Estimating galaxy shape and flux with CNNs

CS 109b module, April, 2020

Doug Finkbeiner
Jun Yin

(Lecture 2)

Overall task (again)

- A galaxy image can be described by a simple model with 5 parameters (a brightness and 4 shape parameters).

- We want to estimate those parameters, given an image (or millions of them).

- We can generate mock images as a function of the 5 parameters to produce a training set.

Your task is to make a training set, train a CNN to do this regression, and then assess its performance on a sample of mock images.

Overall task (again)

- A galaxy image can be described by a simple model with 5 parameters (a brightness and 4 shape parameters).
- We want to estimate those parameters, given an image (or millions of them).
- We can generate mock images as a function of the 5 parameters to produce a training set.

Your task is to make a training set, train a CNN to do this regression, and then assess its performance on a sample of mock images.

Image noise

Real astronomical images have a lot of imperfections that make them different from a theoretical model.

- Noise: There is noise in the readout electronics of a CCD (charge-coupled device, the sensor used in most telescopes).
- Sky background: The night sky is not “dark” — there are many sources of light (reflected city light, airglow, moonlight, …)
- Photon counting is a Poisson process anyway.

Signal to Noise ratio

The uncertainty of your parameter estimate for a given galaxy depends on the noise per pixel in the image, and also on the model you are fitting. (We assume constant noise per pixel, true for the "sky dominated" case).

Assuming the model is correct (it is, because *we are fitting mock data generated with the same model*) then the S/N is basically the total signal over the noise in the pixels the galaxy occupies, weighted by the confidence they came from the galaxy.

Signal to Noise ratio

The signal to noise ratio is

$$SNR = \frac{1}{\sigma_p}\sqrt{\sum p_i^2}$$

Where p_i is the value in pixel i and
σ_p is the noise per pixel.

again, this is for faint objects. For bright objects, you must include the Poisson noise from the object itself, and the SNR scales as sqrt(total flux).

Cramér-Rao bound

The *Cramér-Rao bound* is the best performance of a minimum-variance unbiased estimator.

It answers the question “What is the best we can expect to do?” i.e. “What are the smallest errors we can expect?”

In assessing performance of an estimator (neural net or otherwise!) it is good to have some idea of what to expect. Then you know if you are doing “well” or not.

Cramér-Rao bound

The Cramér-Rao bound is related to the inverse of the second partial derivative of the log likelihood.

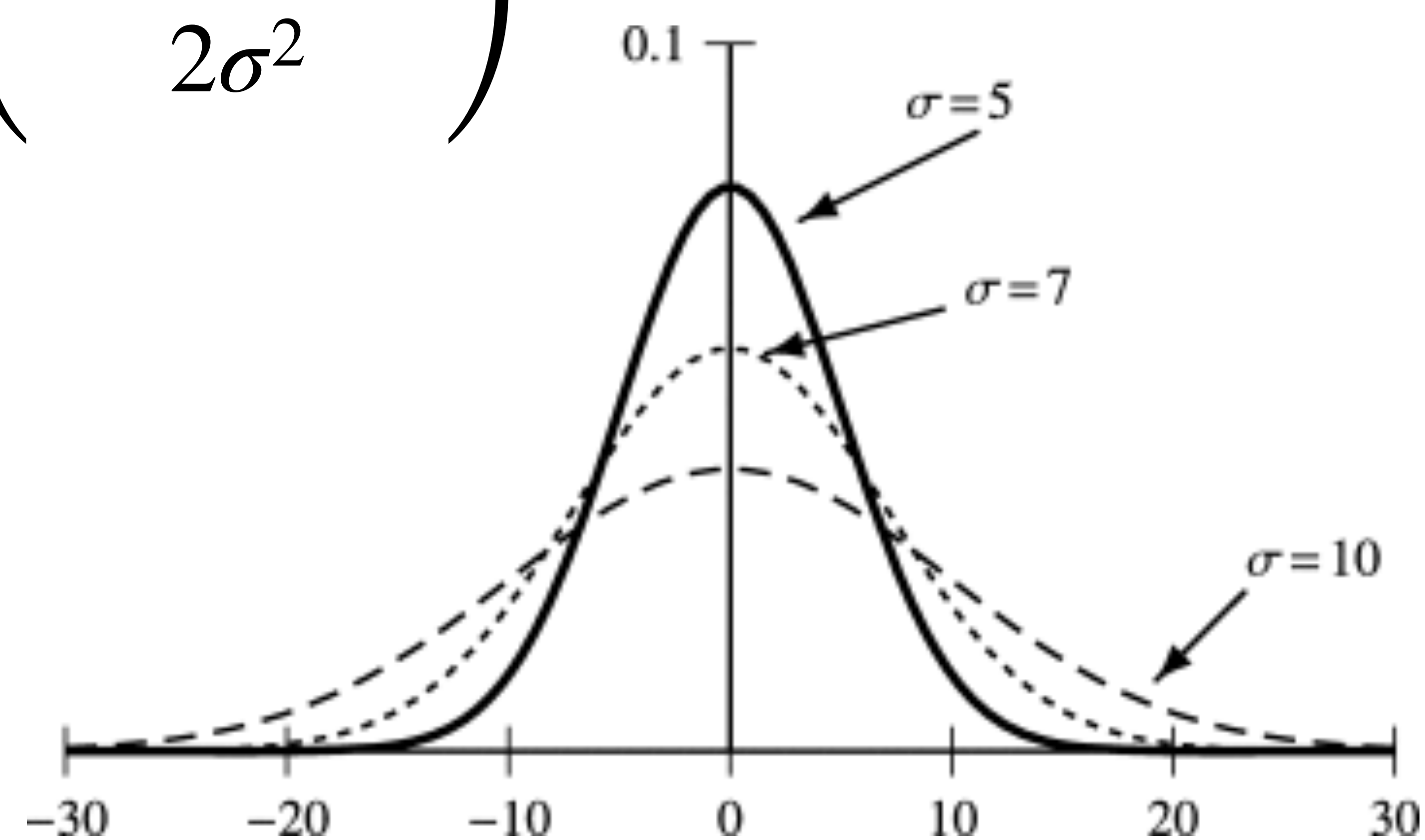
Likelihood refers to the probability of observing some data, x, given some parameter, θ. This is written

$$\mathscr{L} = P(x \mid \theta)$$

Gaussian Likelihood

As a toy problem, consider a parameter with true value θ and an observation of it, x. Life isn't perfect, so $x \neq \theta$. Rather it is draw from a Gaussian distribution centered on θ with a variance σ^2. (The "standard deviation" is σ).

$$\mathscr{L} = P(x|\theta) \propto \exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$$

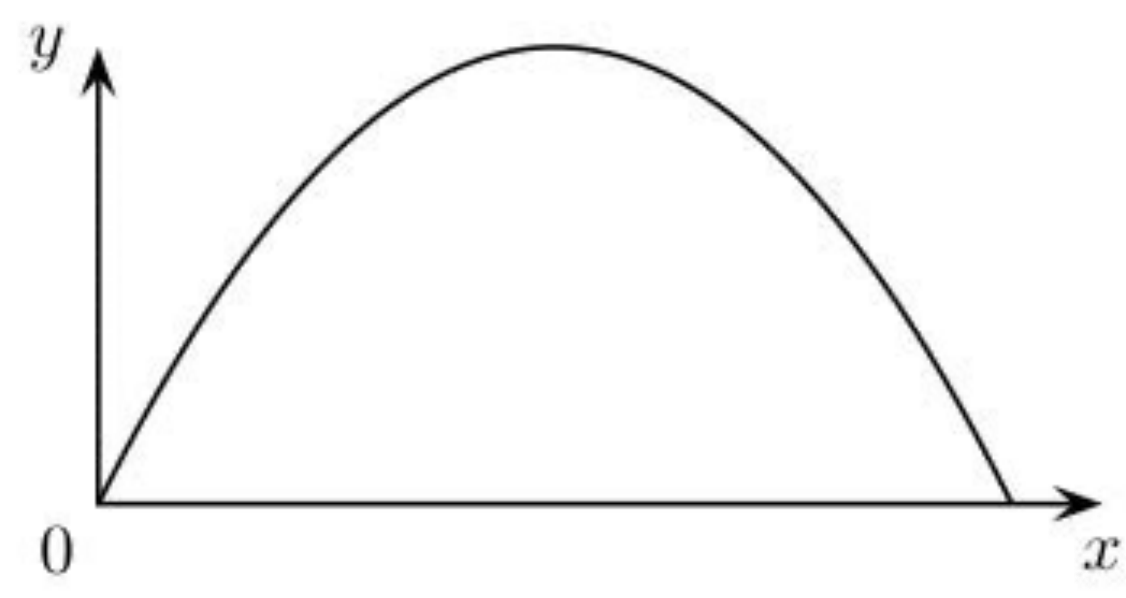


Gaussian Likelihood

The log of a Gaussian is a parabola.

$$\mathscr{L} = P(x\,|\,\theta) \propto \exp\left(\frac{-(x-\theta)^2}{2\sigma^2}\right)$$

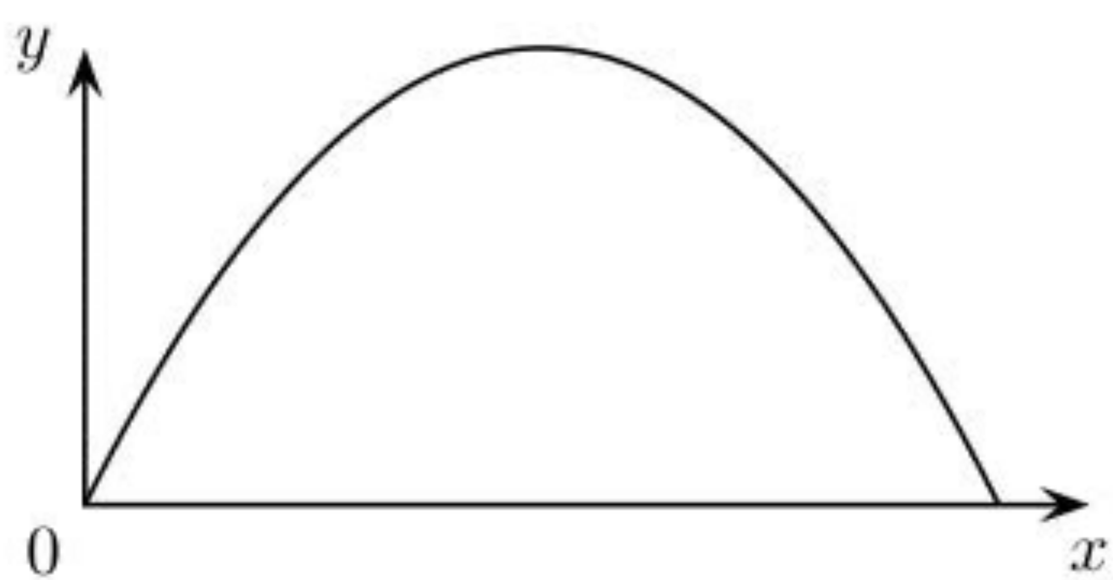
$$\ln \mathscr{L} = -\frac{(x-\theta)^2}{2\sigma^2} + C$$



How narrow / wide is it?

The whole question here is how narrow the parabola is. if the likelihood has a sharp peak, the distribution is very tight, the variance is low, i.e. the error of any give measurement is small.

A broad peak means the variance is larger.



Gaussian Likelihood

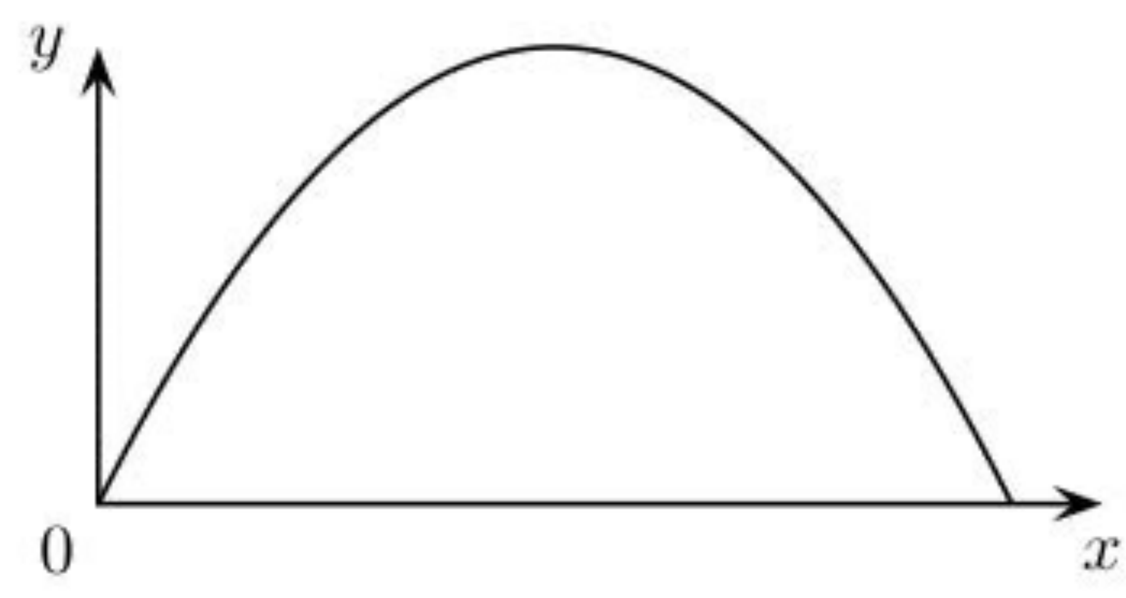
A sly way to get at the “curvature” near the peak is to evaluate the *second derivative* there.

$$\ln \mathscr{L} = -\frac{(x-\theta)^2}{2\sigma^2} + C$$

$$\frac{d\ln(\mathscr{L})}{d\theta} = \frac{x-\theta}{\sigma^2}$$

$$\frac{d^2\ln(\mathscr{L})}{d\theta^2} = \frac{-1}{\sigma^2}$$

$$\sigma^2 = -\left(\frac{d^2\ln(\mathscr{L})}{d\theta^2}\right)^{-1}$$



Gaussian Likelihood

This is the Cramér-Rao bound, the minimum variance possible for an unbiased estimator. (If the likelihood is approximately Gaussian near the peak anyway).

$$\sigma^2 = -\left(\frac{d^2 \ln(\mathscr{L})}{d\theta^2}\right)^{-1}$$

Or for multiple parameters, it is the matrix inverse of the matrix of second partial derivates (for Gaussians).

$$C_{i,j} = -\left(\frac{\partial^2 \ln(\mathscr{L})}{\partial\theta_i \, \partial\theta_j}\right)^{-1}$$

And it is really the “expectation value” of that matrix averaged of realization of the data, but don’t worry about that right now.

Performance Benchmarks

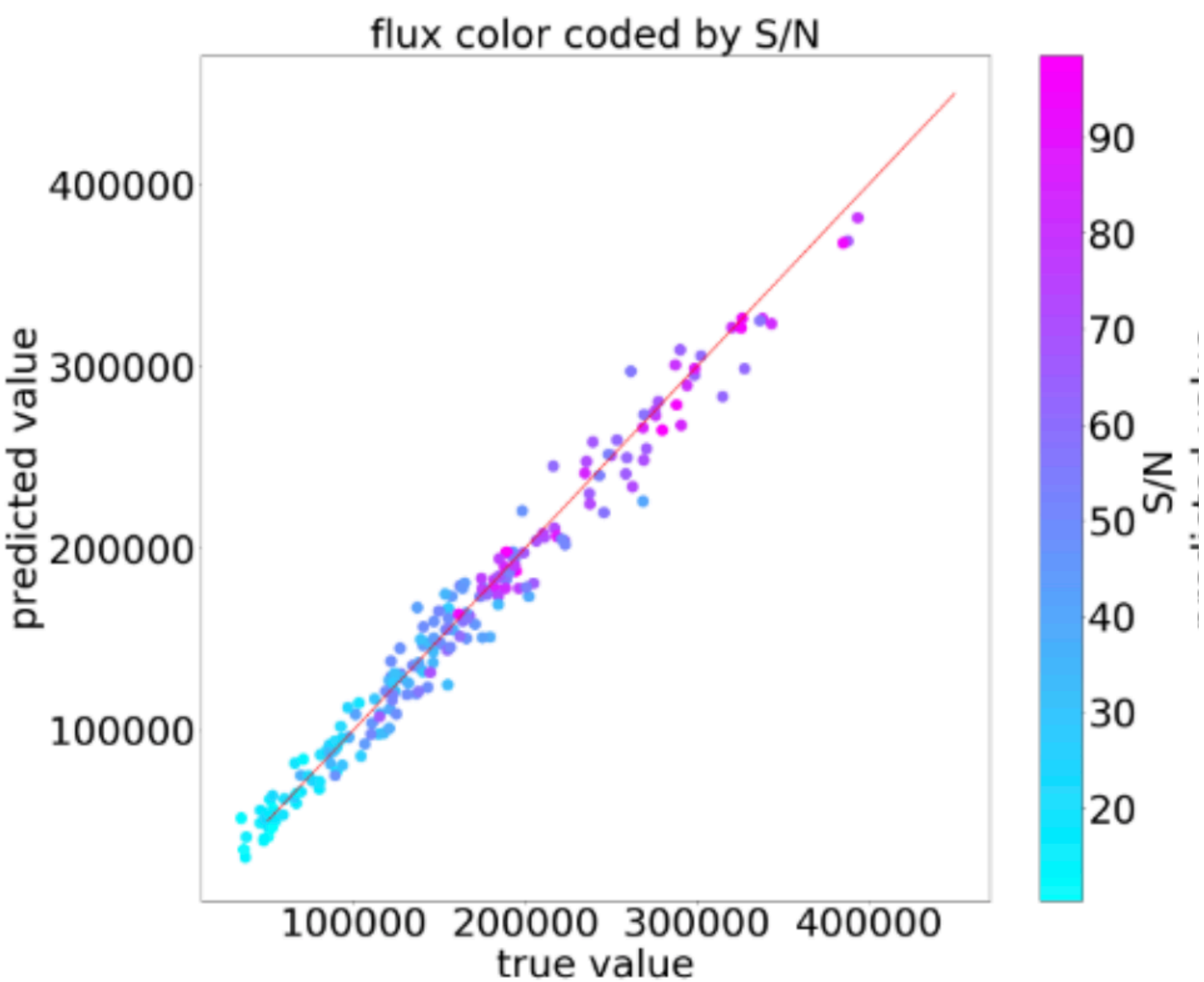
That was too easy, you say? Actually it can be challenging to take second derivatives numerically. Jun Yin already did it for you.

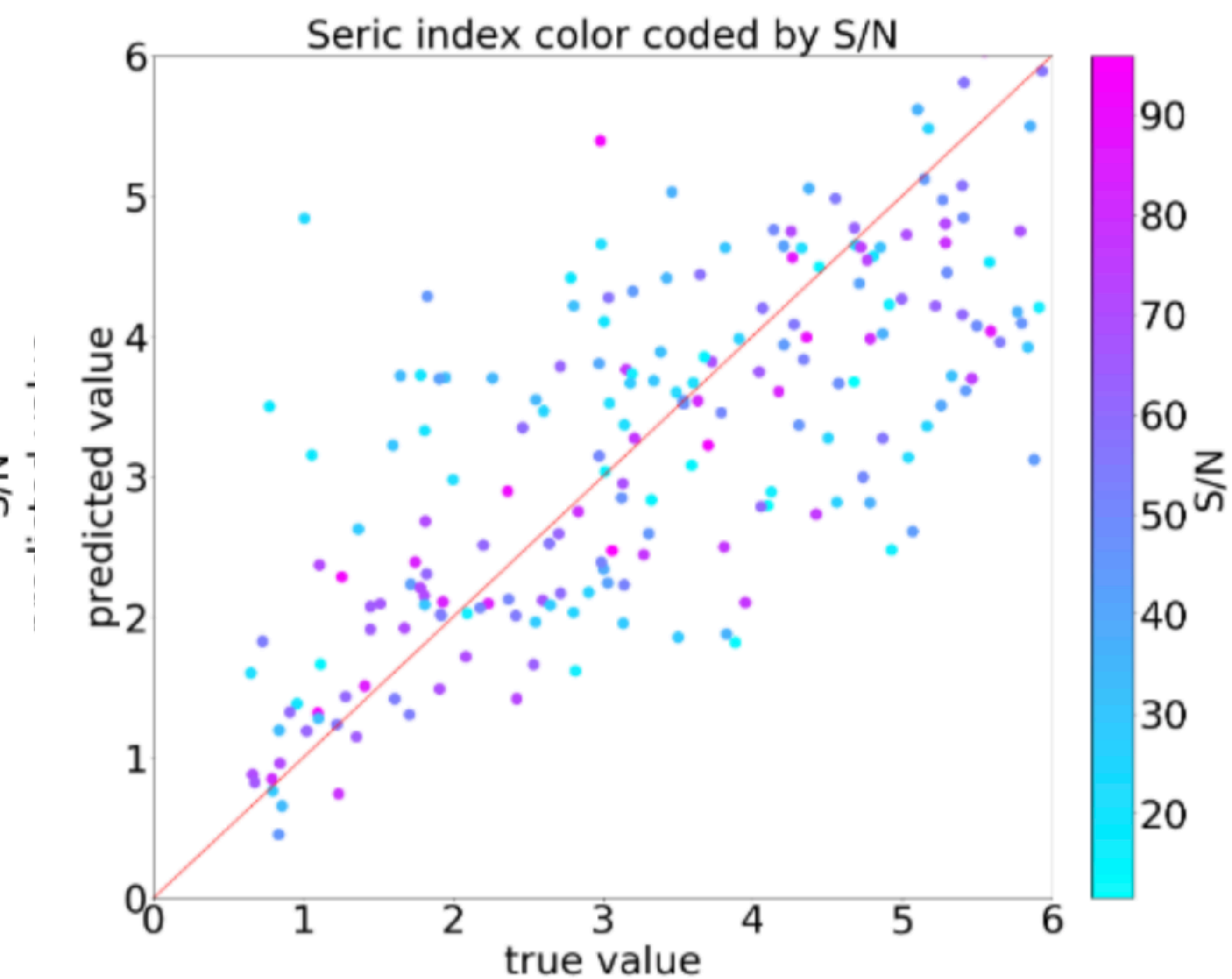
This gives you some idea how well you should do for a S/N=30 or S/N=60 galaxy.

Parameters	Value	CRB	CAE	CRB	CAE
SNR	NA	30	30	60	60
Flux[10^5]	1	0.11	-0.052 ± 0.12	0.056	-0.016 ± 0.072
Sersic Index	3	1.56	-0.57 ± 0.85	0.78	-0.48 ± 0.67
Sersix radius[arcsec]	0.3	0.001	0.0067 ± 0.054	0.028	-0.012 ± 0.036
g1	1	0.11	-0.019 ± 0.10	0.054	0.001 ± 0.053
g2	0.71	0.11	0.046 ± 0.10	0.054	-0.027 ± 0.061

Performance Benchmarks

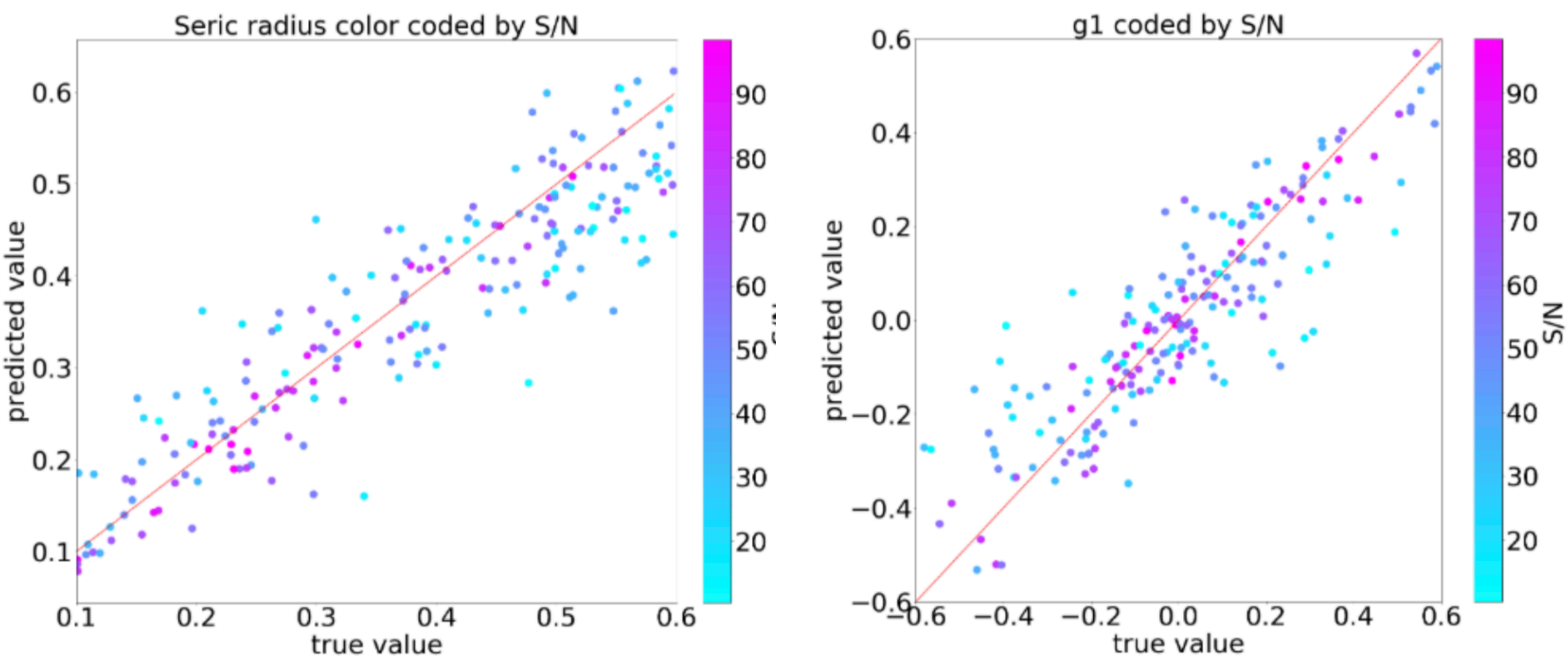
Also from Jun:





Performance Benchmarks

Also from Jun:



Stretch Goals

What happens if you vary the noise in the image?

What if the galaxy is slightly off center (a pixel or 2?)

What if the PSF is different from the training data?

It is important to explore failure modes and see what they look like. Every neural net model will fail in some limit, and mapping out those failures gives the user confidence that — at least in some parameter range — the results are trustworthy.