Scoring of Strong Gravitational Lens Finding Challenge

R. B. Metcalf

1. F_β score

The F_β score is defined as

$$F_\beta = (1+\beta^2)\,\frac{P \times R}{(\beta^2 P + R)} \tag{1}$$

where

$$P = \text{precision} = \frac{TP}{TP+FP} \tag{2}$$

$$R = \text{recall} \;= \frac{TP}{TP+FN} \tag{3}$$

and

$$TP = \text{true positives} \tag{4}$$

$$FP = \text{false positives} \tag{5}$$

$$TN = \text{true negative} \tag{6}$$

$$FN = \text{false negatives} \tag{7}$$

So $F_{\beta=0}$ = precision, $F_{\beta=\infty}$ = recall and $0 \leq F_\beta \leq 1$. $F_\beta = 0$ occurs when there are zero TP and $F_\beta = 1$ if there are no FP and no FN. As β is changed the relative importance of precision and recall changes.

The F_β can be calculated for a specific threshold on the score given by the classifier. Call the threshold p. The measure of the performance of a classifier will be the maximum F_β reaches for any p,

$$F_\beta = \max_p F_\beta(p) \tag{8}$$

Having high precision, or low contamination rate, is important for us because lenses will be rarer in the real data than they are in the simulated data sets by a factor of about 1000. We will use the value $\beta^2 = 0.001$. This will ensure that a high scoring classifier would give a pure sample of lenses on a test set with the real ratios of lenses to nonlenses.

2. What is a lens?

There are a fraction of the simulated lenses that have no background source added. These are signified in the training set with n_sources = 0.

These are not lenses. There are also cases where the source is too dim and the magnification too small for the object to look like a lens. The ability of a finder to find a lens will depend on at least the flux from the lensed image and its magnification (a proxy for the degree of distortion of the image). Other things like the number of companion galaxies and the surface brightness contrast might be important also. We can ask "What is the F_β for lenses with lensed image flux above f? or magnification above μ?" These quantities are given in the training set truth table.

For the sake of getting a single number out, a lens will be defined as a case where the entries in the log file satisfies n_source_im > 0, mag_eff > 1.6 and n_pix_source > 20. Cases with mag_eff < 1.0 or no source added will be considered nonlenses. There are cases that have $\mu > 1.5$ and the table says they have zero lensed images (column 16, n_source_im $= 0$). These are cases where there are no pixels in the source image above the threshold. It might still be possible to detect those so they will initially be considered lenses.

We will also do some cuts on other parameters and plot F_β as a function of those variables, but this should be considered the benchmark. You are welcome and encouraged to use a different definition and see how well you do.

3. supervised vs unsupervised

The simulated training set will never be a perfect simulation of reality and we will not know for sure all the ways it fails. Supervised learning methods can over fit to the training set and then give spurious results on the real data. An alternative is an unsupervised classifier that might be less susceptible to being fooled by idiosyncrasies in the simulated data. Unsupervised methods might also be better at finding unusual rare lenses that might not be well represented in the training set.

So there will be an extra category for lens finders that do not use the truth table of the training in their training. Please be honest about this.

4. colors

Using all the bands will probably make your classifier better. But experience shows that if the color distributions in the simulation are not the correct ones this can make the lens finder very biased and it can miss lenses that have odd colors or confuse companions galaxies with lensed images. Monochromatic lens finders might be useful also even if they don't perform as well on the simulated test set.

5. Einstein radius

In addition to identifying lenses you can also try to measure the Einstein radius. This is defined as

$$R_E = \sqrt{\frac{A_-}{\pi}} \tag{9}$$

where A_- is the area of the region with negative magnification within the cutout. A_- is also given for the training set in the truth table. In this case the MSE (mean squared error) will be used for scoring the submissions.