



Correcting a Confused Classifier

Automating the Assessment of Astronomical Anomalies

Kristoffer Stensbo-Smith¹ (k.stensbo@di.ku.dk), Daniel Mortlock², Christian Igel¹, Kim Steenstrup Pedersen¹

¹Image Section, Department of Computer Science, UCPH; ²Imperial Centre for Inference and Cosmology, Imperial College London

What's the point?

We want to detect astronomical images with defects or unwanted objects.

OK, tell me more!

Astronomers are looking for very old and distant objects, seen when the universe was less than 5% of its current age (as done by Mortlock et al., 2011, Nature, 474, 616). These objects are very faint, and the current detection pipeline results in many false positives. These false positives are what we want to detect and remove.

As a start, we want to classify images as either *true positives* or *false positives*. Later, we will attempt to do multiclass classification to predict exactly which defects (if any) are present in an image.

Why is that so difficult?

- Noisy background.
- Very faint sources.
- CCD flaws can be as small as a couple of pixels wide.
- 5 classes of problems, about 1000 labelled samples.
- Some classes contain very few samples.

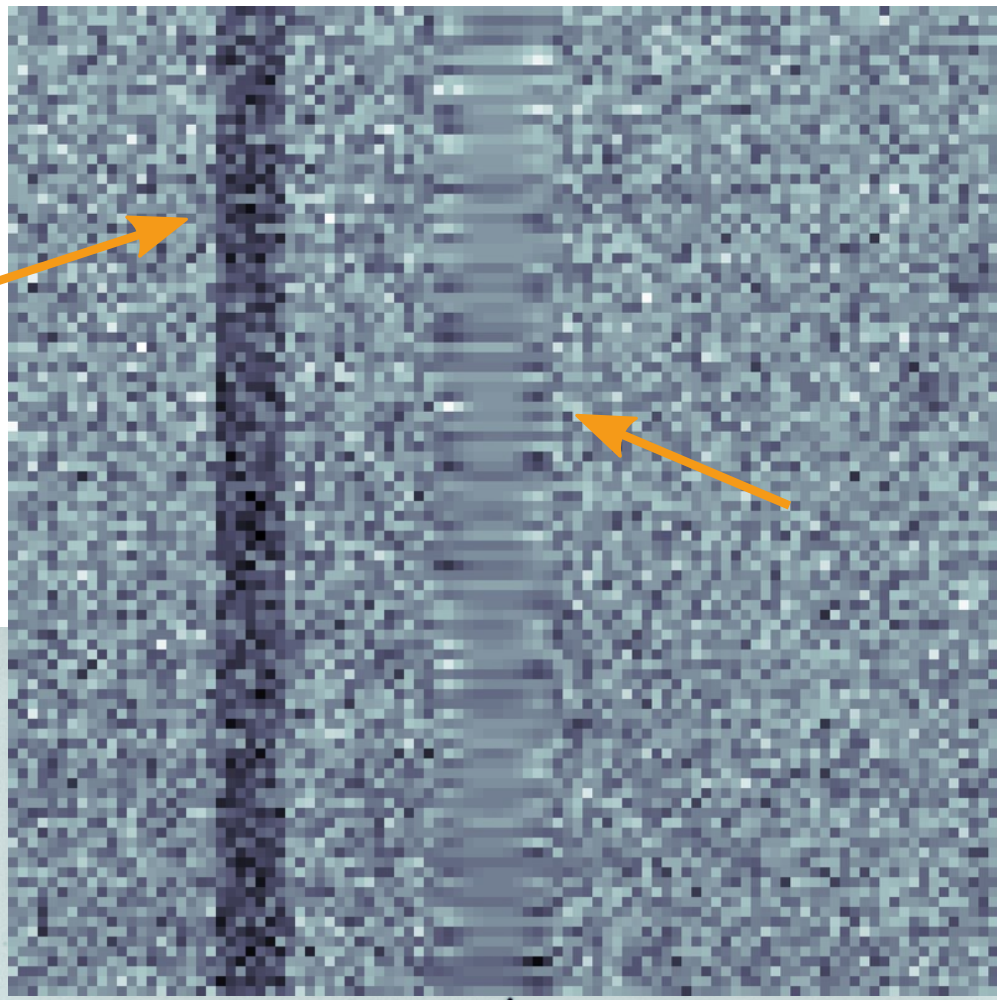
What is an acceptable image?

- There must be:
- no source at the centre,
 - no artefacts in the image,
 - no diffraction spikes pointing towards the centre,
 - no sources (faint or bright) close to the centre.

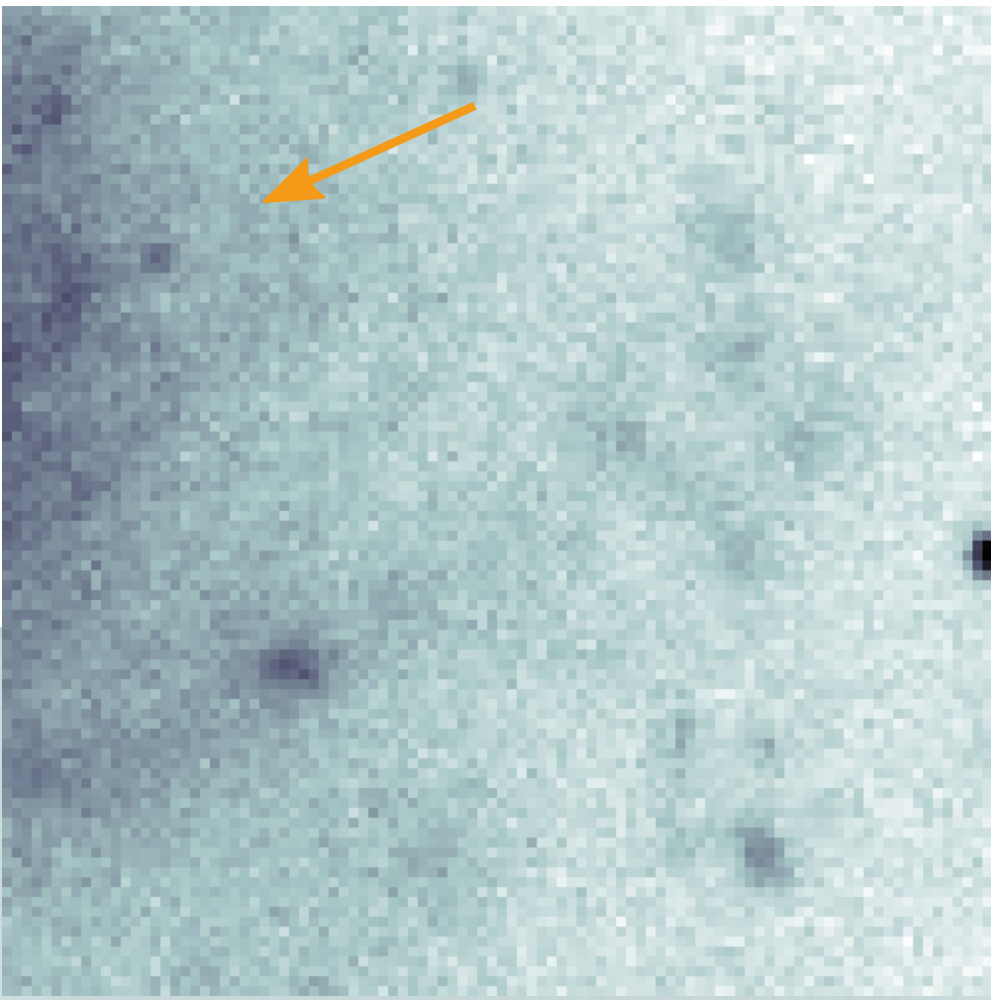
What is not an acceptable image?

Anything else. See the four examples below.

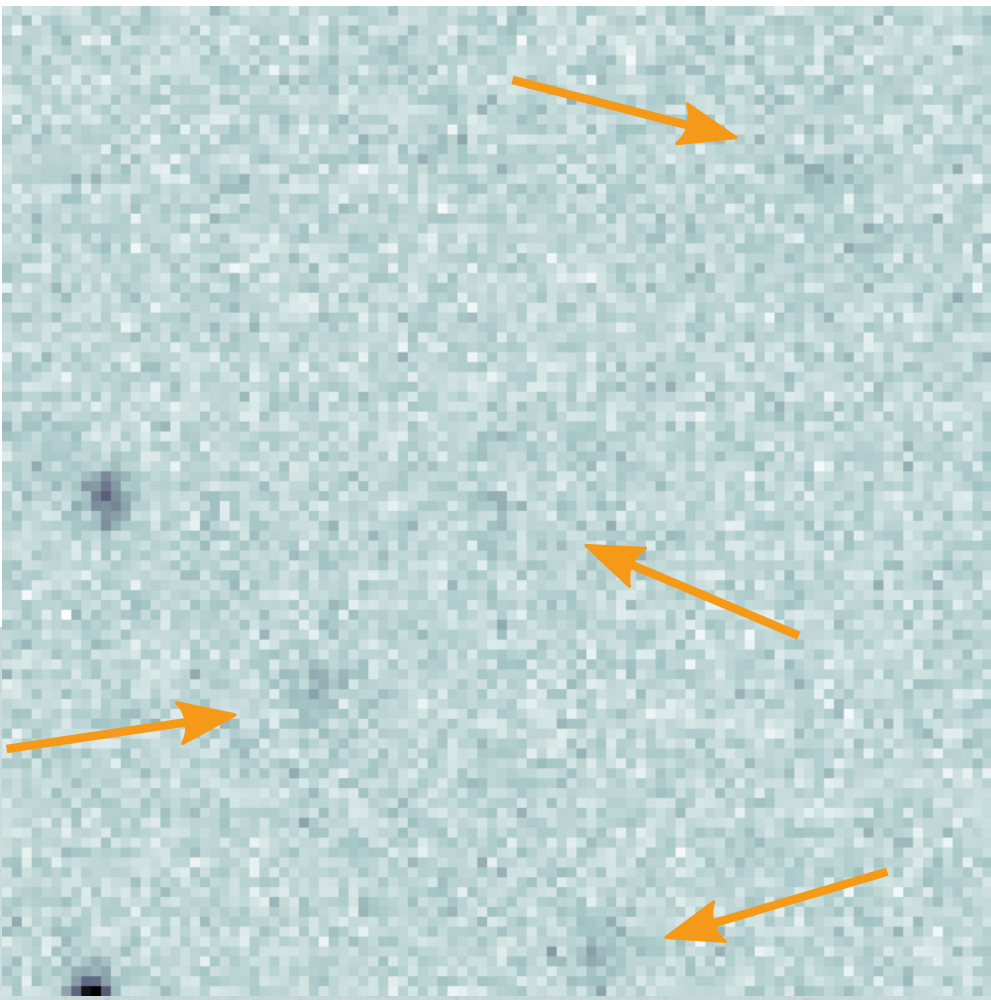
CCD flaws



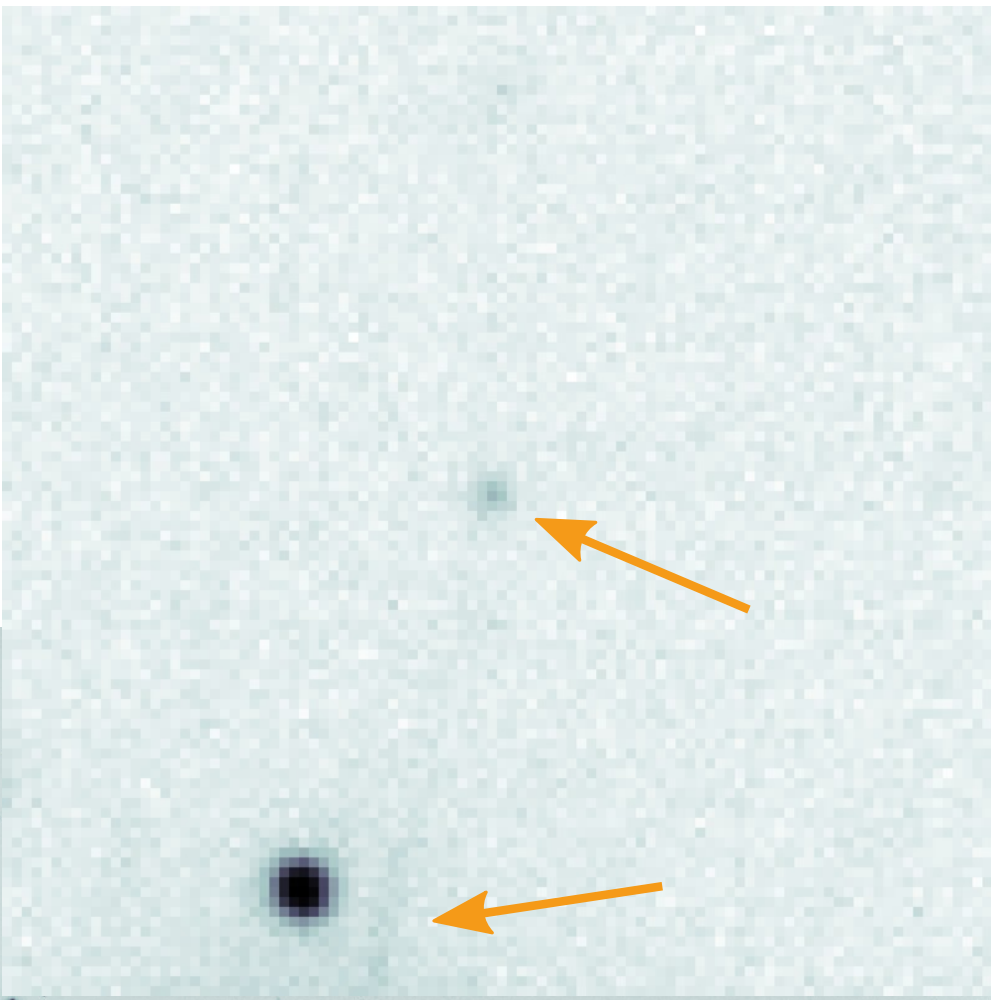
Foreground galaxy



Very faint sources



Faint diffraction spike; source at centre



My wishlist

- How can we take advantage of semisupervised learning?
- Which other features could be informative?
- What would be a good way to do (probabilistic) multiclass classification?

