SF2A 2018
P. Di Matteo, F. Billebaud, F. Herpin, N. Lagarde, J.-B. Marquette, A. Robin, O. Venot (eds)

IDENTIFYING CONTAMINANTS IN ASTRONOMICAL IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

M. Paillassa[1], E. Bertin[2] and H. Bouy[1]

Abstract. In this work, we propose to use convolutional neural networks to detect contaminants in astronomical images. Each contaminant is treated in a one vs all fashion. Once trained, our network is able to detect various contaminants such as cosmic rays, hot and bad pixel defaults, persistence effects, satellite trails or fringe patterns in images of various field properties. The convolutional neural network is performing semantic segmentation: it can output a probability map, assigning to each pixel its probability to belong to the contaminant or the background class. Training and testing data have been gathered from real or simulated data.

Keywords: convolutional neural networks, astronomical image analysis, astronomical image contaminants

1 Introduction

Many scientific results derived from astronomical images are obtained by analysing catalogues of objects that are extracted from those images. Thus, it is a matter of importance to have the most complete and less contaminated source catalogues. But this task is largely complicated by the numerous contaminants that pollute the images. For this reason, we aim to develop methods to identify these contaminants. Each survey pipeline incorporates prior knowledge about its instruments or external tools like LA Cosmic van Dokkum et al. (2012) to ignore contaminated pixels for further analysis. Here we would like to have a tool that is universal, e.g. that would not be tuned for a specific instrument or images. This is why we propose to address this problem using machine learning techniques, in particular through the task of semantic segmentation using supervised learning and convolutional neural networks.

In the following, we present the data we used to train our convolutional network. Then we describe its architecture and show some qualitative results.

2 Data

We chose to use real data as much as possible and take advantage of the private archive of wide-field images gathered for the COSMIC-DANCE survey Bouy et al. (2013). This library includes images from many past and present optical and near-infrared wide-field cameras, hence covering a broad range of detector types and sites. Plus, the COSMIC-DANCE pipeline detected most problematic images including tracking/guiding loss, defocused images or images strongly affected by fringes, providing a very valuable library of real problematic images for the analysis.

To build our training samples, our procedure has been to make sure to have clean images and to add contaminants in it so that we know exactly which pixels are affected by such contaminant. Examples of training samples can be seen in the two first columns in figure 2. The contaminants included in this study are: cosmic rays (red), hot columns (white), bad columns (yellow), bad lines (brown), hot pixels (blue), bad pixels (green), persistence effects (turquoise), satellite trails (orange) and fringe patterns (lighter gray). Plus, the brightest astronomical objects have been separated in an additional class (magenta). Black pixels are pixels that belong to several classes.

[1] Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France.
[2] Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France.

3 Architecture and qualitative results

The model used for the semantic segmentation is similar to SegNet (Badrinarayanan et al. 2017) and consists of two parts. The first part is made of convolutional layers followed by max-pooling downsampling. Indices of max-pooling are kept up and used in the second part which is made of upsampling and convolutional layers. All the convolutional layers are followed by Rectified linear units (ReLUs), except the last one that uses sigmoid to produce the probability maps for each class. The architecture is represented in Fig. 1. It was implemented using the TensorFlow library (Abadi et al. 2016).

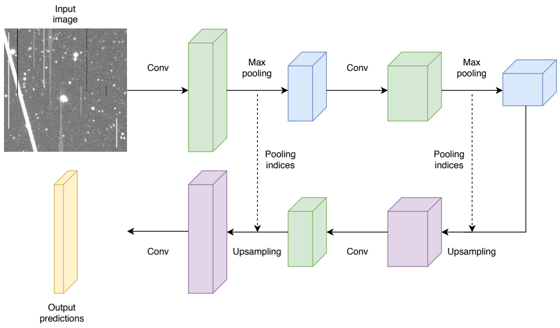

Fig. 1. Architecture of the neural network

The model is trained end-to-end using Adam optimizer and sigmoid cross entropy. The main problem encountered for training is the very strong class imbalance. To circumvent this, each pixel cost is weighted based on its class representation in the training set and those of its closest neighbors.

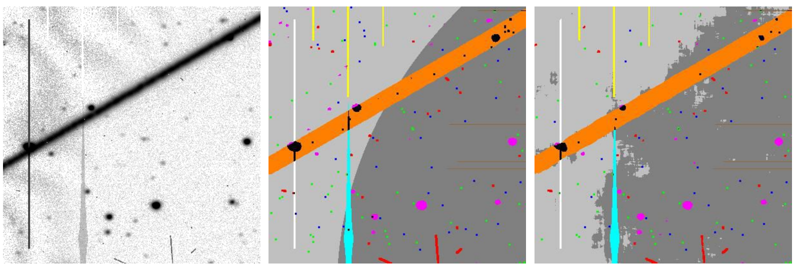

Fig. 2. Left: Input image. **Center:** Ground truth. **Right:** Prediction.

4 Conclusions

We show that we can train convolutional neural networks to identify astronomical contaminants in images. Further work would consist of detecting more contaminants (saturation patterns, reflections) or explore more ways to resolve the strong class imbalance that biases the training procedure.

This research has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 682903, P.I. H. Bouy), and from the French State in the framework of the "Investments for the future" Program, IdEx Bordeaux, reference ANR-10-IDEX-03-02. This research has also received funding from the French National Center for Space Studies (CNES).

References

Abadi, M., Barham, P., Chen, J., et al. 2016, in OSDI, Vol. 16, 265–283

Badrinarayanan, V., Kendall, A., & Cipolla, R. 2017, IEEE transactions on pattern analysis and machine intelligence, 39, 2481

Bouy, H., Bertin, E., Moraux, E., et al. 2013, Astronomy & Astrophysics, 554, A101

van Dokkum, P. G., Bloom, J., & Tewes, M. 2012, Astrophysics Source Code Library