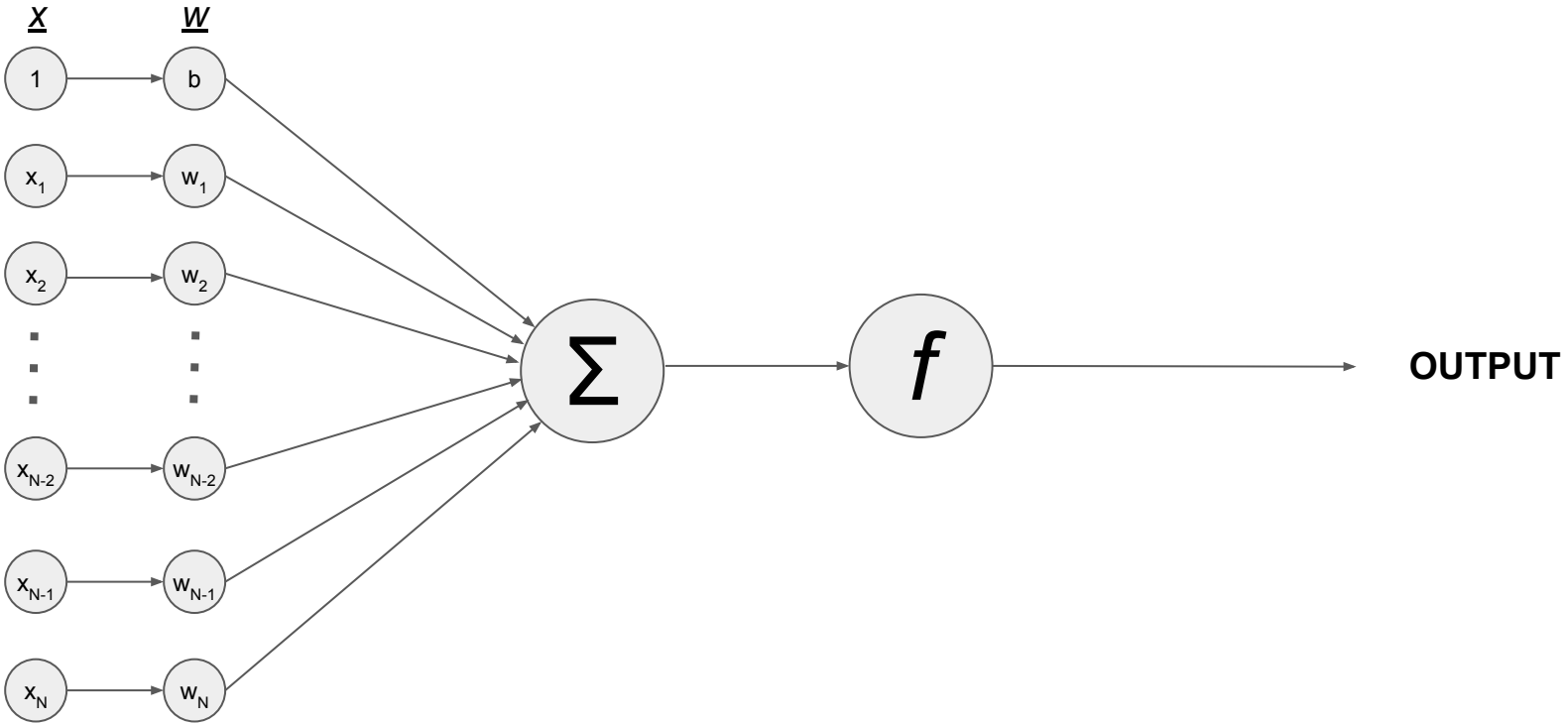
An Application of Machine Learning to Star-Galaxy Classification

Aidan Sedgewick

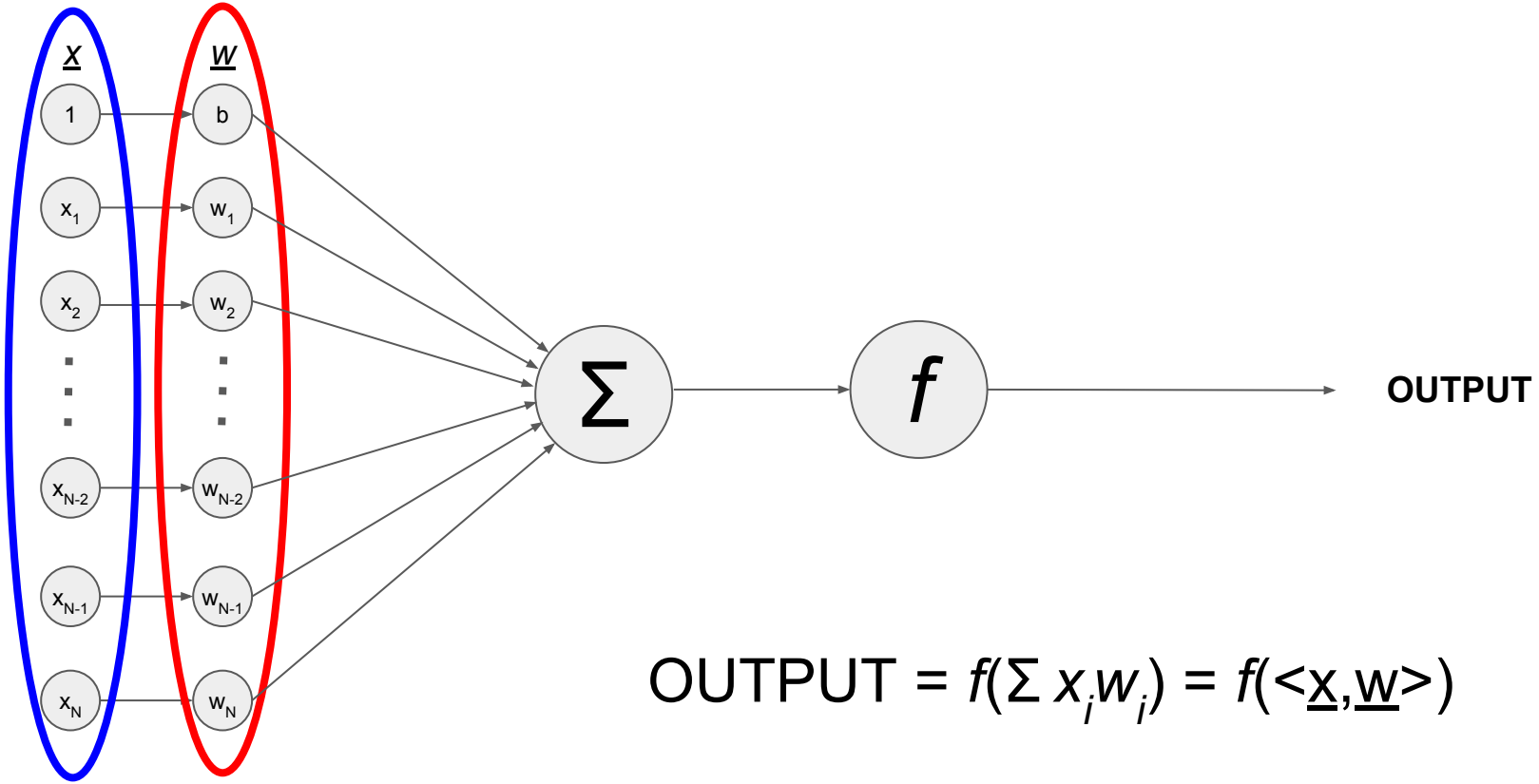
Outline

- What is an (artificial) neural network (ANN)?
- ANN vs CNN
- CNN layers: convolution and pooling
- Application to astronomy (PAU survey star/galaxy classification)

Artificial neuron

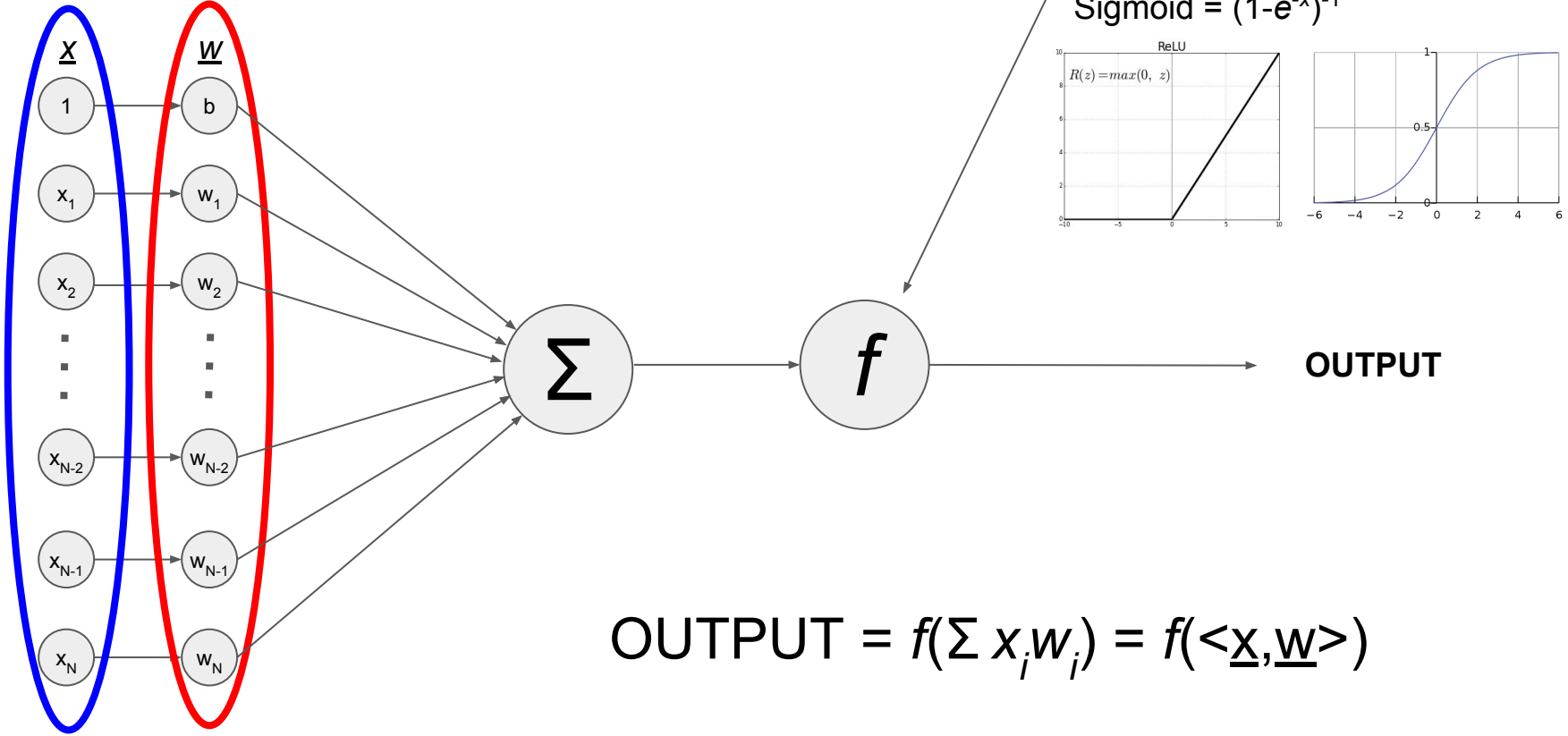


Artificial neuron

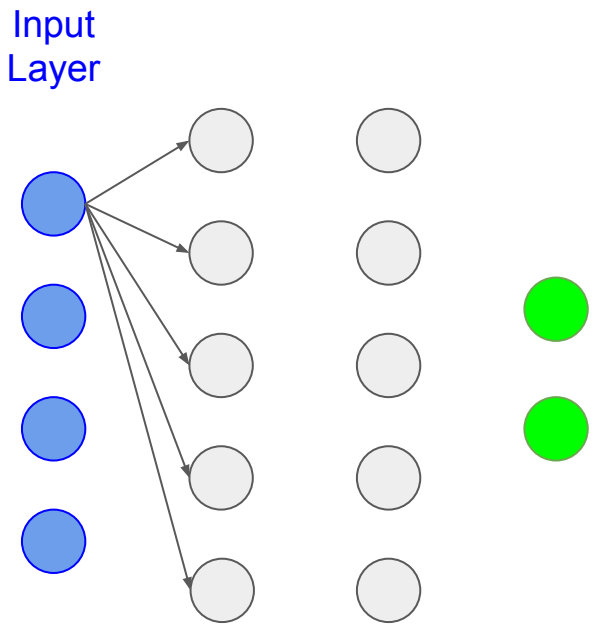


$$\text{OUTPUT} = f(\Sigma\, x_i w_i) = f(\langle \underline{x}, \underline{w} \rangle)$$

Artificial neuron

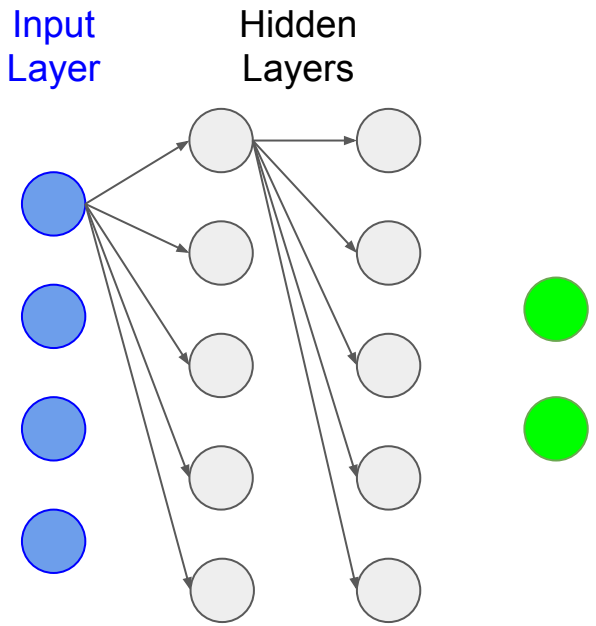


Neural networks



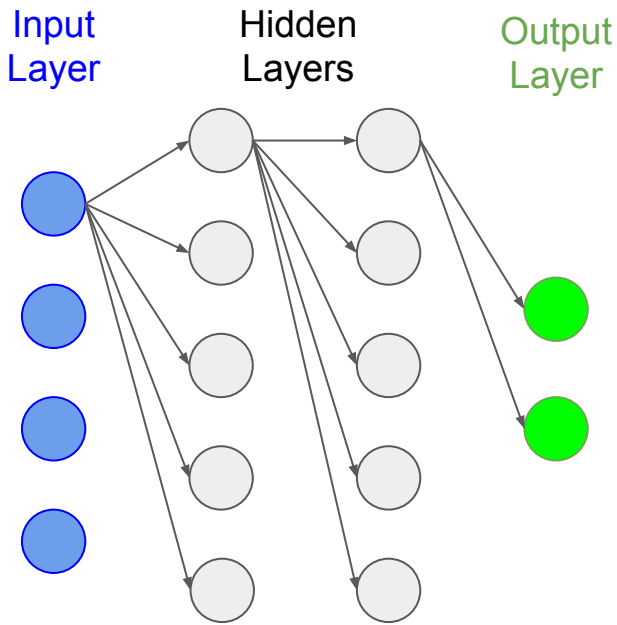
- Can chain artificial neuron outputs to new inputs, and make a ‘network’.

Neural networks



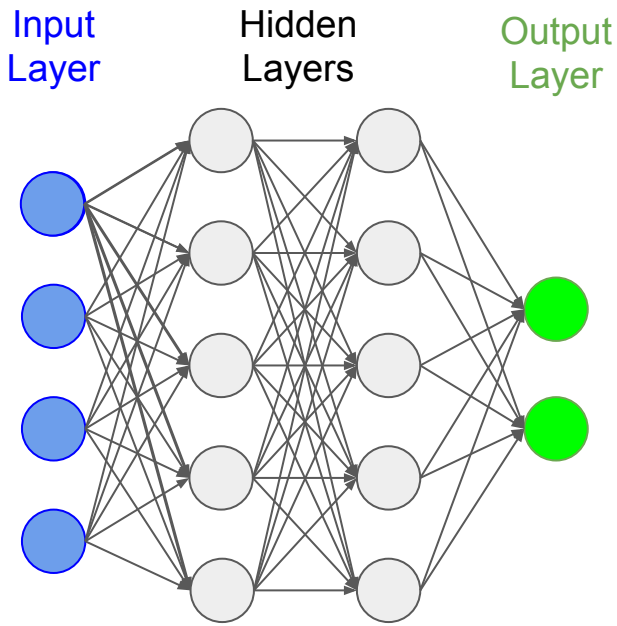
- Can chain artificial neuron outputs to new inputs, and make a ‘network’.

Neural networks



- Can chain artificial neuron outputs to new inputs, and make a ‘network’.

Neural networks



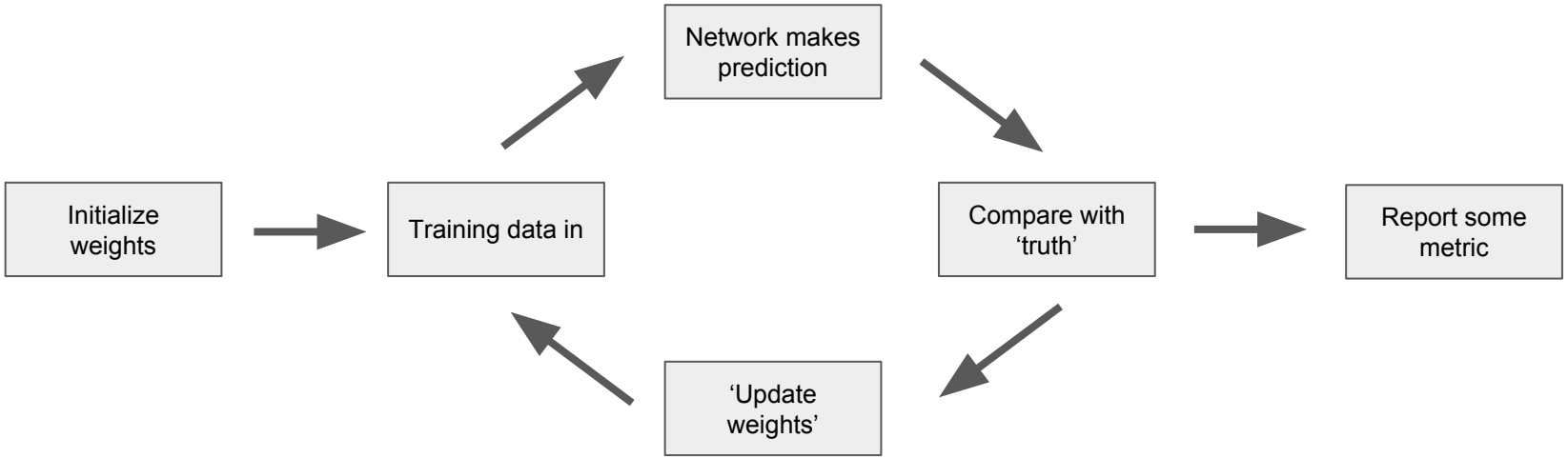
- Fully connected, or ‘dense’ layers.
- Each neuron has its own set of weights $\underline{w}$.

- Number of neurons in a layer?
- Number of layers?

→ *Hyperparameters*

Training parameters

- How to get a network which outputs the ‘correct’ answers?



- How many ‘epochs’ (iterations)?

Convolutional neural network (CNN)

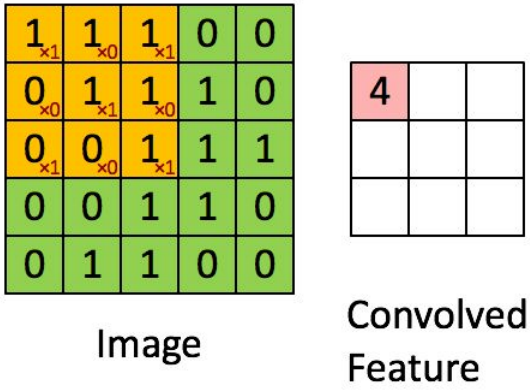
- Often have image data that we'd like to apply ANNs to.
- Need many layers for complex patterns, images with many pixels...
- Huge number of training parameters for ANN!

Use convolutional network!

- Made of two main types of layer: convolutional, and pooling layers...

Convolution layers

- Convolutional layers are *not* fully connected.
- Operate *locally*, so can still detect features.

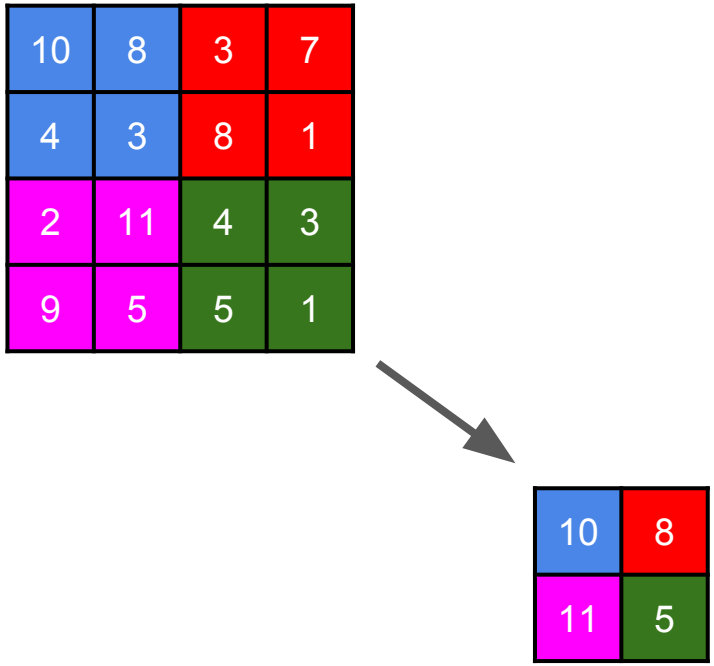


- Kernel: ‘Moving window filter’
- Far fewer parameters to be learned.
- No activation function here.

Kernel =

×1	×0	×1
×0	×1	×0
×1	×0	×1

Pooling layers



- “Serve to reduce dimensionality” - coarse graining of images.
- Max pooling or average pooling are commonly used.

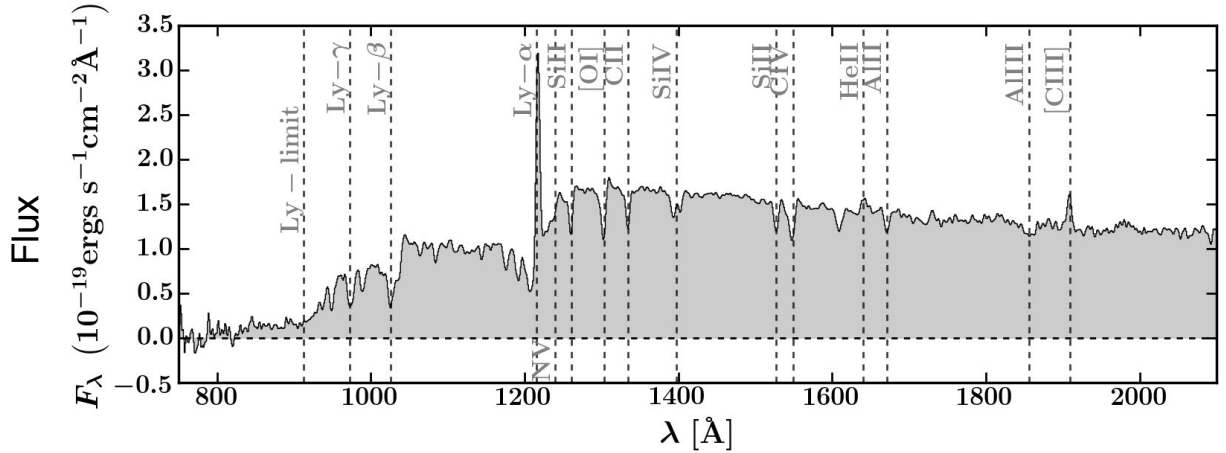
- Usually no trainable parameters in these layers.
- Usually apply activation function *f* at this stage.

PAU Survey

- Physics of the Accelerating Universe Survey
- Provide accurate “photometric redshift” measurements for up to 30 million objects.

Redshift Estimates: spectro-z

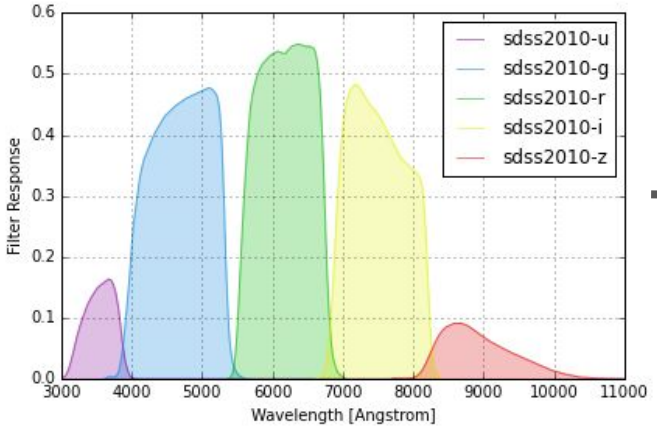
- Most accurate way to obtain redshift estimates is with high-resolution spectra ('spectro-z')
- *Extremely* time consuming, but error of 0.01%



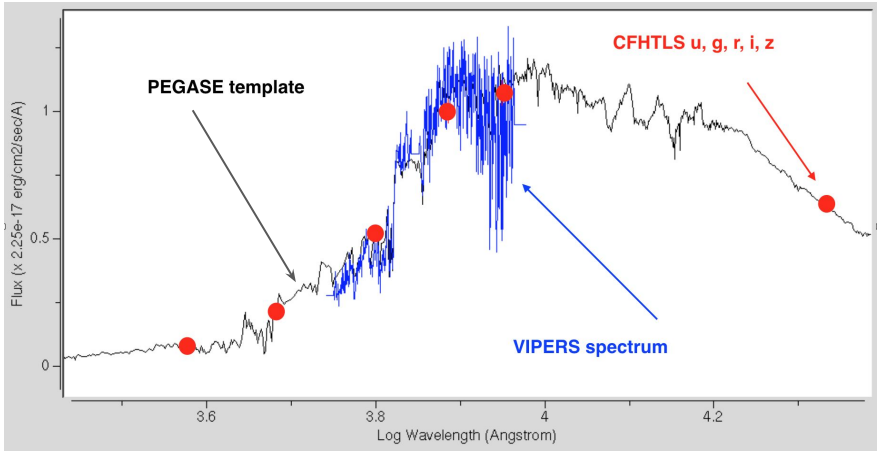
CANDELS-VIMOS

Redshift Estimates: photo-z

- Can obtain rough redshift estimates (~5% error) with few images in different bands ('photo-z') very quickly.



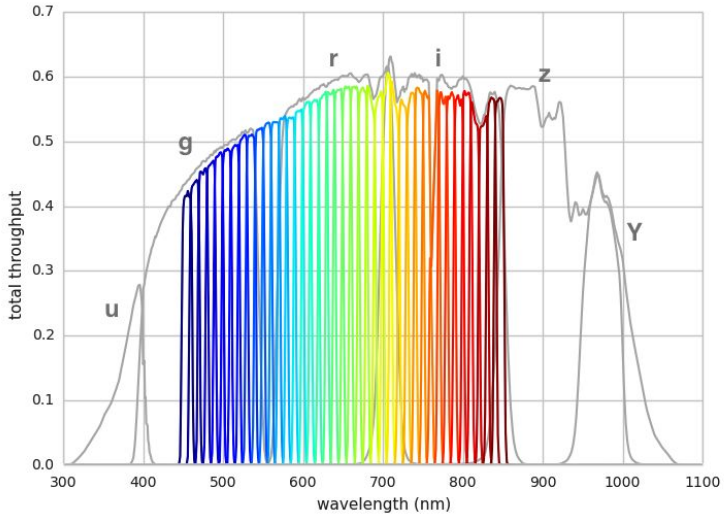
github.com/dkirkby/speclite



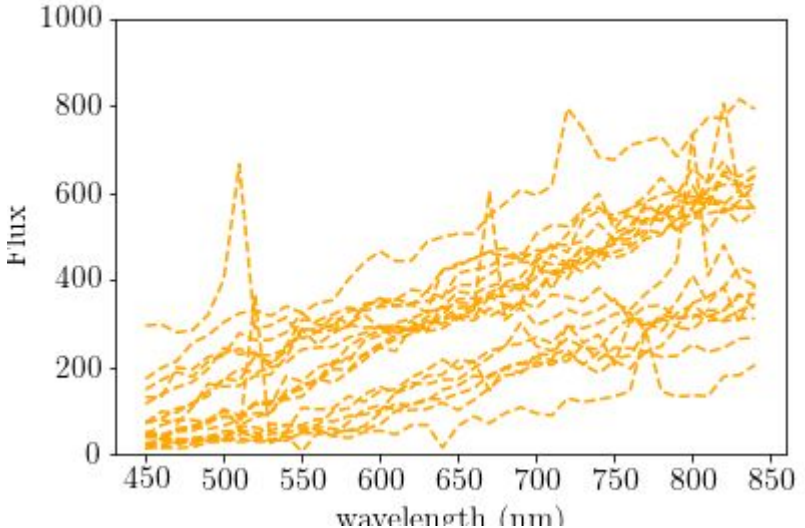
Giorgio Manzoni

PAUS

- Traditionally have ~6 filters
- PAU Cam has 40 narrow band filters!
- Effectively low resolution spectra.



PAUS collab.



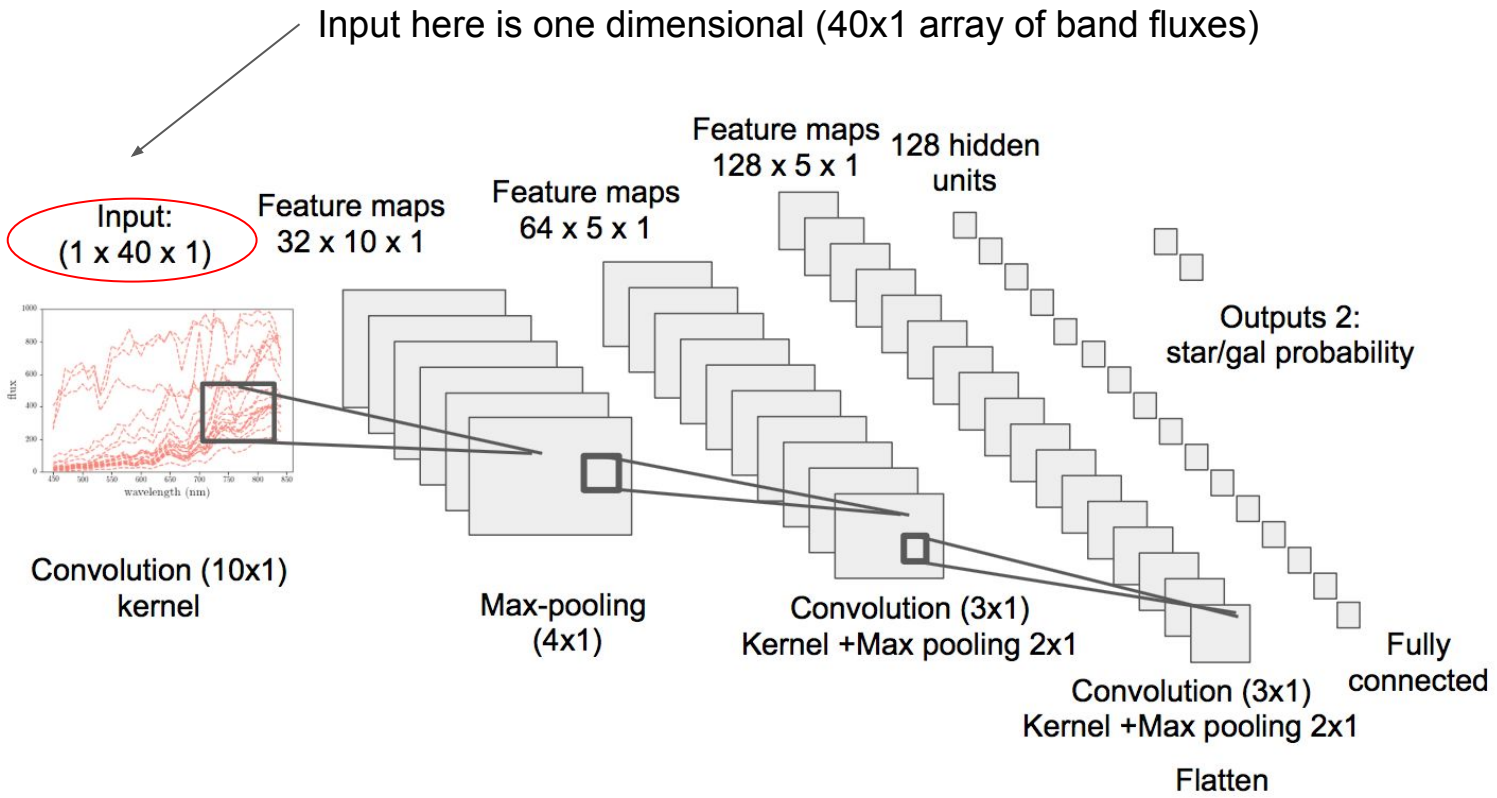
Cabayol '18.

Applying NNs to PAUS

- First, need to determine what type of object, to use relevant template.
- Can we also use the band fluxes to determine this?

- Problem in this paper is star-galaxy classification.
- Use previously imaged fields (*Hubble* data) as a training set - 1.2M objects.
  1.1M galaxies, 30k stars → *strong* class imbalance

Network design



Performance metric

- How to evaluate the performance of the classifier?
- Defining a galaxy detection as a ‘positive’ result:

	Classified as galaxy	**Classified as star**
True galaxy	TP	FN
True star	FP	TN

$$TPR = \frac{TP}{TP + FN}$$

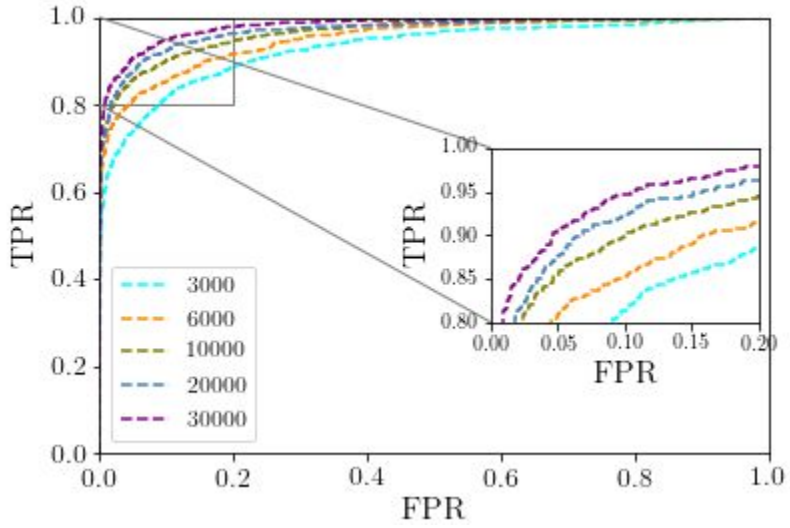
‘Fraction of true galaxies correctly classified as galaxies’.

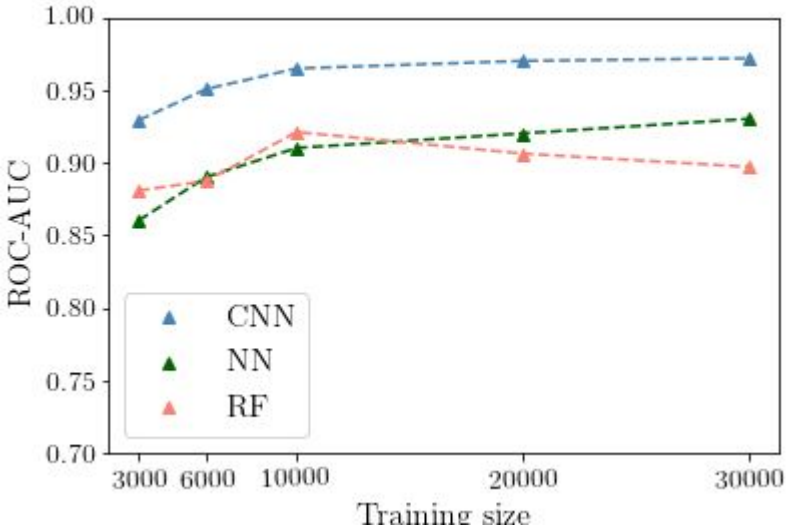
$$FPR = \frac{FP}{FP + TN}$$

‘Fraction of true stars falsely classified as galaxies’.

ROC Curves - training size

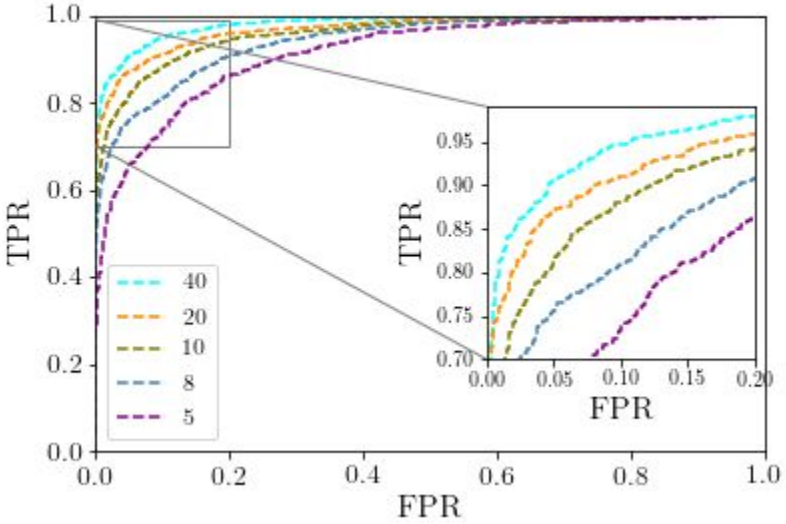
- For some acceptable *FPR*, what is the obtained *TPR*?
- Area under curve is also a useful measure of performance.

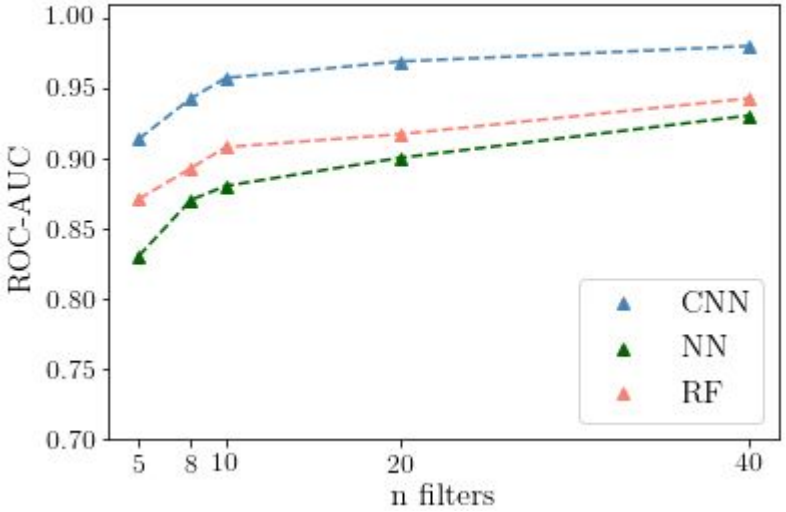




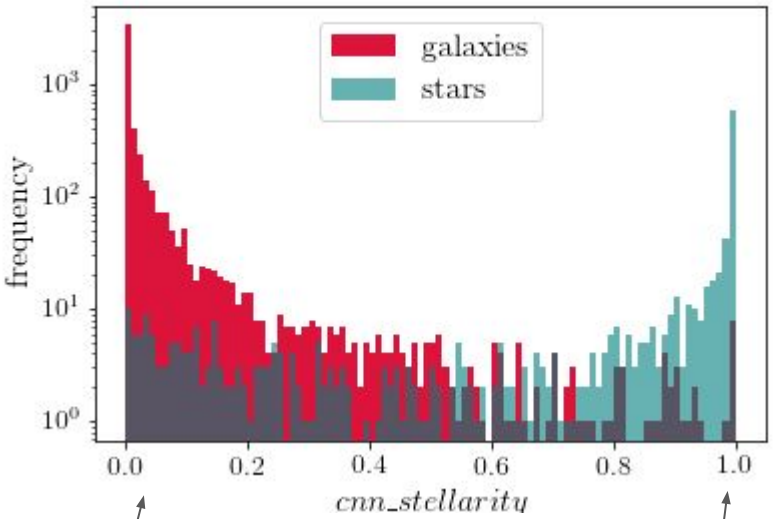
ROC Curves - input size

- Combine consecutive filters to simulate broader bands.



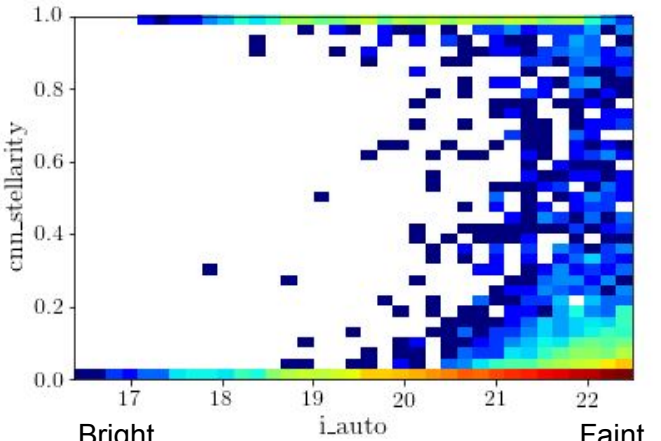


Output distribution



Very confident galaxy prediction

Very confident star prediction



Bright objects

Faint objects

Result

- 99% purity, for 98% completeness.

“Of all the galaxies in the sample, we detect 98% of them. 99% of what we label galaxies truly are galaxies”.