A Closer Look at ClaRAN

Vinay Kerai

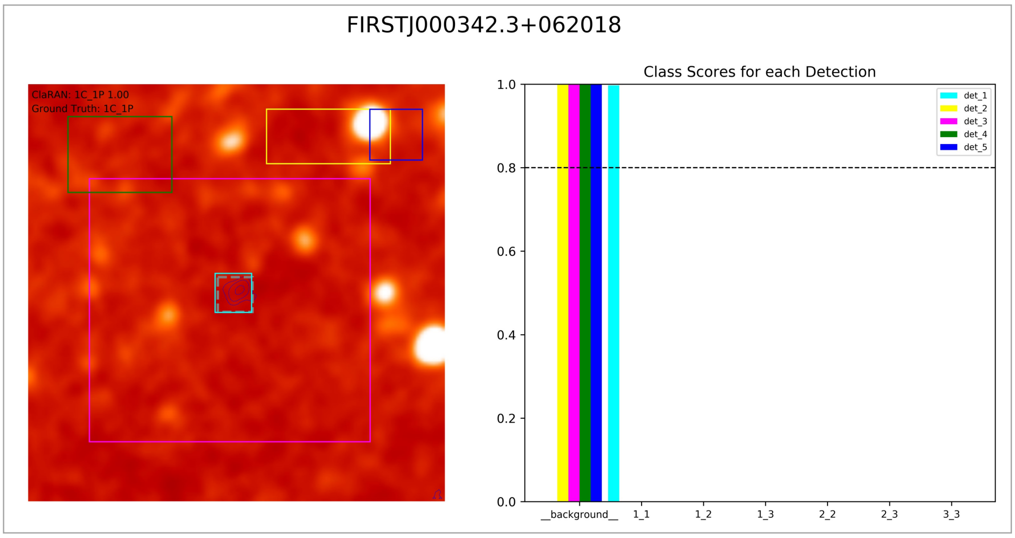
Abstract

ClaRAN is a convolutional neural network that is able to identify and classify radio sources, however in its current state uses an arbitrary 0.80 confidence threshold when finalising classifications. In many cases, it was observed this leads to correct classifications being omitted which is non-ideal. To inspect such cases more closely, additional code was written to be able to visualise the score distribution across classes for each detection by generating histogram plots. This report uncovers the findings made during examination of these plots. Further to this, potential threshold alternatives were trialled and explored, however a robust solution is yet to be found. The limit on the number of detections was raised from 5 to 10, to investigate whether consistent trends lied amongst the score distributions in order to determine if statistical-based solutions are feasible.

Background

ClaRAN's output after applying SoftMax is a score ranging from 0 to 1 for each class. ClaRAN can choose from 7 classes, one of which is a default _background_ class, while the others include 1C_1P, 1C_2P, 1C_3P, 2C_2P, 2C_3P, and 3C_3P. A characteristic of the SoftMax function is that after being applied, the resulting score distribution will sum to 1, hence these scores are often treated as probabilities. If a single class has a score exceeding 0.80, then it is possible to infer that it was the class with the highest score. Theoretically, this threshold could be set at 0.50, because likewise if a score exceeds 0.50 it must be the highest, since the remaining classes will sum to a value less than 0.50. However this is likely to enable random detections to be introduced into ClaRAN's output, hence a much higher value of 0.80 serves to filter out to an extent, detections made on the basis of randomness.

ClaRAN also proposes numerous bounding boxes to highlight regions where it has found a radio source, hence there is a score distribution associated with each of these bounding boxes. Currently ClaRAN keeps 5 bounding boxes, which are chosen using NMS (Non-Maximum Suppression) which considers the likeliness of a box containing a radio source as well as the relative position of other bounding boxes. A simple example of the plots generated is shown here. Each bounding box and score distribution is colour coded to provide clear insight into ClaRAN's choice of output. Additionally included is the ground truth bounding box drawn with a grey dashed line. Printed in the top left corner is ClaRAN's classification and the ground truth classification.



Plots of Interest

Compiled below are some plots of interest along with a short description of some notable observations. It is important to note whilst the detection scores are plotted side by side, meaningful comparison cannot be made across detections since each is made independent of one another.

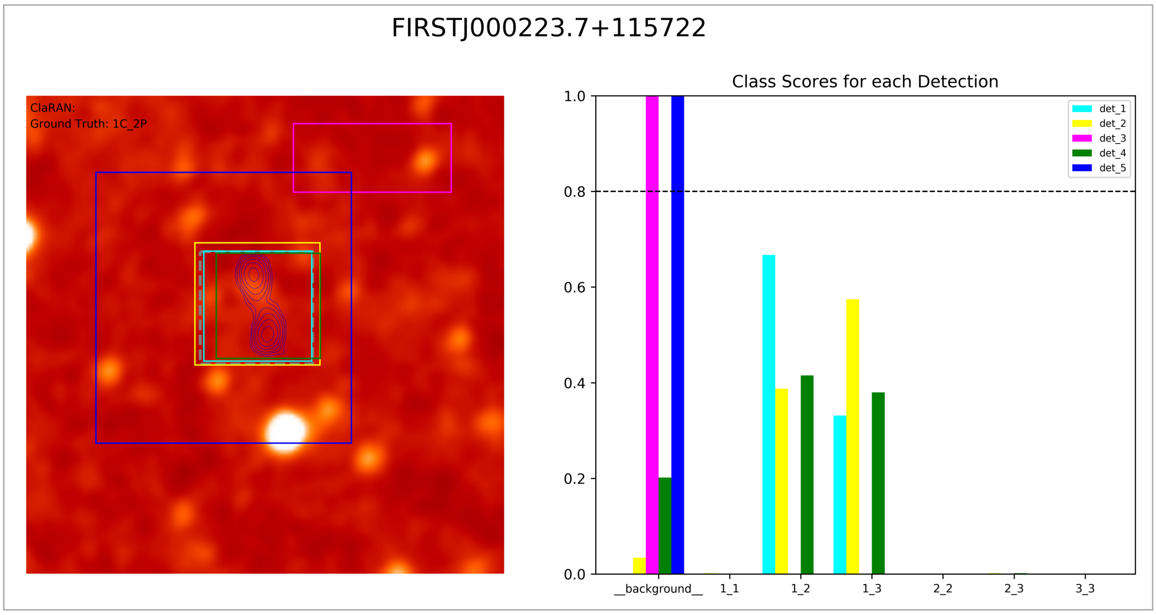


Figure 1

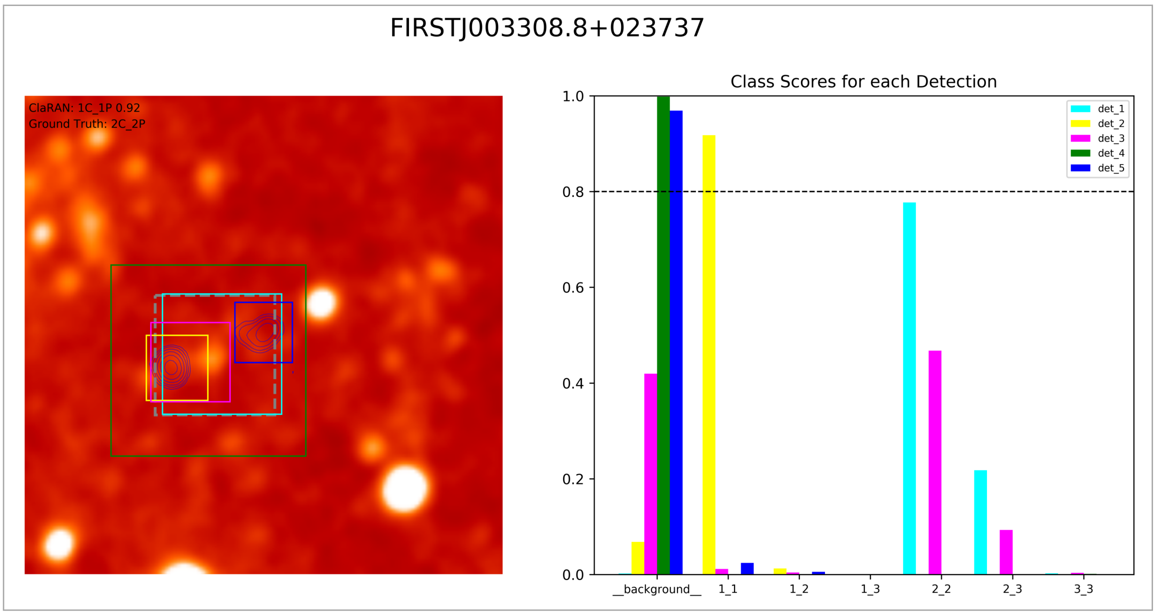


Figure 2

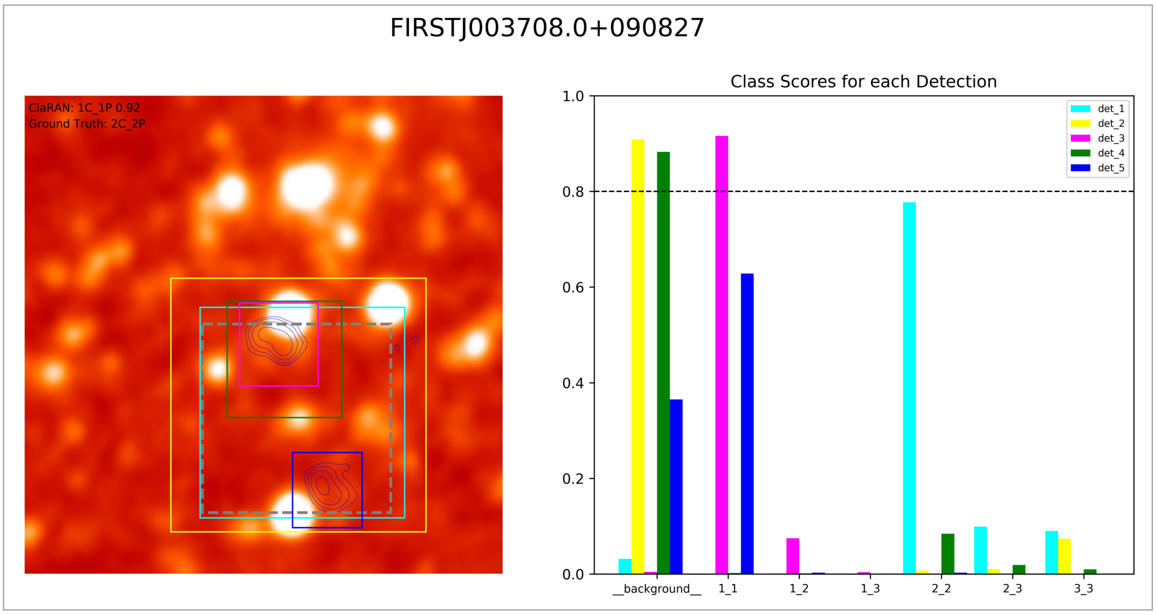


Figure 3

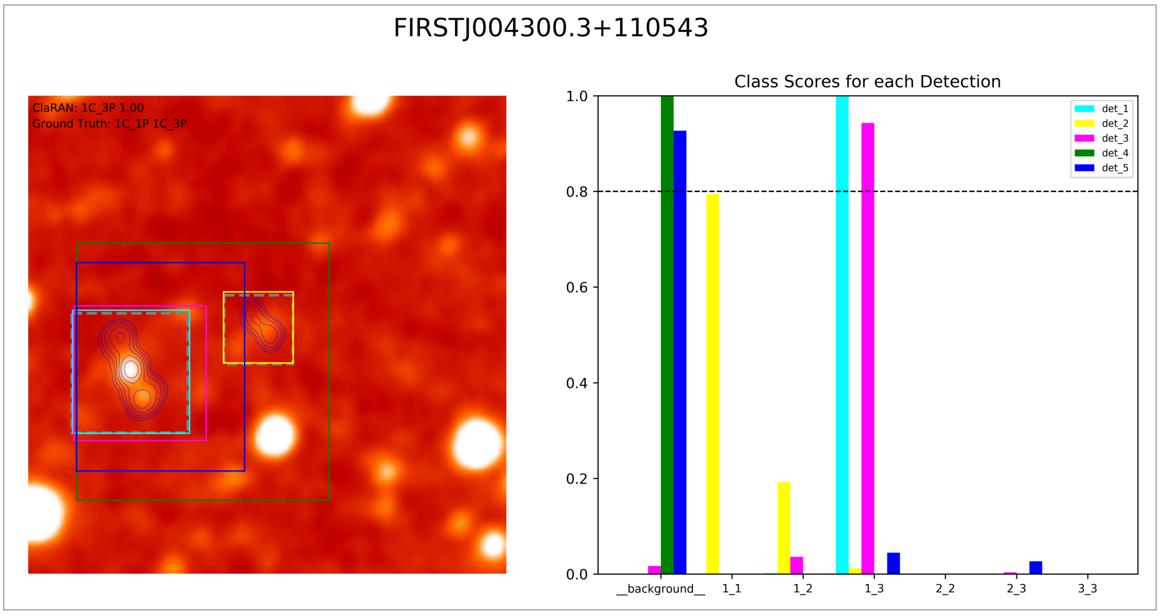


Figure 4

In figure 1, ClaRAN did not display any detected radio sources since no class excluding background had a score exceeding 0.80 across the five detections. The detection with the highest score is detection 1, which made the correct classification as well as proposing a suitable bounding box. This is a case where ClaRAN was able to make the correct classification however it's confidence score was too low for it to be displayed. For cases like these when no individual class exceeds the 0.80 threshold, we can simply take the class with the highest score to be the output. This solution requires no more than a few lines of code to implement but would enable ClaRAN to report an output; but this relies on ClaRAN being confident in background classifications for when truly no radio source lies in the image.

In figure 2 and 3, ClaRAN has output a 1C_1P classification when in fact its second highest detection was correct. In both figures, detection 1 was correct but again fell slightly short of the 0.80 threshold. It may appear as though ClaRAN has a stronger affinity for 1C_1P classifications. This potentially could be attributed to the image set ClaRAN was trained on. The training set contained 3,518 1C_1P images, which is more than the other 5 classes combined (3,460 images). More likely however, this is explained by the fact we cannot make meaningful comparisons across detections. The high score for 1C_1P is based on its corresponding bounding box, and is relative to the other classes in the same detection. In other words, in its own detection it may be slightly likely to be a 1C_1P, but after passing through SoftMax it's relative probability is inflated to the high value observed. An ideal solution would take these five detections and their class score distributions and rank them on the basis of how correct they are likely to be.

ClaRAN is also able to detect multiple radio sources within the same image, as such is the case in figure 4. This figure highlights again why the arbitrary 0.80 threshold may be negatively impacting ClaRAN's performance. Detection 1 is a perfect classification, but detection 2 was also correct since there are indeed two radio sources in the image. If ClaRAN was significantly biased towards 1C_1P classifications it would be expected a high score would be observed here. Simply lowering the threshold may solve the problem in cases such as those in figures 1 and 4 but simultaneously will make no difference in cases such as those in figures 2 and 3.

From what has been observed in these plots, it can be deduced ClaRAN does not have a major issue deep within its network, since generally one of its detections will be correct. By visualising ClaRAN's output in the form of the plots shown above, it was found ClaRAN is able to make the correct detections, they are simply not being displayed. A smarter and more dynamic method of choosing a detection is required in order to improve ClaRAN's performance.

Threshold Alternatives

Two simple alternative methods to using a fixed threshold were investigated to see whether improvements could be made. The first relies on a statistical method using the probabilistic nature of these distributions. Given five independent probability distributions a simple way to combine them is to take the arithmetic mean. Below are results when this method is applied to the images from figure 2 and 3 above.

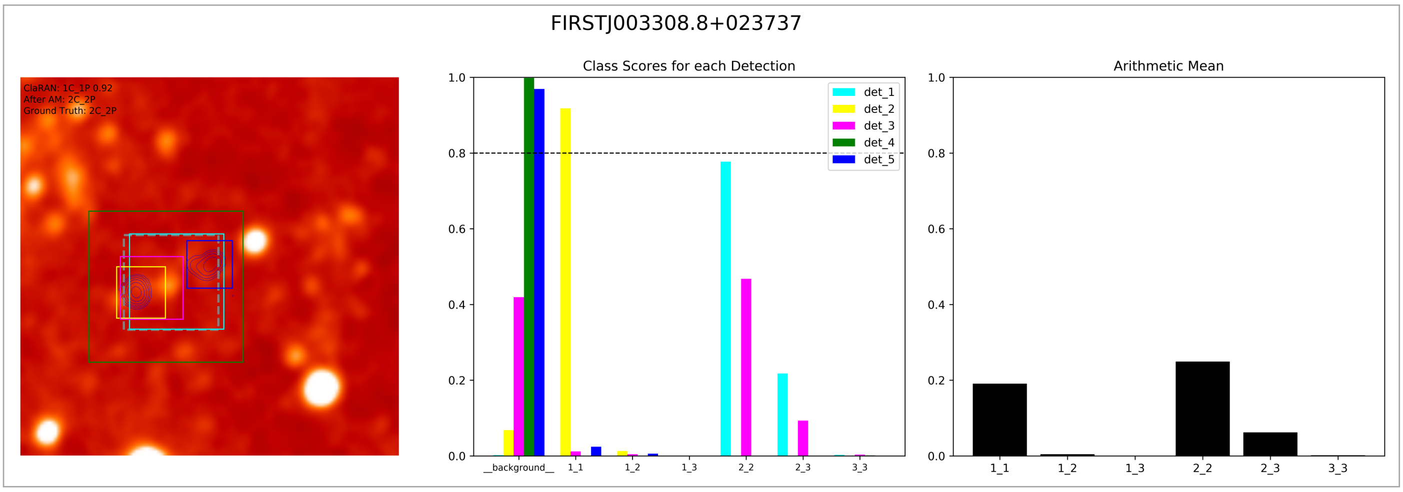


Figure 2 with Arithmetic Mean

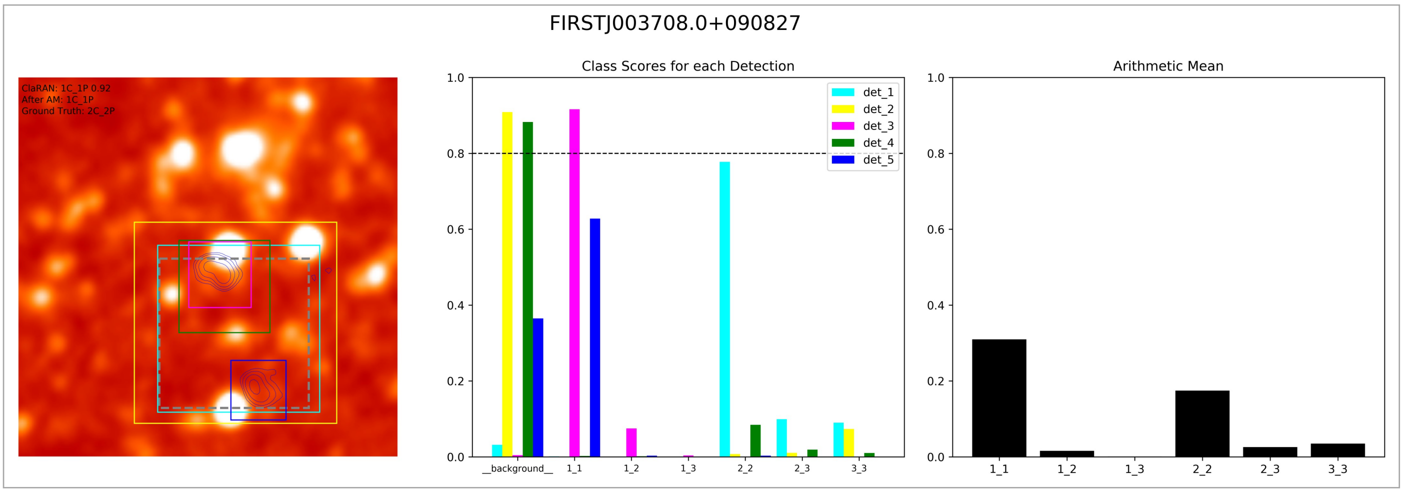


Figure 3 with Arithmetic Mean

For figure 2, the method is successful since ClaRAN made two 2C_2P detections, hence had the greatest mean score across the classes. Unfortunately, this method was unsuccessful for figure 3, since ClaRAN made two 1C_1P detections instead. This method could be improved by taking a weighted mean instead, to place greater importance on certain class detections. Another factor which may have affected this result, is the fact that the detections are chosen by NMS to be diverse, hence these five bounding boxes are not necessarily the five best bounding boxes. This means there may be many more proposed boxes closer to the ground truth, but were excluded after NMS due to the high overlap of similar boxes.

Another method trialled was using logistic regression on the output. Given, the bounding box dimensions and the corresponding score distribution, a value between 0 and 1 was output which can be thought of as a new score representing how likely a particular detection is correct. A flaw in this method was that the inputs were again independent of the other detections, so it was clear more parameters were needed.

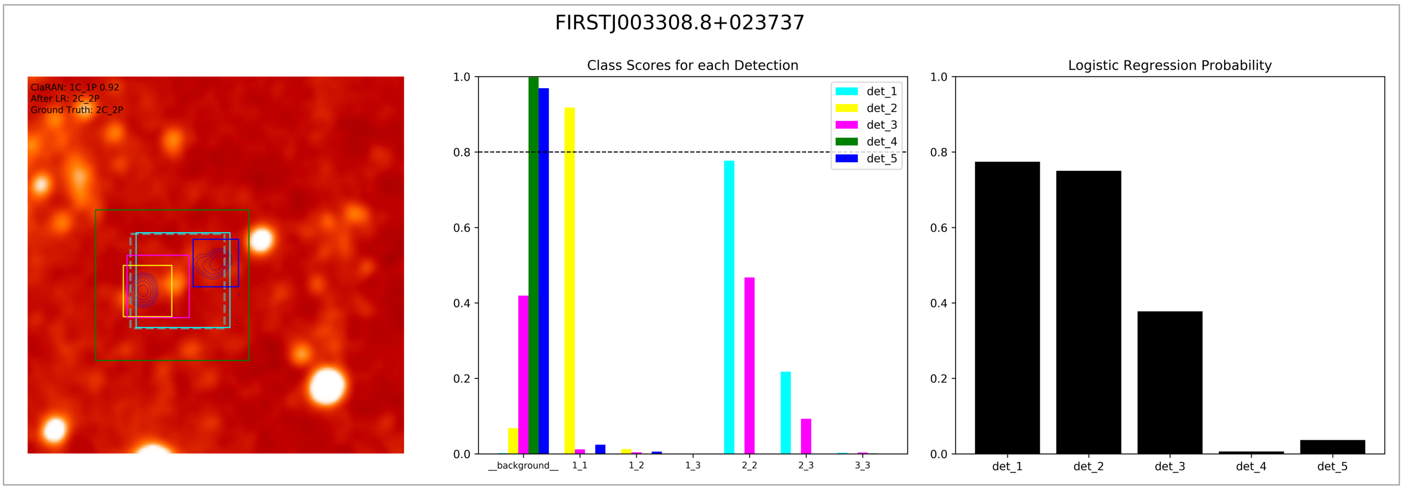


Figure 2 with Logistic Regression

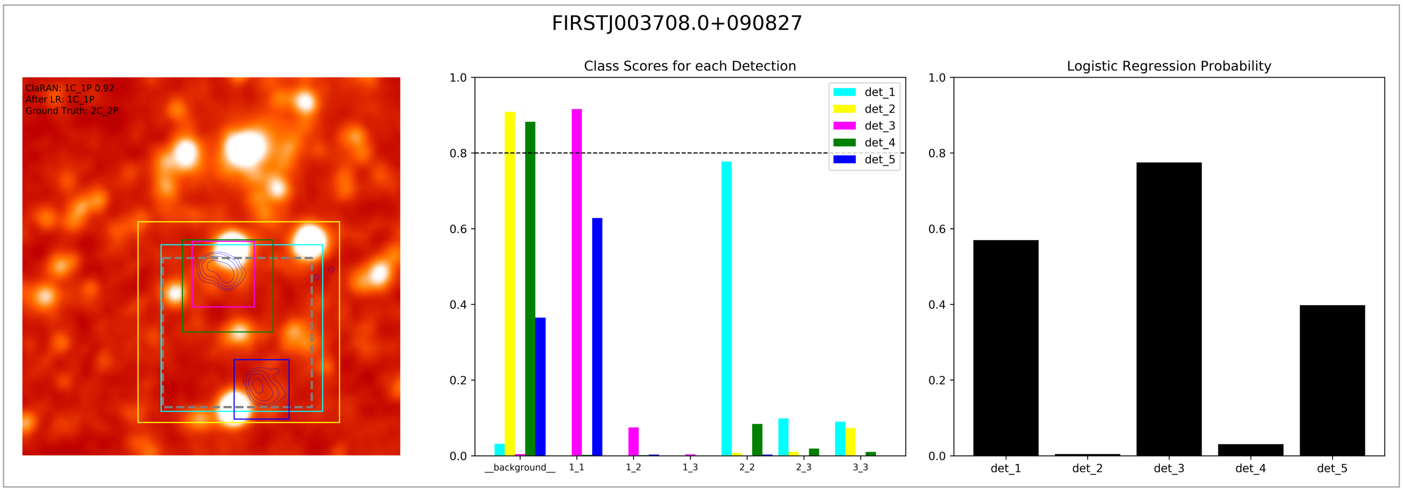


Figure 3 with Logistic Regression

Shown above, this method provided little improvement over using the arithmetic mean. It was successful for figure 2, but unsuccessful for figure 3. However there is the possibility for this method to be improved by including more parameters, especially if they include information about the other score distributions.

Increasing the Number of Detections

Given that statistical-based solutions are easier to implement, the limit on the number of detections was raised from 5 to 10, in order to observe whether trends would present themselves amongst a larger sample set. Taking the images from figure 2 and 3 again, below are the plots with the increased detections. Figure 2 has multiple 2C_2P detections, hence a statistical method would work here, but for figure 3 the added detections contribute little towards the correct classification. Ideally, the trend as observed for figure 2, would be present across input images to ClaRAN, hence it would be possible to filter out the outlier detections.

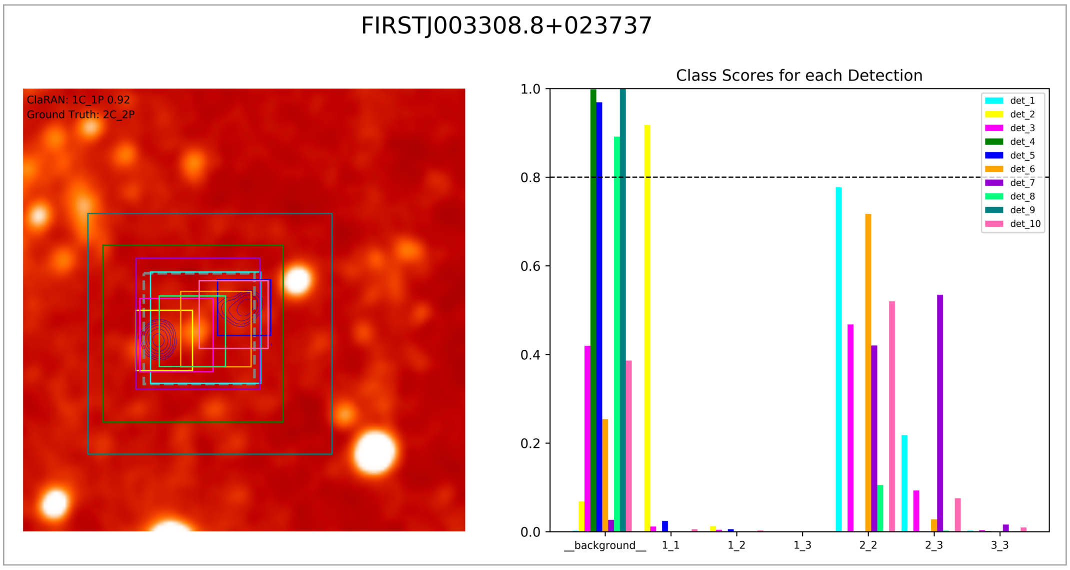


Figure 2 with Increased Detections

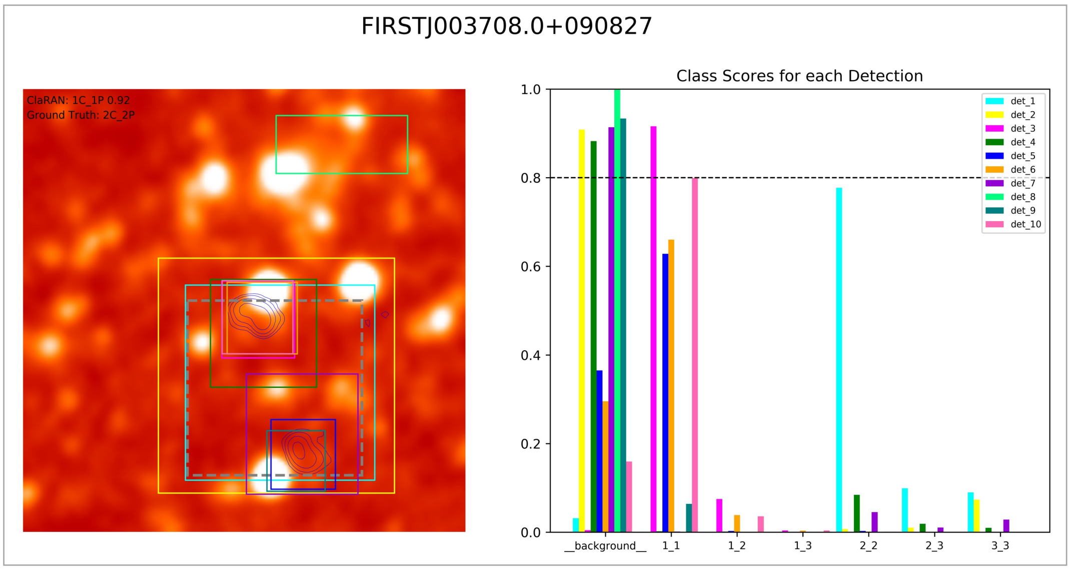


Figure 3 with Increased Detections

Conclusion

In conclusion, much has been learnt by visualising ClaRAN's score distribution as well as proposed bounding boxes. These plots have yielded insight in cases where misclassifications have occurred, and there appears to be promise in improving ClaRAN's performance by modifying how the SoftMax output is handled. Future developments could stem from a recent article, which highlights a unique method involving a per-class soft threshold output alongside the traditional SoftMax output, and then comparing to a pre-determined reference score calculated during training. This is in order to make safe and reliable decisions, and prevents cases where the network has high confidence in the wrong class. If ClaRAN is going to become part of an astronomers toolkit, it must be reliable and a solution such as this could let the user know when the network is uncertain. While the neural network is often seen as a black box, generating these visualisations has enabled unprecedented vision inside ClaRAN, and provides a clearer path for further development.