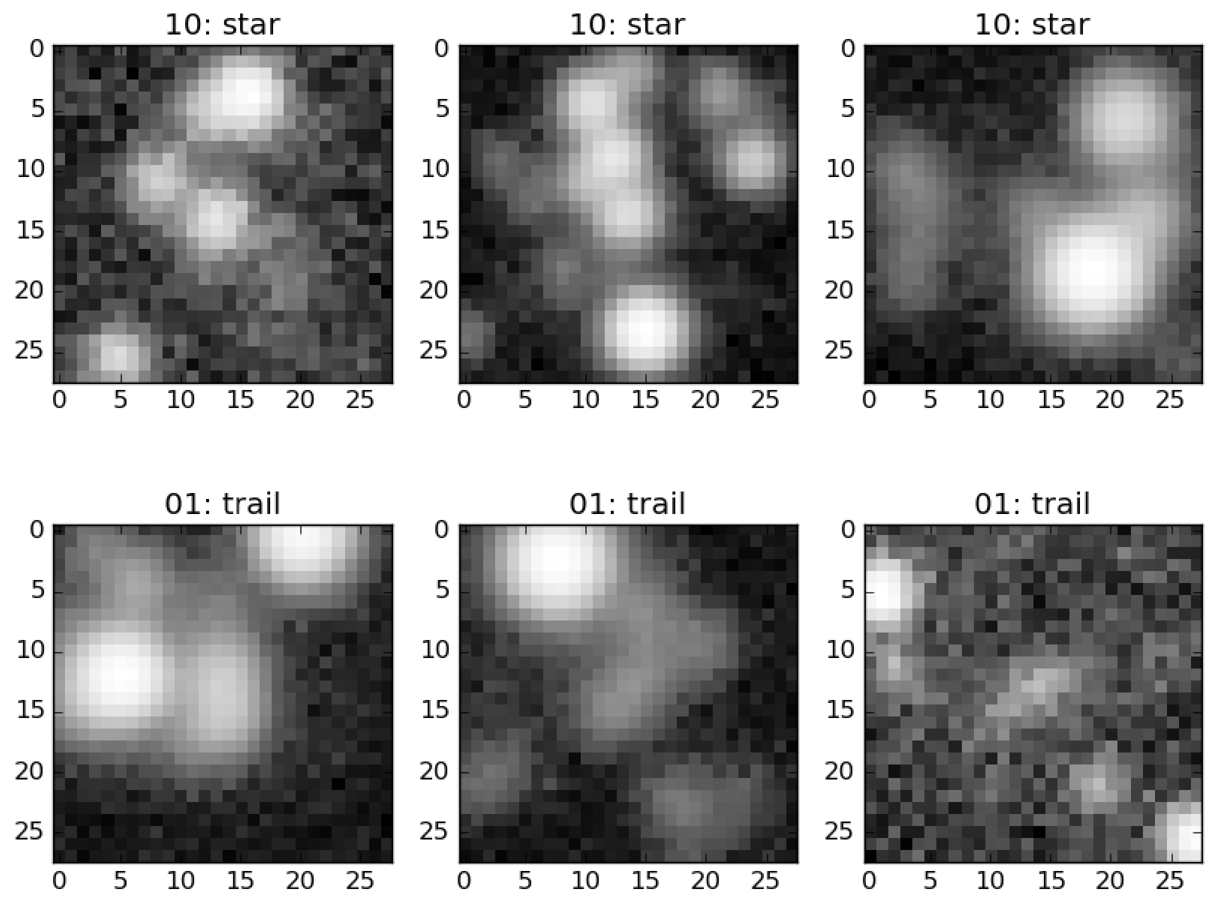
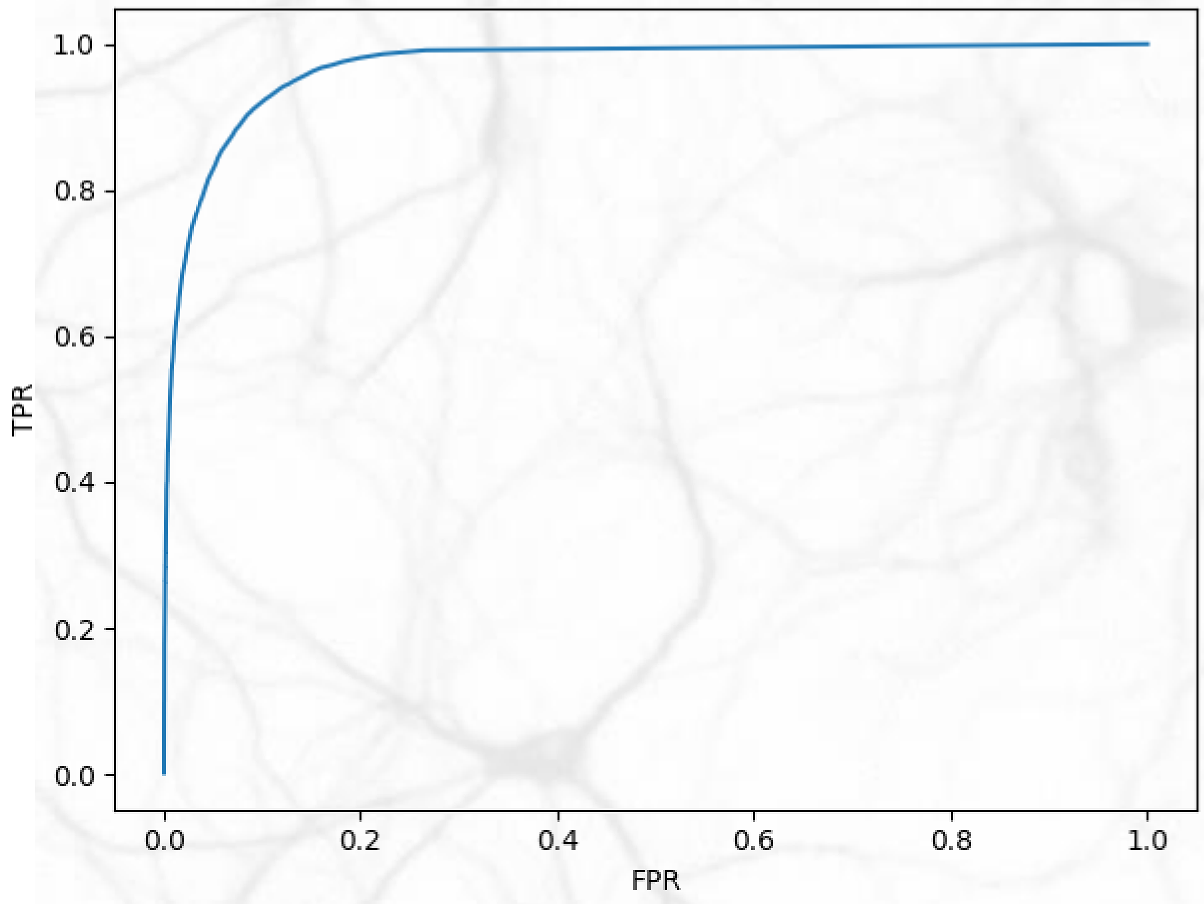
Inferring from astronomical images using Deep Learning

Emmanuel Bertin (IAP)

A typical astronomical image analysis problem

- Example: Identify a specific type of source in a crowded field

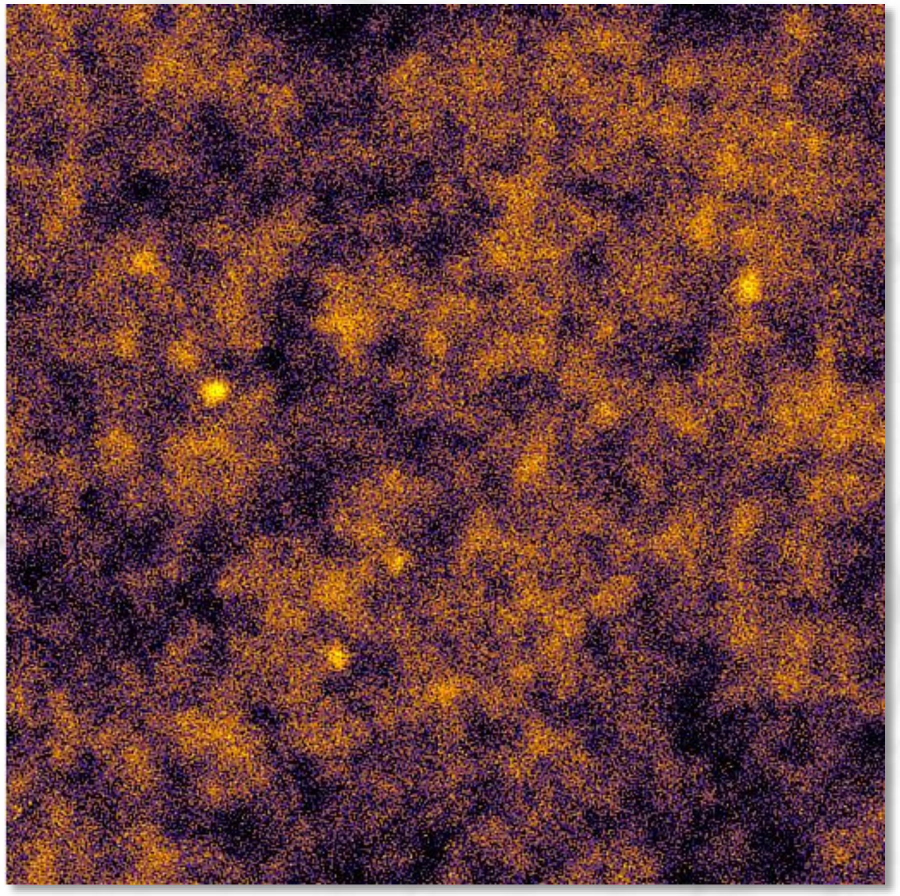




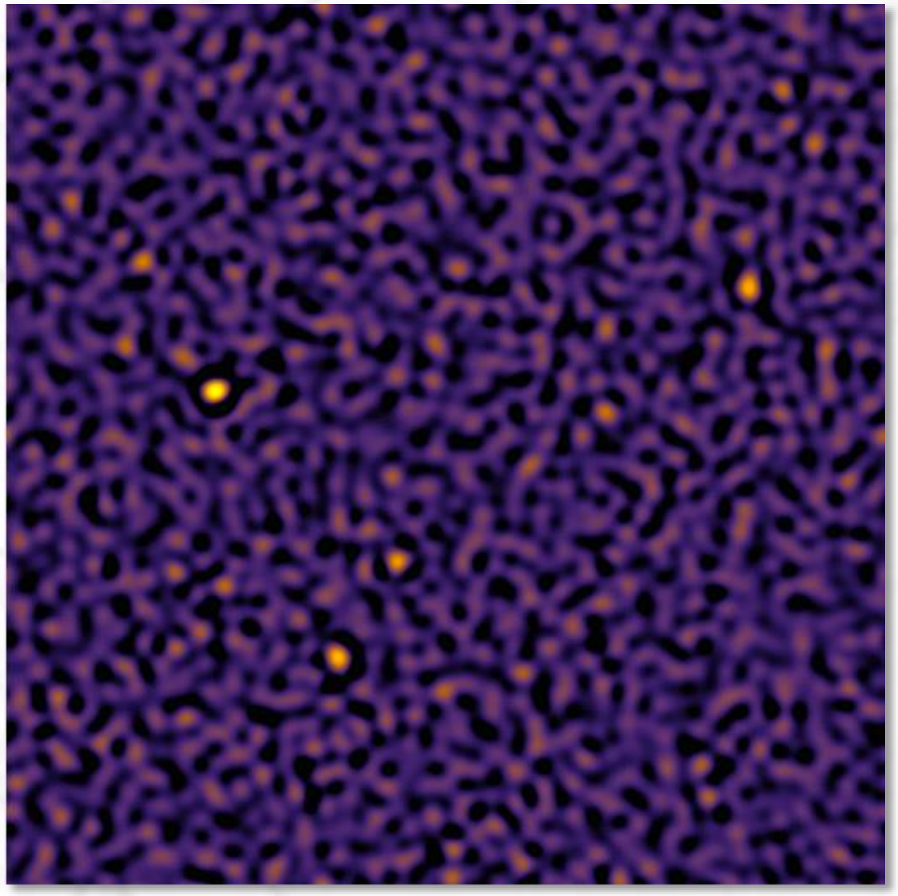
- We want probabilities!

Matched filter

- Optimal linear filter for identifying a known pattern with added wide-sense stationary noise: $H(\boldsymbol{k}) = \frac{\Phi^*(\boldsymbol{k})}{P(\boldsymbol{k})}$
  - Convolve with $h(\boldsymbol{x}) = \phi^\top(\boldsymbol{x})$ for white noise
  - Deconvolve with $\phi(\boldsymbol{x})$ for confusion noise

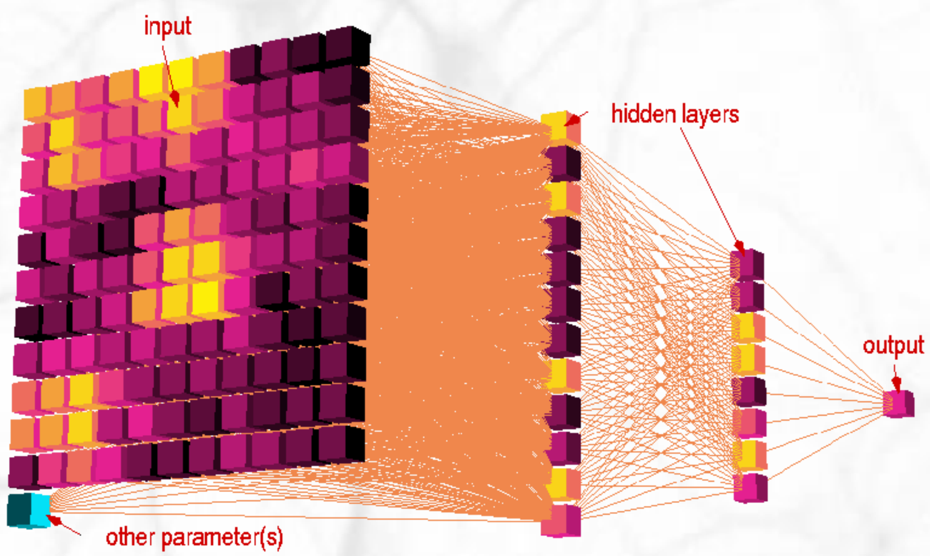


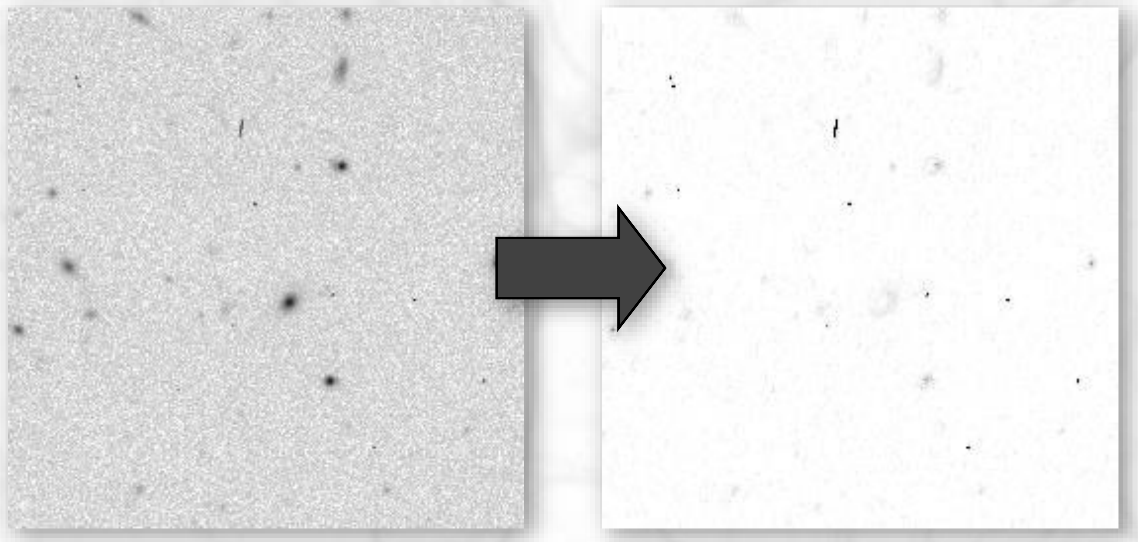
filtered



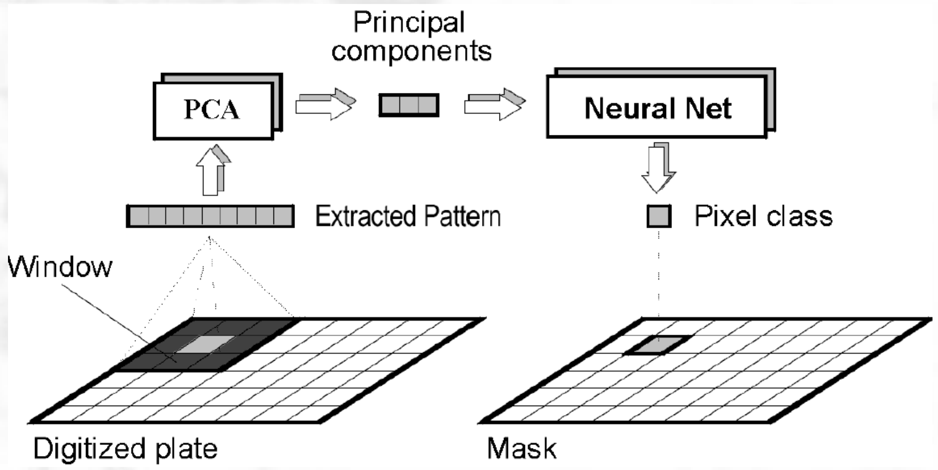
Early AI-driven detection filters

- "EyE" (1998, 2001)
  - Multilayered Perceptron where inputs are compressed pixel values
  - Used in production for detecting CR impacts

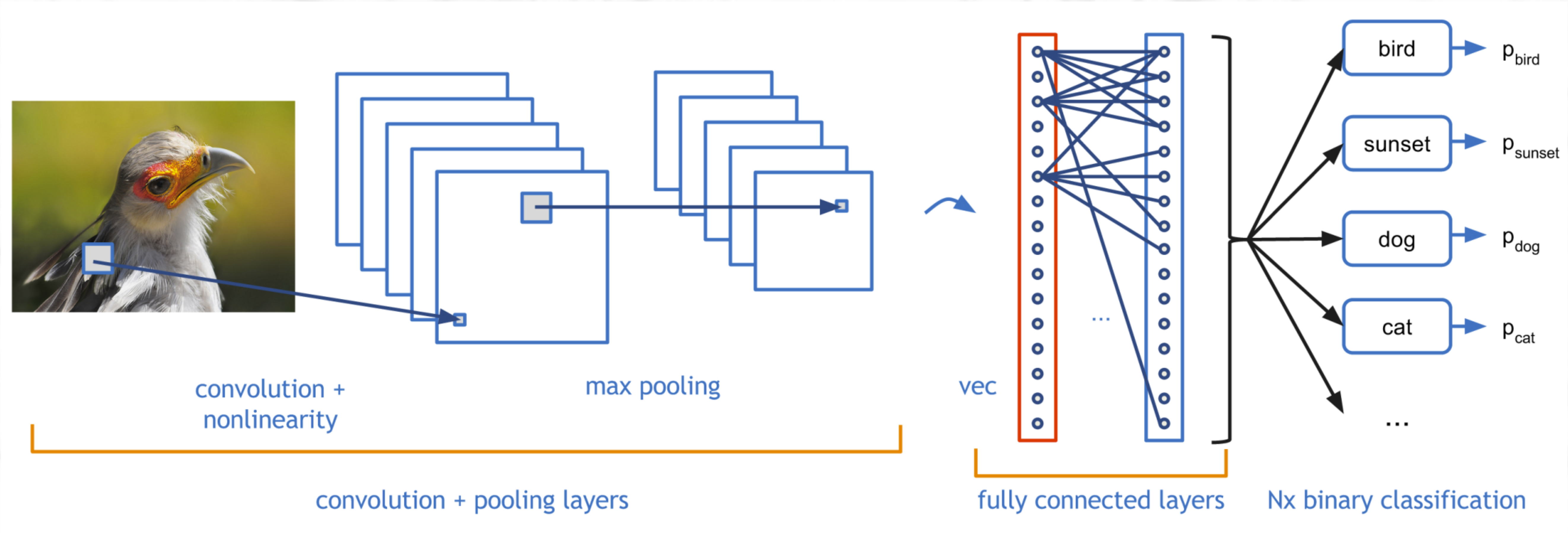




- "NExtractor (Andreon et al. 2000)
  - PCA + multilayered Perceptron



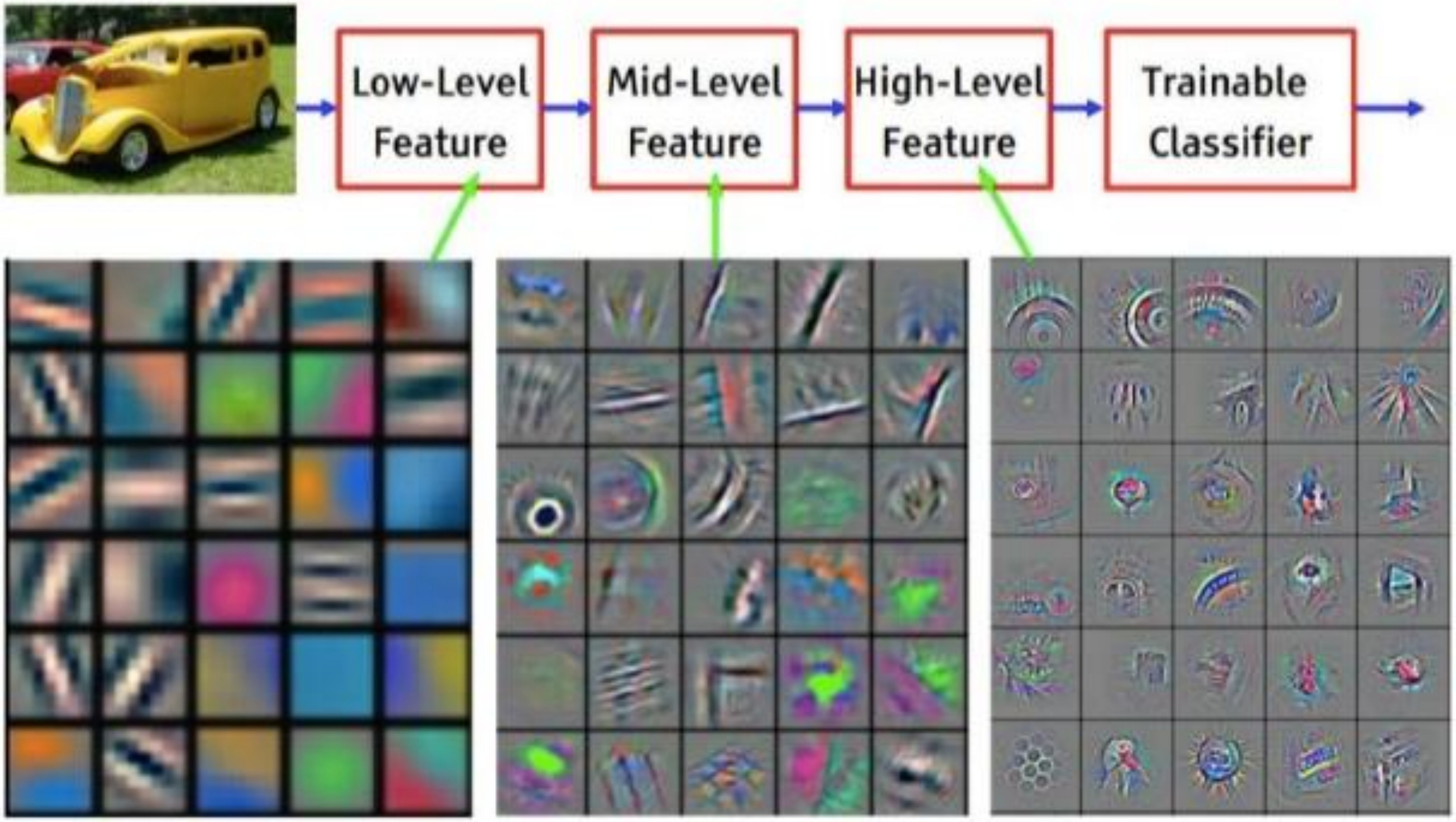
Deep Learning with convolutional layers



- Convergence of several breakthroughs
    - Convolutional nets (CNNs, e.g., LeCun et al. 1998)
    - “Guided” unsupervised learning on individual layers (e.g., Hinton et al. 2006)
    - Learning of sparse representations (e.g., Ranzato et al. 2007)
    - GPU libraries (e.g., cuDNN) 
    - Availability of large datasets with labels throughout the web (e.g. ImageNet)

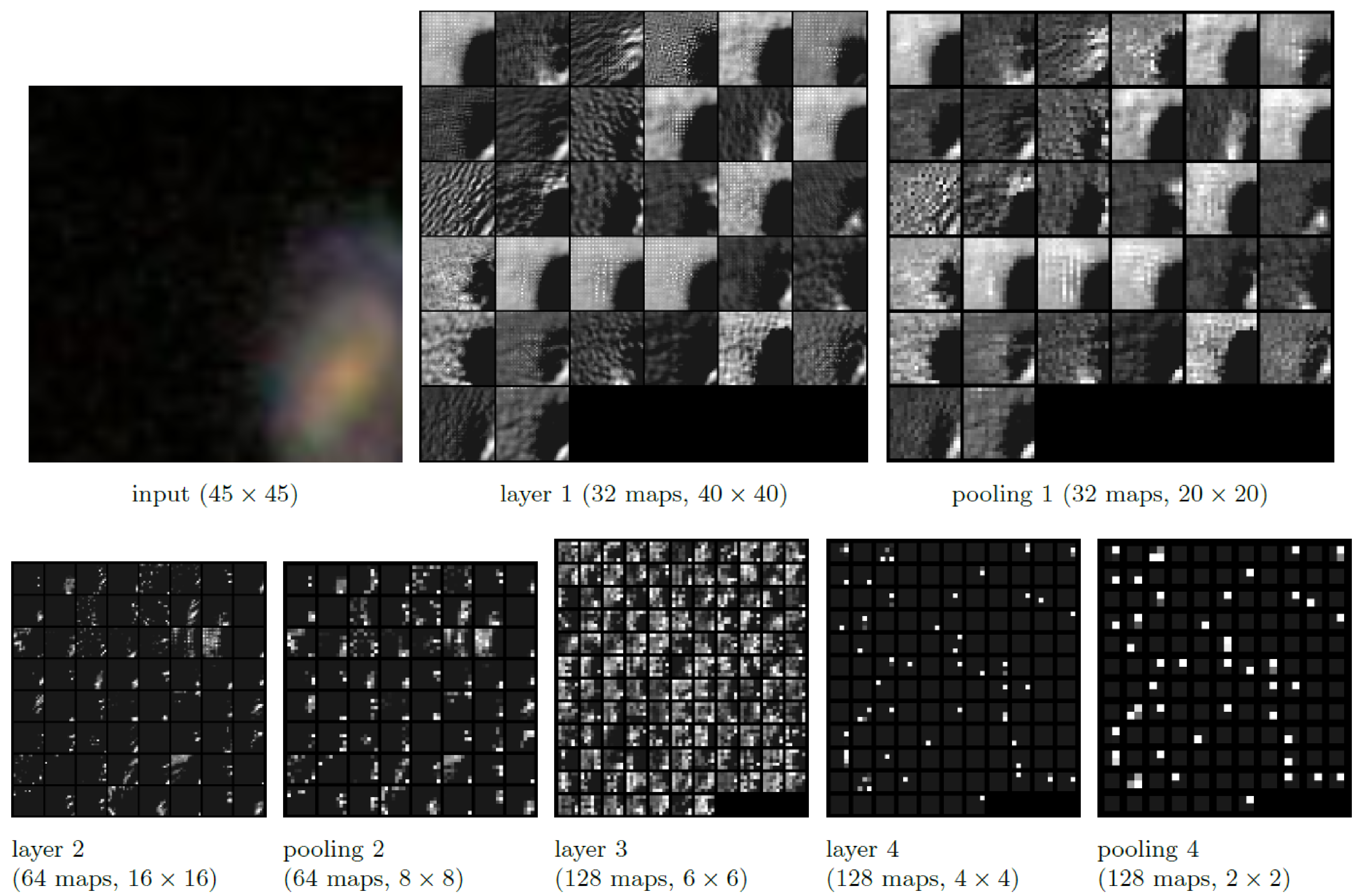
- Research funding by GAFAs motivated by the monetization of big data

Hierarchy of representations



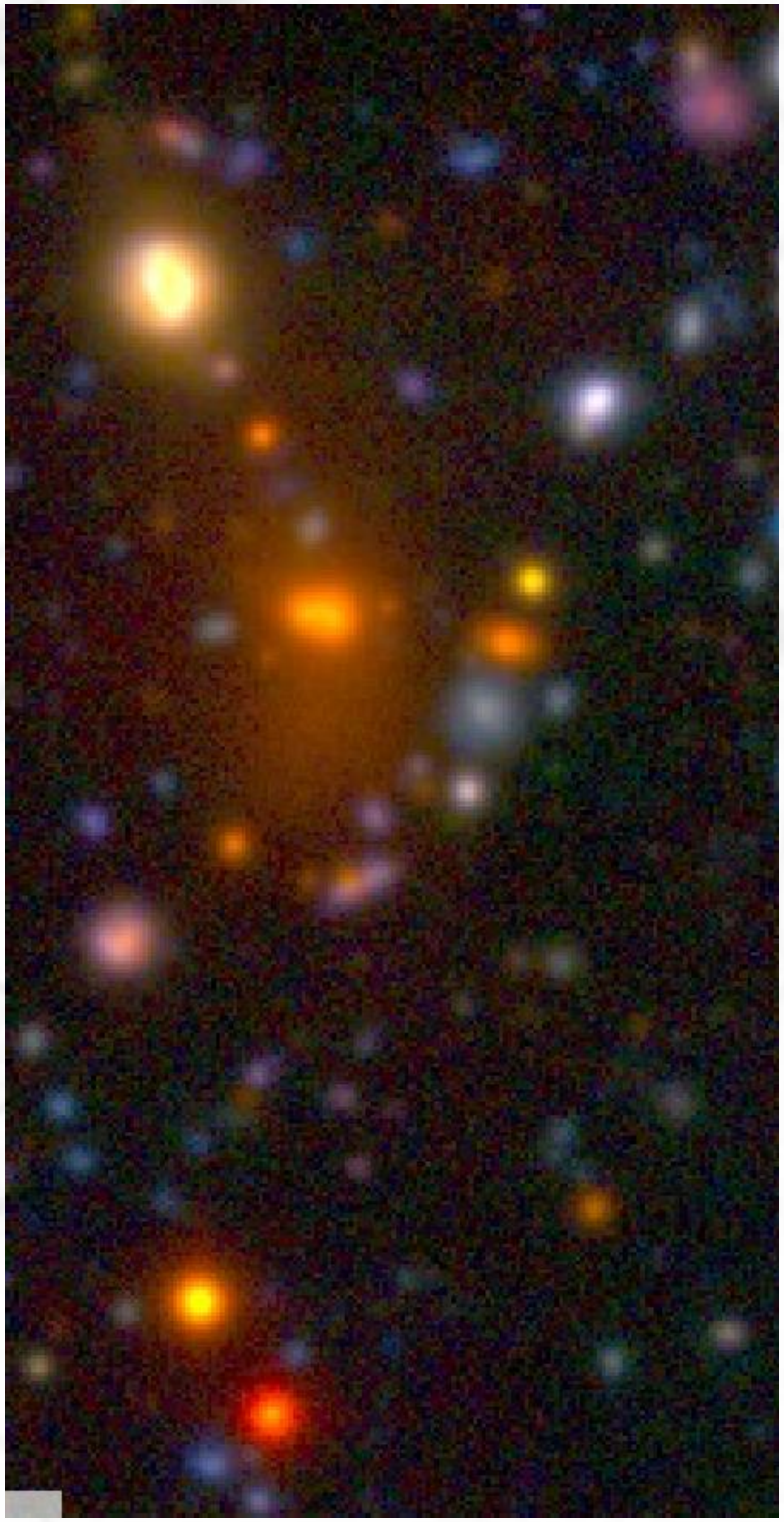
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Application to the Kaggle Galaxy Challenge (Dieleman et al. 2015)



Training with astronomical images

- Deep sky images:
  - Multispectral data are now commonplace
  - Lighting and projection effects not dominant
  - Fuzzy and noisy (SNR~1)
  - Spectral distributions and image formation processes (PSF + noise) can generally be accurately modeled
  - ➢ Augmented reality is an efficient approach

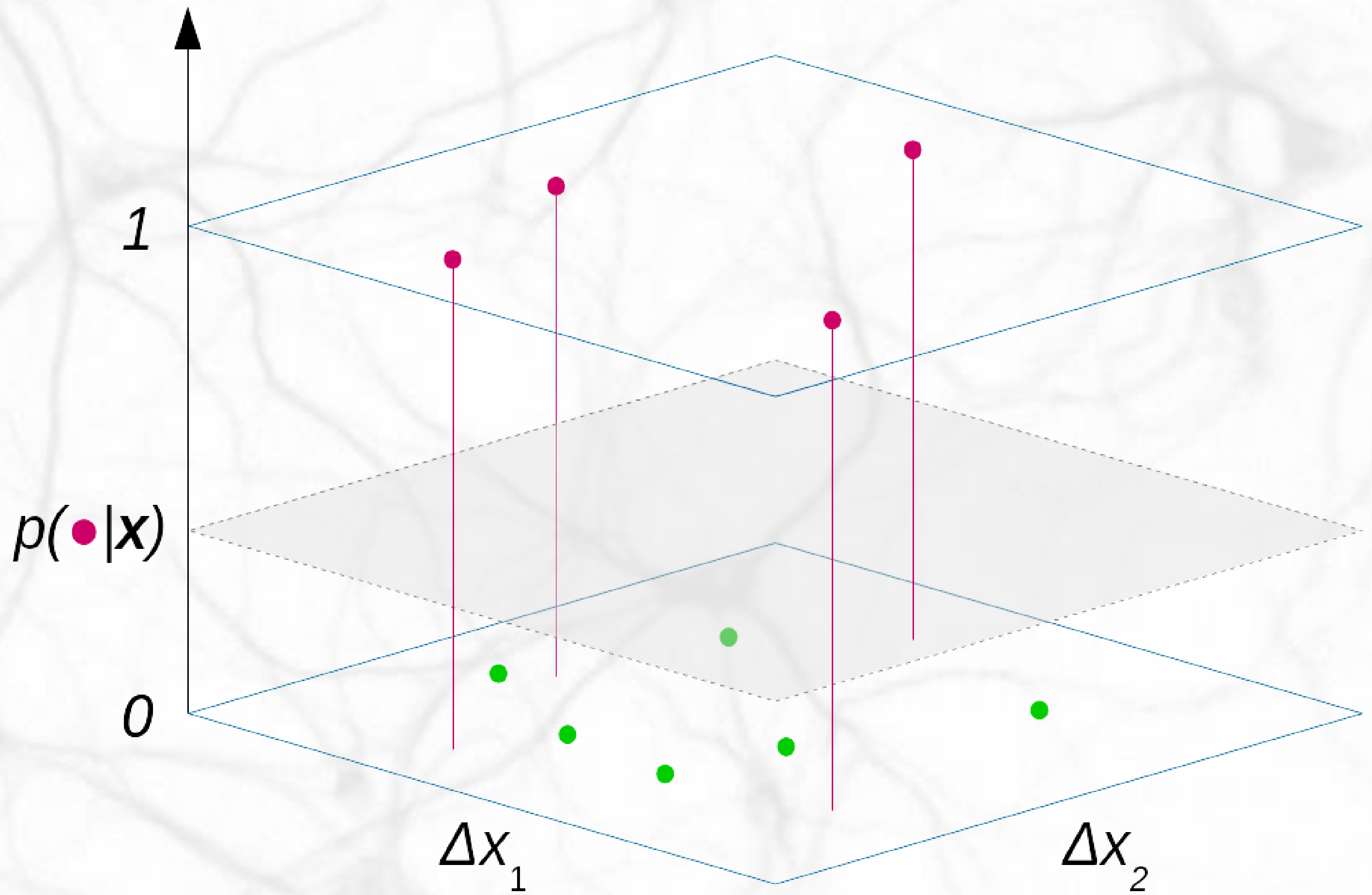


Hudelot et al. 2012

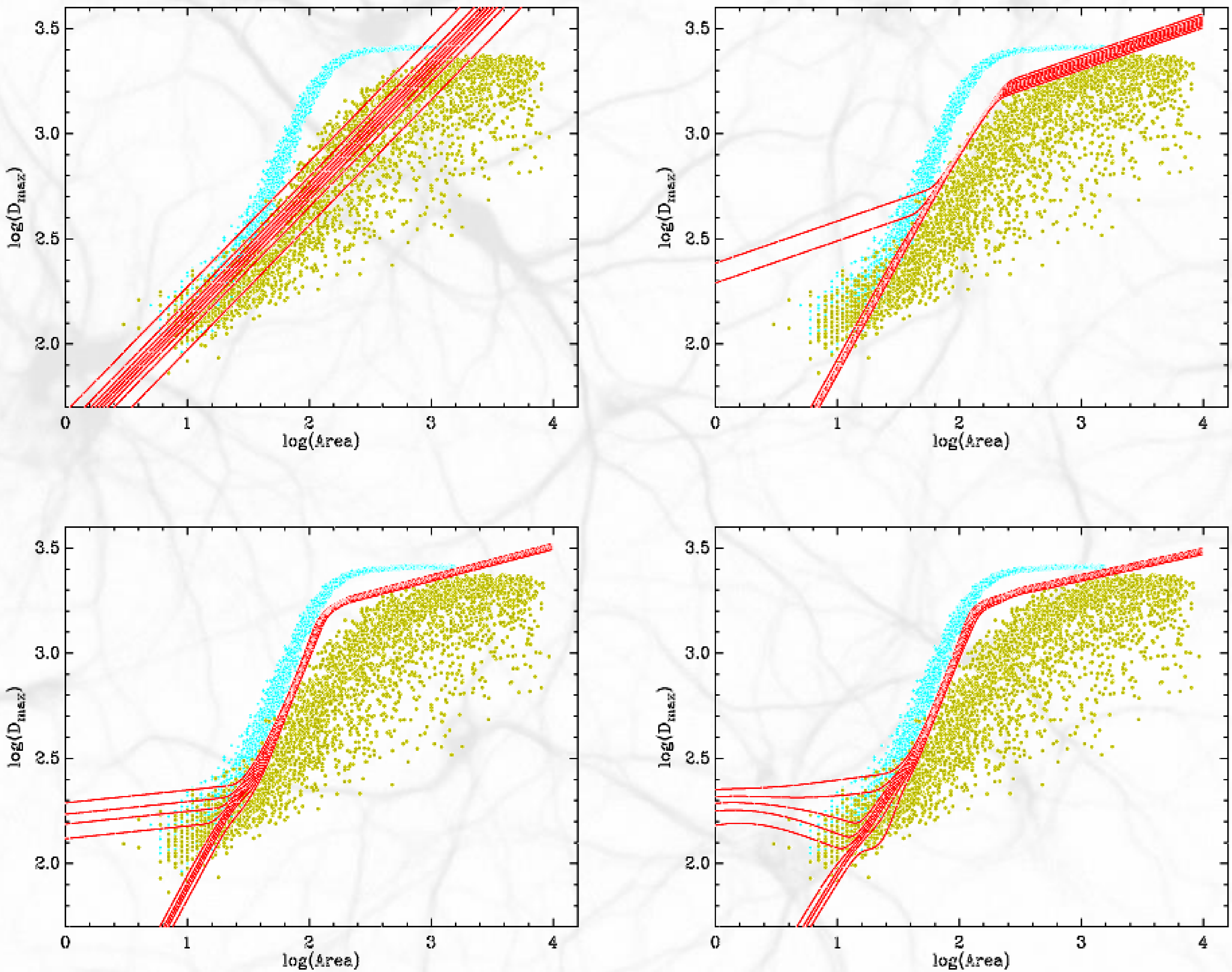
Estimating posterior probabilities

- The output of <u>perfectly trained</u> neural network classifiers can provide direct estimation of the posterior class probabilities (Pearlmutter & Hampshire 1990, Richard & Lippmann 1991, Miller et al. 1991, Rojas 1996, …)
  - Valid for a large range of cost functions (including quadratic and cross-entropy)
  - Requires a sufficiently powerful model with excellent generalization abilities
  - Both a blessing and a curse!
    - Importance of sample selection for training
    - Strong class imbalance shifts decision boundaries
    - “Hidden” priors

$$P(c_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|c_i)P(c_i)}{\sum_j p(\boldsymbol{x}|c_j)P(c_j)}$$



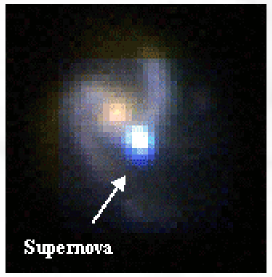
Example



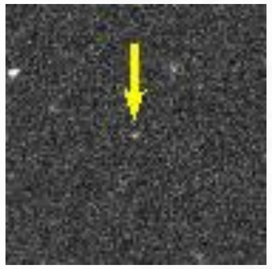
Class imbalance

- Class imbalance is often huge in representative samples
  - Interesting objects among regular ones: $\mathrm{P}{\sim}10^{-6}$ to ${\sim}10^{-5}$:

  - Strong lenses (Jacobs et al. 2018)



  - Galaxies with a detectable SN (Smartt et al. 2018)



  - Trans-Neptunian Objects (Gladman et al. 1998)

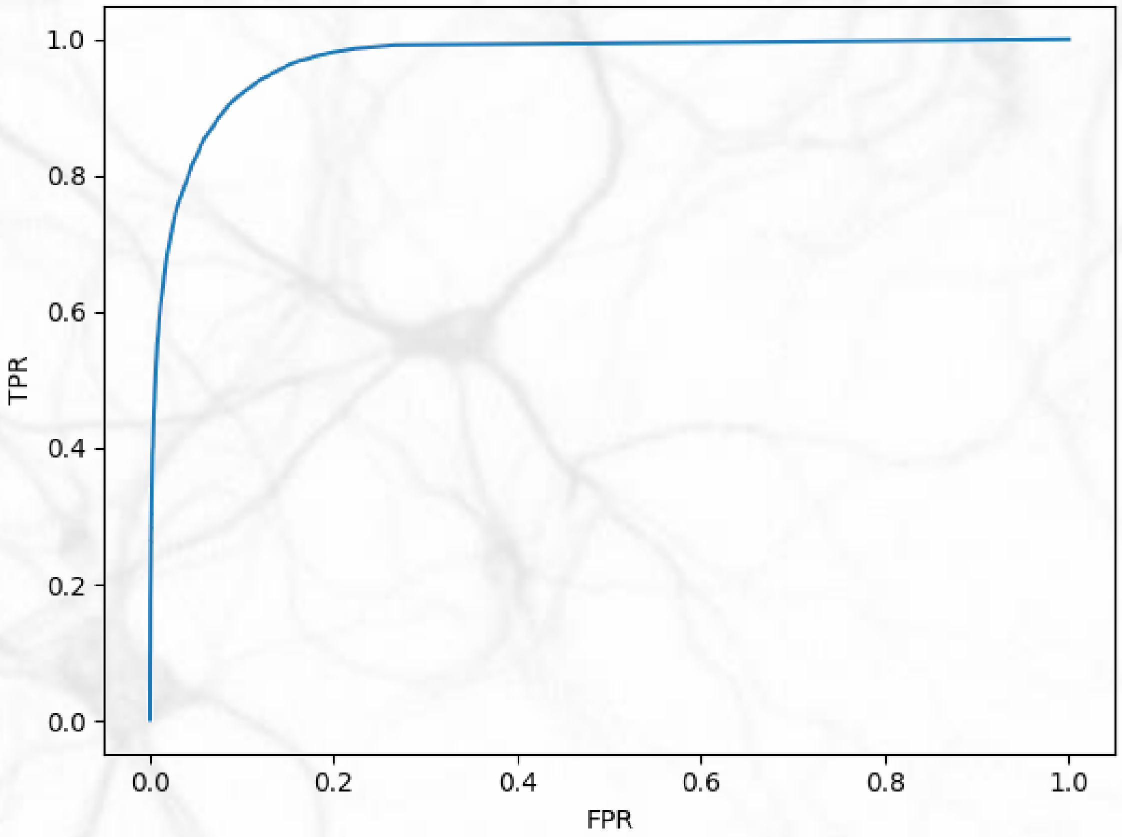
  - Even worse when considering blind searches in images: one often has $\mathrm{P}{\sim}10^{-7}$

Re-balancing training

- Rare events in excess in the training set: $P_T(c)$
- Correct output probabilities using the true prior $P(c)$:

$$P(c|\boldsymbol{x}) = \frac{P_T(c|\boldsymbol{x})P(c)}{P_T(c)\sum_{c'}\frac{P(c')}{P_T(c')}P_T(c'|\boldsymbol{x})}$$

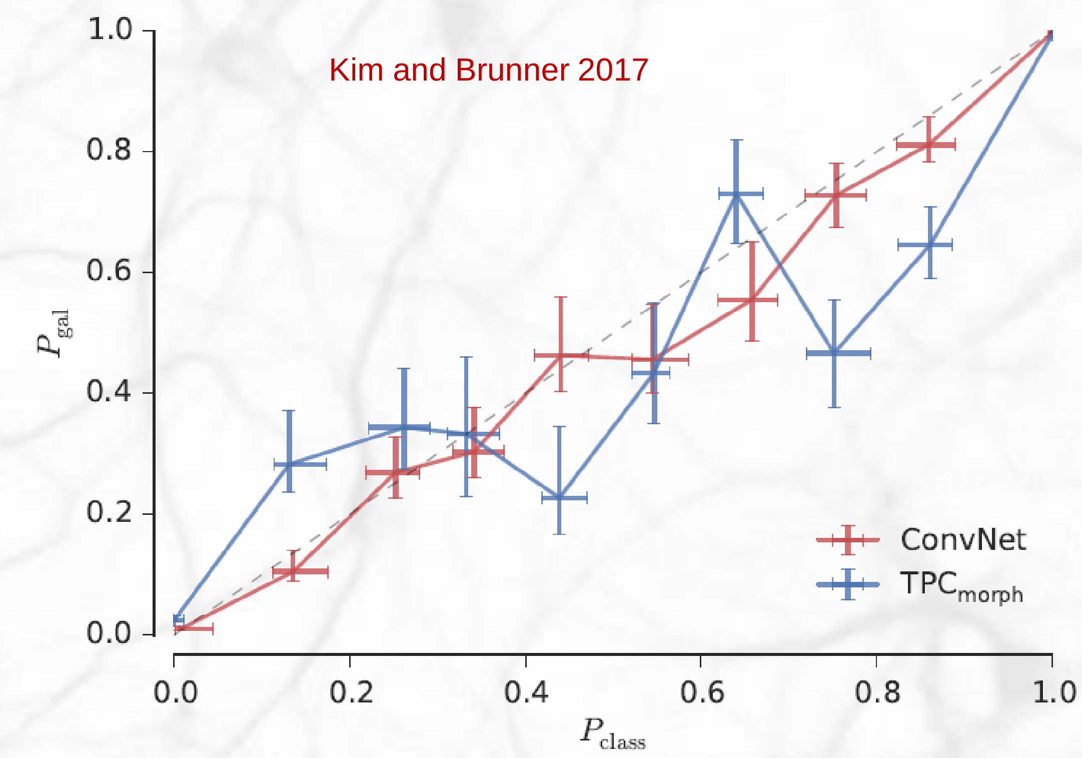
  - Works well in good training conditions (large samples, no mismatch)
  - Can be checked on ROC curves
- Or adjust thresholds directly from ROC curves



Estimating the accuracy of the posterior probabilities

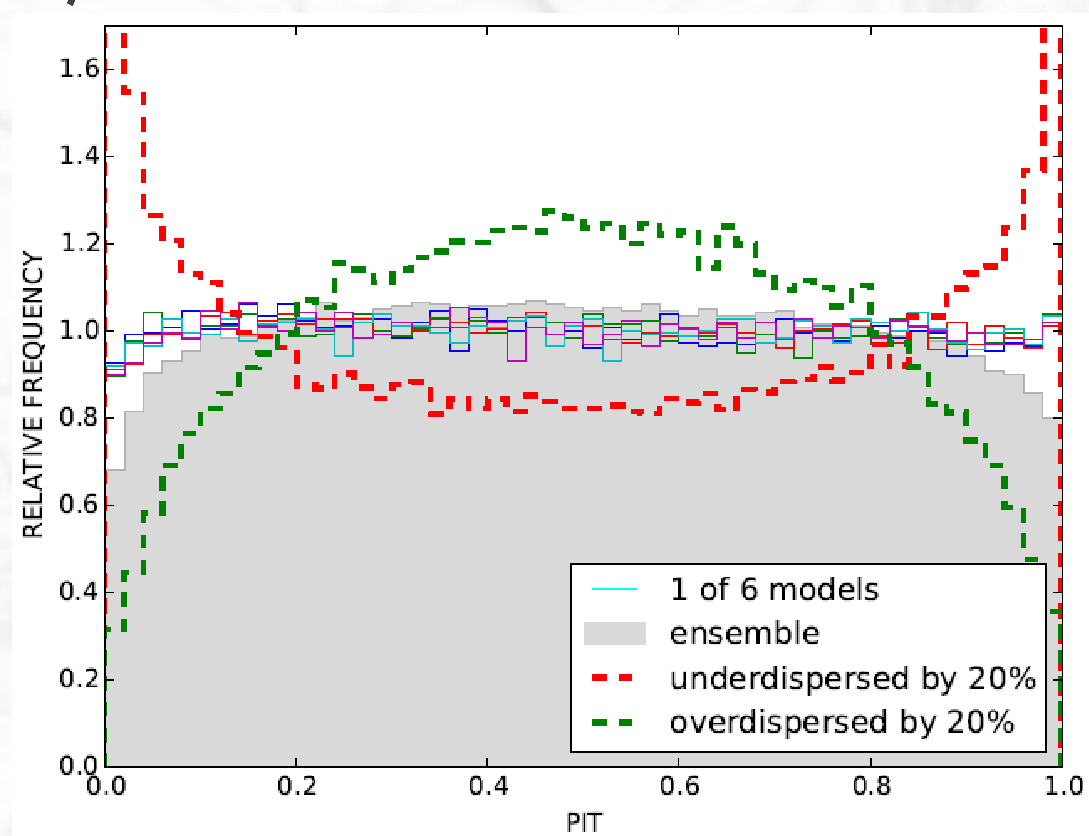
- Two mutually exclusive classes:

Check TPR in intervals of output $P(c|\boldsymbol{x})$



- Probability Integral Transform (Dawid 1984):

$$PIT(c_i) = \sum_{c_j = c_i} \sum_{c=1}^{c=c_j} P(c|\boldsymbol{x}_j)$$



Side note: dealing with the high dynamic range

- Most CNNs are meant to operate with images coming straight from JPEG, PNG, TIFF or even MPG files were the recorded pixel values are "gamma-compressed":

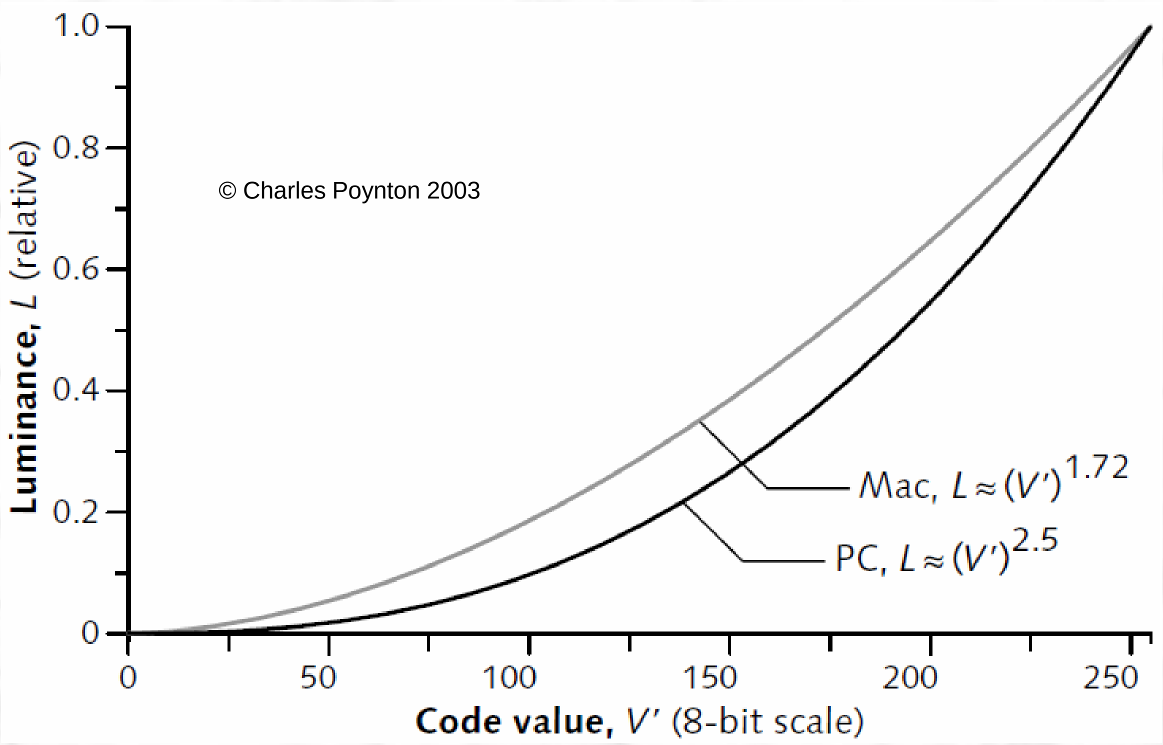
$$x \propto F^{1/\gamma} \text{ with } \gamma \approx 2.2$$

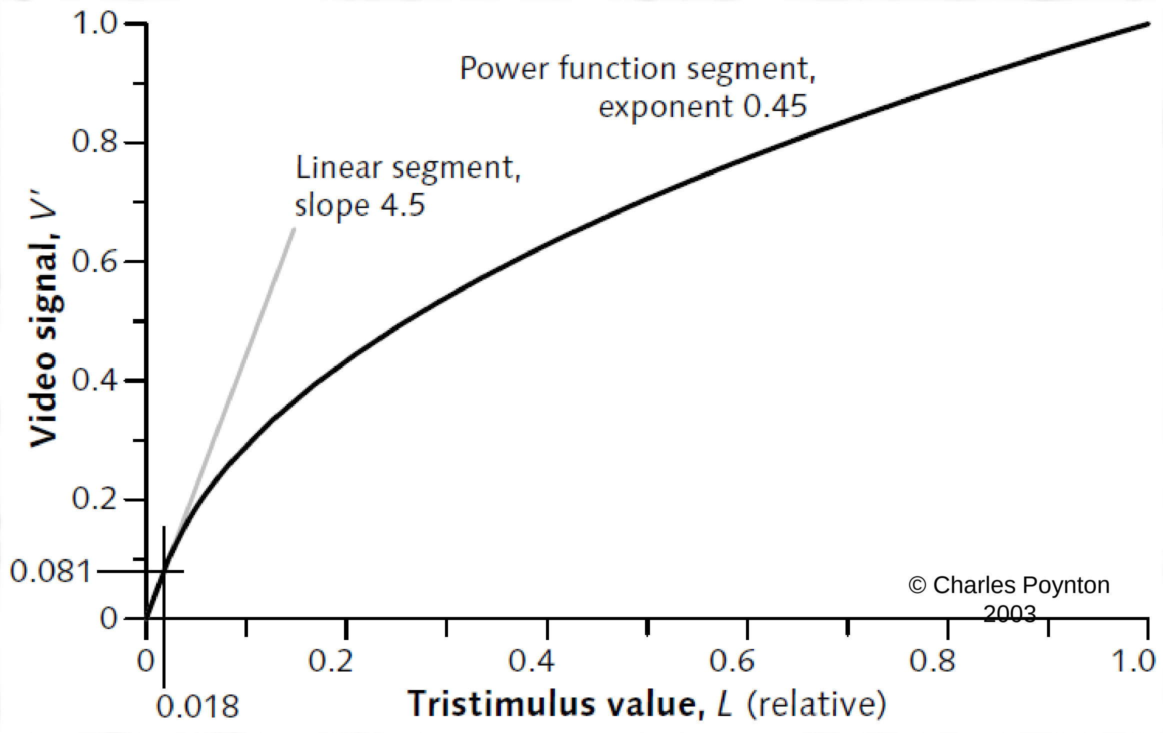
- Whereas in scientific image file formats (e.g., FITS), the recorded pixel values are proportional to the incoming flux:

$$x \propto F$$

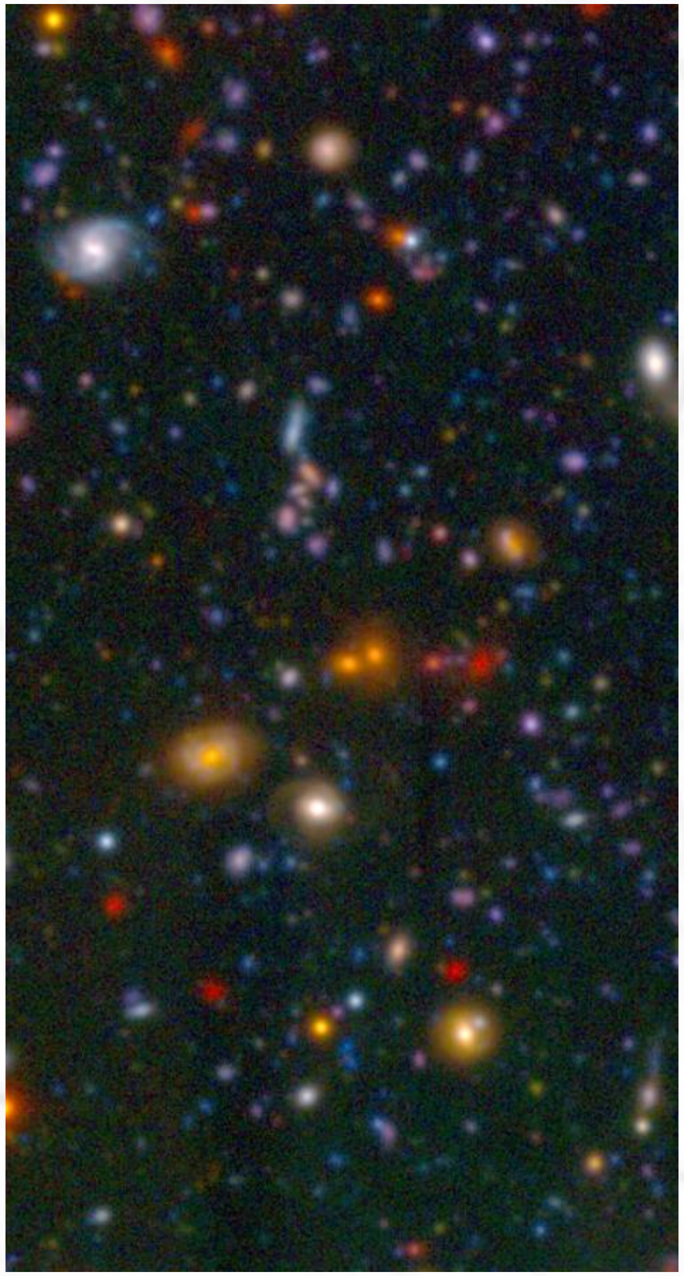
- To help with convergence it is often appropriate to compress the dynamic range using e.g.,

$$x = \operatorname{arcsinh}\left(\frac{F}{\sigma_F}\right)$$





Photometric redshifts



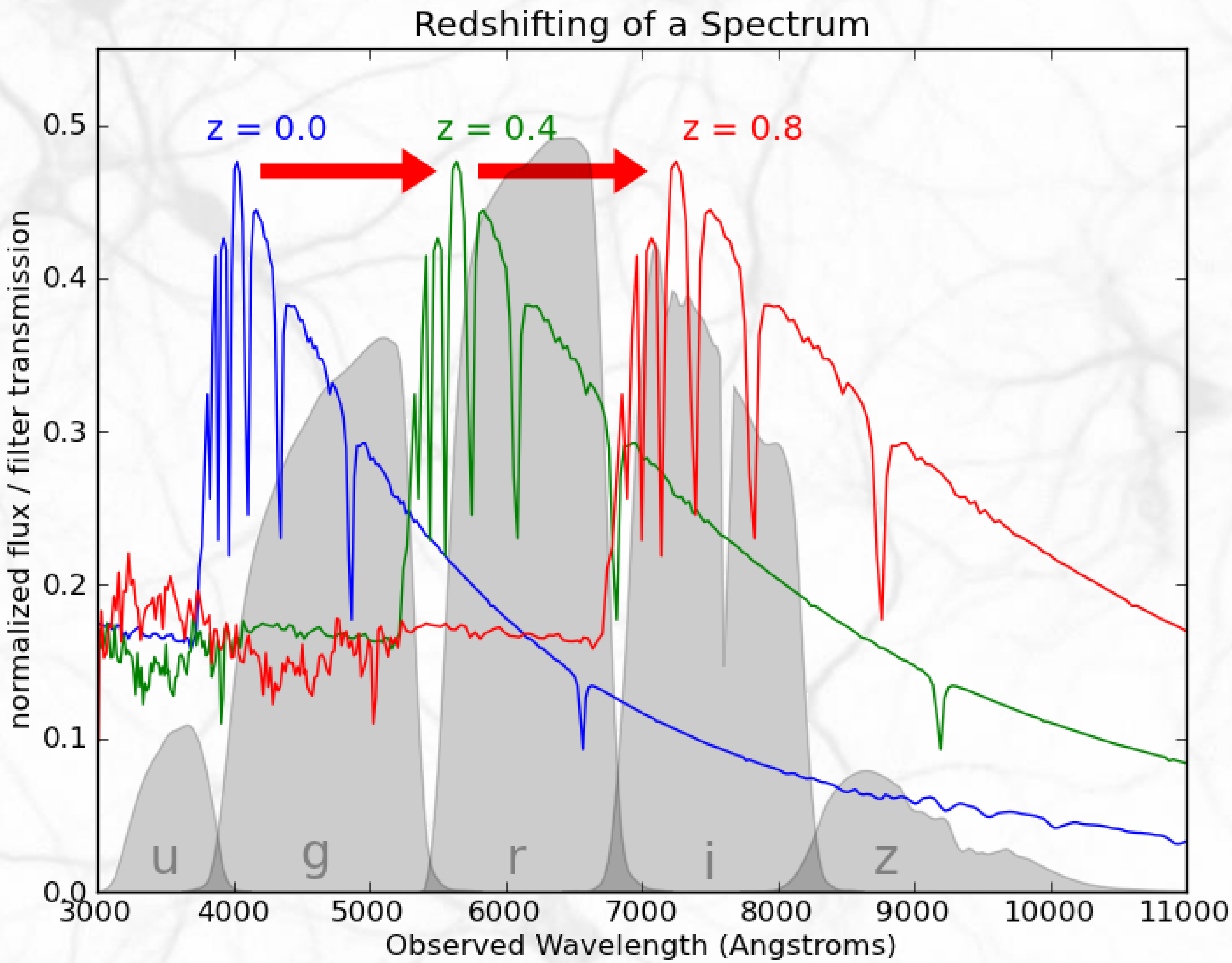


Photo-Z's using a CNN

J. Pasquet, EB, M. Treyer, S. Arnouts, D.Fouchez (Pasquet et al. 2018)

github.com/jpasquet/Photoz

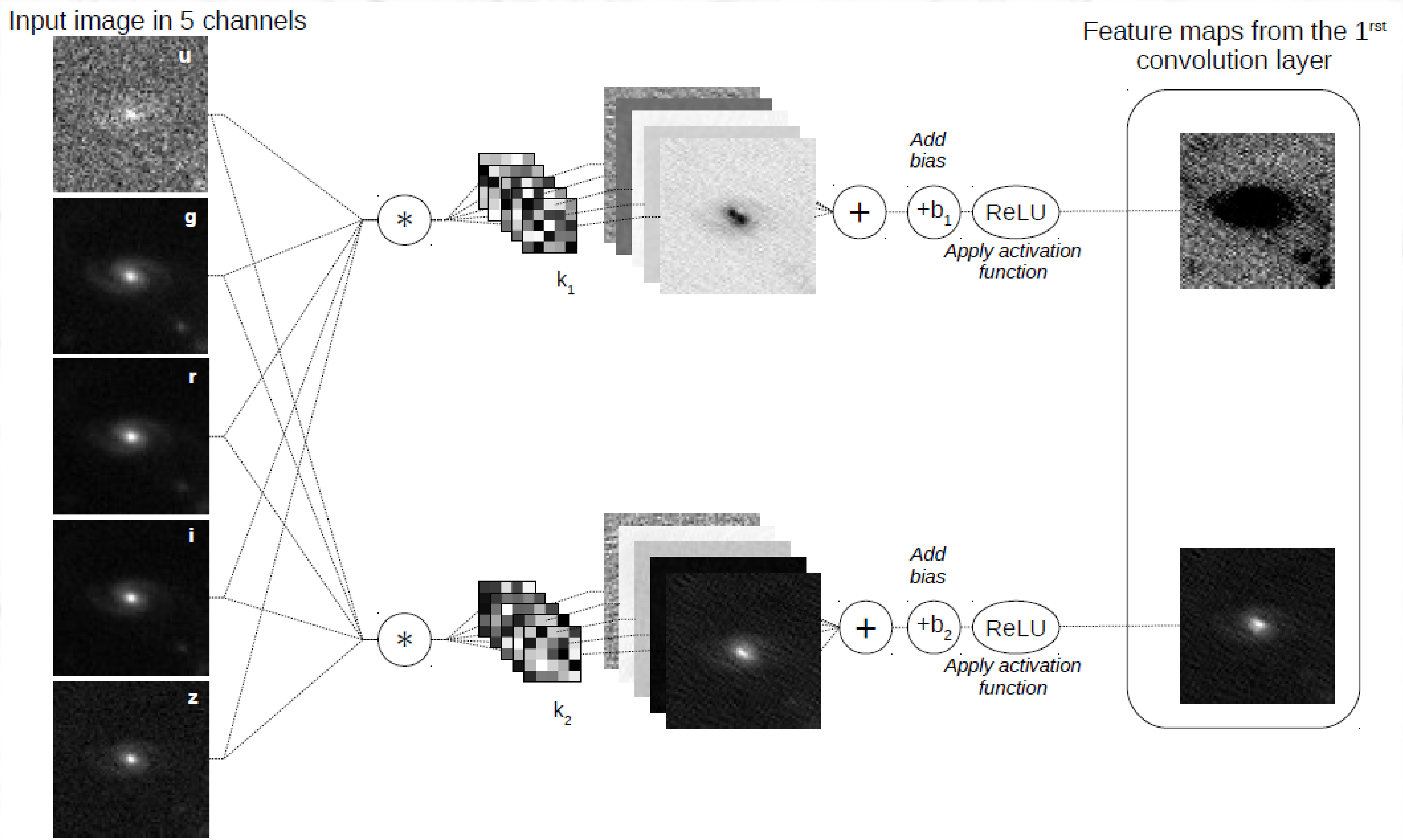


Photo-Z's using a CNN: Inception architecture

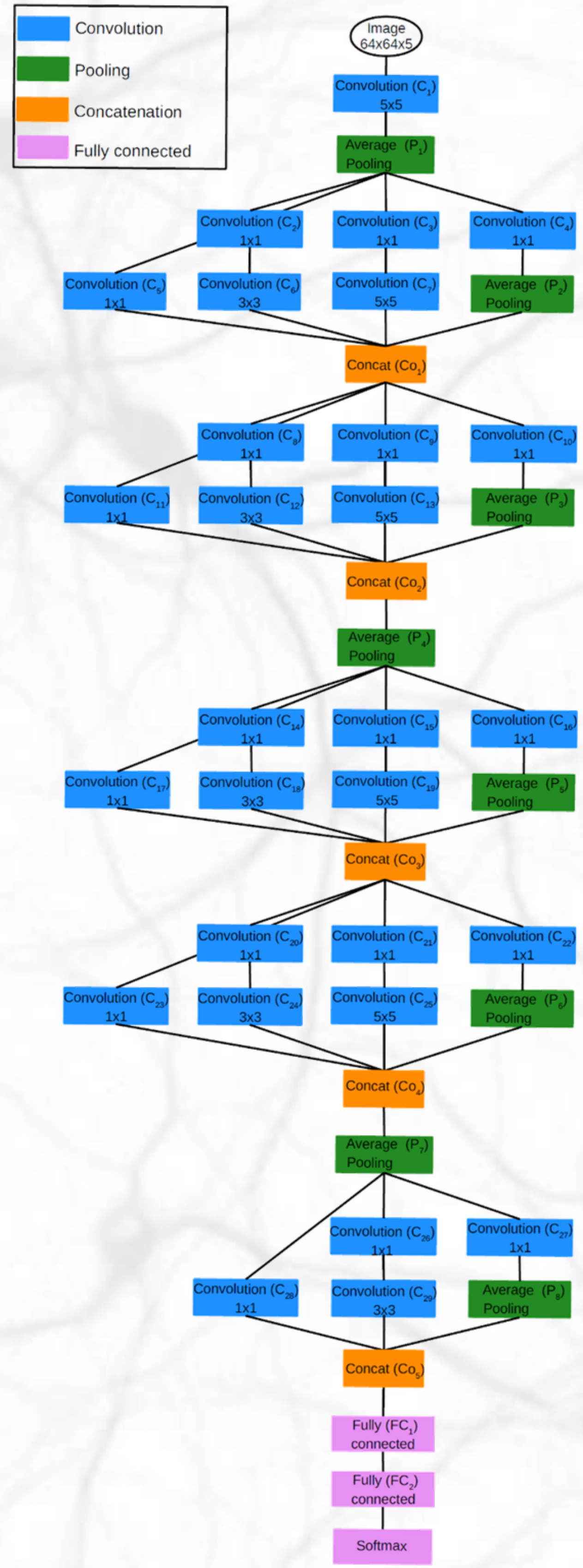


Photo-Z's using a CNN: results

Low z

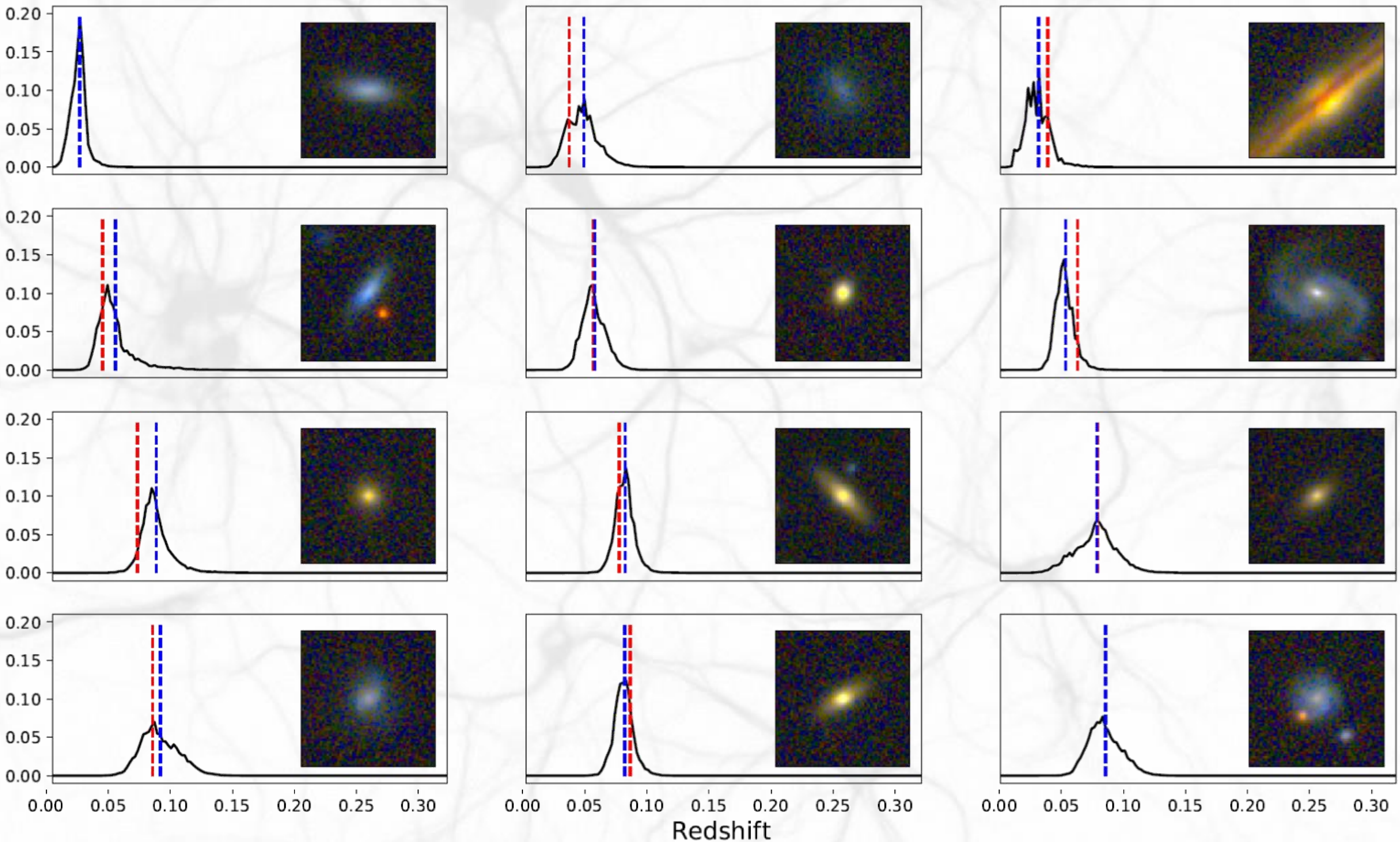


Photo-Z's using CNNs: results (cont)

High z

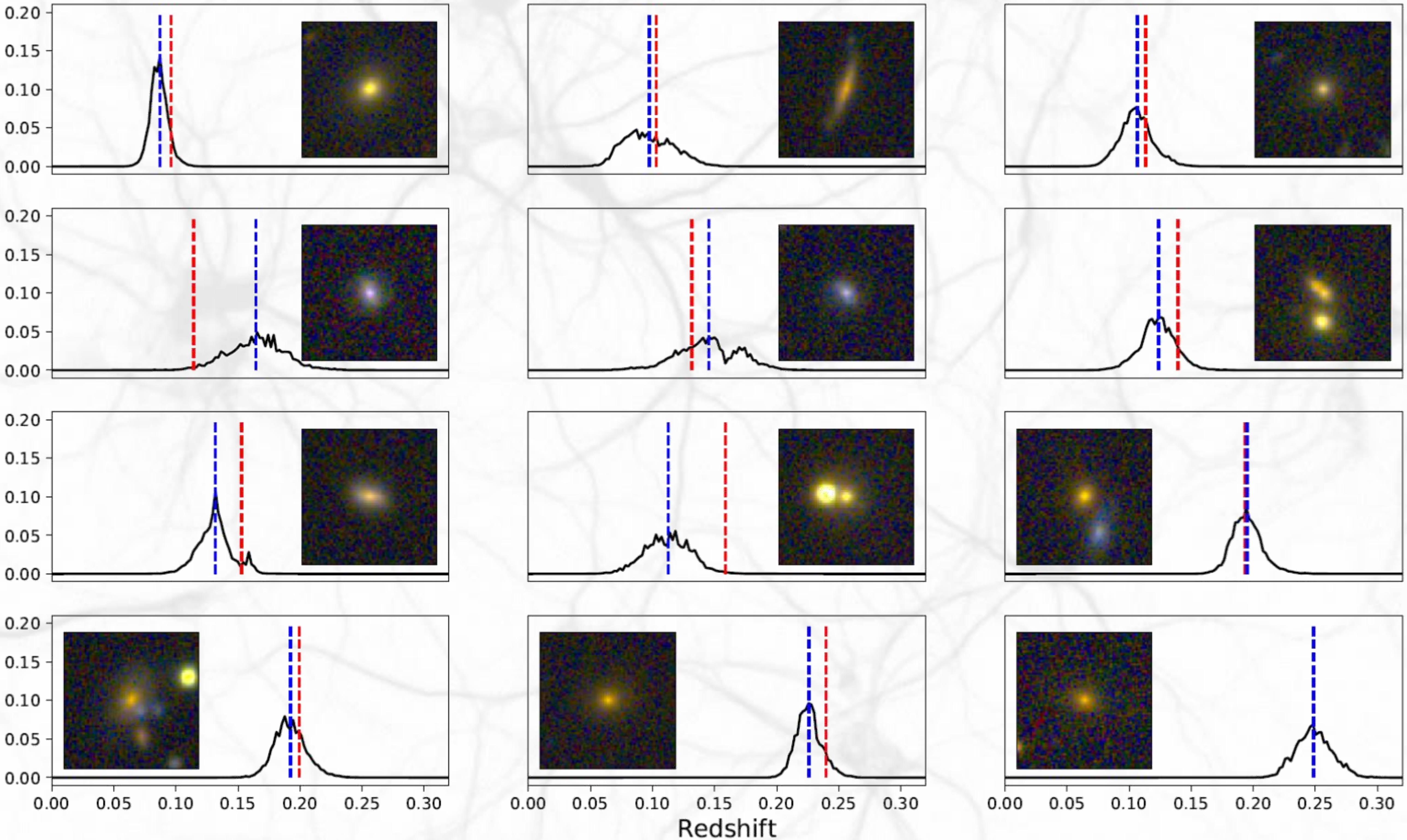


Photo-Z's using CNNs: results (cont.)

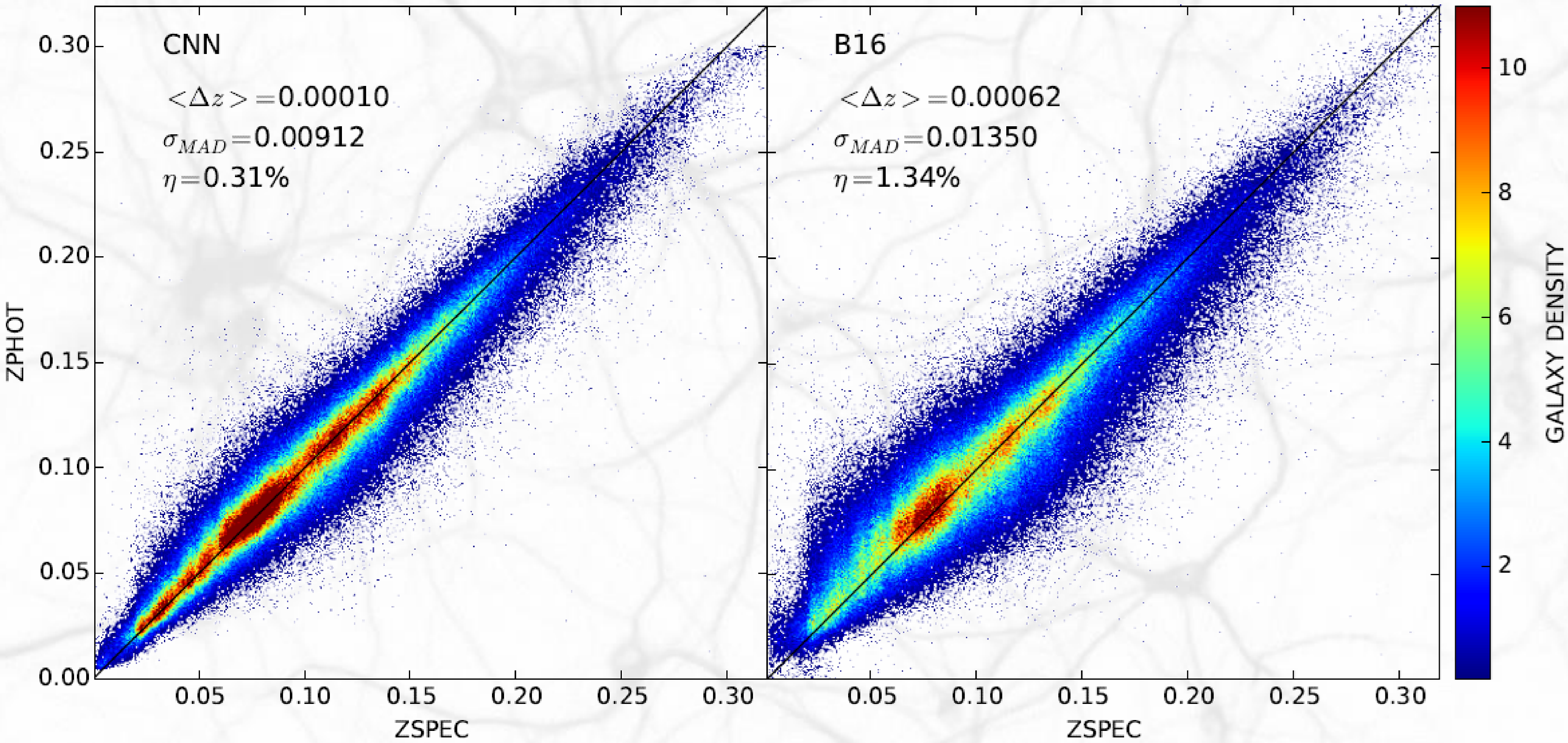


Photo-Z's using CNNs: effect of neighbors

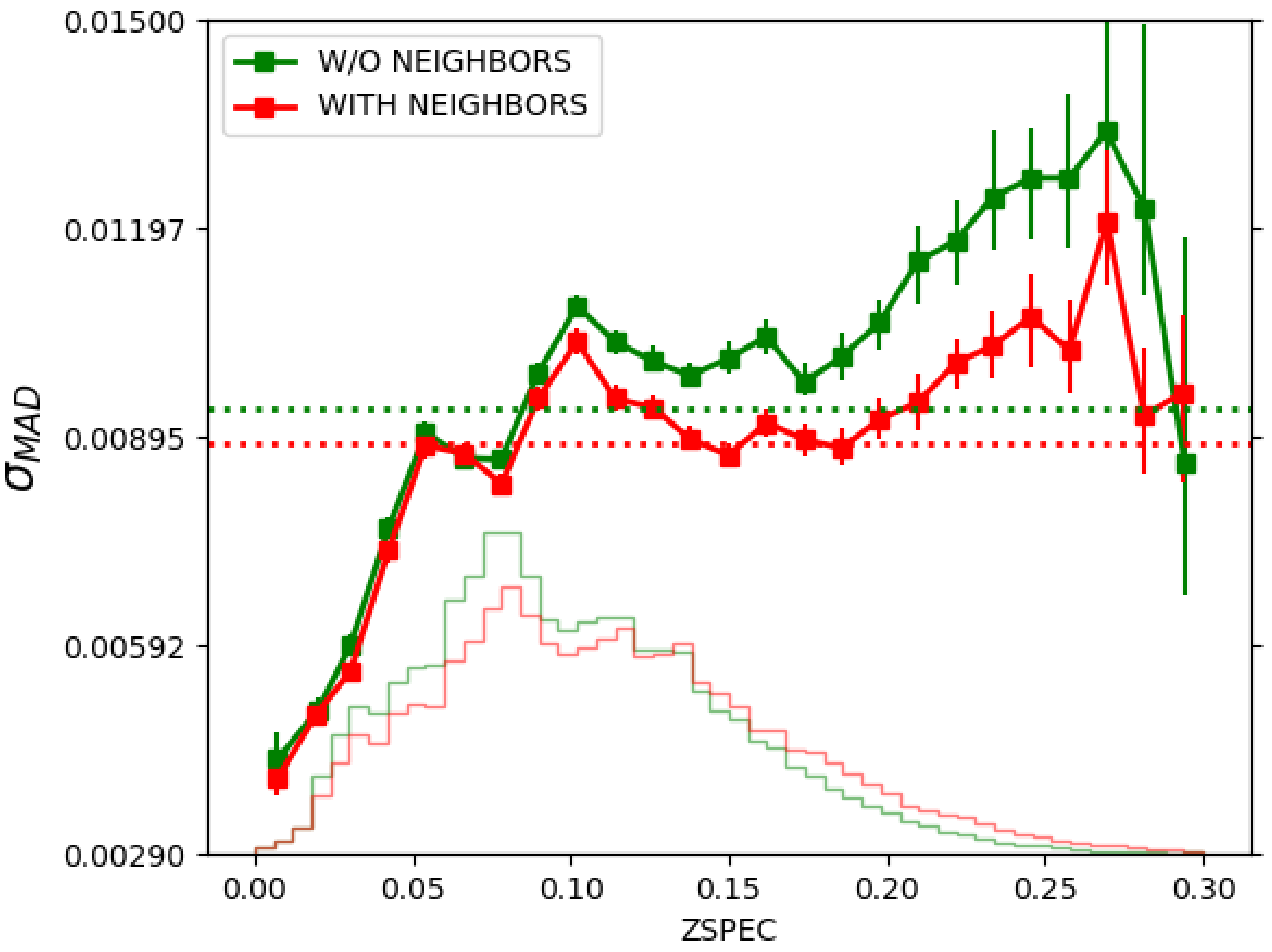
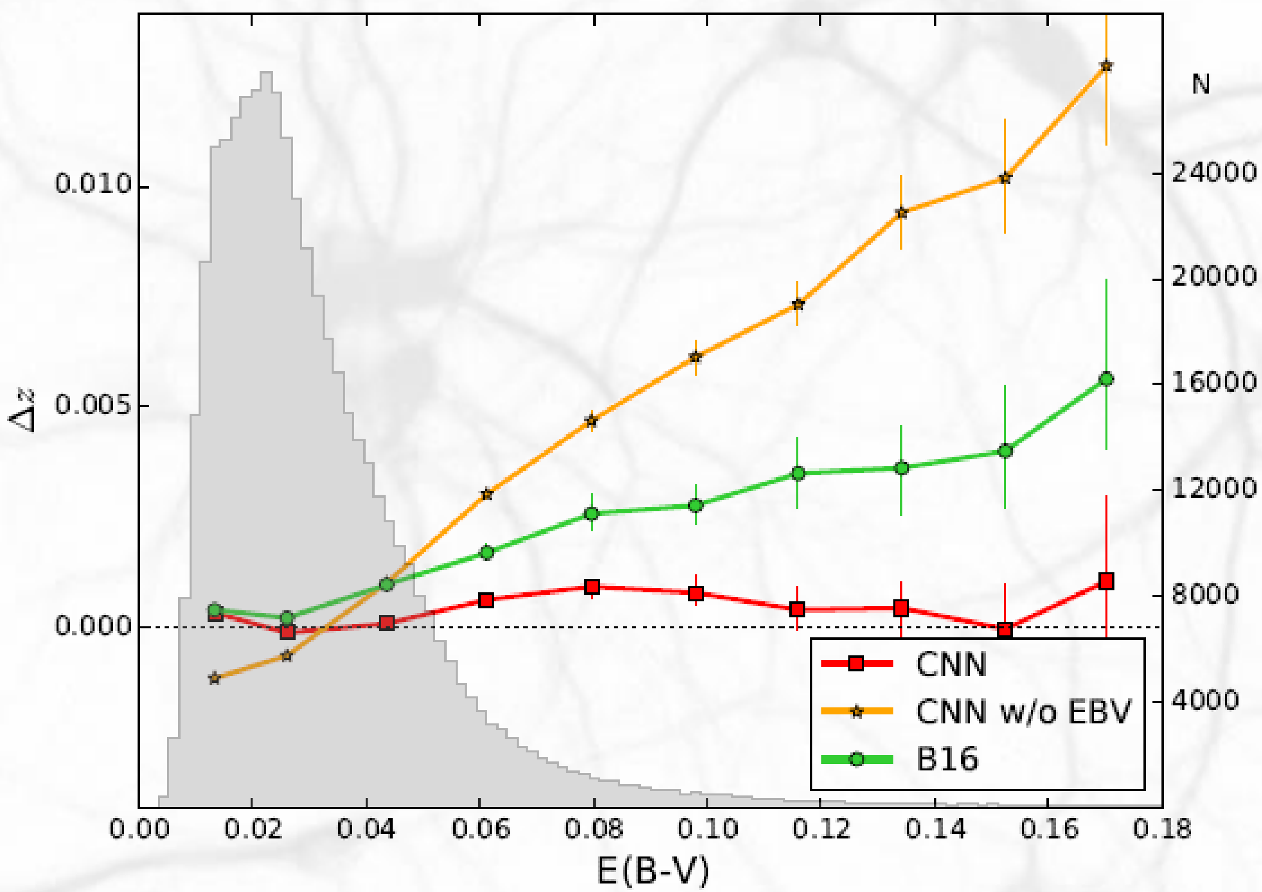


Photo-Z's using CNNs: biases (cont.)



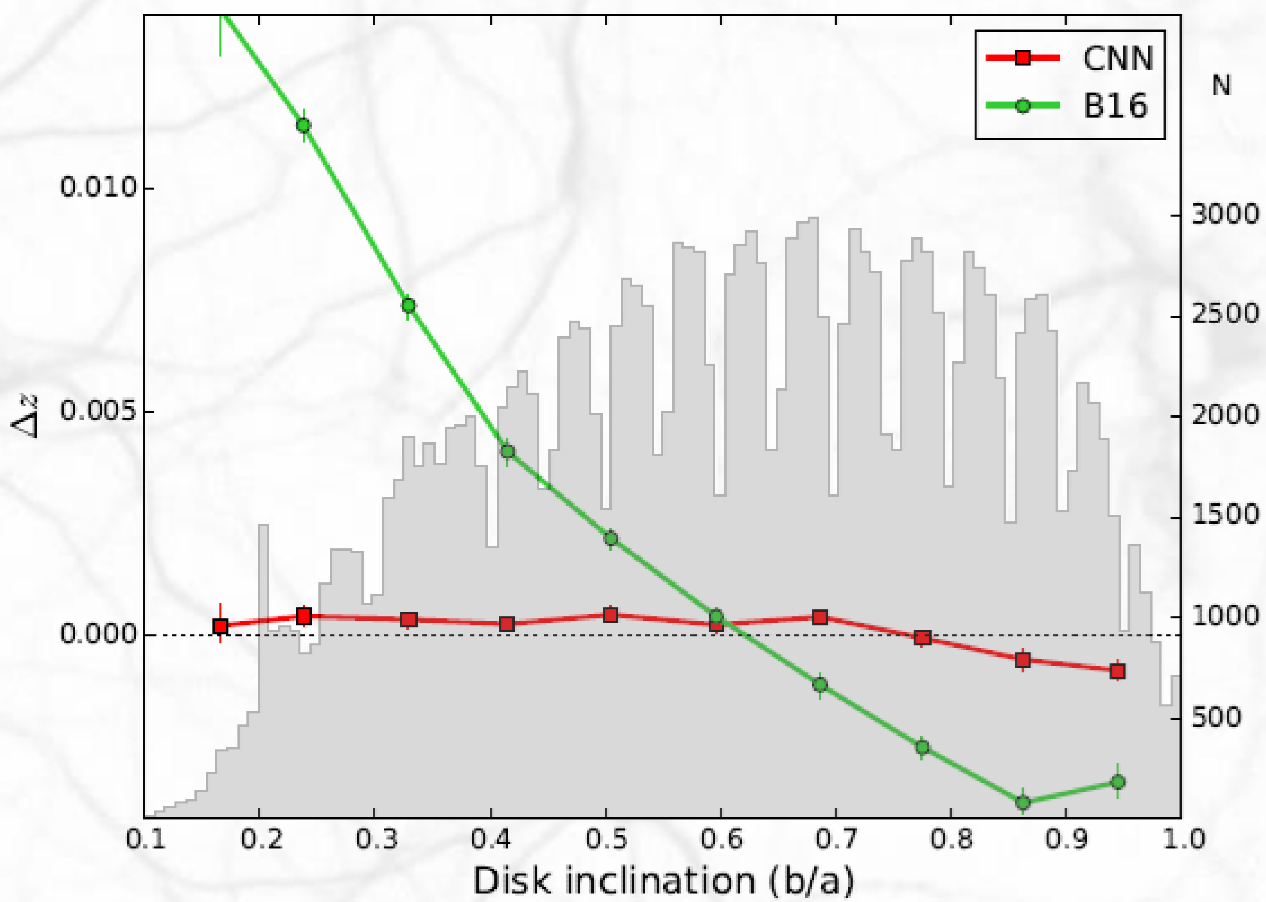


Photo-Z's using CNNs: dispersion

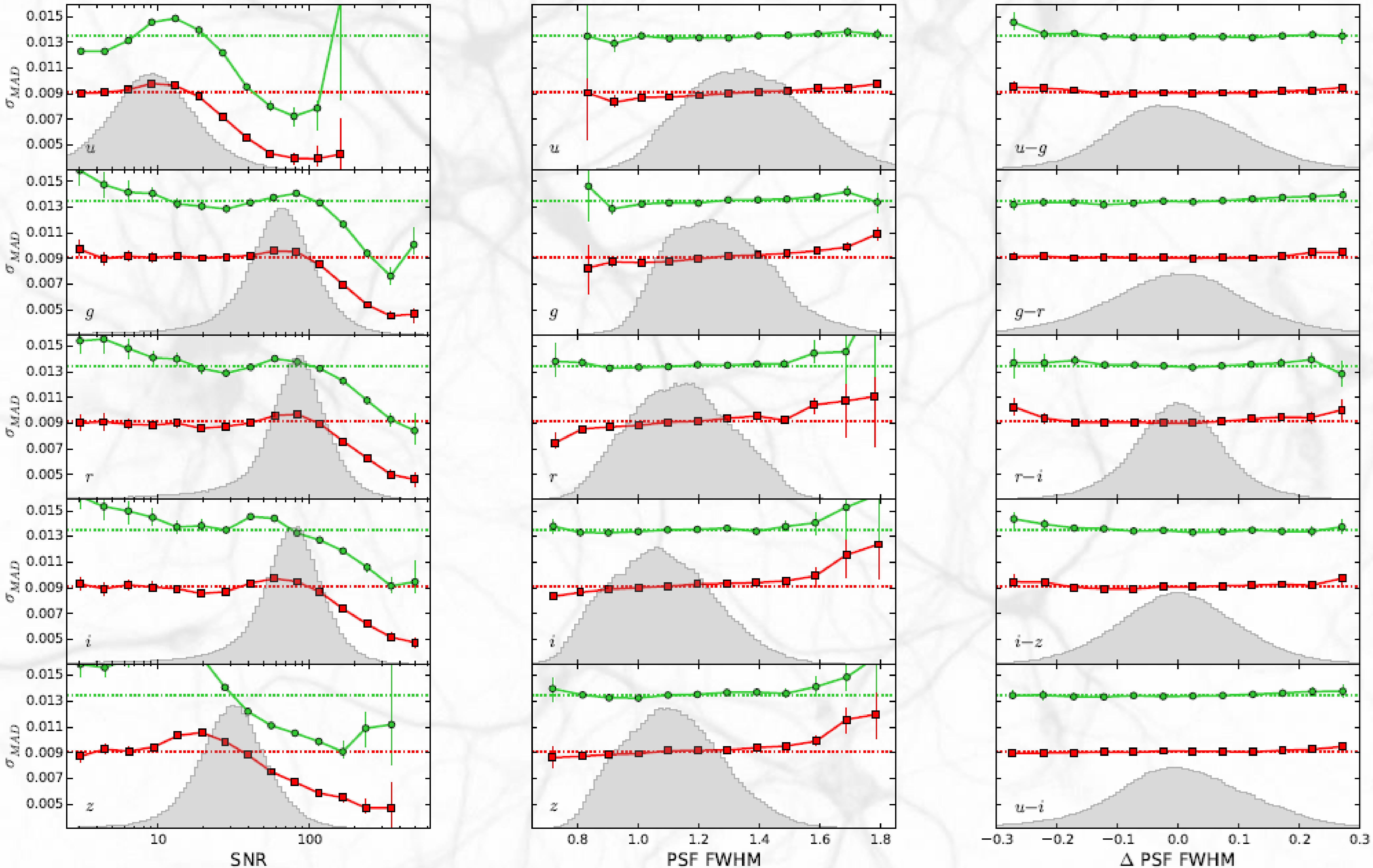
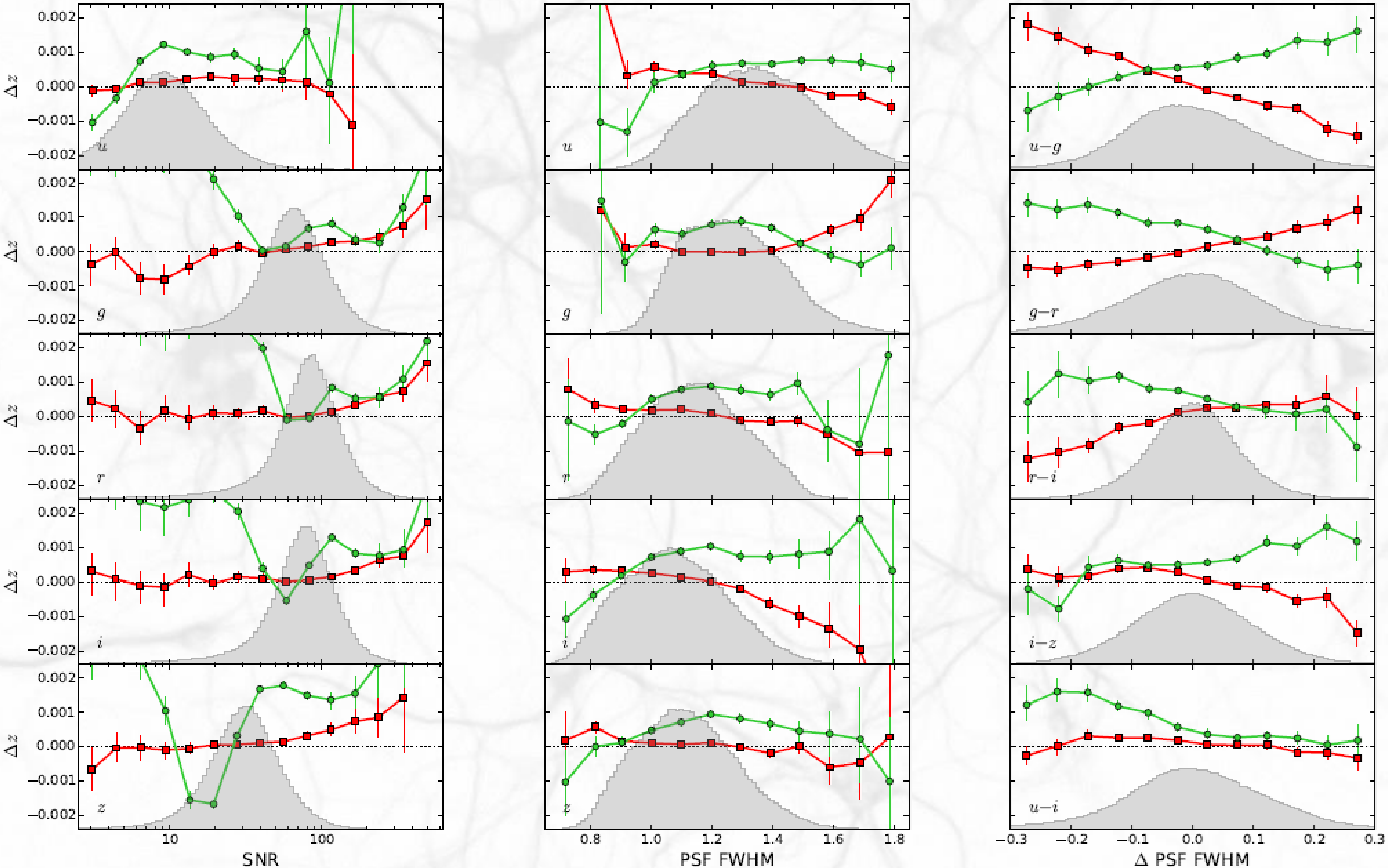


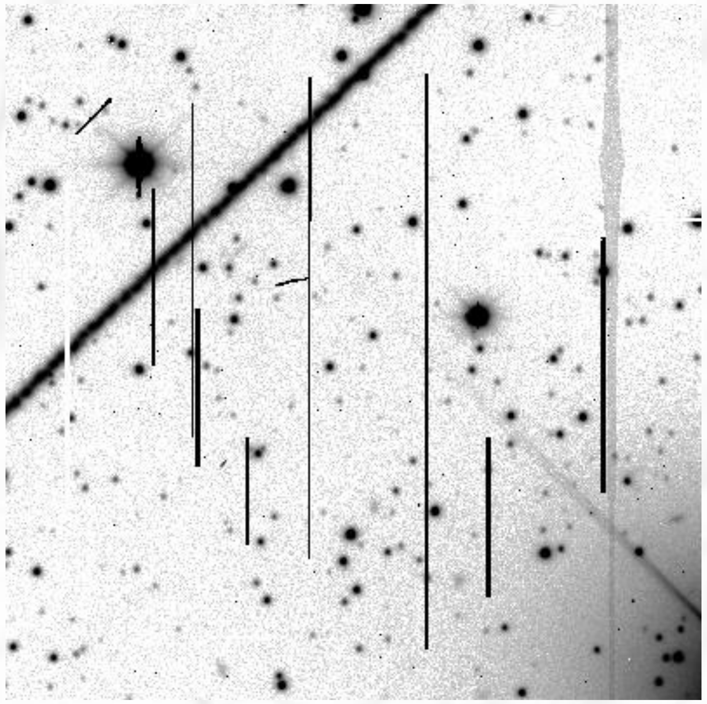
Photo-Z's using CNNs: biases (cont.)

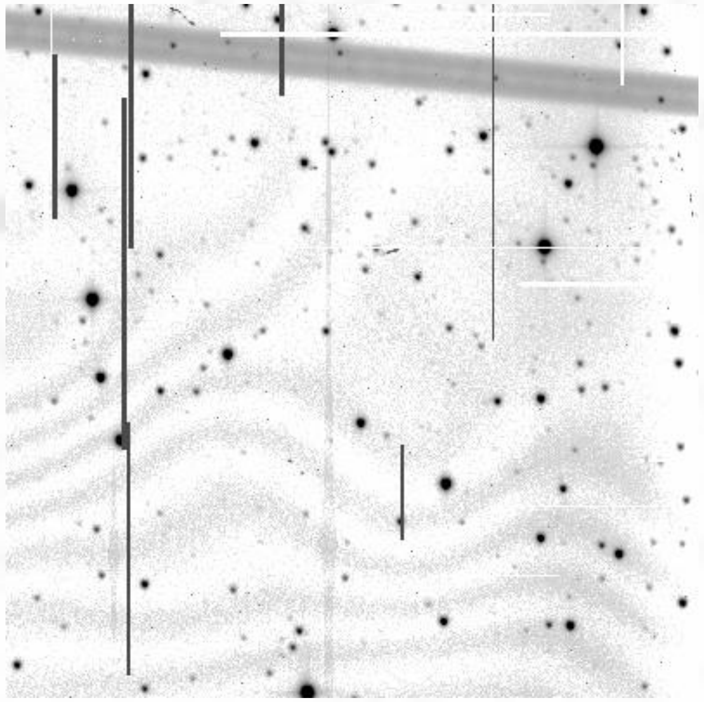


Detection of defects in astronomical images

M.Paillassa, EB, H.Bouy (Paillassa et al. 2019)
github.com/mpaillassa/MaxiMask

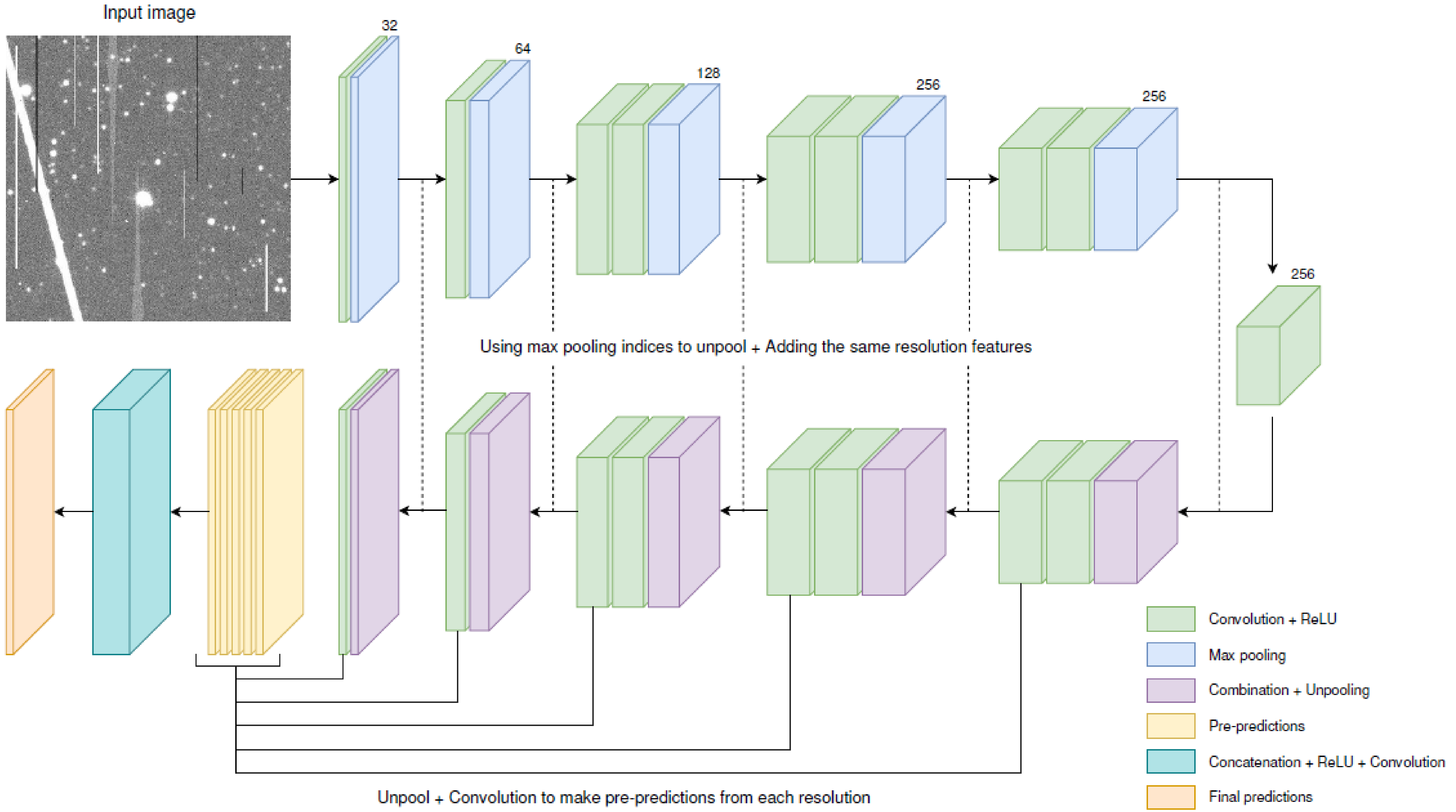
- Petabytes of astronomical image data now freely available online
  - Invaluable for studies of the variable sky
  - Inhomogeneous image quality
    - Many images hardly usable
  - Metadata often missing
- Efficient defect mapping/correction engine have been developed for large imaging survey pipelines
  - Highly tuned
- Here comes **MaxiMask**!



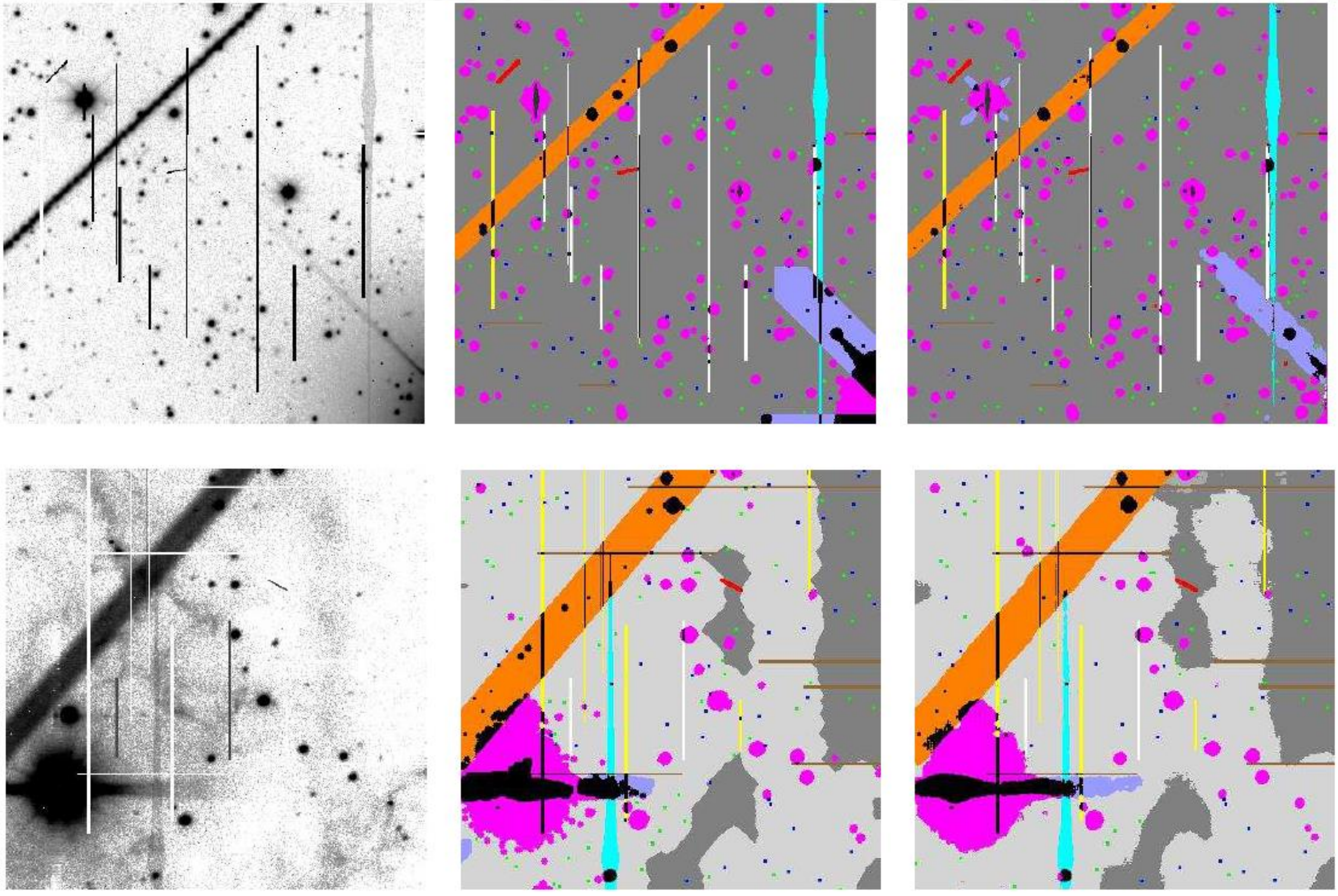


MaxiMask: architecture

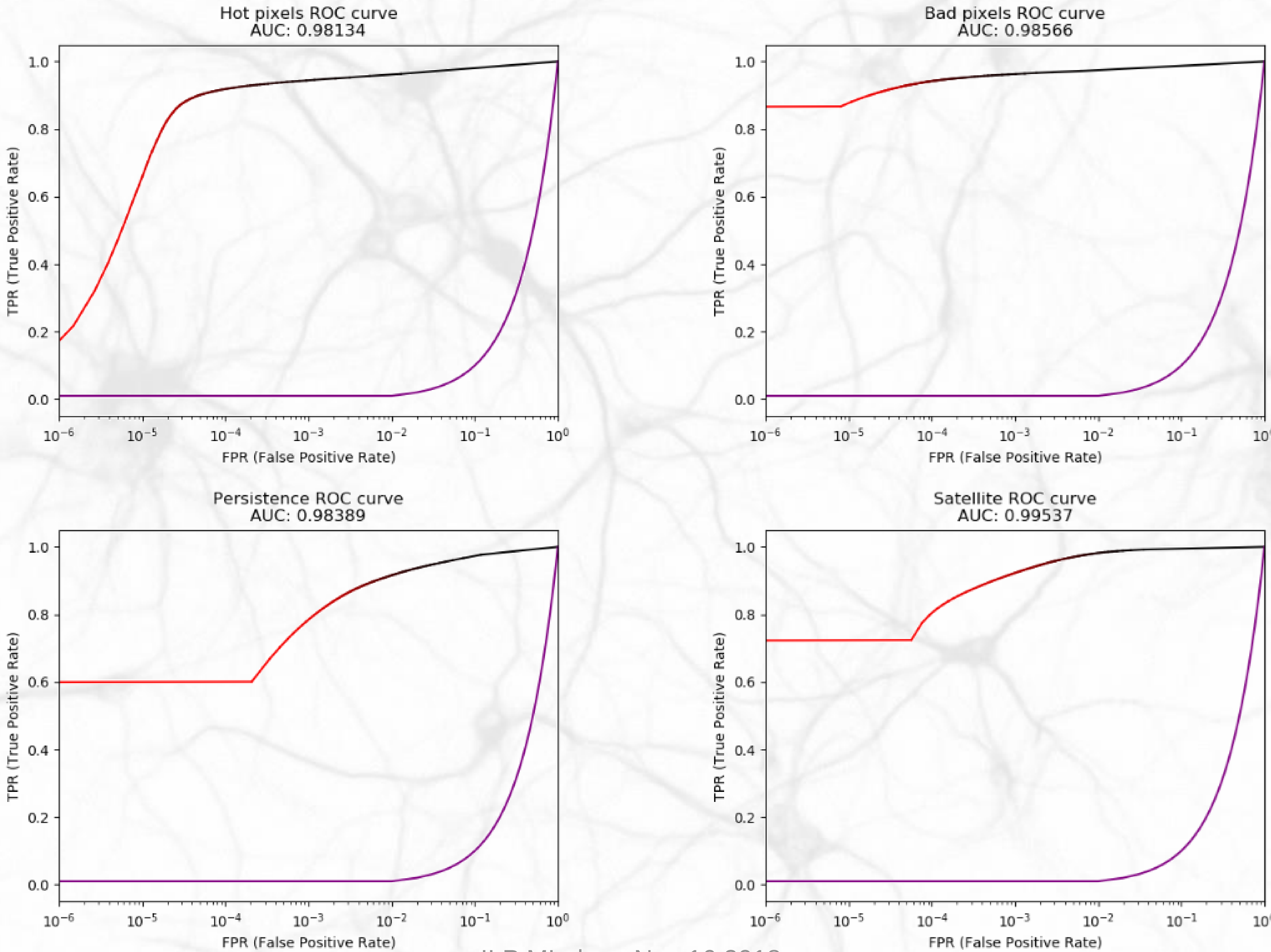
- Derived from Yang et al. (2018) « Highly fused » model



MaxiMask training



MaxiMask ROC curves



Conclusion

- The range of possible applications appears almost endless
- The classification performance of Deep CNN models significantly exceeds that of previous, state-of-the-art "traditional" astronomical image analysis algorithms
  - Working directly on pixel alleviates many difficulties
  - Properly trained models yields accurate PDFs
- Paradigm change in the way we process image data
  - Forward modeling all the way
  - Even "bigger data" than before
  - Knowledge of the physical and instrumental processes more useful than ever!