Let’s get deep.

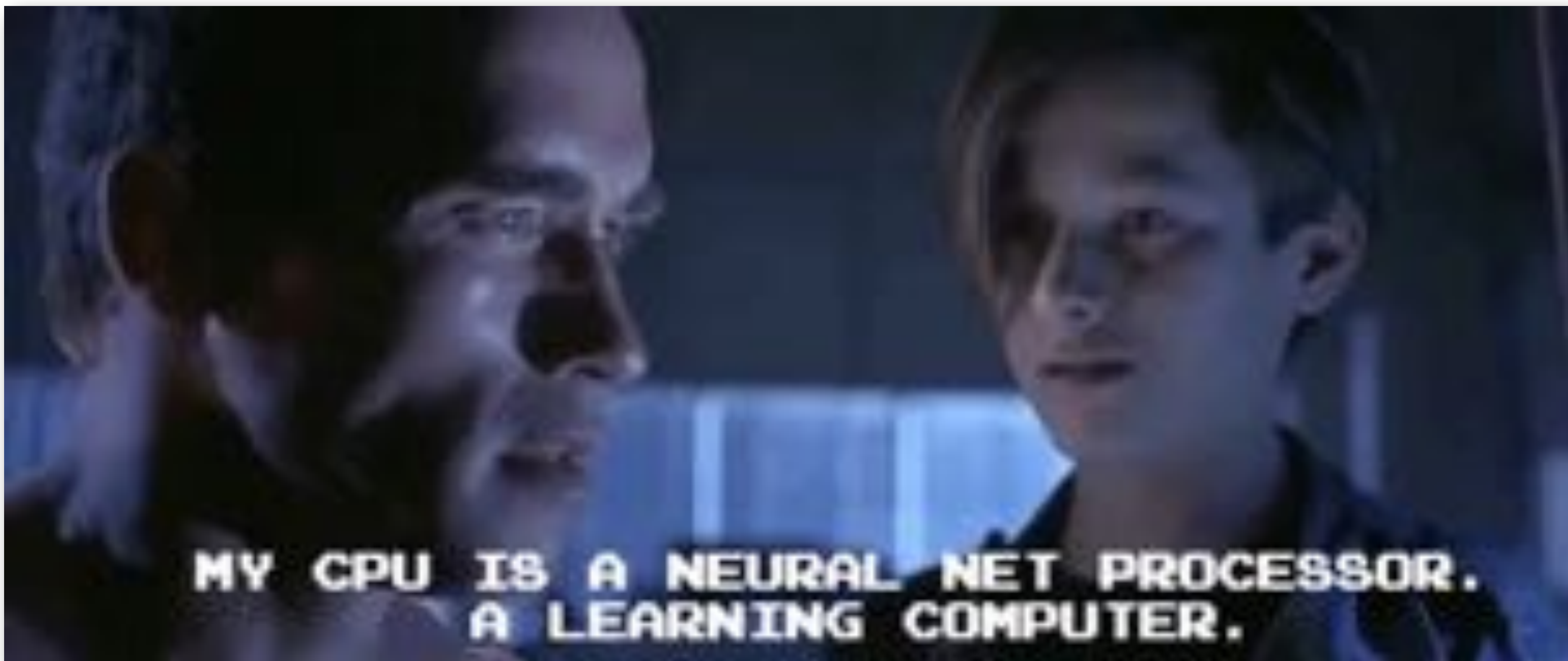
Discussion 2

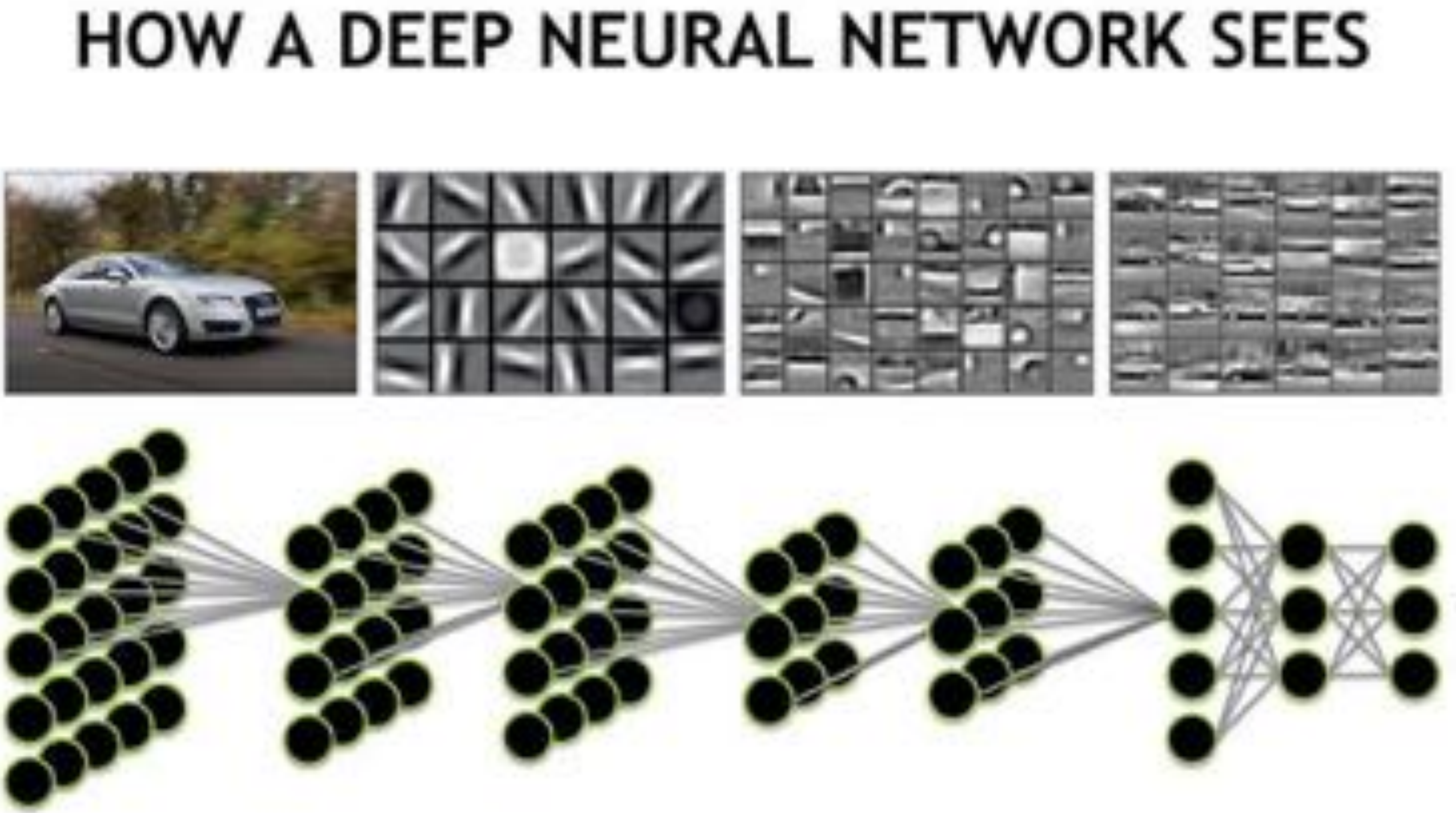
A high-level view of deep learning.

Preview

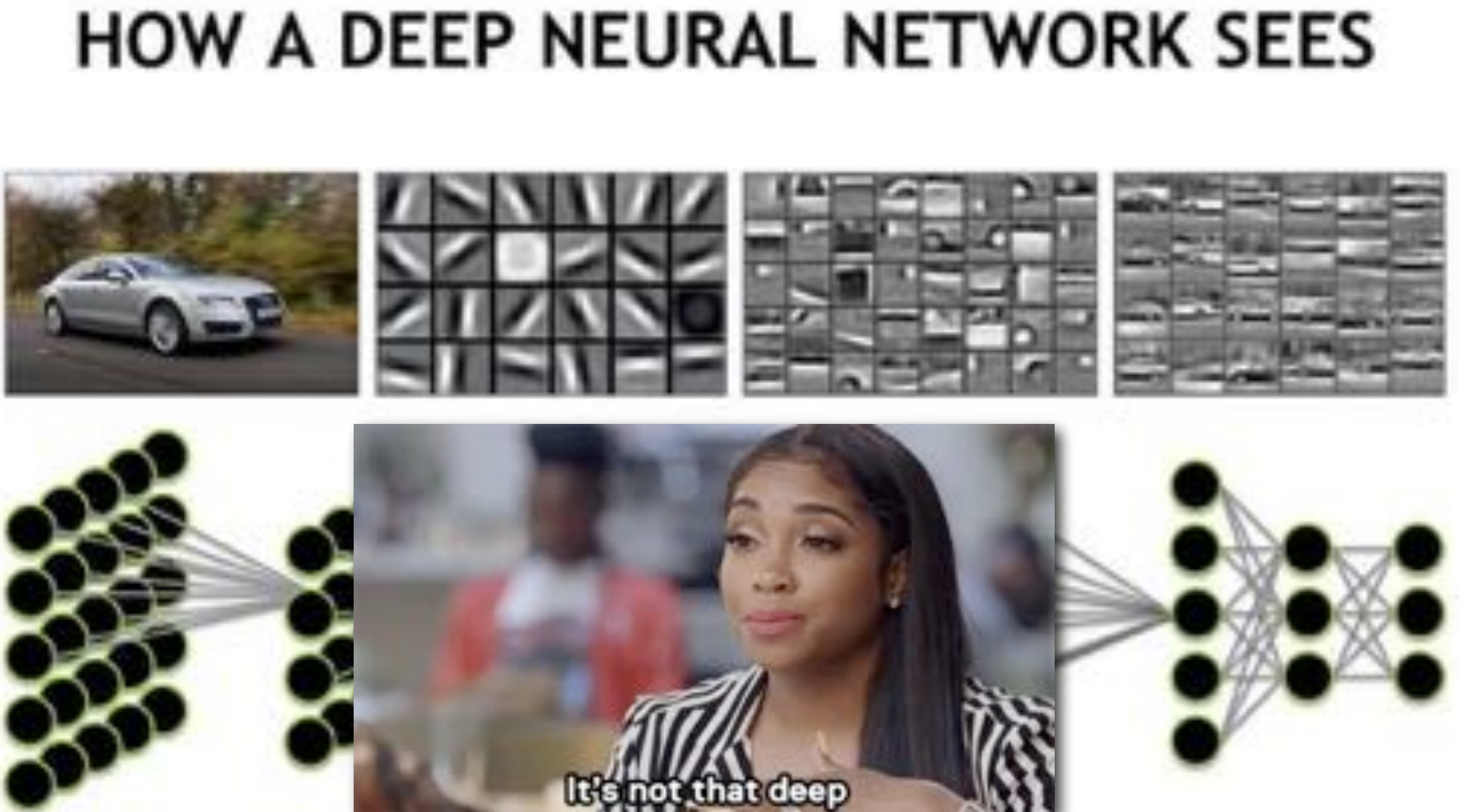
- How Deep Nets see
- Main components of a convolutional neural network (CNN)
- Walking through a CNN
- A zoo of CNN architectures
- A zoo of software
- Applications in Astro
- Broader outlook

<u>Neural Networks:</u>
How do they work?



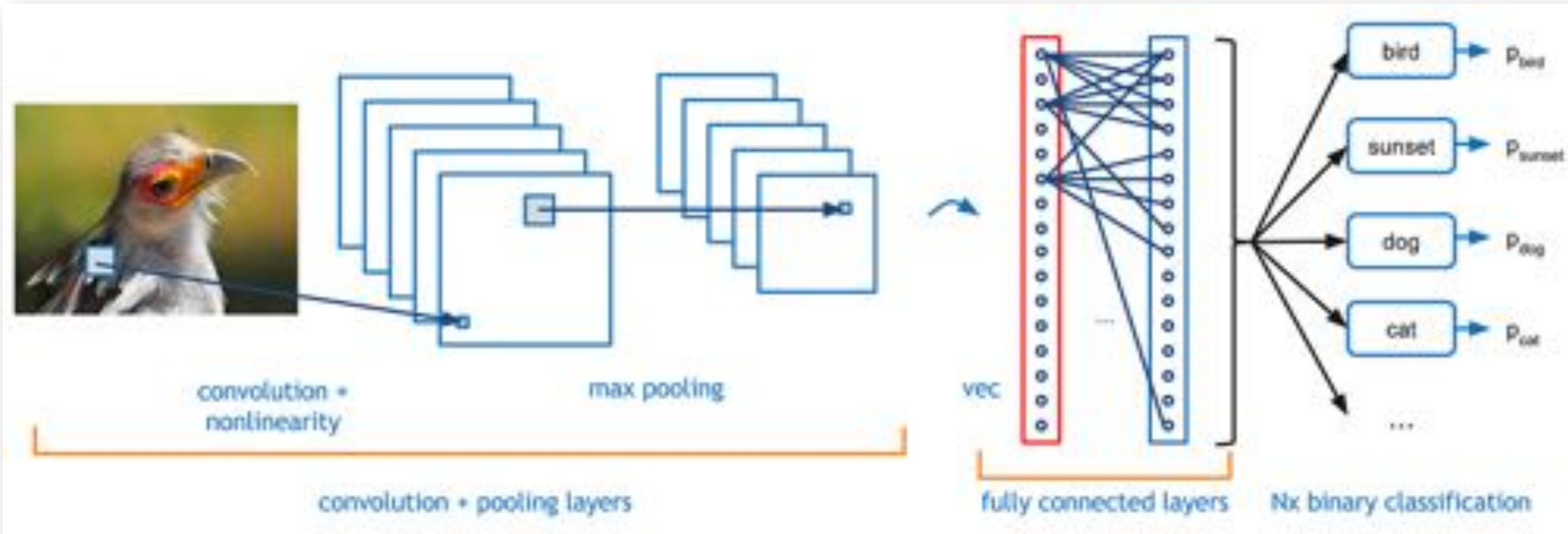


more layers = increasing depth



more layers = increasing depth

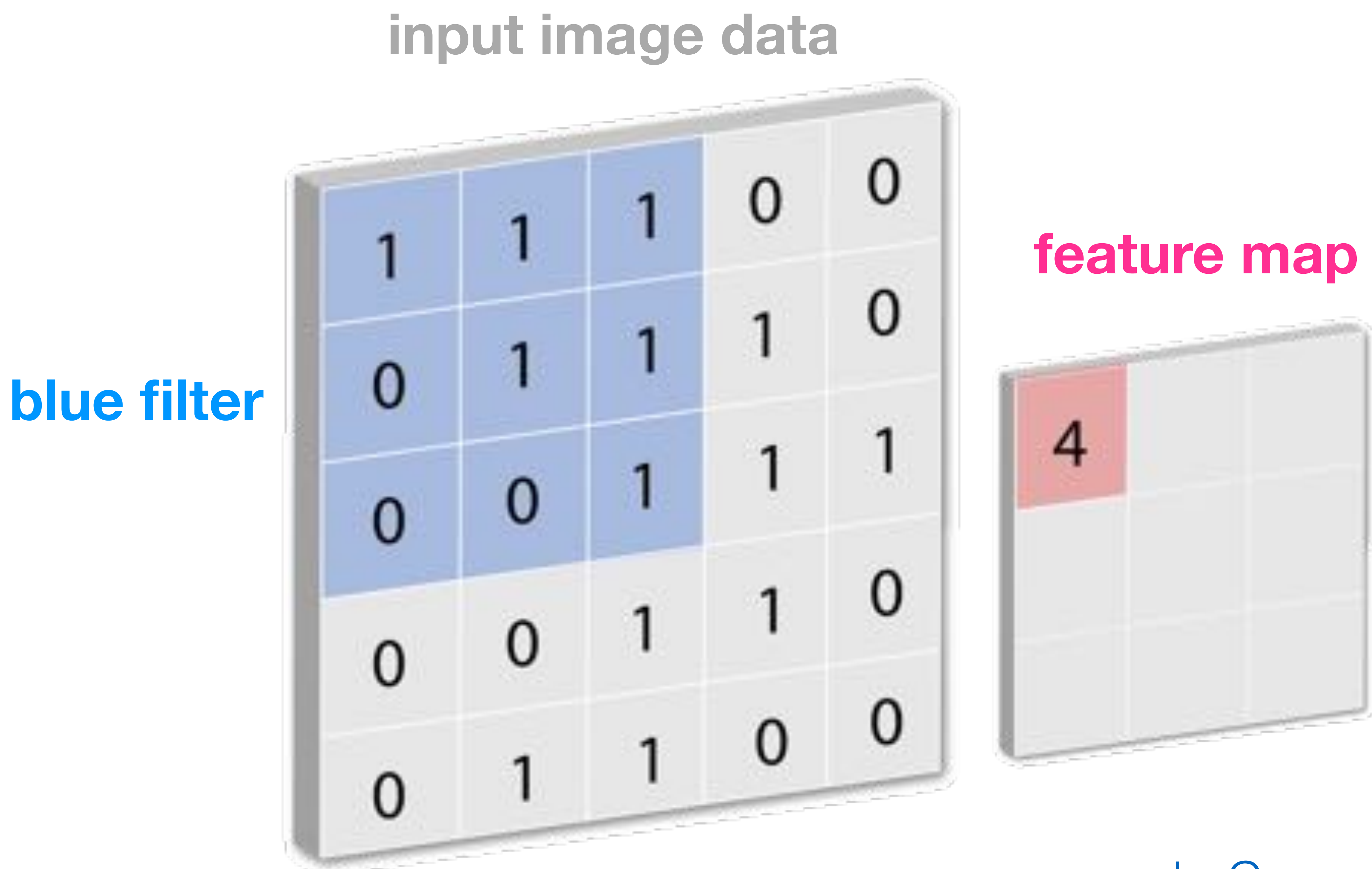
Convolutional Neural Network: It's made of layers.



- **Convolutional**: sharpens some features, blurs others.
- **Activation:** highlights features (talked about in first session)
- **Pooling:** collects (zooms in on) highlighted image regions
- **Dropout:** removes neurons that might get calcified
- **Dense:** correlates features with predictions

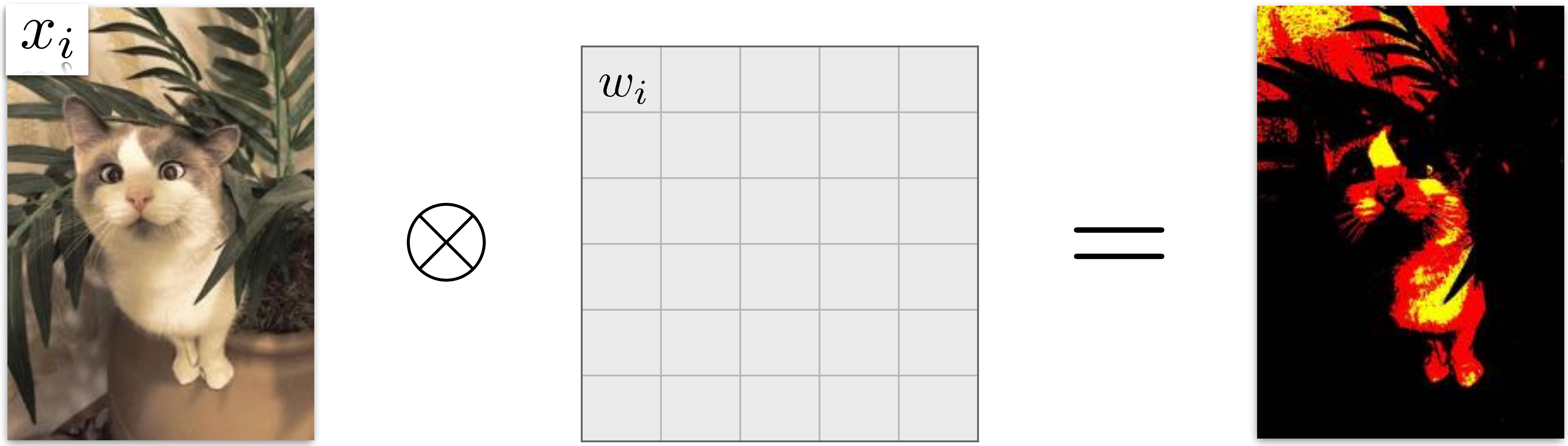
Convolutional Neural Network: weight parameters

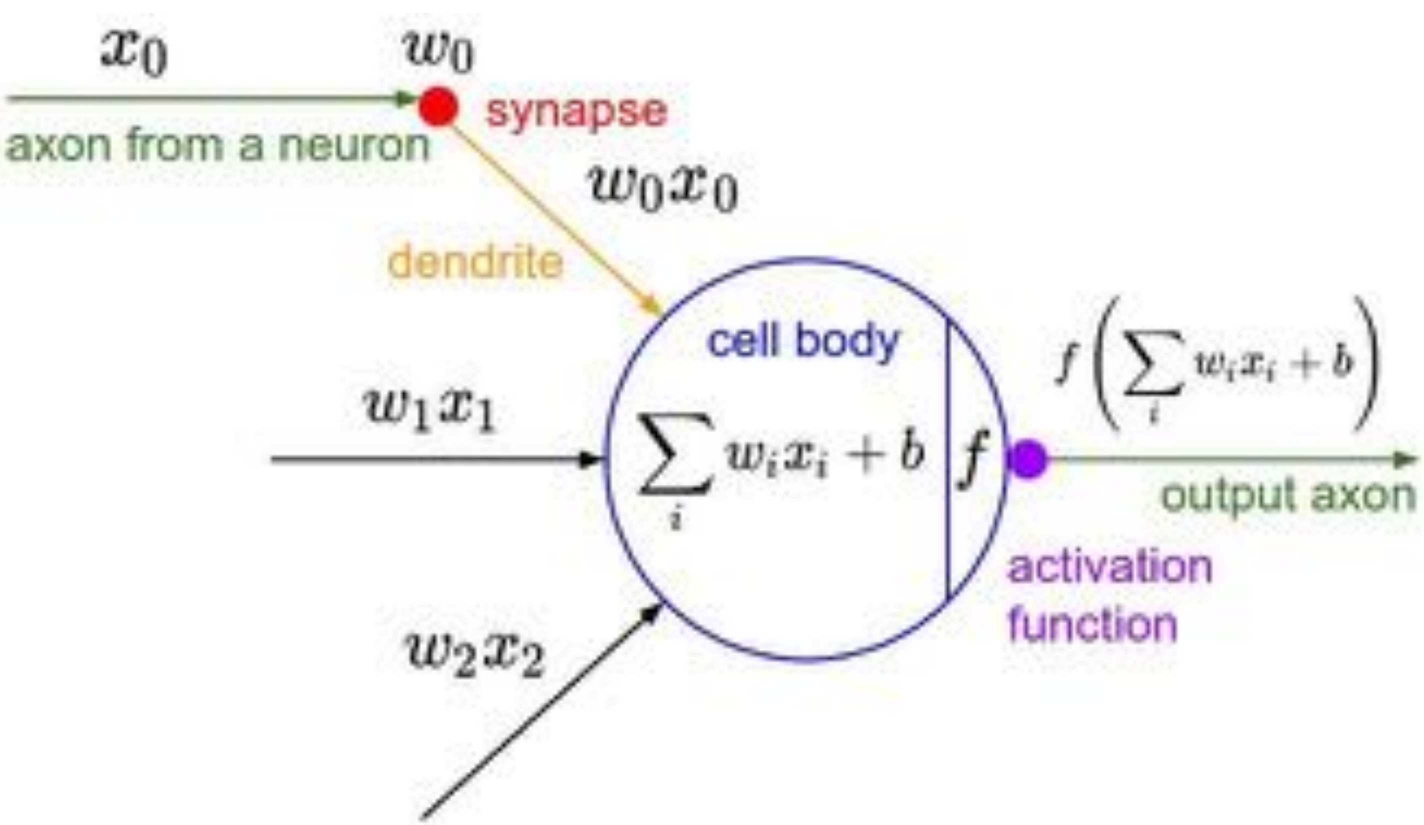
- Each pixel in the **blue filter** is one parameter in the network model
- The resulting **feature map** is the result of the convolution.



LeCun + Bengio, 1995

Convolutional Neural Network: Convolution

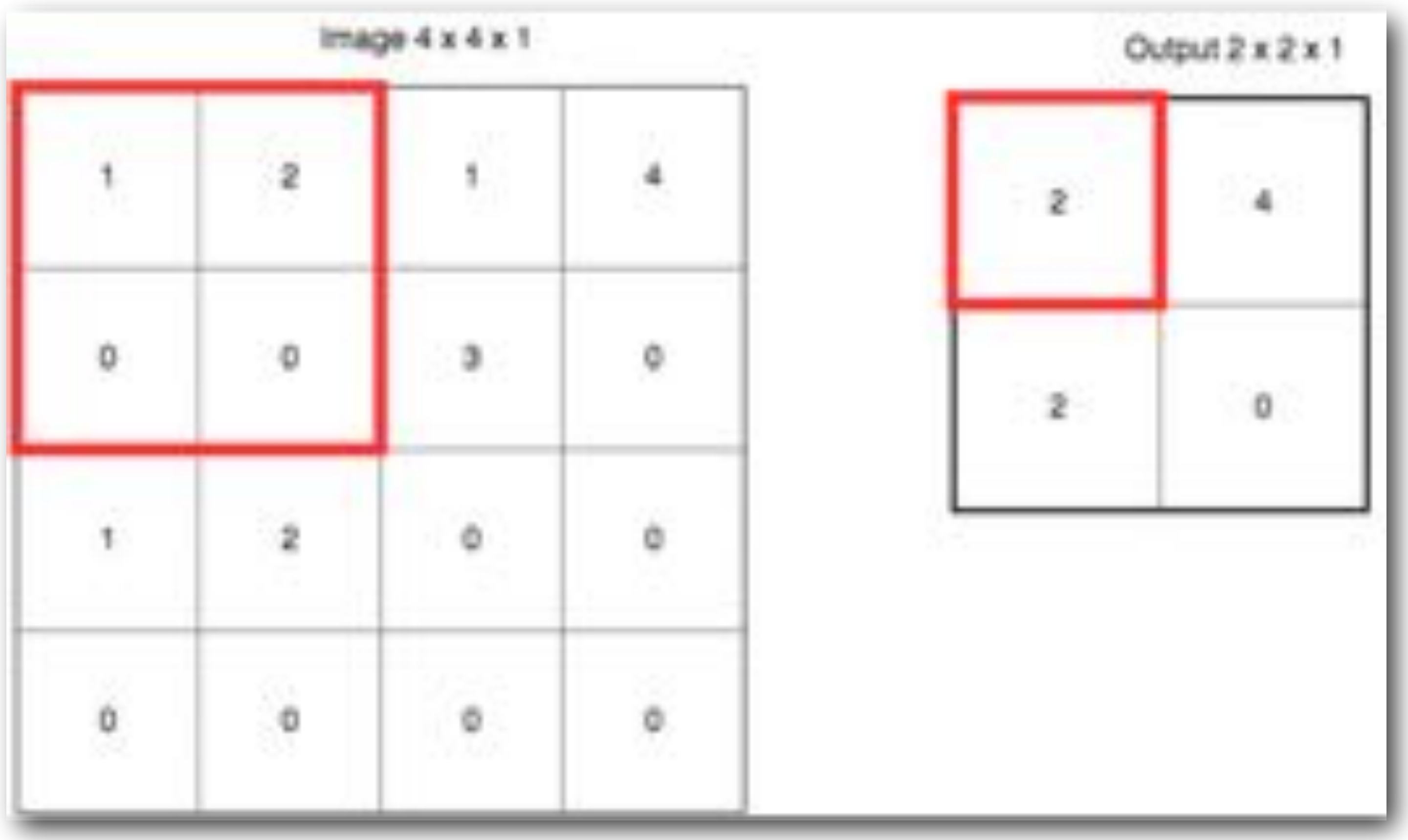




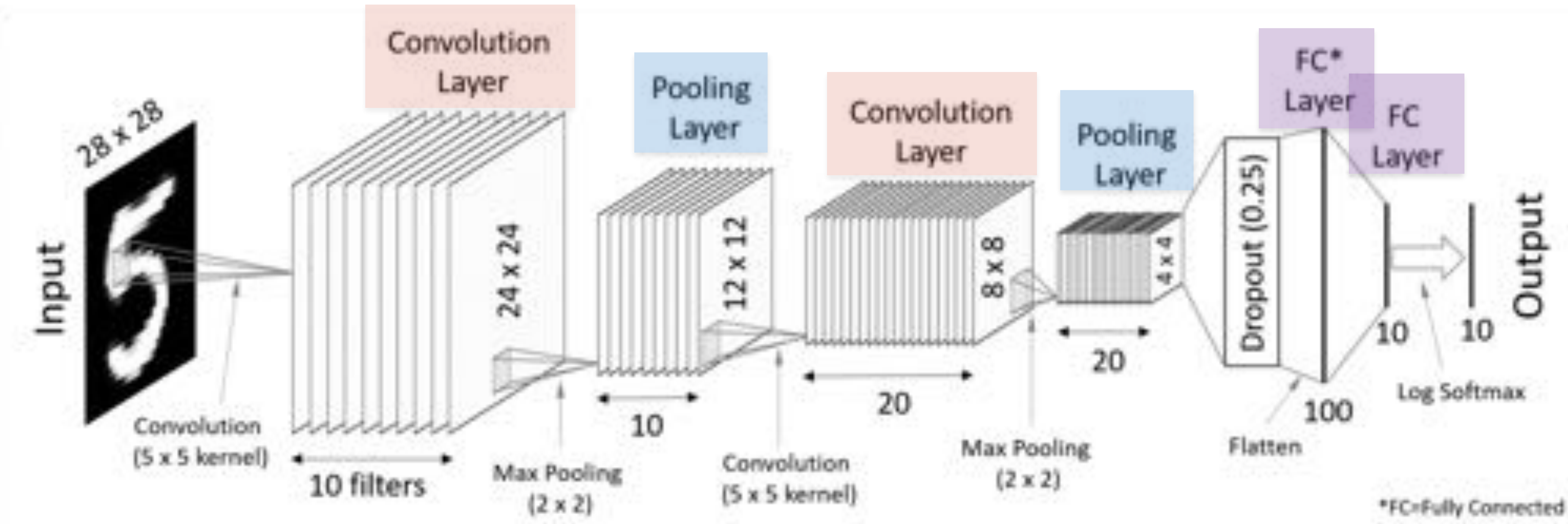
- <u>Each computational neuron</u> is an image filter, where $\mathbf{w_i}$ is the value of a pixel in that filter and a model parameter
- <u>During *convolution and activation*</u>, the model acts on the input image, highlighting features, such as edges or circles.

Pooling

- Reduces the size of the image very quickly.
- Preserves information in the image.
- Doesn't add parameters to the network
- Paper: [Scherer + 2010]



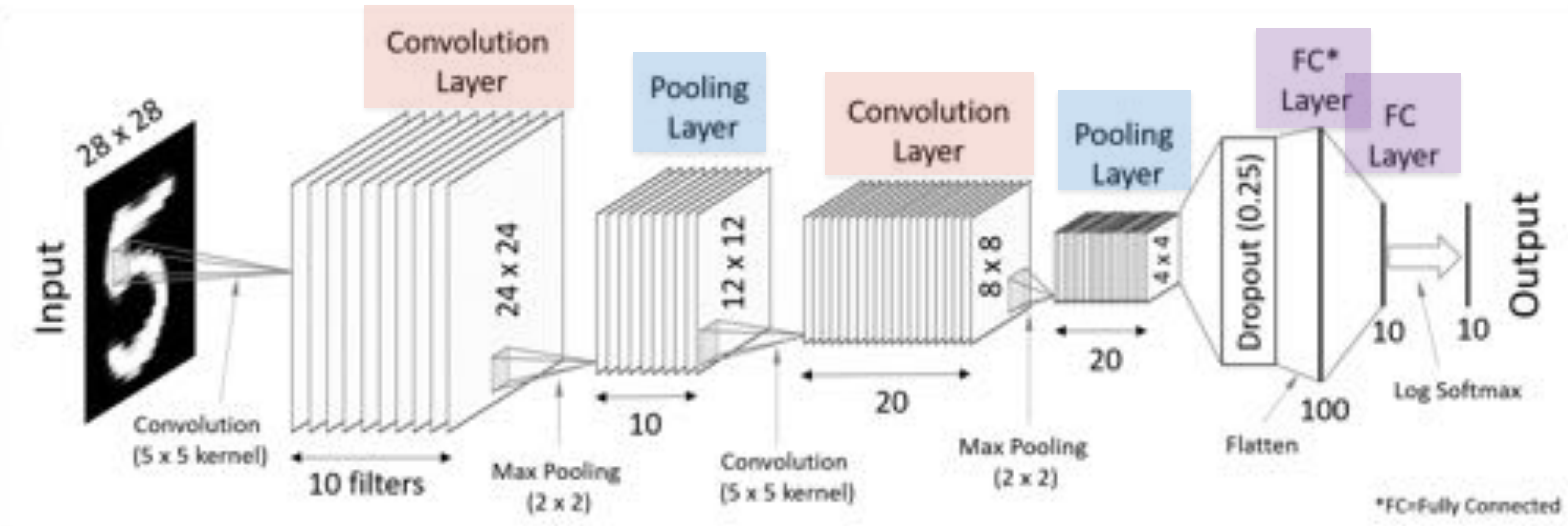
CNN Architecture structure



Learns the representation features

Associates features with labels

CNN Architecture structure



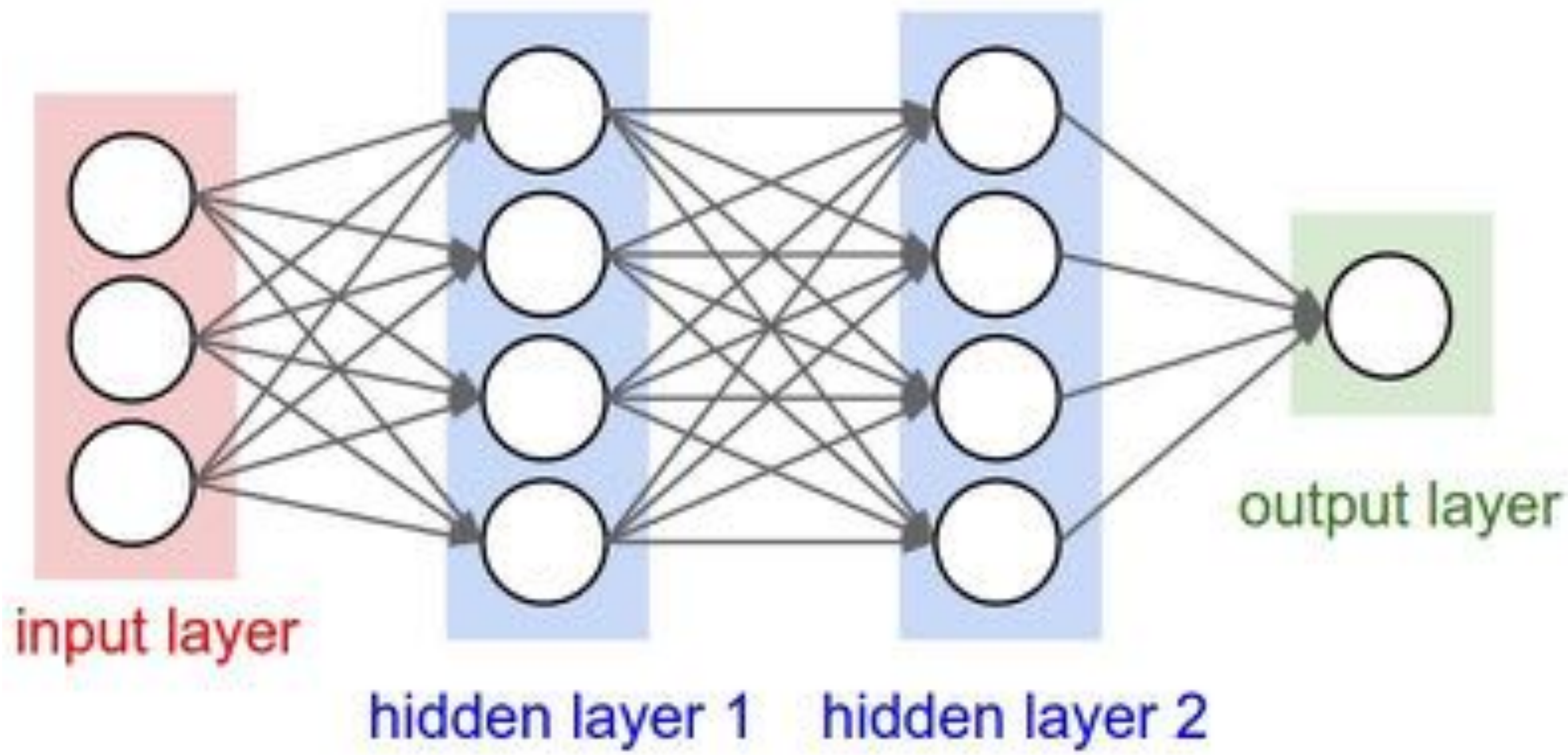
Learns the representation features

Associates features with labels

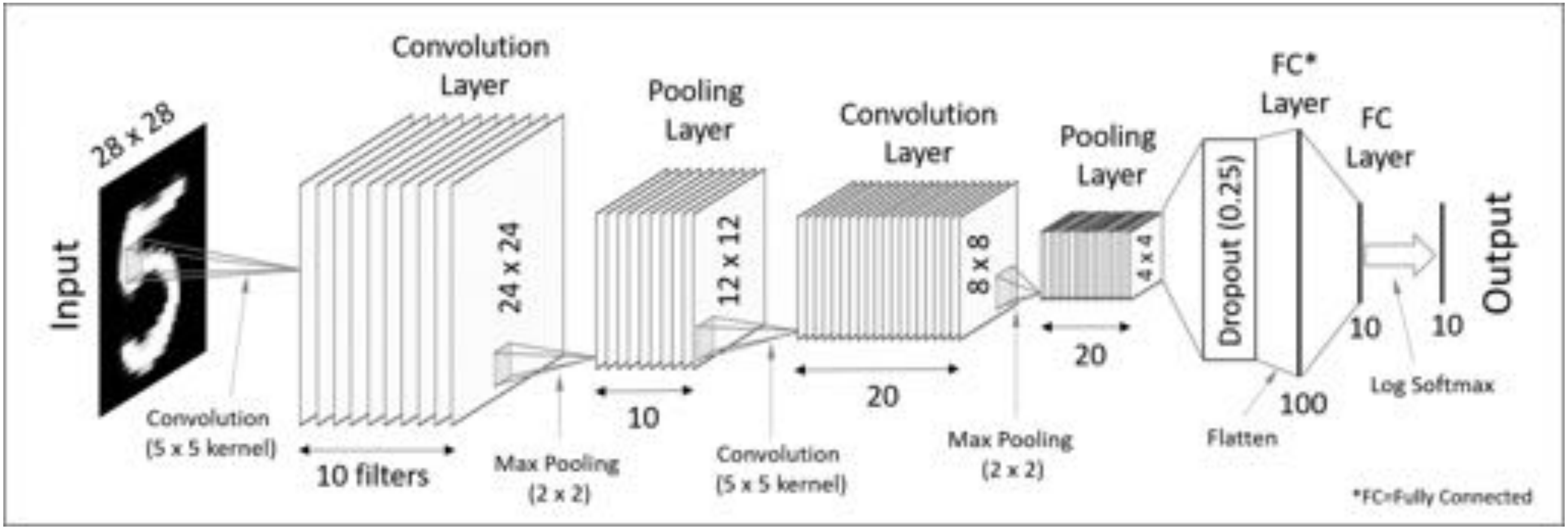
- As the image progresses through the layers, **image features are learned**
  - Through *highlighting* by **activations in convolutions**
  - On *multiple spatial scales* by **pooling**
- Features are then *associated with truth labels* via **fully-connected layers**

A Zoo of Architectures

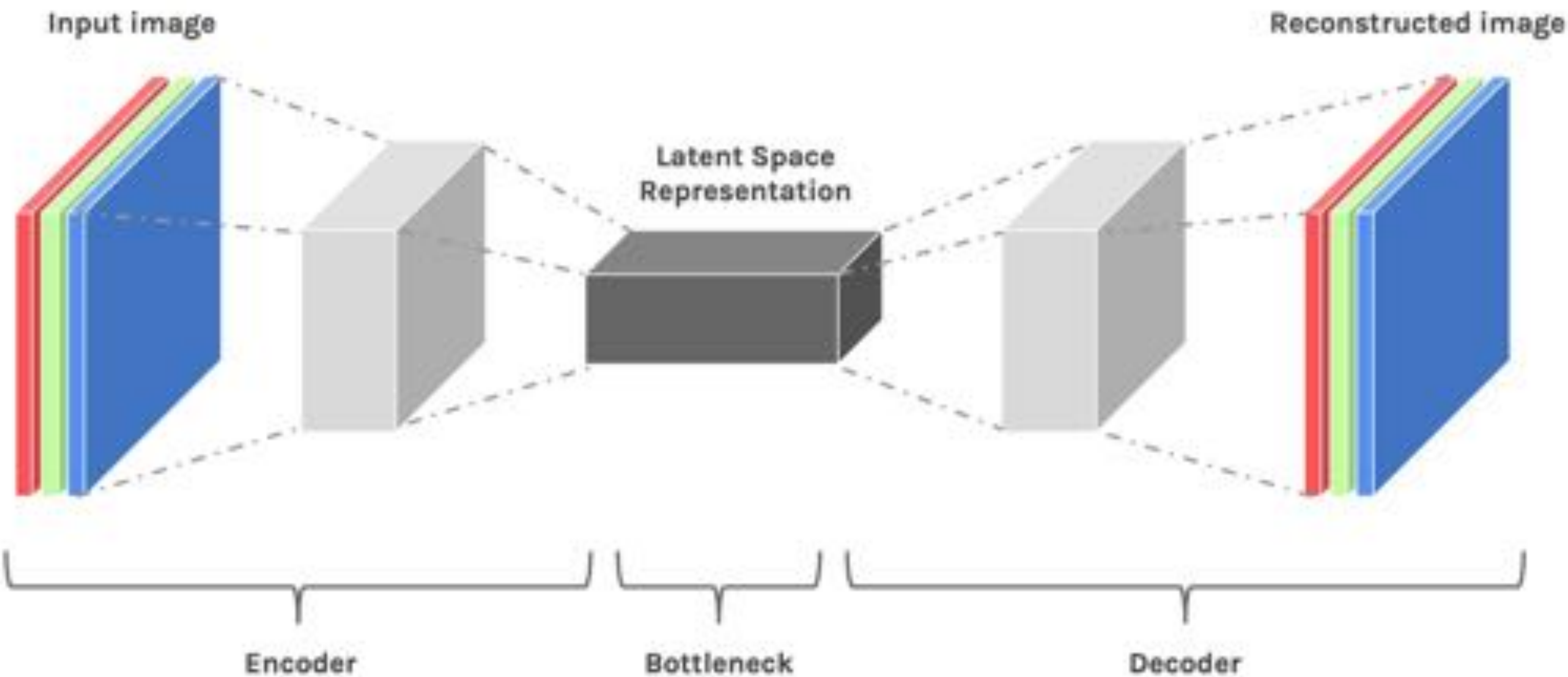
Feed-forward Fully connected Network



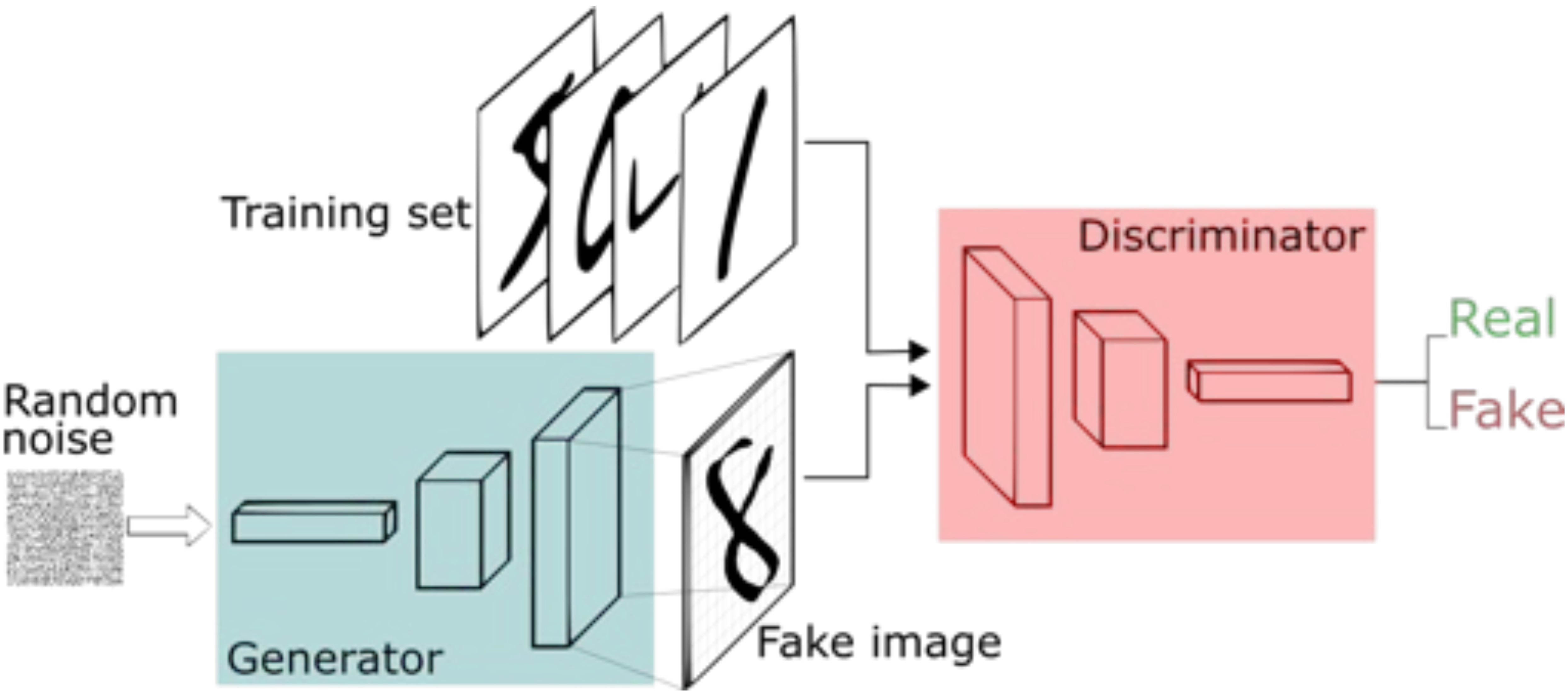
Convolutional Neural Network



AutoEncoder



Generative Adversarial Network (GAN)



- A Generative model in which two networks (Generator and Discrimator) learn simultaneously.
- GANs can create hyper-resolution images (higher res than original)

Generative Adversarial Network (GAN)

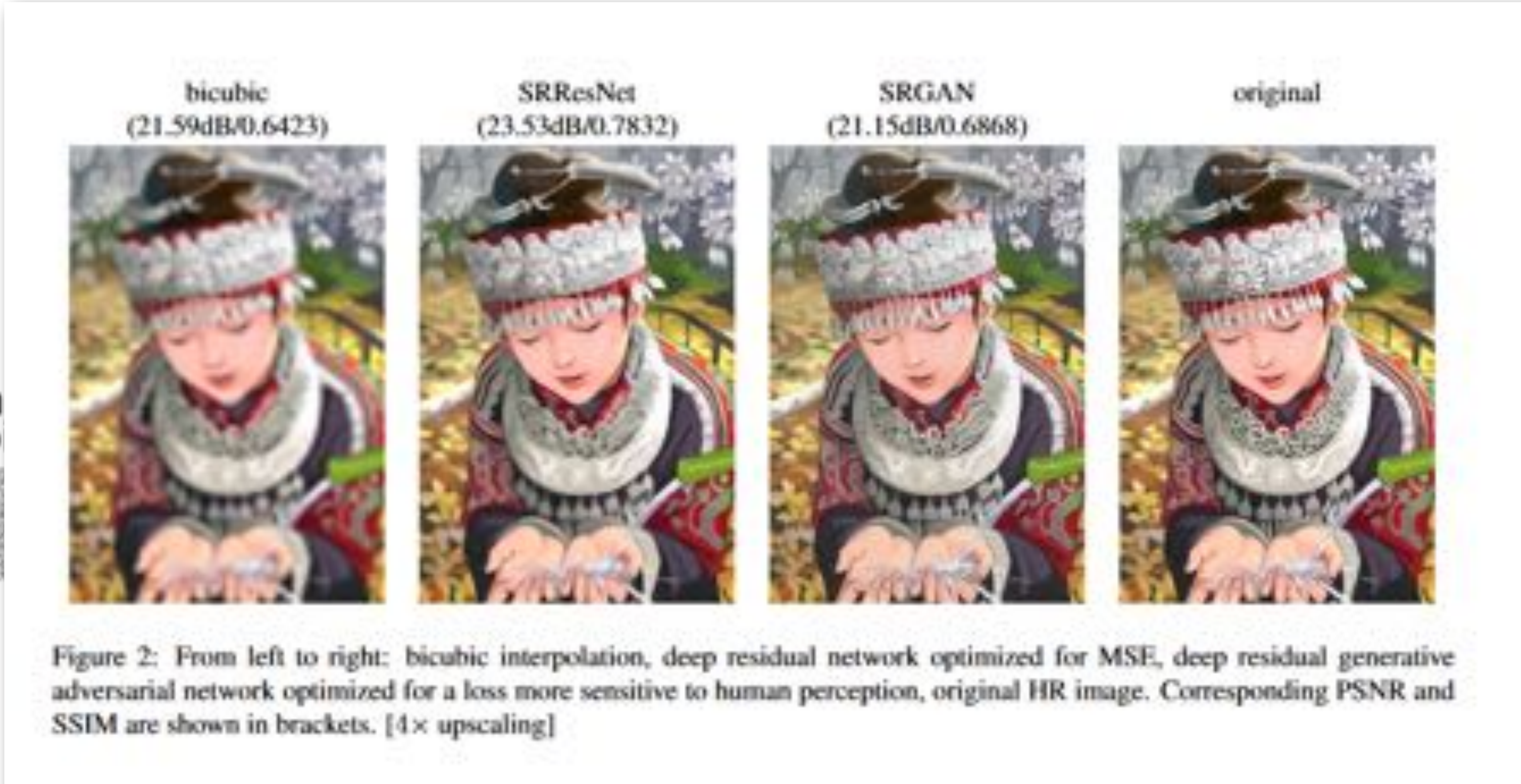
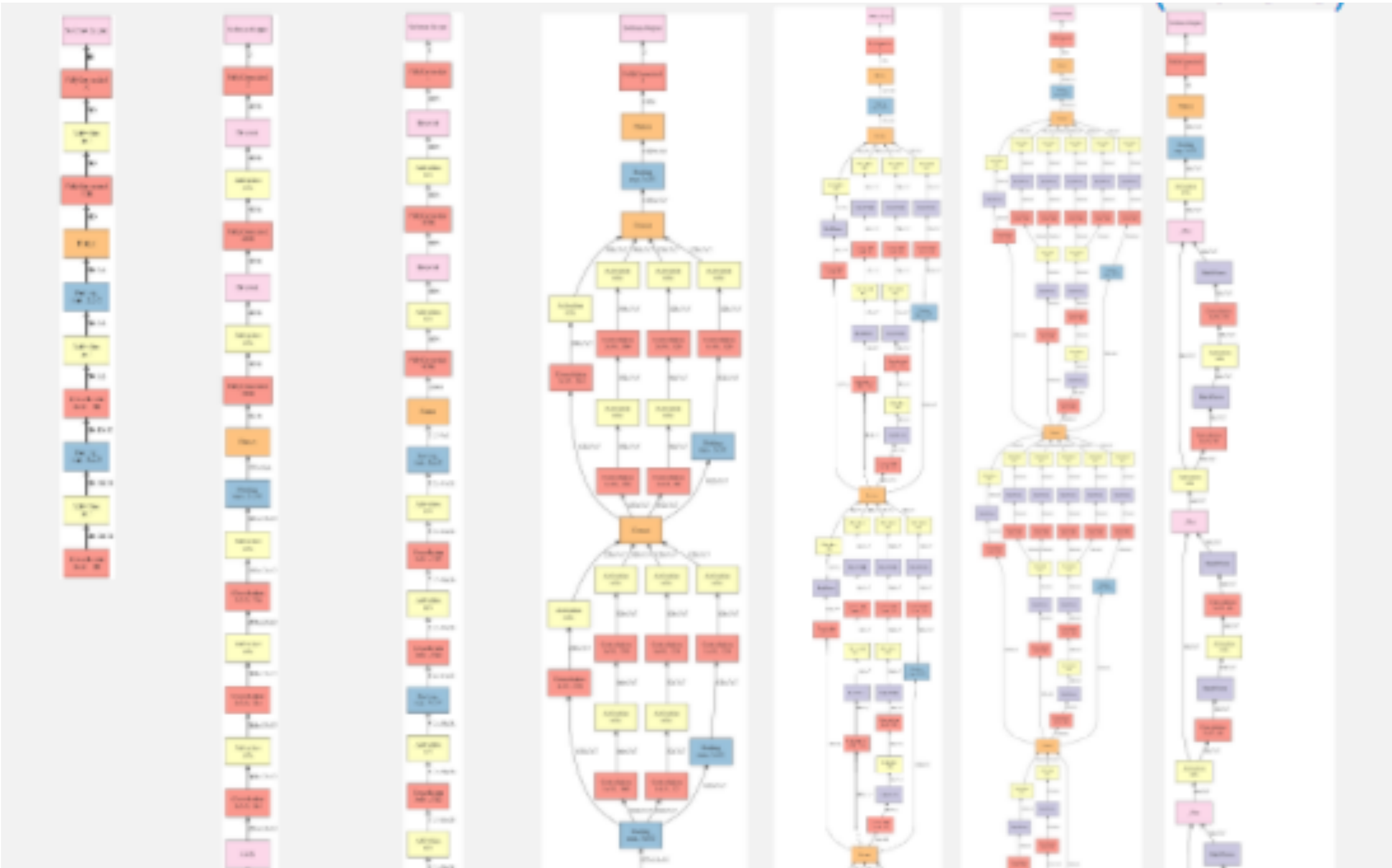


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

- A Generative model in which two networks (Generator and Discriminator) learn simultaneously.
- GANs can create hyper-resolution images (higher res than original)

Evolution of networks

LeNet | AlexNet | VGG | GoogLeNet | Inception BN | Inception V3 | ResNet



Evolution of networks

What happens when Data comes alive?

Ready Player One?

- <u>Software</u>
  - **Scikit Learn**
  - XGBoost
  - **Keras/ TensorFlow**
  - Theano (no longer developed)
  - Lasagne
  - **PyTorch**
  - Caffe
- <u>How to choose!</u>
  - Depends on performance requirements, ease of deployment, familiarity, rest of software stack

Deep Nets meet the Deep Sky:

Is it full of stars or full of …?

Applications to Astrophysics

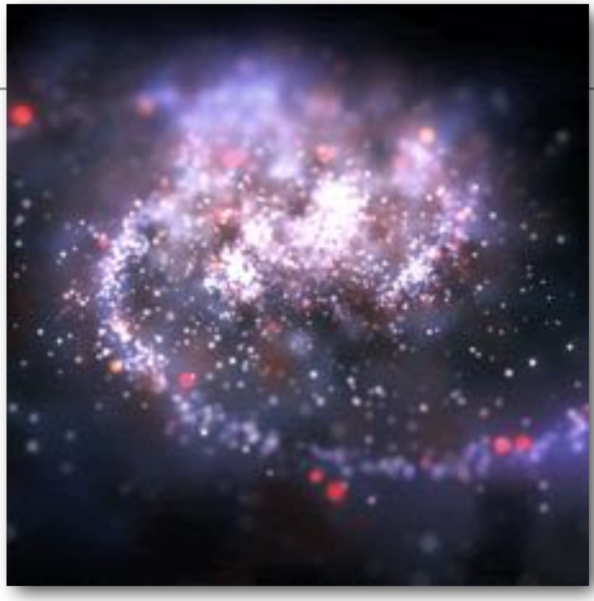
Finding and measuring

Simulations

Observing

Applications to Astrophysics

Finding and measuring

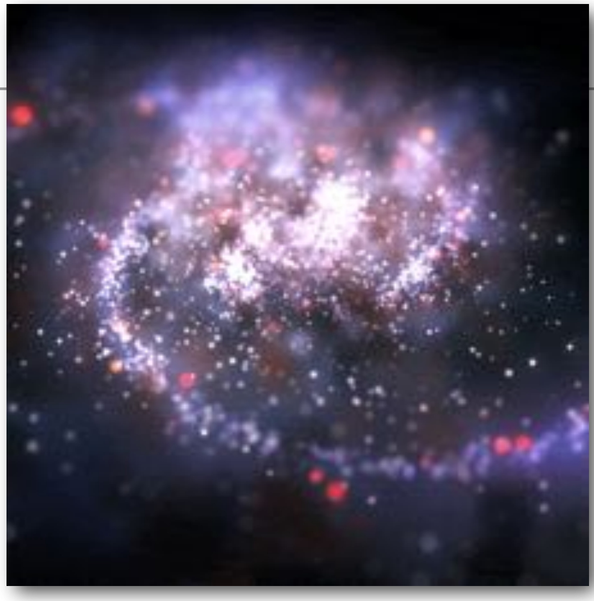


Simulations

Observing

Applications to Astrophysics

Finding and measuring

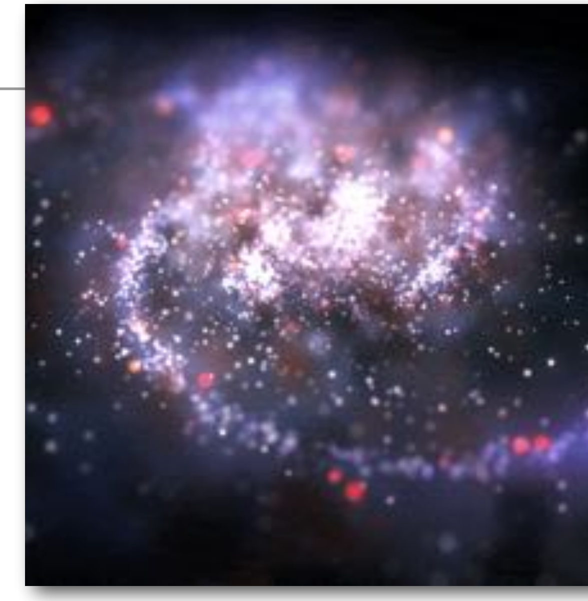


Simulations

Observing

Applications to Astrophysics

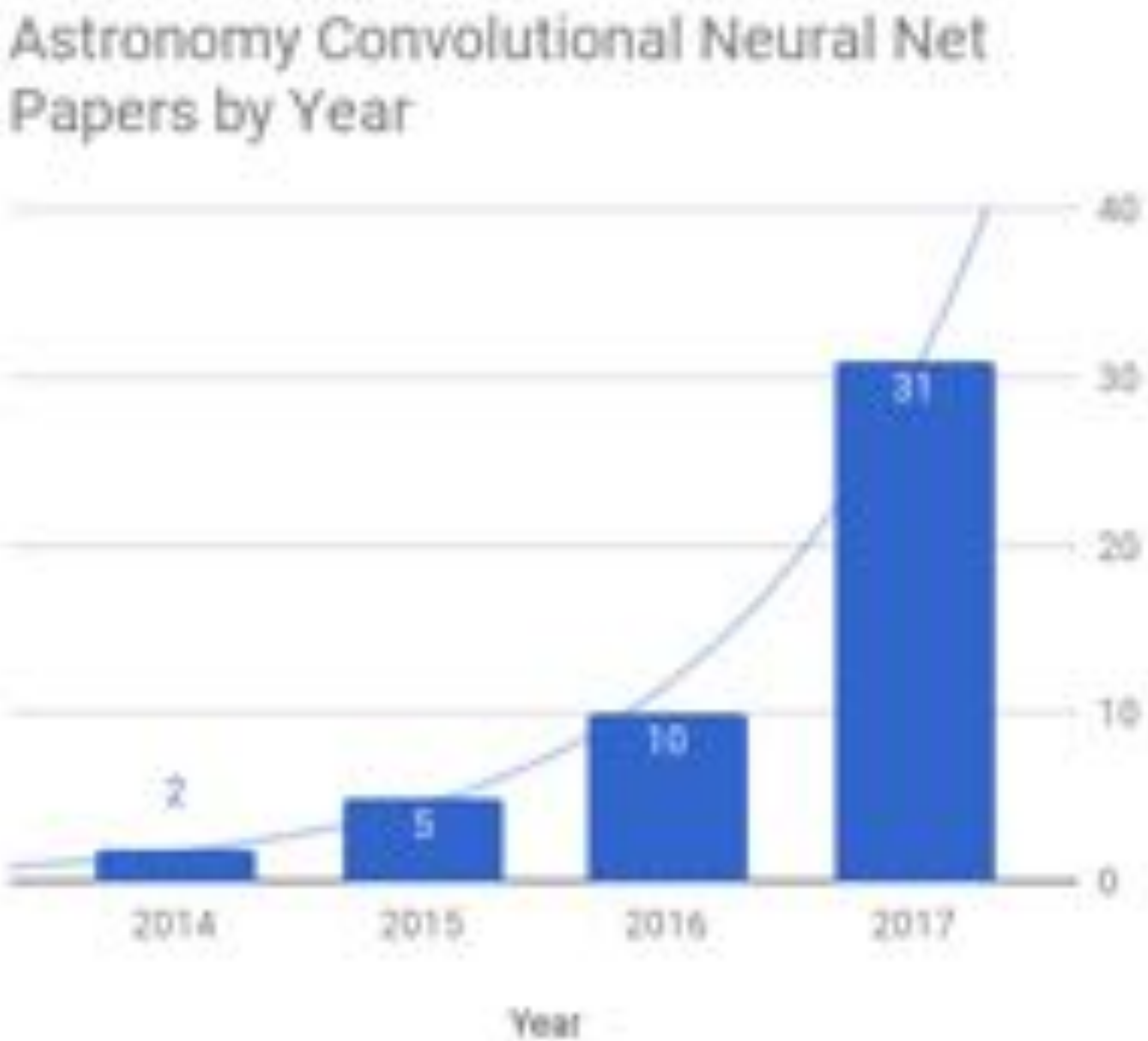
Finding and measuring



Simulations

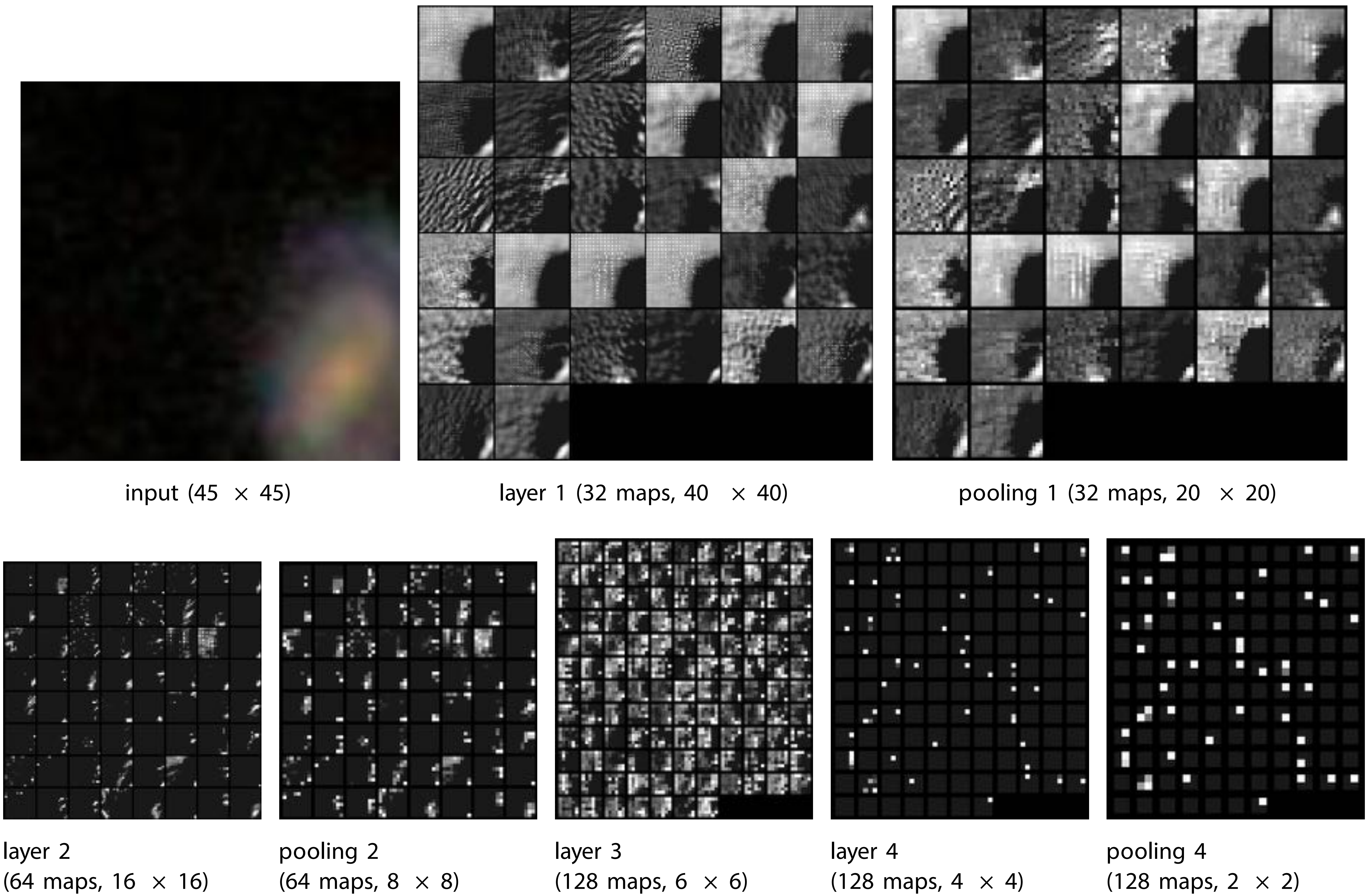
Observing

Publications: Deep Learning in Astro

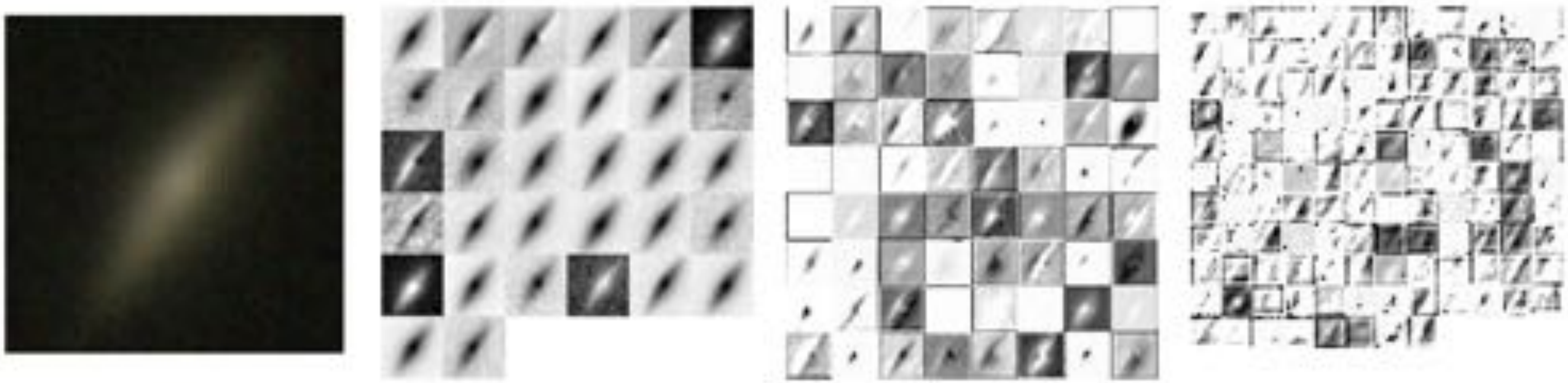


2 DIMENSIONS

Galaxy Morphology classification (Dieleman+2015)

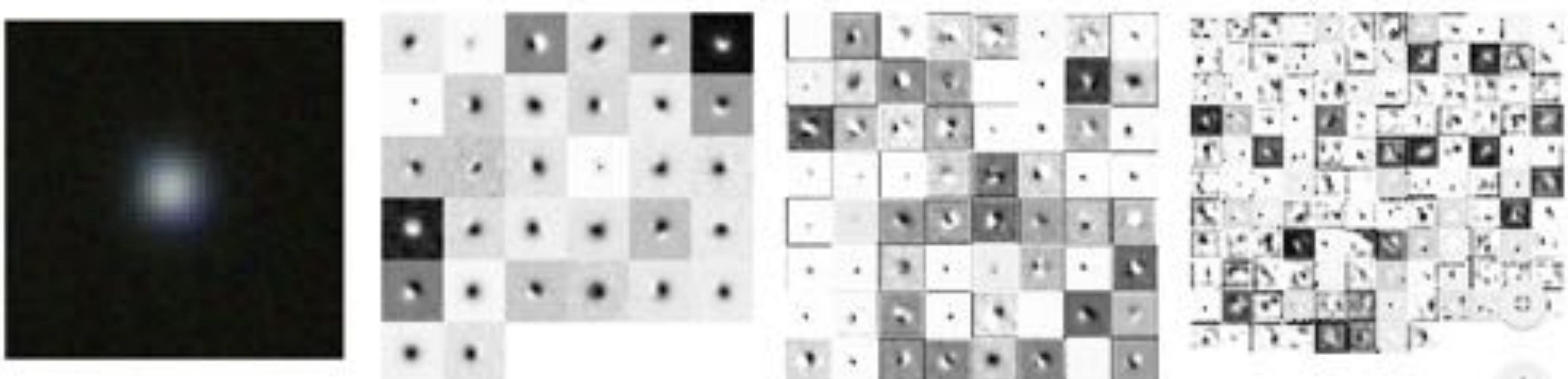


Star-galaxy classification (Kim+Brunner 2016)



(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)

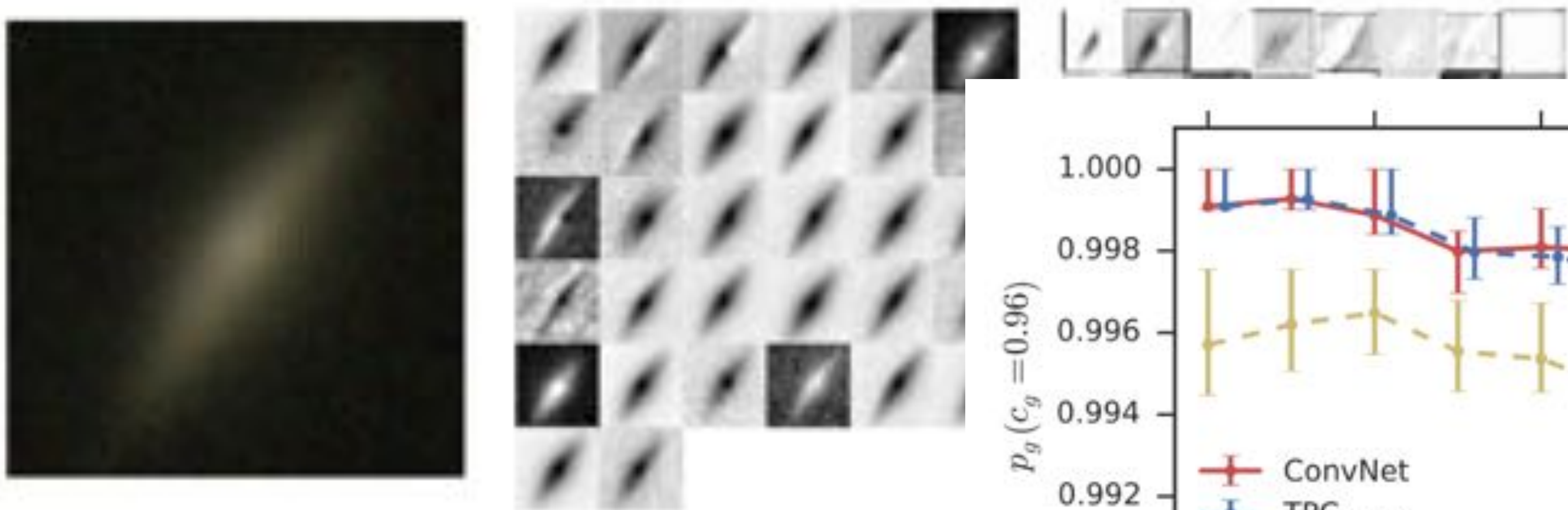
Figure 12. Similar to Figure 6 but for a galaxy in the SDSS data set.



(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)

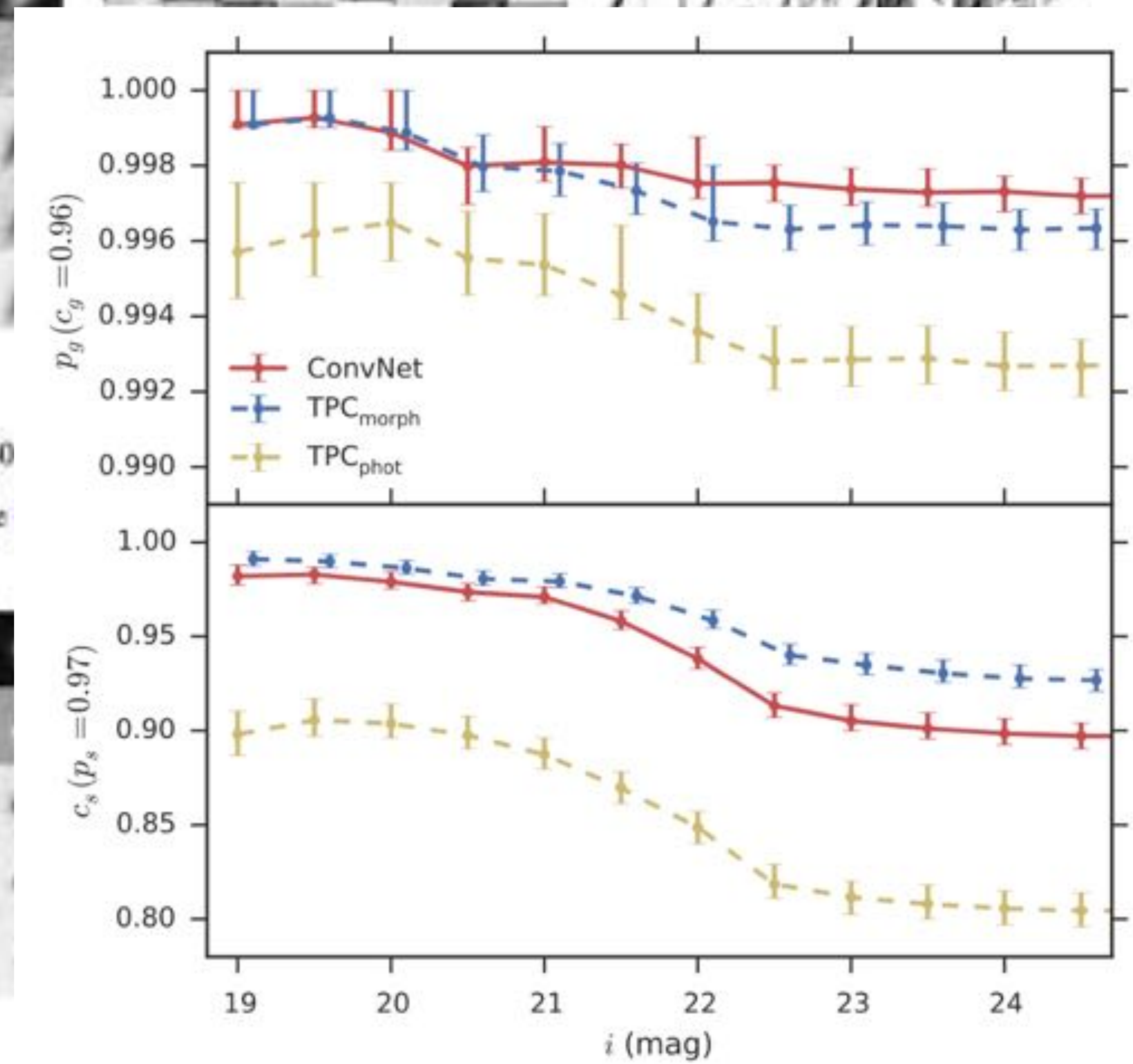
Figure 13. Similar to Figure 12 but for a star in the SDSS data set.

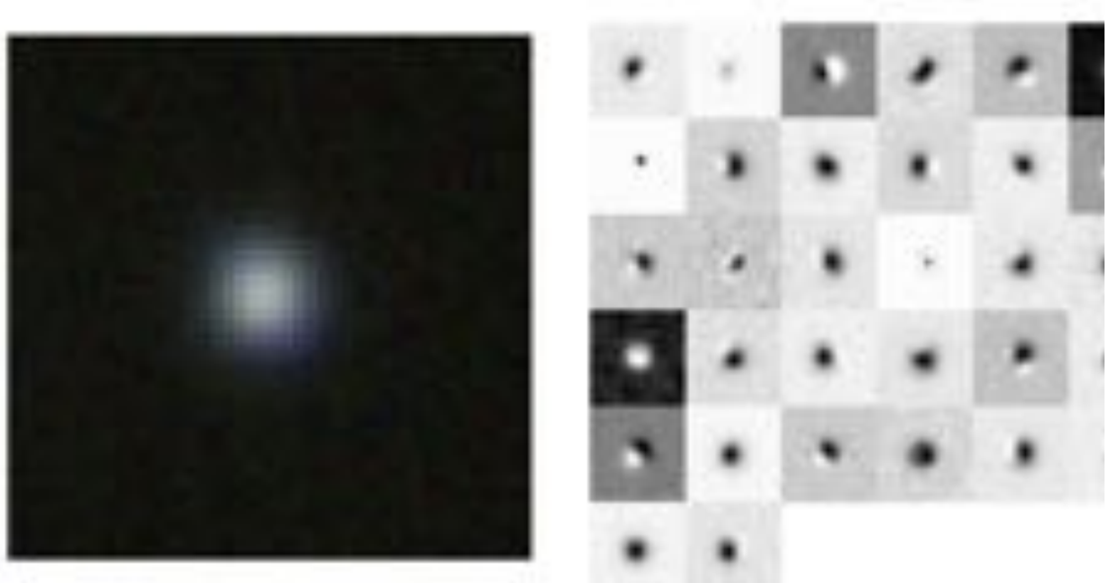
Star-galaxy classification (Kim+Brunner 2016)



(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40

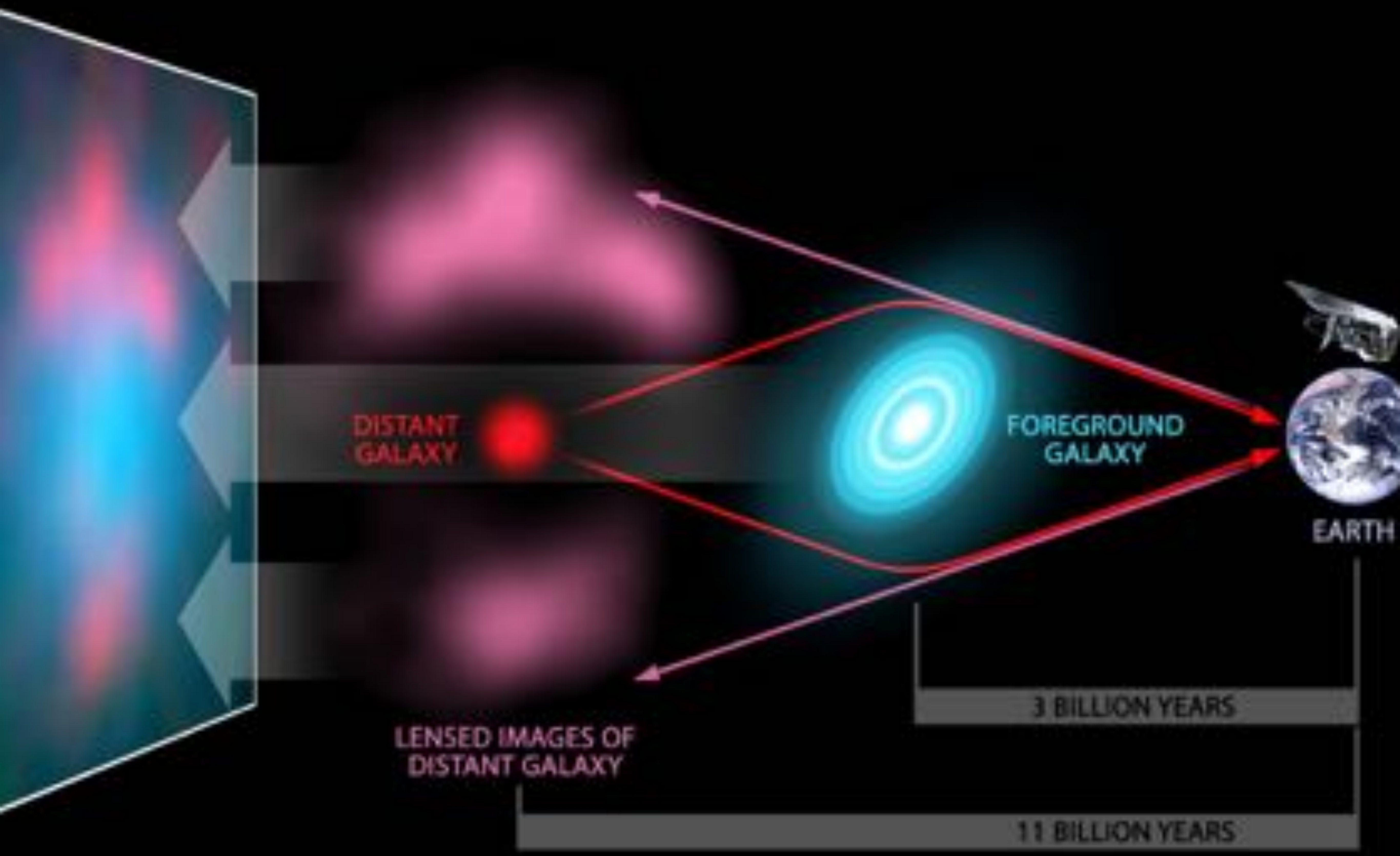
Figure 12. Similar to Figure



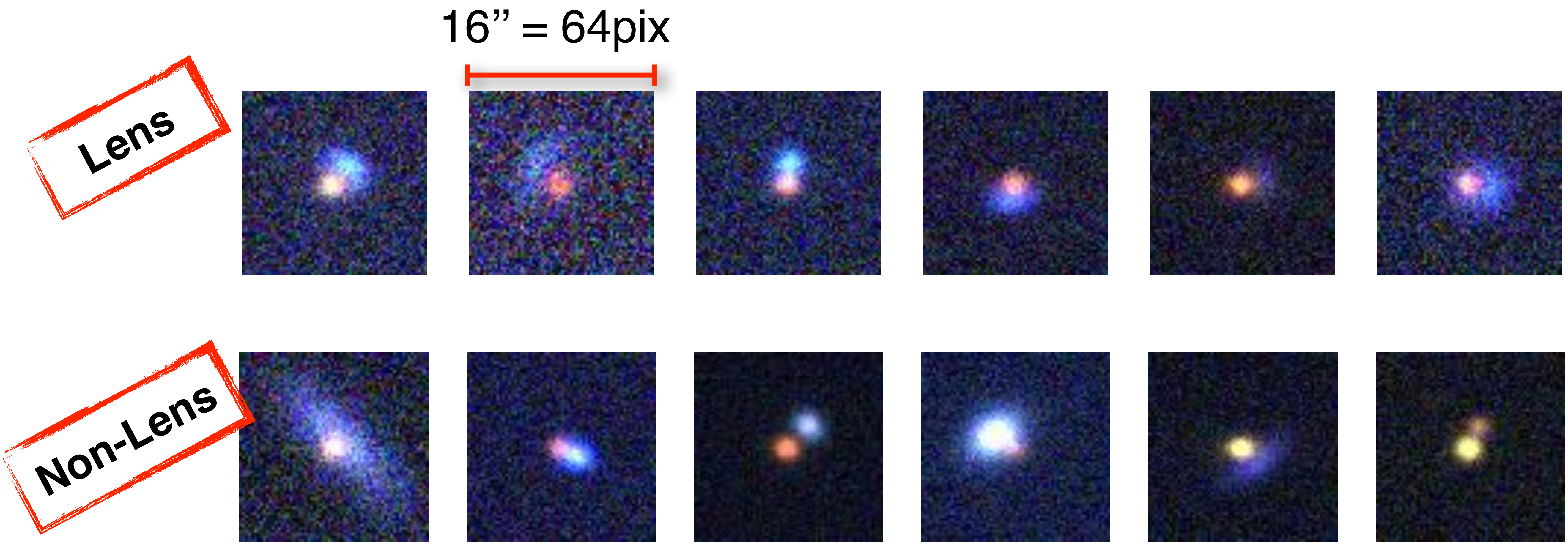


(a) Input (5 bands×44×44) (b) Layer 1 (32 maps×40×40) (c) Layer 3 (64 maps×20×20) (d) Layer 6 (128 maps×10×10)

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


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

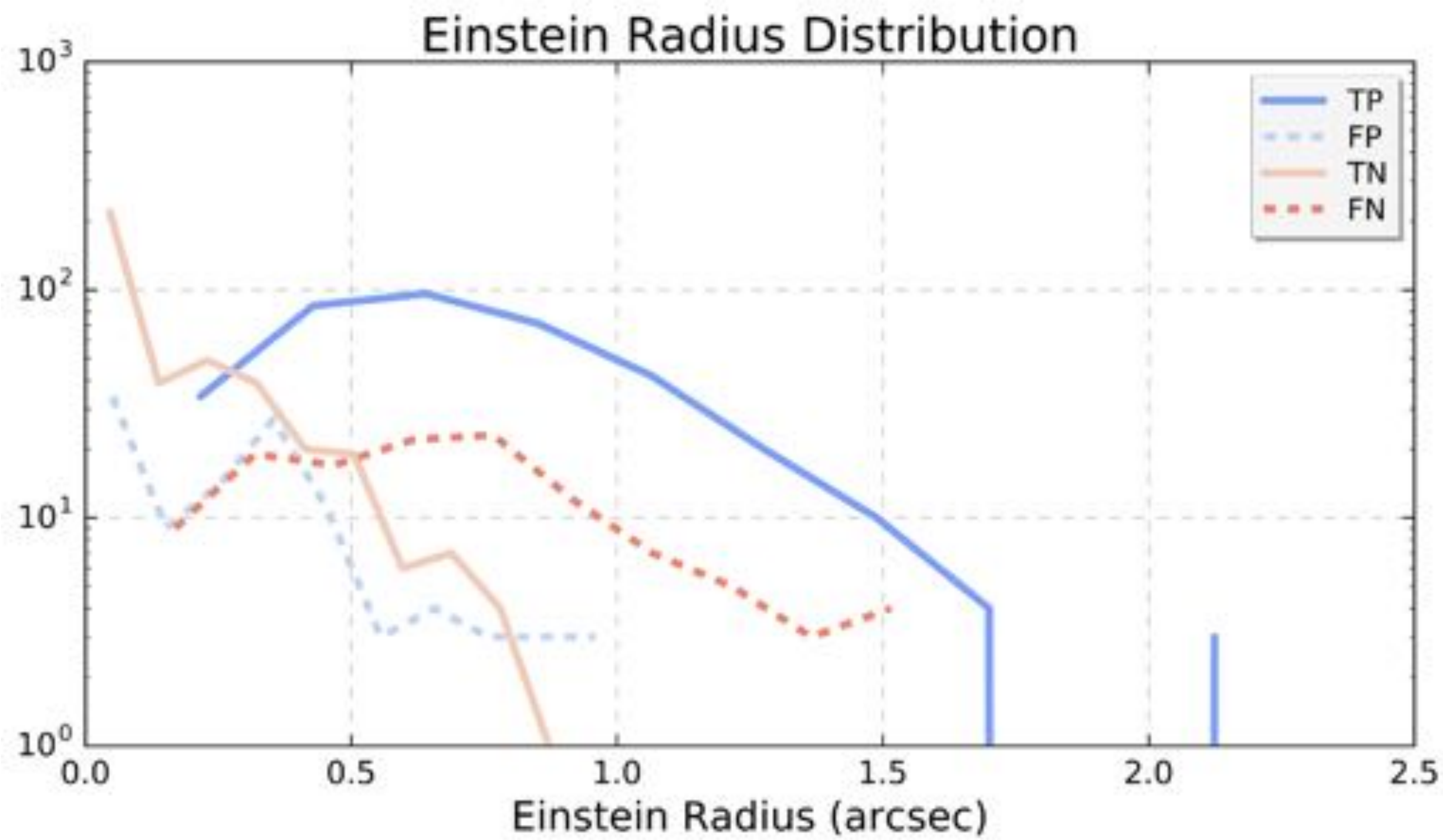
Deep Lensing: Lens Classification

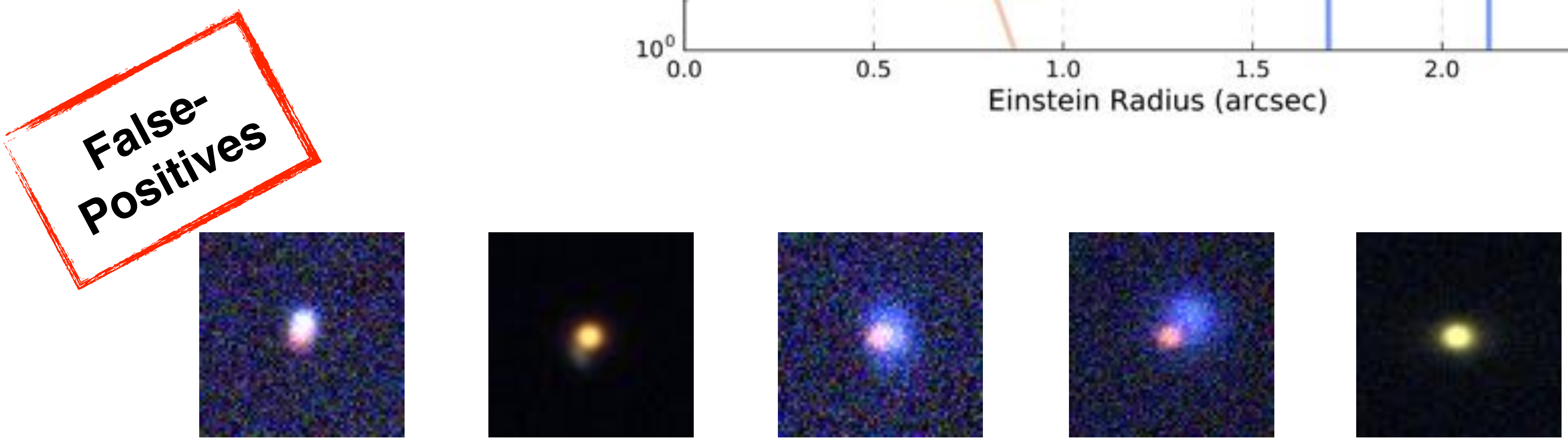


- <u>Simulations for Training Set</u>
  - **Training** 15K objects; 50 epochs
  - **Empirically motivated** density and light profiles of sources and lenses
  - **Mimic DES Survey characteristics:** noise levels, exposure time, PSF, photometry, resolution

Diagnostics:
Einstein Radius

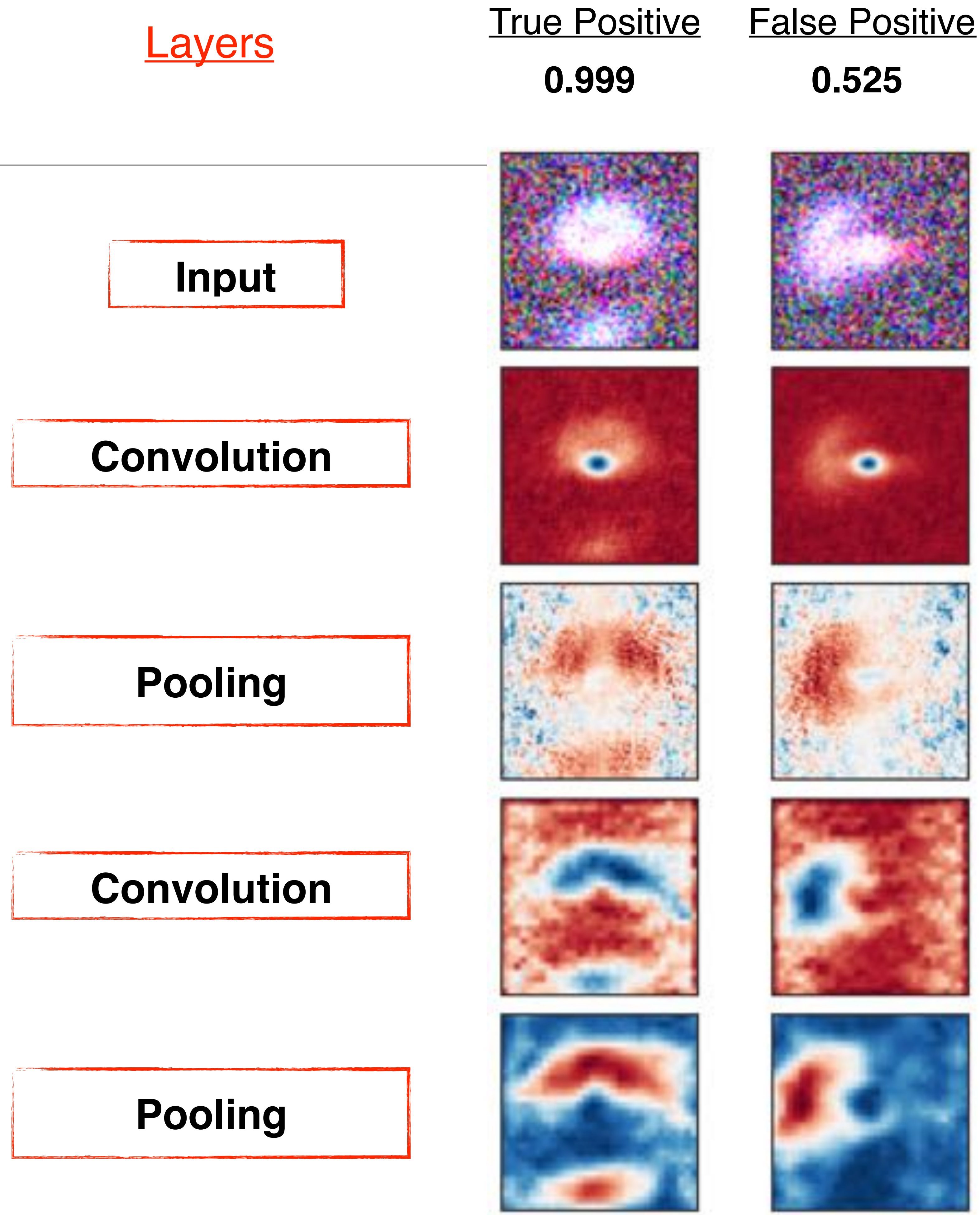
- <u>False-identification rates are higher at small Einstein radius</u>, where there can be more confusion in discerning source image from lens.





What's Inside?

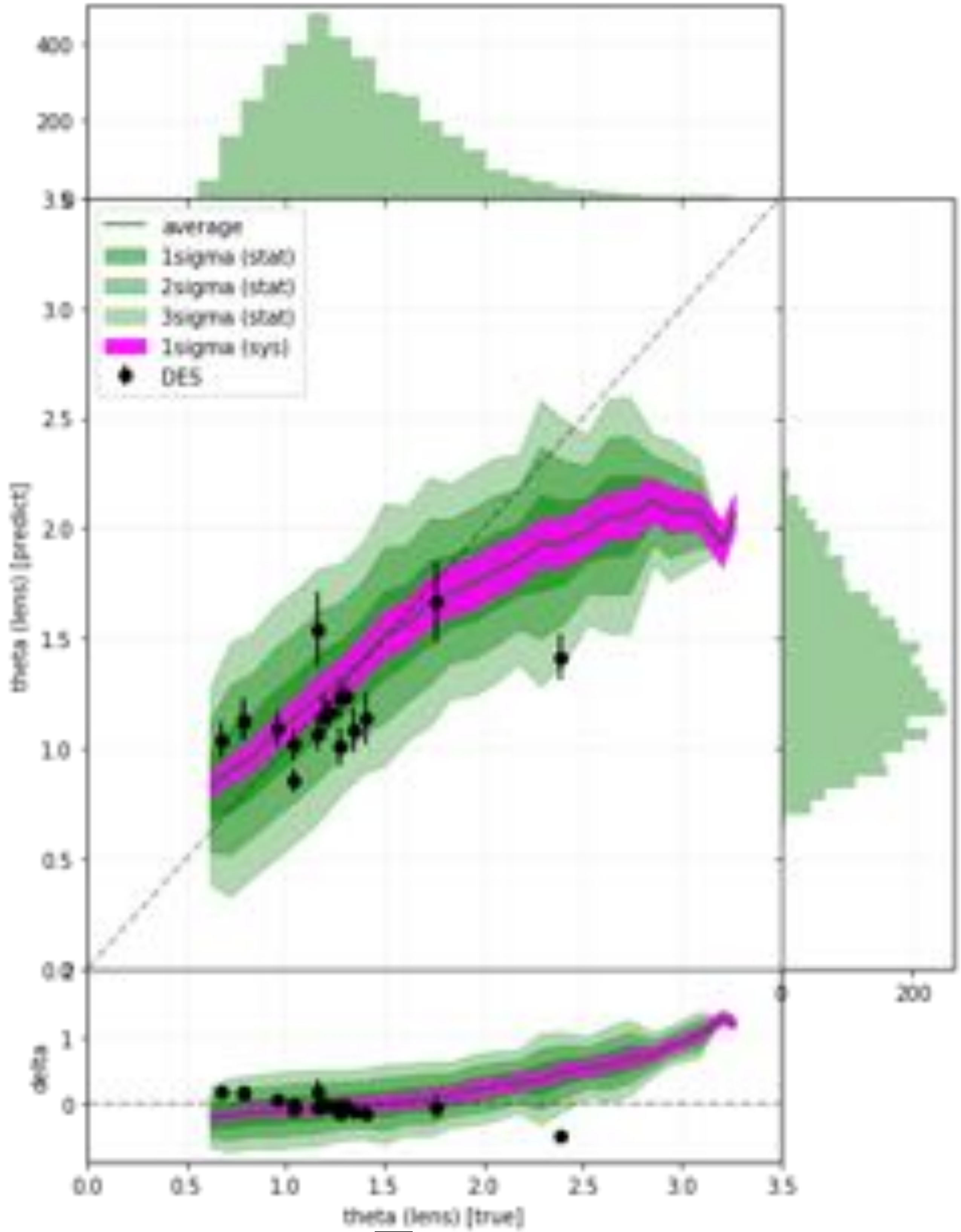
- Each column is a different object and its probability of detection in the network.
  Left: True positive
  Right: False positive
- Convolution layers filter the images to highlight features
- Pooling layers down-sample images, efficiently reducing parameters for modeling
- See also work by Lanusse+17,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



Predicting strong lens parameters: e.g., Regress on Einstein Radius

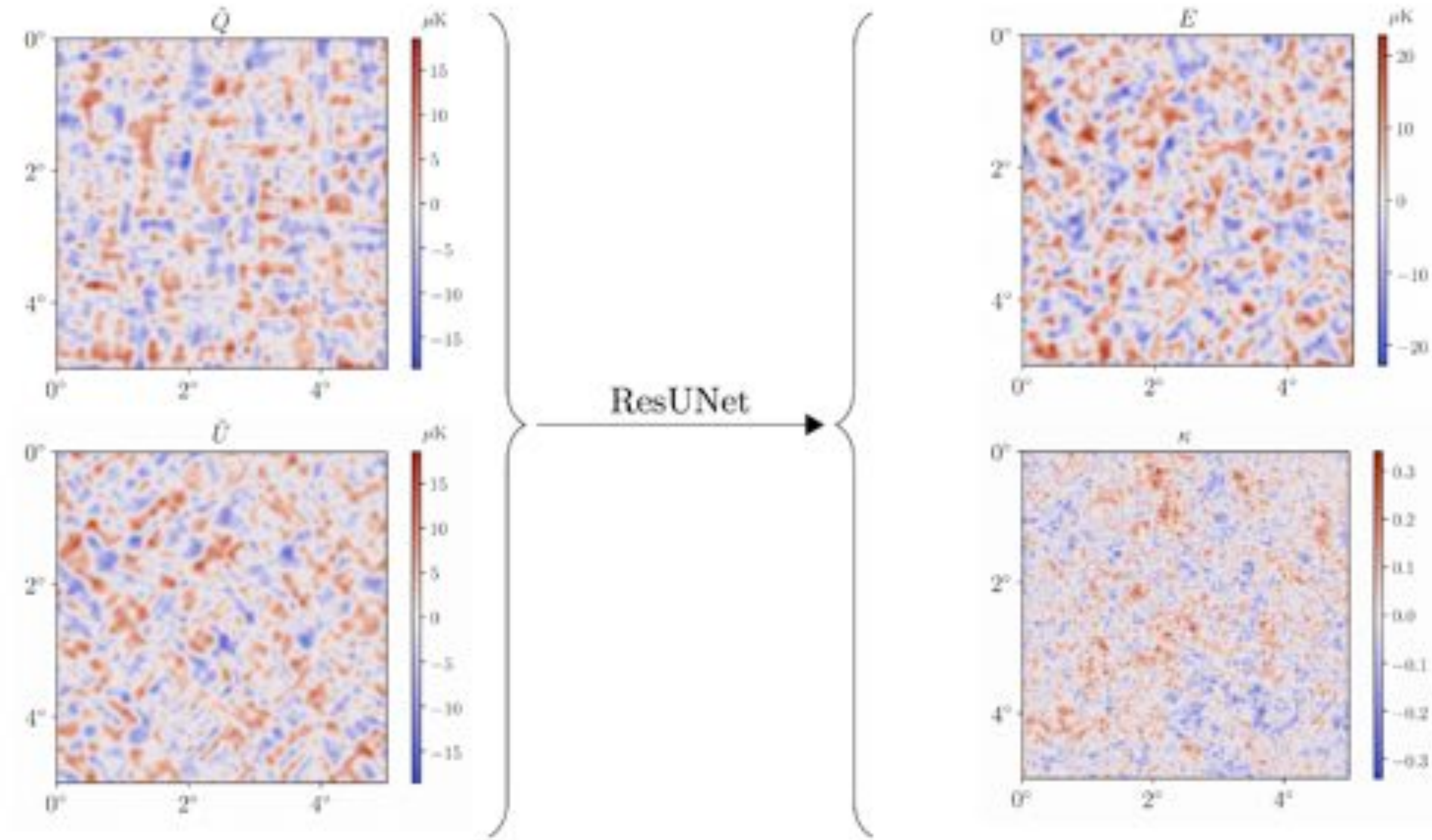
Predict

Difference

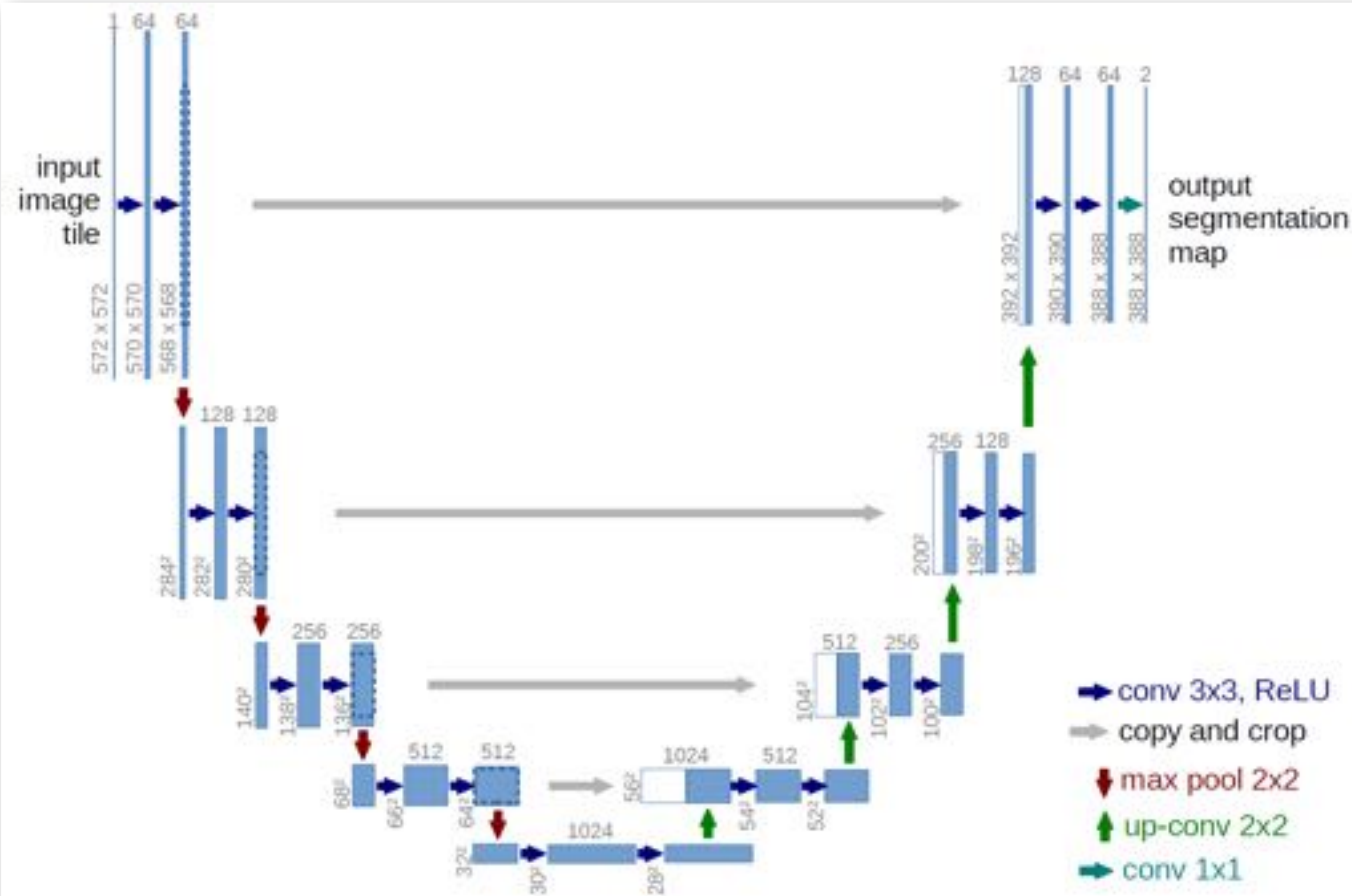


True

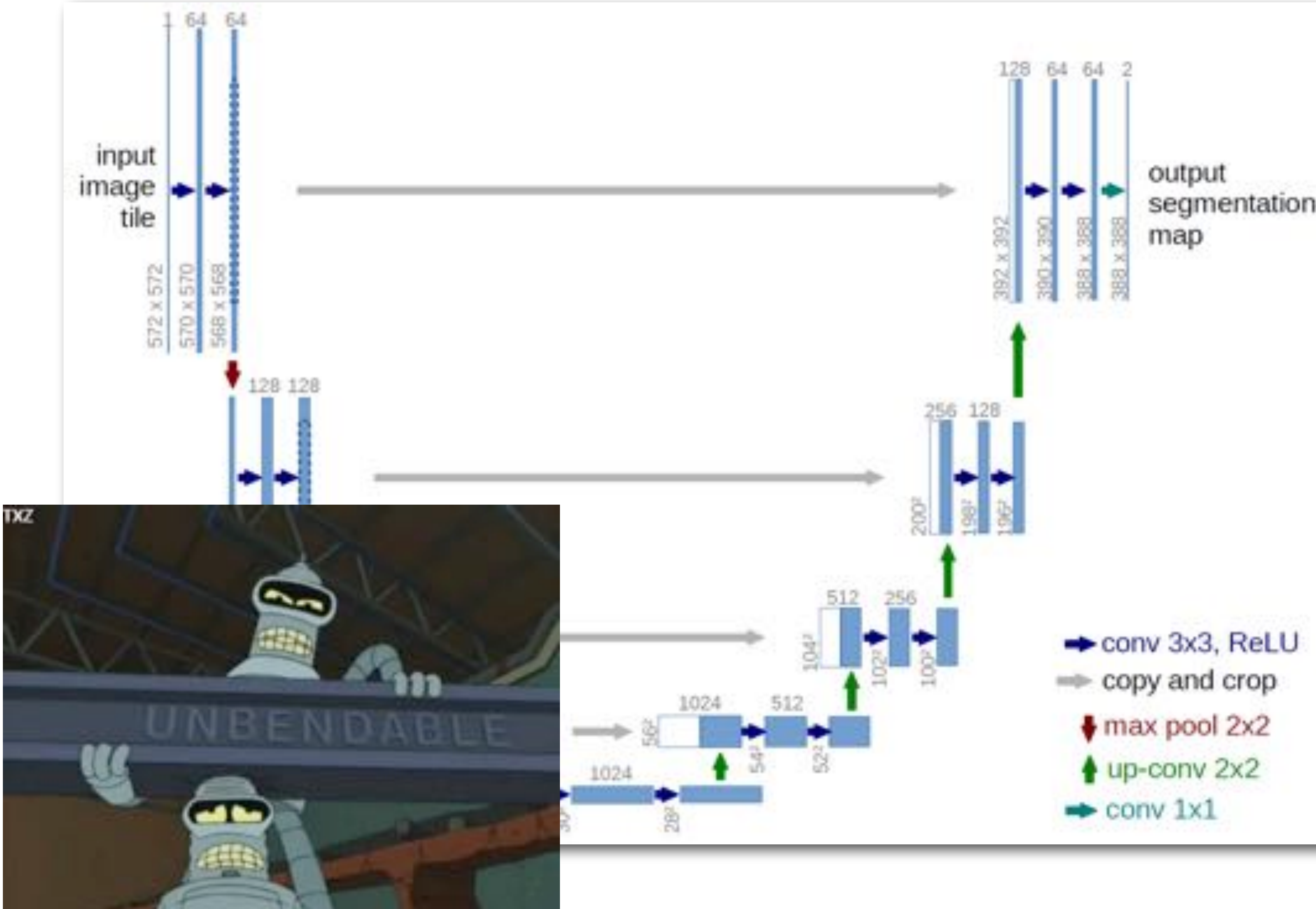
DeepCMB: de-lensing the CMB with UNets (Caldeira+2018)



DeepCMB: de-lensing the CMB with UNets (Caldeira+2018)

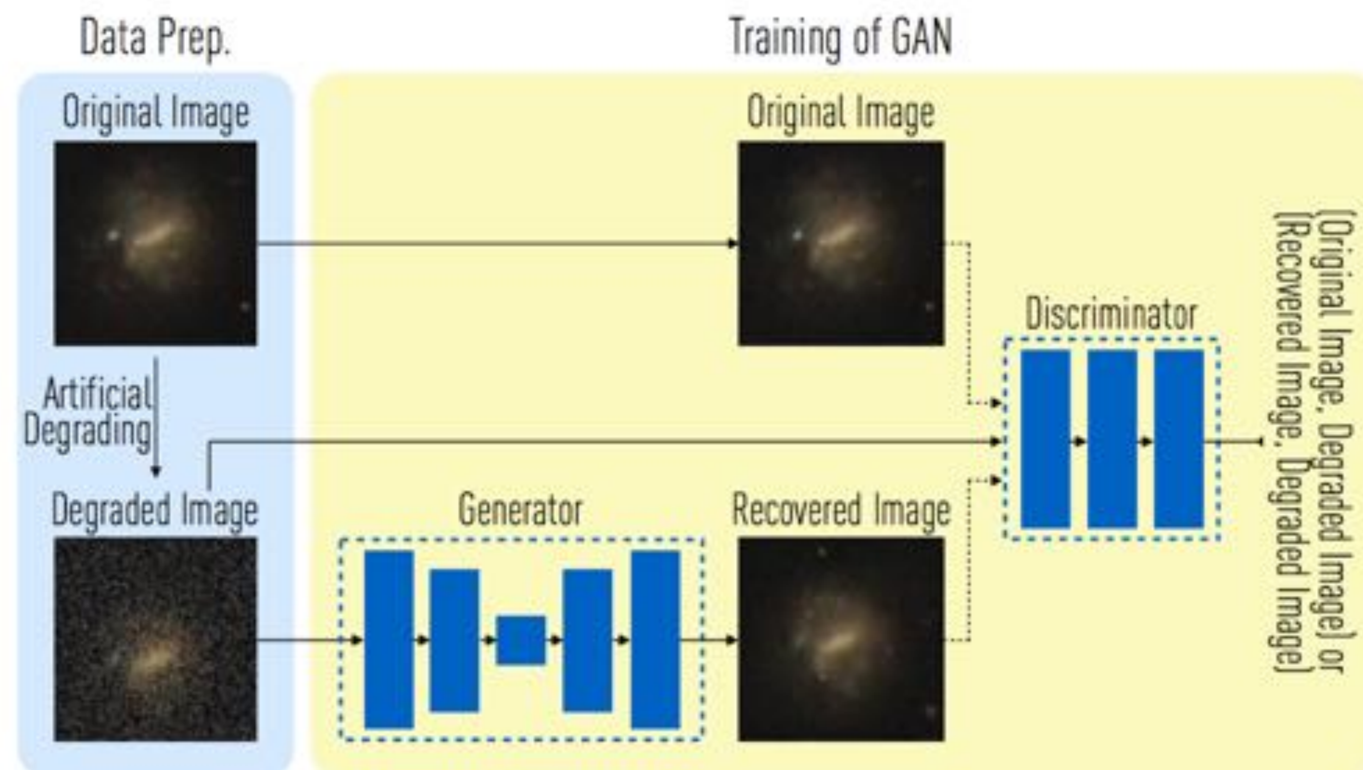


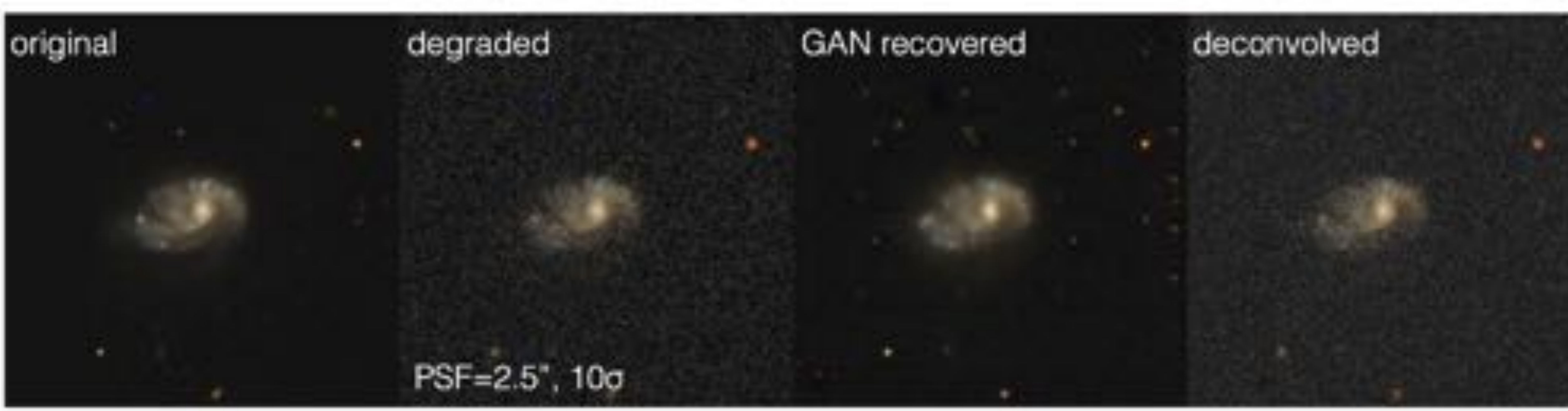
DeepCMB: de-lensing the CMB with UNets (Caldeira+2018)



Galaxy Image Simulation (Schawinski+2017)

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

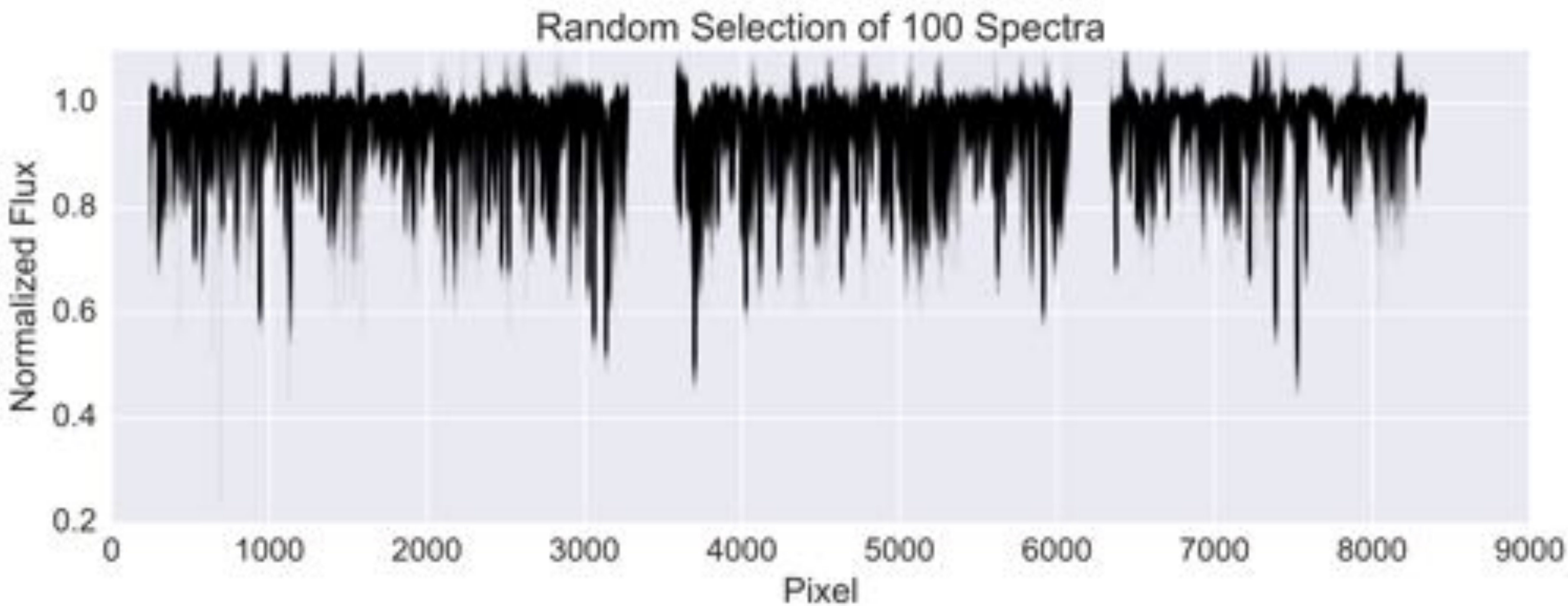




1 DIMENSION

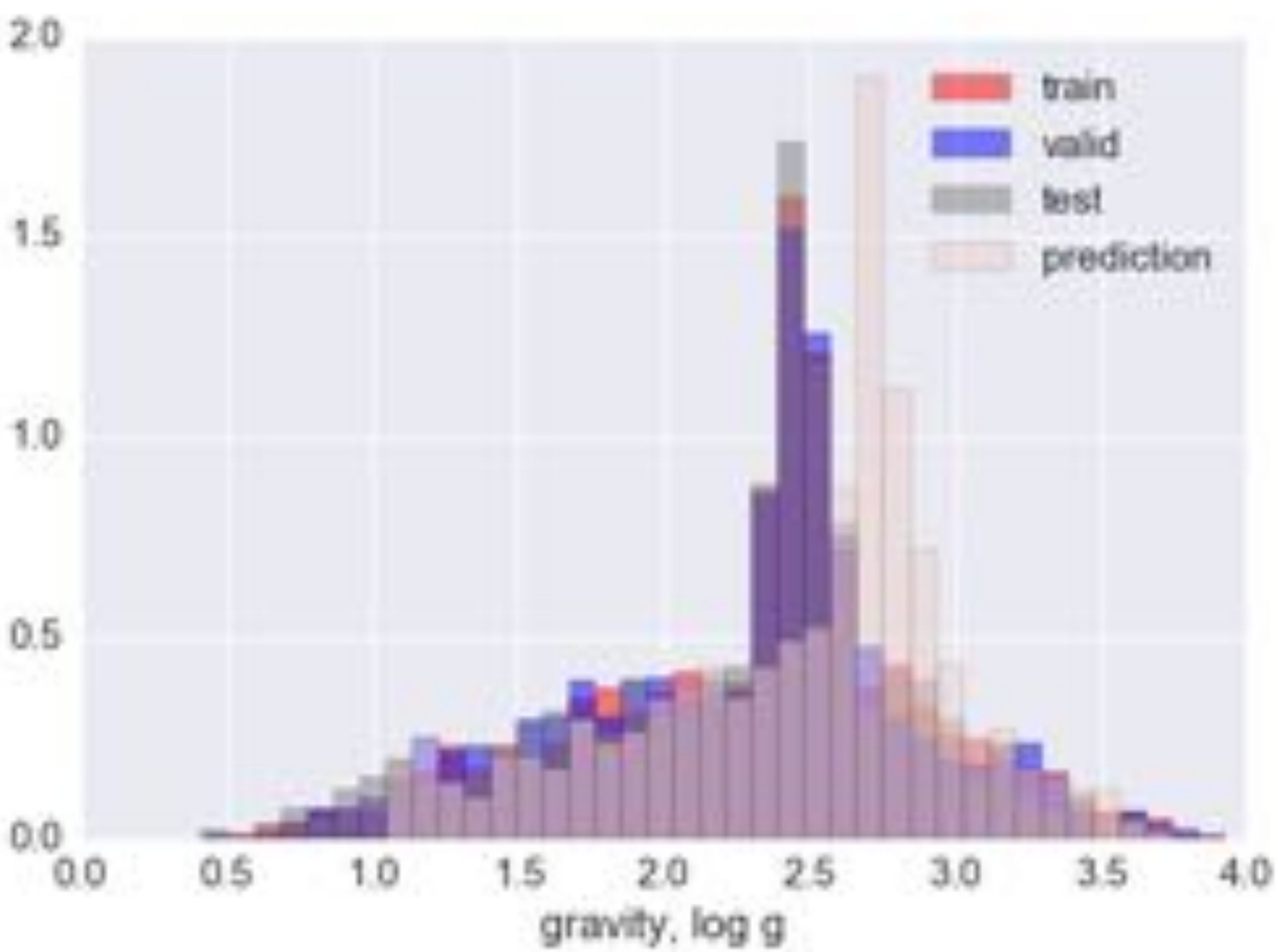
Fitting Stellar Spectra with 1D CNNs (Nord + Price-Whelan)

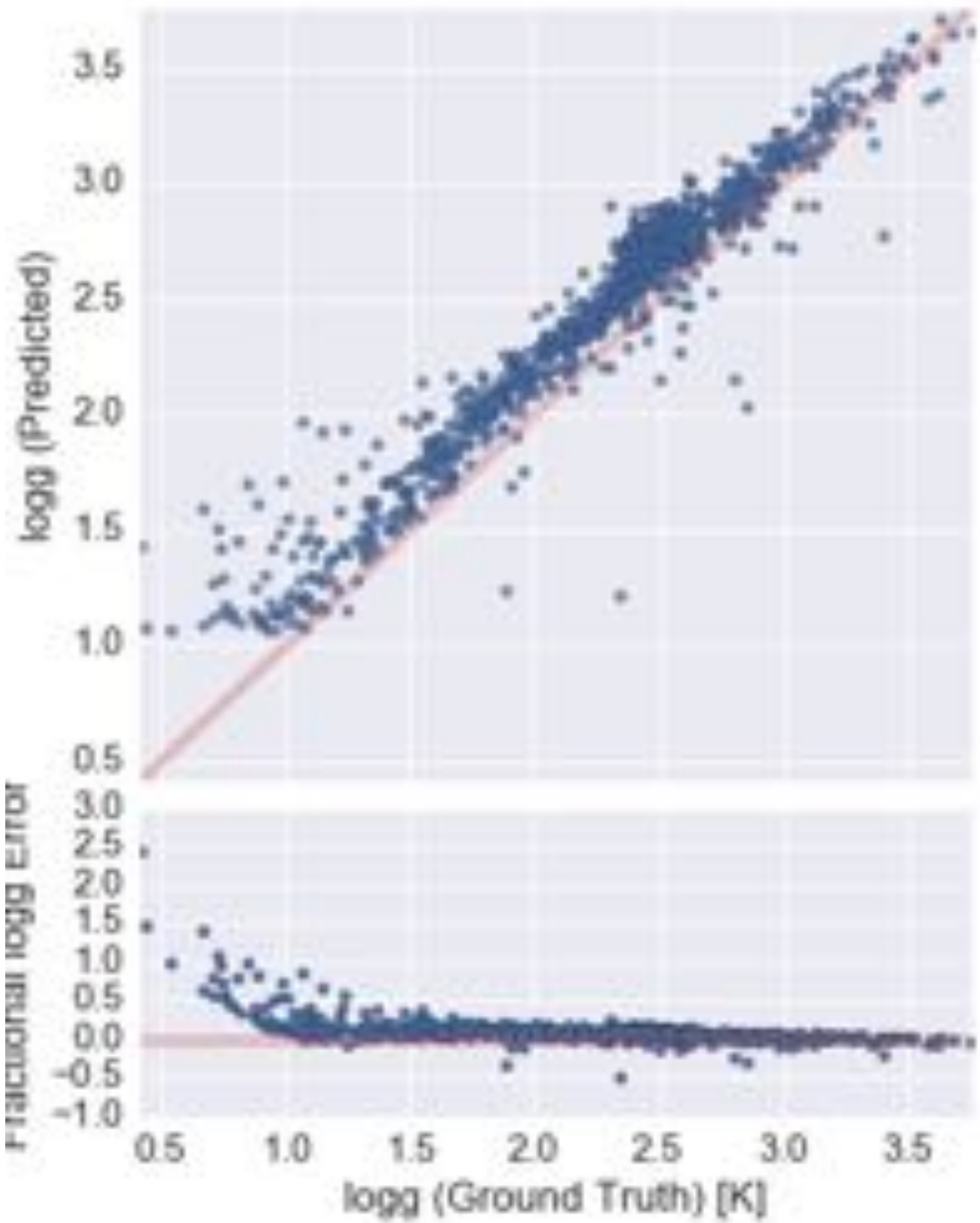
- Fit 1D Apogee stellar spectra with labeled quantities: *Teff, log g, metallicity* (see Ness+2015)
- Architecture
  - 3 conv, 3 pooling, and 1 drop out layer
  - 15 lines of (DL) code, a GPU and 40 minutes of compute time.



Gravity (log g)

- moderate biases
- architecture achieves very low losses
- we may need more diverse training set.





Outlook

- Neural Architecture Search: Neural Networks tailored to your problem
  - Google AutoML
  - ENAS (Stanford)
  - Oak Ridge’s MENDL uses genetic algorithms

Outlook

- Neural Architecture Search: Neural Networks tailored to your problem
    - Google AutoML
    - ENAS (Stanford)
    - Oak Ridge’s MENDL uses genetic algorithms
- Uncertainties
    - Bayesian Neural Nets - can they work for physical physical parameters
    - Current standard: concrete dropout

Outlook

- Neural Architecture Search: Neural Networks tailored to your problem
  - Google AutoML
  - ENAS (Stanford)
  - Oak Ridge's MENDL uses genetic algorithms
- Uncertainties
  - Bayesian Neural Nets - can they work for physical physical parameters
  - Current standard: concrete dropout
- Searching for symmetries
  - Group symmetries govern the the convolutional process.
  - Different group symmetries mean different spaces that can be convolved.
    - Z^2 for translational symmetry
    - SO3 for 3D rotational symmetry

Outlook

- Neural Architecture Search: Neural Networks tailored to your problem
  - Google AutoML
  - ENAS (Stanford)
  - Oak Ridge's MENDL uses genetic algorithms
- Uncertainties
  - Bayesian Neural Nets - can they work for physical physical parameters
  - Current standard: concrete dropout
- Searching for symmetries
  - Group symmetries govern the the convolutional process.
  - Different group symmetries mean different spaces that can be convolved.
    - Z^2 for translational symmetry
    - SO3 for 3D rotational symmetry
- Quantum Computing
  - IBM has a 20-qubit system you can submit a job to.
  - Machine Learning and quantum has significant promise … and peril.

Perspectives and Problem-solving Approaches

Data scientist perspective:
"What is the format of the data, and what are the patterns you might see?"

Scientist perspective:
"What is a model with physical meaning that can describe the patterns in this data?"

DeepSkiesLab.com

DEEP SKIES

Bringing Artificial Intelligence to Astrophysics

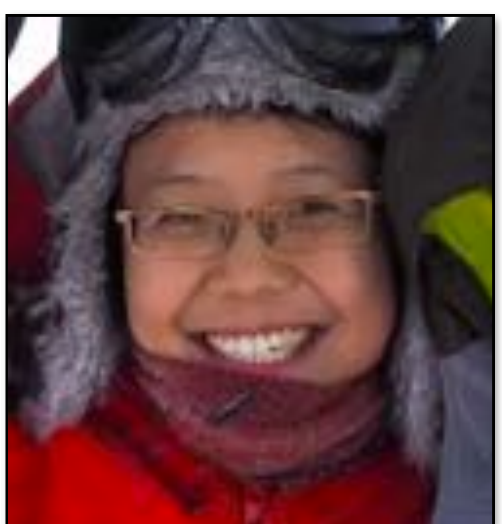
Brian Nord, PhD

Josh Peek, PhD

Camille Avestruz, PhD

Current Projects:

- Strong Lensing
- Early Universe
- Simulations
- Quantum Computing
- Automating Telescopes



Kimmy Wu, PhD

João Caldeira

Shubhendu Trivedi


Wooo let's go!