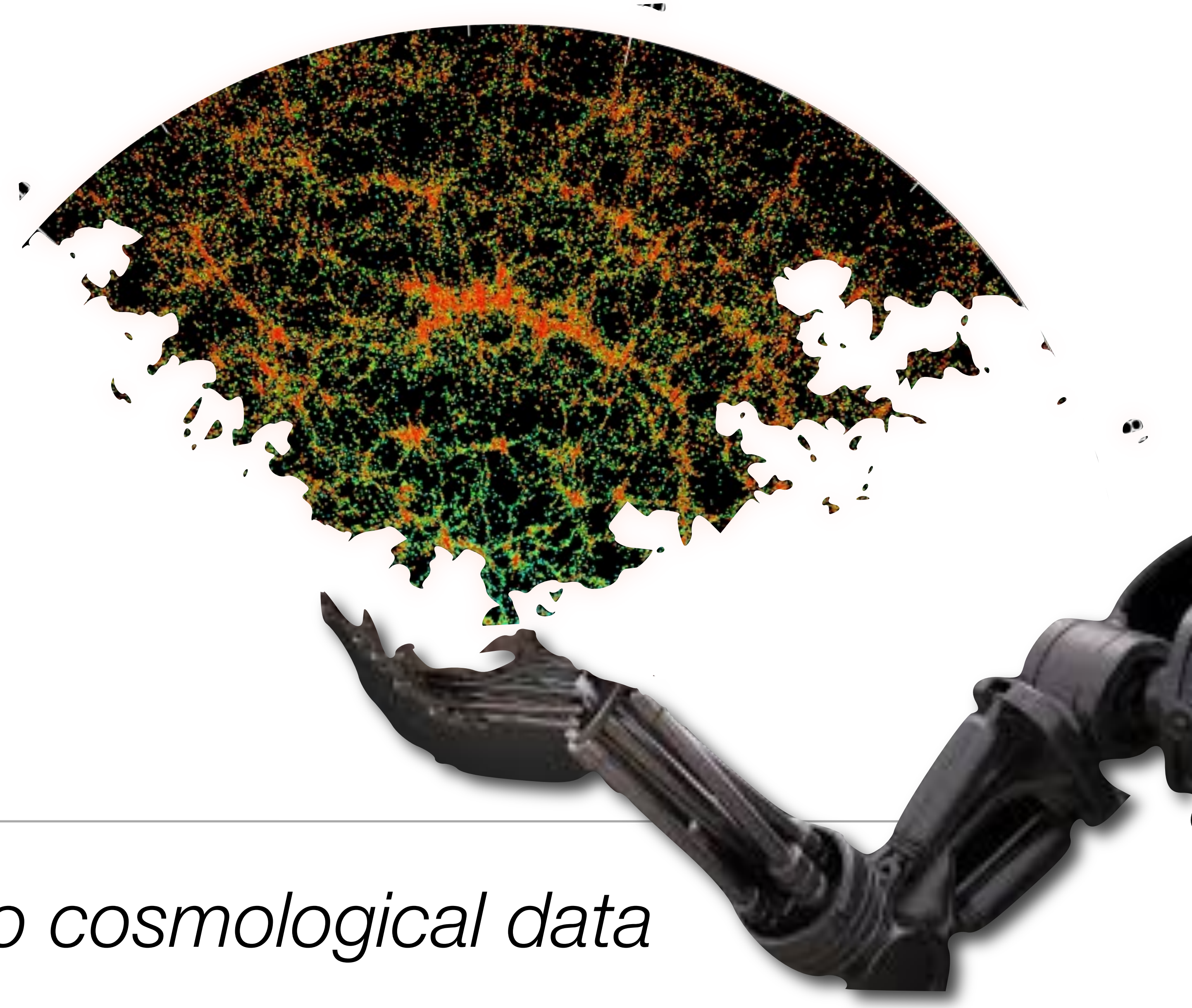


AstroEncoder

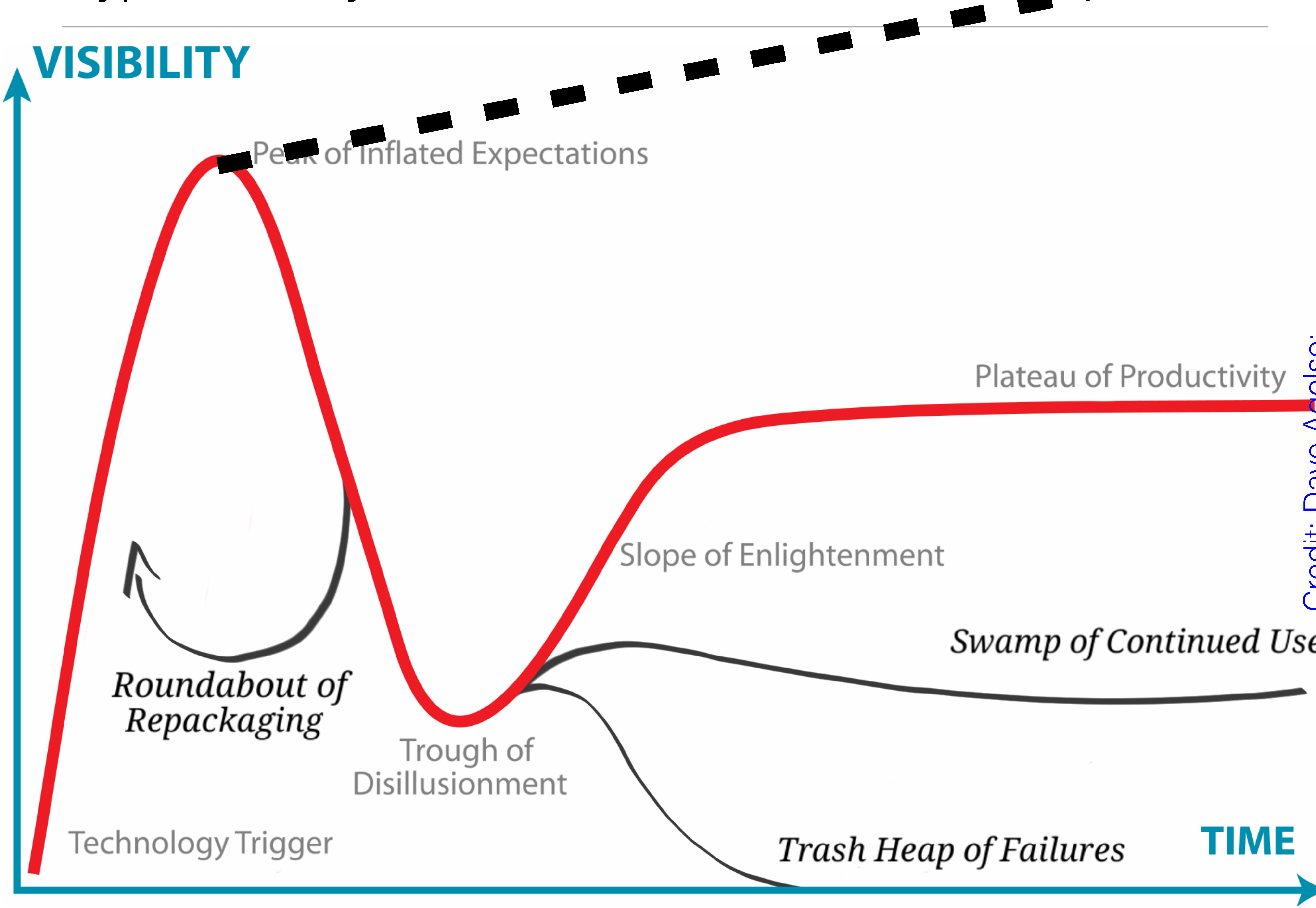
Applications of deep learning to cosmological data

Brian Nord (*@iamstarnord*)
Fermilab + DES
News Perspectives
05 June 2017

Preview

- Deep Learning Methodologies
- Applications:
  - Strong Lensing
  - Stellar Spectrum Modeling
- Outlook: Challenges and Potential

Hype/Visibility Curve



Credit: Dave Agolso:

Rise of the machines

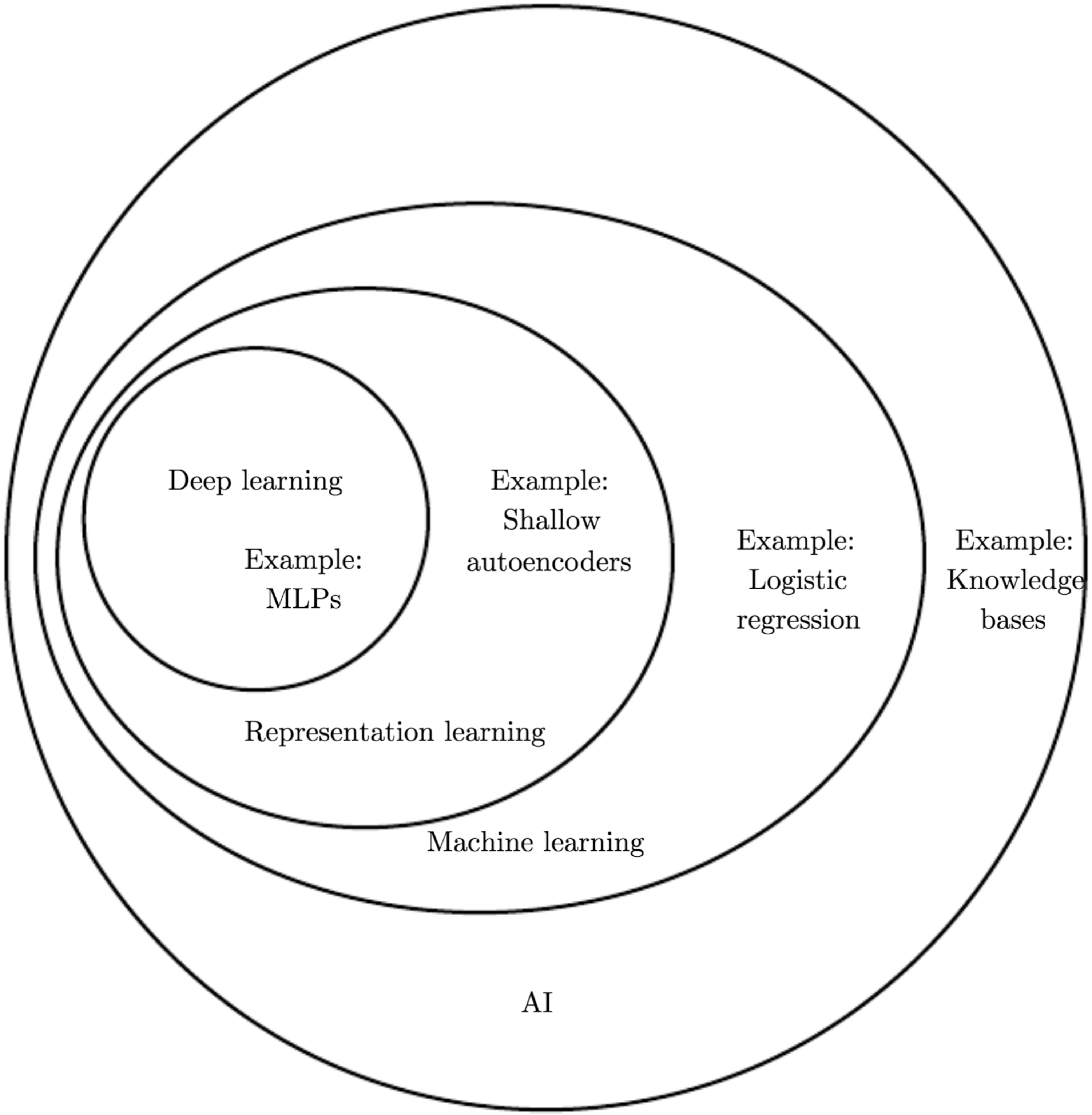
Jet Substructure Classification in High-Energy Physics with Deep Neural Networks

Pierre Baldi,[1] Kevin Bauer,[2] Clara Eng,[3] Peter Sadowski,[1] and Daniel Whiteson[2]

[1] *Department of Computer Science, University of California, Irvine, CA 92697*
[2] *Department of Physics and Astronomy, University of California, Irvine, CA 92697*
[3] *Department of Chemical Engineering, University of California Berkeley, Berkeley CA 94270*

(Dated: April 1, 2016)

Ecosystem



deeplearningbook.org: Goodfellow, Bengio, Courville

How do machines learn?

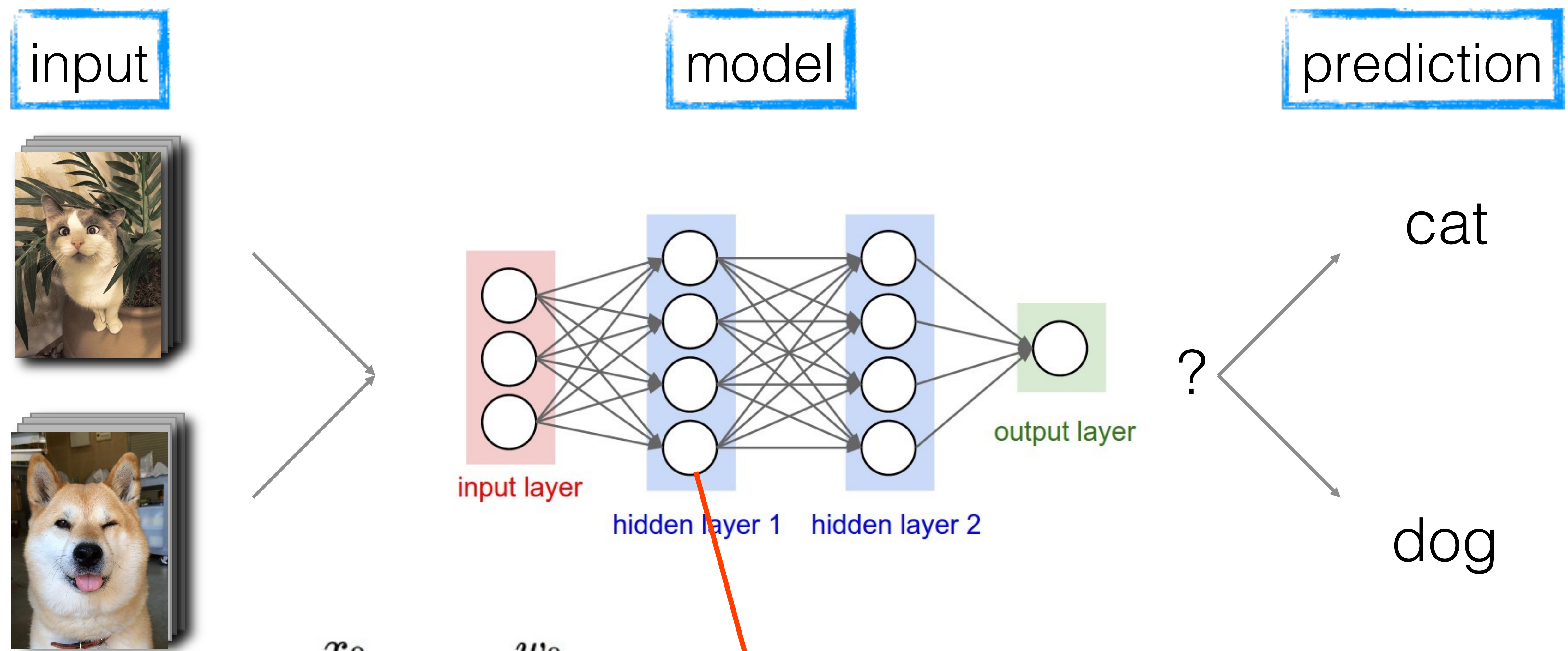
Supervised

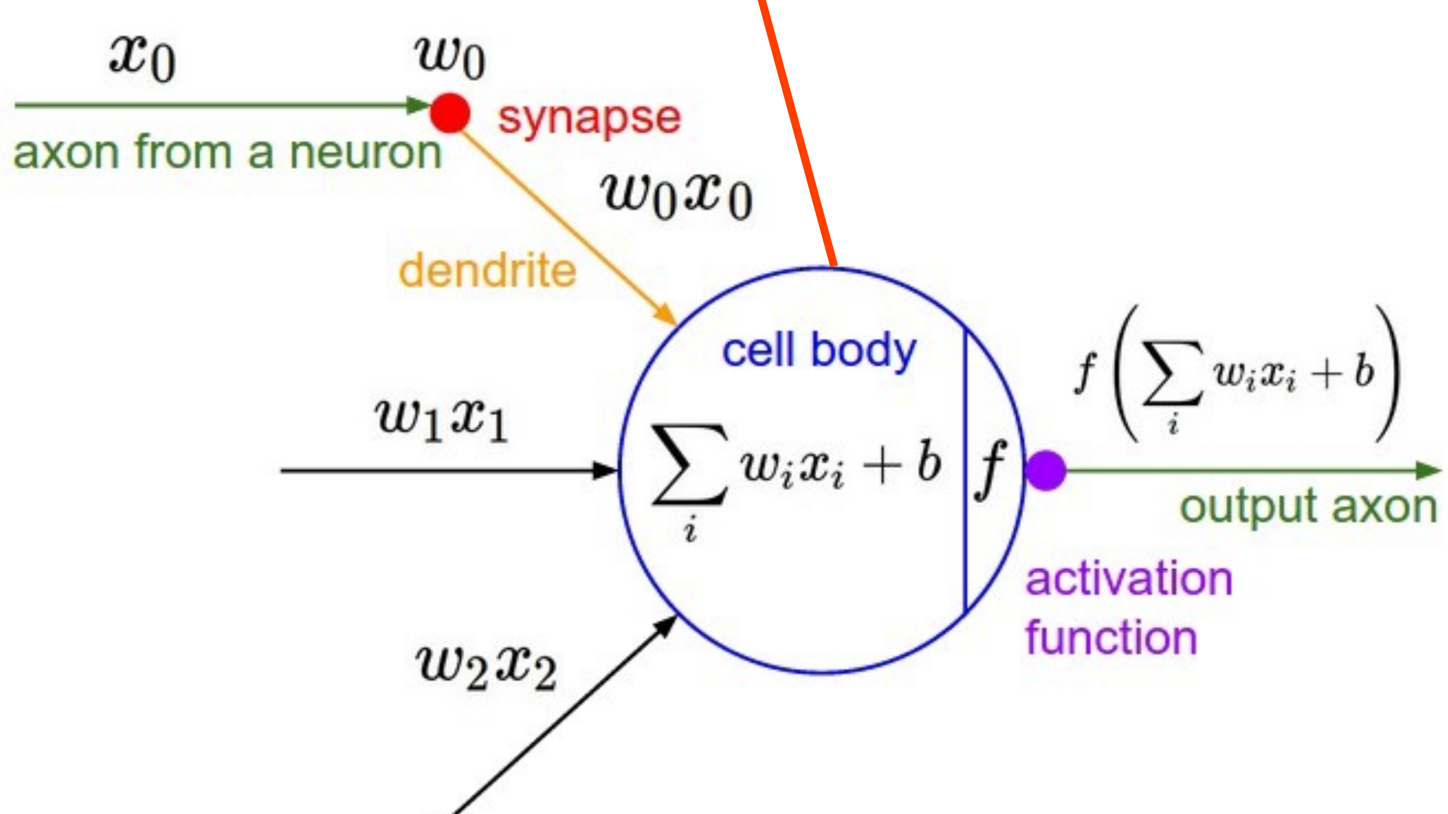
- ***Convolutional Neural Networks***
- *Support Vector Machine*
- *Random Forest*

*Un*supervised

- *Principle Component Analysis (PCA)*
- *K-means clustering*
- *t-Distributed Stochastic Neighbor Embedding (t-SNE)*

Convolutional Neural Network: Overview

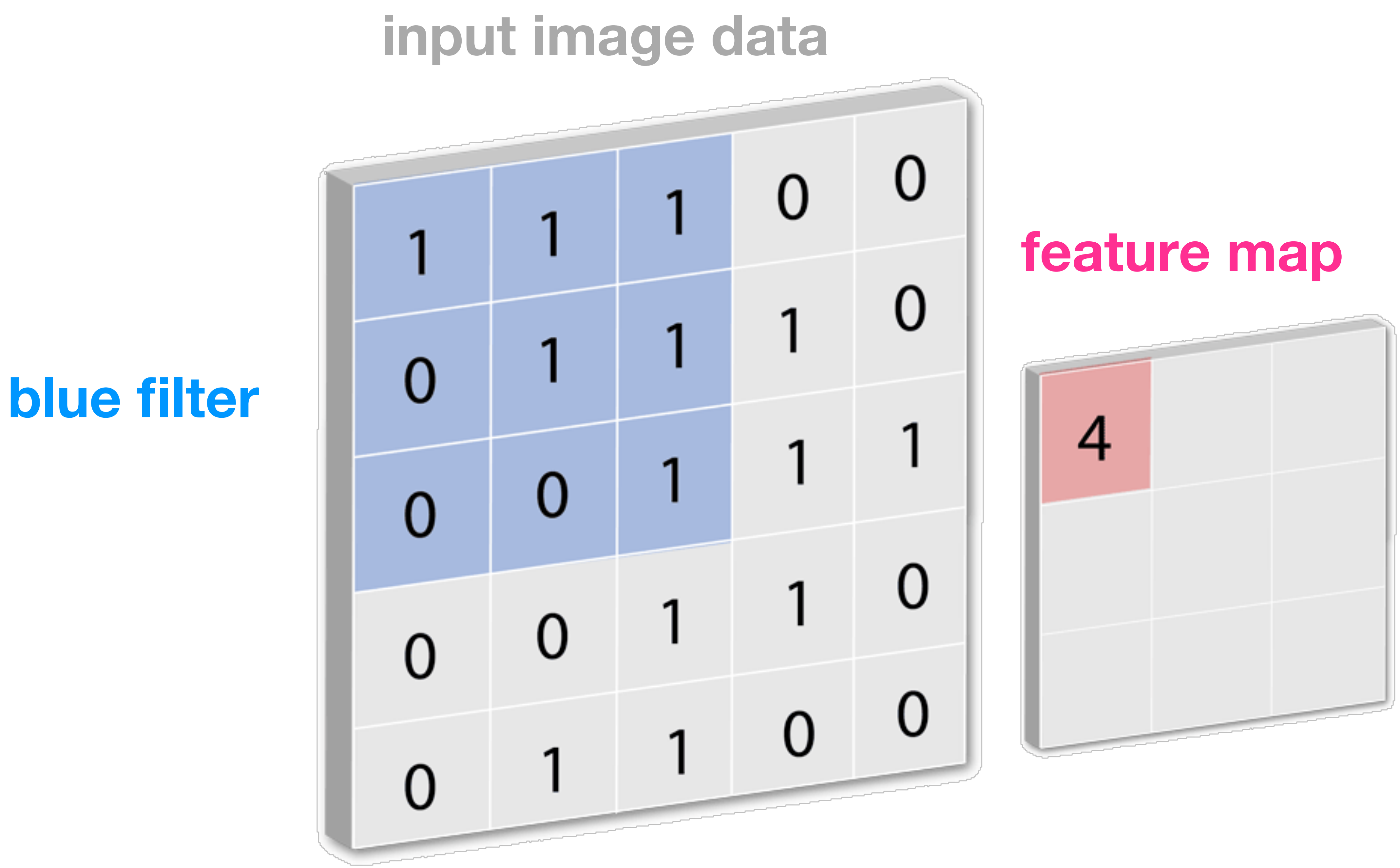




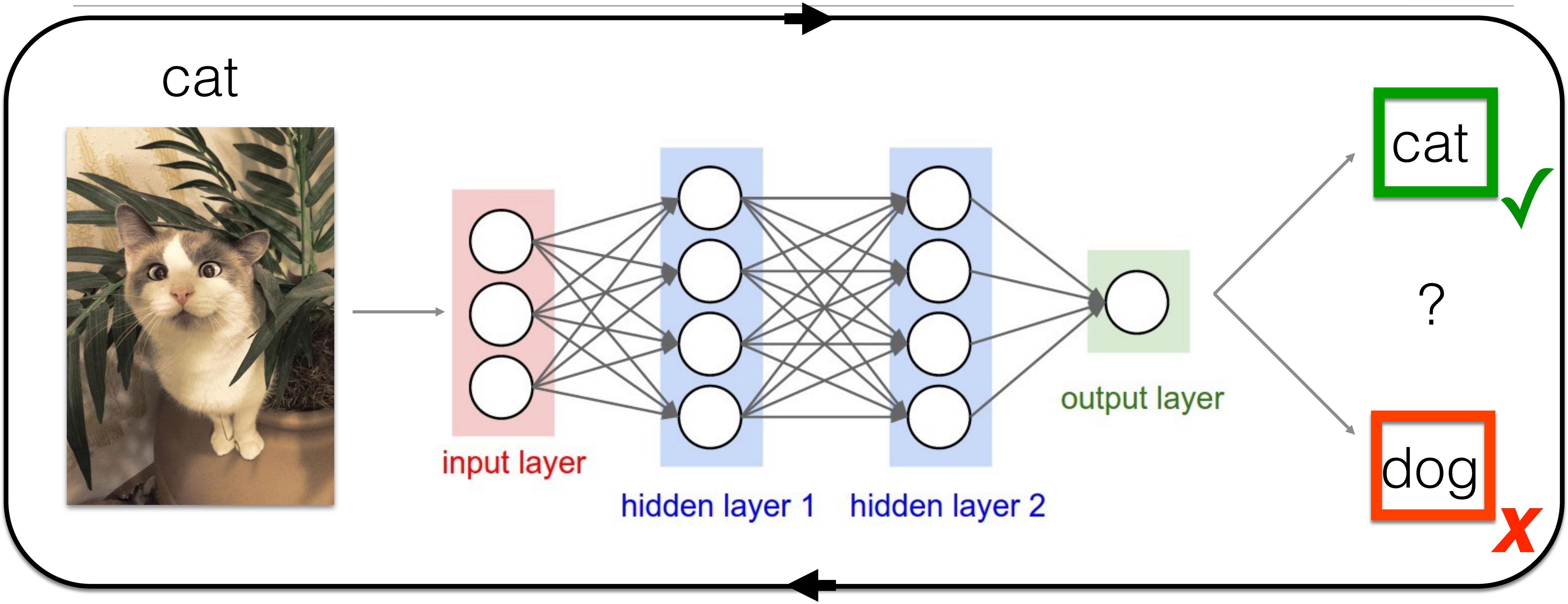
- Typ
  - (
  - / s others.
  - I age regions
  - I

Convolutional Neural Network: Convolution

- Each pixel in the **blue filter** is one parameter in the network model
- The resulting **feature map** is the result of the convolution.



Convolutional Neural Network: Training

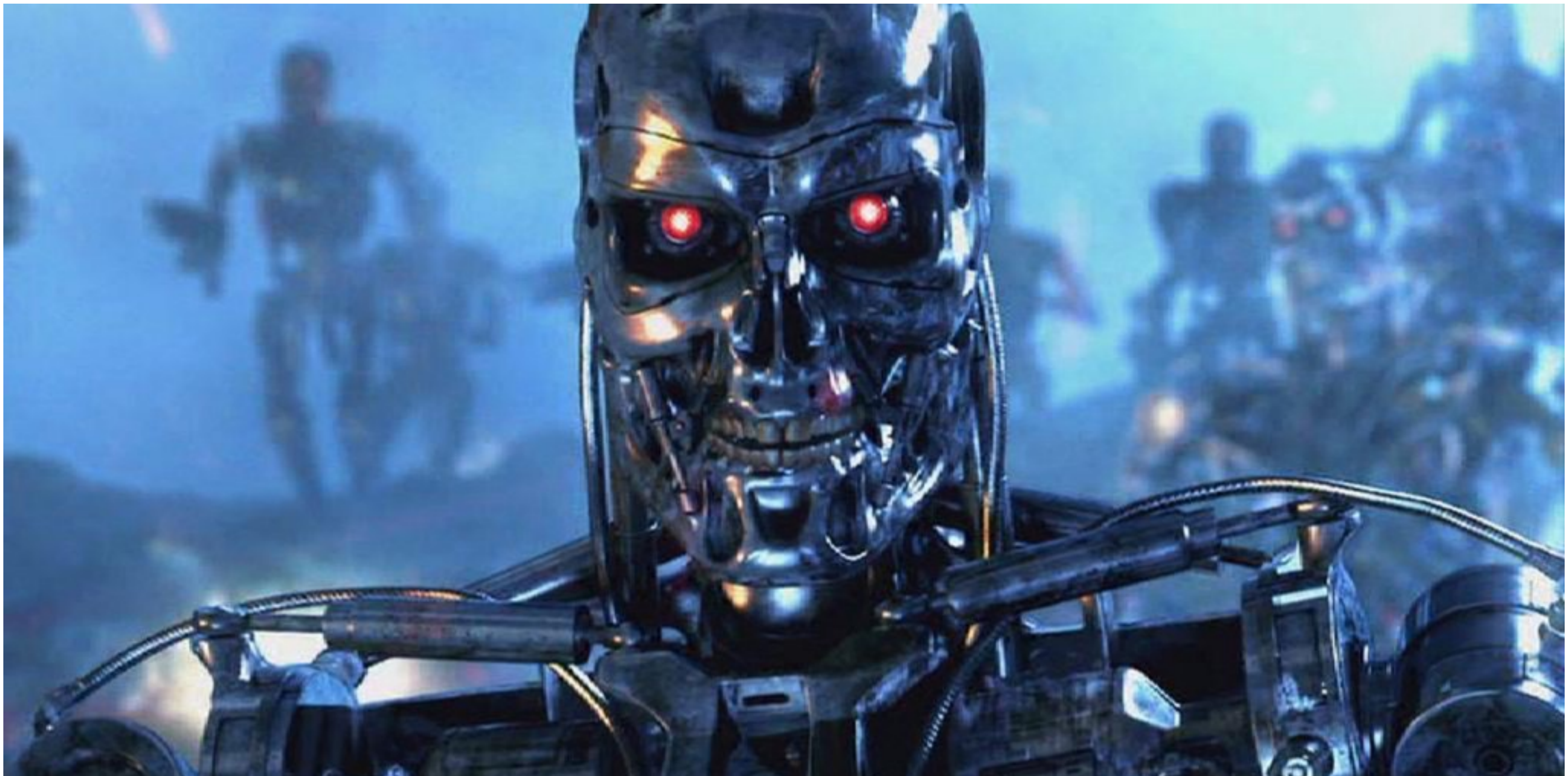


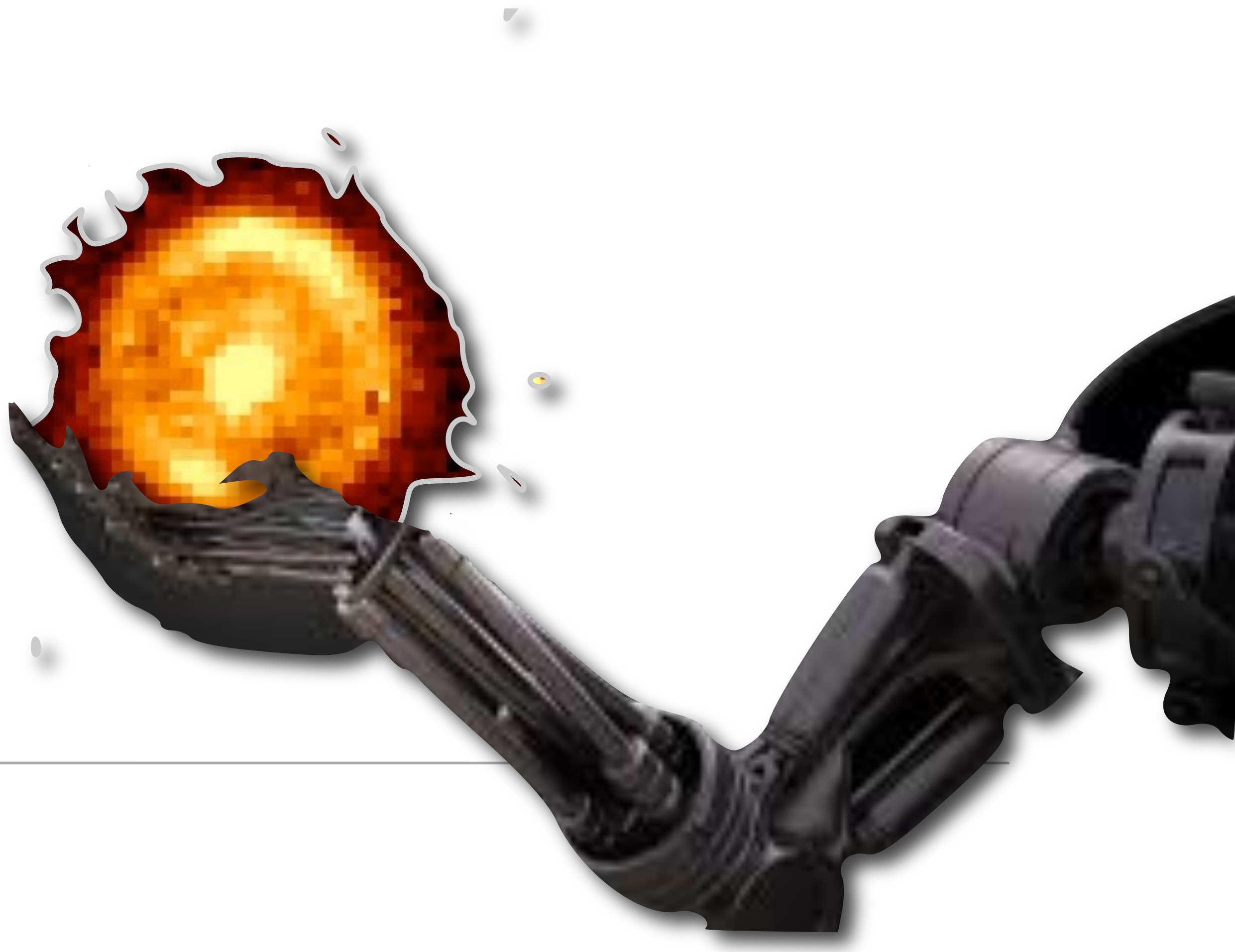
- Minimize error (**E**)
  minimize error between prediction (**f**) and true label (**y**)

$$E(\mathbf{w}) = \sum_{i=1}^{N} \left(y_i - f_{\mathbf{w}}(\mathbf{x}_i)\right)^2$$

- *Stochastic gradient descent* is typically used to optimize **w** by propagating the error back through each layer

Evolution of networks



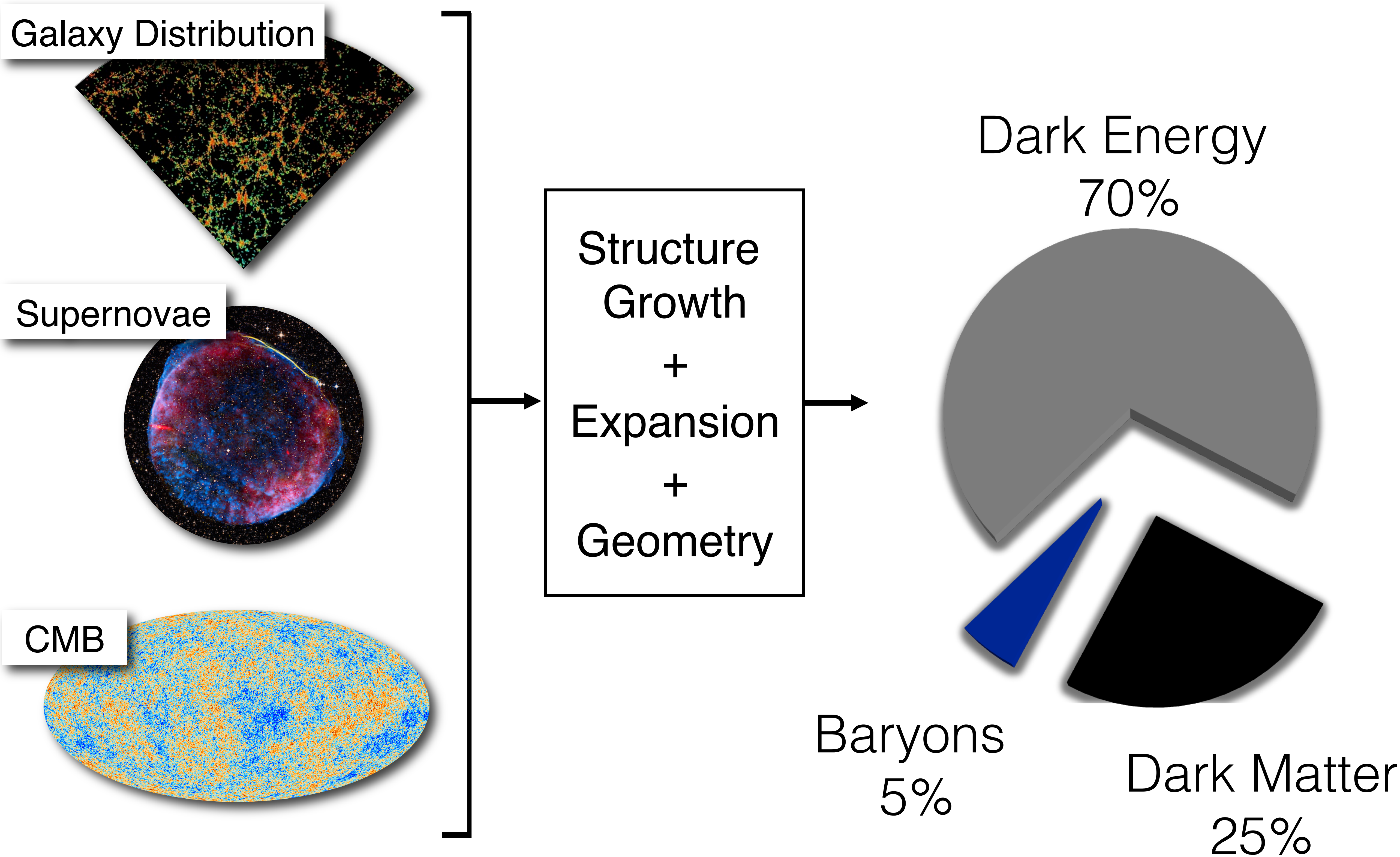


Deep Learning in Astro

Example applications

- Work in coordination with
  - Irshad Mohammed (FNAL)
  - Adrian Price-Whelan (Princeton)

Path to the Modern Cosmological Paradigm

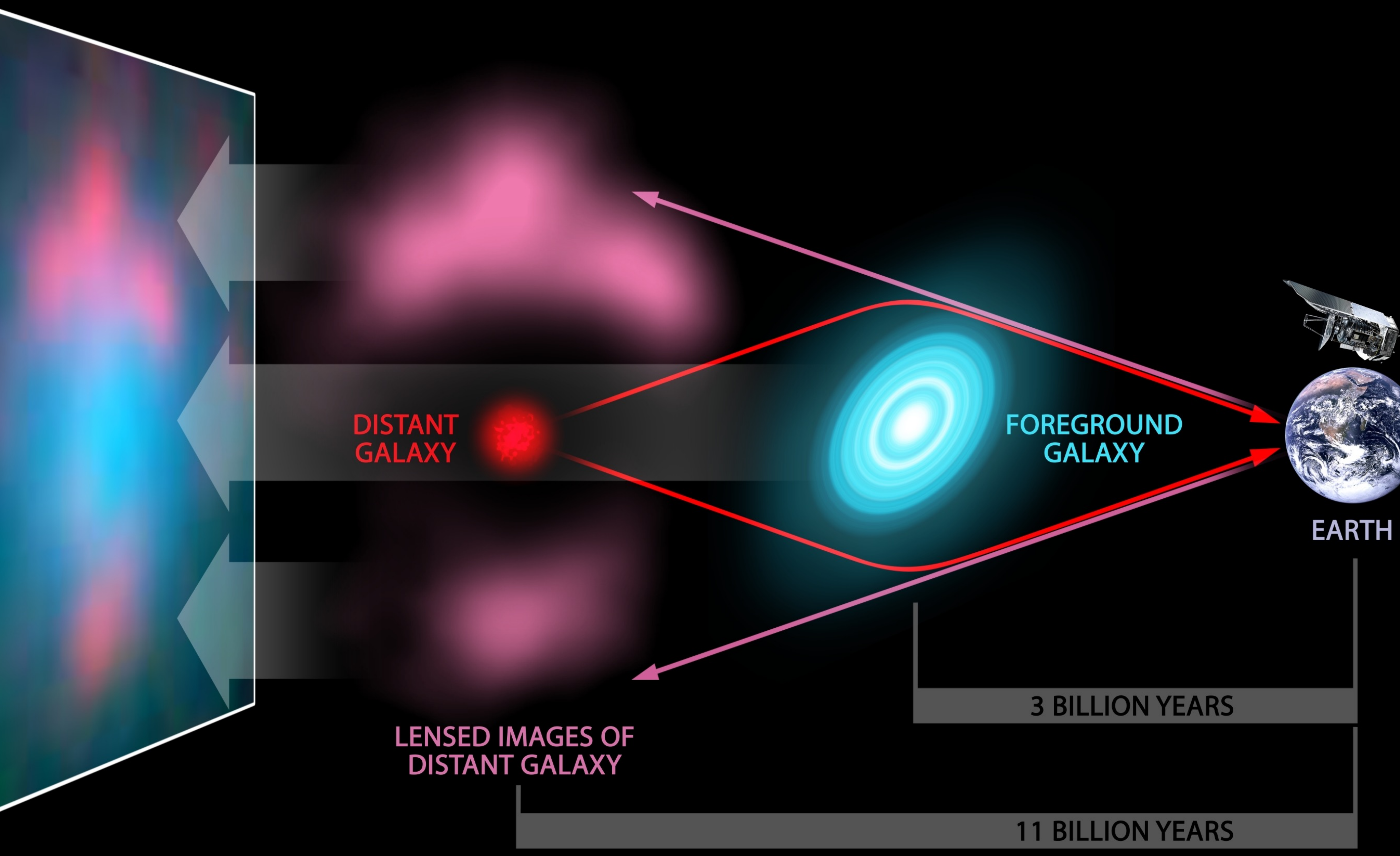


DECam installed in 2012

courtesy Reidar Hahn

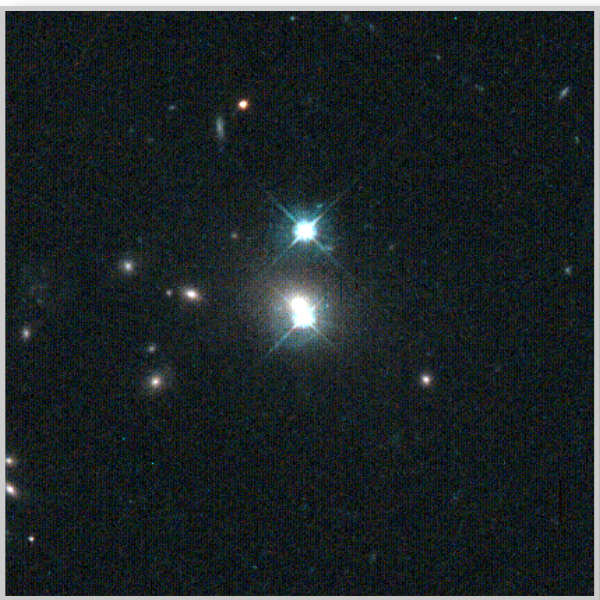
Early DES Data

Position Flux Shape


DISTANT GALAXY
FOREGROUND GALAXY
EARTH
LENSED IMAGES OF DISTANT GALAXY
3 BILLION YEARS
11 BILLION YEARS

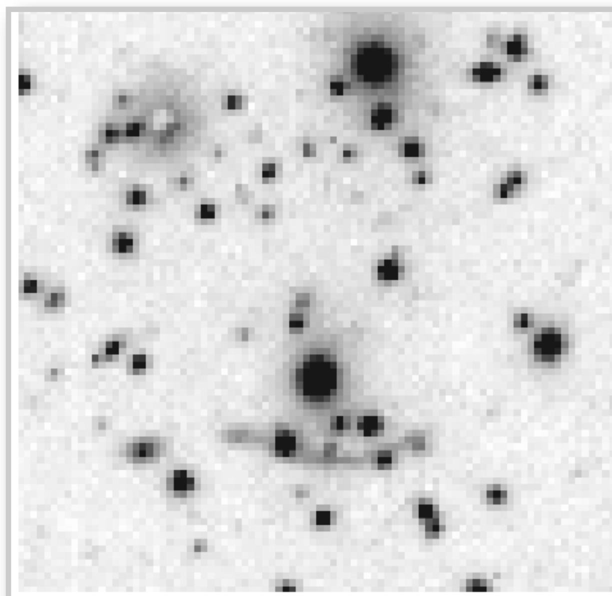
Strong Lensing Milestones

- **1979: Quasar**
  Twin Quasar SBS 0957+561



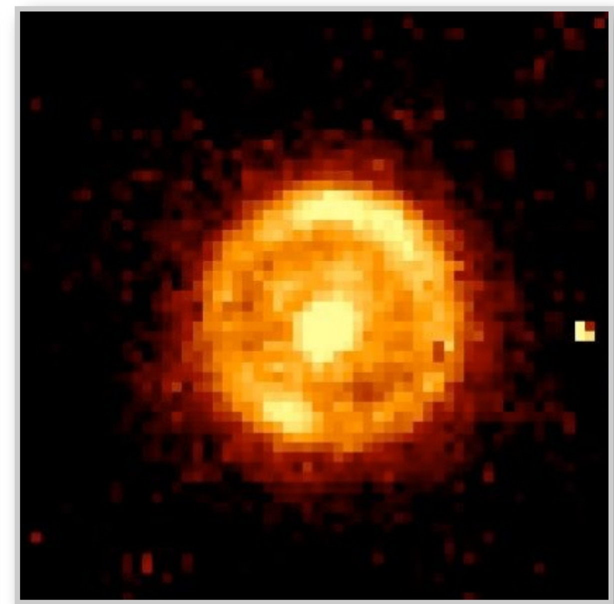
- Walsh, Carswell, Weyman 1979

- **1986: arcs**
  Cluster Abell 370



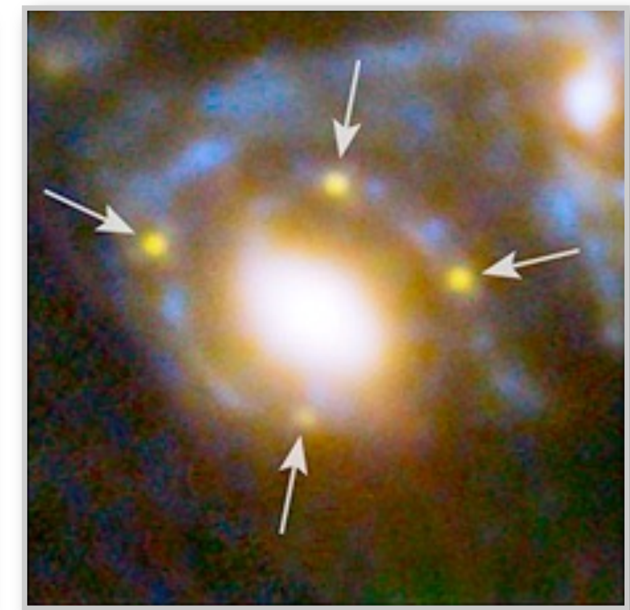
- Lynds & Petrosian 1986; Soucail +1987

- **1998: Einstein Ring**
  Galaxy JVAS B1938+666



- King+1998

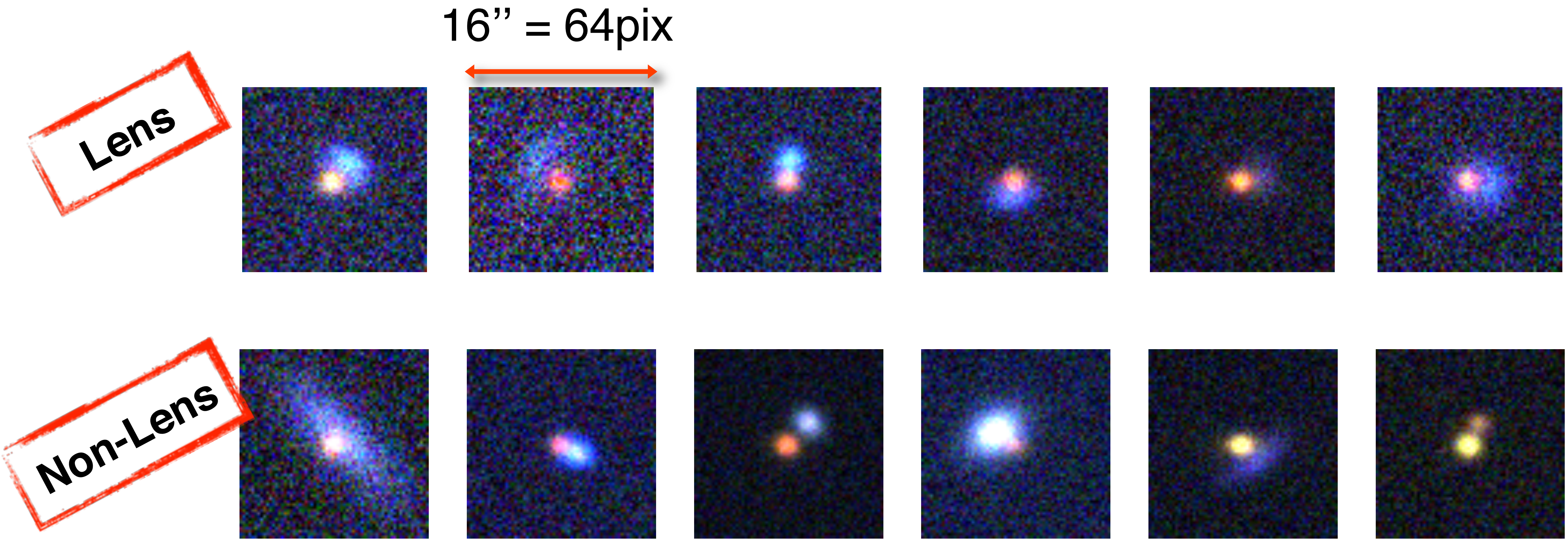
- **2014: Supernova**
  Cluster MACS J1149.6+2223



- Kelly+2014

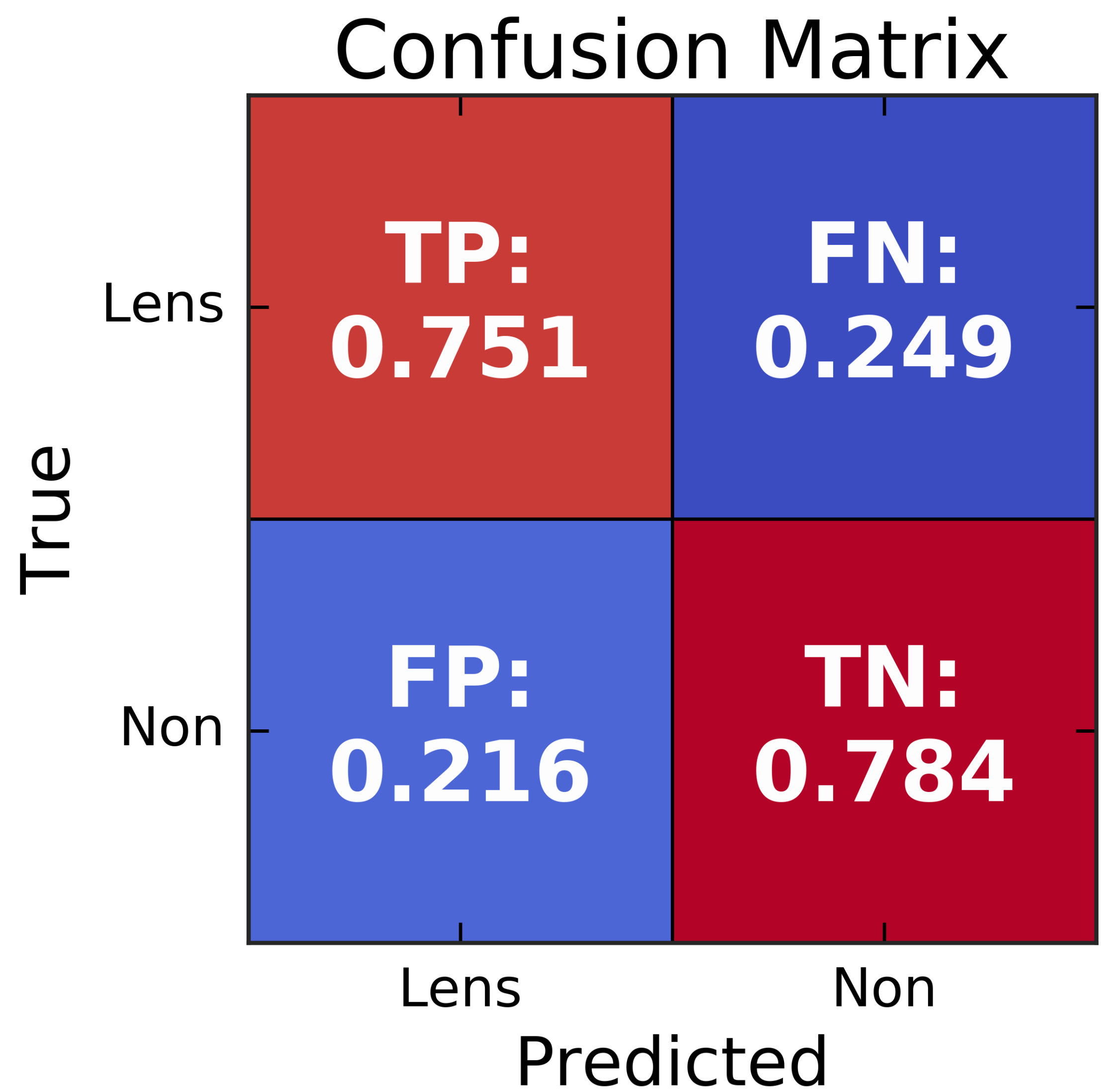
- **~1000** lenses currently exist across all wavelengths
- **~2000** predicted for DES footprint
- **~120,000** predicted for LSST footprint

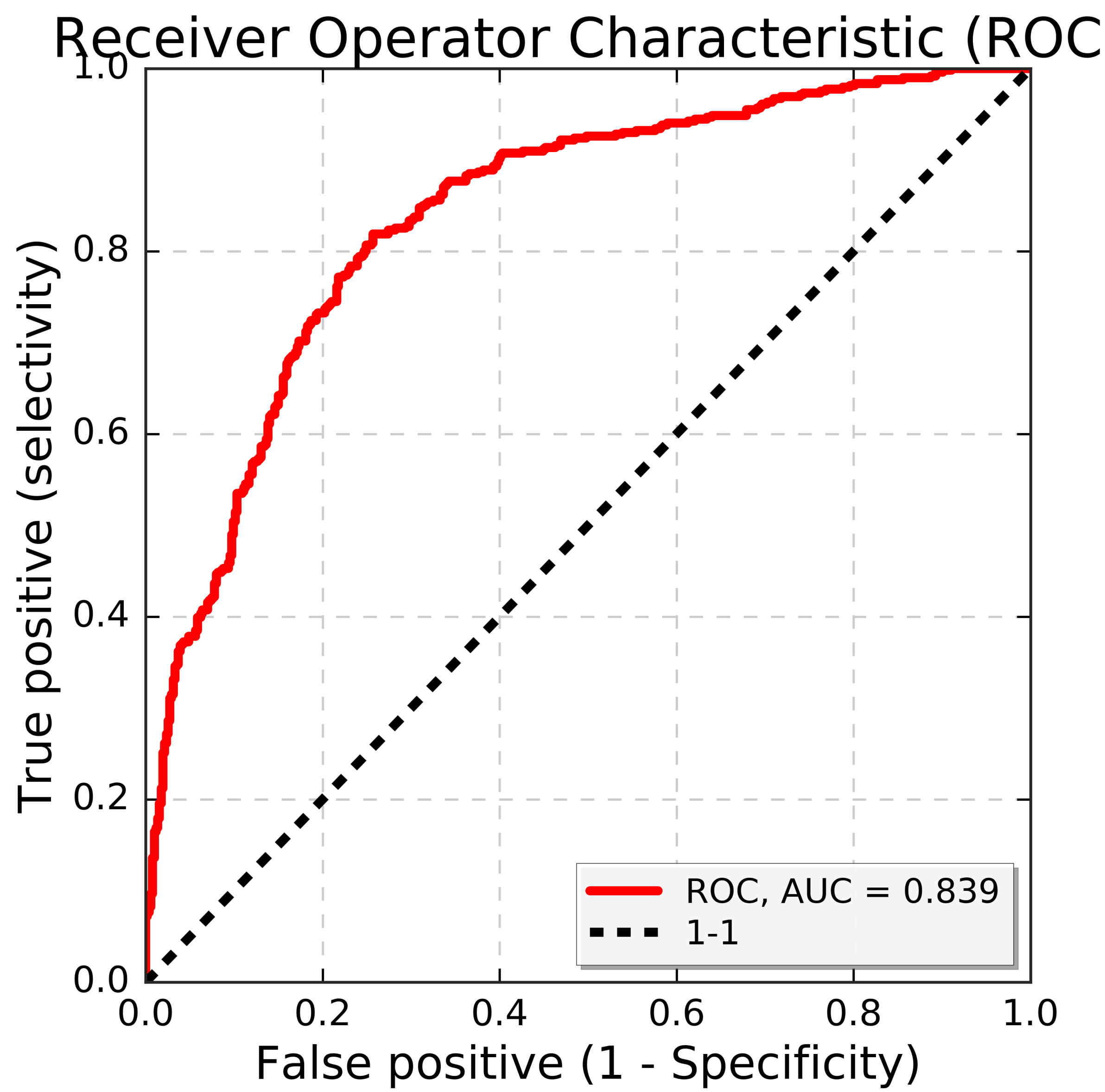
Deep Lensing: Lens Classification (Nord+2017, in prep.)



- <u>Simulations for Training Set</u>
  - **Training** 15K objects; 50 epochs
  - **Empirically motivated** density and light profiles of sources and lenses
  - **Mimic DES Survey characteristics:** noise levels, exposure time, PSF, photometry, resolution

Deep Lensing: Classification results for sims



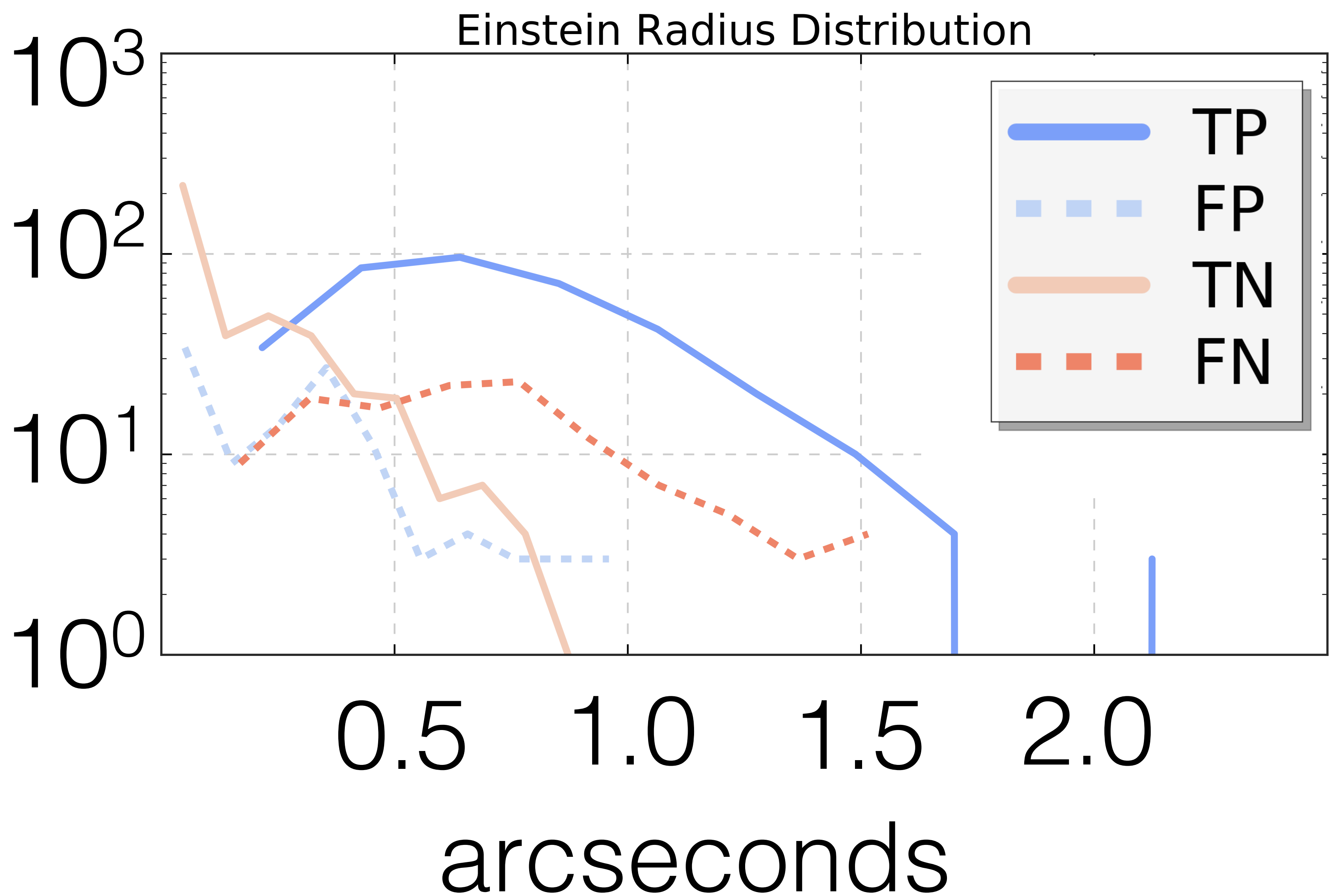


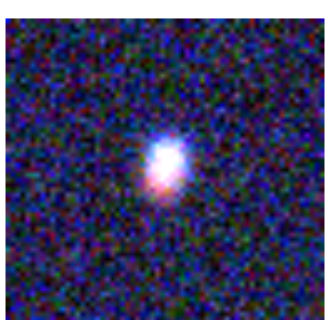
<u>Confusion matrix</u> shows **high precision and recall** when testing on images NOT used for training.

<u>ROC Curve</u> shows the accuracy as the threshold of probability for detection is incremented

Diagnostics: Einstein Radius

- False-identification rates are higher at small Einstein radius, where there can be more confusion in discerning source image from lens.

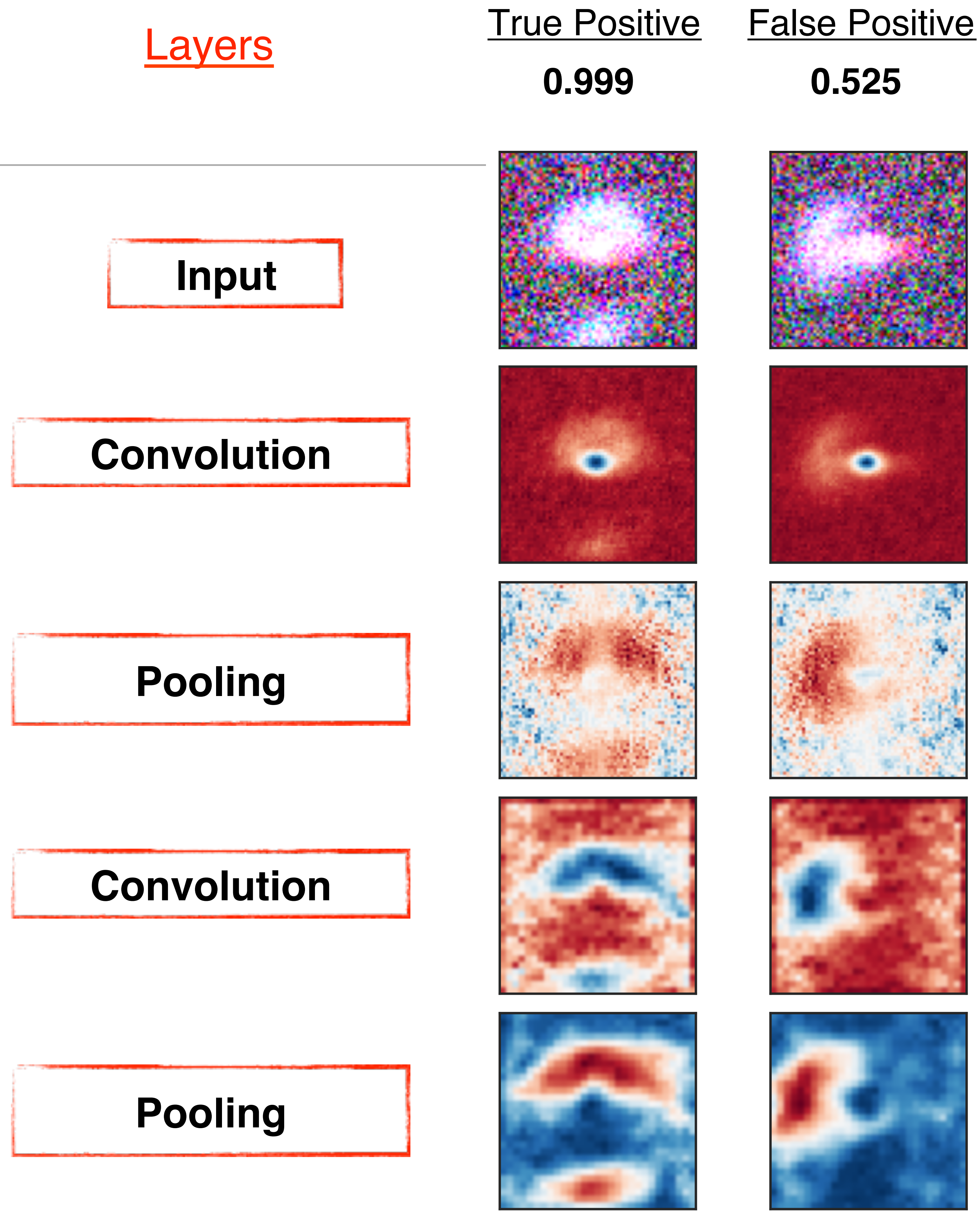






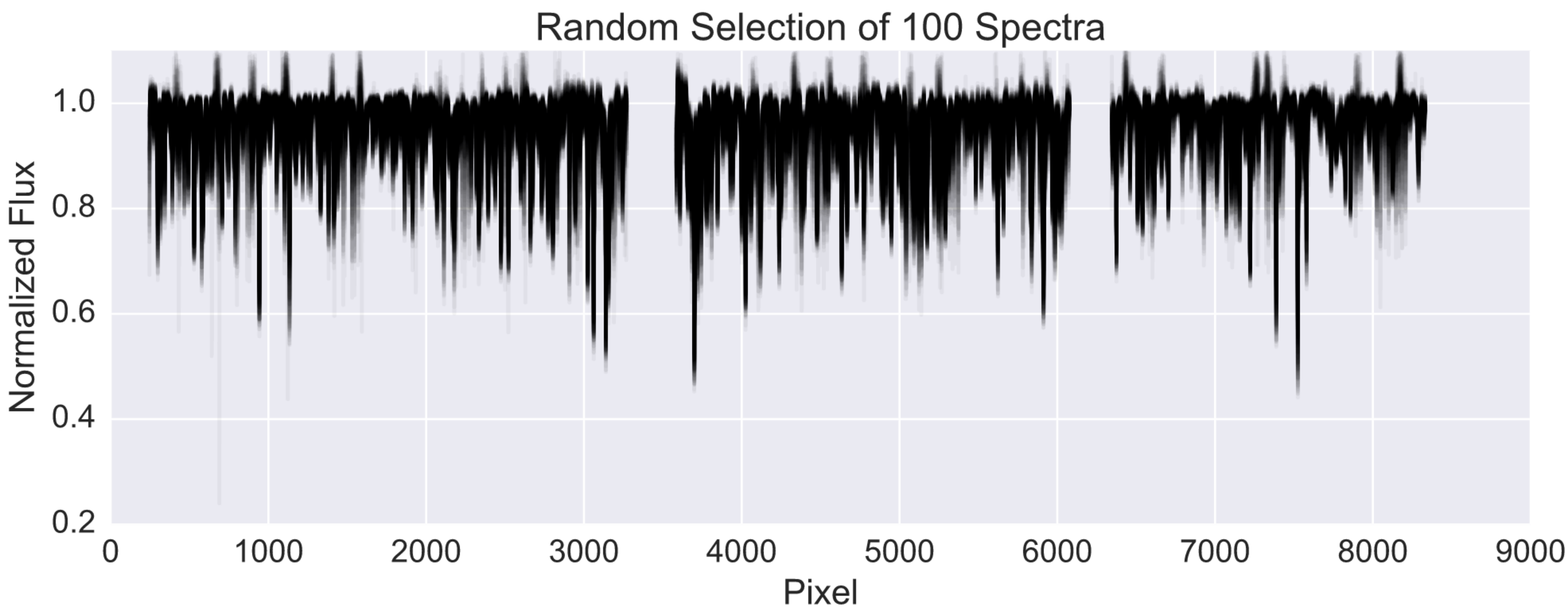
What’s Inside?

- Each column is a different object and its probability of detection in the network.
  Left: True positive
  Right: False positive
- Convolution layers filter the images to highlight features
- Pooling layers down-sample images, efficiently reducing parameters for modeling
- See also work by Lanusse+17, Trejillo+17 for lens-finding with CNNs

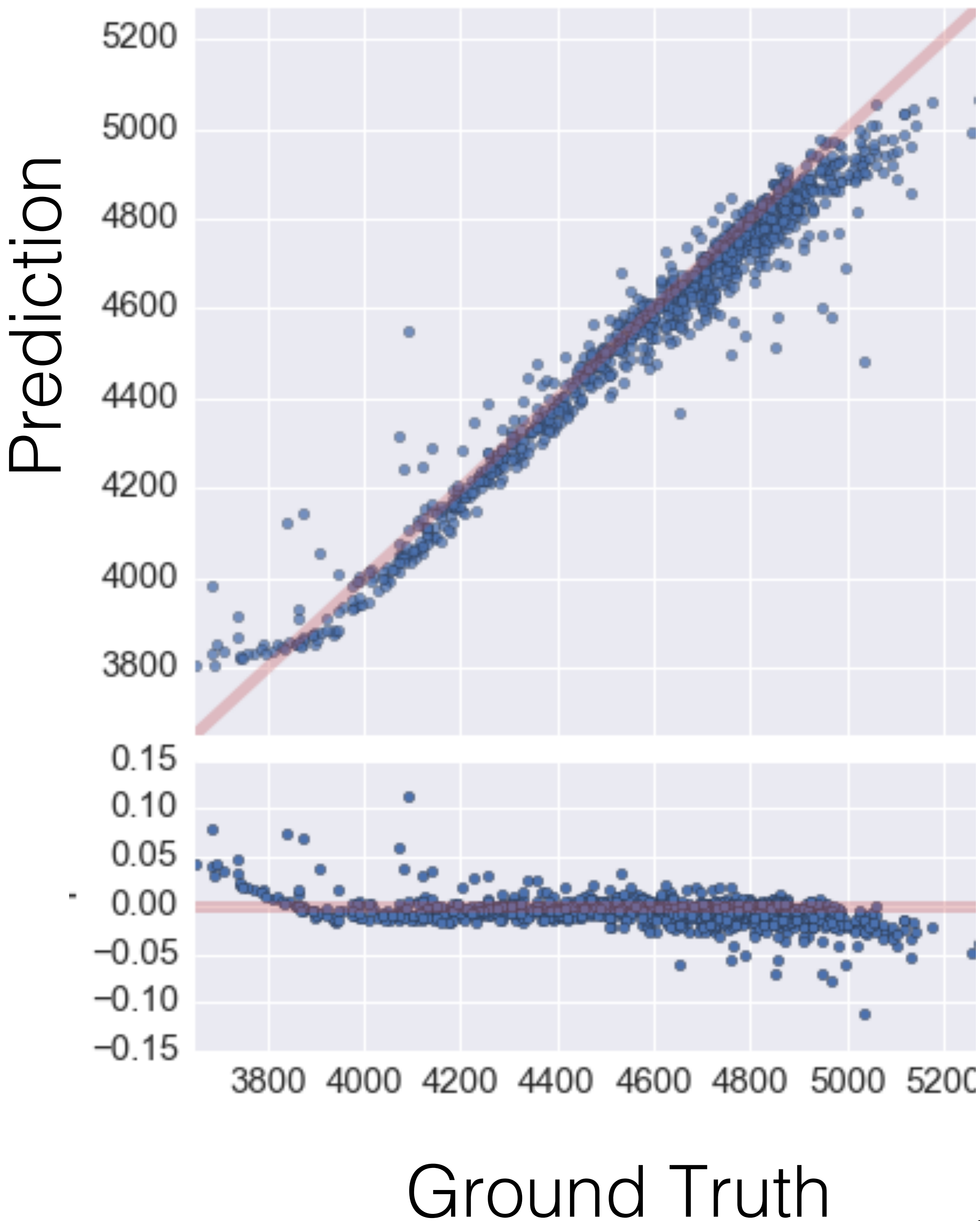


DeepSpec: modeling stellar spectra (Nord, Price-Whelan+2017, in prep)

- Data: Apogee stellar spectra with labeled quantities
  - *Teff, log g, metallicity* (see Ness+2015)
- 1D ConvNets
  - 3 convolution layer, 3 pooling, and 1 drop out layer
  - 15 lines of (DL) code, a GPU and 40 minutes of compute time.



Temperatures (T_{eff} [K])



- Most predictions are < 1% error
- Still require methods to assess uncertainties on predicted Teff

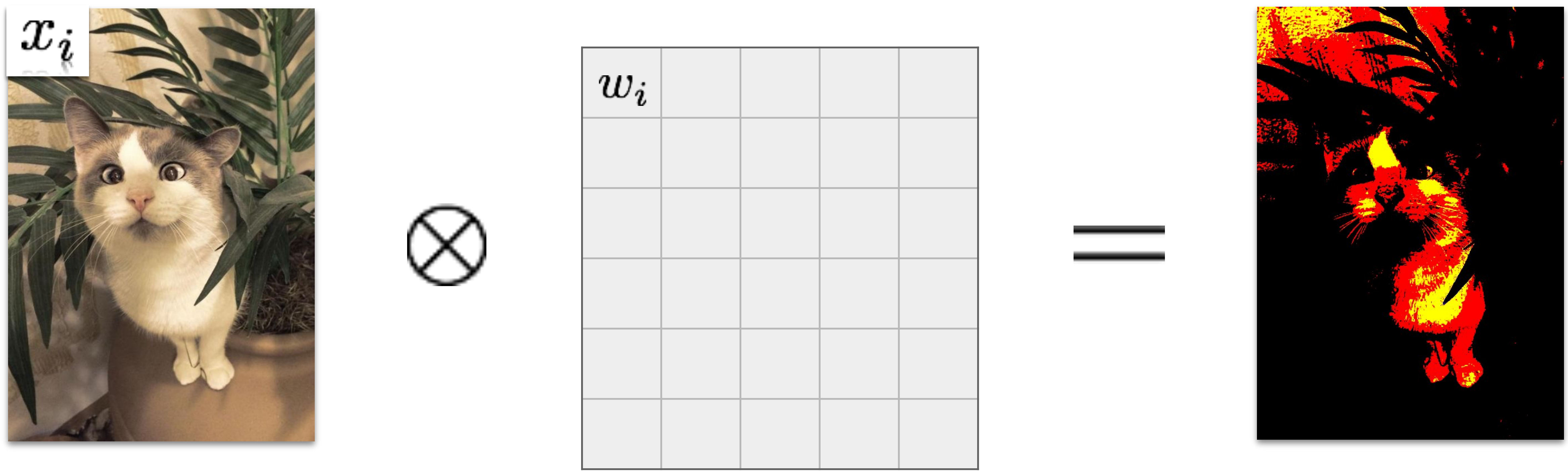
Outlook

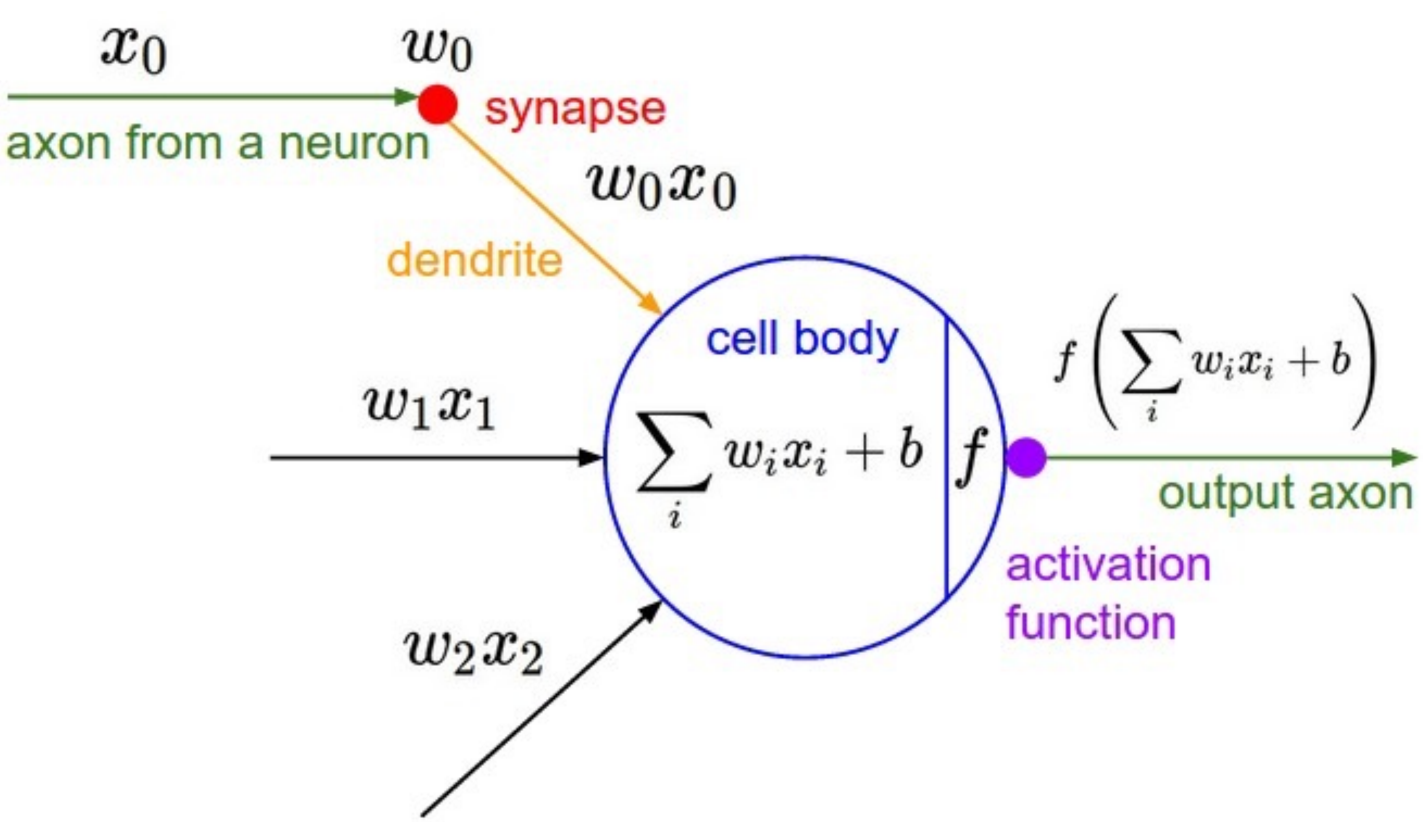
- <u>Caution:</u>
  - data set construction
  - propagation of uncertainties
  - difference between training and test sets
- <u>Opportunity:</u>
  data-driven approaches offer *complementary techniques* and *insights* for exploring big data

Example CNN code:
github.com/bnord/cosmonet

Extras

Convolutional Neural Network: Convolution

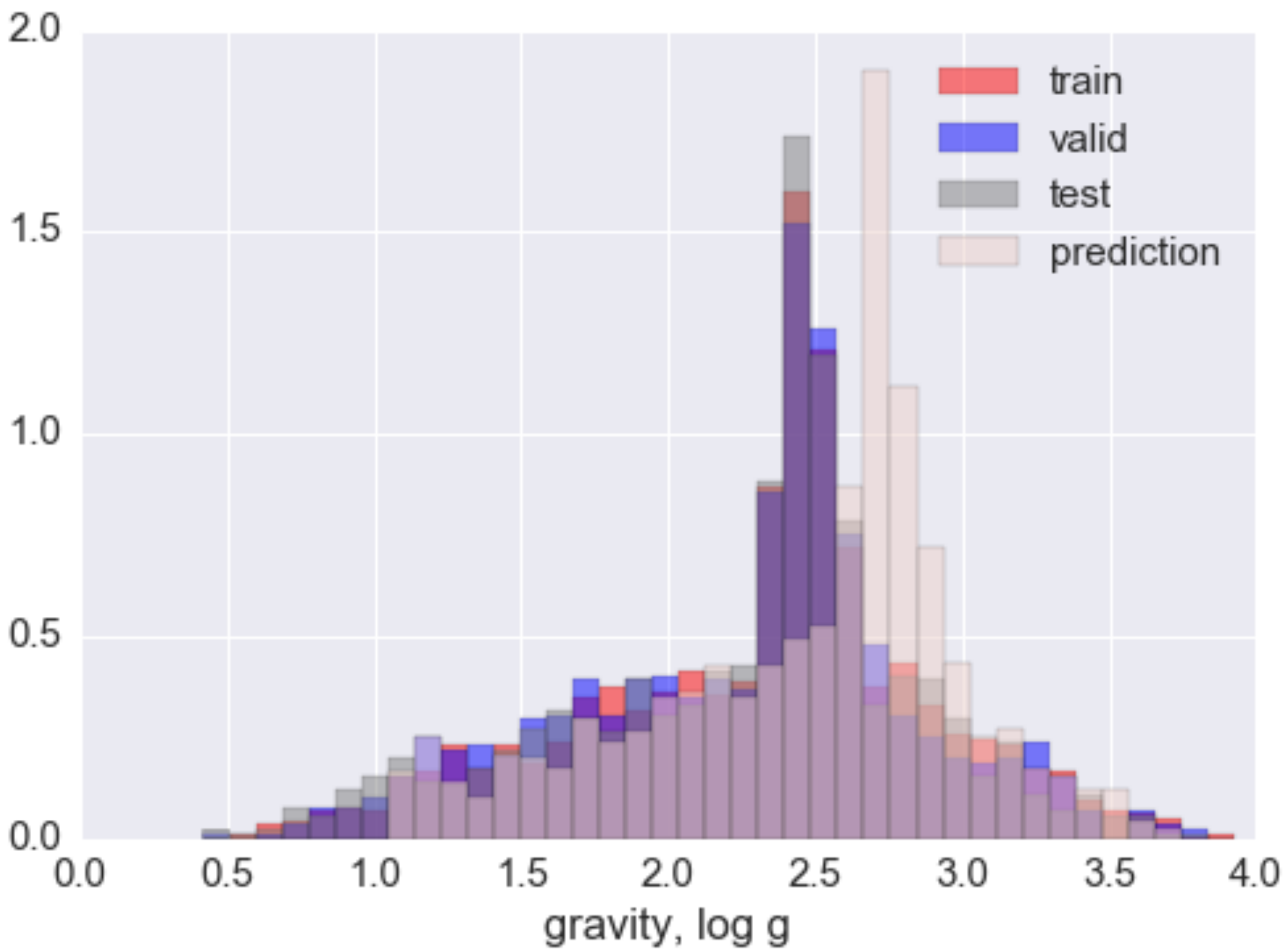


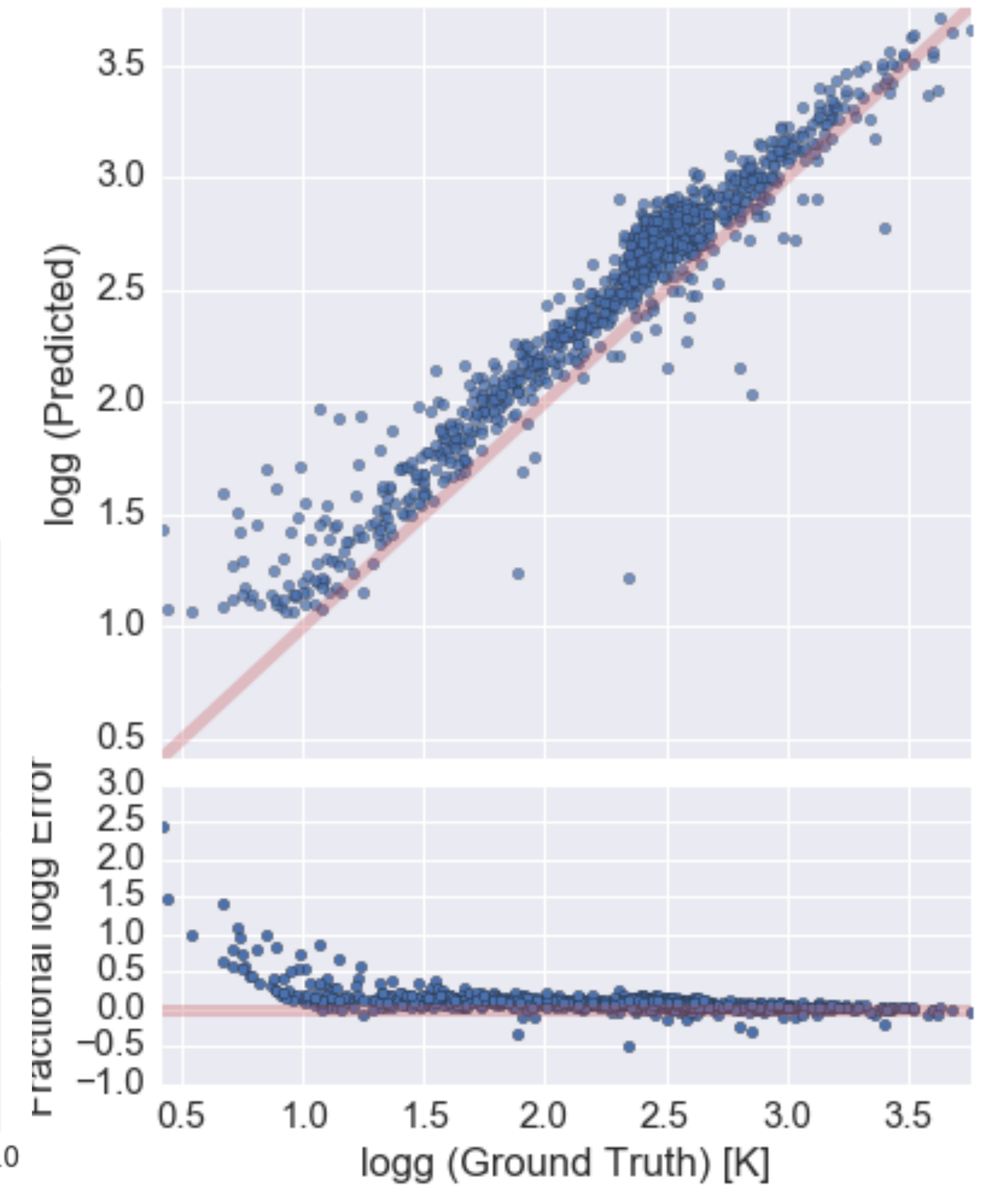


- <u>Each computational neuron</u> is an image filter, where $\mathbf{w_i}$ is the value of a pixel in that filter and a model parameter
- <u>During *convolution and activation*</u>, the model acts on the input image, highlighting features, such as edges or circles.

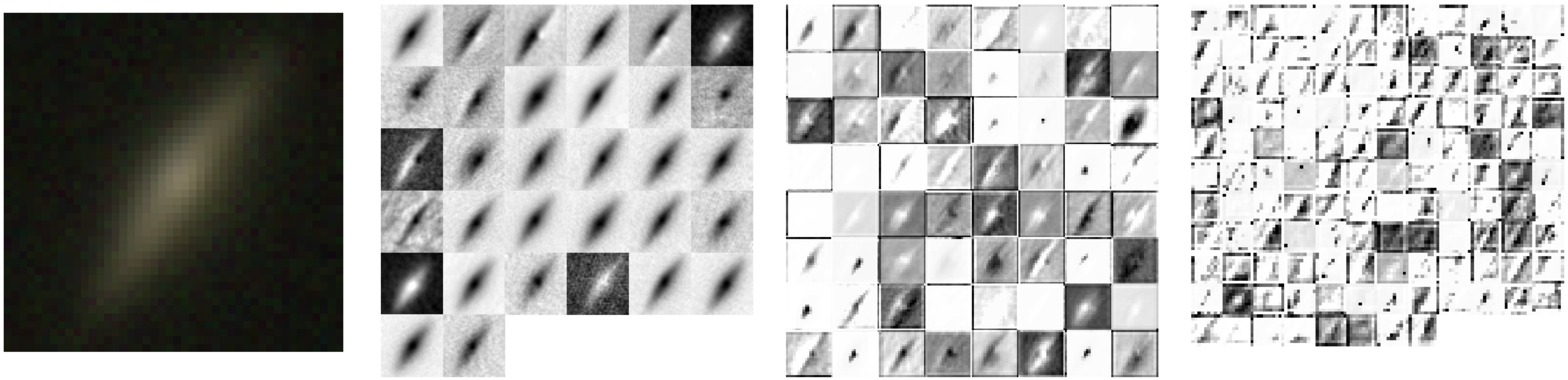
Gravity (log g)

- (same architecture)
- large biases
- architecture achieves very low losses
- we may need more diverse training set.



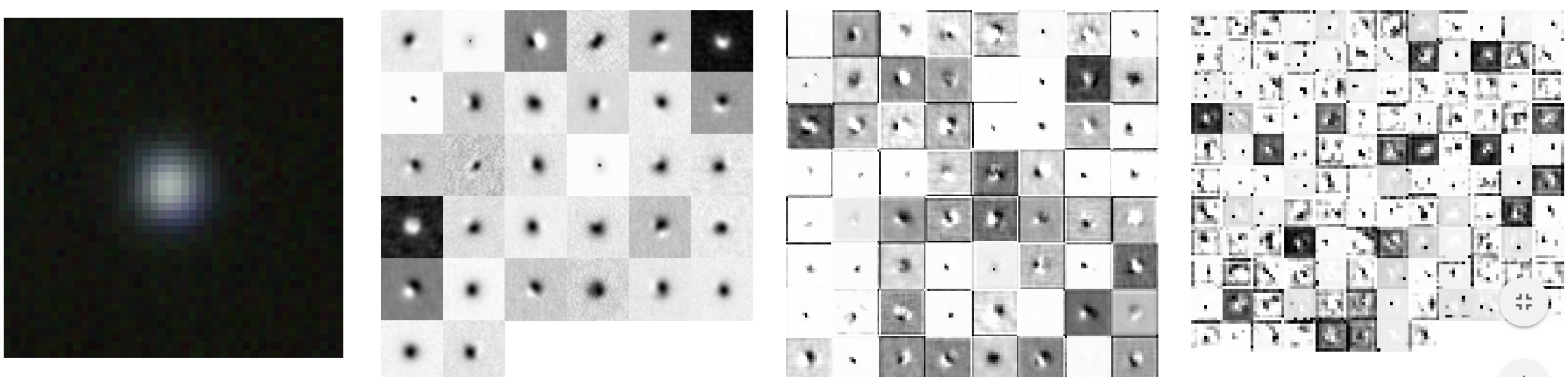


Star-galaxy classification (Kim+Brunner 2016)



(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)

Figure 12. Similar to Figure 6 but for a galaxy in the SDSS data set.

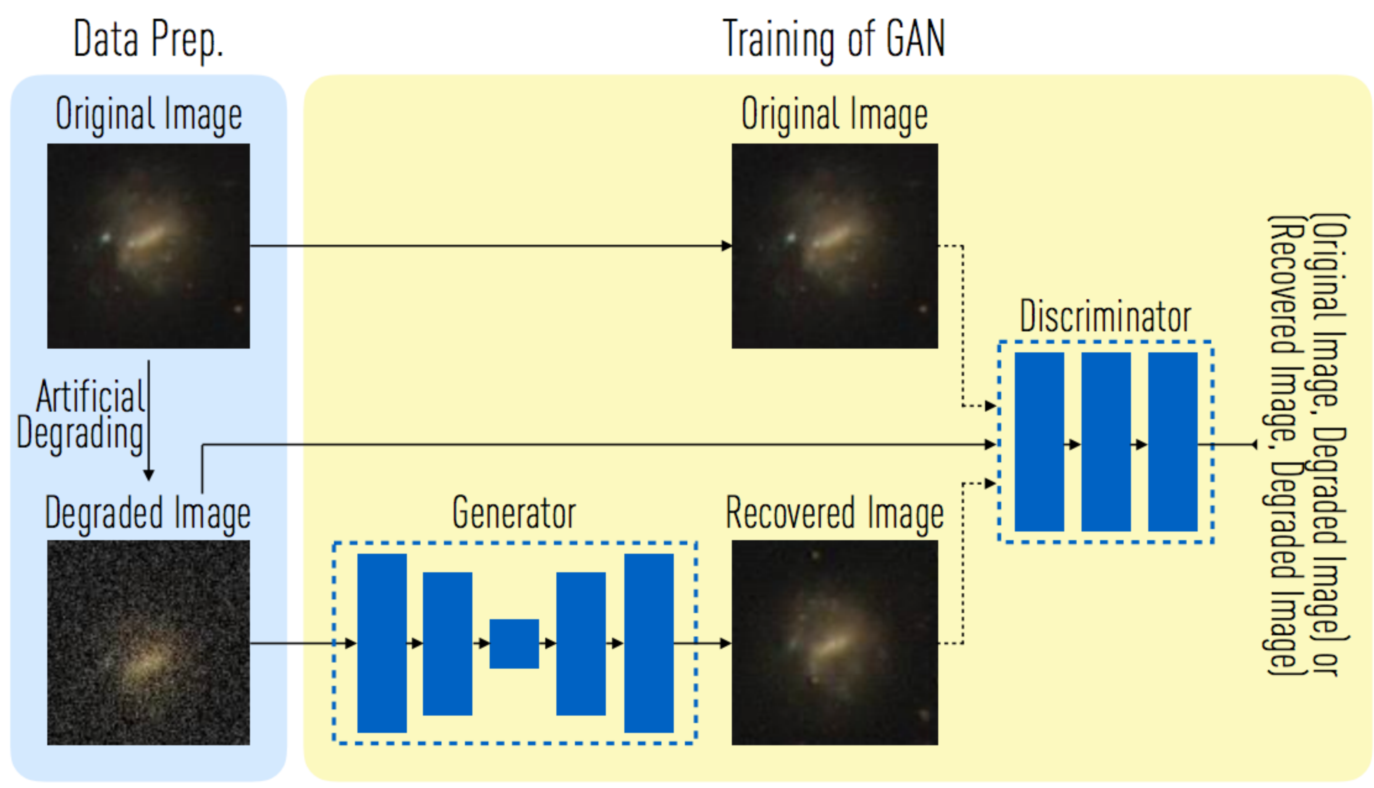


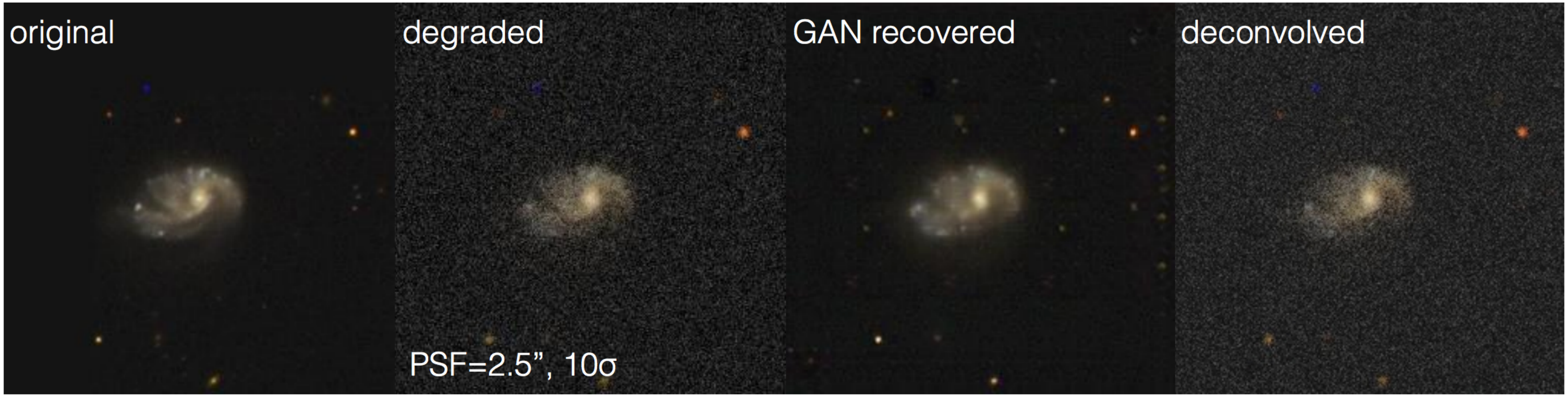
(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

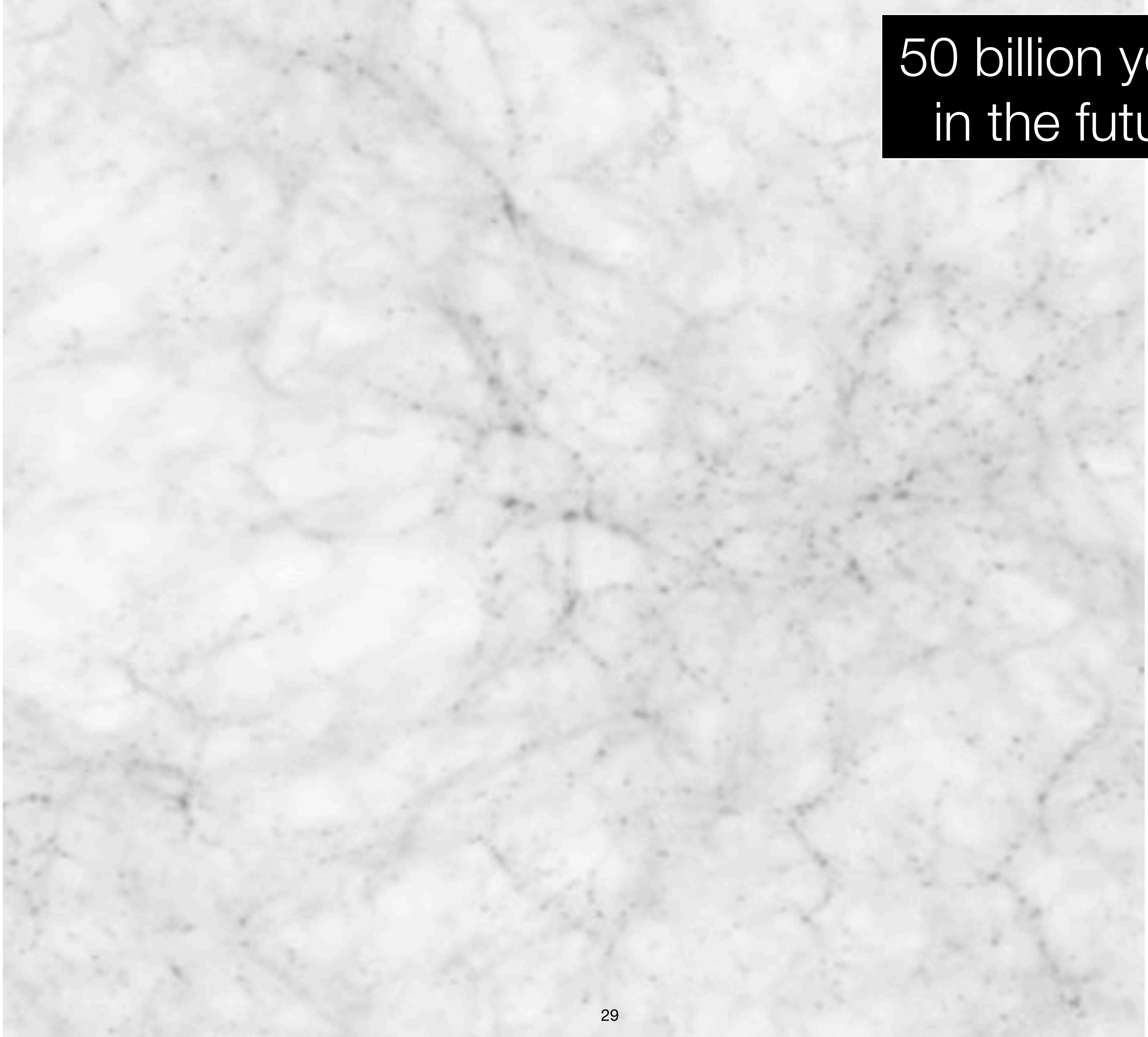
Figure 13. Similar to Figure 12 but for a star in the SDSS data set.

Galaxy Image Simulation (Schawinski+2017)

- **G**enerative **A**dversarial **N**etworks offer an avenue to simulate realistic images of galaxies.
- We currently lack the functionality to propagate errors with these frameworks, leaving us without estimates of noise, let alone the ability to track noise sources.

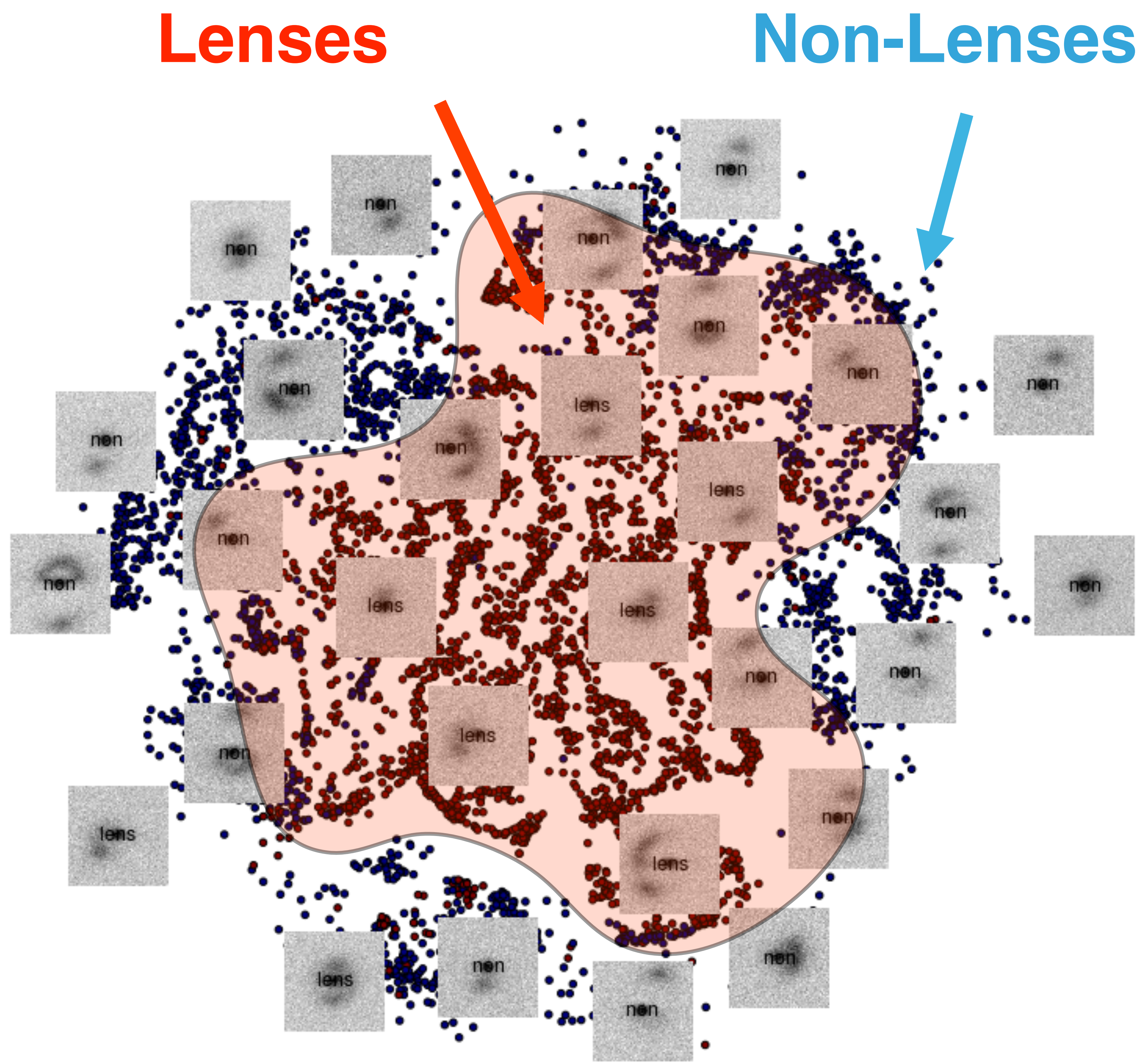




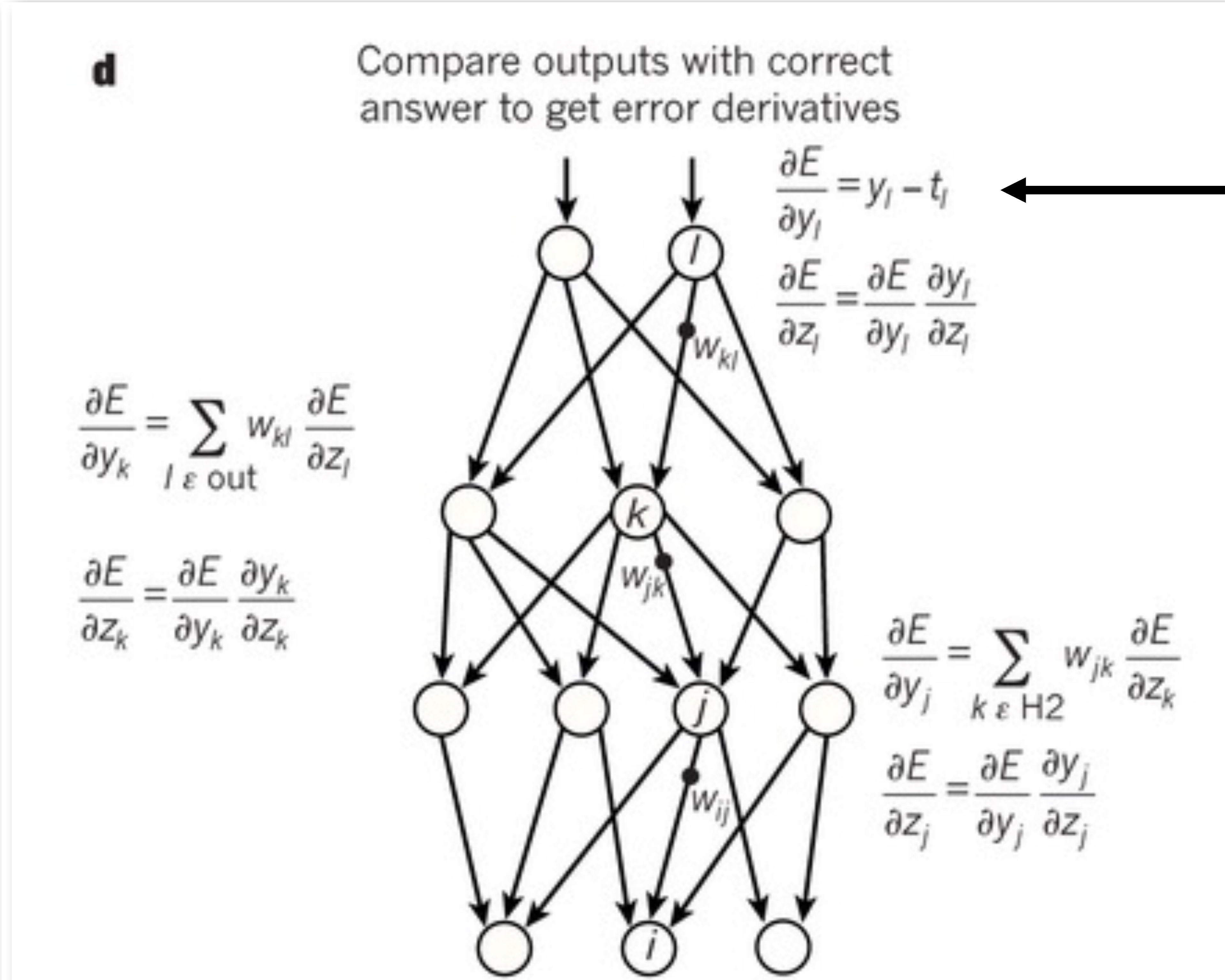

50 billion years
in the future

Unsupervised Learning

- Lenses on the inside, Non-lenses on the outside. well-separated by contour
- t-SNE: algorithm for dimensionality reduction



Backpropagation = Chain Rule



compare prediction to truth

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica


Dark Energy Survey