Galaxy Detection with Neural Networks

Andrii Torchylo, undergraduate, Stanford University, KIPAC
torchylo@stanford.edu

Blending problems

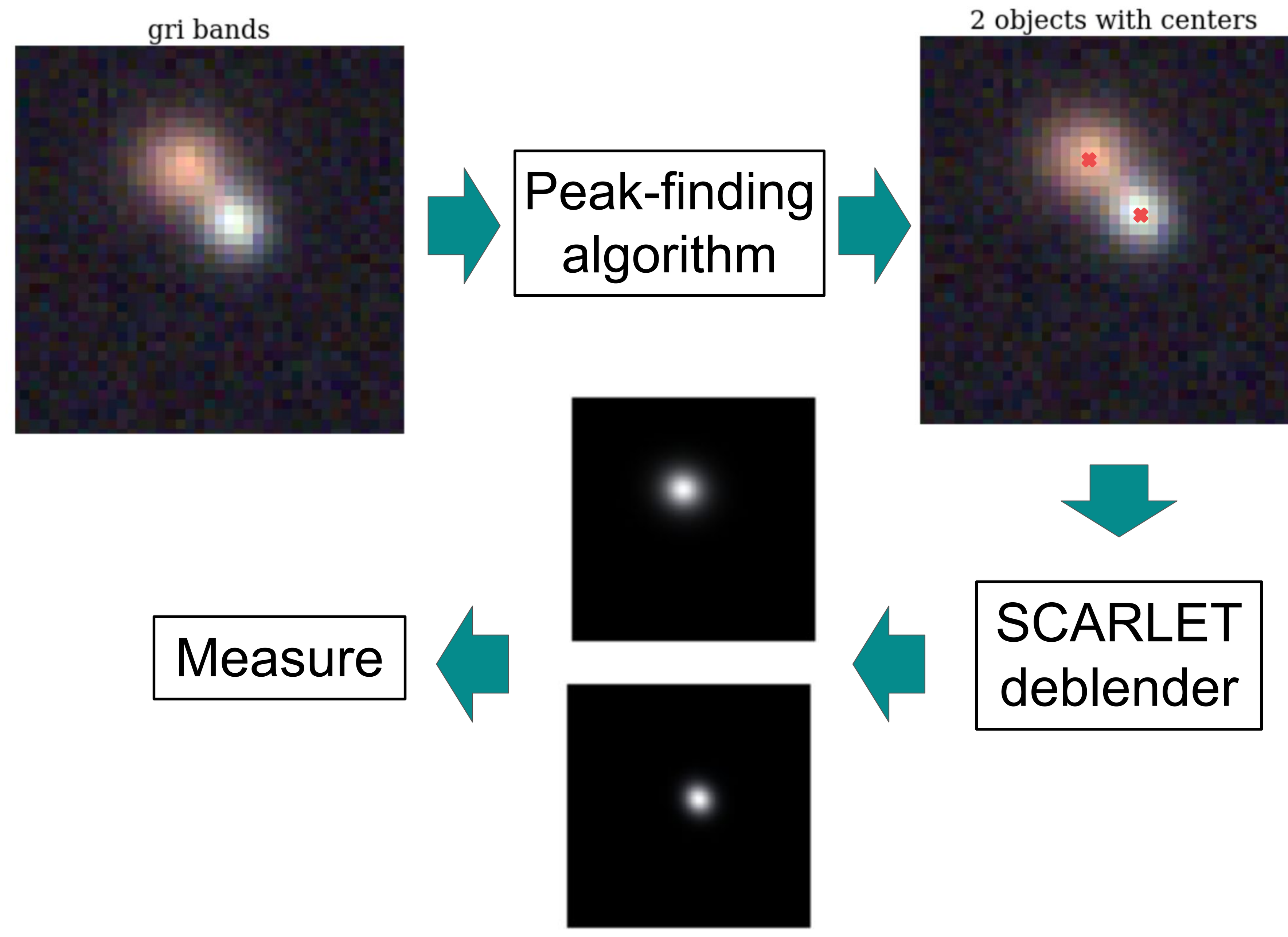
- Because of the increased depth of observations in the LSST, ~63% of galaxies will appear to be "blended" — their fluxes will partially overlap in an image (Sanchez et al. arXiv:2103.02078).
- Effects of galaxy blending can bias measurements of **position**, **flux**, **shear**, and **photometric redshifts** of galaxies, introducing significant uncertainty to our science results.

Current solutions

- **Detection:** typically a peak-finding algorithm (like **SourceExtractor**) is used to find the centroids of stars and galaxies.
- **Deblending:** current state-of-the-art approach is the **SCARLET** deblender (Melchior, Moolekamp et al. arXiv:1802.10157), which separates the flux using the input image and centroid detections.

Limitations

- Deblender performance will degrade significantly if incorrect number of detections is provided.
- Most detection algorithms use a single-band image for detecting sources, or compress multi-band information into one band.



Project 1: Predicting number of sources with ResNet

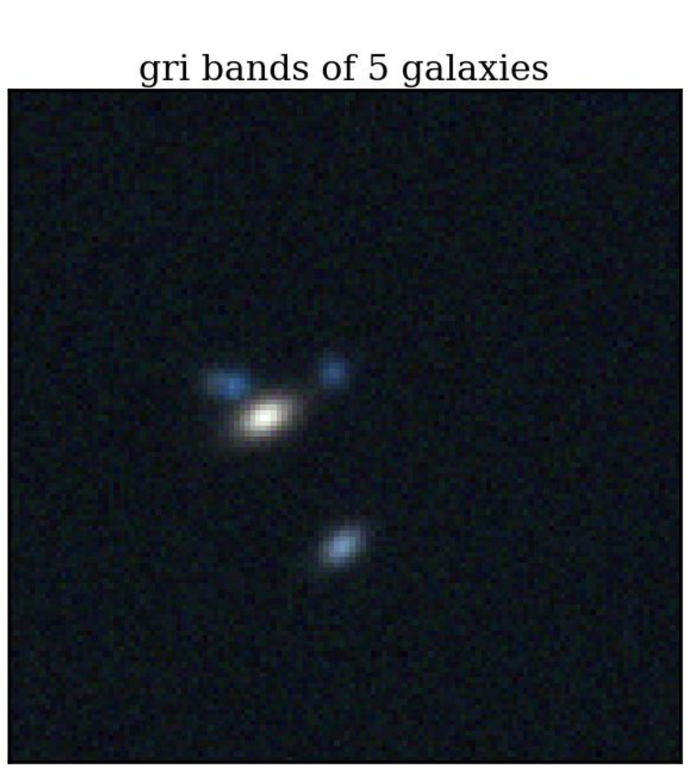
Motivation:

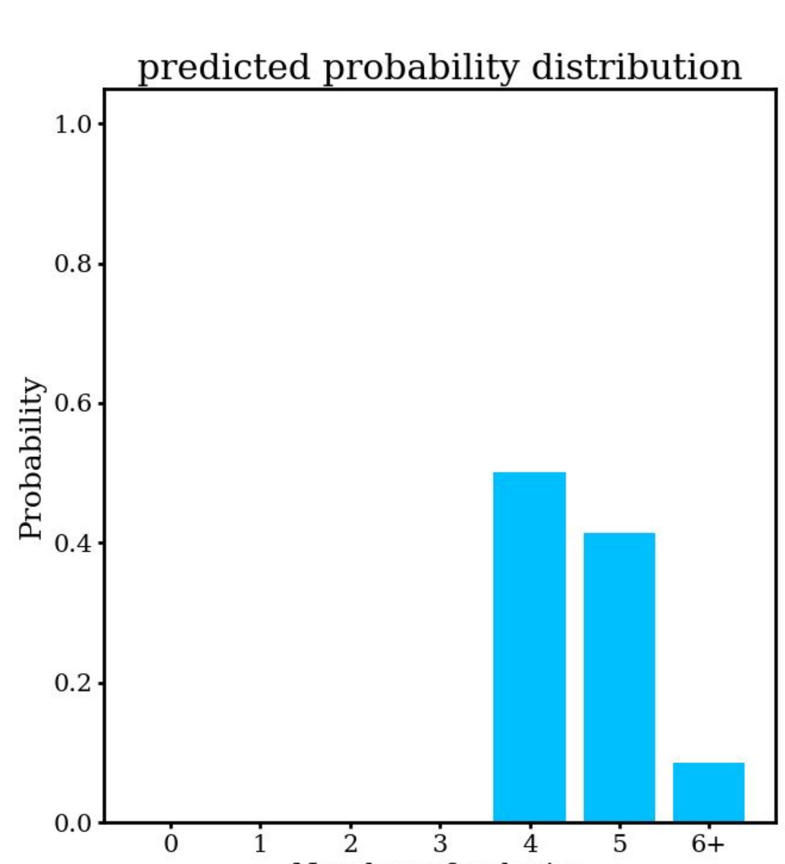
- "Any undetected sources are catastrophic to the deblender" (Fred Moolekamp).
- Want an algorithm that minimizes unrecognized blends — cases when we *detect* fewer sources than are present in the blend.

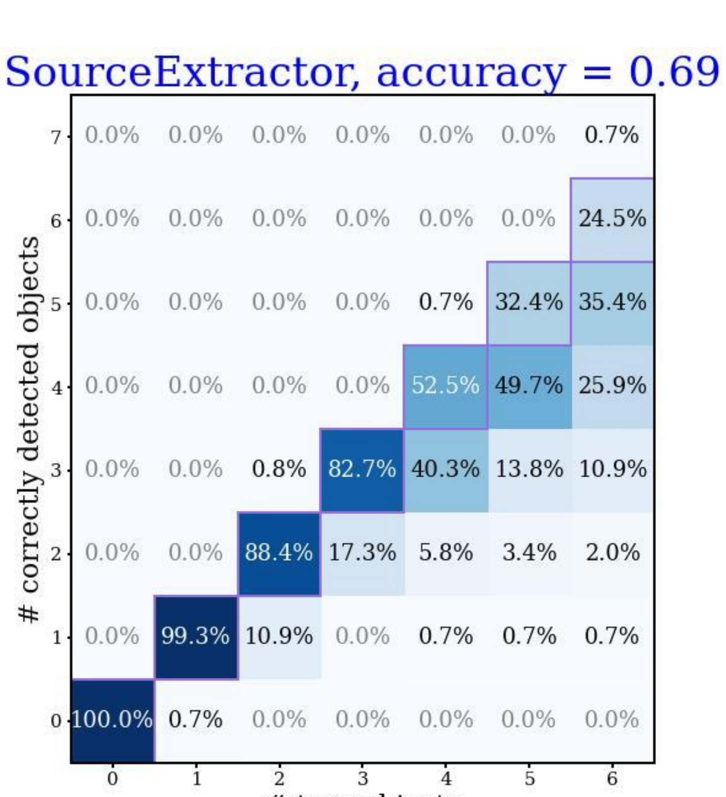
Recognized vs. unrecognized blend
dashed ellipse = detected object(s)

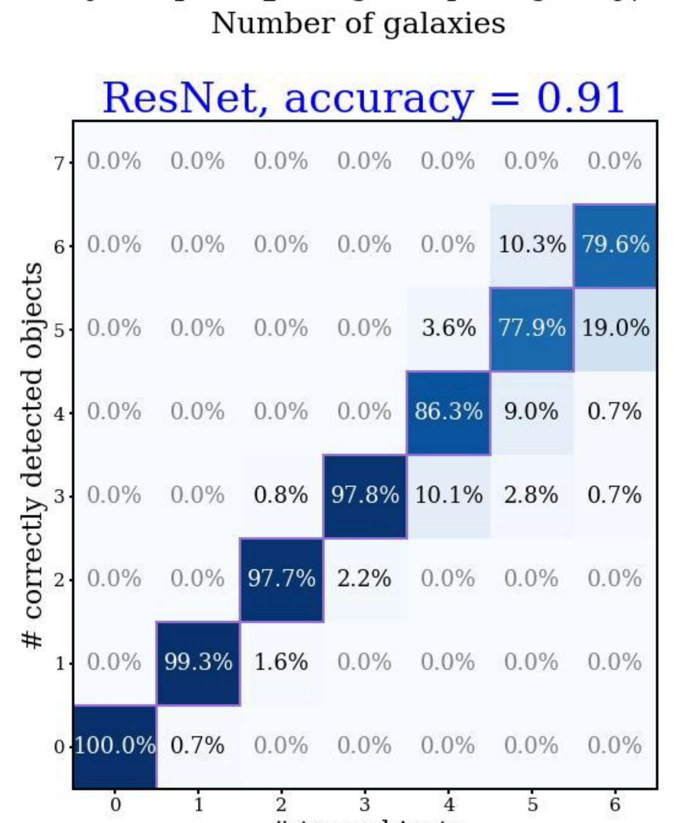
Our Approach:

- We used the DESC **Blending ToolKit** (Mendoza el al. 2022 in prep) to construct a dataset of simulated images for training and evaluation.
- We trained a Convolutional Neural Network model, **ResNet** (He et al. arXiv:1512.03385), to predict the **number of galaxies** on the input **6-band** postage-stamp image.
  - Framed the problem as a multiclass classification problem, where each class is the number of galaxies in the images.
  - Class "6+" represents images that have 6 or more galaxies.







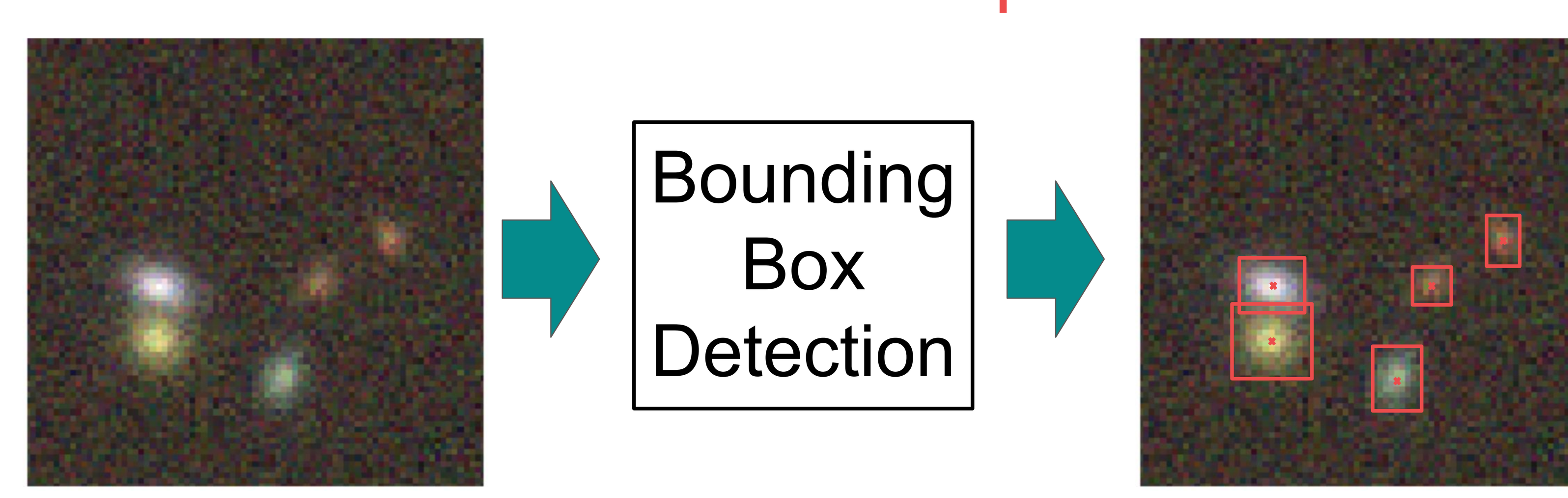


Note: accuracy here is the fraction of correctly classified images

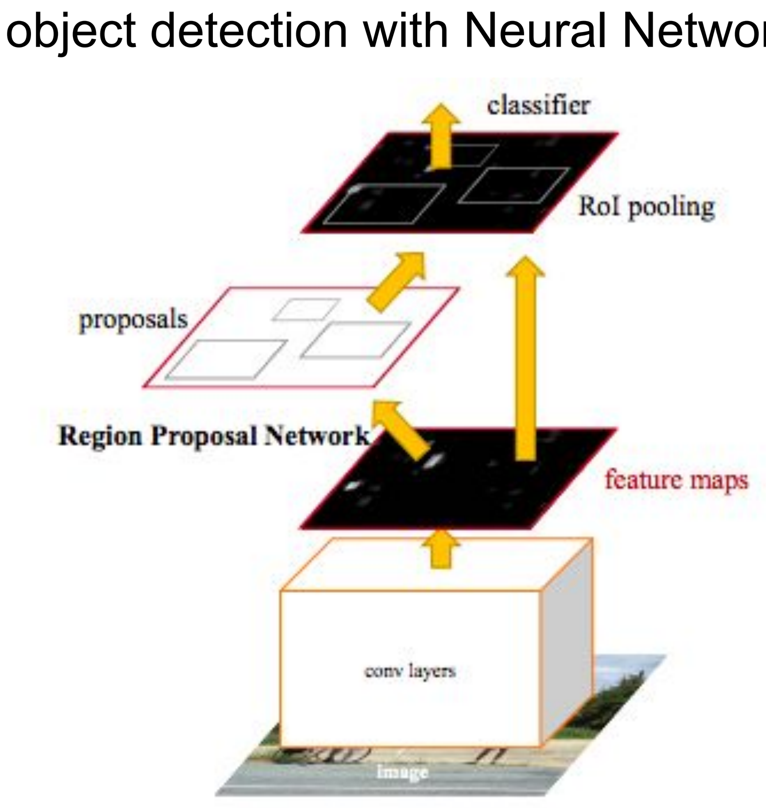
Next Steps:

- Train and evaluate on a more realistic sample of galaxies.
- Study the effects of noise on performance.

Project 2: Galaxy detection and shape measurement with Ellipse R-CNN



Motivation: Bounding box prediction is one of the most popular methods of object detection with Neural Networks, but has limitations for elliptical objects.

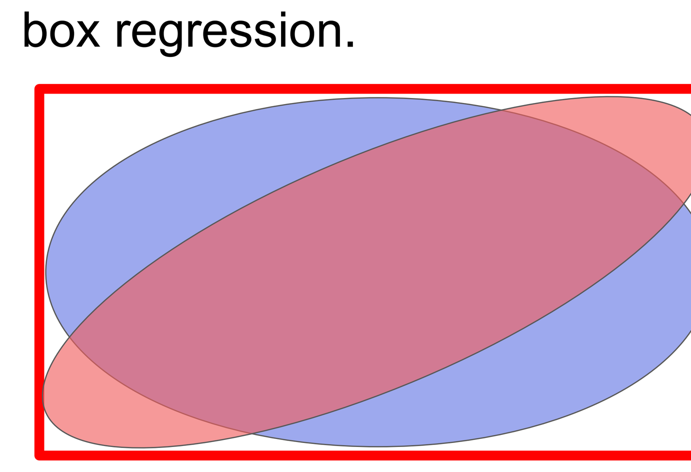


Faster R-CNN architecture (Ren et al., arXiv:1506.01497)

The architecture of most such models is composed of two parts:

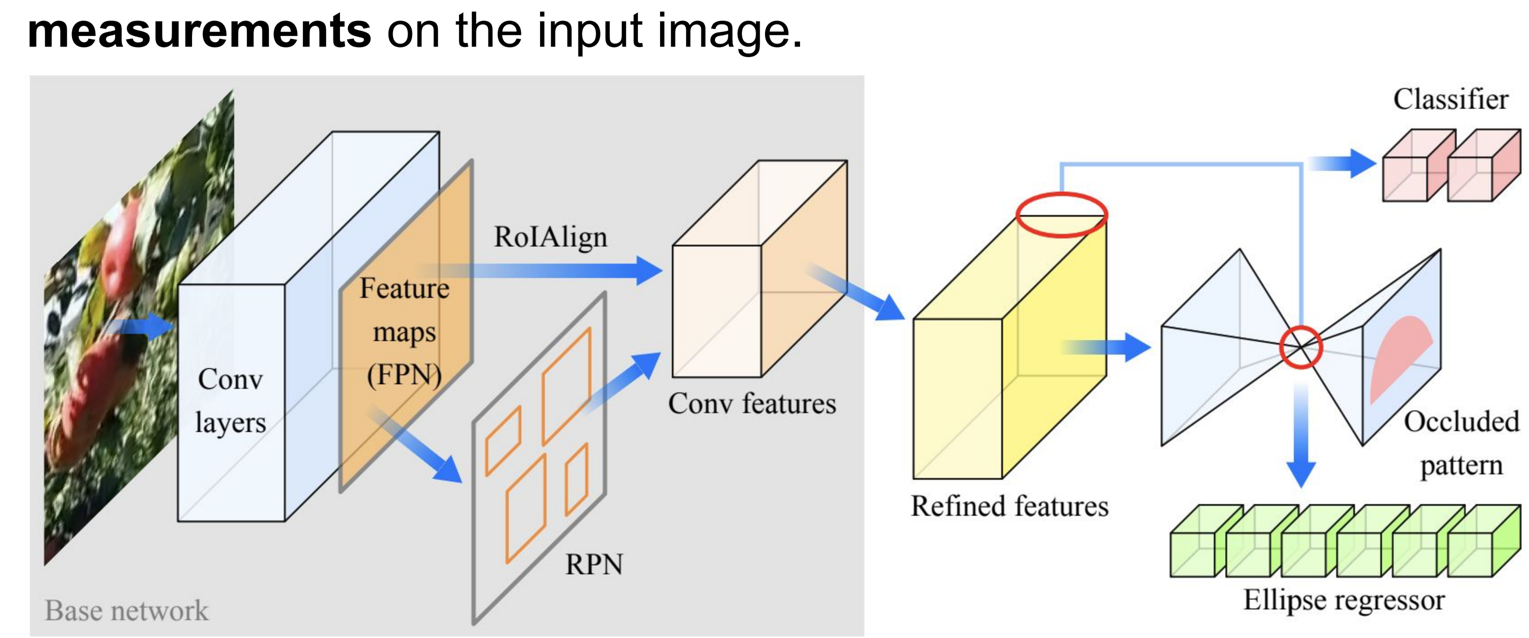
- **Region Proposal Network (RPN):** a small neural network that outputs rectangular regions that are likely to contain an object.
- **Classifier:** a neural network that takes in a region of the image and performs classification (galaxy/no galaxy in our case) and bounding box regression.

Limitation: Bounding boxes do not fully capture the geometry of the galaxies, as many different galaxy shapes can fit in the same bounding box. This can create a bias in the detection algorithm for certain shapes and aspect ratios.



Our Approach:

- We propose to use **Ellipse R-CNN** (Dong et al. arXiv:2001.11584), which is able to predict 5 parameters of the ellipse (x, y, a, b, Θ) and learn occluded patterns in the image.
- Ellipse R-CNN extends the regions proposed by RPN to a larger square with padding, which minimizes shape bias.
- Using this architecture, we can not only **detect** galaxies, and use those detections to initialize a deblender, but also perform direct **shape measurements** on the input image.



Next Steps:

- Optimize for 6-band detection of transparent elliptical galaxies.

I thank my advisors Pat Burchat, Shuang Liang, and Sidney Mau for all of the guidance I received during the summer.
Also, I am grateful to the Physics Undergraduate Summer Research program at Stanford for funding my research

Participant in the LSST Corporation's program for student researchers at the 2022 Rubin PCW.