Applying machine learning methods in astronomy and data centres

Artificial Intelligence in Astronomy

ESO Garching, July 22-26, 2019

Teimoorinia, H., Kavelaars, J. J., Gwyn, S. D. J., Durand, D., Rolston, K., and Ouellette, A.

The Canadian Astronomy Data Centre (CADC)

NRC, Herzberg Astronomy and Astrophysics, Victoria, Canada

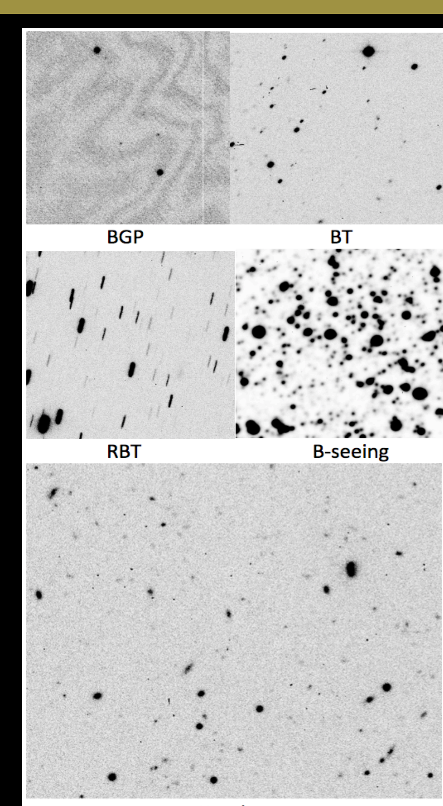
Abstract

We present a Machine Learning (ML) based approach for classifying astronomical images by data-quality via an examination of sources detected in the images and image pixel values from representative sources within those images. This approach uses a small fraction of the image pixels to determine the quality of the observation. The representative images (and associated tables) are ~800 times smaller than the original images, significantly reducing the time required to train our algorithm. The useful information in the images is preserved, permitting them to be classified in different categories, but the required storage is reduced. Using ground-based telescope imaging data, we demonstrate that the method can be used to separate 'usable' images from those that present some problems for scientific projects -- such as images that were taken in sub-optimal conditions. This method uses two different data sets as input to a deep model and provides better performance than if we only used the images' pixel information. The method may be used in cases where large and complex data sets should be examined using deep models. Our automated classification approach achieves 97% agreement when compared to classification generated via manual image inspection. We compare our method with traditional results and show that the method improves the results by about 10%, and also presents more comprehensive outcomes.

Data

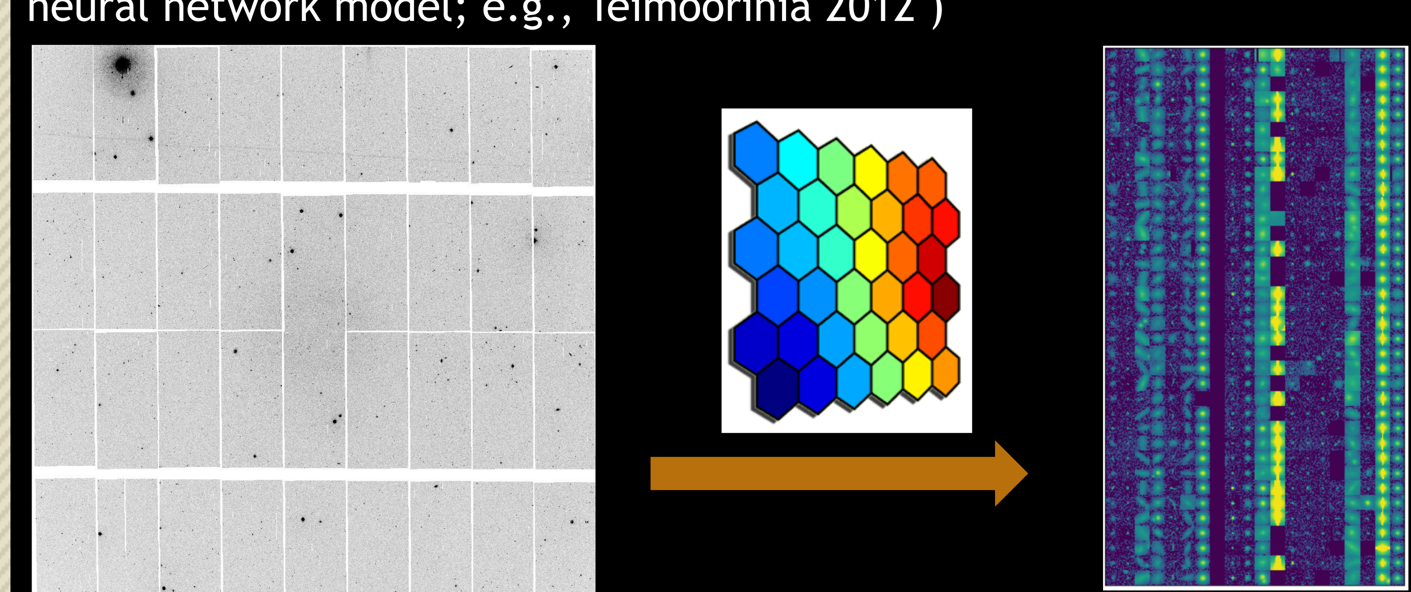
For this analysis, we explore our assessment method using selected images from the MegaCam instrument mounted on the Canada-France-Hawaii Telescope (CFHT) (Boulade et al. 2003).

The right plot shows an example of five different targets for our models used in this paper. They include images with different problems in the background (BGP), bad tracking (BT), really bad tracking (RBT), bad seeing or bad observational conditions (B-Seeing) and an instance of a Good image at the bottom of the figure.



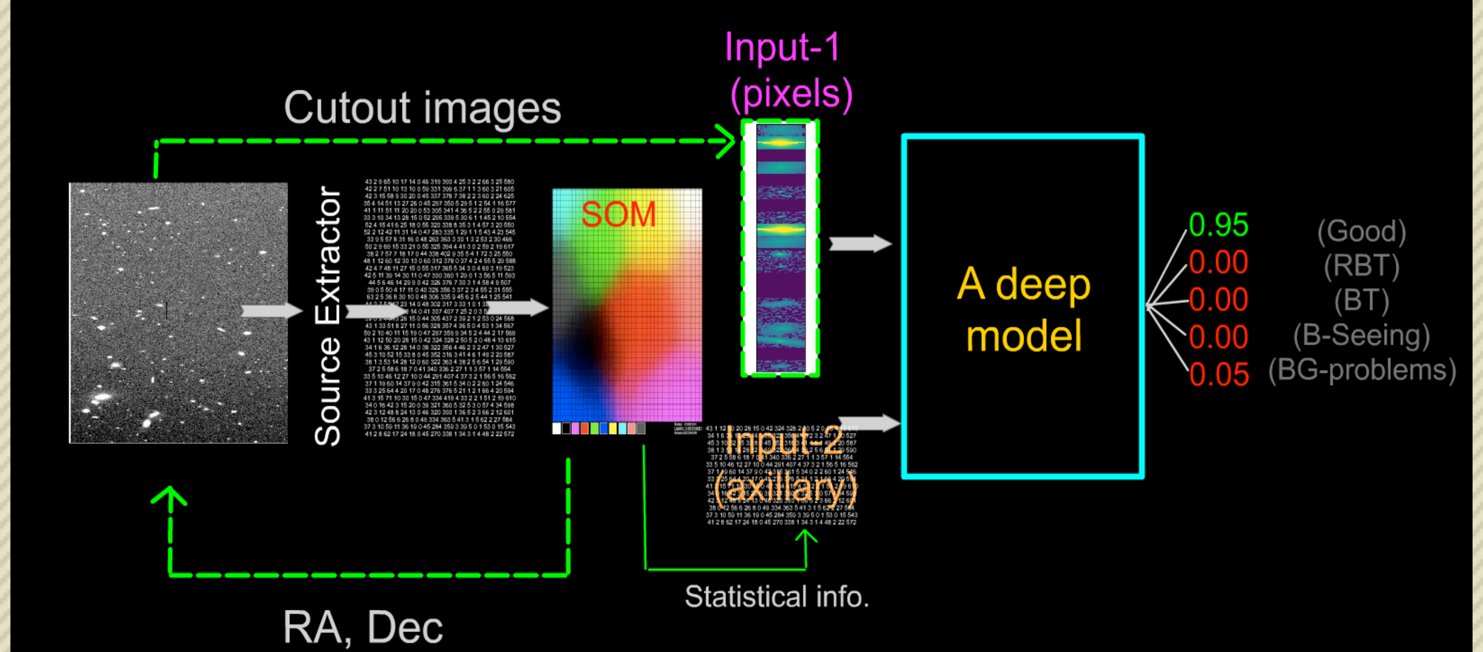
Reducing the size of images

A MegaCam image is huge (contains 36 big sub-images). We use Self-Organizing Maps (SOM; e.g, Rahmani et al. 2018) to create a representative image with a smaller size (more suitable to be fed to a neural network model; e.g., Teimoorinia 2012)



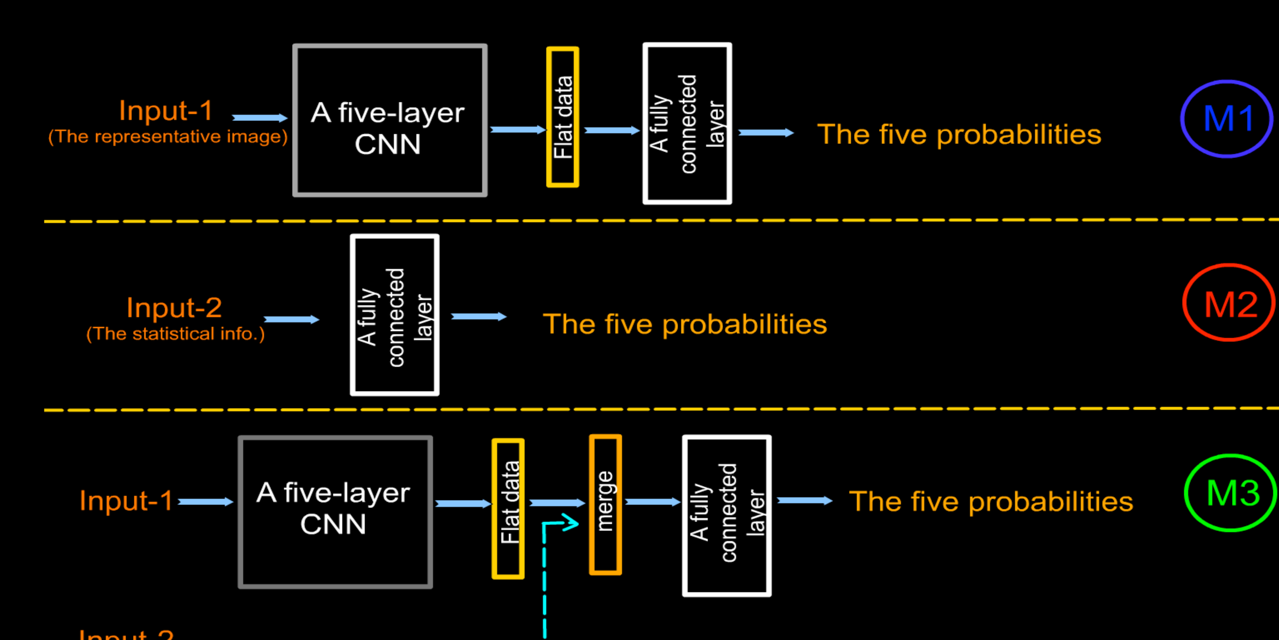
36 x 2048 x 4612 pixels → 36 x 20 x 625 pixels

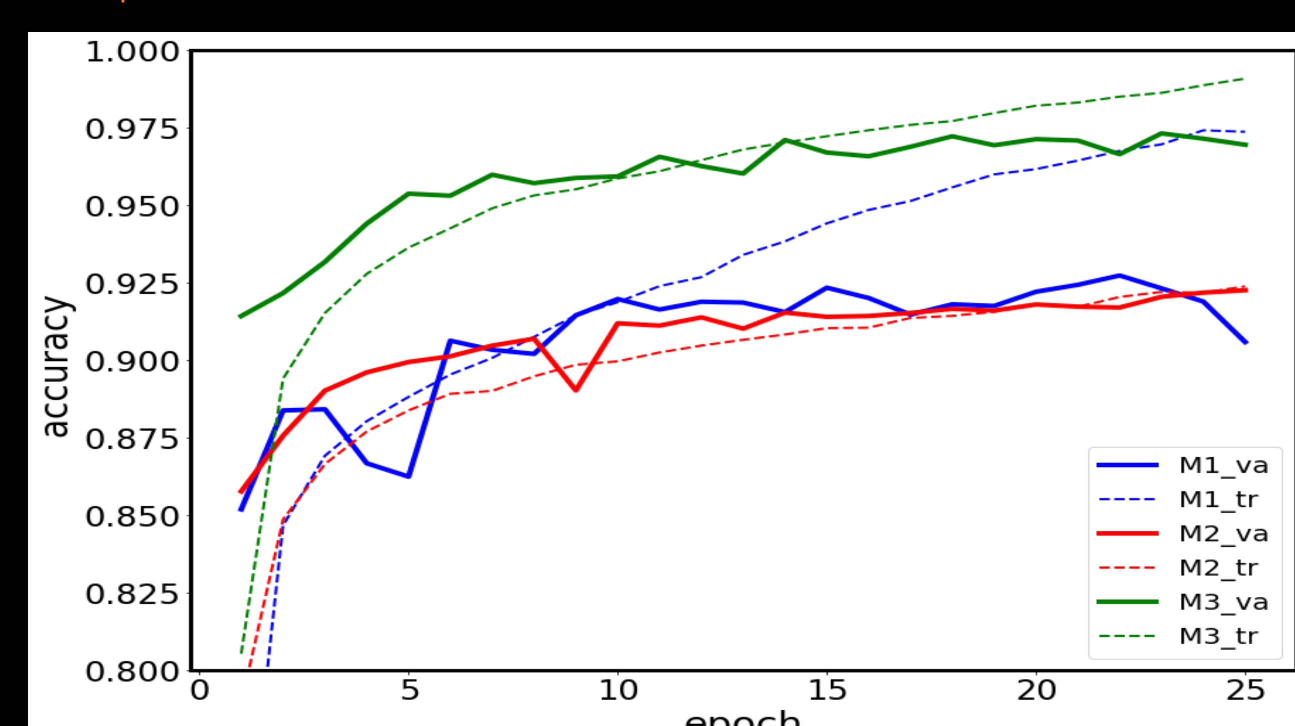
The pipeline used in this work



The plot is the combined model used in this paper. The selected parameter of detected sources from an image (the left image) can be extracted by Source Extractor (SE; Bertin & Arnouts 1996), and the table from SE is then fed to a trained SOM. The SOM provides a set of suitable RA and Dec, of the objects in the table, by clustering the information in the table. The CADC's cutout service cuts the object out of the images (the representative images), and they are sent to a deep model. The representative images (i.e., pixel information) are the primary input (Input-1). Besides, we can obtain statistical information from SOM (the number of similar objects in different clusters). That is more information we will provide to the deep model (Input-2). The five classes are the output of the last model. No direct information from SE is used in the deep model.

The deep models tested in the pipeline

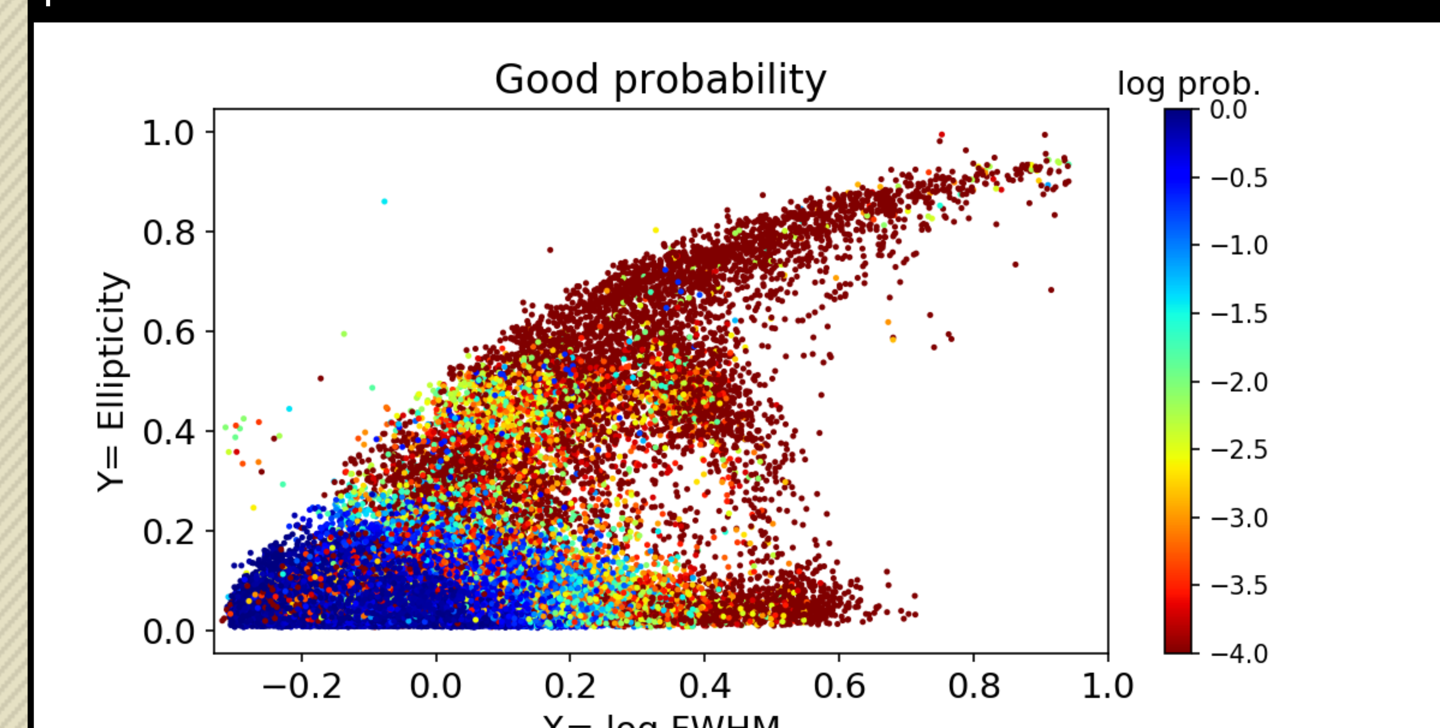




The top plot shows the detailed picture for the deep models, i.e., M1, M2, and M3. We can use three different inputs for the models. M1 indicates a case in which only the representative image (Input-1) is fed to the model. M2 shows a model in which the input is just statistical information from SOM (Input-2). M3 shows a combination of the two inputs. The performance of the three models is shown for both training and validation sets. The performance of M3 shows a significant improvement over the other two models, which use only one input.

Results

Using deep model M3, we classified over 220,000 exposures (more than~ 8,000,000 images) in less than one day of computation. At no time did the performance of the process decay due to fatigue. The following plot is an example that compares our results with traditional methods. To assess the quality of images, the classical methods use parameters such as ellipticity vs. Full-Width Half-Maximum. Each point shows the average of the two parameters measured for the point sources in an image by traditional methods. The colour-coded points show the probabilities that are predicted by our method. The region surrounded by ellipticity <0.2 and log(FWHM)<0.2 contains points that may be considered as 'good' images in traditional methods. However, there are more than 10% of images in this region that are predicted as unusable images (by our method). In visual inspections, they have, generally, problems in the background or have different weird patterns.



Summary

We present a method in which two groups of input data are fed to a deep neural network to classify complex, ground-based telescopic images. The method, significantly, improves the performance. As an example of a complex data set, we use CFHT MegaCam images to explore and demonstrate our approach. We have tested the different sets of exposures and have found that a decision boundary of Good=0.20 provides an accuracy of more than 97% (during a visual inspection). We have also compared our results with those of classical methods and found that our approach improves the outcomes and presents more comprehensive results.

References

-- Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393
-- Boulade O., Charlot X., Abbon P., et al., 2003, vol. 4841 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 72-81
-- Rahmani S., Teimoorinia H., Barmby P., 2018, MNRAS, 478, 4416
-- Teimoorinia H., 2012, AJ, 144, 172

The paper is submitted to MNRAS (in the title of 'Assessment of astronomical images using combined machine learning models')