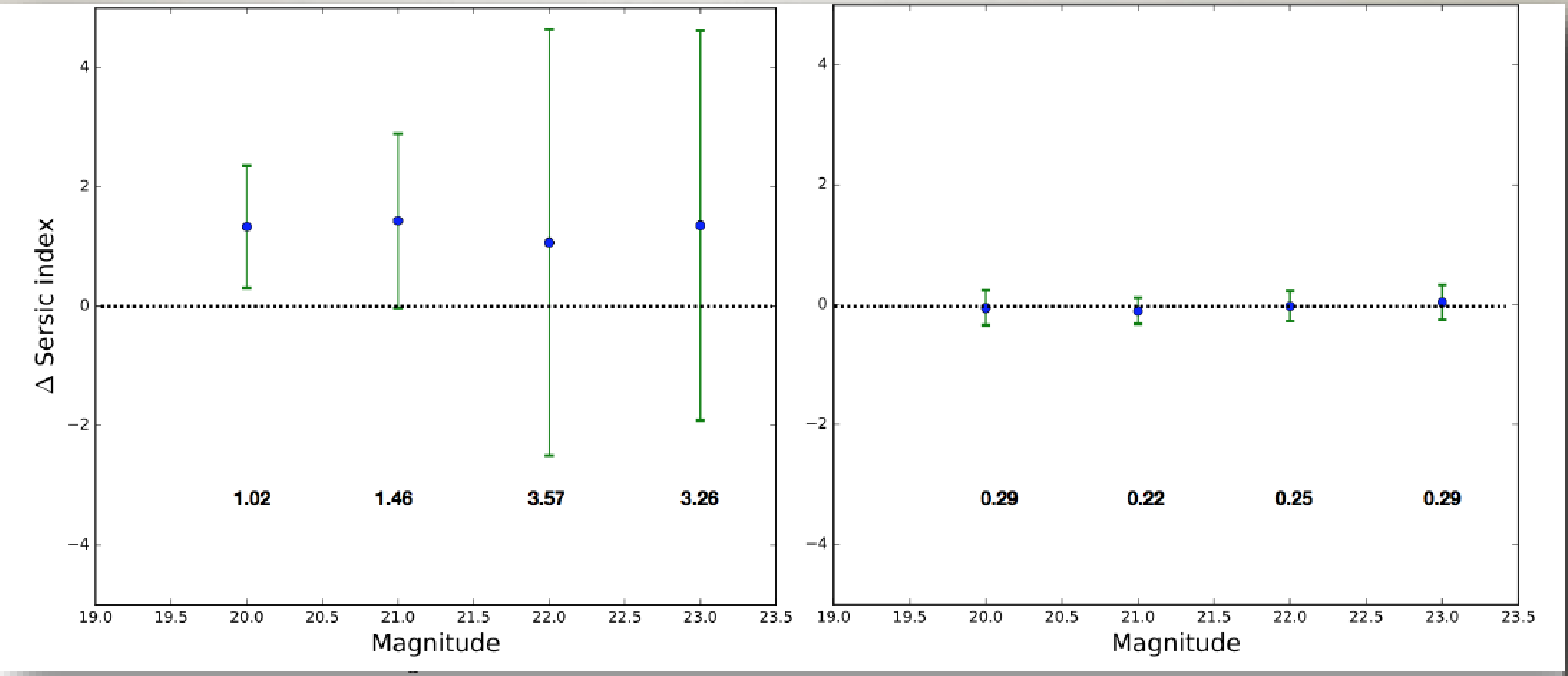
Recovery of galaxy parameters for HST images

Simulation of 31K galaxies (24K training), H band PSF, CANDELS survey noise




Tuccillo, D., Etienne Decencìère, and Santiago Velasco-Forero. "Deep learning for studies of galaxy morphology." *Proceedings of the International Astronomical Union* 12.S325 (2016): 191-196.

# DL for of galaxy morphology (con't)




GALFIT

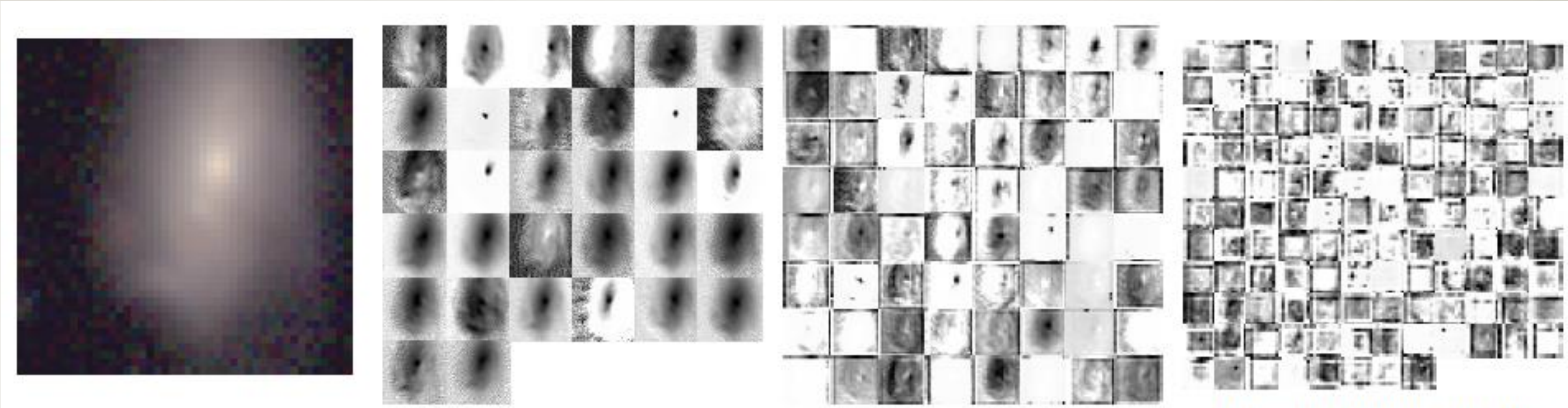
CNN

# CNN: Star-galaxy Classification

| type | filters | filter size | padding | non-linearity | initial weights | initial biases |
|---|---|---|---|---|---|---|
| convolutional | 32 | $5\times5$ | - | leaky ReLU | orthogonal | 0.1 |
| convolutional | 32 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| pooling | - | $2\times2$ | - | - | - | - |
| convolutional | 64 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| convolutional | 64 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| convolutional | 64 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| pooling | - | $2\times2$ | - | - | - | - |
| convolutional | 128 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| convolutional | 128 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| convolutional | 128 | $3\times3$ | 1 | leaky ReLU | orthogonal | 0.1 |
| pooling | - | $2\times2$ | - | - | - | - |
| fully-connected | 2048 | - | - | leaky ReLU | orthogonal | 0.01 |
| fully-connected | 2048 | - | - | leaky ReLU | orthogonal | 0.01 |
| fully-connected | 2 | - | - | softmax | orthogonal | 0.01 |

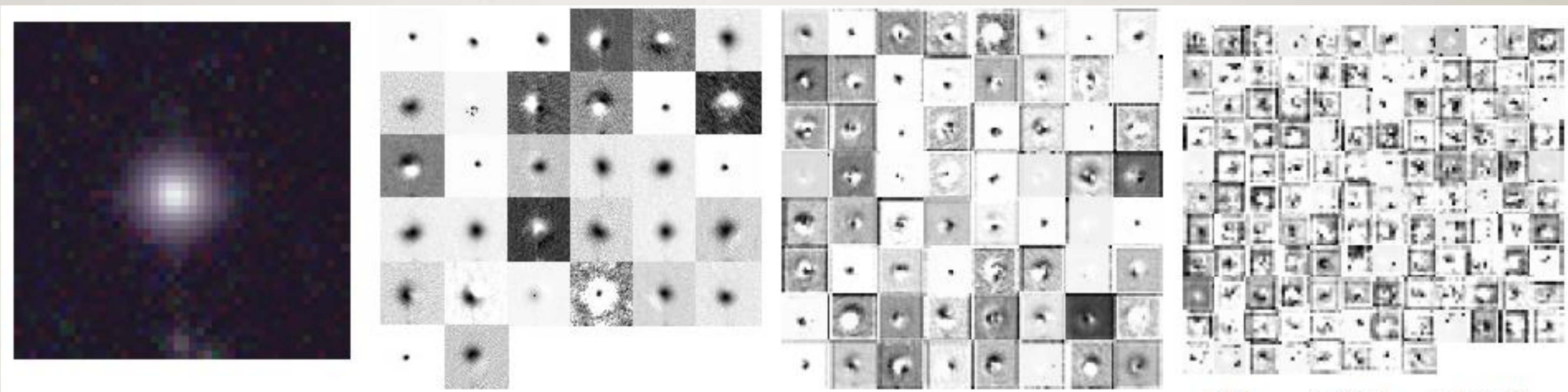
Kim, Edward J., and Robert J. Brunner. "Star-galaxy classification using deep convolutional neural networks." *Monthly Notices of the Royal Astronomical Society* (2016): stw2672.

Galaxy



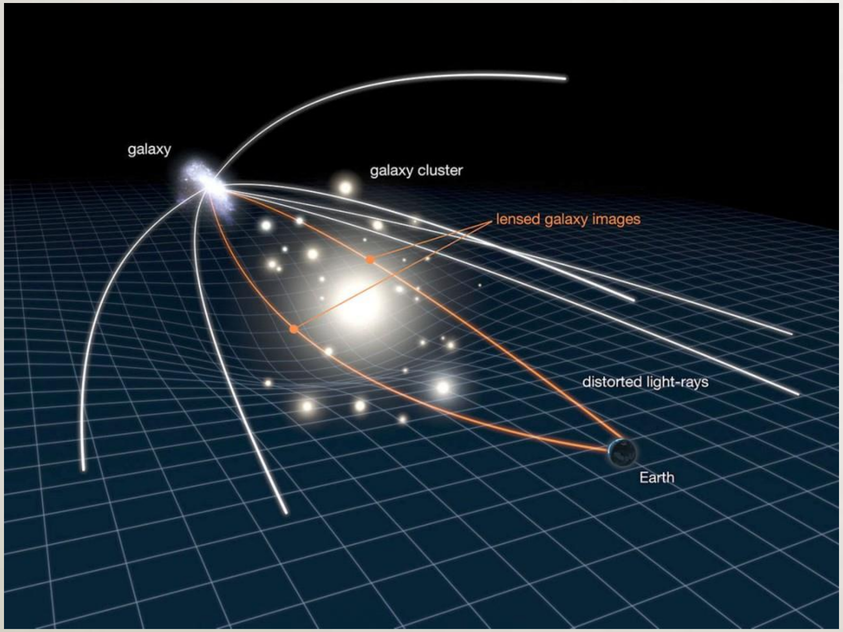
(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

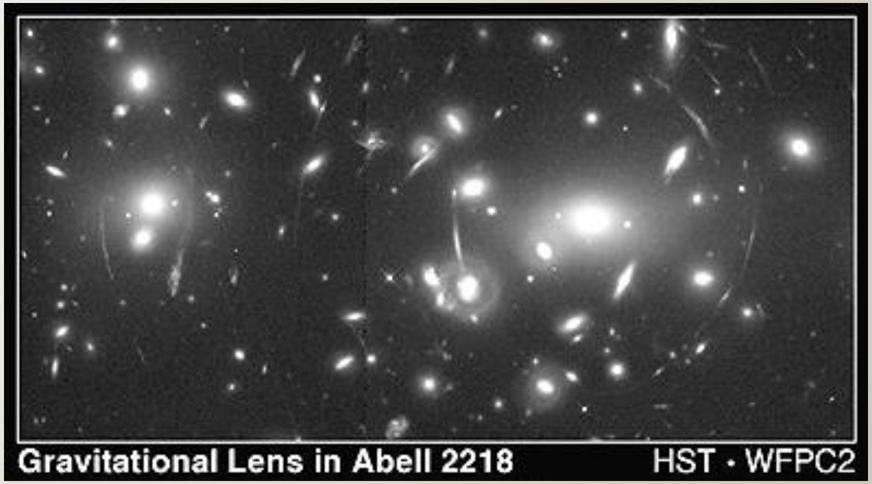
Star



(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)

# Gravitational Lensing

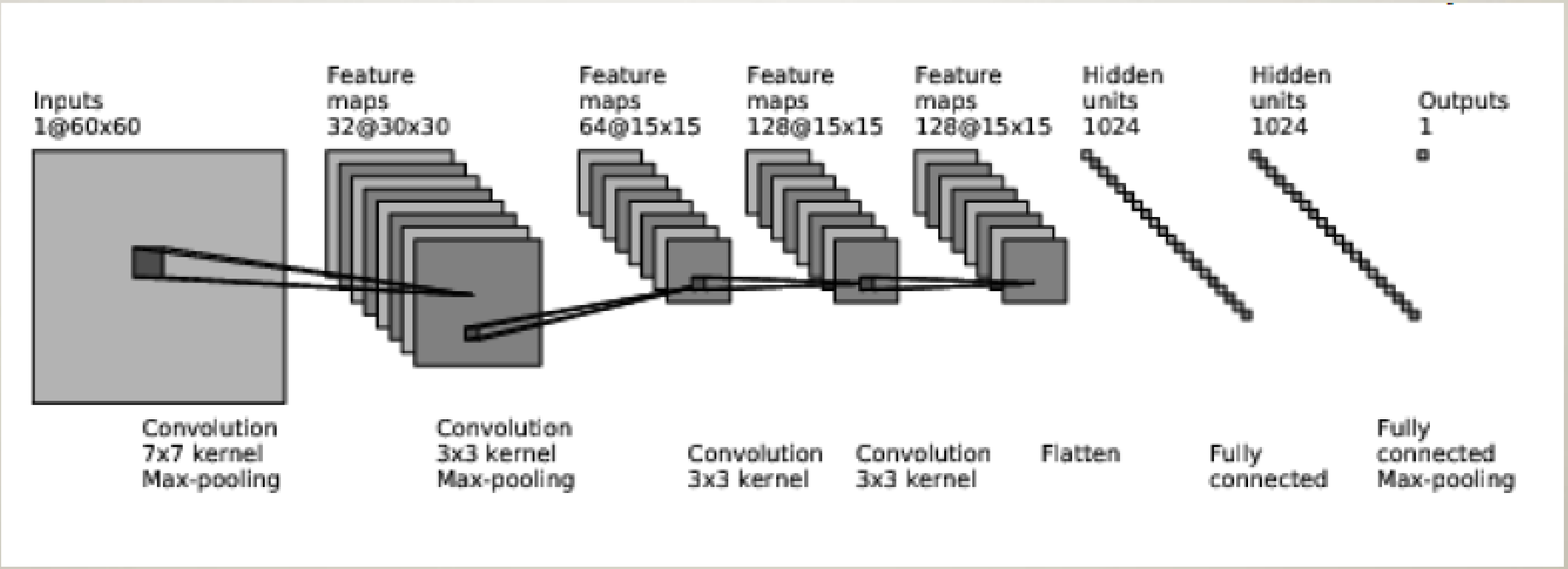


# CNN for lensing

- CNNs in Kilo Degree Survey

21789 colour-magnitude selected Luminous Red Galaxies, of which 3 are known lenses, the CNN retrieves 761 strong-lens candidates and correctly classifies 2/3 of known lenses.




Petrillo, C. E., C. Tortora, S. Chatterjee, G. Vernardos, et al. "Finding strong gravitational lenses in the Kilo Degree Survey with convolutional neural networks." Monthly Notices of the Royal Astronomical Society 472, no. 1 (2017)

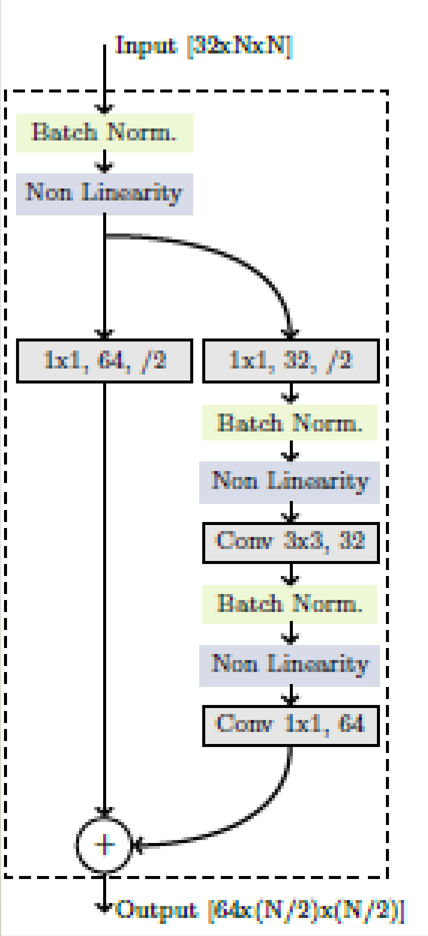
# DeepLens

- Training

  20,000 LSST-like observations

- Testing

  for a rejection rate of non-lenses of 99%, a completeness of 90% can be achieved for lenses with Einstein radii larger than 1.400 and S/N larger than 20 on individual g-band LSST exposures.




Lanusse, Francois, et al. "CMU DeepLens: Deep Learning For Automatic Image-based Galaxy-Galaxy Strong Lens Finding." arXiv preprint arXiv:1703.02642 (2017).

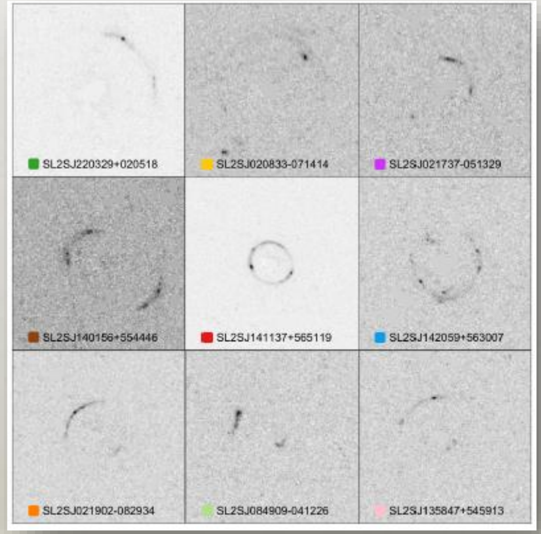
# Detecting strong lensing

- Strong galaxy-galaxy lensing systems
- CA-FR-HA Telescope Legacy Survey (CFHTLS)
- Ensemble of trained DL networks
- Search of 1.4 million early type galaxies selected from the survey catalog as potential deflectors,
- Identified 2,465 candidates (117 previously known lens candidates, 29 confirmed lenses, 266 novel probable or potential lenses and 2097 false positives.

Jacobs, Colin, et al. "Finding strong lenses in CFHTLS using convolutional neural networks." Monthly Notices of the Royal Astronomical Society 471.1 (2017)

# Fast Strong Gravitational Lenses analysis

- Typical ML approaches: single lens → few weeks & experts
- Estimation of lensing parameters via CNN
  - Singular Isothermal Ellipsoid density profile
  - Parameters: Einstein radius, complex ellipticity, the coordinates of the lens center
- Lens removal through ICA



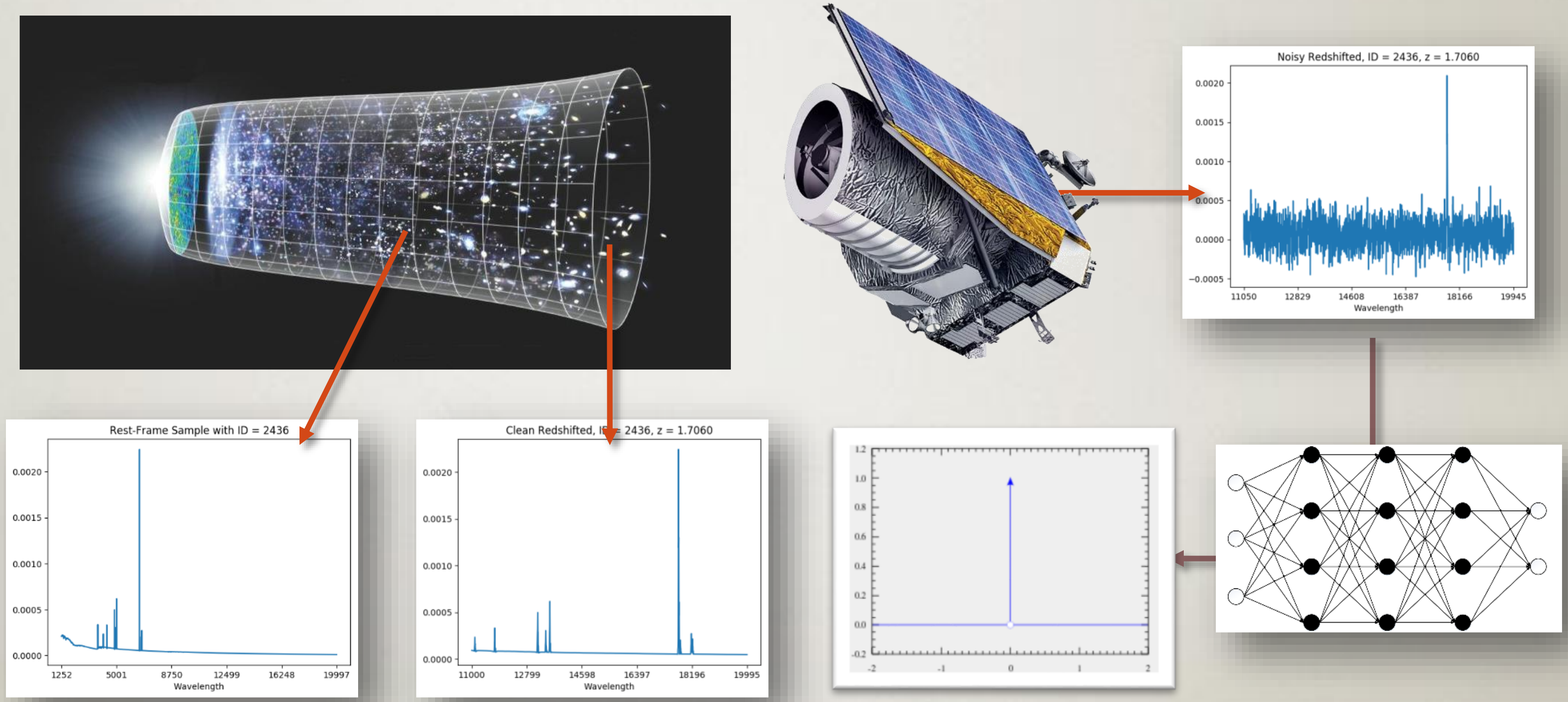

Hezaveh, Yashar D., Laurence Perreault Levasseur, and Philip J. Marshall. "Fast automated analysis of strong gravitational lenses with convolutional neural networks." *Nature* 548.7669 (2017)

# Other applications

- Classifying Radio Galaxies With Convolutional Neural Network[1]
- Deep-HITS: Rotation Invariant Convolutional Neural Network For Transient Detection[2]
- Galaxy surface brightness profile[3]

1. Aniyan AK, Thorat K. Classifying Radio Galaxies with the Convolutional Neural Network. The Astrophysical Journal Supplement Series. 2017 Jun 13
2. Cabrera-Vives G, Reyes I, Förster F, Estévez PA, Maureira JC. Deep-HiTS: Rotation invariant convolutional neural network for transient detection. The Astrophysical Journal. 2017 Feb 10
3. Tuccillo D, Huertas-Company M, Decencière E, Velasco-Forero S, Domínguez Sánchez H, Dimauro P. Deep learning for galaxy surface brightness profile fitting. Monthly Notices of the Royal Astronomical Society. 2017 Dec 11

# Spectroscopic red-shift estimation w/ CNNs




R. Stivaktakis, G. Tsagkatakis, B. Moraes, F. Abdalla, J.-L. Starck, P. Tsakalides, "Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data," EEE Transactions on Big Data: Special Issue on Big Data from Space, 2018

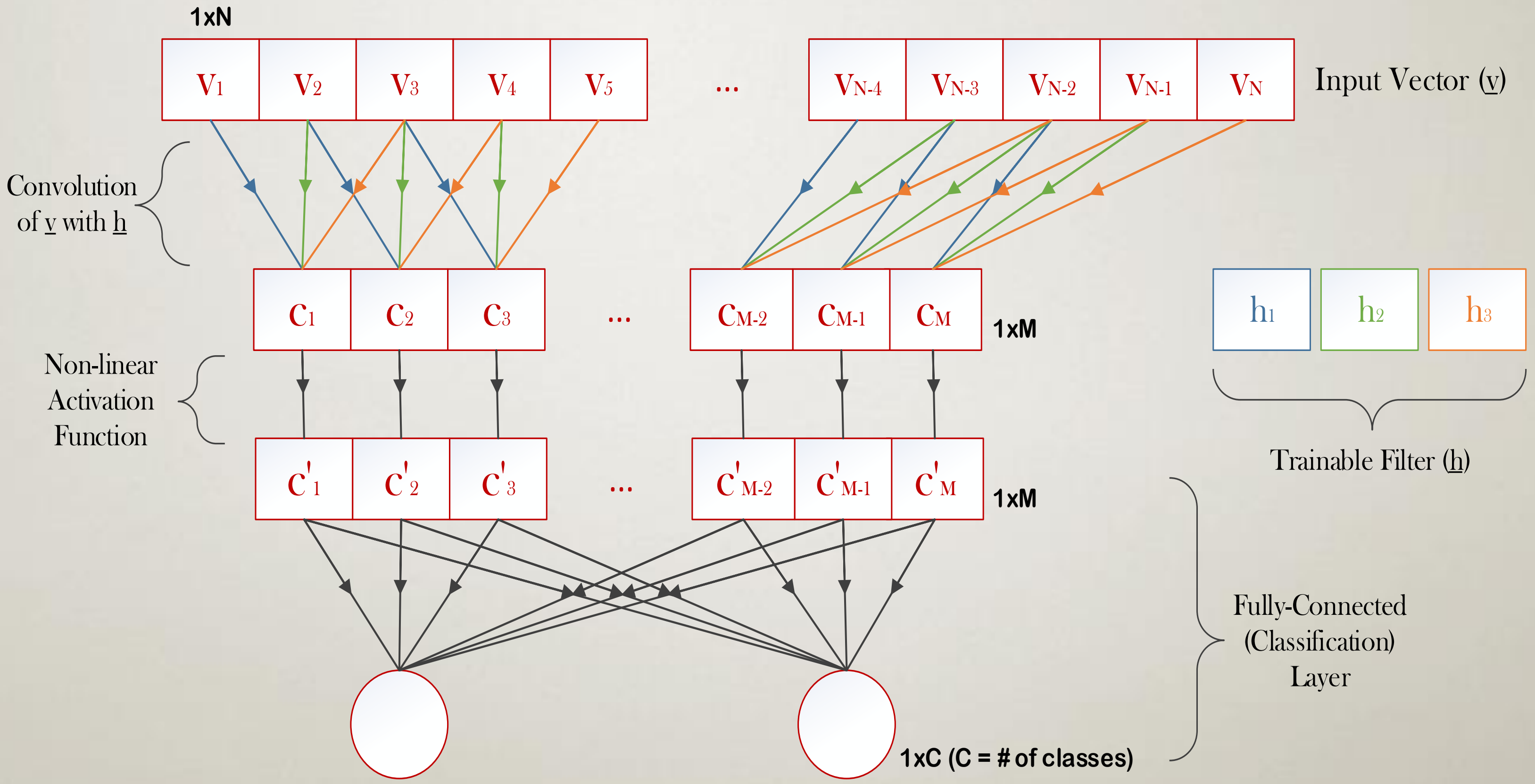
FORTH INSTITUTE OF COMPUTER SCIENCE

UCL

# Methodology

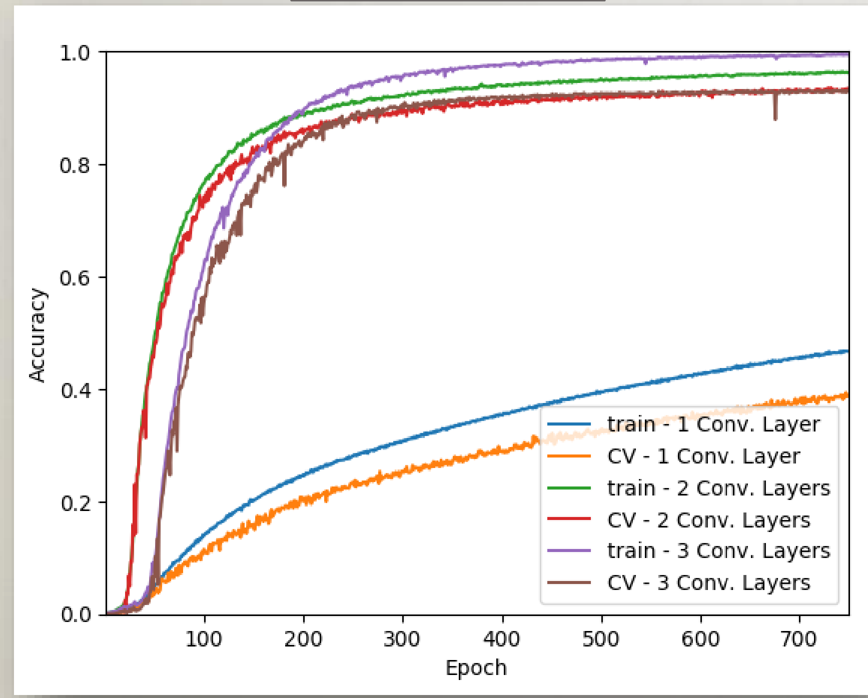
- 13K clean, rest-frame spectral profiles (SEDs)
- Production of randomly redshifted examples from original dataset
  - $z = [1, 1.8)$, similar to Euclid's range
  - $\log(1+z) = \log(\lambda_{observed}) - \log(\lambda_{emit}) \leftrightarrow 1+z = \frac{\lambda_{observed}}{\lambda_{emit}}$
  - Redshift range is split into 800 discrete classes ($\Delta z$=0.001)
  - Addition of white Gaussian noise
- Evaluation on idealistic & realistic observations
  - Impact of the depth of the network
  - Data-driven analysis
  - Address to overfitting conditions
  - Comparison with other popular classifiers

# CNN architecture

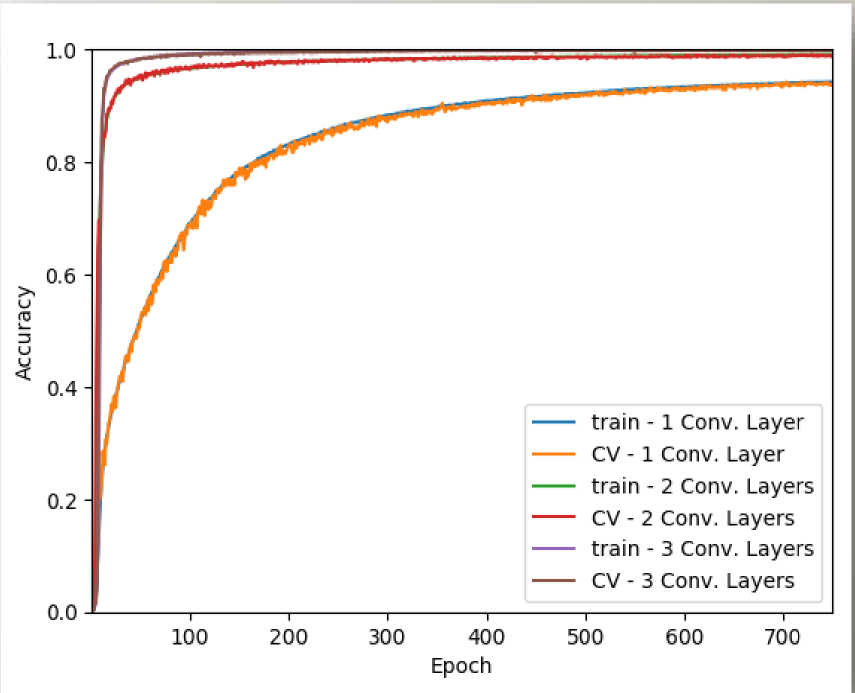


# Impact of data availability

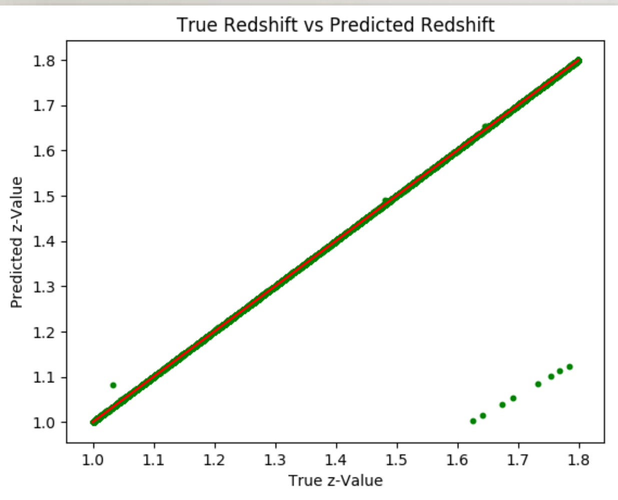
**40K Training**



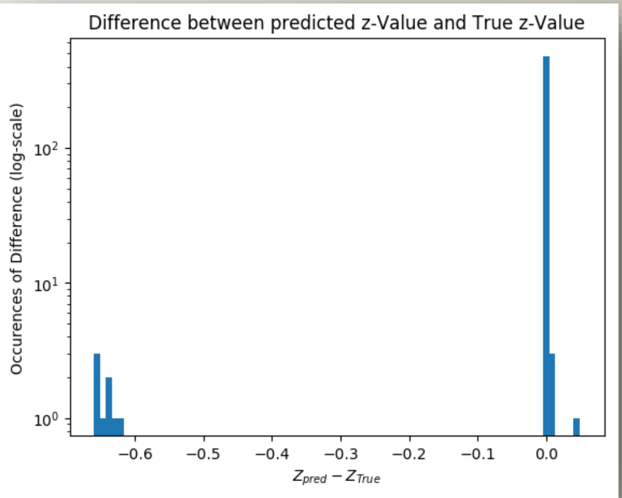

**400K Training**

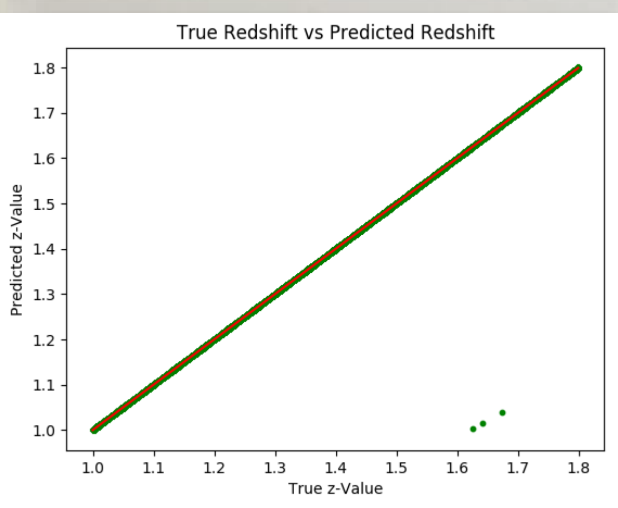


# Impact of architecture

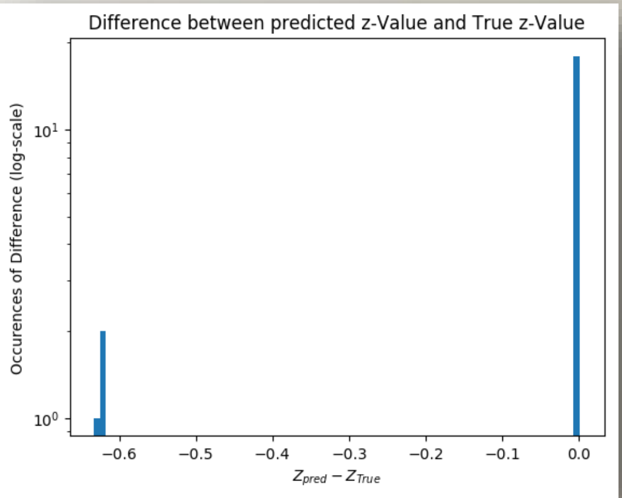



**<u>400,000 Tr. Examples - 1 Conv. Layer</u>**

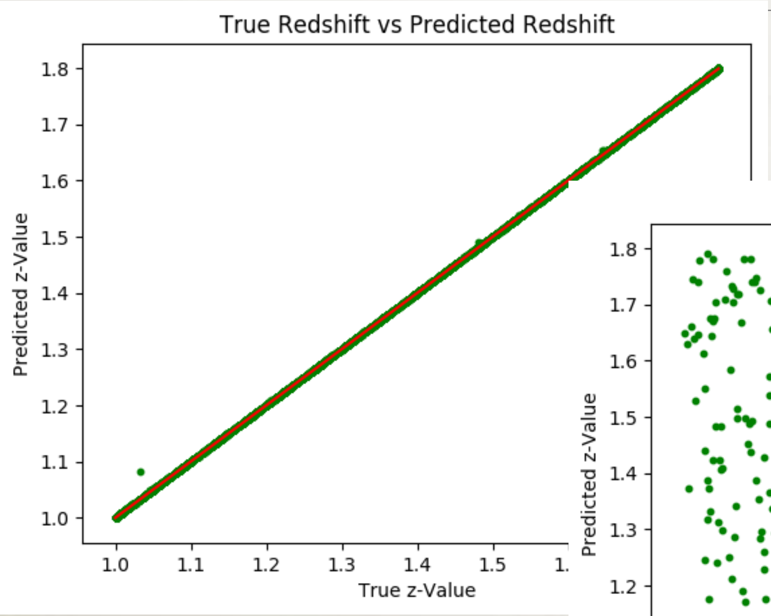


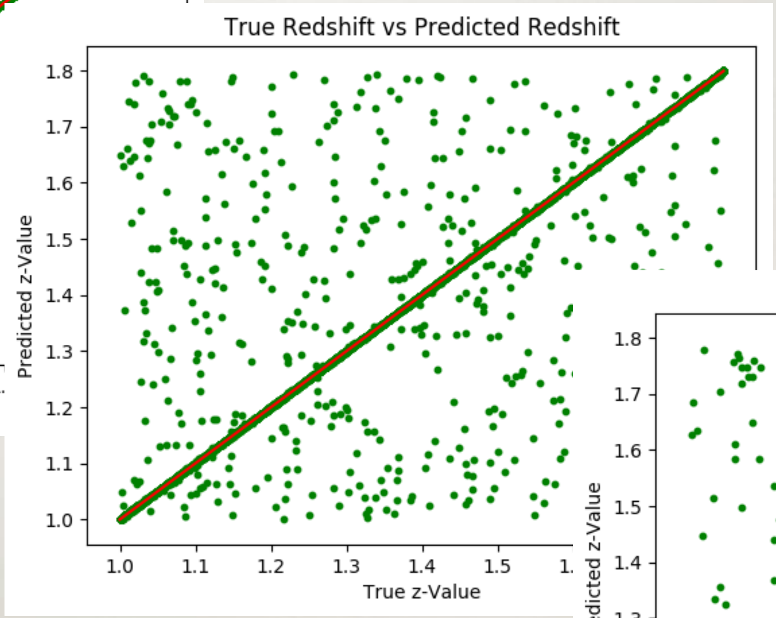

**<u>400,000 Tr. Examples - 3 Conv. Layers</u>**



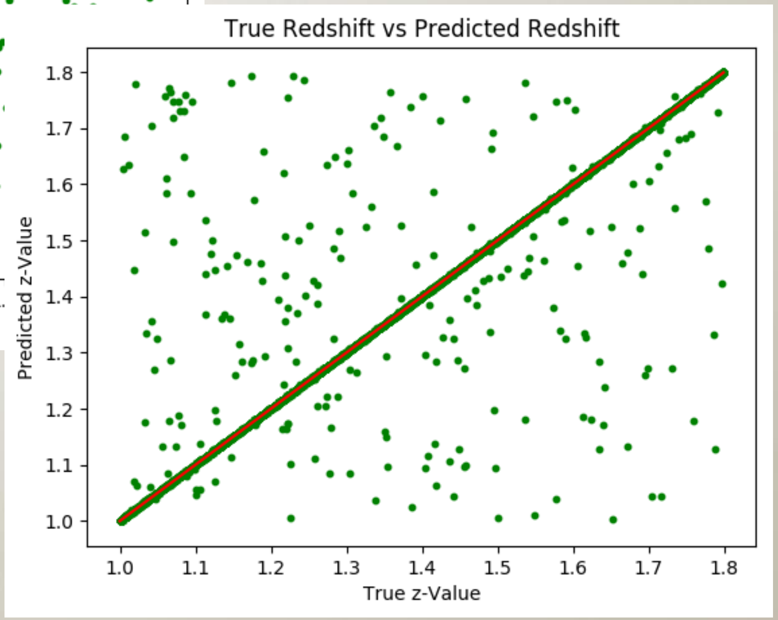
# Impact of noise




Clean - 400K

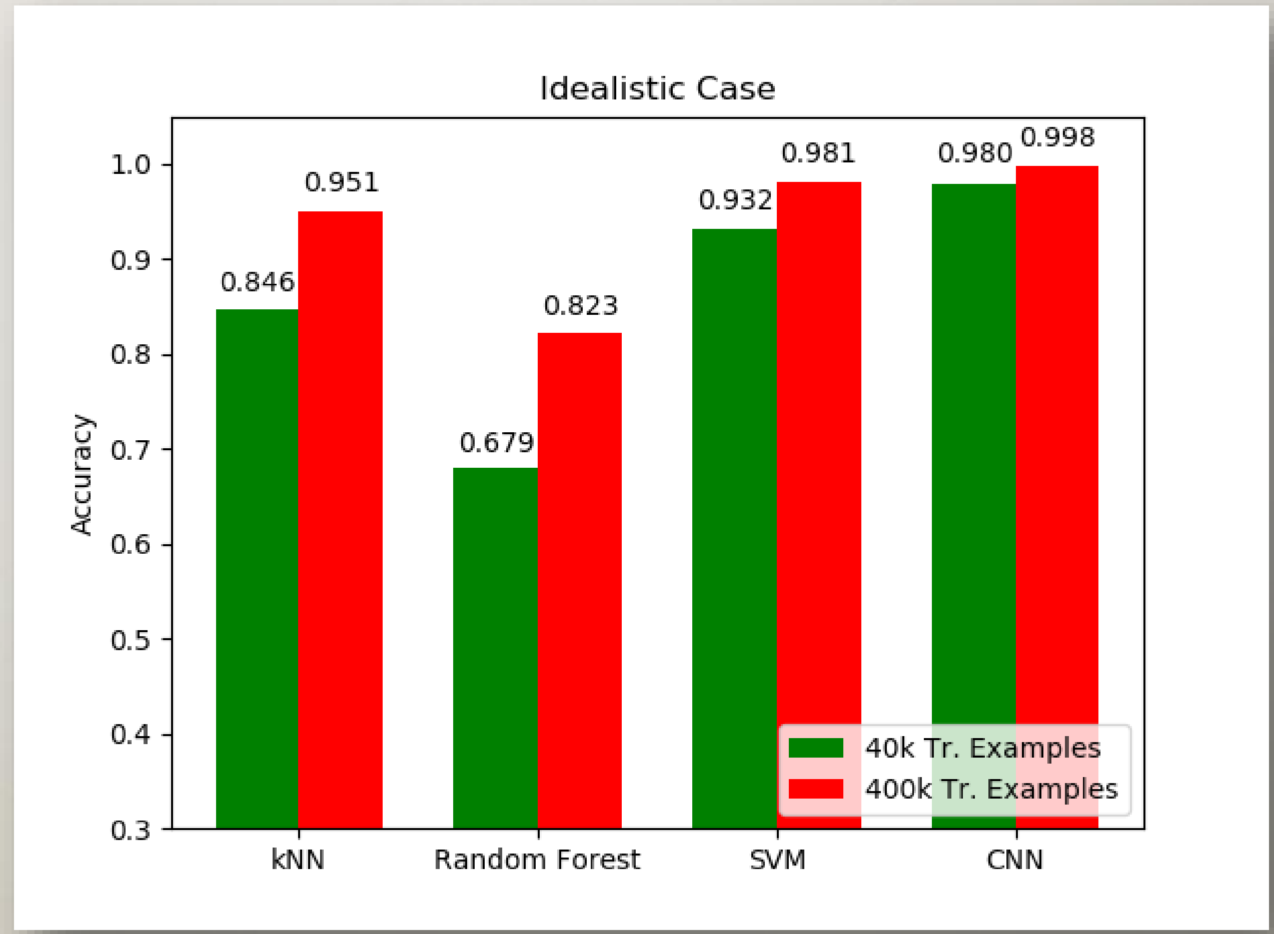



Noisy – 400K

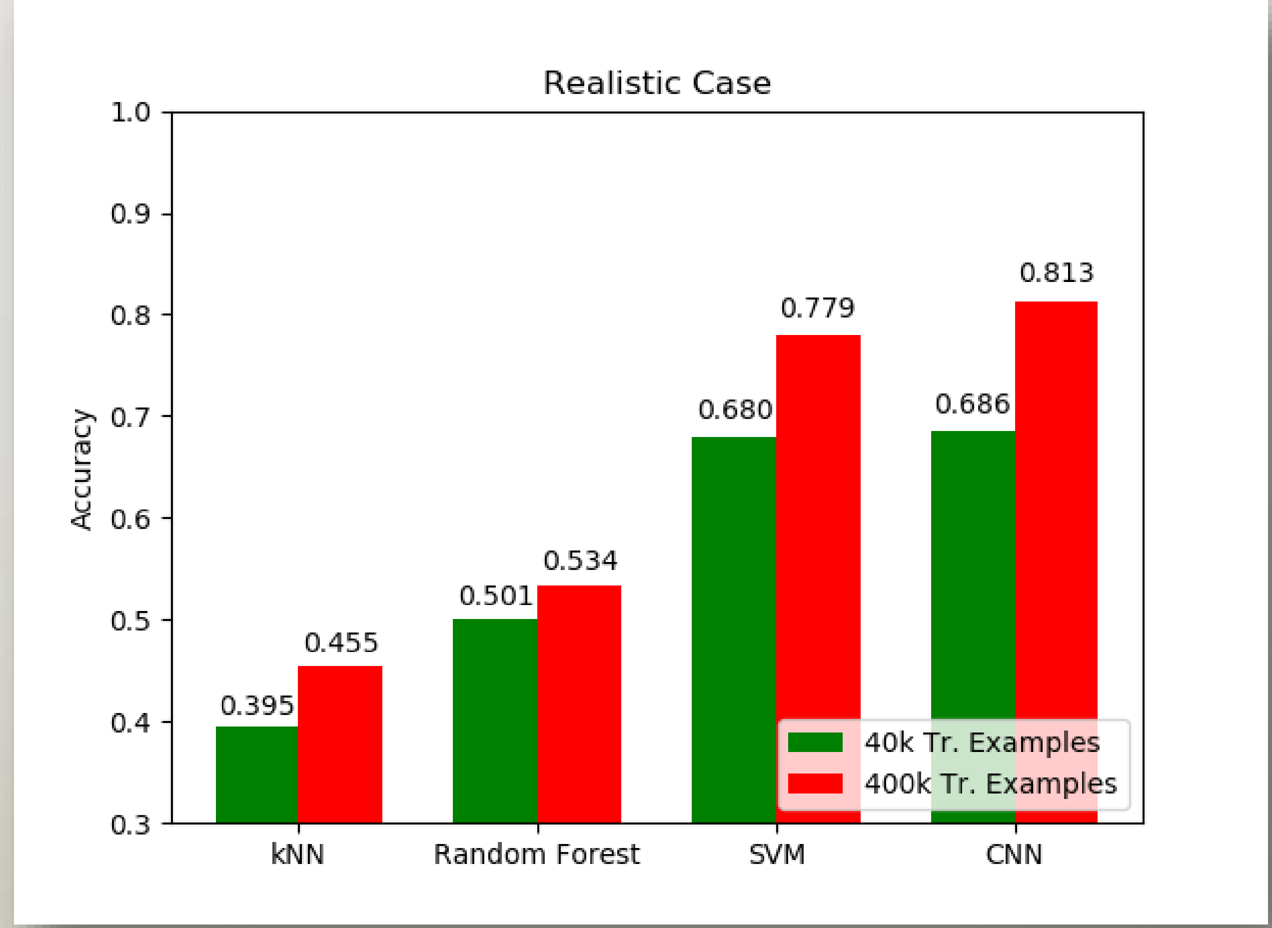


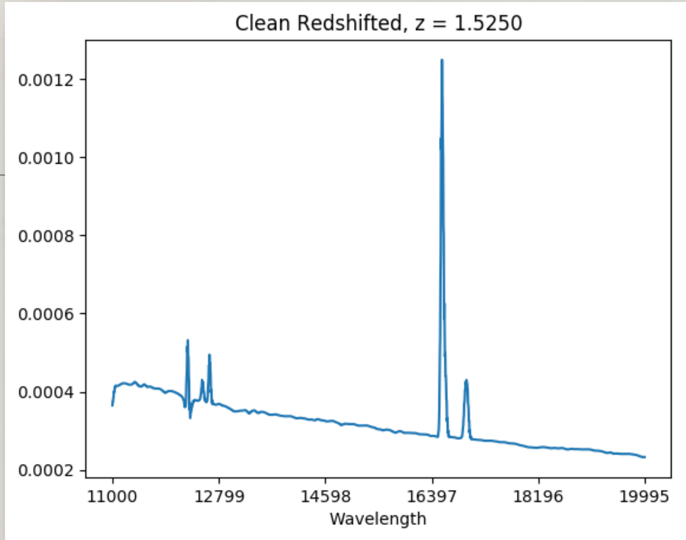

Noisy – 4M

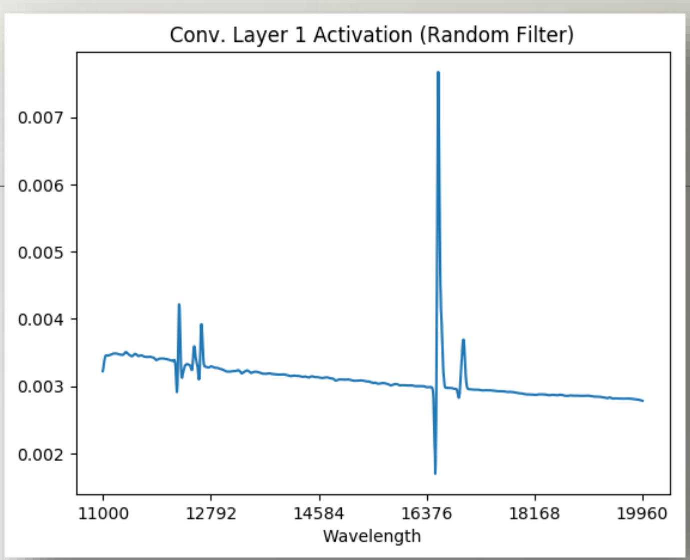
# Comparison with SotA (clean)

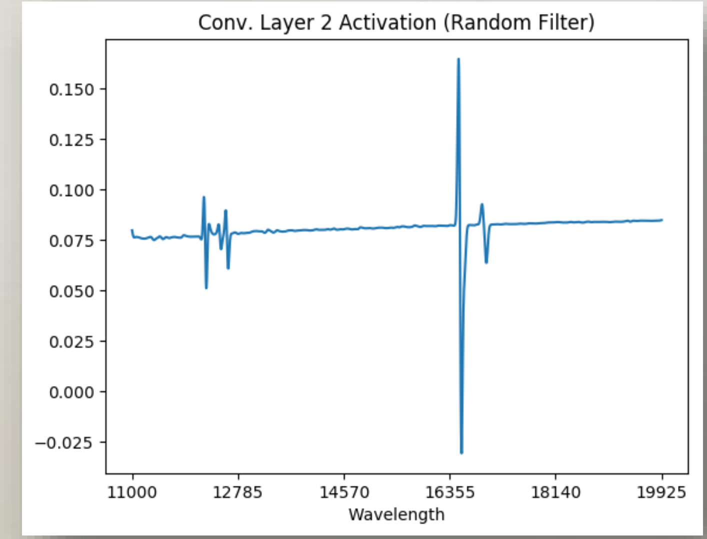


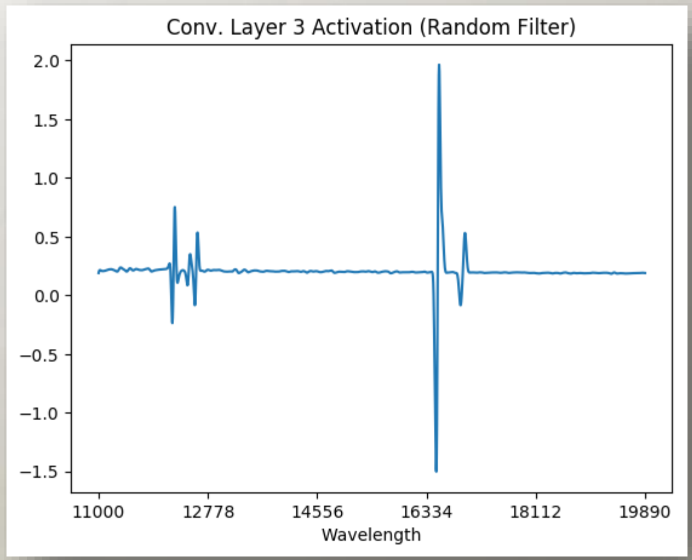
# Comparison with SotA (noisy)




Clean Redshifted, z = 1.5250
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
11000
12799
14598
16397
18196
19995
Wavelength


Conv. Layer 1 Activation (Random Filter)
0.007
0.006
0.005
0.004
0.003
0.002
11000
12792
14584
16376
18168
19960
Wavelength


Conv. Layer 2 Activation (Random Filter)
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
11000
12785
14570
16355
18140
19925
Wavelength


Conv. Layer 3 Activation (Random Filter)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
11000
12778
14556
16334
18112
19890
Wavelength

# A Brave New World…

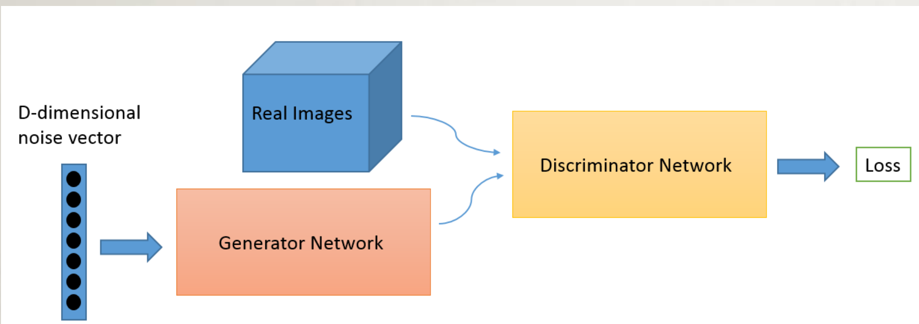
## Yann LeCun's cake

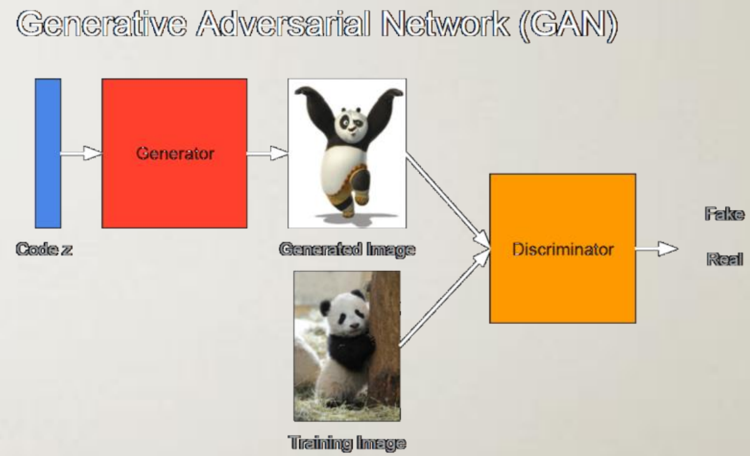
- **"Pure" Reinforcement Learning** (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning** (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Unsupervised/Predictive Learning** (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

# Generative Adversarial Networks

- Generative model produces realistic new samples
- Discriminative model differentiate real vs synthetic samples




$$\min_{G} \max_{D} = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(z)}[1 - \log D(G(\mathbf{x}))]$$

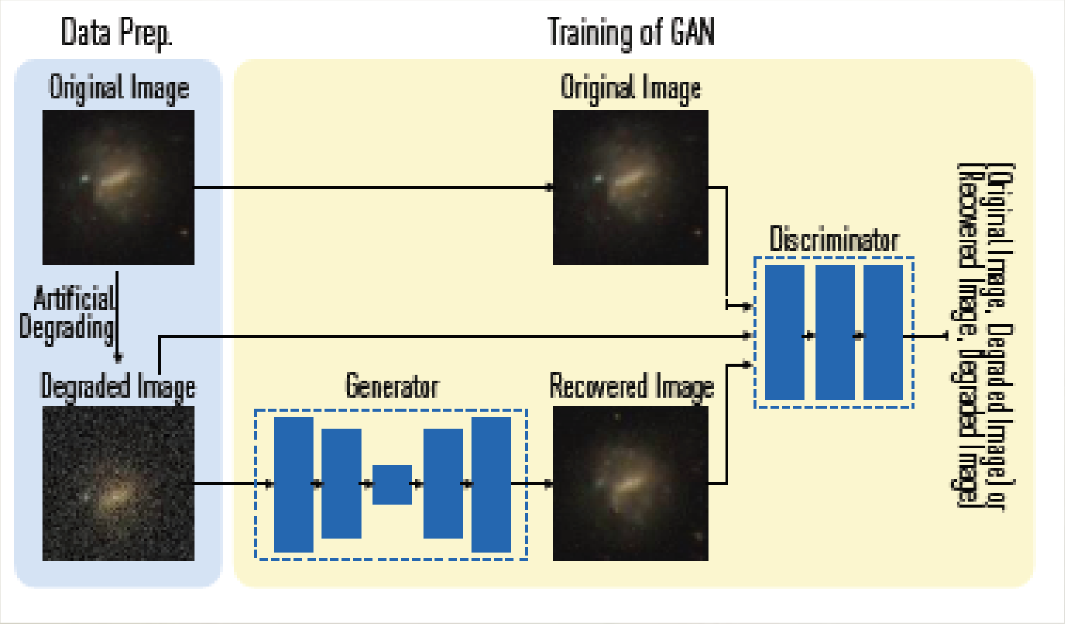
Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. InAdvances in neural information processing systems 2014

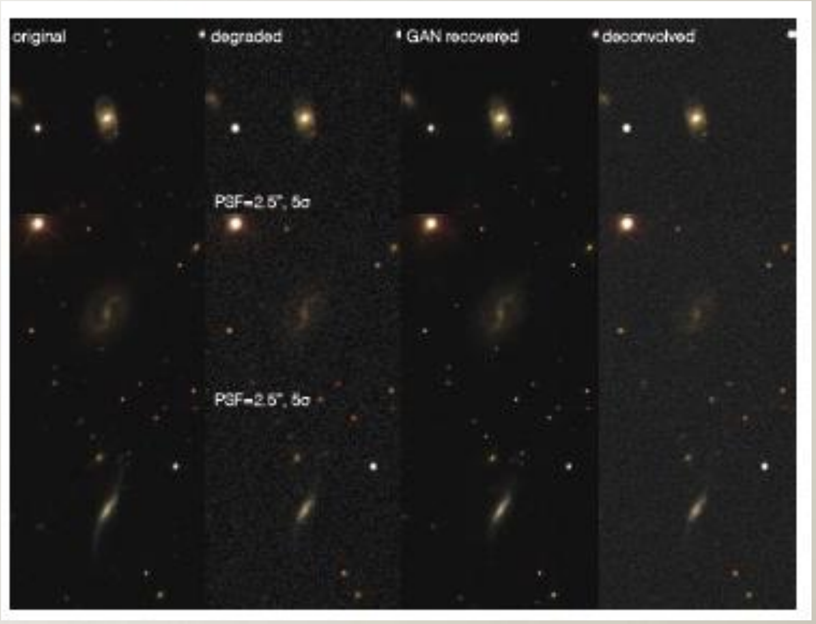
# Dreaming of bedrooms…

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

# GANs for deconvolution

4,550 SDSS images of nearby galaxies at 0:01 < z < 0:02




Schawinski, Kevin, et al. "Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit." *arXiv preprint arXiv:1702.00403* (2017).

# Discussion

- Deep Learning "always" achieves state-of-art performance
- Big Data + Deep Learning = 🥰
- Generative Adversarial Networks → Data Driven Science

<u>Open issues</u>

- Meta-learning / one-shot learning
- Time Domain Astronomy

# Thank you

spl.edu.gr
dedale.cosmostat.org

GREGORY TSAGKATAKIS

SIGNAL PROCESSING LAB

ICS - FORTH